





- Modelo de Redes de Flujo para la Evaluación de la Propuesta de Carrera Docente de Educación 2020. 5  
*Constanza Barrena, Valentina Quiroga, Mario Waissbluth, Andrés Weintraub*
- Programación del Fixture de la Segunda División del Fútbol de Chile mediante Investigación de Operaciones. 27  
*Guillermo Durán, Mario Guajardo, Rodrigo Wolf-Yadlin*
- Aplicación de Técnicas de Web Mining sobre los Datos Originados por Usuarios de Páginas Web. Visión Crítica desde las Garantías Fundamentales, especialmente la Libertad, la Privacidad y el Honor de las Personas. 47  
*Juan D. Velásquez, Lorena Donoso*
- Experiencias Prácticas en la Medición de Riesgo Crediticio de Microempresarios utilizando Modelos de Credit Scoring. 69  
*Cristián Bravo, Sebastián Maldonado, Richard Weber*
- Modelo de Planificación de Producción para un Sistema Multiproducto con Múltiples Líneas de Producción. 89  
*Rodrigo Viveros, Eduardo Salazar*
- Generación de Equipos de Trabajo Mediante Análisis de Redes Sociales e Identificación de Atributos Personales. 103  
*Francisco Molina, Pablo Loyola, Juan D. Velásquez*

R E V I S T A  
**INGENIERIA DE SISTEMAS**

ISSN 0716 - 1174

EDITOR

**Guillermo Durán**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

AYUDANTE DE EDICIÓN

**Ma. Fernanda Bravo**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

COMITÉ EDITORIAL

**René Caldentey**

*Universidad de Chile, Chile*

**Héctor Cancela**

*Universidad de la República, Uruguay*

**Rafael Epstein**

*Universidad de Chile, Chile*

**Luis Llanos**

*CMPC Celulosa, Chile*

**Javier Marengo**

*Universidad Nacional de  
General Sarmiento, Argentina*

**Juan de Dios Ortúzar**

*P. Universidad Católica, Chile*

**Víctor Parada**

*Universidad de Santiago, Chile*

**Oscar Porto**

*GAPSO, Brasil*

**Lorena Pradenas**

*Universidad de Concepción, Chile*

**Nicolás Stier**

*Columbia University, USA*

Financiado parcialmente por el Instituto Sistemas Complejos de Ingeniería, como reconocimiento a la difusión de las materias abordadas y de sus participantes.

Las opiniones y afirmaciones expuestas representan los puntos de vista de sus autores y no necesariamente coinciden con las del Departamento de Ingeniería Industrial de la Universidad de Chile.

Instrucciones a los autores:

Los autores deben enviar 2 copias del manuscrito que desean someter a referato a: Comité Editorial, Revista Ingeniería de Sistemas, Av. República 701, Santiago, Chile. Los manuscritos deben estar impresos en hojas tamaño carta, a doble espacio, deben incluir un resumen de no más de 150 palabras, y su extensión no debe exceder las 25 páginas. Detalles en [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)

Los artículos sólo pueden ser reproducidos previa autorización del Editor y de los autores.

Correo electrónico: [ris@dii.uchile.cl](mailto:ris@dii.uchile.cl)

Web URL: [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)

Representante legal: Máximo Bosch

Dirección: República 701, Santiago, Chile.

Diagramación: Ma. Fernanda Bravo

Impresión: Ka2 Diseño e Impresión

Mail: [contacto@ka2.cl](mailto:contacto@ka2.cl)

---

## Carta Editorial Volumen XXIV

---

Nos es muy grato presentar este nuevo número de la Revista de Ingeniería de Sistemas (RIS) dedicado a temas de frontera en Investigación de Operaciones, Gestión y Tecnología. Queremos agradecer al Instituto Sistemas Complejos de Ingeniería (ISCI) por su colaboración para hacer posible esta publicación.

Este número contiene artículos de académicos y estudiantes de nuestro Departamento de Ingeniería Industrial (algunos de ellos incluso son consecuencia de trabajos finales de grado, tesis de magister o tesis de doctorado), y de investigadores del ISCI. Contamos también en este número con un trabajo de académicos de la Universidad de Concepción.

Nuestro objetivo a través de esta publicación es contribuir a la generación y difusión de las tecnologías modernas de gestión y administración. La revista pretende destacar la importancia de generar conocimiento en tecnología y administración para nuestras problemáticas, junto con adaptar las tecnologías foráneas a las realidades en Chile y de otros de características similares.

Estamos seguros de que los artículos publicados en esta oportunidad muestran formas de trabajo innovadoras que serán de gran utilidad e inspiración para todos los lectores, ya sean académicos o profesionales, por lo que esperamos que esta iniciativa tenga la recepción que creemos se merece.

Guillermo Durán  
*Editor*



---

# MODELO DE REDES DE FLUJO PARA LA EVALUACIÓN DE LA PROPUESTA DE CARRERA DOCENTE DE EDUCACIÓN 2020

---

CONSTANZA BARRENA<sup>\*</sup>  
VALENTINA QUIROGA<sup>\*\*</sup>  
MARIO WAISSBLUTH<sup>\*\*</sup>  
ANDRÉS WEINTRAUB<sup>\*</sup>

## Resumen

*Educación 2020 nace a fines de 2008 como una manifestación desde la sociedad, encabezada por uno de los autores, Mario Waissbluth, para proponer acciones de mejoras en la educación y equidad de acceso en la misma. Educación 2020 presenta una propuesta completa para ser discutida en el plan de gobierno de largo plazo. Un elemento necesario para el análisis, es determinar las inversiones de cada una de las cuatro propuestas centrales planteadas; profesores de excelencia, directores de nivel internacional, condiciones adecuadas en aulas vulnerables y apoderados informados y participativos. En particular, se destaca la necesidad de establecer una capacitación y evaluación constante al cuerpo docente, cantidad de profesores jubilados anualmente, incentivos monetarios, demografía de la población escolar en los niveles de párvulo, básica y media, entre otros, dentro del marco de la formación y capacitación de profesores de excelencia, así como la formación de directivos de nivel internacional y apoderados participativos. El trabajo desarrolla dos modelos matemáticos para evaluar la mejora de la calidad docente, con el objetivo de alcanzar la formación de profesores de excelencia, detallando la transición desde un nivel de desempeño o estado a otro del profesorado, en conjunto con el costo anual de la propuesta. De esta forma, se permite la revisión de diferentes escenarios frente a restricciones presupuestarias por cada nivel educacional: Párvulo, Básica y Media. Permitiendo no sólo obtener los flujos anuales de cada ítem de la propuesta para cada año, sino también la trayectoria y flujo de la dinámica de la evolución del profesorado entre niveles de desempeño en el mismo periodo.*

**Palabras Clave:** *Formación de Educadores, Programación Lineal, Análisis de Costo.*

---

<sup>\*</sup>Departamento de Ingeniería Industrial, Universidad de Chile.

<sup>\*\*</sup>Departamento de Ingeniería Industrial, Universidad de Chile. Educación 2020.

---

## 1. Introducción

---

La educación es una variable que ha sido fuertemente criticada en el último tiempo, debido a su alta injerencia en el desarrollo y oportunidades a los que pueden acceder los chilenos. Es el sector municipal, el que se encuentra más cuestionado, como consecuencia de una gran variedad de incentivos que no apuntan a mejorar necesariamente el desempeño de sus profesores, sino más bien la antigüedad de los mismos y otros elementos de carácter administrativo.

Es dentro de este contexto que Educación 2020 presenta una propuesta, la que por primera vez proviene desde la ciudadanía hacia el gobierno. Un problema al que se enfrenta Educación 2020 es el establecer en mayor detalle la inversión requerida para cada uno de sus pilares y en particular de las propuestas que los componen. Es decir, pese a que se han establecido valores tentativos, éstos sólo responden a órdenes de magnitud y reflejan una idea somera de lo necesario, más bien para fines de ejemplificación. Bajo este escenario el lograr esclarecer el detalle de la inversión es un punto vital, para tener un proyecto sólido y bien fundamentado, que sea de consideración para su integración al presupuesto de gobierno para el MINEDUC en corto y largo plazo. Por tanto, cabe destacar que el ámbito de mejora de la calidad docente represente el mayor volumen de la inversión necesaria, siendo por su parte, el eje de la propuesta.

En este trabajo se desarrolla la evaluación de la inversión asociada al pilar de “Profesores de Excelencia”, reflejando a su vez de manera simple, los flujos que muestran cómo van evolucionando en el tiempo, desde un nivel de desempeño al siguiente, los docentes que se desempeñan en el sector estatal. En particular se hace énfasis en la formación y perfeccionamiento del profesorado una vez ingresados al sistema educacional, evaluaciones, mejoras salariales progresivas, paquetes por desvinculación y el logro de una calidad en docencia de estándares internacionales a través de capacitaciones e incentivos a la superación constante, para el período 2010-2020. La trascendencia de resolver lo anterior, radica en que según las estimaciones establecidas por Educación 2020, es precisamente este factor el que representa el mayor monto de inversión, dado que es imperante mejorar la precaria calidad de los docentes actuales, quienes continuarán educando a los alumnos año a año, mientras en forma paralela se progresa en la formación de las nuevas generaciones de profesores. A su vez, diversos estudios internacionales, confirman que es precisamente el profesor y su desempeño en el aula, quien hace la mayor diferencia en las oportunidades a las que pueden acceder luego un niño, por lo que es de suma importancia abordar esta área.

Para lograr lo anterior, se crearon dos modelos; uno tiene por objetivo el

cálculo de la demanda anual de docentes de Párvulo, Educación Básica y Media por separado y en detalle a nivel comunal para las 346 comunas de Chile. El segundo, el más importante, toma esta demanda por cada nivel educacional como un input y entrega para cada año estudiado, el detalle del flujo de costos y dinámica de transiciones de mejora de desempeño docente. Es importante destacar, que además es posible modelar las salidas por retiros de jubilación, voluntario y desvinculación por calificación insatisfactoria, los que son puntos centrales de la propuesta. De esta forma, este segundo modelo es un reflejo de la propuesta de una renovada Carrera Docente, con foco en el desempeño de profesores ligado a los resultados de sus alumnos.

Los supuestos incorporados, reflejan los objetivos establecidos en la propuesta de Educación 2020 para lograr profesores de excelencia, así como factores políticos y contractuales que deben ser considerados. Para el caso del cálculo de demanda de docentes, no se incorporó el factor geográfico por simplicidad. Pese a esto, se considera que un docente debe desempeñarse 44 hrs. a la semana en su cargo, considerando una jornada completa, cumpliendo una proporción entre horas lectivas y no lectivas de 75/25, lo cual es un parámetro que es variado año a año para lograr un objetivo de 60/40 al 2020. También se supone que el promedio de niños por sala es de 35 alumnos, siendo el IVE (Índice de Vulnerabilidad Escolar) un índice que permite determinar cuantos de las matrículas por comunas responden a la clasificación de niño vulnerable, manteniéndose relativamente constante desde el 2010 al 2020. Este último es un parámetro que permite establecer al 2020 objetivos de reducción de niños vulnerables por sala.

Por su parte, el modelo matemático de Carrera Docente considera la evolución entre niveles de ejercicio que puede tener un docente, si es capacitado o no, teniendo por máximo un escalón por año. Lo anterior, se ve restringido para los dos menores niveles de desempeño, ya que por lo deficiente de sus conocimientos en ese momento, se requiere de un tiempo de al menos 2 años para pasar al siguiente nivel. Además, se considera mantener la evaluación docente que califica al 25 % del total de profesores en cada ocasión, agregando una prueba de habilitación que restringe el ingreso al sistema municipal, para mejorar la calidad de los educadores que ejercen en los municipios. También se tiene por parámetros un objetivo lineal de calidad docente, que al 2020 alcanzaría un 45 % del profesorado en nivel Experto, un 40 % como Intermedio y el restante 15 % entre Fundamentales e Insatisfactorios, detrás de una inversión en capacitación e incentivos monetarios ligados al desempeño. Finalmente, y según la necesidad de entregar una mejora en la carrera docente en cuanto las desvinculaciones, se agregan bonos por jubilación que actúan como un soporte para la salida de aquellos, que pese a no ejercer de manera óptima no pueden renunciar por no tener una jubilación suficiente.

---

## 2. Descripción del Problema

---

La educación de los chilenos es un problema complejo que se ha tratado en variadas oportunidades, pero pese a los esfuerzos realizados, aún no se ha logrado asegurar la equidad en la calidad de la misma. Este punto es crítico y debe solucionarse con un plan de largo plazo integrado a programas de gobierno. Existen estadísticas alarmantes en este sentido, de las cuales hay que hacerse cargo como país:

- El 6 % de los alumnos provenientes de las familias con más recursos no superan el promedio general internacional [1].
- 2 de cada 5 escolares terminan cuarto medio sin comprender lo que leen [1].
- La brecha entre el 20 % mejor y peor resultado del SIMCE se incrementó en 13 puntos en lenguaje y 16 en matemáticas entre 1996 y 2006 [1].
- Cerca del 36 % de los estudiantes que ingresan a estudiar pedagogía no saben leer un gráfico. Siendo casi el 35 % de los egresados aún incapaces de hacerlo [1].
- El gasto público en Chile es cerca de un tercio de la de países industrializados en educación básica y llega a ser un cuarto para educación secundaria [2]

En este marco es que Educación 2020 establece una propuesta basada en cuatro pilares; profesores de excelencia, directores de nivel internacional, condiciones adecuadas en aulas vulnerables y apoderados informados y participativos, considerando el aumento de la proporción de las horas no lectivas frente a las lectivas para dar a los docentes el tiempo necesario de preparar clases, atender a sus alumnos y corregir pruebas. Si bien el problema afecta a todo el sistema educacional, el foco se encuentra en los colegios municipalidades, los cuales son los más afectados en la actualidad en cuanto al desempeño de sus profesores. De los casi 3,5 millones de niños y jóvenes en el sistema escolar, cerca del 40 % pertenecen al sistema municipal como lo muestra la Figura 1.

Este trabajo se centra en el primero de estos puntos, creando una herramienta a través de un modelo de redes que evalúa el costo anual y total del sistema de capacitación, remuneración, evaluación y desvinculación (ya sea por motivos voluntarios, jubilación o desempeño), durante el período 2010 al 2020 para profesores del sector municipal, respondiendo a las exigencias de calidad establecidas. De esta forma no sólo es posible conocer los costos anuales,

### Distribución de matrículas

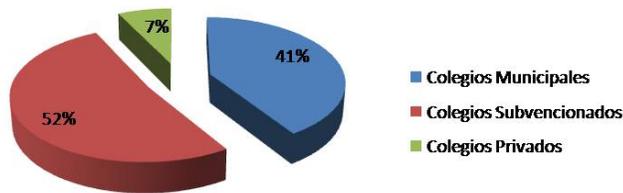


Figura 1: Distribución de matrículas hasta 2010 según número de alumnos.

sino que a su vez la dinámica de transición de los profesores desde un nivel de desempeño a otro.

Para apoyar el modelo de redes, se desarrolló un modelo de programación lineal que calcula la cantidad de profesores necesarios año a año según la demografía proyectada del país, considerando una proporción de horas lectivas vs no lectivas de un 60 % y 40 % respectivamente, sumado a la reducción de niños vulnerables por sala.

La información necesaria para plantear los escenarios discutidos, corresponden a datos obtenidos principalmente de 2008. Si bien es muy factible que estos datos puedan variar, el resultado sigue siendo significativo y referente a la realidad que se desea plasmar. El problema principal radica en que la regulación del estatuto docente vigente, permiten que profesores con menor desempeño en la evaluación docente, logren incluso percibir un salario mayor que aquellos catalogados como destacados. De esta forma, no sólo no se retribuye la excelencia ni se fomenta la constante formación de los educadores, sino que de desincentiva al ingreso de profesores con altas habilidades. La influencia del propio trabajo en el aula de un profesor es tan decisiva para los alumnos, como es mostrado en la Figura 2, por lo que resulta ser un debate altamente relevante. Si se tiene a dos alumnos de 8 años con desempeño de un percentil 50, es decir medio, pero uno de ellos es educado en lo que sigue por un profesor de alto rendimiento, mientras el otro en paralelo tiene por educador a uno de bajo rendimiento, al cabo de 3 años las diferencias son abismantes. El alumno que ha recibido las enseñanzas del profesor de mejor desempeño logra superarse y llegar al percentil 90, mientras su compañero empeora aún más su rendimiento, bajando del percentil 50 al 37 en igual tiempo.

Este trabajo considera como se mencionó en líneas previas, el área de formación y perfeccionamiento de profesores una vez ingresados al sistema educacional, evaluación, mejoras salariales progresivas, paquetes por desvinculación y el logro de una calidad en docencia de estándares internacionales a través de capacitaciones e incentivos a la superación constante, para el período 2010-

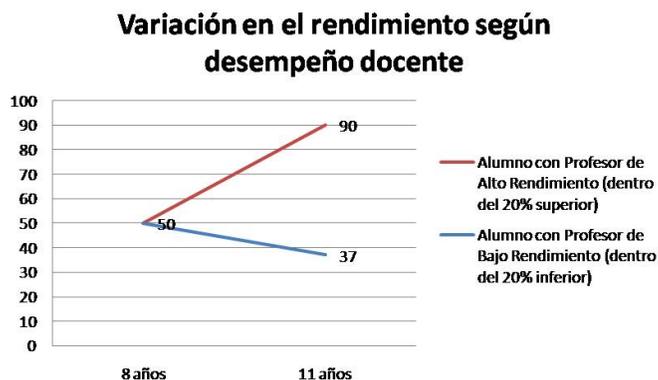


Figura 2: Influencia del desempeño docente en los resultados de sus alumnos. Fuente: Sander & Rivers Cumulative and Residual Effects on Future Student Academic Achievement, McKinsey.

2020. La trascendencia de resolver lo anterior, radica en que según las estimaciones establecidas por Educación 2020, es precisamente este factor el que representa el mayor monto de inversión dentro de su portafolio de proyectos, dado que es imperante mejorar la calidad de los docentes actuales, quienes continuarán educando a miles de chilenos año a año, mientras en forma paralela se progresa en la formación de las nuevas generaciones de profesores. Particularmente en este rubro, no se ha confeccionado un modelo similar hasta el momento, por lo que la herramienta entregada pretende ser un primer paso a un área de investigación y desarrollo en el que se potencie el soporte de metodologías y modelamiento para la toma de decisiones.

El modelo de redes que refleja la propuesta de Carrera Docente ataca el problema anterior, siendo capaz de entregar el detalle del flujo de transición en el nivel de desempeño de los docentes, en conjunto con los flujos de inversión necesarios para lograr los objetivos de calidad, remuneraciones e incentivos planteados al 2020 Por su parte, un modelo de programación lineal simple permite el cálculo período tras período de la cantidad de profesores requeridos según la demografía de la población y el progresivo avance hacia una proporción 60/40 entre horas lectivas y no lectivas, esto a través del período de 11 años que comprende el proyecto. Esto permite entregar las demandas totales de profesores a cumplir al modelo principal como uno de las restricciones angulares.

Para encontrar la solución al problema se desarrolló una metodología apoyada en la programación de una herramienta que permita introducir los insumos necesarios de manera amigable y fácil, entregando las tablas de datos en un formato que permita utilizarse directamente a través de AMPL-CPLEX para la solución y optimización de la evaluación de la propuesta de mejora docente en el largo plazo. Esta herramienta corresponde a una planilla Excel

que confecciona al ingresar los datos determinados la estructura de datos para el modelo. Lo mismo se realizó para el modelo de demanda de profesores. De esta forma, luego del levantamiento de información, se calibraron los parámetros definidos en cada modelo, los que muestran los elementos claves de la propuesta de Carrera Docente. Entre estos se destacan, las proyecciones de sueldo máximo y mínimo actuales y proyectados al año 2020, bonos de salida, jubilaciones, capacitaciones y evaluaciones, incluyendo la prueba de habilitación docente que serviría de filtro de entrada al sistema y la ya recurrente evaluación docente. En cuanto al cálculo de la demanda anual, los escenarios dependen principalmente de la cantidad de alumnos vulnerables por curso y proporción de horas lectivas y no lectivas. Esta parametrización puede realizarse a nivel de educación de Párvulo, Básica y media, flexibilizando aún más la toma de decisiones.

El estudio entrega una solución para cada uno de los 11 periodos pertenecientes al tramo 2010 -2020 en cuanto al costo anual de la propuesta y la dinámica de transición y salida entre un estado y otro, en conjunto con el posible déficit de alumnos a escuelas pedagógicas que deben ingresar al sistema para cumplir con los objetivos de calidad planteados año a año. En cuanto a la demanda por cada nivel educacional, el detalle obtenido es a nivel comunal, para las 346 comunas del país por cada año estudiado.

---

### 3. Enfoque del Modelo de Cálculo de Demanda

---

Como parte de la propuesta de Educación 2020, se encuentra la importancia de la proporción docente/alumno. Con especial énfasis en los alumnos vulnerables, reduciendo la cantidad de los mismos desde 35 a 25 por sala al 2020. Junto a esto último, se suma la necesidad de entregar a los docentes condiciones adecuadas para la preparación de sus clases. Es por esto, que para todos los profesores del sistema municipal, sin diferenciar según la clasificación de sus alumnos, se tiene por objetivo que al final del ciclo ya descrito, la proporción horas lectivas/no lectivas pase del actual 75/25 a un 60/40 como lo indican los estándares internacionales. El problema radica entonces, en la inexistencia de una herramienta o datos que reflejen diferentes escenarios que reflejen lo anterior, permitiendo determinar la cantidad de docentes que por ende se requerirán año a año en cada caso. Esto es relevante, ya que tiene directa influencia en los costos asociados a la propuesta. Por ejemplo, si al reducir la proporción de horas lectivas/no lectivas desde 75/25 a 60/40 el 2020, se tiene un aumento en la demanda de docentes, se tendrá entonces un incremento importante en la inversión, para cumplir la promesa de remuneraciones que es parte de la carrera docente. Esto, se suma a la necesidad de revisar la factibilidad de la propuesta en cuanto al número de docentes que ingresan año a

año al sistema vs los que salen del mismo.

El presente modelo, considera una reducción gradual del actual número de niños vulnerables por sala, así como de horas lectivas/ no lectivas año a año. De esta forma, se calcula la cantidad de salas que deben recibir educación municipal por comuna a nivel nacional desde el 2010 al 2020, siguiendo la proyección de población entregada por el INE. Con esta base, y considerando la jornada laboral descrita en el portal Docentesmás se procede a calcular los requerimientos de profesionales en cada uno de los tres niveles estudiados; Párvulo, Básica y Media, por comuna. Como supuestos se tiene que:

- Se establece una meta lineal al 2020 de tener 25 niños vulnerables por sala.
- Al 2020 se mantendrán los 35 niños por sala que actualmente hay, para aquellos no vulnerables.
- Se establece proyección lineal al 2020 de lograr proporción de hrs lectivas/ no lectivas de un 60/40.
- Actualmente la proporción es de 75 / 25.
- Datos de proyección por edad entregados por el INE al 2020.
- Jornada Completa para los docentes.
- Educación Básica y Pre Básica cumplen 38 hrs pedagógicas/ semana
- Jornada Laboral total 44 hrs.
- Se entregarán resultados por comuna, por grupo:
  - Media: 15-18 años.
  - Básica: 7-14 años.
  - Párvulo: 4-6 años.

Para su realización, se creó una interfaz en Excel debido a su familiaridad y facilidad de uso. Ésta genera tanto la estructura de datos necesaria, como el vínculo y entrega de datos a través del solucionador CPLEX. Se eligió este solucionador dada su facilidad para el manejo de grandes volúmenes de datos.

### 3.1. Modelo de Programación Lineal

Para obtener la demanda por docentes por comuna y año, se desarrolló un modelo simple de programación lineal, con el fin de poder hacer un análisis sencillo del problema. Dadas Restricciones sobre requerimientos de demanda de alumnos vulnerables y no vulnerables, el modelo determina el número de

profesores requeridos por comuna a nivel de párvulo, educación básica y educación media. Este modelo permite determinar los requerimientos de profesores por año y comuna entre el 2010 y el 2020 (ver detalles en [3]).

Considerando los datos de horas electivas, jornada laboral y horas pedagógicas a cumplirse por nivel educacional según lo establecido en la Jornada Completa a ser completada el 2010, se obtiene el número de profesores necesarios por sala para cada nivel, lo que multiplicado por el número de salas totales (de niños vulnerables y no vulnerables) de cada comuna, define demanda de total profesores por comuna.

Para el cálculo de salas por comuna año a año, se recogieron de la base de datos de la JUNAEB, el Índice de Vulnerabilidad de cada una de las 346 comunas del país. A su vez, se obtuvo el porcentaje de matrículas que representa el sistema municipal sobre el total en el que es calculado el IVE (establecimientos particular subvencionados más municipales). De esta forma, se calcula para cada nivel educacional, la cantidad de matrículas municipales, sobre ello, cuántas corresponden a niños vulnerables y cuantos no lo son. Después, se toman los objetivos de niños por sala para ambas condiciones y se divide el resultado anterior por este. Es así como se tiene el total de salas a ser cubiertas, lo que se repite hasta el 2020, para cada uno de los tres niveles educacionales ya descritos. Es importante destacar que la proporción de matrículas municipales y el IVE se suponen constantes, ya que no han mostrado variación importante en los últimos 3 años, junto con la inexistencia de estudios que permitan tener una estimación más exacta para el período 2010-2020.

### 3.2. Resultados

El modelo de demanda de profesores, calcula para cada nivel educacional y año perteneciente al período 2010-2020, la necesidad mínima de docentes. Para ello, asume un promedio de niños por sala, una proporción de horas lectivas vs no lectivas, cantidad y características de de horas pedagógicas (de 45 minutos) según sea Párvulo, Educación Básica o Media, todos parámetros los cuales pueden ser variados. Se define como caso base, aquel en que se consideran 44 horas semanales como jornada laboral, y horas pedagógicas de 45 minutos para el cumplimiento de la jornada escolar completa en cada nivel de escolaridad, utilizando sólo la proyección demográfica entregada por el INE. De esta forma, se obtiene que en el año 2020 se proyecta tener una reducción en las matrículas total cercana a 237.000 alumnos (Figura 3) como consecuencia del cambio en la demografía del país. Lo anterior se traduce en una reducción en la demanda de 5410 docentes para el 2020, en la Figura 4, se muestra el requerimiento de docentes para distintos escenarios de atención de alumnos.

Así se ve que si se conserva la estructura actual, se produce un reducción pequeña de requerimientos de profesores, producto de la reducción demográfi-

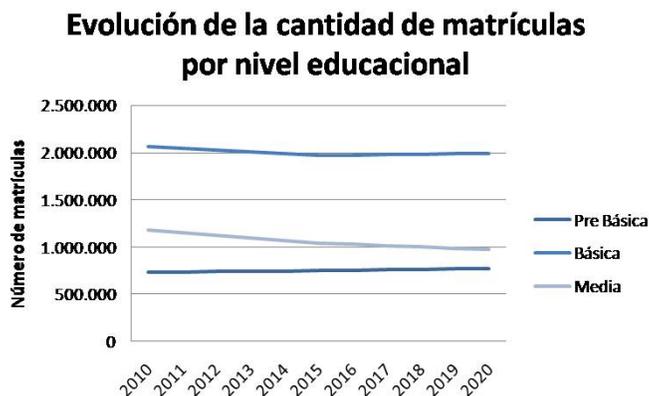


Figura 3: Variación de matrículas Párvulo, Básica y Media para el total de los sectores Privado, Municipal y Particular Subvencionado.

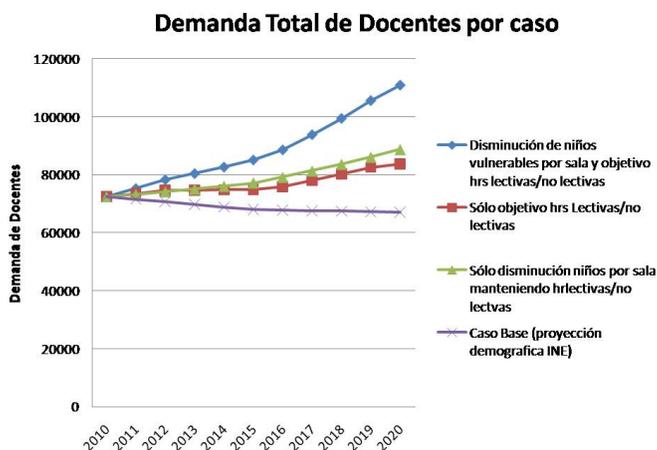


Figura 4: Estudio de demanda total de docentes por caso según parametrización.

ca de alumnos. Al aplicarse mejoras de atención, aumenta el requerimiento de profesores. La reducción de alumnos por clase permite pasar a 60/40 la programación de horas lectivas a no lectivas y las dos medidas simultáneamente.

Se encontró que el determinante principal en la variación de la demanda, para cada uno de los niveles estudiados, es la disminución de niños por sala (refiriéndose de manera única a quienes pertenecen a la clasificación de vulnerables). También se observó que la mayor variación en volumen de matrículas viene dado por aquellas del sector de Media, lo que concuerda con la evolución de las edades de la población.

Es importante recordar que todos los resultados del modelo son entregados no sólo de manera anual, sino que a su vez por comuna a nivel nacional. De esta forma, es posible observar por ejemplo los casos de 3 de las comunas más importantes de nuestro país Antofagasta, Concepción y Santiago las cuales

presentan un alto volumen de estudiantes, así como un porcentaje no menor de niños vulnerables dentro del sistema educacional municipal (68 %, 73 % y 67 % respectivamente), el cuál por su parte representan más del 52 % de los colegios en estas localidades, llegando incluso a un 67.7 % para Antofagasta. Maipú por su parte, es la comuna con mayor cantidad de estudiantes en el país, llegando a 36580 escolares de Párvulo, 99348 en Educación Básica y 56627 de Enseñanza Media. Pese a esto, es interesante ver que la proporción de establecimientos municipales sólo llega a un 23.9 % con un índice de vulnerabilidad de 63 %. Otro es el de San Esteban, comuna en que la totalidad de las instituciones pertenecen a la educación pública, con un alto IVE de un 90 %. Dicho lo anterior, si se analiza el caso base mostrado en la Figura 5, es posible destacar que si bien en su mayoría hay un descenso en la cantidad de profesores requeridos a lo largo del tiempo, para Maipú, el efecto contrario se debe a que según las estimaciones del INE la población entre 4 y 18 años crecerá, al contrario de lo que ocurre en las demás comunas.

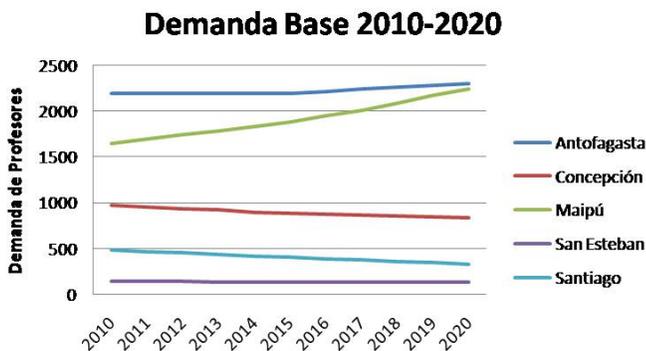


Figura 5: Análisis de demanda total para el caso base. Ejemplo para comunas: Antofagasta, Concepción, Maipú, San Esteban y Santiago.

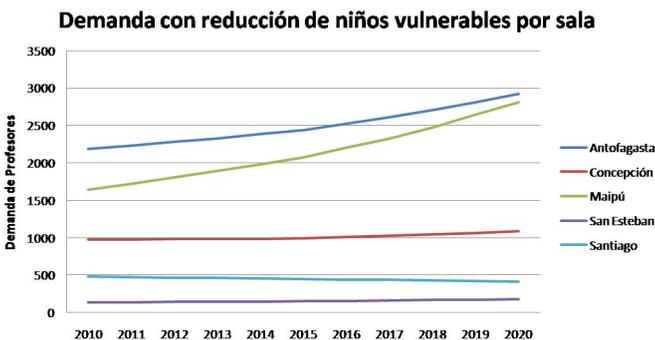


Figura 6: Análisis de demanda total bajo el efecto de la reducción de niños vulnerables por sala. Ejemplo para comunas: Antofagasta, Concepción, Maipú, San Esteban y Santiago.

Es interesante destacar que en aquellas comunas donde se combina una alta responsabilidad de la entidad pública sumada a un índice de vulnerabilidad considerable, como es el caso de Antofagasta, Concepción y Santiago, el efecto obtenido en el cambio de tendencia es mucho mayor, que lo que ocurre por ejemplo con Maipú, localidad en la cual la presencia estatal es menor. Para San Esteban, las consecuencias son proporcionalmente más importantes, debido a que el peso de la vulnerabilidad y la dependencia de la alcaldía son de gran importancia.

El modelo de demanda si bien es bastante simple, representa un buen complemento para el modelo de carrera docente, logrando una proyección del requerimiento de profesores por cada nivel educacional e incluso comunal. Esta información, si bien puede ajustarse en futuros trabajos con factores demográficos, los resultados del modelo principal no debieran variar de manera significativa, al agrupar la demanda en los tres estratos de educación.

---

## 4. Enfoque del Modelo de Cálculo de Carrera Docente

---

El modelo de capacitación y formación docente, refleja el flujo de profesores (separados por nivel de educación Párvulo, Básica y Media) desde un nivel de desarrollo y otro a través del tiempo que comprende el período 2010-2020. Este flujo entre estados corresponde al reflejo de la propuesta de carrera docente ya mencionada anteriormente, que incluye tanto los resultados de la Evaluación Docente (que cataloga a los docentes como Destacados, Competentes, Básicos e Insatisfactorios según su nivel de manejo en los 20 criterios que conforman los cuatro dominios del Marco para la Buena Enseñanza), junto con el propio desarrollo a través de post-títulos y responsabilidades adquiridas. Se encuentra caracterizado entonces, por arcos que unen parejas de nodos, los que a su vez representan cada uno de los niveles de desempeño posible que puede tener un profesor a lo largo de su proyección. Cada nodo está identificado con un índice mostrando el año al que pertenece y nombre del nivel; Experto, Intermedio, Fundamental, Insatisfactorio, Principiante de Primer y Segundo año de desempeño. De esta forma, la unidad básica del problema es un nodo, que se encuentra completamente caracterizado. Los arcos de transición entre un estado de perfeccionamiento y otro, muestran las inversiones anuales necesarias para lograr el objetivo de obtener una educación de alta calidad, centrada en profesores de nivel avanzado al año 2020.

El modelo muestra cuantos profesores cambian a cada estado, minimizando el costo total de la propuesta desde el año 2010 al 2020, considerando como objetivo el cumplimiento de la excelencia en la calidad docente al 2020, así como las metas parciales anuales en cuanto a la cantidad de docentes en cada nivel

para cada uno de los períodos. Todo lo anterior lo realiza considerando las estadísticas y costos de salida por retiros por jubilación, voluntarios e involuntarios, siendo estos últimos aquellos en los que quienes al recibir por segunda vez una calificación como Insatisfactorios deben abandonar el sistema. Cabe destacar, que un punto fundamental de la propuesta es la mejora salarial de los profesionales a medida que vayan subiendo de nivel. Es por esto que se establece una mejora lineal en las remuneraciones para cada uno de los estratos de la red, con el objetivo de llegar a un tope cercano a los \$ 1.500.000 para el nivel experto y un mínimo superior a los \$ 600.000 para los Principiantes e insatisfactorios en 2020.

Estos valores se encuentran aún en discusión, por lo que se puede realizar cálculos de sensibilidad para posibles variaciones. Una condición básica establecida es la formación de docentes de alta calidad, los cuales entrarán a cursar una carrera de alto nivel. La propuesta implica que incentivados por los mejores ingresos y condiciones, se confirmará un ingreso de mejores alumnos. De esta manera, a partir del 2016 ingresarían al sistema profesores de buen nivel. La red ya descrita se duplicará, para que a partir del año 2016 (en promedio una carrera de educación tarda 6 años en ser completada), un grupos de profesionales con un alto nivel de formación entre al mercado. El modelo será cargado con datos correspondientes a cada nivel de educación (Párvulo, Básica y Media), para poder obtener los costos de la propuesta de la institución en cada uno de ellos. Por lo tanto se tendrán tres resultados con las características previas para los casos de Párvulo, Básica y Media. Por último, es necesario hacer notar que, como punto de partida para 2010 del modelo de capacitación y formación docente, establece supuestos en relación a los niveles de la Carrera Docente.

- Todos los profesores en nivel Insatisfactorio y Fundamental son capacitados.
- Un profesor puede subir de nivel en un año gracias a una capacitación de calidad.
- De un año a otro sólo puede progresar en un nivel de calidad.
- 25 % de los profesores son evaluados cada año a través de la Prueba de Evaluación Docente.
- Se realiza una prueba de habilitación docente que comienza a rendirse en 2011, para todos quienes entren al sistema municipal. Esta evaluación realiza durante el primer año de desempeño, siendo similar en su forma ala actual prueba de evaluación docente
- Entre un 2,5 % y un 3,9 % del total de profesores se jubila cada año,

en cada uno de los tres niveles de educación que se estudian; Párvulo, Básica y Media.

- El 1.5 % se retira voluntariamente
- Para reflejar como una mejora paralela en las aulas en la que los profesores son formados, se tiene una red paralela que el 2016 permiten ingresar al sistema profesores mejor capacitados. Estos nuevos docentes, pueden comenzar en niveles superiores de desempeño en su primer año de ejercicio, con menores restricciones al subir al nivel siguiente de desempeño.
- Un profesor Insatisfactorio, puede llegar a Experto en 5 años mínimo si se capacita todos los años.
- Los profesores pueden mantenerse, subir de nivel o salir del sistema entre un año u otro.
- Aquellos profesores evaluados 2 veces Insatisfactorios, deben salir del sistema educacional municipal.
- Se corren 3 versiones del modelo: uno para Párvulo, uno para Básica y otro para Media, que si bien poseen las mismas restricciones, los datos de input son diferentes, respetando la realidad de cada uno. De esta manera, se tiene una mayor flexibilidad para genera una propuesta más robusta y flexible frente a restricciones de presupuesto.

Siguiendo lo anterior, el presente modelo considera 5 niveles;

- **Experto:** Último nivel de la red que puede ser alcanzado, reflejando un manejo completo de los contenidos evaluados en la Evaluación Docente, así como de estudios de postgrado. Se les aplica la Evaluación Docente cada 4 años.
- **Intermedio:** Tercer nivel del trayecto en la escala, representa a quienes son catalogados como Competentes en la Evaluación Docente y además poseen conocimientos adquiridos mediante diplomados u otros estudios personales acreditados. Al igual que en el caso anterior, son evaluados cada 4 años.
- **Fundamental:** Se refiere a quienes logran un nivel de Básicos en la Evaluación Docente, pudiendo o no tener estudios de post-títulos acreditados. Al ser determinados Básicos, son capacitados por el estado de manera necesaria durante el año siguiente al que rindieron la prueba de Evaluación Docente , la que se repite en su caso también cada cuatro años.

- **Insatisfactorio:** Representa el primer nivel de la malla, siendo equivalente al de la Evaluación Docente, en el que quienes obtienen este desempeño, tienen un bajo dominio de los criterios del Marco para la Buena Enseñanza, por tanto deben de capacitarse obligatoriamente para volver a ser evaluados al año siguiente.
- **Principiantes:** Aquellos profesionales del área de educación que tienen un año o menos de desempeño en el sistema estatal. Son quienes para poder entrar a trabajar a los colegios Municipales deben de rendir durante este primer año de trabajo, la Prueba de habilitación, la cual tiene características similares a la de Evaluación Docente (la cual realizan durante su segundo año de trabajo), con el fin de determinar si pueden ejercer en colegios de este tipo de dependencia. Se dividen en principiantes de primer y segundo año de desempeño.

Un ejemplo de las dinámicas de transición entre niveles es mostrado en la Figura 7.

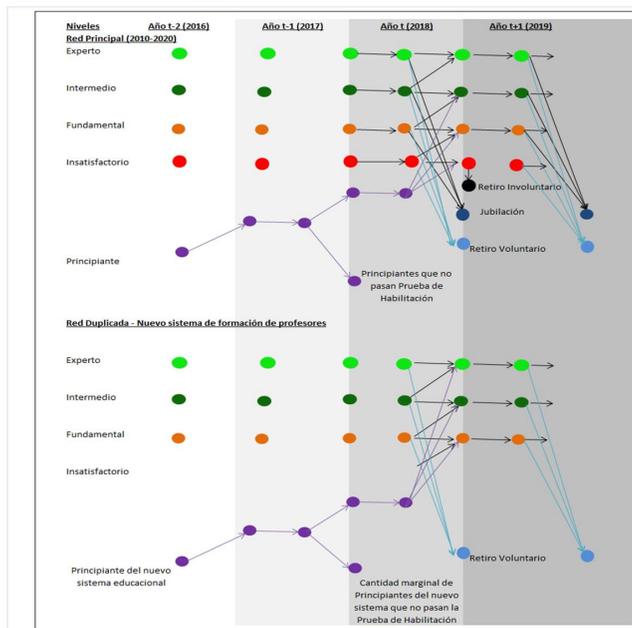


Figura 7: Dinámica de flujo según nivel de desempeño.

## 4.1. Modelo de Programación Lineal

El modelo de red descrito, comprende los años del período 2010-2020 ambos inclusive. La metodología escogida es resultado de un análisis frente a las herramientas disponibles. Como característica principal se encuentra su facilidad para amoldarse a problemas de programación lineal y de reflejar los movimientos entre un estado y otro de las variables a considerar, las que en este caso se refieren a cantidad de profesores.

Las principales variables del modelo son:

- Cantidad de profesores en cada estado o nivel ya descrito de la red en el año  $t$ .
- Cantidad de profesores que se mantienen o suben al estado siguiente entre  $t$  y  $t + 1$ . Ello como consecuencia de la carrera docente explicada en líneas anteriores.
- Cantidad de docentes que deben salir en  $t$  del sistema municipal, como consecuencia de haber obtenido dos evaluaciones Insatisfactorias consecutivas.
- Cantidad de profesores se retiran del sistema entre  $t$  y  $t + 1$ , por motivos de jubilación o de forma voluntaria (ambas identificados de manera separada) y de qué nivel provienen.
- Cantidad de educadores que como consecuencia de un deficiente resultado en su Prueba de Habilitación no les es posible ejercer en el sistema municipal y por tanto deben salir de él.
- Cantidad de profesores que dan por segunda vez la Prueba de Habilitación. Se refiere al grupo de profesores que por segunda vez consecutiva dan la Prueba de Habilitación con el fin de entrar al sistema municipal. Dado esto, no implica que por ello, pasen necesariamente esta vez la evaluación.
- Costo Anual de la propuesta en USD (560,14CLP/USD).

Por otro lado, como fue mencionado en líneas anteriores, se cuenta con una red duplicada en la cual se refleja la entrada e interacción entre un estado y otro, para un grupo de profesionales con un mejorado estándar educacional, quienes se titularían a partir del año 2016. A su vez, ambas redes en conjunto cumplen con los requerimientos de demanda y la factibilidad de que los docentes decidan de manera voluntaria abandonar el sistema educacional municipal. En este caso, las principales variables involucradas son:

- El Flujo de Transición de la red duplicada: representa la cantidad de profesores que se mueven desde un estado a otro, sin poder realizar descensos de nivel. Su principal diferencia con el flujo de transición de la red anterior, es que no permite que los Principiantes no pasan a ser Insatisfactorios, dado su conocimiento y habilidades adquiridas.
- El Flujo de Mantención de la red duplicada: Refleja la cantidad de profesores en cada nivel por cada año.

- El Flujo de Salida voluntaria de la red duplicada: Muestra cuantos docentes de cada nivel se retiran voluntariamente año a año, como parte de la cota total calculada para ambas redes en conjunto.
- El Flujo de Retiro por una calificación deficiente en la Prueba de Habilitación para la red duplicada: Tal como la red principal, responde mostrar la cantidad de educadores que no pasan la prueba de evaluación, pese a provenir de un mejorado sistema de formación. Debido a esto, se utilizará para observar escenarios en que una porción pequeña del total obtenga calificación insuficiente o que todos superen la prueba en cuestión.
- Cantidad de profesores que dan por segunda vez la Prueba de Habilitación en la red duplicada: Se refiere al grupo de profesores que por segunda vez consecutiva dan la Prueba de Habilitación con el fin de entrar al sistema municipal, provenientes del sistema educacional representados por la red duplicada. Dado esto, no implica que por ello, pasen necesariamente esta vez la evaluación.

Los costos asociados al proceso son:

- Costo de la Evaluación Docente del año  $t$ . Refleja el costo total, multiplicando el número de profesores que se evalúan por su costo.
- Costo de la prueba de Habilitación del año  $t$ . Al igual que con la prueba de Evaluación Docente, recoge la cantidad de profesores que ingresan al sistema municipal y el costo de la prueba de habilitación propuesta.
- Costo total de las capacitaciones brindadas en el año  $t$ .
- Muestra la inversión requerida cada año para poder brindar a los docentes sus remuneraciones según nivel de desempeño.
- Resume de todos los costos por salida, como bonos por jubilación o desvinculación cada año del horizonte.
- Represente el costo asociado a la formación de mejores docentes provenientes de un sistema de educación universitaria de calidad superior.
- Esta variable considera todas las condiciones de borde del modelo que deben reflejarse como parte del año 2020.

Las principales restricciones que involucran a las variables descritas corresponden a conservaciones de flujos y cumplimiento de demandas por nivel de desempeño; es decir que cada año se vea cumplida la cantidad de docentes en cada uno de los niveles explicados como reflejo de la mejora de la calidad de profesores, para lo cual entre un año y otro deben ser capacitados. En cuanto a la dinámica de ingreso al sistema, se hacen efectivas las tasas de aprobación para la prueba de Habilitación y la incorporación a los niveles de Insatisfactorio y Fundamental (desde el estado de Principiantes de segundo año de desempeño) para la red principal y hasta el 2015, incluyéndose la posibilidad de entrar al estado Intermedio desde el 2016 en adelante. Por su parte, para el caso de la red duplicada todos los niveles de desempeño son factibles desde

el segundo año de desempeño como consecuencia de pertenecer a una mejor formación inicial de los docentes que recién ingresan al sistema municipal. Finalmente, es de suma importancia recordar que como parte fundamental de la propuesta se refleja el hecho que quienes sean dos consecutivas veces catalogados como Insatisfactorios, deben abandonar la red, para lo cual puede agregarse un bono de retiro que si bien en la actualidad no existe, es posible de integrarse a análisis futuros al tratarse de un parámetro considerado en el código.

Así, es posible obtener luego a través de la programación en el software AMPL, con la resolución por CPLEX, el cálculo de manera anual del costo de la propuesta y el número de profesores que cambian de estado y desde qué nivel se mueve al siguiente. Es importante resaltar que gracias a la alta flexibilidad presentada por la herramienta, es pueden manejar un gran número de escenarios factibles.

## 4.2. Resultados

En esta sección se presentan los resultados analizados para 5 casos diferentes. Para cada uno de los niveles educacionales; Párvulo, Básica y Media, se establecieron casos base por separado considerando los salarios máximos y mínimos actuales según datos entregados por Educación 2020, los cuales fueron reajustados por un 3% anual, desde el 2010 al 2020, a modo de proyección del de la moneda real esperada como único factor de variación conocido (con un tipo de cambio de 560,14 a correspondiente al actual utilizado para el presupuesto 2010 del gobierno). A su vez, se consideran las demandas por docentes calculadas por el modelo de demanda ya explicado, incluyendo sólo la influencia de la demografía entregada por el INE, en el período estudiado. En cuanto al objetivo de calidad, comprendido como la proporción del total de profesores que representa cada nivel de desempeño (Experto, Intermedio, Fundamental e Insatisfactorio) se mantiene el actualmente utilizado que responde al promedio de los resultados de los últimos cuatro años de realización de la Evaluación Docente. Por último, consideramos los incrementos en costo para cada escenario en relación a un caso base en que se considera la mantención de las condiciones actuales y un incremento de 3% anual en las remuneraciones de los profesores.

La reducción de demanda de profesores por factores demográficos implica que los costos del caso base crecen lentamente entre 2010 y 2020.

La reducción de niños vulnerables resulta ser un factor importante en cuanto a la variación de los costos, lo cual se explica por el aumento de los requerimientos de profesores para poder cubrir el aumento de salas que se deben crean. Este factor supera al efecto de la sola reducción de la proporción horas lectivas/no lectivas, siendo este último uno de los puntos fuertes de la propues-

ta. Es por esto, que es recomendable evaluar reducir la cantidad de alumnos vulnerables por sala, sólo en uno de los tres niveles de educación, como lo sería por ejemplo en Párvulo, ya que es donde se obtienen las mayores diferencias en el nivel de aprendizaje.

Por su parte, se analizó la posibilidad de buscar un objetivo menos ambicioso para el nivel de calidad al 2020, siendo más reservado, pero quizás más factible en su ejecución pensando en las dificultad de formar en 10 años un número elevado de profesionales expertos (Tabla 1). De esta forma, se obtuvo un posible ahorro de casi un 4%.

	<b>Inicio 2010</b>	<b>Objetivo al 2020</b>	<b>Objetivo Alternativo</b>
Expertos	8 %	45 %	15 %
Intermedios	58 %	40 %	55 %
Fundamentales	32 %	15 %	28.6 %
Insatisfactorios	1.4 %	0.10 %	1.4 %

Tabla 1: Objetivo de Calidad Docente como porcentaje del total de profesores en el sistema Municipal cada año.

En cuanto a los sueldos, como era de esperarse representan otro de los factores críticos en la propuesta. Esto es directa consecuencia al aumento de dos ítems al tener un sistema que gratifica el desempeño individual:

- I. Por una parte el sostenido aumento del monto de las remuneraciones, que trabajan como incentivos para que los buenos docentes permanezcan como tales en las aulas, al tiempo que se ayudan a atraer alumnos con habilidades a las carreras de pedagogía.
- II. Gracias a los programas de capacitaciones, se aumenta la proporción de profesores con buen desempeño en las salas de clase. De esta forma el ejercicio al estar ligado a las remuneraciones, con un sistema de capacitaciones e incentivos se espera logre los niveles de calidad en las aulas que se planifican, resulta en este aumento de costos.

Se identifican una gran cantidad de escenarios posibles, frente a los que se contrasta la propuesta sensibilizando parámetros. Por ejemplo, si se decide aumentar al doble los costos de las capacitaciones hoy consideradas, las que por su parte recaen en entidades independientes y acreditadas en forma de diplomados o magister, el aumento es significativo, de 4.2%.

La Tabla 2 muestra los resultados de correr los distintos escenarios. Como se ve, los costos anuales son crecientes. Además tienden a estabilizarse después del año 2020. Los valores marcados como GAP INVERSION reflejan los costos adicionales, por sobre el plan base, de las distintas mejoras propuestas. Así tenemos que los gastos aumentan desde el caso mas básico de mejora de sueldos y reducción e horas lectivas , que para el año 2020 significa un costo

adicional de \$1.2 billones de dólares sobre el escenario base, hasta el escenario que incluye todas las mejoras, reducción de horas lectivas, mejora de remuneraciones , reducción de número de alumnos por clase en niños vulnerables, mejora e intensidad de las capacitaciones, con un costo anual de 2.3 billones de dólares.

GAP In- versión	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Caso 1	\$100	\$189	\$267	\$356	\$438	\$526	\$736	\$841	\$962	\$1.081	\$1.016	\$6.713
Caso 2	\$118	\$246	\$365	\$504	\$642	\$798	\$1.158	\$1.358	\$1.603	\$1.854	\$2.136	\$10.781
Caso 3	\$142	\$270	\$388	\$526	\$663	\$818	\$1.336	\$1.552	\$1.774	\$2.025	\$2.310	\$11.804
Caso 4	\$116	\$235	\$343	\$468	\$591	\$729	\$1.067	\$1.242	\$1.458	\$1.672	\$1.912	\$9.835
Caso 5	\$103	\$198	\$282	\$379	\$471	\$570	\$807	\$930	\$1.072	\$1.215	\$1.377	\$7.402

Tabla 2: Resultados modelo de Carrera Docente. Reflejo del gap de inversión anual versus caso base. Montos en millones de dólares.

Caso 1: Reducción Hras. Leactivas/No Lectivas

Caso 2: Reducción Hras. Leactivas/No Lectivas y de niños vulnerables por sala.

Caso 3: Reducción Hras. Leactivas/No Lectivas y de niños vulnerables por sala + Capacitación Doble.

Caso 4: Reducción Hras. Leactivas/No Lectivas y de niños vulnerables por sala + Obj. Alternativo.

Caso 5: Reducción Hras. Leactivas/No Lectivas y de niños vulnerables por sala PB + sólo reducción HLNL Básica y Media.

---

## 5. Conclusiones

---

Para la evaluación a la propuesta de Educación 2020 para la carrera docente, se establecieron los requerimientos en la dinámica de flujos para reflejar lo que sería el camino a una mejora sustancial en la calidad de la educación impartida, soportada por inversiones en el área de salarios, capacitaciones y bonos de retiro por jubilación, con el fin de entregar una señal clara de la promesa de una proyección digna y gratificante del oficio. Si bien la consecuente mejora en la calidad de la educación tiene consecuencias directas en las oportunidades de los alumnos del sistema de educación municipal, este trabajo no pretende cuantificarlos, siendo por tanto el paso a investigaciones futuras. La metodología escogida, responde a la necesidad de plasmar fácilmente junto con los montos requeridos, el número de docentes en cada nivel de desempeño y su evolución entre un año y otro. Así se permite evaluar tanto los flujos de costos como de los mismos docentes, como parte del sistema educacional en un período de 11 años.

Como factores críticos están los sueldos objetivos en conjunto con las inver-

siones en capacitaciones, lo que afecta mayormente a la inversión en el nivel de Párvulo, quienes actualmente reciben los menores salarios. Además, se halla el efecto de la demanda de docentes, lo que refleja las decisiones de reducir o no la cantidad de niños vulnerables por sala y/o mejorar la proporción de horas lectivas/no lectivas. Por su parte, el reflejar un aumento en los sueldos, representa un mensaje potente con el compromiso en la mejora de la estructura de sueldos y por lo tanto un punto a atracción para atraer estudiantes de alto desempeño a las carreras de pedagogía. Relacionado con ello, se encuentra la factibilidad de la propuesta, en concordancia con la necesidad de docentes. El modelo supone que se encuentra disponible el volumen necesario de profesionales para entrar al sistema municipal, pero a su vez calcula el déficit de los mismos como la diferencia entre los que se requiere que ingresen a su primer año de desempeño (considerando la proporción que podrá aprobar la prueba de habilitación, suponiendo un 60 % los 3 primeros años) y el 50 % del promedio de titulados según los datos de la DIVESUP, manteniendo este último en el tiempo al no contarse con una proyección a futuro. De esta forma, al combinar la reducción de niños por sala y de proporción hrs lectivas/no lectivas, se tiene en promedio una carencia anual de 565 y 3402 profesores de nivel Párvulo y Básico respectivamente, como consecuencia de la baja tasa esperada de aprobación de la prueba de habilitación. Esto demuestra la necesidad de atraer alumnos de alta calidad y de mejorar la formación de los mismos, para revertir este índice de aprobación, reduciendo esta diferencia. Por su parte, el caso de la educación media, no presenta grandes déficits, debido a que la demografía disminuye el impacto del incremento de la demanda de docentes por los efectos de la reducción en horas lectivas/no lectivas y de niños por sala. Cabe recordar, que en los últimos años, la cantidad de estudiantes de pedagogía se ha visto incrementada de manera importante, lo que significa que en el futuro próximo, la cantidad de titulados será mayor a la utilizada en este momento para calcular esta diferencia. Es importante recordar que al suponer un objetivo de calidad que podría ser llamado más conservador, pero con mayores probabilidades de lograrse, la inversión incremental se reduce en un 3.9 %, lo que es un factor importante en la toma de decisiones.

Si bien la inversión requerida no deja de ser importante, nuestro país posee un gasto público en educación de un 2,4 % del PIB, lo cual en comparación al 3,5 % promedio de los países de la OECD representa un amplio margen de crecimiento posible. De esta forma, un incremento como el de la propuesta de rebaja de hrs. lectivas/no lectivas, sumado a una mejora en la cantidad de alumnos vulnerables por sala, propuesta de carrera docente con objetivos de calidad que apuntan a un 85 % de los docentes en los niveles Expertos e Intermedios, tiene un anual incremental al año 20 de 2.3 billones de dólares. Este valor se puede reducir por ejemplo si se disminuye linealmente el número de niños vulnerables sólo en Párvulos.

Por último, no debe dejar de mencionarse que una arista importante es que se está modelando el comportamiento y mejor dicho, la mejora de las capacidades de personas, por lo que los supuestos en torno a esta parea fueron brindados por expertos para reflejar de mejor forma la realidad. Cabe notar además que las cifras indicadas corresponden a números aproximadas solamente, dada la relativa poca exactitud de múltiples parámetros considerados, como ya se observó.

**Agradecimientos:** Los autores agradecen al apoyo de la Fundación Educación 2020 y al Instituto Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16).

## Referencias

- [1] Educación 2020. Hoja de Ruta 2010-2020, Se acabó el recreo. Santiago, Chile. 2010.
- [2] Amar, M. Equidad, Calidad y Derecho a la Educación en Chile: Hacia un nuevo rol del Estado. Biblioteca Nacional de Chile. Serie N°9. Chile. 2007.
- [3] Barrena C. Modelo de Redes de Flujo para la Evaluación de la Propuesta de Carrera Docente de Educación 2020. Memoria Ingeniería Civil Industrial, Universidad de Chile. 2010.
- [4] Beyer, H. y Araneda, P. Hacia un Estado más Efectivo en Educación: una mirada a la regulación laboral docente. 2009.
- [5] Departamento de Estudios y Desarrollo . Indicadores de la Educación en Chile. División de Planificación y Presupuesto. Chile. 2007.
- [6] Departamento de Estudios y Desarrollo. Indicadores de la Educación en Chile 2006. División de Planificación y Presupuesto. Chile. 2006.
- [7] Mizala, A. y Romaguera, P. Determinación de factores explicativos de los resultados escolares en educación media en Chile. Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Universidad de Chile. 2000.
- [8] Unicef, MINEDUC Gobierno de Chile. ¿Quién dijo que no se puede? Escuelas Efectivas en Sectores de Pobreza. Chile. 2004.

---

# PROGRAMACIÓN DEL FIXTURE DE LA SEGUNDA DIVISIÓN DEL FÚTBOL DE CHILE MEDIANTE INVESTIGACIÓN DE OPERACIONES

---

GUILLERMO DURÁN\*

MARIO GUAJARDO\*\*

RODRIGO WOLF-YADLIN\*\*\*

## Resumen

*Presentamos en este trabajo la aplicación de técnicas de Investigación de Operaciones a la programación del fixture de la Segunda División del fútbol de Chile. Este fixture debe cumplir una serie de condiciones solicitadas por la Asociación Nacional de Fútbol Profesional (ANFP), entidad que organiza el torneo. El criterio geográfico es particularmente importante, debido a que la disposición de algunos equipos en lugares extremos del país implica largos desplazamientos, a menudo realizados por vía terrestre. Abordamos el problema mediante un modelo de programación lineal entera que define cuándo y dónde se juega cada partido del torneo, sujeto a que todas las condiciones se cumplen. Para las instancias más difíciles, desarrollamos un modelo adicional también de programación lineal entera que genera patrones de localías y los asigna a los equipos, previo a la ejecución del modelo que define la programación de los partidos. Los fixtures así generados han sido exitosamente utilizados en los cinco torneos de Segunda División que se han disputado entre 2007 y 2010, reemplazando la metodología aleatoria que la ANFP utilizaba anteriormente. Durante este período, el tipo de torneo ha sufrido diversas modificaciones, incluyendo un cuádruple round-robin y un torneo en dos etapas que considera fases regionales y nacionales. Hemos debido entonces adaptar nuestros modelos temporada tras temporada, según el tipo de torneo. Esta aplicación marca un nuevo avance en el uso de Investigación de Operaciones para la gestión del fútbol chileno, que se suma a otros proyectos desarrollados en la misma línea en los últimos años.*

**Palabras Clave:** Fútbol, Patrones de Localías, Programación Entera, Programación en Deportes.

---

\*Departamento de Ingeniería Industrial, FCFM, Universidad de Chile y Departamento de Matemática, FCEN, Universidad de Buenos Aires.

\*\*Department of Finance and Management Science, Norwegian School of Economics and Business Administration.

\*\*\*Departamento de Ingeniería Industrial, FCFM, Universidad de Chile.

---

## 1. Introducción

---

La aplicación de técnicas de optimización a la programación de ligas deportivas ha recibido atención durante, al menos, las últimas cuatro décadas. El interés práctico y la difícil resolución a instancias de tamaño real de los problemas de programación en deportes han cautivado la atención de académicos en Investigación de Operaciones (IO) y áreas relacionadas. En fútbol, uno de los deportes más populares del mundo, la literatura que reporta el uso de técnicas de programación en deportes en ligas del mundo real es relativamente reciente. La primera aplicación es reportada en [17], donde se muestra la programación de la liga de fútbol de Holanda en la temporada 1989/1990. Esta liga constaba de 18 equipos, que se enfrentaban todos contra todos una vez, i.e., en un torneo *round robin*. Diferentes condiciones fueron consideradas, abarcando aspectos comerciales, deportivos y de organización. Las restricciones son divididas en duras, que deben cumplirse obligatoriamente, y blandas, que deseablemente deben cumplirse, a las cuales se asigna un peso que representa su importancia. La función objetivo busca maximizar el peso total que logra la solución. El problema es abordado mediante la implementación de una heurística constructiva. La solución encontrada cumple el 80 % de las condiciones requeridas para el torneo de la temporada 1989/1990.

Una aplicación a la liga de fútbol de Argentina fue realizada por E. Dubuc en 1995 (documentada en [13]), pero fue utilizada sólo en el campeonato de ese año. El tipo de torneo era un *round robin* de 20 equipos y el objetivo fue cumplir con los requerimientos de la TV. El esquema de solución consistió de la implementación de una heurística basada en *simulated annealing*.

Bartsch et al. programaron las ligas de fútbol profesional de Alemania y Austria [2]. La liga alemana consideraba 18 equipos, que jugaban un torneo doble *round robin* espejado (la segunda mitad del torneo es equivalente a la primera, invirtiendo las localías). La liga de Austria consideraba 10 equipos, que jugaban un torneo cuádruple *round robin*, no espejado. Para ambas ligas, varias condiciones fueron requeridas, incluyendo criterios de atractivo para el público, de equidad deportiva y organizativos. Estas restricciones son clasificadas en duras y blandas. La función objetivo busca minimizar el costo de penalización de restricciones violadas. La metodología de solución para generar los fixtures consiste de la implementación de una heurística. La liga alemana usó esta solución en la temporada 1997/1998, y la liga de Austria la empleó en seis oportunidades, entre 1997 y 2003.

Durán et al. reportan la programación del campeonato de Primera División del fútbol profesional chileno desde 2005 a 2007 mediante el uso de técnicas de IO [4] (posterior a la publicación de dicho artículo, los autores han seguido

programando todos los torneos de la liga hasta la actualidad). La Primera División constaba entonces de 20 equipos, divididos en cuatro grupos de cinco equipos cada uno. El torneo se componía de dos fases. La primera fase estaba organizada como un *round robin*, es decir los 20 equipos jugaban todos contra todos una vez. A la segunda fase avanzaban los dos mejores equipos de cada grupo, en total ocho equipos, que se enfrentaban en series de *playoffs* para determinar al campeón. Las condiciones que el fixture debía cumplir incluían distintos aspectos, tales como restricciones geográficas, atractivo para los simpatizantes y requerimientos de la TV. El enfoque de resolución se basa en programación lineal entera y procedimientos básicos para la asignación de patrones de localías a los equipos.

Rasmussen programó la principal liga de fútbol de Dinamarca en la temporada 2006/2007 [14]. La liga consistía de 12 equipos, que se enfrentaban entre sí tres veces, constituyendo así un torneo triple *round robin*. Una variedad de condiciones fueron consideradas, tales como requerimientos geográficos y restricciones asociadas a los equipos principales. Estas condiciones son divididas en duras y blandas. La función objetivo busca minimizar las penalizaciones de no cumplir ciertas restricciones. El enfoque de solución usa descomposición de Benders y técnicas de generación de columnas, y conduce a soluciones buenas en poca cantidad de tiempo.

Goossens and Spijksma programaron la principal liga de fútbol de Bélgica en la temporada 2007/2008 [8]. El torneo consistió de 18 equipos que jugaron un doble *round robin* espejado. El número total de *breaks* (repetición de la condición de local o visitante de un equipo en fechas consecutivas) debía ser mínimo. Varias condiciones más fueron consideradas, tales como requerimientos de la TV y balance de los efectos de *carry-over* (un equipo  $i$  recibe una unidad de efecto *carry-over* del equipo  $j$ , si  $i$  juega en la fecha  $k + 1$  contra el mismo rival que  $j$  enfrentó en la fecha  $k$ ). Las restricciones del problema son clasificadas en grupos de acuerdo a su nivel de prioridad, asignando penalizaciones. El enfoque de solución utiliza un modelo de programación entera mixta que asigna equipos a un fixture canónico preestablecido (generado para balancear los efectos de *carry-over*), satisfaciendo las restricciones de fixture espejado y minimizando el número de *breaks*. Luego, procedimientos de búsqueda local son utilizados para mejorar la solución, en términos de minimizar la penalidad total de las restricciones violadas.

Más recientemente, otros dos casos de aplicaciones han sido reportados. Flatberg et al. programaron la liga de fútbol de Noruega usando un modelo de programación entera [6], que considera en su objetivo los efectos de *carry-over*. Esta liga juega un doble *round robin* no espejado de 14 equipos. Ribeiro y Urrutia programaron la liga de fútbol de Brasil en 2009 [16], utilizando también un enfoque de programación entera que busca maximizar el número de partidos atractivos entre equipos de elite que pueden ser transmitidos por

televisión abierta. La liga de Brasil está compuesta de 20 equipos que juegan un torneo doble *round robin* espejado.

Hasta lo que conocemos, los trabajos citados en los párrafos anteriores son las únicas aplicaciones reportadas en la literatura que han sido usadas en la práctica en ligas de fútbol del mundo real. Hay otros artículos contextualizados en ligas de fútbol, pero que han permanecido a nivel teórico. También suponemos que hay otras aplicaciones en el mundo real que no han sido reportadas en artículos académicos. Mediante comunicación personal, sabemos que técnicas de optimización han sido utilizadas en la programación de ligas de fútbol en República Checa, desde la temporada 2002/2003 a la fecha [7]; Polonia, en las temporadas 2007/2008 y 2008/2009 [11]; y Honduras, en el torneo 2010 [5].

Existe también una extensa literatura en la programación de torneos en otros deportes. Para una revisión extensa a la literatura, incluyendo aspectos teóricos y prácticos, referimos al lector a un par de artículos recientes ([15], [10]). Instancias de prueba para distintos torneos del mundo real pueden encontrarse en [12].

Las aplicaciones de IO a la programación de torneos de fútbol parecen ser pocas, considerando que actualmente asociaciones de 208 países están afiliadas a la Federación Internacional de Fútbol Asociado (FIFA). Es presumible entonces que el uso de técnicas manuales aún impera en la programación de muchas ligas alrededor del mundo. En este artículo, reportamos un nuevo caso de aplicación: la Segunda División del fútbol chileno. Esta liga, al igual que la de Primera División, está organizada por la Asociación Nacional de Fútbol Profesional (ANFP).

La utilización de técnicas de programación en deportes en el fútbol chileno comenzó en el torneo de Primera División 2005. Debido a los buenos resultados obtenidos, dos años más tarde la ANFP nos solicitó apoyar también la programación del torneo de Segunda División. De este modo, hemos programado los cinco torneos de Segunda División que se han jugado desde 2007 hasta la actualidad.

Las características del torneo de Segunda son diferentes al de Primera División. Aun más, el campeonato de Segunda División ha sufrido diversas modificaciones en los últimos cuatro años. De este modo, los modelos y metodologías de resolución que utilizábamos para el campeonato de Primera División no fueron suficientes para programar el de Segunda. Además de adaptar la modelación para la asignación de partidos, hemos debido explorar más profundamente la generación de los patrones de localías. Para las instancias más difíciles del torneo de Segunda División hemos entonces desarrollado dos modelos de programación lineal entera que se resuelven en etapas sucesivas: primero, un modelo para generar los patrones de localías de los equipos y posteriormente, un modelo que decide la asignación de partidos. Este procedimiento permite resolver rápidamente todas las instancias en que hemos

trabajado.

La aplicación de nuestros modelos en el torneo de Segunda División ha sido un nuevo paso adelante en el uso de técnicas de IO para la gestión del fútbol chileno, logrando impactos cualitativos y cuantitativos importantes, que han satisfecho plenamente las expectativas de las autoridades de la ANFP y de los clubes. En la práctica, el uso de técnicas de IO ha reemplazado la metodología básica que anteriormente utilizaba la ANFP para programar el torneo y su impacto positivo ha hecho que su uso perdure hasta nuestros días.

Consideramos que las principales contribuciones de este trabajo son tres: la expansión del uso de IO en la gestión del fútbol chileno; la generación de un mecanismo de solución en dos etapas que permite resolver rápidamente las instancias más difíciles de la programación de la Segunda División y que incluso pudo ser adaptado para resolver de manera más eficiente el fixture de Primera; y, por último, la versatilidad de los modelos que han permitido resolver los distintos tipos de campeonato que ha tenido la Segunda División en los últimos años.

El artículo está organizado de la siguiente manera. En la Sección 2 se describe la organización del fútbol en Chile. En la Sección 3 se muestran las condiciones genéricas y los distintos tipos de torneo de Segunda División que se han disputado en los últimos años. En la Sección 4 se describen los métodos de resolución del problema, mientras que la última Sección está dedicada al impacto y las conclusiones.

---

## 2. El fútbol profesional chileno

---

El fútbol profesional chileno consta de dos ligas: la Primera División, en la que actualmente participan los 18 mejores equipos, y la Segunda División, compuesta por otros 14 equipos. Estas son las únicas ligas de fútbol profesional en Chile. Anualmente, los dos mejores equipos de la Segunda División son promovidos al torneo de Primera División del año siguiente, mientras que los dos últimos de la Primera División son relegados a la Segunda División. Adicionalmente, cada fin de año los equipos que obtuvieron el tercer y el cuarto lugar de la Segunda División disputan una Promoción en series de *playoffs* con los equipos en las posiciones 16 y 15 de la Primera División, respectivamente. Los dos vencedores de estos enfrentamientos juegan en Primera División al año siguiente, mientras que los dos equipos perdedores juegan en la Segunda División.

Lógicamente, todos los equipos aspiran a jugar en Primera División, que es la más popular entre los simpatizantes y, por lo tanto, la más atractiva para los auspiciantes y la TV. No obstante, la Segunda División es muy relevante debido al dinamismo que imprimen las posibilidades de ascenso a la Primera

División de un año a otro. Además, en la Segunda División la gran mayoría de los equipos son de fuera de Santiago, representando a ciudades en las cuales el fútbol suele jugar un rol protagónico en el panorama de actividades de entretenimiento. Luego, estos equipos se vuelven fuertemente populares entre sus habitantes, generando un sentimiento de identificación especial, y despiertan el interés de auspiciadores privados regionales. De hecho, los 14 equipos que actualmente componen la Segunda División son de ciudades distintas a Santiago, mientras que los 7 equipos profesionales de Santiago juegan en la Primera División. El atractivo del torneo ha progresivamente cautivado el interés de la TV, que durante los últimos torneos ha transmitido también algunos partidos de la Segunda División.

Una de las tareas de la ANFP es la programación del fixture de cada torneo de la Primera y Segunda División. Previo a 2005, la ANFP utilizaba un fixture canónico en la programación de todos sus torneos. Este fixture es una asignación predeterminada de tantos números como equipos en una planilla que determina en qué orden se enfrentan. La ANFP aleatoriamente asignaba equipos a estos números, generando así el fixture utilizado en los torneos. Esta modalidad les permitía realizar una programación que cumplía los requerimientos básicos del torneo, pero no permitía considerar una serie de otros criterios, lo que conllevaba a múltiples quejas de parte de los equipos.

Para el torneo de la Primera División 2005 las autoridades de la ANFP solicitaron a investigadores de la Universidad de Chile, entre los que se encontraban los autores de este artículo, apoyar la programación del fixture mediante técnicas de IO. Se desarrolló entonces un modelo de programación lineal entera, que incorporó requerimientos de la ANFP, de los clubes y de la TV. Las condiciones consideradas integraron múltiples aspectos tales como geográficos, económicos, de equidad deportiva y de atractivo para el público. El resultado fue exitoso, causando impacto positivo en todos estos aspectos.

Debido a los buenos resultados obtenidos en la Primera División, la ANFP nos solicitó apoyar la programación del fixture del torneo de Segunda División en 2007. Naturalmente, un enfoque similar al que utilizamos para abordar el problema de la Primera División nos permitió encarar el problema de Segunda, aunque las diferentes características de los torneos y de los requerimientos de los equipos nos obligaron a generar nuevos modelos que capturaran la nueva problemática. Más aun, la modalidad del campeonato de Segunda División ha ido variando temporada a temporada. En todos los casos, hemos generado el fixture del torneo usando modelos de programación lineal entera, adaptando nuestra modelación y metodología de resolución según el tipo de torneo, la cantidad de equipos y los requerimientos de la ANFP. En las siguientes secciones describiremos las características más importantes de la aplicación y su evolución a lo largo de los años.

---

### 3. Programación de la Segunda División

---

La temporada de la Segunda División del fútbol chileno parte en febrero de cada año. La programación del torneo se prepara con un par de meses de anticipación. La ANFP fija de antemano las fechas calendario en que se debe disputar cada ronda (o fecha) del torneo. Generalmente cada ronda se disputa los fines de semana (de viernes a domingo), aunque a veces puede ser fijada para mitad de semana (de martes a jueves). Posteriormente, nosotros generamos una propuesta de asignación de los partidos a las fechas, considerando una serie de restricciones. Algunas condiciones son solicitadas por los equipos a la ANFP, otras son determinadas directamente por la ANFP y otras atienden a características propias de cada tipo de torneo.

#### 3.1. Condiciones genéricas

A continuación describiremos aquellas condiciones que se han repetido en prácticamente todos los torneos de la Segunda División desde 2007 a 2010, con mínimas variaciones, y por eso las llamamos *condiciones genéricas*.

- **Restricciones básicas**

1. Cada equipo juega un partido en cada fecha.
2. Cada par de equipos se enfrenta entre sí un cierto número  $m$  de veces, que depende del tipo de torneo.

- **Restricciones sobre breaks**

Decimos que un equipo tiene un *break* en la fecha  $k$  si en las fechas  $(k-1)$  y  $k$  juega en ambas de local o en ambas de visita.

3. Cada equipo puede jugar a lo más un cierto número  $l_{home}$  *breaks* de local y  $l_{away}$  *breaks* de visita.
4. En determinadas fechas, ningún equipo puede tener un *break*. Esta condición generalmente se impone sobre la segunda y la última fecha del torneo, por una razón de equidad deportiva (no es deseable que los equipos partan o terminen el torneo con dos partidos consecutivos de local o de visita).



Figura 1: Ubicación geográfica de los 14 equipos de la temporada 2010 de la Segunda División del fútbol profesional de Chile.

### ■ Restricciones geográficas

Chile es el país más largo del mundo, con una longitud de poco más de 4,300 kilómetros. El equipo San Marcos de Arica, de la zona norte del país, debe desplazarse, por ejemplo, 2,250 kms. al sur para jugar de visita contra Curicó Unido, un equipo de la zona central, y 3,080 kms para jugar de visita contra Puerto Montt. El mapa de la Figura 1 muestra la ubicación geográfica de los equipos participantes en la temporada 2010.

Debido a que sus recursos son limitados, los equipos de la Segunda División generalmente no pueden viajar en avión, un medio de transporte altamente más costoso que viajar por vía terrestre. Por lo tanto, a pesar de las largas distancias, los equipos suelen desplazarse en bus para disputar sus partidos de visita (las personas que viajan son en total alrededor de 25, incluyendo jugadores, cuerpo técnico, cuerpo médico y dirigentes). En la práctica, todos los equipos vuelven a casa inmediatamente después de jugar un partido de visita, si es que la próxima fecha se disputará el fin de semana siguiente. Si es que la próxima fecha es en tres días más (por ejemplo, de domingo a miércoles) y el equipo tiene un *break* de visita, el mismo puede preferir quedarse afuera de casa para esperar el partido

siguiente, si es que económicamente le resulta conveniente. Generalmente, esos equipos son los de lugares extremos del país. En la generación del fixture de la Segunda División es entonces sumamente importante considerar restricciones geográficas. Algunas de las que hemos utilizado en los torneos son las siguientes:

5. Balancear a lo largo del torneo el número de viajes que un equipo dado realiza a zonas relativamente *alejadas*. Para ello, definimos clusters de equipos según su locación geográfica. La confección de estos clusters depende, por supuesto, de los equipos participantes en cada torneo, pero generalmente consideramos tres: Norte, Centro y Sur. Luego, imponemos cotas  $t_{low}^{a,b}$  y  $t_{up}^{a,b}$  para el número mínimo y máximo de veces que un equipo de un cluster puede jugar en otro cluster en un cierto número  $s$  de fechas consecutivas del torneo.
6. Programar *buenos* viajes cuando hay fechas a mitad de semana. Un viaje bueno ocurre cuando un equipo visita en 2 fechas consecutivas, siendo una de estas entre semana, zonas lejanas de su ciudad. Por ejemplo, cuando un equipo del Norte (Sur) juega dos partidos de visita consecutivos, uno de estos es un día miércoles en el Centro y el otro es en el Sur (Norte) el fin de semana inmediatamente anterior o posterior. De esta manera, los equipos aprovechan de realizar dos visitas en zonas lejanas en un mismo viaje, pues el costo de alojamiento fuera de casa mientras esto ocurre es menor al ahorro generado en concepto de transporte.
7. Evitar viajes largos seguidos cuando un equipo debe jugar dos partidos consecutivos de visita. Para ello, imponemos que si un equipo juega dos partidos de visita consecutivos, al menos uno de estos debe jugarlo contra un equipo de su cluster. Excepciones a esta regla pueden ocurrir cuando hay fechas a mitad de semana, según mencionamos en la condición anterior.

#### ■ Restricciones de complementariedad

8. Pares de equipos complementarios son definidos para equipos que se requiere jueguen en secuencias de localía complementarias (si uno juega de local en una fecha dada, el otro juega de visita en dicha fecha, y viceversa). Este requerimiento se debe o bien a que los equipos utilizan de local un mismo estadio, a razones de seguridad (para evitar que los simpatizantes de equipos de una misma ciudad estén jugando de local la misma fecha), o a que es deseable mantener actividad futbolística todos los fines de semana en una misma zona.

### ■ Restricciones de localía-visita

9. Fijación de algunas localías, debido a peticiones particulares de los equipos. Por ejemplo, cuando un equipo está de aniversario puede preferir jugar el partido correspondiente a esa fecha como local, para atraer a más simpatizantes al estadio y así generar mayores ingresos.
10. Fijación de algunas visitas, debido a peticiones particulares de los equipos. Por ejemplo, cuando un equipo no tiene disponible su estadio porque el mismo está en reparación o debido a algún evento, puede solicitar jugar el partido correspondiente a esa fecha como visita.

### ■ Restricciones sobre partidos atractivos

11. Fijación de algunos partidos en fechas determinadas, según requerimientos de la ANFP. Esto puede deberse, por ejemplo, a la rivalidad histórica de algunos equipos, que cuando se enfrentan juegan los llamados partidos “clásicos”. Debido al especial atractivo que despiertan en los simpatizantes estos partidos, la ANFP prefiere programarlos para el final del torneo, cuando posiblemente serán definitorios.

### ■ Restricciones sobre patrones de localías

12. Si un equipo juega un número par de partidos en un torneo, se requiere que la mitad de ellos sea de local y la mitad de visita. (Si el número de partidos que un equipo juega por torneo es impar, entonces el número de partidos que juega de local y de visita difieren sólo en uno.)
13. Ningún equipo puede jugar más de dos partidos de visita en tres partidos consecutivos.
14. Ningún equipo puede jugar más de dos partidos de local en tres partidos consecutivos.

## 3.2. Tipos de torneos

En 2007 la dirigencia de la ANFP, elegida cada cuatro años mediante una elección en que votan los presidentes de los clubes profesionales, experimentó un cambio importante. El torneo de la Segunda División ha sufrido desde entonces diversas modificaciones, en pos de alcanzar un sistema de torneo que

sea atractivo y a la vez permita el financiamiento de los clubes. Actualmente, la dispersión de los equipos a lo largo del país y su presupuesto limitado ha terminado por condicionar la modalidad en que se juega el torneo. En los últimos cuatro años, dos tipos de torneos, con ciertas variaciones, han sido utilizados: un cuádruple *round robin* y un torneo en dos etapas de distinto tipo, separando a los equipos en la primera etapa por razones geográficas. Estos tipos de torneo difieren de lo que se hace en la Primera División del fútbol de Chile (que actualmente disputa un único torneo doble *round robin* a lo largo del año), y de lo que la literatura ha reportado para torneos de fútbol, según revisamos en la Sección 1.

A continuación revisaremos brevemente las características particulares más relevantes de los torneos de cada temporada.

**Temporada 2007.** En 2007, la Segunda División del fútbol chileno se componía de 11 equipos. El Torneo se dividió en cuatro cuartos. En cada cuarto, cada par de equipos se enfrentó entre sí exactamente una vez. Dado el número impar de equipos, en cada fecha un equipo debía quedar libre. El segundo cuarto es el espejo del primer cuarto, i.e., si el equipo  $i$  jugó de local contra el equipo  $j$  en la  $n$ -ésima fecha del primer cuarto, el equipo  $j$  debe jugar de local contra el equipo  $i$  en la  $n$ -ésima fecha del segundo cuarto. Análogamente, si el equipo  $i$  quedó libre en la  $n$ -ésima fecha del primer cuarto, el mismo equipo  $i$  debe quedar libre en la  $n$ -ésima fecha del segundo cuarto. Una condición de simetría establece que el tercer cuarto debe ser igual al primer cuarto, y el cuarto cuarto es igual al segundo. El torneo conforma entonces un tipo de torneo que denominamos cuádruple *round robin* doblemente espejado con un número impar de equipos. En total, se jugaron 220 partidos en un total de 44 fechas.

**Temporada 2008.** El Torneo 2008 fue muy similar al de 2007. La diferencia principal es que participaron 12 equipos, por lo tanto, ningún equipo quedó libre en ninguna fecha. En total, se jugaron 264 partidos en un total de 44 fechas.

**Temporada 2009.** La temporada 2009 constó de 14 equipos y se dividió en 2 torneos: el de Apertura y el de Clausura, siendo uno el espejo del otro. El campeón del Apertura y el campeón del Clausura jugaron un *playoff* cuyo ganador ascendió a la Primera División. El perdedor de esta serie accedió a la Promoción para jugar por un lugar en Primera. Además, el equipo que sumó más puntos en toda la temporada, incluyendo ambos torneos, también ascendió a Primera División, mientras que el segundo también accedió a la Promoción.

Cada torneo (Apertura y Clausura) constó de dos fases. En primer lugar se jugó la Fase Zonal y luego la Nacional. En la Fase Zonal, los 14 equipos que componen la Segunda División fueron divididos en dos grupos de siete equipos, de acuerdo a las proximidades geográficas existentes entre ellos. Se

generó así un *Grupo Norte* y un *Grupo Sur*. En cada grupo se jugó un torneo *round robin* con un número de localías balanceado para todos los equipos. Como el número de equipos en cada grupo fue impar, un equipo de cada grupo debía quedar libre en cada fecha. Por su parte, la Fase Nacional fue un torneo *round robin* en el cual se enfrentaron entre sí los 14 equipos, acumulando el puntaje obtenido en los grupos..

Como restricción adicional, la ANFP nos solicitó invertir localías entre la Fase Zonal y Nacional, esto es, si un equipo  $i$  es local contra el  $j$  en la Fase Zonal, el  $j$  debe ser local contra el  $i$  en la Fase Nacional. En la práctica, esta restricción dificultó la resolución del problema en comparación a los torneos de las otras temporadas, según comentaremos en la Sección 4. En total, durante el año 2009 se jugaron 266 partidos en un total de 40 fechas.

**Temporada 2010.** En la temporada 2010 se ha vuelto a jugar un solo torneo en el año, rescatando algunos elementos de los dos tipos de torneo utilizados en los años previos. El torneo se divide en dos etapas: una primera Fase Zonal y luego una Fase Final. En la Fase Zonal los 14 equipos son divididos en dos grupos (Norte y Sur) de siete equipos cada uno, de acuerdo a su proximidad geográfica. En cada grupo se disputa un cuádruple *round robin* doblemente espejado con un número impar de equipos. Luego los cuatro primeros equipos de cada grupo pasan a jugar la Fase Final del torneo que consiste en un doble *round robin* espejado de ocho equipos. Los puntos obtenidos en la Fase Zonal son borrados al iniciarse la Fase Final (o sea, los 8 equipos empiezan esta fase con 0 puntos). Para la Fase Final, si bien no se sabe qué equipos la disputarán, ya que depende del resultado de la fase anterior, sí se conoce que serán los cuatro primeros de ambos grupos. Esto permite generar de antemano un fixture que incorpora, entre otras, restricciones de carácter geográfico. La razón por la cual la ANFP solicita hacer también esta programación antes de iniciar todo el torneo es que la Fase Final comienza unos pocos días después de finalizadas las fases zonales.

Una condición adicional que la ANFP nos solicitó en este torneo fue que las fechas en que algunos equipos juegan de local o visita dependían del fixture de partidos de los equipos de Primera División que juegan en las mismas ciudades, con el fin de que hubiese un balance de localías en ellas. Esta finalidad es similar a lo que buscan las restricciones de complementariedad, aunque en este caso es importante notar que el fixture de Primera División es programado antes y en forma independiente del de Segunda División. En total, en la temporada 2010 se juegan 224 partidos en un total de 42 fechas.

Para todas las temporadas entre 2007 y 2010, la meta del proyecto ha sido encontrar un fixture que satisfaga todas las restricciones requeridas. El problema principal es entonces netamente de factibilidad. En el Torneo 2010, además de satisfacer las condiciones, la ANFP nos solicitó como medida deseable que

en la Fase Final concentráramos el mayor número posible de *partidos atractivos* en las tres últimas fechas del torneo, entendiendo por partido atractivo a uno que enfrente a los equipos que salieron primeros o segundos en cada Grupo Zonal. Para efectos de la modelación, en el Torneo 2010 incorporamos entonces como función objetivo maximizar la cantidad de partidos atractivos en las tres últimas fechas.

---

## 4. Métodos de resolución

---

Abordamos el problema mediante un modelo de programación lineal entera, que llamaremos *Match Scheduling Model* (MS-Model). La variable de decisión principal es una variable binaria  $x_{i,j,k}$  que toma el valor 1 si el equipo  $i$  juega de local contra el equipo  $j$  en la fecha  $k$ . Su implementación y resolución las hemos realizado año a año utilizando el solver de ILOG, CPLEX.

En la práctica, la programación de los torneos cuádruple *round robin* doblemente espejados en 2007 y 2008 no fue particularmente difícil. Mediante una corrida del MS-Model encontramos una solución factible en sólo un par de minutos, lo que creemos se explica por dos razones. Primero, el número de equipos (once o doce) es relativamente bajo, y por lo tanto el tamaño del problema es reducido. Segundo, las condiciones de espejado y simetría reducen significativamente el espacio de búsqueda. En efecto, es posible separar el problema, de tal manera que generar un fixture para el primer cuarto es suficiente para determinar el fixture del torneo completo. Por ejemplo, una condición para una fecha  $k_2$  del segundo cuarto puede ser expresada en términos de su fecha correspondiente en el espejo del primer cuarto, digamos  $k_1$ , teniendo en cuenta que, debido a la condición de espejado,  $x_{i,j,k_1} = x_{j,i,k_2}$ . A su vez, una condición en una fecha  $k_3$  del tercer cuarto puede ser expresada en términos de su fecha análoga en el primer cuarto, teniendo en cuenta que, debido a las condiciones de simetría,  $x_{i,j,k_1} = x_{i,j,k_3}$  (la relación es análoga para fechas en el segundo y cuarto cuarto).

La programación del torneo 2010 tampoco fue difícil de lograr. El tipo de torneo permite separar la programación de la Fase Zonal en dos sub-problemas independientes cuya dimensión es relativamente baja (hay sólo siete equipos en cada grupo). La programación de la Fase Final fue programada en forma independiente de la Fase Zonal y es menos restricta, dado que los equipos que la juegan no se conocen *a priori*. Su dimensión es también relativamente baja (la juegan ocho equipos). La resolución a optimalidad con la función objetivo definida en la sección anterior (o sea, con la programación de la mayor cantidad de partidos considerados atractivos en las 3 últimas fechas), se consigue en cuestión de segundos.

En cambio, los dos torneos de 2009 fueron más difíciles de programar, debido a la dependencia entre la Fase Zonal y la Fase Nacional. Consideramos distintas alternativas de abordar el problema.

Una primera alternativa es dividir el problema en dos sub-problemas, uno para resolver la Fase Zonal y otro para resolver la Fase Nacional por separado. La ventaja de esta división es que tenemos que abordar problemas más pequeños. El problema de la Fase Zonal a su vez puede ser descompuesto en dos, según los grupos en que se dividen los equipos. Para cada grupo, el MS-Model tiene alrededor de 500 variables binarias y 500 restricciones, mientras que el problema de la Fase Nacional tiene alrededor de 3,000 variables binarias y 1,500 restricciones. El inconveniente de este enfoque es que los sub-problemas no son independientes, hay que conectarlos, debido a la restricción de invertir localías para los partidos entre equipos  $i$  y  $j$  que se enfrentan en ambas fases; a la condición que indica el máximo número de *breaks* por equipo (notar que un *break* de visita, por ejemplo, puede ocurrir cuando un equipo juega de visita la última fecha de la Fase Zonal y la primera fecha de la Fase Nacional); y a la condición de no jugar 3 partidos de local o visita consecutivos (como en el caso anterior, podría pasar por ejemplo que un equipo termine la Fase Zonal con un *break* de local y empiece la Fase Nacional nuevamente como local). En esta alternativa de dividir el problema en dos sub-problemas, debemos optar por resolver primero la Zonal y luego la Nacional, o al revés. Una solución para una de las fases podría llevar a infactibilidad para la otra, sin que necesariamente el problema completo sea infactible.

Una segunda alternativa es resolver la Fase Zonal y la Fase Nacional al mismo tiempo. La dimensión del modelo entero en este caso es bastante más grande, con alrededor de 5,000 variables binarias y 2,500 restricciones. Sin embargo, la ventaja de esta alternativa es que, si encontramos una solución factible, es una solución para el problema completo, sin necesidad de tener que conectar soluciones de dos sub-problemas.

Para ninguna de las dos alternativas logramos encontrar soluciones factibles utilizando directamente la formulación del MS-Model por sí solo, ni con los procedimientos básicos de asignación de patrones que utilizábamos para el torneo de Primera División (ver la sección de “computational solution” de [4]).

En este trabajo para la Segunda División desarrollamos entonces otro modelo de programación lineal entera, mediante el cual generamos los patrones de localía. Lo denotamos como *Pattern Generating Model* (PG-Model).

La programación del torneo la realizamos entonces en base a dos modelos que se resuelven en etapas sucesivas. Primero, el PG-Model construye y asigna los patrones de localías de cada equipo, considerando parcialmente las restricciones del problema. Luego, el MS-Model decide qué equipos se enfrentan en cada fecha, considerando que todas las restricciones del problema se cumplan.

Para esta segunda etapa, condiciones como las restricciones de complementariedad y las de patrones de localías ya están aseguradas por la resolución del PG-Model, por lo tanto, las corridas del MS-Model pueden ser simplificadas omitiendo estas restricciones. Este tipo de descomposición en etapas es común en los problemas de programación en deportes, aunque las metodologías para su implementación utilizadas en la literatura son variadas (una revisión profunda a este respecto puede ser encontrada en la sección 4 de [15]).

Cuando implementamos ambos modelos logramos encontrar soluciones factibles para ambas alternativas de enfoque del problema de la Temporada 2009. Para la alternativa de resolver las Fases por separado, obtener soluciones para la Fase Zonal usando el MS-Model, se logra en cuestión de segundos. Luego resolvimos la Fase Nacional usando la asignación de patrones obtenida al resolver el PG-Model, y después aplicando una versión simplificada del MS-Model sólo con las restricciones aún no aseguradas por la asignación de patrones lograda a través del PG-Model. De este modo, encontramos una solución factible para la Fase Nacional en alrededor de 30 minutos. Esta solución de la Fase Nacional conectada con las soluciones previas de las Fases Zonales constituyen el fixture final del torneo. Creemos interesante comentar que mediante el proceso inverso no conseguimos buenos resultados. Es decir, cuando resolvimos primero la Fase Nacional (también usando el PG-Model), y luego intentamos buscar soluciones a la Fase Zonal que invirtiera las localías de la solución que obteníamos para la Fase Nacional, no conseguimos una solución factible. La razón de esto probablemente sea que al fijar la solución encontrada para la Fase Nacional (que es la más grande) restringía exageradamente el espacio de búsqueda para la Fase Zonal, a tal punto que no logramos conseguir soluciones factibles por este mecanismo.

Para la alternativa de resolver la Fase Zonal y Nacional simultáneamente, la implementación del PG-Model y del MS-Model nos condujeron a encontrar soluciones factibles en alrededor de 10 minutos. En la práctica, este tiempo de resolución ha sido funcional y, por lo tanto, este es el mecanismo que utilizamos para generar los distintos fixtures que se le propusieron a la ANFP para el torneo de 2009. Notar que en la programación del fixture para un torneo dado realizamos un proceso de iteración con la ANFP que dura al menos un par de semanas. En este proceso, surgen una serie de condiciones que originalmente ni ellos ni los equipos tenían consideradas, por lo que la construcción del fixture envuelve varias corridas de los modelos.

A partir del PG-Model que implementamos para poder resolver el problema de la Segunda División 2009, desarrollamos una versión similar para programar el fixture de Primera División, obteniendo significativas mejoras en los tiempos de las corridas. De este modo, esta metodología de resolución por medio de la aplicación de dos modelos sucesivos está en uso en el fixture de la Primera División desde 2010.

---

## 5. Impacto y conclusiones

---

Los fixtures generados mediante la metodología presentada en este artículo han sido utilizados en los cinco torneos de Segunda División del fútbol chileno que se han disputado entre 2007 y 2010. La aplicación ha generado un impacto positivo tanto a nivel cuantitativo como cualitativo.

A pesar de que la medición del impacto cuantitativo no es directo y es prácticamente imposible aislarlo de efectos exógenos, algunas observaciones pueden ser realizadas. Un factor relevante, que ha significado ahorros para los equipos debido a un mejor fixture, ha sido la inclusión de viajes buenos para los equipos de los extremos del país. En cada torneo hubo alrededor de cinco viajes buenos, excepto en 2009 en que todas las rondas se jugaron en fines de semana. Según estimaciones de la ANFP, cada viaje bueno significa a un equipo ahorrar aproximadamente US\$4,000. Este ahorro es cerca del 13 % de la planilla mensual de un equipo de Segunda División, que ronda los US\$30,000, y del 5 % de la recaudación anual por concepto entradas de equipos como Puerto Montt, del extremo sur del país, que en sus partidos de local percibe en promedio unos US\$75,000 por año.

Por su parte, el balance de los viajes a lo largo del torneo permite a los equipos una mejor organización de sus gastos a través del año. El distanciamiento de viajes largos es un aporte a la distribución mensual de los recursos de los equipos, sobre todo para aquellos cuyo presupuesto es más ajustado y dependiente de ingresos por partido de local y contratos de auspicios cuyo pago se concreta mensualmente.

Los cambios realizados por la ANFP al tipo de torneo en los últimos dos años han sido diseñados para que, a la vez de ofrecer un torneo atractivo para los simpatizantes, los equipos realicen menos viajes de un extremo del país al otro, ahorrando costos y permitiendo que los jugadores se desgasten menos. El apoyo a la generación del fixture mediante el uso de modelos de IO han facilitado la introducción de dichos cambios.

Un estudio interesante sobre formatos de ligas de fútbol ha sido reportado en [9], en el contexto de la liga de fútbol belga. En el caso de Chile, los sucesivos cambios al formato del torneo de Segunda División pueden abrir una buena oportunidad de estudio, dado que la ANFP aún no está convencida de cuál es el formato ideal. Al menos en lo que respecta a la programación del torneo, para nosotros ha resultado desafiante enfrentarnos a un nuevo formato de torneo cada año.

Nos parece pertinente comentar que la temporada de Segunda División 2007, primer torneo programado con nuestro apoyo, vio un aumento en la

asistencia de público promedio por partido de aproximadamente 10% sobre la temporada 2006, cuando la ANFP aún utilizaba su metodología tradicional de programación aleatoria. No podemos atribuir la plenitud de este aumento a un mejor fixture, pero creemos que sí contribuye a explicarlo. Por ejemplo, en el Torneo 2007 un importante partido clásico de la Segunda División disputado por los equipos Rangers de Talca y Curicó Unido, fue programado para la última fecha de cada cuarto. Esta solicitud nos la realizó la ANFP pensando que el atractivo de este partido sería mayor si se jugaba en una fecha avanzada del torneo, en que podría tornarse decisivo. En efecto, sucedió que Rangers disputó de local el último partido del torneo 2007 con buenas chances de ascender a Primera División y lograrlo frente a su rival más clásico imprimió un sabor especial al partido. A dicho partido asistieron 10,006 simpatizantes, muy por sobre la asistencia promedio de ese año que Rangers tuvo como local (2,761 simpatizantes/partido, sin contar este último encuentro). La recaudación correspondiente a este partido fue de US\$57,600, más de siete veces los US\$8,000 promedio por partido que Rangers recaudó como local en esa temporada. Aun más, este fue el segundo partido de mayor recaudación de toda la liga. La situación de Rangers en el campeonato fue causa importante de estos números, pero creemos que también lo fue el rival clásico que tenía que enfrentar. Difícilmente la asignación aleatoria de antaño hubiera contribuido al espectáculo y a las finanzas de esta manera.

Por otro lado, el impacto cualitativo del proyecto se ha manifestado en varios aspectos. Debido a que los requerimientos incorporados son conocidos por todos los equipos, el proceso de generación del fixture es más *transparente*. Como la ANFP considera peticiones de todos los equipos para incorporarlas a la programación, el sistema es *justo*. La adaptación de nuestros modelos a los distintos cambios de los torneos organizados por la Segunda División desde 2007, da cuenta de la *flexibilidad* de nuestro enfoque. La programación apropiada de los partidos también genera un mayor *atractivo* de los torneos, lo cual es bien percibido por los simpatizantes. Los partidos clásicos programados en fechas apropiadas otorgan una motivación adicional no sólo a los simpatizantes, sino también a los jugadores. También la metodología contribuye a la *seguridad*, incorporando aspectos como que un equipo juegue de visita en una fecha en que hay otros eventos masivos en su ciudad. Por ejemplo, la cuarta fecha del Torneo 2009 coincidió con el festival musical en Viña del Mar, cuya seguridad demanda una alto volumen de dotación policial en la zona. Consideramos en las restricciones de localía-visita entonces como condición que los cuatro equipos que juegan en la zona de influencia de esa ciudad no fueran programados para jugar de local en dicha fecha.

En general, los gerentes de la ANFP y de los equipos, los simpatizantes y la prensa han recibido positivamente esta aplicación, por lo cual nuestra metodología goza de credibilidad en el medio futbolístico. Desde que las técnicas

de IO han sido aplicadas al torneo de Segunda División en 2007, la ANFP no recibe quejas de los equipos sobre el fixture.

En la realización de este trabajo ha sido crucial nuestra alianza con los gerentes de la ANFP. Durante la ejecución del proyecto, antes de cada torneo, la inclusión de condiciones al modelo es realizada en un trabajo muy cercano a ellos. Durante estos años han entendido perfectamente que un modelo de Investigación de Operaciones es un apoyo a la toma de decisiones sobre la programación del fixture, cuyo uso no les hace perder control sobre la programación de partidos; todo lo contrario, les permite incorporar condiciones que de acuerdo a su metodología anterior no podían considerar. Este entendimiento ha sido un aspecto clave para el éxito del proyecto y su perduración en el tiempo. Creemos que este es uno de los factores más difíciles de lograr en la puesta en uso de un modelo de programación en deportes, quizás aun más complicado que la resolución de los problemas técnicos propiamente tal. En acercamientos preliminares que hemos tenido con otras asociaciones deportivas, la predisposición negativa al cambio por parte de los tomadores de decisiones es la principal barrera a la incorporación de nuevas metodologías. Por el contrario, los gerentes de la ANFP nos han comentado que actualmente no se imaginan realizar el fixture de sus torneos sin el apoyo de modelos de IO.

Esta aplicación en la Segunda División surgió como extensión natural del exitoso desarrollo que hemos estado realizando junto a la ANFP para el fixture de la Primera División, para la cual hemos programado los fixtures de los 12 torneos disputados entre 2005 y 2010. “Nuestra satisfacción con estos modelos que apoyan la programación de nuestros partidos es altísima; esperamos en el futuro continuar con la utilización de estos sistemas para ambos campeonatos, Primera y Segunda División”, expresa Harold Mayne-Nicholls, Presidente de la ANFP.

Destacamos que el uso de técnicas de IO se ha extendido a lo largo de los últimos años a distintas áreas de la gestión del fútbol chileno. A la programación de ambos campeonatos profesionales mediante técnicas de programación en deportes, se suma el uso de herramientas similares para la generación del fixture de tres torneos juveniles entre 2005 y 2010, y para la propuesta de fixture para la Eliminatoria Sudamericana del Mundial 2010 que la ANFP llevó a la Confederación Sudamericana de Fútbol (CONMEBOL) en 2007. Los buenos resultados obtenidos en estos proyectos nos han llevado recientemente a trabajar junto a la ANFP en el desarrollo de un modelo de optimización para una asignación más justa y eficiente de árbitros a los partidos de fútbol, que está siendo testeado en el torneo de Segunda División y en los torneos juveniles de 2010 [1], con la idea de aplicarlo a futuro también en la Primera División.

Asimismo, la experiencia en programación en deportes en Chile nos permitió extender su aplicación a Argentina. Desde 2007 se ha programado la liga

profesional de vóley de Argentina utilizando herramientas similares [3], mientras que ya se han iniciado conversaciones con las asociaciones respectivas de dicho país para programar las ligas de fútbol y basquet.

**Agradecimientos:** Agradecemos a la Asociación Nacional de Fútbol Profesional de Chile, y en particular a su Presidente Harold Mayne-Nicholls y su gerente de competiciones René Rosas. Este proyecto no hubiera sido posible sin su apoyo. Agradecemos también a Andrés Weintraub, quien nos motivó a escribir este artículo y a Fernando Alarcón, por sus interesantes comentarios que permitieron mejorar este trabajo. Este proyecto está parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería. El primer autor es parcialmente apoyado por los subsidios ANPCyT PICT-2007-00518 y UBACyT X069 (Argentina); y por el subsidio FONDECyT 1080286 (Chile).

## Referencias

- [1] Alarcón, F., G. Durán, M. Guajardo. 2009. Un modelo de asignación de árbitros para el torneo de fútbol chileno y un enfoque de resolución en base a patrones. *Revista Ingeniería de Sistemas* 23:125–143.
- [2] Bartsch, T., A. Drexl, S. Kroger. 2006. Scheduling the professional soccer leagues of Austria and Germany. *Computers and Operations Research* 33(7):1907–1937.
- [3] Bonomo, F., A. Cardemil, G. Durán, J. Marengo, D. Saban. 2010. Scheduling the Argentine volleyball league: A real-world application of the Traveling Tournament Problem with couples of teams. Submitted to *Interfaces*.
- [4] Durán, G., M. Guajardo, J. Miranda, D. Sauré, S. Souyris, A. Weintraub, R. Wolf. 2007. Scheduling the Chilean Soccer League by Integer Programming. *Interfaces* 37(6):539-552.
- [5] Fiallos, J., M. Licona, J. Pérez, F. Sabillón. 2010. Personal communication.
- [6] Flatberg, T., E. J. Nilssen, M. Stlevik. 2009. Scheduling the topmost football leagues of Norway. *23rd European Conference on Operational Research, Book of Abstracts*. Bonn, Germany, p.240.
- [7] Froncek, D., M. Meszka. 2010. Personal communication.
- [8] Goossens, D., F. Spieksma. 2009. Scheduling the Belgian Soccer League. *Interfaces* 39(2):109–118.

- [9] Goossens, D., J. Belin, F. Spieksma. 2008. Comparing league formats with respect to match unimportance in Belgian football. *Proceedings of the 7th International Conference on the Practice and Theory of Automated Timetabling (PATAT)*. Montreal, Canada.
- [10] Kendall, G., S. Knust, C. C. Ribeiro, S. Urrutia. 2009. Scheduling in sports: An annotated bibliography. *Computers and Operations Research* 37(1):1–19.
- [11] Meszka, M. 2010. Personal communication.
- [12] Nurmi K., T. Bartsch, F. Bonomo, D. Briskorn, G. Durán, D. Goossens, J. Kyngas, J. Marenco, C. Ribeiro, F. Spieksma, S. Urrutia, R. Wolf. 2010. A Framework for a Highly Constrained Sports Scheduling Problem. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol III, IMECS 2010, March 17-19*, Hong Kong, 1991-1997.
- [13] Paenza, A. 2006. *Matemática...Estás Ahí? Episodio 2 Siglo XXI*, Buenos Aires, Argentina, 194-206.
- [14] Rasmussen, R. V. 2008. Scheduling a triple round robin tournament for the best Danish soccer league. *European Journal of Operational Research* 185(2):795–810.
- [15] Rasmussen, R. V., M. A. Trick. 2008. Round robin scheduling - a survey. *European Journal of Operational Research* 188(3):617-636.
- [16] Ribeiro, C., S. Urrutia. 2009. Scheduling the Brazilian soccer tournament by integer programming maximizing audience shares under fairness constraints. *23rd European Conference on Operational Research, Book of Abstracts*. Bonn, Germany, p.240.
- [17] Schreuder, J. 1992. Combinatorial aspects of construction of competition Dutch professional football leagues. *Discrete Applied Mathematics* 35:301–312.

---

# APLICACIÓN DE TÉCNICAS DE WEB MINING SOBRE LOS DATOS ORIGINADOS POR USUARIOS DE PÁGINAS WEB. VISIÓN CRÍTICA DESDE LAS GARANTÍAS FUNDAMENTALES, ESPECIALMENTE LA LIBERTAD, LA PRIVACIDAD Y EL HONOR DE LAS PERSONAS

---

JUAN D. VELÁSQUEZ\*  
LORENA DONOSO\*\*

## Resumen

*Web mining es la aplicación del data mining a los web data para la extracción y descubrimiento automático de información y conocimiento. Dependiendo del tipo de web data a procesar, web mining se divide en tres grandes categorías: contenido, estructura y uso. El análisis de estos datos permite a las instituciones significativas mejoras en la estructura y contenido de los sitios web corporativos, así como la aplicación de complejos sistemas informáticos destinados a personalizar la experiencia del usuario en el sitio que visita. El presente artículo muestra una revisión científico-técnica de los fundamentos del web mining, sus principales técnicas, métodos y algoritmos, con especial atención en aquellos que permiten extrapolar las preferencias de navegación y contenidos de los usuarios que visitan un sitio web determinado, para finalmente contrastar su operación con la regulación vigente a nivel nacional e internacional.*

**Palabras Clave:** *Web Mining, Privacidad, Regulación.*

---

\*Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

\*\*Centro de Derecho Informático, Facultad de Derecho, Universidad de Chile, Santiago, Chile.

---

## 1. Introducción

---

La World Wide Web o simplemente **La Web** [1] es tal vez el mayor portento tecnológico que el hombre haya desarrollado jamás. Su impacto en nuestra sociedad ha sido tal que se le ha comparado con la invención de la rueda o el descubrimiento del fuego.

Desde los orígenes de la Web, la creación de un sitio no ha sido un proceso fácil. Muchas veces se requiere de un equipo multidisciplinario de profesionales abocados a una sola misión: asegurar que el contenido y la estructura del sitio le son atractivos al usuario. Lo anterior se ha abordado con relativo éxito en el ámbito de la “*personalización de la Web*”, concepto que es la clave del éxito para obtener una adecuada participación en el mercado electrónico, mantener la vigencia del sitio y sobre todo, lograr la tan ansiada y difícil fidelización del cliente digital.

La personalización implica que de alguna forma se puede obtener información respecto de los deseos y necesidades de las personas, para luego preparar la oferta correcta en el momento correcto [7]. Lo anterior plantea la necesidad de efectuar estudios previos para analizar la respuesta del consumidor ante un determinado estímulo, por ejemplo, los muy utilizados “*focus group*”, donde un grupo de personas, que son la muestra representativa de un conjunto mayor, entrega su opinión respecto de lo que percibe en un producto o servicio.

Pensando en una esquema como el anterior, tal vez la solución para entender mejor al cliente digital sería someterlo a varias encuestas de opinión vía e-mail o al llenando formularios electrónicos. Sin embargo, la práctica ha demostrado que los usuarios no gustan de llenar formularios, contestar e-mails con preguntas, etc., a menos que se trate de algún amigo o familiar que quiera ayudar en el análisis, lo cuál no sería un caso real.

Los datos originados en la Web o web data, prácticamente corresponden a todos los datos que se han originado a lo largo de la historia de la computación. En efecto, aquí se encuentran los hipervínculos entre páginas web y sus los contenidos, que pueden ser imágenes, sonidos, videos, texto libre, etc. A lo anterior, se debe agregar datos acerca de la navegación del usuario en los sitios que visita, específicamente la IP desde donde accedió y el tipo de navegador utilizado.

La ley 19.628 sobre protección de la vida privada, establece ciertas restricciones al procesamiento de datos personales, por lo que de entrar los web data en esta categoría, es importante analizar hasta que punto su procesamiento está conforme a la regulación vigente. Si adicionalmente se consideran otros datos que los mismos usuarios pueden develar en blogs, foros o sistemas similares, tales como vinculaciones políticas, vida sexual, origen racial, ideologías

o convicciones religiosas, etc. los web data también entrarán en la categoría de datos sensibles, según lo consigna la letra “g” del artículo 2º de la mencionada ley.

En esencia, los algoritmos, técnicas y métodos que comprende el web mining, son utilizados en el procesamiento masivo de datos, lo cual requiere una automatización parcial o total de todas las operaciones a fin de obtener resultados en cuestión de horas o días. En consecuencia, el análisis de los web data utilizando técnicas de web mining cuenta con todos los requisitos necesarios para ser estudiando a partir de la regulación nacional e internacional que hasta el momento se ha desarrollado. En particular, es de suma importancia revisar ¿hasta dónde este afán por analizar al usuario en la Web no se transforma en una persecución? [2].

El presente artículo comienza con la sección 2, la cual aborda el fenómeno Internet y Web desde sus orígenes hasta nuestros días, explicando a grandes rasgos su funcionamiento, los datos que se pueden recolectar y cuales estarían directamente relacionados con información relativa a las personas naturales. A continuación, en la sección 3, se analizan la operación de las técnicas, algoritmos y metodologías propuestas en web mining, del punto de vista de la restricción a la libertad y vulneración de la privacidad de los usuarios de sitios web.

Por su parte, la sección 4 profundiza en los aspectos jurídicos relacionados con el tratamiento de los datos originados en la Web, también conocidos como Web Data. Finalmente, en la sección se presentan las principales conclusiones y recomendaciones que se han obtenido a lo largo de este trabajo.

---

## 2. Internet y la Web

---

Es importante hacer la distinción entre la Web e Internet, ya que son conceptos distintos pero que a menudo se confunden. Internet representa a la red de redes que permite la interconexión de dispositivos que se encuentran a nivel local, con sus similares en una región diferente, a través del envío y recepción de los datos que viajan en paquetes o datagramas. La Web es el conjunto de páginas y objetos relacionados que se vinculan entre si a través de hipervínculos. A un conjunto de páginas web se le denomina sitio web y es administrado por una aplicación conocida como servidor web, la cual utiliza a Internet como lugar físico para transferir las páginas web y otros objetos asociados. De acuerdo a la definición dada por su creador Tim Berners-Lee en 1989, la “*World Wide Web es el universo de información accesible en la red, una encarnación del conocimiento humano*” [1].

## 2.1. Datos originados en la Web

La Web es el conjunto de archivos (páginas) que se relacionan a través de hipervínculos, almacenados en los servidores ubicados alrededor del mundo, para lo cual se utiliza un mecanismo de direccionamiento global de documentos y de otros recursos conocido como URL. Cada una de estas páginas posee un contenido representado a través de objetos como texto, imágenes, sonidos, películas o vínculos a otros sitios web.

La Fig. 1 muestra en forma simple el funcionamiento de la Web. El servidor web o web server (1) es un aplicación que está en ejecución continua, atendiendo requerimientos (4) de objetos web, es decir, el conjunto de archivos que conforman el web site (3) y los envía (2) a la aplicación que hace la solicitud, generalmente un web browser (6). En general estos archivos son imágenes, sonidos, películas y páginas web que conforman la información visible del sitio. Las páginas están escritas en Hyper Text Markup Language (HTML), que en síntesis es un conjunto de instrucciones, también conocidas como tags (5), acerca de cómo desplegar objetos en el browser o dirigirse a otra página web (hyperlinks). Estas instrucciones son interpretadas por el browser, el cual muestra los objetos en la pantalla del usuario [3].

Cada uno de los tags presentes en una página, son interpretados por el browser. Algunos de estos hacen referencia a otros objetos en el web site, lo que genera una nueva petición en el browser y la posterior respuesta del server. En consecuencia, cuando el usuario digita la página que desea ver, el browser, por interpretación secuencial de cada uno de los tags, se encarga de hacer los requerimientos necesarios que permiten bajar el contenido de la página al computador del usuario.

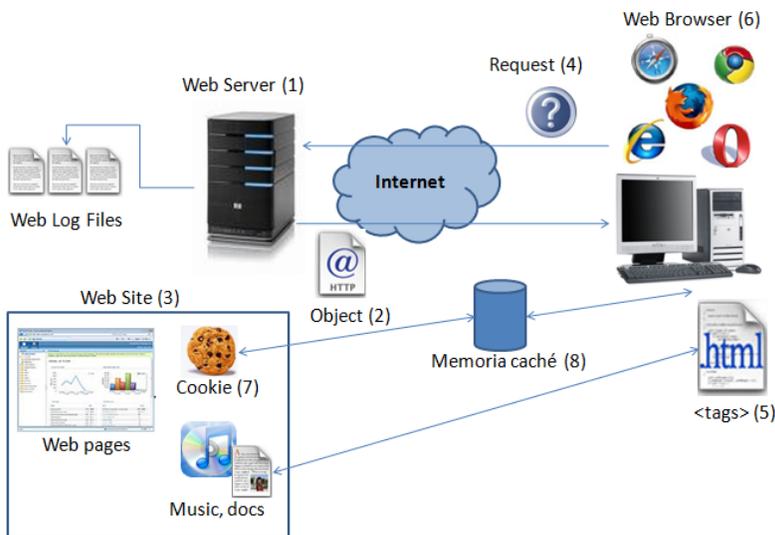


Figura 1: Modelo básico de operación de la Web

La interacción anterior, ha quedado registrada en archivos conocidos como web log files [3], con lo cual es posible saber aproximadamente qué objetos fueron requeridos por un usuario, reconstruir su sesión y en la práctica realizar un verdadero seguimiento a sus actividades de navegación, analizando los contenidos visitados, el tiempo que se ha invertido en ello, qué información no atrae su interés, etc. La Fig. 2 muestra un ejemplo del contenido y estructura de un archivo de web log.

N°	IP	ID	Access	Time	Method/URL/Protocol	Status	Bytes	Referer	Agent
1	164.77.129.50	-	-	12/Apr/2003:23:47:44	GET /img/tab.gif HTTP/1.1	200	89	http://www.thebank.cl	MSIE 6.0; Windows 98
2	200.28.206.200	-	20	12/Apr/2003:23:48:31	GET transa/info.htm HTTP/1.1	200	144	/infoeco/info.html	MSIE 4.0.1; Windows 95
3	200.86.248.170	-	-	12/Apr/2003:23:48:37	GET /img/gen.gif HTTP/1.1	304	0	/ofert/wines/	MSIE 6.0; Windows 98
4	66.249.65.97	-	-	12/Apr/2003:23:48:41	GET /index.htm HTTP/1.1	200	88	-	Googlebot/2.1; google.com/bot.html
5	216.241.8.179	-	31	12/Apr/2003:23:50:03	GET /tx/infoeco/card.htm HTTP/1.1	200	210	/tx/infoeco/prom/	MSIE 6.0; Windows NT 5.1
6	164.77.129.50	-	-	12/Apr/2003:23:48:34	GET /tx/infoeco/ HTTP/1.1	200	186	/tx/infoeco/card.htm	MSIE 6.0; Windows 98
7	200.28.206.200	-	20	12/Apr/2003:23:51:13	GET transa/account.htm HTTP/1.1	200	180	/transa/info.htm	MSIE 4.0.1; Windows 95
8	216.241.8.179	-	31	12/Apr/2003:23:51:23	GET /tx/infoeco/ind.htm HTTP/1.1	200	300	/tx/infoeco/card.htm	MSIE 6.0; Windows NT 5.1
9	200.86.248.170	-	-	12/Apr/2003:23:51:41	GET /prom/wine.html HTTP/1.1	404	0	/ofert/wines/	MSIE 6.0; Windows 98
10	164.77.129.50	-	44	12/Apr/2003:23:52:04	GET /tx/infoeco/ind.htm HTTP/1.1	200	186	/tx/infoeco/	MSIE 6.0; Windows 98

Figura 2: Estructura de un web log file

La estructura estándar del archivo web log queda definida entonces por la dirección IP del dispositivo desde donde se solicita la página, los parámetros IP y Access que permiten una forma de identificación de la sesión, time que especifica la fecha y momento en que se hizo el requerimiento, el tipo de método utilizado para solicitar la página, el estado (status) de la solicitud, la cantidad de bits transmitidos, la URL de la última página visitada y el tipo de aplicación utilizado para navegar (Agent)

Otra forma de capturar el comportamiento del usuario en una sesión es a través de las cookies (7) que se almacenan en el disco duro del cliente a través del Browser o Navegador, y que guardan parte de la información de la página que visitó, usando la memoria de rápido acceso o memoria caché (8). Estas cookies son generadas a pedido del servidor, y se utilizan tanto para el control de usuarios, por ejemplo cuando se pide una contraseña en algún sitio, como para ver el comportamiento de navegación de los usuarios. Hay que dejar en claro que este mecanismo no identifica a una persona en particular, sino a un tipo de usuario que navega en un sitio web en un determinado Browser.

Como se puede ver, cada registro da cuenta de los movimientos de un usuario en un sitio web. En consecuencia, y en forma casi anónima, los datos generados en el sitio web son tal vez la mayor encuesta que podría tener una empresa por sobre sus eventuales clientes, permitiendo un análisis profundo de sus preferencias de información, las cuales están directamente relacionadas con las características de los productos y servicios ofrecidos.

## 2.2. Otros datos presentes en la Web

Para acceder a los datos existentes en la Web, primero se requiere una conexión a Internet, la cual se realiza a través de diversos métodos, siendo

uno de los más populares el de la contratación de un Internet Service Provider (ISP), empresas que utilizando un medio de comunicación, como puede ser el espectro radioeléctrico, TV-cable, telefonía convencional, etc. brindan el acceso de un abonado a la gran red [12].

Independiente del método, toda navegación o interacción del usuario con la Web, se realiza a partir del envío de paquetes de datos, que por construcción poseen la dirección IP de origen y destino de la comunicación. Esta estructura permite un monitoreo muy eficaz y eficiente de las sesiones que establecen los usuarios que navegan por la Web.

Variados son los dispositivos y aplicaciones usados para brindarle conectividad a los usuarios y también para restringir su navegación en la Web. Algunos de estos son:

- Router: Dispositivo por excelencia que permite la conectividad de una Red de Área Local (LAN) a la Red Externa (WAN).
- Proxy. Se trata de un espacio de memoria, comúnmente discos duros en un computador dedicado, para el almacenamiento de todas las páginas web que se visitan desde una red local.
- Firewall. El dispositivo de seguridad por excelencia de una red local. Se trata de un conjunto de aplicaciones de seguridad que se ejecutan preferentemente en un computador dedicado a estos menesteres, que por arquitectura aísla a la red local de la Internet.

Los dispositivos antes analizados, permiten recopilar datos sobre toda la navegación que los usuarios realicen en la Web, con lo cual es posible realizar un trabajo muy profundo de monitoreo de sus actividades. Sin embargo, estos datos quedan en posesión del dueño del dispositivo, es decir, si se desea acceso a la navegación de los usuarios que tienen contratado un determinado ISP, habrá que solicitar estos datos al dueño del servicio, permiso que, salvo una orden judicial, es muy difícil de obtener.

En vista de estas restricciones, se han desarrollado otros métodos para recolectar parte de los datos originados en la Web:

- Web Crawlers. También conocido como *Spyder Robots*, son aplicaciones que recorren la Web de forma automática y sistemática, almacenando información de los sitios web mediante el uso de sus hipervínculos.
- Spyware: Es una aplicación cuya función es espiar las actividades de navegación de un usuario desde un computador conectado a Internet. Esta aplicación, claramente no se instala con la venia de los usuarios, lo cual la transforma en un ente extraordinariamente invasivo, amén que ocupa los recursos del computador para su funcionamiento (RAM,

CPU, etc.). Su instalación se realiza cuando el usuario sin saber navega por sitios que contienen spywares.

Estos métodos se orientan en parte a capturar toda la información que se pueda de las transacciones y navegación que realiza un usuario. De un punto de vista comercial, entender el comportamiento de compra de los clientes a través de la Web es vital, en términos de mejorar el contenido del sitio utilizado para el negocio, ya sea Business to Business (B2B), Business to Consumer (B2C), Peer to Peer (P2P) o cualquier otra variación que requiera del análisis de los registros de un visitante vía web.

En el caso particular de los usuarios que realizan compras por Internet, generalmente se almacenan datos personales, ya sea de una cuenta de usuario o tarjetas que permitan realizar este tipo de compras. Las personas han hecho uso gradual de este medio como un sistema de compras, dada la inseguridad (intangibilidad del producto), verificación de la compra, detalles del producto a entregar, plazos, etc. Por esto es que la idea es adaptarse a las necesidades de los clientes, haciéndolos sentir que la compra es personalizada (por ejemplo se puede hacer sugerencias basándose en el historial del cliente), fácil, rápida y que agrega valor. Este último es muy importante por ejemplo para las empresas, ya que a través de Internet se ha eliminado parte de la cadena de suministro. Un ejemplo claro es Amazon, ya que no tiene necesidad de tener un inventario en exhibición de sus libros, acortando la cadena, eliminando a los elementos que no aportan y disminuyendo de esta forma los costos asociados. Otro ejemplo son las transacciones que realizan los bancos e instituciones financieras, que al ser vía electrónica ahorran en impresiones, personal y tiempo tanto para la empresa como para los clientes.

---

### 3. La dirección IP como dato personal

---

Hemos dicho que en la ley se define datos personales como cualquier información relativa a personas naturales identificadas o identificables. Al respecto cabe señalar que esta definición es similar a las que se han adoptado en otros países, por lo que nos podrá ayudar a dilucidar si una dirección IP puede ser considerada un dato personal no obstante no haya pronunciamientos expresos en nuestra legislación o jurisprudencia.

Conceptualizando una dirección IP, la RFC 791<sup>1</sup>, la dirección Internet es “una dirección de origen o destino de 4 octetos (32 bits) formada por un campo de Red y un campo de Dirección Local”. Ello nos denota que la finalidad del establecimiento de estas direcciones es el reconocer máquinas interconectadas a través del protocolo IP, y no necesariamente a las personas que están

---

<sup>1</sup><http://www.rfc-es.org/rfc/rfc0791-es.txt> [consulta: 31.03.2010]

operándolas. Sin embargo, con la masificación de Internet y su desarrollo como sistemas de intercambio de la información, se ha cuestionado la posibilidad de que estas direcciones sean consideradas un dato personal.

Una primera aproximación al respecto nos llama a reparar en que la definición de dato personal exige que la persona a quien se refiere el dato personal sea al menos *“identificable?”* esto es, que exista la posibilidad de identificarla, sin importar las dificultades técnicas y/o económicas que la determinación implique. Asimismo, se ha sostenido que la identificabilidad es un atributo cambiante en el tiempo, en donde tiene mucho que decir el avance científico y tecnológico. En efecto, hace unos años atrás no era posible identificar a una persona específica a través de una muestra biológica, en cambio hoy en día es perfectamente factible hacerlo. Con esto queremos señalar que un dato, por ejemplo de dirección IP, si hoy, por las condiciones de mercado y/o tecnológicas no es atribuible a una persona determinada o determinable, es posible que mañana si pueda ser considerado como tal. Siendo así la pregunta natural es ¿qué criterio debemos aplicar al respecto?. En derecho comparado, ya en 2003, la Agencia Española de Protección de datos (AEPD), mediante informe 327/03 sostuvo que las direcciones IP, tanto fijas como dinámicas, son datos de carácter personal, decisión que basa en los siguientes argumentos:

- a) Es factible identificar por medios razonables a los usuarios a los que se asigna una dirección IP fija o dinámica, por parte de los proveedores de acceso a Internet y los administradores de redes locales.
- b) Con la asistencia de terceras partes responsables de la asignación de la dirección IP se puede identificar a un usuario de Internet por medios razonables.
- c) Existe la posibilidad de relacionar la dirección IP del usuario con otros datos de carácter personal, de acceso público o no, que permitan identificarlo, especialmente si se utilizan medios invisibles de tratamiento para recoger información adicional sobre el usuario, tales como cookies con un identificador único o sistemas modernos de minería de datos.

Este pronunciamiento si bien ha sido controvertido desde la óptica técnico/económica, en el sentido que la aplicación del estatuto jurídico de los datos personales, implica que se les deba aplicar medidas de seguridad al menos de nivel básico, ha sido en general acatada y sostenida en el tiempo. Claro está en este entorno, en todo caso, que no serán considerados dato personal las IP disociadas ya sea porque han sido sometidas al proceso de disociación y/o que se hayan generado disociadas, esto es, que no sea posible por ningún medio atribuirla a una persona determinada y/o determinable.

De su parte, en el seno de la Unión Europea, el grupo del artículo 29 (que es aquel referido al tratamiento de datos personales), el año 2000 se

había pronunciado en este mismo sentido a través del documento de trabajo 5063/00/ES/Final (wp37), titulado **Privacidad en Internet: enfoque integrado comunitario de la protección de datos en línea**. Esta opinión fue ratificada a través de dictamen 04/2007 sobre el concepto de datos personales y recientemente fue aplicada en un caso concreto que ha suscitado bastante polémica, relativo a las actividades de tratamiento de datos de IP de algunos buscadores de Internet. En este documento se enfatiza que *“a menos que el prestador de servicios de Internet sepa con absoluta certeza que los datos corresponden a usuarios que no pueden ser identificados, tendrá que tratar toda información IP como datos personales para guardarse las espaldas?”*.

Un tercer documento relevante de la Unión Europea nos lleva a las mismas conclusiones. Se trata de la Directiva de comunicaciones electrónicas 2002/58 del Parlamento y del Consejo, a cuyo respecto el Grupo del Artículo 29 propone su modificación a través de dictamen 2/2008, en la que profundiza aún más sobre las consecuencias de la consideración de las direcciones IP como datos personales, proponiendo que en la directiva en comento se incluya la obligación de los ESP de notificar los incidentes de seguridad de datos personales a los usuarios *“interesados”*. Proponen que esta obligación se extienda no sólo a los proveedores del servicio de acceso a Internet, sino a todos los proveedores de servicios de la sociedad de la información. A su turno, respecto de los interesados, el grupo 29 propone que sean considerados como tales no sólo los abonados, sino todas aquellas personas cuyos datos se han visto efectivamente comprometidos por la violación de seguridad.

Concluyendo, con independencia de las consideraciones técnicas que podamos realizar, nos parece meridianamente claras las siguientes conclusiones:

- a) Siempre que no sea posible sostener la imposibilidad de que una dirección IP sea atribuible a una persona, habrá de dársele el tratamiento de un dato personal, aplicando la legislación correspondiente.
- b) Será de responsabilidad de quien trata estos datos el acreditar que no existen medios razonables para atribuir a una persona esos datos personales. En consecuencia, a este sujeto le corresponde probar la disociación del dato, lo que implica invertir la carga de la prueba.

---

## 4. Minería de datos de la Web

---

Web mining es el concepto que agrupa a todas las técnicas, métodos y algoritmos utilizados para extraer información y conocimiento desde los datos originados en la Web (web data). Parte de estas técnicas apuntan a analizar el comportamiento de los usuarios, con miras a mejorar continuamente la estructura y contenido de los sitios que son visitados.

Detrás de tan altruista idea, es decir, ayudar al usuario a que se sienta lo mejor atendido posible por el sitio web, subyacen una serie de metodologías para el procesamiento de datos, cuya operación es al menos cuestionable, desde el punto de vista de la privacidad de los usuarios de un sitio web determinado [13, 14]. Entonces surge la pregunta de ¿hasta donde el deseo por mejorar continuamente lo que se ofrece a través de un sitio web puede vulnerar la privacidad de quien lo visita?

#### 4.1. Limpieza y preprocesamiento de los web data

Los datos originados en la Web o web data, corresponden esencialmente a tres fuentes [3]:

1. **Contenido:** Son los objetos que aparecen dentro de una página web, por ejemplo las imágenes, los textos libres, sonidos, etc.
2. **Estructura:** Se refiere a la estructura de hipervínculos presentes en una página.
3. **Uso:** Son los registros de web logs, que contienen toda la interacción entre los usuarios y el sitio web.

Los web data deben ser pre procesados antes de entrar en un proceso de web mining, es decir, son transformados en vectores de características que almacenan la información intrínseca que hay dentro de ellos [6, 16].

Aunque todos los web data son importantes, especial atención reciben los web logs, ya que ahí se encuentra almacenada la interacción usuario sitio web, sus preferencias de contenido y en síntesis su comportamiento en el sitio. Por esta razón, y concediendo de que es posible que en los otros web data se pueda albergar información que identifique a los usuarios, nos concentraremos esencialmente en los web logs, como fuente de mayor controversia al momento de analizar el comportamiento de los usuarios.

La primera etapa, entonces, corresponde a la reconstrucción de la sesión del usuario a partir de los datos existentes en los registros de web log. Este proceso se denomina **sesionización** [3].

Cabe señalar que ciertos sitios han cambiado su estructura con el propósito de identificar a sus visitantes [17]. Una primera estrategia consiste en implementar un sistema username/password, que promueva el registro de los usuarios a cambio de nuevos servicios. Sin embargo, sólo es posible reconstruir perfectamente las sesiones de los registrados, quedando los no registrados en el anonimato. Otra estrategia consiste en utilizar páginas dinámicas en el sitio. Con ellas, cada solicitud de abrir una página genera un identificador único para el usuario, sin embargo, ello obliga a reconstruir el sitio y trae complejidades para identificar qué está realmente viendo el visitante, dadas las direcciones URL dinámicamente generadas [6].

## 4.2. Técnicas, algoritmos y métodos usados en web mining

El concepto web mining, agrupa a todas las técnicas, algoritmos y metodologías utilizadas para extraer información y conocimiento desde los web data, entre los cuales se cuentan [9, 17]:

1. **Self Organizing Feature Maps (SOFMs)**: Esta herramienta tiene una estructura semejante a las redes neuronales, pero en este caso el aprendizaje se da de manera competitiva, es decir, las neuronas compiten para ser activadas, y sólo lo hace una a la vez. La idea de este aprendizaje es que se compara un elemento con la red con el fin de encontrar la neurona más similar, o neurona ganadora. A partir de lo anterior se generan grupos de neuronas o clusters cuyas características son similares.
2. **K-Means**: Este algoritmo se basa en la determinación de grupos o clusters dentro de un conjunto de datos. Para su funcionamiento se necesita como parámetro el número esperado de grupos ( $k$ ). Cada uno de estos clusters estará representado por un centroide, que es el elemento cuyas características se parecen más a las de su conjunto (Obtenido mediante una medida de similitud). Este método tiene una alta performance, por lo que es posible repetirlo varias veces con distintos parámetros.
3. **Árboles de Decisión**: Esta técnica se basa en la estimación de un resultado y toma de decisiones a partir de datos conocidos. La idea es identificar los atributos mínimos con los cuales se pueda deducir un resultado, clasificando los datos en una estructura de árbol y moviéndose a través de las ramas.
4. **Support Vector Machines (SVMs)**: En comparación con las redes neuronales, tiene la ventaja de ser menos propensos al sobre aprendizaje, por lo tanto pueden mantener un gran número de características y datos sin preocuparse de la complejidad del problema. La idea básica de esta herramienta es trabajar con ciertas funciones efectivas (Funciones de Kernel) que permitan tratar los datos a otro nivel dimensional y de esta forma trabajar con modelos complejos.
5. **Algoritmos Inspirados en la Vida**. Se trata de una nueva familia de algoritmos cuya operación está basada en cómo ciertas especies, bacterias y la misma evolución con cambios genéticos, tratan de sobrevivir y perpetuarse en la vida.

### 4.3. Análisis de la operación de las técnicas de minería de datos

Las técnicas de web mining analizadas, utilizan como entrada de datos los web data preprocesados y en forma de vectores de características. Como ya se ha comentado antes, de todos los posibles web data, son los registros de log los que más información aportan para realizar un análisis del comportamiento de los usuarios en un sitio web [16]. Los otros web data: contenido y estructura, pueden ser usados como complemento para hacer más certera la aplicación de técnicas como clustering, clasificación y la estadística.

El resultado que más interesa a las empresas dueñas de sitios web orientados al comercio electrónico, es la creación de sistemas que permitan mejorar la experiencia de los usuarios en el sitio a partir de la personalización de su navegación, lo cual se logra fundamentalmente a través de recomendaciones en línea, respecto de qué deben ver o por donde deberían dirigir su navegación. Lo anterior no elimina la posibilidad de que también se hagan recomendaciones a los administradores del sitio respecto de modificaciones que se deben hacer durante su mantención, es decir, sin usuarios concurrentes.

#### 4.3.1. Procesamiento de los registros de web log

Previo al uso de estos registros, se requiere aplicar un proceso de reconstrucción de la sesión de los usuarios: la sesionización.

Desde un punto de vista de la privacidad de los web data, todo apunta a que el análisis del comportamiento del usuario debe hacerse utilizando estrategias de reconstrucción de la sesión que no ligen directamente a un ser humano con el usuario web [3, 6]. Sin embargo, la extracción de patrones de navegación y preferencia de los usuarios, siempre puede ser utilizada como una forma indirecta de extrapolar el comportamiento de un visitante en un sitio web, que a través de la personalización de sus contenidos [10], puede atentar contra el libre albedrío del usuario, toda vez que la información que verá no será toda la que puede ver, de eso la lógica informática del sitio se va a encargar, tal como “*el gran hermano*” que vela por lo bueno y lo malo que se le permite ver a las personas.

Luego, asumiendo que sólo se trabajará con datos que identifican sesiones, pero no personas, se construyen los vectores de características. Los más usados, contienen información sobre la página visitada, el tiempo que el usuario gasta por página y sesión, más alguna referencia al objeto que se está visitando [18].

#### 4.3.2. Procesamiento de los contenidos en una página web

En una página web se pueden encontrar todos los contenidos desarrollados a lo largo de la historia de la computación, con una variada posibilidad de

formatos. Entonces, el análisis de estos datos se vuelve un proceso no trivial, que requiere de un preprocesamiento y representación de la información previo.

El primer tipo de dato a analizar es el texto libre, el cual corresponde a todo lo que esté escrito y que se haya consignado en una página web, ya sea a través de un archivo enlazado o dentro de la misma página. Estos textos deben ser transformados a un formato numérico, el cual considera que existen palabras más importantes que otras, que se puede reducir un conjunto de palabras a la idea central y que es posible prescindir de algunas estructuras morfológicas [15].

Los otros tipos de datos a analizar, corresponden a imágenes, sonidos y videos. Por lo pronto el desarrollo de herramientas de web mining para estos formatos se encuentra en sus primeras etapas de investigación, siendo necesario recurrir a los metadatos, esto es datos por sobre los datos, que permitan procesar el entorno de estos objetos. Por ejemplo, si se está buscando información sobre una persona, específicamente su fotografía o la escena donde aparece en un video en la Web, será necesario conocer datos adicionales, consignados en los textos que acompañan al objeto, tal como su nombre, edad, etc.

#### **4.3.3. Procesamiento de la estructura de hipervínculos**

El análisis de la estructura de hipervínculos, apunta principalmente a la extracción de información respecto de la importancia de una página en la Web, la identificación de comunidades y el ranqueo de la información recuperada por alguno de los motores de búsqueda. Con esta conocimiento es posible mejorar notablemente la búsqueda de información que realiza el motor para su usuario.

Por construcción, los motores de búsqueda realizan periódicamente una actualización de su base de datos de páginas web, esto es, revisar la Web y recuperar los objetos que han variado en un sitio desde su última visita, respetando la política de seguridad que se haya configurado en el servidor web que mantiene al sitio, es decir, si un objeto no tiene permiso para ser recuperado por el motor, entonces dicha operación no se lleva a cabo.

Bajo la premisa anterior, se puede pensar que el motor de búsqueda sólo puede realizar operaciones de análisis de las páginas que hayan podido ser recuperadas de manera directa, es decir, sin la necesidad de recurrir a algún mecanismo de seguridad como puede ser la aplicación de una clave de acceso. Entonces, la regla es que sólo aquello que es público puede ser buscado en la Web, por lo que la responsabilidad de la publicidad de los contenidos de un sitio queda expresamente consignada a quienes lo mantienen.

Durante todo el período de la Web, también conocido como 1.0, eran los dueños y administradores de los sitios los encargados de publicar la información que se haría pública en el ciberespacio. Sin embargo, con el advenimiento de la Web 2.0, algo cambió radicalmente. Ahora son los usuarios los que han

tomado el control de la publicación de información que muchas veces les es privada, pero que quieren mostrar al mundo, por ejemplo a través de un blog, foro, facebook, etc. . De inmediato surgen varias interrogantes:

- ¿De quién son los datos? ¿Del usuario que lo publicó o del dueño del sitio?.
- Si un usuario quiere borrar algo que el mismo publicó, ¿existen los canales directos para hacerlo?.
- Si alguien publicó en un sitio información que daña la honra de una persona ¿a quién se le obliga a eliminar la página y dar las compensaciones necesarias? ¿al dueño del sitio o al usuario?.

La premisa de que la responsabilidad de publicación de información en un sitio es de quien lo mantiene, no está clara, por cuanto ahora son los usuarios los que pueden crear sus propios contenidos [4].

#### 4.3.4. Análisis de la operación de los sistemas de recomendación

La próxima generación de sitios web, estará fuertemente influenciada por la capacidad que estos tengan para adaptar su estructura y contenido a las necesidades de información que tenga el usuario, ya sea durante su navegación o luego que esta se haya realizado [17]. Este nuevo tipo de sistemas incorpora módulos de personalización del sitio, los cuales tienen su realización práctica en la recomendación de qué visitar, buscar o simplemente observar que se le hace a los usuarios de un sitio.

Es en la preparación de la recomendación donde más se puede vulnerar la privacidad del usuario [13], ya que se requiere de un seguimiento de sus acciones en el sitio, para poder clasificarlo en el grupo adecuado y preparar la recomendación de navegación que más se ajuste a lo que el sistema cree que el usuario anda buscando en el sitio.

Las técnicas usadas en web mining para analizar el comportamiento del usuario en la Web, trabajan con miles de sesiones, sin importar quién es la persona que generó una determinada sesión. Aquí se aplica el principio estadístico de que el comportamiento de una persona es aleatorio, por lo tanto no sirve para conjeturar nada. Sin embargo, el comportamiento colectivo siempre marca una tendencia, por lo que se puede extrapolar y usar como un estimador probabilístico aceptado.

Finalmente, la preparación de la acción de personalización claramente limita el libre albedrío del usuario que visita el sitio, por cuanto implica limitar su exposición a contenidos que *“tal vez no le son de interés”*. En la práctica, esta limitación no ha sido mal recibida por los usuarios, lo cual no quita que igual sea una invasión en la privacidad del visitante del sitio. Sin embargo,

en el ciberespacio, ¿existe el libre albedrío?, claramente somos dueños de ir donde queramos, pero en la mayoría de los casos lo hacemos influenciados por una recomendación de un motor de búsqueda, así que al menor podemos decir que el libre albedrío estaría limitado a lo que “*el gran hermano*” tecnológico quiera mostrarnos [11, 5].

---

## 5. Aspectos jurídicos del tratamiento de web data

---

La idea básica que subyace detrás del web mining es la extracción de información y conocimiento desde un conjunto de web data. Dependiendo del tipo de web data a minar, el algoritmo de web mining puede estar altamente relacionado con los datos personales del usuario. Lo anterior plantea muchas interrogantes, sobre todo en lo referente a la privacidad del usuario.

### 5.1. Marco legal para el análisis de la privacidad en la Web

Partamos analizando los archivos de web log, en especial la dirección IP desde donde accedió el usuario al sitio web. Este parámetro en combinación con otros datos existentes en el registro de web log, ha sido frecuentemente utilizada para identificar la sesión del usuario. Debido a la posibilidad de que se pueda relacionar identificar a la persona a través de la dirección IP que utiliza para navegar por la Web, es que en la UE se está comenzando a considerar a la IP como un dato personal. En España, la Ley Orgánica 15/1999, en su artículo 3a define al dato personal como “*cualquier información concerniente a personas físicas identificadas o identificables.*”

El TCP/IP versión 4, que es el protocolo con que en la actualidad opera Internet, fue concebido para identificar un computador conectado a la red. Hay que recordar que Internet es una *red de redes*, así que para identificar un computador, primero se identifica a qué red pertenece. De esta forma, las direcciones IP están compuestas de cuatro números (rango entre 0 y 255 cada uno) con los que se identifica la red y el computador dentro de esta.

Entonces, por construcción la dirección IP no fue creada para identificar a la persona detrás del computador. Mucho menos ahora que existen sistemas que permiten a varios usuarios acceder a Internet, usando la misma IP y que los ISP entregan direcciones dinámicas, es decir, sólo relacionan una sesión de usuario mientras este está conectado e incluso más, es posible que durante la sesión, el usuario experimente cambios en la IP asignada. Sin embargo, si se realizan los cruces de datos adecuados, se puede llegar a una aproximación respecto de quien sería la persona que en un determinado momento, estaba conectada desde un computador, usando una IP específica.

Asumiendo, entonces, que a través de la dirección IP sólo se puede identificar la sesión y no a la persona que hay detrás, los algoritmos de web mining se orientan a extraer información desde los web data para analizar comportamientos de usuarios en determinados momentos del día, es decir, un mismo usuario se puede comportar diferente en momentos diferentes, con lo cual se argumenta que no se estaría analizando a la persona, sino más bien a grupos de personas para extrapolar comportamientos colectivos [9].

Ahora bien, ¿qué es la privacidad?. La RAE define el término como “*ámbito de la vida privada que se tiene derecho a proteger de cualquier intromisión*”. Y en Internet, ¿este concepto tiene sentido?. En este artículo no se ahondará en el contexto filosófico de la privacidad, sino que se fijarán límites sólo en lo referente al control de la información respecto de uno mismo, es decir, la capacidad que tiene el individuo de proteger los datos que le son propios de su persona. Entonces, la privacidad puede ser violada cuando los datos personales son obtenidos, usados, procesados y diseminados, especialmente sin el consentimiento de su titular. En este contexto, es donde el web mining tendría su mayor accionar, ya que el usuario no tendría la mas mínima idea de que información referente a su persona puede estar siendo procesada.

A partir del uso de algoritmos de web mining, se pueden extraer patrones respecto del comportamiento de grupos de usuarios en la Web. En este sentido el valor del “*individualismo*” podría verse afectado. Este concepto se relaciona con el de privacidad por cuanto muchos sistemas que usan los patrones extraídos a través del web mining, tienden a clasificar a los usuarios y a tomar decisiones en base a cuan parecido es su comportamiento respecto de un grupo, por ejemplo, este usuario se comporta como aquellos que pertenecen al grupo de los amantes del rock, entonces las páginas a mostrarle en su navegación son sólo las referentes a ese tipo de música. Lo anterior claramente coarta toda posibilidad al usuario de que pueda tomar decisiones respecto de que en realidad quiere ver [16].

Otro punto muy importante a dejar en claro, es que los registros web log, identifican a usuarios y no necesariamente a personas determinadas, no obstante pudieran ser determinables y por esta vía ingresan al estatuto jurídico de los datos personales. Es decir, lo que los web log identifican directamente es un ente que tiene acceso a una dirección IP desde la que se conecta, fecha y hora de visita, las páginas que visitó, etc., para los efectos de determinar si, cuando trabajamos con estos datos lo hacemos con datos personales deberemos previamente cuestionarnos sobre si existen medios razonables para identificar efectivamente a las personas “*detrás*” de esos usuarios. Si la respuesta es sí: estaremos frente a un dato personal.

La utilización de mecanismos de identificación, tales como las conocidas cookies, podría establecer una relación directa entre el ser humano y el usuario web. Sin embargo, es posible que un tercero use el computador de una persona

y sin deseárselo la suplante en el sitio web que visita, ya que estaría usando la misma cookie que su antecesor.

Otro caso de vinculación usuario web/persona se produce en los ISP. En efecto, cuando contratamos el servicio Internet, datos personales respecto de nosotros quedan consignados en un contrato. Luego para una determinada sesión, el ISP sabe al menos a través de que conexión el cliente está navegando por la Web. Sin embargo, dado que una conexión puede ser compartida, es decir, varias personas saliendo por un mismo lugar, nuevamente no es posible vincular una determinada sesión a un usuario.

La ley alemana para la *legítima interceptación* obliga a los ISPs a mantener todas las transacciones que han realizado los usuarios a través de sus sistemas, en el caso de que el gobierno las necesite para realizar una investigación criminal. En Chile, el decreto 142 de 2005 de la Subtel señala que los ISPs deben mantener un registro, no inferior a seis meses, de las conexiones a Internet que realicen sus abonados.

También se da el caso de que los ISP pueden ser restringidos en su operación, por ejemplo, en Holanda, este servicio es considerado como una telecomunicación más, es decir, tienen que obedecer lo estipulado en la nueva ley de Telecomunicaciones de 1998, el cual estipula que los ISP están obligados a borrar o hacer anónimo, todos los datos relacionados con el tráfico generado por sus suscriptores una vez que estos finalizan la llamada. La aplicación de cualquier técnica o algoritmo de extracción de información por sobre los datos generados por a través ISP, sólo se puede realizar previa autorización expresa del cliente.

La tendencia mundial en mejores prácticas para el tratamiento de los web data, especifica que se debe [8]:

- Informar al usuario que está entrando en un sistema informático el cual por construcción almacenará datos respecto de su navegación en el sitio y que dichos datos pueden ser usados para hacer estudios posteriores.
- Obtener el consentimiento explícito del usuario para realizar una operación de personalización del sitio web que visita. Por ejemplo ¿desea usted que le enviemos sugerencias de navegación?.
- Proveer una explicación sobre las políticas de seguridad que se aplican para mantener los web data que se generen en el sitio.

Estas prácticas, son un marco mínimo de requerimientos para asegurar una adecuada privacidad del usuario en el tratamiento de los web data.

En Chile, la ley 19.628 sobre datos personales, consagra como tales a *los relativos a cualquier información concerniente a personas naturales, identificadas o identificables*<sup>2</sup>. En su sentido amplio, los web data no estarían con-

---

<sup>2</sup>ey sobre protección de datos de carácter personal. Ver <http://www.bnc.cl/>, 2002

templados como dato personal, salvo los referentes a las direcciones IP que podrían ser utilizadas, en combinación con otros datos para identificar a la persona detrás de la sesión del usuario. Entonces, el tratamiento de los web data podría estar regulado por la citada ley, siendo el responsable del banco de datos, el administrador o dueño del sitio web que el usuario visita.

## 5.2. Privacidad y libertad de navegación en la personalización de la Web

La personalización de la Web es la rama de la investigación en Web Intelligence dedicada a ayudar al usuario a que pueda encontrar lo que busca en un sitio web [7, 17]. Para esto, se han desarrollado sistemas informáticos que ayudan a los usuarios a través de sugerencias de navegación, contenidos, etc. y más aun, entregan información valiosa a los dueños y administradores de sitios para que realicen cambios en su estructura y contenido, siempre con la idea de mejorar la experiencia del usuario, haciéndolo “*sentir*” como si fuese el visitante más importante del sitio, con una atención personalizada. Para lograr lo anterior, se han desarrollado múltiples esfuerzos tendientes a extraer información desde los web data que se generan con cada visita del usuario a un sitio, siendo los trabajos en web mining, los que han concentrado la mayor atención de empresas e investigadores en los últimos años.

Primero que todo, hay que dejar en claro el fin último que persigue el uso del web mining: aprender del comportamiento de los usuarios en la Web, para mejorar la estructura y contenido de un determinado sitio, personalizando la atención del usuario [16].

Como se puede apreciar, el fin es bastante altruista, siempre orientado a satisfacer al usuario y en el fondo a ayudarlo a encontrar lo que busca. Ahora bien, el exceso de *ayuda* no sólo puede molestar al usuario, sino que además, para ayudarlo mejor, se requiere de más y más datos, conocer sus preferencias y en buenas cuentas, intrometerse en su privacidad y limitar la cantidad de contenidos que puede ver de un sitio.

Existe evidencia empírica que los sitios que incorporan sistemas de personalización de sus contenidos, logran establecer una relación de lealtad con sus visitantes. Sin embargo, el precio a pagar es permitir que el sistema se inmiscuya en aspectos relacionados con las actividades del usuario en el sitio, sus hábitos anteriores de navegación o de pares parecidos, etc. En algunos casos, el usuario puede llegar a experimentar una verdadera sensación de invasión su privacidad, lo que se traduce en otra razón más por la cual un usuario no visita un sitio web que personaliza la información que muestra a sus visitantes, es decir, el *remedio fue peor que la enfermedad*, por lo que el desarrollo de este tipo de sistemas se está tomando con cautela, más allá de las implicancias legales que puede traer el vulnerar la privacidad de los actos de los

visitantes de un sitio. Adicionalmente, los sistemas de personalización tienden a mostrar lo que se cree es bueno e interesante para el usuario, coartándole su libertad de navegación y restringiéndolo sólo a lo que el sistema considera que es importante o necesario que vea.

Entonces ¿hasta qué punto la personalización de la Web es invasiva de la privacidad de los usuarios? [6], ¿Se coarta la libertad de navegación al ocultar o sólo mostrar ciertos contenidos, dependiendo de lo que el sistema estime es conveniente para el usuario?. La percepción dependerá mucho de las características culturales de cada país o más aun, comunidad de individuos. La solución a la cual más se ha recurrido, es realizar encuestas de opinión a los usuarios de los sitios, pero que van más de acorde a las bondades que trae la personalización, sin explicar en detalle el cómo se logra.

La creación de sistemas para personalizar la navegación en la Web, limita el libre albedrío, por cuanto asume que el usuario no es lo suficientemente avezado como para encontrar información por si solo y necesita ayuda, que al final se transforma en una imposición sublime sobre qué debe ver. Entonces, ¿dónde está el punto de balance entre vulnerar privacidad, coartar la libertad de navegación y ayudar efectivamente al usuario?. Tal vez la solución sea muy simple, y todo pase por preguntarle al usuario si necesita apoyo y explicarle que para ayudarlo se requiere involucrarse un poco más en su vida privada. Lamentablemente lo anterior en la Web es complicado, ya que muchas preguntas cansan al usuario y es ineficaz.

### 5.3. Comentarios finales

En la Ley 19.628, artículo 2º letra “e” se define como dato estadístico a aquel que “*en su origen, o como consecuencia de su tratamiento, no puede ser asociado a un titular identificado o identificable*”. En este sentido, las técnicas de preprocesamiento y limpieza de web data, que son aplicadas como paso previo al web mining, pueden eliminar cualquier indicio que identifique o permita identificar a los usuarios, transformando de esta forma al web data en un dato estadístico. El problema es que si se realiza esta práctica, se minimiza el beneficio potencial que las empresas pueden obtener respecto del uso que los visitantes les dan a sus sitios web corporativos.

Del punto de vista de la investigación científica, el uso de web data de corte estadístico no es un problema, por cuanto lo que se estudia es el comportamiento de grupos de usuarios utilizando los datos consignados en las sesiones y los contenidos de las páginas web, con lo que se salvaguardaría la privacidad de los usuarios.

Respecto de la afectación a la libertad de navegación, esta viene dada principalmente por los sistemas de personalización y adaptación de la Web. Si bien es cierto que tanto del punto de vista científico como de negocio no es necesari-

ria la identificación del usuario para dar una recomendación de navegación o la reestructuración en línea de los sitios, de todas maneras se producirá una reducción de los posibles contenidos que el usuario podrá ver. En este sentido, es necesario que al menos se de la posibilidad al usuario para que libre y soberanamente decida si quiere ser ayudado por un sistema informático a encontrar lo que busca o si desea hacerlo por su propia cuenta.

El desarrollo actual y futuro de Internet y la Web, estará estrechamente relacionado con los avances en telecomunicaciones. Entonces se hace necesario analizar la normativa vigente para las telecomunicaciones, por cuanto los datos originados en la Web, se transmiten a través de redes de computadores, cuyos medios justamente son regulados por ley. En efecto, en la ley General de Telecomunicaciones 18.168<sup>3</sup>, se entiende por telecomunicación a *“toda transmisión, emisión o recepción de signos, señales, escritos, imágenes, sonidos e informaciones de cualquier naturaleza, por línea física, radioelectricidad, medios ópticos u otros sistemas electromagnéticos”*.

Lo primero antes de establecer la normativa a aplicar en el caso de los web data, desde el ámbito de la ley 18.168, es establecer a qué servicio correspondería su transmisión dentro de Internet. Yendo al meollo mismo, habría que analizar que tipo de servicio es el acceso a la gran red, pues es ahí donde comienza la generación de los datos. En base a la definición de **Servicio Complementario**, que se consagra en la ley como *“servicios adicionales que pueden ser prestados por concesionarios de servicio público o terceros, mediante la conexión de equipos a redes públicas. No requieren autorización previa de ninguna autoridad”*, se puede argumentar que el acceso a Internet es un servicio complementario de telecomunicaciones, por lo que las normas y principios de esta rama del derecho le son plenamente aplicables, especialmente los principios de libertad y secreto de las comunicaciones, neutralidad de red, acceso universal, protección de usuarios, etc., respecto de los cuales corresponde velar a la Subsecretaría de Telecomunicaciones (Subtel), la cual a su vez debe estar encargada de velar por la protección de los derechos de los abonados que le pagan a una empresa por usar el enlace que le permite llegar a la gran red, sin perjuicio de otras instancias que le permitan al abonado estampar algún reclamo.

---

## 6. Conclusiones

---

La identificación de una persona a partir de los web data que se recolectan en un sitio web, no es factible en su totalidad. Utilizando el actual protocolo de comunicaciones de Internet: IP versión 4, a lo más se puede identificar la sesión de un usuario, es decir, un ente que en un determinado momento

---

<sup>3</sup>Ley general de telecomunicaciones. Ver <http://www.bnc.cl/>, 1982

está navegando en un sitio web.

El uso de las técnicas de web mining para la extracción de información y conocimiento desde los web data, puede vulnerar la privacidad de los usuarios que visitan un sitio web. Ahora bien, existen formas de minimizar esta vulneración hasta lo estrictamente necesario y con el consentimiento del usuario para ayudarlo en su búsqueda de información en un sitio, comenzando por cómo se pre procesan los web data y finalizando en la forma en que se entregan las recomendaciones de navegación y preferencias de contenido.

Finalmente, la pregunta que motiva este artículo tiene como respuesta de que efectivamente, las herramientas de web mining pueden ser el soporte tecnológico como para que se vulnere la privacidad de los usuarios y se coarte su libertad de navegación a través de sistemas que buscan la personalización de su experiencia en un sitio web. Por lo tanto, lo que se debe hacer es promover un conjunto de buenas prácticas para hacer un trabajo limpio, ético y que salvaguarde las garantías fundamentales de todos los involucrados. No se recomienda bajo ningún precepto la creación de una nueva regulación que sólo actuaría en casos puntuales, que en muy poco tiempo quedaría obsoleta y lo que es peor, detendrá el desarrollo científico en un área tan importante como lo es el futuro de la Web.

### Agradecimientos

El primer autor agradece el aporte del Instituto Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16).

### Referencias

- [1] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret. The world wide web. *Communications of ACM*, 37(8):76–82, 1994.
- [2] P. Carrasco-Jiménez. *Análisis Masivo de Datos y Contraterrorismo*. Tirant lo Blanch, Valencia, España, 2009.
- [3] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Journal of Knowledge and Information Systems*, 1(1):5–32, 1999.
- [4] M. Corripio-Gil-Delgado. *Regulación Jurídica de los Tratamientos de Datos Personales realizados por el Sector privado en Internet*. Agencia de Protección de Datos, Madrid, España, 2000.
- [5] P.L. Murillo de la Cueva and J. L. Piñar-Mañas. *El Derecho a la Autodeterminación Informativa*. Fundación Coloquio Jurídico Europeo, Madrid, España, 2009.

- [6] M. Eirinaki and M. Vazirgannis. Web mining for web personalization. *ACM Transactions on Internet Technology*, 3(1):1–27, February 2003.
- [7] W. Kim. Personalization: Definition, status, and challenges ahead. *Journal of Object Technology*, 1(1):29–40, 2002.
- [8] A. Kobsa. Tailoring privacy to users needs. In *In Procs. of the 8th International Conference in User Modeling*, pages 303–313, 2001.
- [9] Z. Markov and D. T. Larose. *Data Mining the Web: Uncovering Patterns in Web Content, Structure and Usage*. John Wiley and Sons, New York, USA, 2007.
- [10] M.D. Mulvenna, S.S. Anand, and A.G. Büchner. Personalization on the net using web mining. *Communications of the ACM*, 43(8):123–125, 2000.
- [11] M. Muñoz-Campos and H. Soto-Arroyo. *Derecho de Autodeterminación Informativa*. Editorial Jurídica Continental, San José, Costa Rica, 2005.
- [12] W. Stallings. *SNMP, SNMPv2, and CMIP: the practical guide to network management*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1993.
- [13] H.T. Tavani. Informational privacy, data mining, and the internet. *Ethics and Information Technology*, 1:137–145, 1999.
- [14] A. Vedder. Privacy and confidentiality. medical data, new information technologies, and the need for normative principles other than privacy rules. *Law and Medicine*, 3:441–459, 2000.
- [15] J.D. Velásquez and P. González. Expanding the possibilities of deliberation: The use of data mining for strengthening democracy with an application to education reform. *The Information Society*, 26(1):1–16, 2010.
- [16] J.D. Velásquez and V. Palade. A knowledge base for the maintenance of knowledge extracted from web data. *Knowledge-Based Systems*, 20(3):238–248, 2007.
- [17] J.D. Velásquez and V. Palade. *Adaptive Web Site*. IOS Press, Amsterdam, Netherland, 2008.
- [18] J.D. Velásquez, R. Weber, H. Yasuda, and T. Aoki. Acquisition and maintenance of knowledge for web site online navigation suggestions. *IEICE Transactions on Information and Systems*, E88-D(5):993–1003, 2005.

---

# EXPERIENCIAS PRÁCTICAS EN LA MEDICIÓN DE RIESGO CREDITICIO DE MICROEMPRESARIOS UTILIZANDO MODELOS DE CREDIT SCORING

---

CRISTIÁN BRAVO\*  
SEBASTIÁN MALDONADO\*  
RICHARD WEBER\*

## Resumen

*Todas las instituciones financieras que ofrecen crédito a sus clientes deben abordar el problema de estimar cuánto del dinero otorgado retornará a la entidad y a qué clientes ofrecerles crédito. Sistemas de Credit Scoring se han desarrollado de manera exitosa para determinar la probabilidad que un cierto cliente falle en devolver el crédito que le ha sido otorgado. En el presente trabajo se describen los modelos desarrollados para dos organizaciones financieras nacionales para microempresarios, ajustando los pasos del proceso KDD (Knowledge Discovery in Databases) a sus necesidades particulares. El documento presenta las experiencias obtenidas a partir de estos proyectos y explica en detalle como se resolvieron los problemas asociados a las características particulares de los microempresarios en Chile. La mayoría de los proyectos asociados al proceso KDD son de naturaleza estática. Sin embargo, con el paso del tiempo los modelos comienzan a perder la capacidad de explicar los fenómenos para los que fueron construidos inicialmente. Debido a los requerimientos de las entidades financieras se desarrollaron módulos para el seguimiento y la recalibración de los modelos. En particular, se proponen técnicas estadísticas con el fin de determinar cuándo los cambios en las características de la población pueden afectar el desempeño del modelo. Durante el desarrollo de las soluciones se pudo obtener un conocimiento importante sobre el comportamiento de los clientes. Algunos descubrimientos fueron sorprendentes, mientras otros confirmaron las nociones que tenían los expertos. La utilización de estos sistemas en las operaciones diarias puede reducir la tasa tanto de falsos positivos como de falsos negativos, lo que se traduce en menores costos y una mayor cobertura en los mercados respectivos.*

**Palabras Clave:** *Credit scoring, Regresión logística, Microempresarios.*

---

\*Departamento Ingeniería Industrial, Universidad de Chile

---

## 1. Introducción

---

En el escenario actual, los avances tecnológicos han permitido un desarrollo importante en la automatización de la decisión sobre la aceptación o rechazo de una solicitud de crédito mediante modelos analíticos, evitando el otorgamiento bajo criterios ambiguos, lo que en algunos países se considera una práctica ilegal.

Los modelos analíticos requieren de información cuantitativa potencialmente útil para su construcción. Si bien la posibilidad de obtener esta información es cada vez más simple, gracias al importante aumento de la capacidad de almacenaje y la disponibilidad de mejores herramientas para el manejo de datos, el proceso de extracción de información relevante a partir de los datos disponibles sigue siendo complejo y costoso. Las técnicas utilizadas para esta tarea se engloban bajo el concepto de Minería de Datos (*data mining*).

La modelación de la falla financiera, tanto en personas como en empresas, ha sido un problema altamente estudiado en la literatura. Desde el comienzo de los años sesenta, de acuerdo a los trabajos de Beaver [3] y Altman [1], se han desarrollado modelos matemáticos y estadísticos que buscan predecir el desempeño que tendría una persona si se le otorgase crédito mediante la asignación de un puntaje estimado a partir de la información del cliente. Este problema se conoce como *Credit Scoring* [13].

Si bien los modelos de Credit Scoring han sido ampliamente estudiados en la literatura, su aplicación al segmento de los microempresarios no es directa debido a que éstos representan un grupo diferente en relación al resto de los clientes, ya sea en términos de tamaño, ingresos o estructura social.

El presente trabajo se enfoca en el problema de medición de riesgo crediticio de microempresarios mediante modelos de Credit Scoring, resumiendo los resultados y experiencias obtenidas en dos proyectos, uno para una entidad financiera gubernamental [6] y otro para una institución privada, en los últimos siete años.

La estructura de este trabajo es la siguiente: La sección 2 define el marco teórico y muestra los principales avances en la modelación de la problemática asociada a Credit Scoring, destacando el desafío que representa el segmento de los microempresarios al momento de efectuar Credit Scoring tradicional. El desarrollo del modelo propuesto, junto con algunas estrategias para el seguimiento y recalibración del modelo se presentan en la sección 3. La sección 4 presenta los principales resultados del trabajo. Finalmente, la sección 5 muestra las conclusiones del trabajo.

---

## 2. Credit Scoring Aplicado a Microempresarios

---

Los microempresarios en Chile representan un sector importante de la economía, los cuales presentan características especiales que los hacen diferentes a las empresas que se estudian comúnmente en Credit Scoring tradicional. Estas particularidades se describen en esta sección. Adicionalmente, este marco teórico introduce los conceptos asociados al problema de Credit Scoring, junto con la descripción del proceso dentro del cual se encuentran insertas las herramientas de clasificación. Este proceso conoce como KDD (*Knowledge Discovery in Databases*, [7]). Se coloca especial énfasis en las etapas de preparación de los datos, selección de atributos y seguimiento de los modelos que son los elementos centrales de este artículo.

### 2.1. Microempresarios en Chile

En Chile, un microempresario se define como una empresa muy pequeña, con un máximo de nueve trabajadores, un ingreso por ventas mensuales promedio no superior a los 200 UF o unos US\$ 8.000 (para el caso de los programas de capacitación y asesoría, el umbral de ventas mensuales se disminuye a 150 UF o unos US\$ 6.000) y cuentan con activos fijos menores a 500 UF o unos US\$ 20.000, de acuerdo a la definición utilizada por el Fondo de Solidaridad e Inversión Social (FOSIS, [www.fosis.cl](http://www.fosis.cl)). Esta definición es la más utilizada en nuestro país, puesto que muchas otras instituciones la usan como referencia.

Los microempresarios representan un pilar fundamental de la economía nacional, ya que el 81% de las 707.634 empresas formales existentes el año 2004 pertenecen a esta categoría, porcentaje que presenta una escasa variación en la última década. Sin embargo, esto no se ve reflejado en las ventas, ya que este sector representa solamente el 3.4% de la participación de las ventas totales de este año [5].

Si bien las microempresas consideran un conjunto de negocios con alto grado de heterogeneidad, es posible describirlas en función de ciertas características comunes [5]:

- Por lo general corresponden a negocios familiares o trabajadores autoempleados.
- Representan organizaciones con bajos rendimientos, generalmente ineficientes en el abastecimiento de materias primas, comercialización, manejo contable y financiero.
- son de gestión conservadora y salarios bajos. Como promedio la venta de las microempresas chilenas es de 456 UF anuales, o sea, US\$16.000

aproximadamente, monto muy pequeño que genera inestabilidad a sus empleados y propietarios.

Si bien las microempresas se encuentran presentes en prácticamente todas las actividades económicas, su mayor representación está en aquellos sectores de menor potencialidad y mayores barreras de entrada, como es el caso del comercio. Las cifras indican que el 77% de las microempresas se concentra en cuatro sectores económicos: comercio, servicios, transporte y agricultura [5].

El mercado de los microempresarios, usualmente apoyado financieramente por gobiernos e iniciativas de la Unión Europea, se ha transformado en un negocio atractivo para bancos y otras instituciones crediticias. Sin embargo, el mercado presenta características de riesgo únicas que no han sido abordadas por los modelos de riesgo tradicionales, surgiendo la necesidad de crear modelos ad-hoc y atrayendo el interés de tanto investigadores como corporaciones privadas. En particular, los microempresarios chilenos presentan ciertas cualidades que deben ser tenidas en cuenta al momento de desarrollar modelos de Credit Scoring, tales como:

- Los microempresarios usualmente tienen un presupuesto limitado, debido a su menor ingreso. Debido a esto, la variable ingreso, que es un candidato natural para formar parte de los modelos de riesgo, suele presentar una escasa capacidad discriminativa.
- Existe un limitado conocimiento de las variables que los caracterizan, siendo necesario un estudio detallado de sus características y necesidades, con el fin de ofrecerles crédito de manera responsable y sin caer en prácticas discriminatorias que nacen de la incertidumbre.

Debido a estas razones, las técnicas de Credit Scoring tradicionales deben ser adaptadas con el objetivo de reflejar la realidad presentada y crear las condiciones adecuadas tanto para ellos como para las mismas instituciones financieras. Esta es una preocupación tanto de los gobiernos como de las instituciones privadas, y esta experiencia busca entregar resultados aplicables para ambos.

## 2.2. Definición del problema

Hasta hace no mucho tiempo, la decisión de entregar créditos se basaba en el juicio humano para determinar el riesgo de no pago del postulante a crédito en base a los atributos de éste. Sin embargo, el crecimiento de la demanda por crédito ha llevado a desarrollar métodos formales y objetivos para ayudar a los proveedores del crédito a decidir a quién otorgar crédito y a quién no. Este enfoque fue introducido en los años 40 y con los años se ha desarrollado significativamente. En los años recientes, la alta competencia de la industria

financiera, los avances en la computación y el crecimiento exponencial del tamaño de las bases de datos han llevado a estos métodos a transformarse en una importante herramienta en la industria.

Credit Scoring se define formalmente como un método cuantitativo que se utiliza para predecir la probabilidad de que un aspirante a crédito o un cliente de la entidad crediticia existente deje de pagar el crédito o bien no lo haga una vez que lo reciba [11]. Su objetivo es ayudar a los proveedores de créditos a cuantificar y manejar el riesgo financiero relacionado con el otorgamiento de créditos, para así tomar decisiones de forma rápida y objetiva.

Credit Scoring tiene múltiples beneficios que incumben no sólo a las entidades crediticias, sino también a los beneficiarios del crédito. Por ejemplo, Credit Scoring ayuda a reducir la discriminación porque provee un análisis objetivo del mérito del postulante para recibir un crédito. Esto les permite a los proveedores enfocarse sólo en la información relacionada con la asignación del crédito y así evitar subjetividad. Cuando se le niega un crédito a un cliente en los Estados Unidos, la *Equal Credit Opportunity Act* exige a la institución financiera proveer las razones de por qué fue rechazado. Razones vagas o indefinidas son ilegales, por lo que variables que puedan llevar a discriminación tales como raza, sexo o religión no pueden ser incluidas en estos modelos [11].

Credit Scoring ayuda también a acelerar y a hacer más consistente el proceso de asignación de créditos, permitiendo su automatización. Esto reduce significativamente la necesidad de intervención humana y por ende los costos asociados a este proceso. Más aún, Credit Scoring puede ayudar a las instituciones financieras a determinar la tasa de interés que deben cobrar a sus clientes y para valorizar portafolios [14]. A clientes con mayor riesgo se les cobra una tasa de interés más alta. Esto ayuda a la entidad a manejar sus cuentas de manera más efectiva y provechosa en términos de utilidades.

Finalmente y gracias a los avances de la tecnología, se han desarrollado modelos para Credit Scoring más efectivos. En consecuencia, entidades crediticias utilizan esta información generada para formular mejores estrategias de cobranza y utilizar sus recursos más eficientemente. En particular, Credit Scoring ayuda a empresas aseguradoras a realizar una mejor predicción de las reclamaciones, controlar el riesgo de manera efectiva y determinar el precio de los seguros de manera adecuada. Esto les permite ofrecer mayor cobertura a más clientes a un precio equitativo, reaccionar rápido ante los cambios del mercado y obtener ventajas competitivas.

El problema principal que se aborda corresponde a definir si un cliente que presenta características  $\mathbf{X}$  va a caer en una situación de falla financiera dentro de un futuro cercano y no devolverá íntegramente el crédito otorgado. Para ello, es necesario contar con características que sean relevantes para el estudio y que permitan medir el fenómeno. En particular, se busca encontrar aquel vector  $\mathbf{X}$  de características tal que permita predecir la probabilidad de ocurrencia de un

fenómeno binario  $y$ , en este caso si el cliente falla en la devolución del crédito, con un margen de error razonable. Matemáticamente, podemos expresar el objetivo según la ecuación 1.

$$p(y = falla|\mathbf{X}) = f(\mathbf{X}) \quad (1)$$

Donde  $p(y = falla|\mathbf{X})$  corresponde a la probabilidad que la empresa caracterizada por  $\mathbf{X}$  no pueda cumplir sus compromisos financieros y  $f(\mathbf{X})$  corresponde a una función matemática que aproxima la probabilidad a partir de los datos disponibles.

### 2.3. Proceso KDD

Se describirá a continuación el proceso KDD, el cual representa el proceso completo de extracción del conocimiento en base de datos [7]. El cumplimiento de los pasos del proceso KDD permite llegar a modelos con un mejor desempeño y evita incurrir en errores de modelación, por ende será utilizado como guía para el desarrollo de este proyecto. El proceso KDD se puede aplicar usando métodos estadísticos como la regresión logística. Los pasos del proceso KDD son la consolidación de datos, el pre-procesamiento de los datos, el minado de los datos y la interpretación de los patrones encontrados, como se observa en la figura 1.

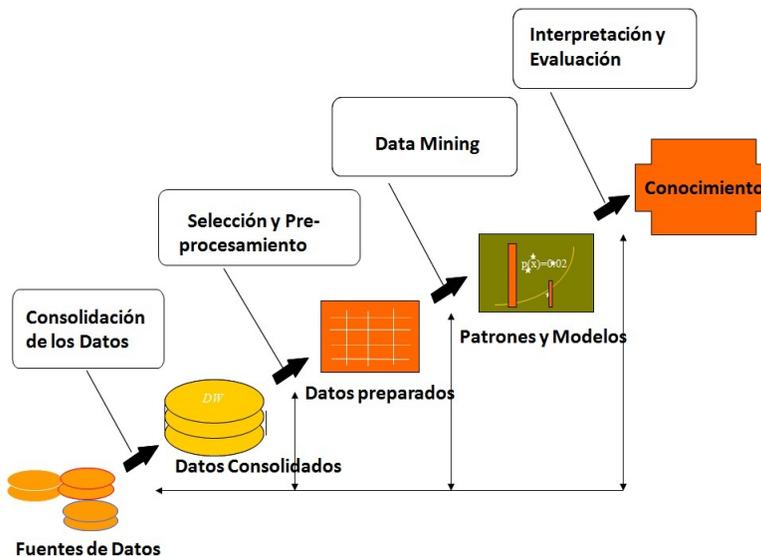


Figura 1: Proceso KDD

- *Consolidación de datos:* Para poder comenzar a analizar y extraer información útil de los datos es preciso, en primer lugar, disponer de ellos.

Esto en algunos casos puede parecer trivial, partiendo de un simple archivo de datos, sin embargo en otros, es una tarea muy compleja donde se debe resolver problemas de representación, de codificación e integración de diferentes fuentes para crear información homogénea.

- *Selección de atributos*: Para la construcción de modelos de clasificación se desea utilizar la menor cantidad de atributos posibles de manera de obtener un resultado considerado aceptable por el analista. Sin embargo, el problema radica en la elección y el número de atributos a seleccionar, debido a que esta elección determina la efectividad del modelo de discriminación construido. Este problema se conoce como *selección de atributos* y es combinatorial en el número de atributos originales [10].
- *Pre-procesamiento de datos*: El propósito fundamental de esta fase es el de manipular y transformar los datos en bruto, de manera que la información contenida en el conjunto de datos pueda ser descubierta. En esta etapa se consideran pasos como limpieza de datos ausentes o incorrectos, reducción de la información y transformación de los datos con el fin de adecuarlos al método de minería de datos.
- *Aplicación del método de minería de datos*: La aplicación de un algoritmo de aprendizaje tiene como objetivo extraer conocimiento de un conjunto de datos y modelar dicho conocimiento para su posterior aplicación en la toma de decisiones.
- *Interpretación y Evaluación*: En esta etapa se realizan distintas pruebas como análisis de sensibilidad y validación con distintas muestras para probar la robustez del modelo, así como la interpretación de los patrones minados.

En todas las etapas del proceso KDD es fundamental la cooperación con un experto del negocio como se mostrará más adelante.

## 2.4. Seguimiento de modelos

Una vez implementado el modelo de Credit Scoring desarrollado, la tarea siguiente, desde el punto de vista estadístico, es cuidar que el resultado obtenido mantenga su capacidad de discriminar entre los clientes que no pagan el crédito (*defaulters*) y los que sí lo hacen. Esta problemática no ha recibido mucha atención en la comunidad, aún cuando el no atenderla lleva a consecuencias graves en el uso. Se ha discutido [8, 9], por ejemplo, que la falta de actualización y mantenimiento de los modelos de riesgo de los bancos estadounidenses fueron una de las causas que precipitaron la crisis Sub-Prime de los años recientes. Debido a peticiones de varias instituciones financieras desarrollamos diferentes

enfoques de seguimiento de modelos de Credit Scoring basados en la regresión logística [4]. Para definir el problema, es posible identificar los cambios que pueden llegar a afectar de manera significativa la capacidad predictiva de un modelo:

- Capacidad discriminante de las variables: Para que una variable sea incluida en un modelo de regresión logística es necesario que esta discrimine entre las dos clases en estudio. Por discriminar se entiende el hecho que la distribución (media, desviación, etc.) de la variable sea distinta para cada una de las clases, de tal forma que a distintos valores de ella se obtengan distintas capacidades discriminantes. Este entonces corresponde a la primera condición que debe ser chequeada al momento de revisar cambios en el modelo.
- Distribución de las variables: Los supuestos básicos del modelo indican que cada una de las observaciones  $x_i$  es extraída de un conjunto  $X$  tal que se distribuye en base a una función  $f(x_i)$  desconocida, pero idéntica para cada elemento. Este supuesto trae como consecuencia que los parámetros extraídos tengan aplicabilidad sólo mientras se tienen variables extraídas desde esta distribución, sin embargo, las distribuciones de las variables tienden a cambiar en el tiempo, pues la población modifica su comportamiento. Este fenómeno se observa por ejemplo en el riesgo crediticio, dónde empíricamente cada dos años se observan cambios en la población suficientes para impactar en el modelo [13].
- Capacidad discriminante del modelo en su conjunto: El cambio más drástico que puede tener una población puede volver el modelo en su conjunto no discriminante, si bien cada variable por separado puede mantener esta capacidad.

Se han desarrollado algunas aproximaciones teóricas por otros autores para identificar estos cambios en modelos de clasificación. Dentro de nuestro conocimiento, el enfoque más cercano al aquí detallado corresponde al trabajo realizado por Zeira *et al.* [15], el cual desarrolla test estadísticos para el caso general de modelos en el cual el error de validación se distribuye normal y las variables poseen un comportamiento tal que se puedan construir estadísticos a partir de sus distribuciones.

---

### 3. Metodología Propuesta y Experiencia

---

La construcción de los modelos se realizó siguiendo el proceso KDD. De acuerdo a esto, las experiencias más importantes se presentan siguiendo el orden señalado en este proceso (sub-sección 2.4), incluyendo el trabajo realizado para el seguimiento de los modelos.

#### 3.1. Definición del Problema y Construcción de la base de datos

Todo proyecto parte con una definición clara de los objetivos del problema. En esta etapa es necesario definir la variable objetivo que se utilizará para clasificar, donde se consideran distintas condiciones de morosidad y se definen umbrales que separan los clientes etiquetados como “buenos” o “malos” en términos de su comportamiento crediticio. Es muy importante que este proceso se lleve a cabo en conjunto con la entidad financiera, debido a que los objetivos suelen variar. Por ejemplo, una entidad financiera estatal presenta una mayor preocupación por temas como la cobertura, a diferencia de instituciones privadas, donde la ganancia es de mayor preocupación.

Dentro de esta primera etapa se deben identificar además las fuentes de datos que son potencialmente útiles de acuerdo a los objetivos del problema y proceder a la adquisición de variables. Este proceso puede resultar complejo ya que la información relevante puede venir de diferentes fuentes. Para una institución financiera se contaban con más de 150.000 registros de créditos en un tramo de diez años, descritos por más de 100 variables. Para la segunda institución se disponían de aproximadamente 8.000 observaciones en un intervalo de tiempo de cuatro años. Sin embargo, el conjunto de atributos disponibles era de más de 650 variables. Esta entidad contaba con un sistema de riesgo traído del extranjero que no alcanzó los resultados esperados debido a que la realidad de los microempresarios difiere de manera drástica de país en país, surgiendo la necesidad de estudiar a fondo sus características en el caso particular.

Las fuentes de datos pueden ser de distinta naturaleza. A continuación se presenta una clasificación de las fuentes de datos más importantes:

- Bases de datos internas: Estas bases de datos incluyen, entre otros, la información personal del cliente, su historial crediticio con la entidad e indicadores preexistentes.
- Bases de datos externas: Muchas veces es posible obtener información de fuentes ajenas a la entidad, como la deuda en otras entidades financieras como bancos, casas comerciales o entidades privadas (DICOM).

- Variables e indicadores derivados: Más de 200 variables fueron construidas a partir de otras, tales como *ratios* de ingresos y deudas.

Considerar modelos con créditos a plazos muy diferentes puede introducir un sesgo, debido a que créditos a más largo plazo tienen asociado generalmente un monto mayor y por ende un riesgo implícito más alto, independiente de las características del cliente que lo recibe. Debido a esto, resulta importante diferenciar los clientes en distintos segmentos de riesgo y/o de acuerdo a condiciones similares. Esto último es más relevante aún cuando se cuenta con clientes antiguos para la compañía que presentan un historial de crédito, versus clientes nuevos sin información en muchas variables potencialmente útiles. A modo de ejemplo, una compañía contaba con créditos con plazos de hasta 10 años. Para esta entidad el universo se segmentó en 5 niveles distintos, de acuerdo a si los clientes eran nuevos o antiguos y en tres niveles de plazo (corto-mediano-largo), donde los dos segmentos de largo plazo se unificaron ya que presentaban características similares.

### 3.2. Pre-procesamiento de los Datos

Una vez con los datos provenientes de distintas fuentes consolidados en una matriz con los créditos las filas y sus atributos en las columnas, los siguientes pasos consisten en la limpieza de los datos y la selección de variables. Se desarrolla una metodología de cinco pasos con este propósito:

1. Concentración y análisis de valores perdidos: Con el fin de descartar rápidamente atributos irrelevantes, las variables muy concentradas en un único valor (en más de un 99% de los casos) y atributos con más de un 30% de valores perdidos fueron eliminados. La racionalidad de este segundo criterio es reducir el número de observaciones que deban ser eliminadas debido a valores perdidos.
2. Análisis univariado: Las variables fueron testeadas de manera individual si presentaban independencia con respecto a la variable objetivo. En particular, se utilizaron los tests de Kolmogorov-Smirnov para variables continuas y Chi-cuadrado para variables discretas. Si las variables estudiadas no presentaban diferencias al ser agrupadas en las dos categorías de la variable objetivo (por ejemplo, si la edad de los clientes buenos fuera estadísticamente similar en distribución a la edad de los clientes etiquetados como malos) se eliminaban del estudio.
3. Análisis Multivariado: Para poder estudiar la contribución de una variable en el método de clasificación, las variables restantes se utilizaron en un árbol de decisión sin poda, es decir, considerando todas las posibles relaciones entre variables que presentan algún tipo de comportamiento

discriminante. Las variables no incluidas en el árbol de decisión se excluyeron del estudio.

4. Limpieza y Transformación final: Las variables seleccionadas representaban un 20 % de las originales. La base de datos poseía un número pequeño de valores perdidos (menos de 1 %) que fueron eliminados. Las variables finales se transformaron para adecuarlas al modelo de clasificación. Las variables categóricas fueron agrupadas de acuerdo a criterios comunes (por ejemplo, los distintos giros de negocio se agruparon por giro primario) y finalmente fueron binarizadas. Para esta etapa de transformación de variables es esencial la comunicación con la contraparte, principalmente en la agrupación de categorías y en la construcción de indicadores que son potencialmente relevantes a priori en base a la información que maneja la entidad financiera.

### 3.3. Elección y Construcción del Modelo de Clasificación

El método de clasificación elegido para llevar a cabo la tarea de Credit Scoring corresponde a la regresión logística, el cual es uno de los más populares en la modelación del riesgo crediticio [13], habiendo sido utilizado con éxito en diferentes países. La regresión logística cuenta con varias ventajas en comparación con otros métodos de clasificación, tales como un buen desempeño predictivo (si bien algunos modelos avanzados de minería de datos, como las redes neuronales y Support Vector Machines, suelen presentar mejores resultados debido a la capacidad de modelar complejas funciones no lineales, esta diferencia no suele ser significativa [2]), simplicidad al momento de implementar e interpretar el modelo y robustez dado que no requiere de supuestos muy estrictos sobre los datos.

Formalmente, la regresión logística pronostica un evento dicotómico  $y_i$  en base a la información de  $N$  variables independientes  $(x_1, \dots, x_N)$ . El método busca determinar la probabilidad de ocurrencia del evento dicotómico en función de la información contenida en las variables independientes, asumiendo una relación funcional como se muestra en la siguiente ecuación:

$$p(\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^N \beta_i x_i)}} \quad (2)$$

Lo anterior expresa que la probabilidad de ocurrencia del evento que se estudia (denotado por  $p(\mathbf{x})$ ) es función de los valores de las variables independientes  $\mathbf{x} = (x_1, \dots, x_N)$ . De esta manera, cuando se quiere ajustar un modelo de regresión logística a un conjunto de observaciones  $(\mathbf{x}_i, y_i)$ ,  $i = 1, \dots, m$ .

Lo más común es estimar el valor de los coeficientes  $(\beta_0, \dots, \beta_N)$  de acuerdo al método de máxima verosimilitud. En términos generales, el método de máxima verosimilitud encuentra los valores de los parámetros desconocidos que

maximizan la probabilidad de obtener el conjunto de datos observados. De esta manera se encuentran los estimadores de los parámetros y con ello se genera el modelo predictivo buscado.

Los estadísticos  $\beta$  se pueden interpretar como la influencia que tienen las variables en la probabilidad de que el cliente sea “malo” en términos de su comportamiento crediticio, asumiendo que  $y_i = 1$  equivale a un cliente etiquetado como malo e  $y_i = 0$  a uno etiquetado como bueno. Por ejemplo, si el signo de un estadístico  $\beta_i$  en particular es positivo quiere decir que esa variable explicativa es directamente proporcional a la probabilidad de no pagar íntegramente el crédito recibido. Si el signo es negativo, en cambio, la relación es inversa. En caso de que un estadístico  $\beta_i$  sea cercano a cero, la variable no afectará en la probabilidad de falla y por ende se considera irrelevante para el modelo, recomendándose su eliminación.

La pregunta que surge ahora es si, una vez encontrado los estimadores  $\beta$ , ¿son éstos estadísticamente diferentes de cero? Para responder esta pregunta se construyeron test estadísticos para determinar, con un cierto nivel de significancia, si un estadístico  $\beta_i$  es estadísticamente diferente de cero, en base a su valor y su desviación estándar muestral asociada. Este estadígrafo se conoce como test de Wald y sigue una distribución  $\chi^2$ . A partir de este test se utilizó una metodología *backward* de eliminación de atributos, que consiste en considerar todas las variables en el modelo y eliminar de manera secuencial la variable más irrelevante de acuerdo al test de Wald. Este procedimiento se repite hasta que se cuente con sólo variables relevantes en el modelo de acuerdo al test.

### 3.4. Metodología de Seguimiento

Una vez obtenidos los parámetros  $\beta$  a partir de los métodos anteriores, es posible definir el problema de seguimiento a ser resuelto. Obviamente, para realizar seguimiento a los modelos es necesario disponer de una nueva base de datos con créditos otorgados utilizando el modelo estadístico. Se dispone entonces de:

- Datos originales  $\mathbf{X}$  y parámetros originales  $\beta_j$  para cada variable  $x_j$  presente en el modelo.
- Nuevo conjunto de datos  $\mathbf{X}'$ , asociado a nuevos casos  $\mathbf{x}'_i \in \{x'_{i1}, \dots, x'_{iJ}\}$ .
- Probabilidades de *default*  $p(\mathbf{x}'_i)$  calculadas con el modelo estadístico.

Uno de los puntos interesantes de este problema es que no se cuenta con las salidas reales  $y'_i$ , es decir, no se sabe si el cliente caracterizado por  $\mathbf{x}'_i$  pagó el crédito o no lo realizó. Lo que se propone hacer, bajo estas condiciones, es

estimar una salida predicha  $\hat{y}'_i$  para cada elemento en base a las políticas internas (punto de corte decidido) para cada compañía. Dado que lo que interesa en este caso es decidir si el modelo sigue discriminando o no, es razonable asumir que sus salidas siguen siendo relevantes y a partir de ellas realizar test acerca de la desviación obtenida.

Para realizar el seguimiento en regresión logística, una de las metodologías desarrolladas con anterioridad corresponde a la utilización de un test estadístico que permita detectar si han ocurrido cambios que sean de consideración.

Para cada parámetro  $\beta_j$ , la estimación de la regresión logística entrega un intervalo de confianza para el parámetro. Este intervalo corresponde al estimador de la regresión ajustado por la desviación estándar asociada al parámetro, así se tienen límites posibles para el parámetro dados por  $\beta_j^{inf} = \beta_j - 2\sigma_{\beta_j}$  y  $\beta_j^{sup} = \beta_j + 2\sigma_{\beta_j}$ , correspondientes a los máximos y mínimos valores que puede tomar el estimador a un 95% de confianza.

El procedimiento para probar si la nueva distribución  $X'$  se ajusta a lo que el modelo es capaz de manejar corresponde a generar un nuevo conjunto de parámetros  $\beta'$  a partir del conjunto de datos  $\{X', y'\}$ . El costo de re-entrenar una regresión logística utilizando programas computacionales actuales es muy bajo, por lo que generar este nuevo conjunto es poco costoso.

Con los nuevos parámetros  $\beta'_j$  y las distribuciones estándar encontradas para estos nuevos parámetros ( $\sigma'_{\beta'_j}$ ) es posible construir un estadístico para los valores poblacionales de  $\beta'_j$ . Si la muestra posee un tamaño grande, se tiene que:

$$\frac{\beta'_j - \beta_{ref}}{\sigma'_{\beta'_j}} \rightsquigarrow t \quad (3)$$

Utilizando esto, es posible definir dos test estadísticos para medir si el nuevo parámetro se encuentra dentro de los intervalos de confianza anteriormente definidos.

$$\begin{array}{l} H_0 : \beta'_j = \beta_{inf} \quad y \quad H_0 : \beta'_j = \beta_{sup} \\ H_a : \beta'_j < \beta_{inf} \quad y \quad H_a : \beta'_j > \beta_{sup} \end{array} \quad (4)$$

Esta aplicación permite revisar si los nuevos parámetros se encuentran al interior del intervalo de confianza determinado por los parámetros antiguos, utilizando para ello la nueva estimación realizada. Se espera no rechazar las hipótesis nulas para ambos casos, donde el valor crítico para el estadístico  $t$  con infinitos grados de libertad está dado por 1,645 para el test unilateral para el límite superior y de -1,645 para el test de unilateral asociado al límite inferior. Esta aplicación debe cumplir con los siguientes requisitos:

- Se debe contar con suficientes casos en la muestra. Esto es importante por dos razones, en primer lugar, el número debe ser lo suficientemente grande para poder estimar parámetros, y en segundo lugar, la expresión (3) sólo se cumple si existe una cantidad alta de casos en muestra, es decir, el estimador  $t$  efectivamente presenta infinitos grados de libertad. En general, estos test de seguimiento se recomienda realizarlos cada tres o seis meses, de tal forma de acumular suficientes casos en muestra.
- Se deben almacenar los datos de cada caso de forma metódica. Esta es una recomendación obligatoria para cualquier aplicación real, los nuevos casos deben ser almacenados manteniendo sus variables, la probabilidad predicha y la clase seleccionada o el punto de corte utilizado para estimarla.

---

## 4. Resultados

---

Para mostrar los resultados del modelo se utilizarán los resultados reales de una de las instituciones en las que se aplicó la técnica para medir los índices de riesgo de las solicitudes. Esta institución entrega créditos a microempresarios dedicados a actividades agrícolas o ganaderas. Las variables utilizadas en la muestra de modelos aquí presentados se calcularon para clientes nuevos, sin historial crediticio, y para aquellos que si lo tenían. Estas corresponden a:

- Tenencia de propiedad: Quién es dueño del terreno. Se representa como una variable categórica, lo que implica que se modela con variables binarias, dejando una de las categorías como referencia. Existen cuatro clases: Propia (cat. base), Mediería (Tenen\_Med), Arrendado (Tenen\_Arr) y Otros(Tenen\_Otro).
- Región: Región del país dónde habita el microempresario. Categorización depende de universo.
- Edad: Edad del cliente. Puede ser transformada en el logaritmo de la edad si éste aumenta la capacidad discriminante.
- Predios: Cantidad de predios que posee el microempresario. Tres categorías: Un predio (base), dos predios (Predios\_Dos), más de dos predios (Predios\_Mas).
- Rubro del microempresario: Categorizado según universo.
- Asociadas a créditos: Variables describiendo la situación crediticia del cliente. Se dividen en dos tipos, la cantidad de créditos en la entidad

(dos variables enteras distintas, créditos cerrados y créditos vigentes) y el plazo promedio de los créditos que ha tomado.

- Asociadas a la mora: Determinan la propensión a caer en mora de los clientes. Son tres variables, si cayó en mora en alguno de los créditos que ha tomado con la institución (Con\_Mora\_Ant), el porcentaje total de las cuotas que ha pagado que cayeron en mora (Porc\_Mora) y el máximo de días que alguna cuota pasó en mora (Max\_Mora).
- Ajustes: Si han ocurrido condonaciones, ajuste de intereses, o renegociaciones, los montos asociados a las pérdidas se almacenan en la variable Ajustes.

#### 4.1. Aplicación del Modelo

El modelo fue aplicado a cinco universos distintos, obteniéndose parámetros y ajustes diversos para cada caso. En las tablas 1 y 2 se muestran los resultados<sup>1</sup> para los universos de clientes antiguos (es decir, que ya tuvieron algún otro crédito que fue pagado) con créditos de largo plazo (ALP) y de clientes nuevos con créditos de corto plazo (NCP).

Variable	$\beta$	$\sigma_{\beta_j}$	P-Valor
Tenen_Otro	0,3821	0,0966	0,0001
Tenen_Arr	0,7091	0,1805	0,0001
Tenen_Med	0,5312	0,1105	0,0000
Region_Z2	-1,0832	0,1328	0,0000
Region_Z3	-0,6011	0,1081	0,0000
Edad	-0,0053	0,003	0,0761
Predios_Dos	-1,4547	0,1029	0,0000
Predios_Mas	-2,4743	0,2232	0,0000
rubro_agric	0,0078	0,1069	0,9419
rubro_cer_prad	-0,3500	0,1169	0,0027
Constante	0,6477	0,2069	0,0017

Tabla 1: Coeficientes  $\beta$ , desviación estándar y significancia de los parámetros para clientes nuevos con créditos a corto plazo

En cada universo tanto los parámetros como las variables cambian, y ejemplifican la diferencia entre un score de comportamiento de pago (*behavioral scoring*) de uno que no lo es. El cambio en la cantidad de variables presentes

<sup>1</sup>Se han eliminado algunas variables por razones de protección de los resultados de nuestros clientes.

Variable	$\beta$	$\sigma_{\beta_j}$	P-Valor
Region_Z5	-0,6514	0,0612	0,0000
Region_Z6	-0,1875	0,0648	0,0040
Tenen_Med	0,2913	0,0700	0,0000
Tenen_Arr	0,8870	0,1666	0,0000
Tenen_Otro	0,4296	0,0595	0,0000
Dos_Predios	-0,5653	0,0531	0,0000
Mas_Predios	-0,9672	0,0648	0,0000
Inedad	-0,9950	0,0898	0,0000
Creditos_Cerrados	-0,0441	0,0074	0,0000
Creditos_Vigentes	-0,1509	0,0248	0,0000
Duracion_Creditos	0,0950	0,0237	0,0000
Con_Mora_Ant	17,3080	0,0593	0,0000
Porc_Mora	0,1636	0,0852	0,0550
Mora_Max	0,0019	0,0001	0,0000
Constant	16,1410	0,3695	0,0000

Tabla 2: Coeficientes  $\beta$ , desviación estándar y significancia de los parámetros para clientes antiguos con créditos a largo plazo

entre los distintos tipos de modelos es relevante, pues en los modelos de comportamiento tienen un peso mucho mayor los historiales de crédito del cliente y, sobre todo, las moras que haya manifestado al interior de la empresa. La conclusión que se desprende de estas tablas es que la variable principal para determinar el comportamiento de estos clientes es cuán ordenados son en sus cuentas y su propensión a desordenarse, aunque sea poco tiempo. Este efecto es aún más relevante que el ingreso del microempresario, pues la gran mayoría poseen ingresos concentrados en un pequeño intervalo de ganancias (sección 2.1)

lo que no permite diferenciar en gran manera. Esta es una diferencia importante con respecto al segmento de personas clásico, donde los indicadores de deuda y, sobre todo, las proporciones de deuda e ingreso, son variables fundamentales.

La diferenciación entre un score de comportamiento y uno que evalúa solicitudes también tiene un impacto en la capacidad de predicción del modelo, como se puede observar en la tabla 3.

El ajuste de los modelos, de todos modos, es razonable para ambos universos. El pago o no pago de un compromiso crediticio corresponde a un fenómeno social de alta complejidad, por lo que se esperan resultados con un rango entre 60-80% de efectividad global.

Universo	No Defaulters	Defaulters
NCP	64,54 %	77,80 %
ALP	76,10 %	72,20 %

Tabla 3: Porcentaje de acierto para cada universo, por pagadores (No Defaulters) y no pagadores (Defaulters).

## 4.2. Seguimiento

Para el experimento de seguimiento, se dividió la muestra en créditos otorgados entre los años 2000 a 2004 y los otorgados con posterioridad de esta fecha. La institución que otorga los créditos conocía la ocurrencia de un cambio entre estos años, por lo que se esperaba que los test entregaran una diferencia significativa. Los resultados se observan en la tabla 4.

Variable	$\beta'$	$\sigma_{\beta'_j}$	$\beta$	Lim. Inf.	Lim Sup.	t Inf.	t Sup
Region_Z8	,696	,100	,794	,592	,995	1,03	-2,98
Region_Z9	,340	,089	,394	,208	,580	1,48	-2,71
Tenen_Med	-,785	,077	-,177	-,335	-,019	<b>-5,83</b>	-9,92
Tenen_Arr	-1,136	,138	-,364	-,682	-,047	<b>-3,30</b>	-7,92
Tenen_Otro	-,728	,080	-,264	-,432	-,096	<b>-3,71</b>	-7,93
Predios_Dos	1,150	,069	,151	,005	,298	16,56	<b>12,32</b>
Predios_Mas	1,845	,089	,495	,325	,665	17,01	<b>13,21</b>
Ajustes	,467	,085	,057	-,004	,117	5,54	<b>4,11</b>
Creditos_Cerrados	,132	,011	,085	,069	,100	5,75	<b>2,97</b>
Porc_Mora	-1,506	,103	-1,548	-1,795	-1,302	2,81	-1,98

Tabla 4: Resultados para el modelo de seguimiento. En negrillas aquellos cambios significativos.

Diversas variables presentan cambios significativos, destacando aquellas asociadas a los predios, pues son variables categóricas cuyo significado está unido al valor de las demás variables que forman las clases. El test detecta correctamente cambios en los intervalos asociados y, como era de esperarse, los cambios se ven reflejados para todas las categorías.

Para ejemplificar el hecho que las variables señaladas por el test sí detectan cambios relevantes, la figura 2 muestra la situación asociada a la variable Creditos\_Cerrados, con una clara desviación entre ambos años.

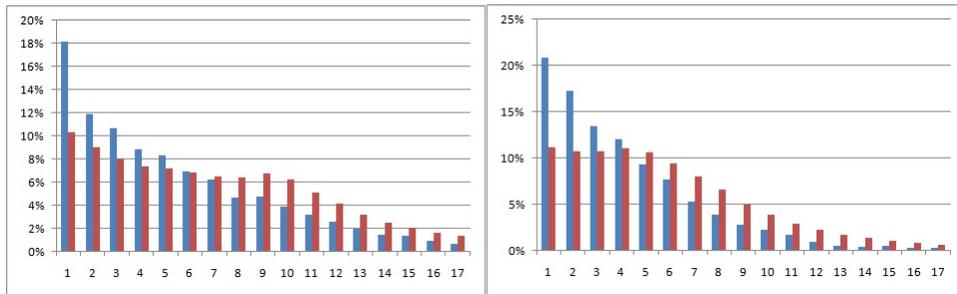


Figura 2: Cambios observados en la variable `Creditos_Cerrados` para los años 2000-2004 (izq.) y 2005 en adelante (der.).

---

## 5. Conclusiones

---

Los créditos a microempresarios corresponden a un mercado de gran importancia para Chile y Latinoamérica, pues las microempresas representan una parte significativa de las fuentes de trabajo, lo que hace que el otorgarle créditos no sea ya un negocio asociado a programas de apoyo, sino una oportunidad real y formal para mejorar las condiciones que enfrentan. A partir de esto, la necesidad por modelos de riesgo que estén adecuados a los fenómenos sociales que estas empresas enfrentan ha surgido en los últimos años.

Los microempresarios presentan características únicas que provocan que los estudios de riesgos sean asociados completamente a la realidad de los países donde se desarrollan, impulsando la investigación científica y social como la que se presenta en este trabajo.

Desde el punto de vista de los modelos, las técnicas clásicas siguen entregando buenos resultados, pero es en la selección y construcción de variables donde se hace la diferencia con los modelos de riesgo clásicos. Situaciones como que el ingreso que posee el microempresario no sea relevante para la determinación del pago o no pago del crédito destaca como una de las razones por las que requieren estudios en profundidad. Es en las variables que muestran solidez financiera (tenencia y cantidad de bienes, por ejemplo) u orden en los pagos que realizan donde se encuentra la información que permite determinar la ocurrencia del fenómeno en estudio.

Otra necesidad importante que presentan estos modelos corresponde a realizar un seguimiento detallado del funcionamiento de éstos, pues los microempresarios están inmersos en un mercado muy volátil, siendo muy sensibles a vaivenes de la economía y presentando, por la naturaleza de su operación, un dinamismo mucho más grande que el que presentan las empresas de mayor tamaño. Así, el desarrollar herramientas que permitan determinar el momento

cuando ha ocurrido un cambio que daña la capacidad predictiva del modelo es una interrogante atractiva para los investigadores del área. El modelo aquí presentado cumple con este objetivo, siendo simple de implementar y entregando muy buenos resultados.

En cuanto a los resultados de la medición, se observan ajustes totalmente en línea con lo que se observa en las bancas de personas y de empresas, lo que avala el uso a nivel global tanto por parte de instituciones gubernamentales como privadas. A medida que aumente el interés por parte de privados para otorgar estos créditos, se hará más relevante su estudio y permitirá mejorar las condiciones que enfrentan estas empresas, sobre todo en países desarrollados.

**Agradecimientos:** El primer y segundo autor desean agradecer a CONICYT por las becas que permiten la realización de esta publicación. Este trabajo fue parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16).

## Referencias

- [1] Altman, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23, 589-609. 1968.
- [2] Baesens, B, Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. y Vant-hienen, J. Benchmarking state of the art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54(6):627-635, 2003.
- [3] Beaver, W. H. Financial ratios as Predictors of Failure. *Journal of Accounting Research* 4, 71-111, 1966.
- [4] Bravo, C., Maldonado, S. y Weber, R. Seguimiento en Modelos de Regresión Logística. *Revista de Ingeniería Industrial* Año 8, N° 2: 31-44. 2009.
- [5] Bravo, F. y Pinto, C. Modelos predictivos de la probabilidad de insolvencia en microempresas chilenas. *Contaduría Universidad de Antioquia* 53, 13-52. 2008.
- [6] Coloma, P., Weber, R., Guajardo, J. y Miranda, J. Modelos analíticos para el manejo del riesgo de crédito. *Trend Management* 8: 44-51, 2006.
- [7] Fayyad, U. Data mining and knowledge discovery- making sense out of data. *IEEE Expert-Intelligent Systems and Their Applications* 11:20-25, 1996.

- [8] Gerardi, K. S., Lehnert, A., Sherlund, S. M. y Willen P. S. Making Sense of the Subprime Crisis. *Public Policy Discussion Paper of the Federal Reserve* 09-1, Bank of Boston, 2009.
- [9] Gerding, E. F. The Outsourcing of Financial Regulation to Risk Models and the Global Financial Crisis: Code, Crash, and Open Source *Washington Law Review*, Forthcomming, 2010.
- [10] Maldonado, S. y Weber, R. A wrapper method for feature selection using Support Vector Machines. *Information Sciences* 179 (13), 2208-2217, 2009.
- [11] D. Martens, B. Baesens, T. Van Gestel y J. Vanthienen. Comprehensive Credit Scoring Models using Rule Extraction from Support Vector Machines. *European Journal of Operational Research* 183(3): 1466-1476, 2006.
- [12] Ohlson, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18, 109-131, 1980.
- [13] Thomas, L. C. A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting* 16(2), 149-172. 2000.
- [14] Yang, J. y Liu, G. The evaluation of classification models for credit scoring. *Arbeitsbericht Institut für Wirtschaftsinformatik, Georg-August-Universität Göttingen*. 2, 2002.
- [15] Zeira, G., Last, M. y Maimon, O. Segmentation on Continuous Data Streams Based on a Change Detection Methodology. En: *Advanced Techniques in Knowledge Discovery and Data Mining*, pp. 103-126, Springer. 2005.

---

# MODELO DE PLANIFICACIÓN DE PRODUCCIÓN PARA UN SISTEMA MULTIPRODUCTO CON MÚLTIPLES LÍNEAS DE PRODUCCIÓN

---

RODRIGO VIVEROS E.\*  
EDUARDO SALAZAR H.\*\*

## Resumen

*Se presenta un modelo para la planificación de la producción en un sistema multiproducto con múltiples líneas de producción de la industria de fabricación de medios de acero para la molienda de minerales. A partir de un enfoque tradicional de planificación de la producción, se desarrolla un modelo de programación lineal, el cual está basado en el pronóstico de ventas, inventarios iniciales de productos terminados y materias primas, capacidad de producción instalada, como también de las rutas alternativas de producción de los productos. El modelo permite que el déficit de producto pueda suplirse en periodos posteriores cuando exista disponibilidad de recursos. El modelo entrega las tasas de producción por línea de producción, la utilización de la capacidad instalada, el balance de inventario de productos y materias primas, como también la planificación de las adquisiciones de materias primas, en un horizonte de planificación anual dividido en periodos mensuales.*

**Palabras Clave:** *Planificación de producción, Multiproducto, Programación Lineal, Optimización.*

---

\*Ingeniero de Planificación Moly-Chile S.A. Concepción, Chile.

\*\*Departamento de Ingeniería Industrial Facultad de Ingeniería, Universidad de Concepción, Concepción, Chile.

---

## 1. Introducción

---

La planificación de la producción en un enfoque clásico, se plantea de manera jerárquica en cuanto a sus decisiones y plazos involucrados, en el que se logra una integración vertical entre los objetivos estratégicos, tácticos y operativos, los que idealmente también deben expresar la relación horizontal entre las diferentes áreas de la compañía.

Varios autores, como Chase [3] y Domínguez Machuca [4] y [5], plantean las bases de la planificación global, siendo estos últimos quienes consideran explícitamente esta integración en ambos sentidos. Otro enfoque más reciente proporcionan Heizer y Render en [9] y [10], incluyendo además de los aspectos integradores en los dos sentidos, los desafíos actuales de la dirección de operaciones con una perspectiva global de la empresa.

Numerosos autores han propuesto diversos modelos de optimización para resolver el problema de planificación agregada. Desde una perspectiva clásica Hax y Candea [8], describen y clasifican algunos de estos enfoques para resolver este problema, con la participación del uso de técnicas de programación matemática, procedimientos heurísticos y técnicas de búsqueda. En este mismo contexto Nam y Logendran[12], hacen una exhaustiva revisión de los modelos y métodos de planificación en su conjunto, presenta un esquema de clasificación de las técnicas de planificación agregada en dos grandes grupos: las técnicas que alcanzan soluciones óptimas y las que no las garantizan. Pochet y Wolsey [13], constituyen una referencia general, ilustran una amplia variedad de problemas de planificación de producción tanto en su modelamiento a través de programación lineal entera mixta como algorítmico para su resolución, .

En la resolución de problemas de planificación agregada propiamente tal, Gazmuri y Arrate [6] desarrollan un modelo determinístico que resuelve el problema de planificación agregada de una empresa de electrodomésticos. Este problema se plantea como un modelo de programación lineal entera mixta, cuya función objetivo es maximizar el beneficio considerando la dotación de mano de obra como factor determinante de la capacidad de producción, asumiendo una demanda conocida para cada periodo y producto. A partir de este modelo, que es tomado como base, Albornoz y Contesse [2] aborda un problema real de planificación agregada, que contempla la elaboración de múltiples productos bajo demanda incierta, y lo resuelven mediante el uso de diversos modelos de optimización robusta. A los modelos robustos, suman otros modelos determinísticos que toman en cuenta la incertidumbre en la demanda de una manera más sencilla. Ambos enfoques consideran la fuerza laboral como una variable del problema, con sus repercusiones en la capacidad de producción y en los costos.

Entre otros modelos de planificación agregada que han sido aplicados a la industria de los alimentos con estructura de demanda estacional, en Tadei *et al.* [14], que para el caso de una empresa que fabrica múltiples productos, algunos estacionales y otros no, proponen un modelo de programación lineal entera mixta para resolver el problema de la planificación agregada de la producción en un horizonte de un año. El objetivo del modelo es minimizar los niveles de stock necesarios, sin permitir ruptura de stock, para satisfacer una demanda conocida, determinando la cantidad de personal necesario para cada mes del año. De este modo es posible obtener la cantidad de personal que se debe subcontratar en los meses de demanda alta. En este mismo contexto, Takei y Mesquita [15], proponen para empresa fabricante de alimentos (helados) con tres familias de productos, también con alta demanda estacional, un modelo de planificación agregada para un año mediante programación lineal entera mixta, cuyo objetivo es maximizar el margen bruto, sin permitir ruptura de stock. El modelo determina las tasas de producción mensual y los niveles de inventario de productos terminados, así como la mano de obra necesarios para cumplir los planes de producciones.

Gomes da Silva *et al.* [7], consideran tres aspectos importantes para tener en cuenta en la planificación agregada de la producción: el beneficio de la compañía, la satisfacción del cliente, y el ambiente de trabajo. Proponen el desarrollo de un modelo de programación lineal entera mixta con múltiples criterios, esto es, tres funciones objetivo por separado: la primera maximiza el beneficio de la compañía, la segunda minimiza los retrasos de los pedidos de los clientes, y la tercera minimiza las variaciones en la fuerza laboral. Todas estas a lo largo de un horizonte de tiempo de un año dividido en periodos mensuales.

También existen modelos de planificación que tienen un enfoque orientado a operaciones con multi-plantas, donde se desprende claramente el enfoque de integración de toda la cadena de suministro. Jolayemi y Olorunniwo [11] elaboran un modelo determinístico de planificación de producción y transporte de multi-plantas y multi-almacenes de capacidad variable, cuyo objetivo es maximizar los beneficios, integrando en él los problemas de producción transporte y capacidad de almacenamiento, permitiendo la subcontratación y procurando abarcar todos los aspectos de la cadena de suministros. En el análisis de este caso no se han considerado estructuras de costos decrecientes para el transporte.

Timpe y Kalltrath [16], proponen un modelo de planificación de producción para la industria química mediante programación lineal entera mixta que busca abarcar toda la cadena de suministro, incorporando como elemento adicional el uso de las diferentes escalas de tiempo existentes entre la producción y la comercialización, en un sistema productivo de multi-planta, cuya demanda de productos considera el origen de los productos suministrados, incluyendo la

posibilidad de satisfacer la demanda con producción externa. El modelo responde a una variedad de objetivos, por lo que abarca varias funciones objetivo.

Vercelis [17] y Aghezzaf [1] han desarrollado modelos de planificación de producción y de asignación de capacidades junto con la gestión de almacenes para problemas de gran tamaño en ambientes multi-planta. Se desarrolla un modelo de programación lineal entera mixto, que para ser resuelto se introducen técnicas de relajación matemática y/o relajación de Lagrange, para luego aplicar una técnica heurística de resolución basado en el modelo relajado.

Este trabajo está organizado de la siguiente manera: en la sección dos, se presenta una breve descripción del problema de planificación. La formulación y detalles del modelo de optimización es tratada en la sección tres. Algunos de los principales resultados del modelo son analizados en la sección cuatro, y finalmente en la sección cinco se establecen las conclusiones de este trabajo.

---

## 2. Descripción del Problema

---

La planificación de la producción es un proceso continuo cuyo objetivo es determinar anticipadamente decisiones que permitan optimizar el uso de los recursos productivos. Se refiere a las decisiones tácticas para determinar las actividades y los recursos de la empresa a mediano plazo. El plan agregado debe ser coherente con el plan estratégico a largo plazo, y su desagregación progresiva determina planes y programas operativos, a corto plazo. Su enfoque es la determinación de la cantidad de producción, los niveles de inventarios y la cantidad de recursos necesarios con la finalidad de satisfacer la demanda para un horizonte temporal de planificación específico de mediano plazo.

Este trabajo se realizó para la planificación de producción de una empresa chilena fabricante de bolas de acero para la molienda de minerales, líder en el mercado nacional con distribución de sus productos a diferentes zonas geográficas, incluido el MERCOSUR, Norteamérica y Sudáfrica. El departamento de planificación de producción es responsable de planificar la producción y las compras de materias primas en base al análisis de los pronósticos de demanda, determinando niveles de producción por línea, uso de capacidad instalada y niveles de inventarios de productos y materias primas para cada mes del horizonte de planificación. Los diferentes productos fabricados se diferencian por sus dimensiones físicas expresadas en el diámetro de las bolas de molienda. Su paleta de productos es de trece artículos.

La empresa opera con estrategia de "make to stock", debido a que la capacidad se hace variable a lo largo del horizonte de planificación (12 meses), presentando sus clientes (mercado) una demanda relativamente estable. A su vez los productos son de uso general, esto es, todos los clientes tienen demandas de productos similares en sus respectivas faenas.

La empresa no puede variar su capacidad de producción en el corto plazo y, como consecuencia, la admisión de trabajadores temporales no implica cambios en la capacidad instalada. La empresa opera actualmente con cuatro turnos de producción, esto es, tres turnos operan las plantas en turnos de ocho horas cada uno y uno descansa, de acuerdo a un programa de rotación establecido. Esto significa que el funcionamiento de las plantas es permanente, incluyendo domingos y festivos y trae como consecuencia, que desde la perspectiva de este trabajo, la mano de obra no se considera como un factor de cambio en la capacidad productiva.

---

### 3. Modelo de Planificación

---

Para desarrollar el modelo de planificación agregada de producción y abastecimiento de materias primas, se consideran las previsiones de demanda de bolas de acero como dato de entrada, así como los inventarios iniciales de materias primas, y de productos, los costos variables y las tasas de producción y rendimiento de la materia prima. El modelo propuesto considera un horizonte de planificación de doce meses y las soluciones deben ser revisadas mensualmente.

Los siguientes supuestos fueron considerados en el modelo desarrollado:

- La capacidad disponible corresponde a los días de cada mes en tres turnos diarios de producción. Por razones de costo se descuenta el tiempo en el que existen restricciones en el uso de potencia eléctrica en horario de punta (modulación eléctrica), y el tiempo destinado a los mantenimientos programados que implican una parada de planta.
- Los operarios están asignados a cada línea de producción.

El modelo de planificación agregada propuesto involucra todas las variables de decisión para la planificación de la producción, las que se describen a continuación:

- Volumen de producción por diámetro de producto en cada mes del horizonte de planificación.
- Volumen de compras por diámetro de materias primas (barras de acero) en cada mes del horizonte de planificación.
- Necesidades de materias primas (barras de acero) para la producción, por diámetro de barras, en cada mes del horizonte de planificación.
- Niveles de inventario final de producto por diámetro en cada mes del horizonte de planificación.

- Niveles de inventario final de materias primas (barras) por diámetro en cada mes del horizonte de planificación.

Debido a que la programación de la producción se realiza posterior al proceso de planificación, el modelo no sugiere las secuencias en que los diferentes productos se producirán en las líneas de producción. La programación de producción, junto al proceso de determinación de los lotes de producción se considera para un trabajo futuro.

### 3.1. Definición de parámetros y variables de decisión

En esta sección se definen los parámetros y las variables de decisión del modelo generalizado de planificación de producción, considerando  $N$  productos,  $J$  líneas de producción,  $L$  materias primas y  $T$  períodos en el horizonte de planificación.

Se utilizan los siguientes índices (y sus variaciones) para identificar productos, líneas de producción, materias primas y períodos:

- $i$  producto (bola de diámetro  $i$ );  $i = 1, \dots, N$
- $j$  línea de producción;  $j = 1, \dots, J$
- $l$  materia prima (barras de diámetro  $j$ );  $l = 1, \dots, L$
- $t$  periodo (mes);  $t = 1, \dots, T$

Los parámetros y variables de decisión relevantes para el problema de planificación de producción se especifican a continuación, mientras que el modelo de programación lineal se presenta en la sección 3.2.

#### Parámetros

- $c_{ij}$  costo de producción del diámetro  $i$  en la línea  $j$  [US\$/ton].
- $h_i$  costo de inventario de producto de diámetro  $i$  [US\$/ton/mes].
- $e_i$  costo (penalización) de escasez de productos del diámetro  $i$  [US\$/ton/mes].
- $b_l$  costo de inventario de materia prima (barras) tipo  $l$  [US\$/ton/mes].
- $d_{it}$  demanda de producto (bola) de diámetro  $i$  durante el mes  $t$ .
- $p_{ij}$  tasa de producción de producto (bola) de diámetro  $i$  en línea  $j$  [ton/hora].
- $cap_{jt}$  horas disponibles para producir en la línea  $j$  en el período  $t$  [horas].
- $r$  constante de rendimiento metálico de las materias primas.
- $s_{lt}$  suministro máximo de materia prima  $l$  en periodo  $t$  [ton].
- $\pi_{i0}$  inventario inicial del producto  $i$  [ton].
- $\pi_{iT}$  inventario final de producto  $i$  al final del horizonte de planificación [ton].
- $\alpha_{l0}$  inventario inicial de barras tipo  $l$  [ton].
- $a_{il}$  vale 1 si producto  $i$  puede fabricarse a partir de barras tipo  $l$ , 0 en caso contrario.
- $T$  horizonte de planificación [mes].

#### Variables de decisión

- $x_{ijt}$  producción del diámetro  $i$  en la línea  $j$  en período  $t$  [ton].  
 $y_{lt}$  materia prima del tipo  $l$  a recepcionar en el periodo  $t$  [ton].  
 $z_{lt}$  necesidades de materias primas del tipo  $l$  en el periodo  $t$  [ton].  
 $\pi_{it}$  balance de inventario del producto  $i$  en el periodo  $t$  [ton].  
 $\pi_{it}^+$  inventario físico del producto  $i$  en el periodo  $t$  [ton].  
 $\pi_{it}^-$  inventario faltante (escasez) del producto  $i$  en el periodo  $t$  [ton].  
 $\alpha_{lt}$  inventario final de barras  $l$  en periodo  $t$  [ton].

### 3.2. Modelo de programación lineal

El objetivo del modelo es minimizar el costo de la planificación (ecuación 1) en el horizonte de planificación, expresado como el costo de fabricación (costo de conversión, es decir, costo de transformación de barras a bolas), costo de inventario de productos y su penalización por escasez, más el costo de mantener materia prima en inventario.

$$\text{mín } C_p = \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^J c_{ij} \cdot x_{ijt} + \sum_{t=1}^T \sum_{i=1}^N h_i \cdot \pi_{it}^+ + \sum_{t=1}^T \sum_{i=1}^N e_i \cdot \pi_{it}^- + \sum_{t=1}^T \sum_{l=1}^L b_l \cdot \alpha_{it} \quad (1)$$

$$\pi_{i,t-1} + \sum_{j=1}^J x_{ijt} - d_{it} = \pi_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (2)$$

$$\pi_{it} = \pi_{it}^+ - \pi_{it}^- \quad i = 1, \dots, N; t = 1, \dots, T \quad (3)$$

$$\pi_{iT} = \frac{D_{iT}}{2} \quad i = 1, \dots, N; \quad (4)$$

$$\sum_{i=1}^N x_{ijt} \cdot \frac{1}{p_{ij}} \leq \text{cap}_{jt} \quad j = 1, \dots, J; t = 1, \dots, T \quad (5)$$

$$\alpha_{l,t-1} + y_{lt} - z_{lt} = \alpha_{lt} \quad l = 1, \dots, L; t = 1, \dots, T \quad (6)$$

$$y_{lt} \leq s_{lt} \quad l = 1, \dots, L; t = 1, \dots, T \quad (7)$$

$$z_{lt} = r \cdot \sum_{i=1}^N a_{il} \sum_{j=1}^J x_{ijt} \quad l = 1, \dots, L; t = 1, \dots, T \quad (8)$$

$$\alpha_{lT} = \sum_{t=1}^T z_{lt} / T \quad l = 1, \dots, L \quad (9)$$

$$x_{ijt}, \pi_{it}^+, \pi_{it}^-, y_{lt}, z_{lt}, \alpha_{lt} \geq 0, \quad \pi_{it} \in \mathfrak{R}, \quad (10)$$

$$i = 1, \dots, N; j = 1, \dots, J; t = 1, \dots, T$$

El conjunto de restricciones del modelo son representados por las ecuaciones (2) a la (9) que en síntesis se describen de la siguiente manera: la ecuación (2) representa el balance de inventario de productos terminados; la planificación de inventario físico y escasez es descrita por la ecuación (3); y los requerimientos de inventario de productos al término del horizonte de planificación son expresados por la ecuación (4), de acuerdo a los requisitos de la política de inventarios de la empresa, equivalente a la mitad de la demanda del periodo  $T$  (último periodo).

La ecuación (5), plantea las restricciones de capacidad de producción de cada una de las líneas, de modo que la producción a fabricar relacionado con los diferentes rendimientos asociados a las posibilidades de producción que poseen los equipos no puede superar las horas disponibles de cada línea.

El balance de materias primas se expresa a través de la ecuación (6), el inventario inicial de materias primas más las adquisiciones y uso de éstas, derivado de la producción en cada periodo, es igual al inventario final de materias primas. La ecuación (7) establece los límites de adquisición de materias primas, esto es, las compras totales a realizar a cualquier proveedor de barras no puede superar la oferta de suministro de cada uno. La ecuación (8) muestra las necesidades de materias primas en función de la fabricación de cada producto. La ecuación (9) el inventario final de materia prima, la que se expresa como inventario esperado de materias primas por cada tipo en planta, igual al promedio de utilización de materia prima por periodo.

---

## 4. Resultados

---

Con el fin de validar el modelo, las pruebas se llevaron a cabo considerando una planta con tres líneas de producción, vale decir, el modelo se probó en una de las plantas del sistema productivo existente, la que actualmente produce 13 productos a partir de 10 tipos de materias primas, los tiempos de producción y los patrones de demanda son tomados en un caso tipo. La dimensión del problema implementado corresponde a 1020 variables continuas y 610 restricciones. El modelo se implementó en **MS Excel** <sup>TM</sup> con el uso de **Premium Solver Platform** <sup>TM</sup>.

Las pruebas simuladas para el periodo enero – diciembre, permiten verificar el comportamiento de algunas variables para el periodo planificado: niveles de producción, niveles de inventarios, uso de la capacidad instalada, oportunidades donde se genera escasez de productos y los costos asociados. Se evaluó el comportamiento del modelo para un caso basado en la demanda proyectada para el año 2009. Se analizaron los resultados del modelo desde los siguientes puntos de vista: niveles de producción, planificación de los inventarios de productos terminados y de materias primas en el horizonte de planificación, y costos de operación y de inventarios.

En la tabla (1) se presenta el resultado, donde el modelo entrega en forma agregada las tasas mensuales de producción, niveles de inventarios finales por mes, y uso de la capacidad, expresada en porcentaje de utilización de las líneas de producción, y la planificación de compra de materias primas. Se observa que se produce una escasez de producto de diámetro 1" (ver tabla (2)), esto nos muestra una característica destacable del modelo el cual prioriza la fabricación de aquellos diámetros donde se incurre en menores costos y tienen mayores velocidades de producción. Al analizar el diámetro donde se produce la escasez, se trata de un diámetro que demanda más recursos económicos y además es el diámetro que posee la menor velocidad de producción, y por ende, demanda mayor cantidad de horas en la fabricación por lote.

Periodo	Producción (ton)	Utilización (%)			Inventarios (ton)		Compra de Materias Primas (ton)
		Línea 1	Línea 2	Línea 3	Productos	Materias Primas	
Enero	11,879	100.0	100.0	100.0	1,133	13,577	25,698
Febrero	12,171	100.0	100.0	100.0	-74	13,577	12,419
Marzo	11,206	62.9	50.4	96.6	58	13,577	11,434
Abril	11,520	86.5	100.0	100.0	0	13,577	11,755
Mayo	11,772	82.3	88.0	85.8	456	13,577	12,012
Junio	13,091	100.0	100.0	100.0	2,705	13,577	13,358
Julio	12,571	100.0	100.0	100.0	1,040	13,577	12,828
Agosto	14,184	100.0	100.0	100.0	1,943	13,577	14,474
Septiembre	13,072	100.0	100.0	100.0	2,711	13,577	13,339
Octubre	15,567	100.0	100.0	100.0	4,639	13,577	15,885
Noviembre	15,409	100.0	100.0	100.0	6,691	13,577	15,723
Diciembre	17,223	100.0	100.0	100.0	12,282	13,577	17,574
<b>Total</b>	<b>159,665</b>	<b>94.0</b>	<b>94.8</b>	<b>98.4</b>	<b>12,282</b>	<b>13,577</b>	<b>176,500</b>

Tabla 1: Planificación agregada y utilización de líneas de producción.

Otra característica a destacar del modelo, surge del análisis de los saldos de producto terminado, en el cual se observa que el modelo es capaz de anticipar la producción para cumplir con la demanda. En la tabla 2 se presentan los saldos de productos para cada periodo del horizonte de planificación y donde se observa que al existir saldos con valores positivos mayores a cero en los meses anteriores a diciembre, tales cifras demuestran generación de stock.

Inventarios (ton)	Total	1.0	1.25	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.25	5.5	6.0
Inicial	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Enero	1,133	-43	0	0	0	278	0	898	0	0	0	0	0	0
Febrero	-74	-74	0	0	0	0	0	0	0	0	0	0	0	0
Marzo	58	58	0	0	0	0	0	0	0	0	0	0	0	0
Abril	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mayo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Junio	991	745	0	0	0	0	0	246	0	0	0	0	0	0
Julio	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agosto	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Septiembre	469	469	0	0	0	0	0	0	0	0	0	0	0	0
Octubre	1,126	1,126	0	0	0	0	0	0	0	0	0	0	0	0
Noviembre	2,111	1,334	0	0	0	0	0	777	0	0	0	0	0	0
Diciembre	6,141	451	171	324	121	350	2,508	585	170	0	1,294	0	167	0

Tabla 2: Inventarios de productos terminados.

Del análisis de los resultados de la optimización, se observó a través de la comparación de los resultados del modelo con datos reales, que el sistema es capaz de incrementar la productividad operativa en un 3.6 %, con una reducción de costos de un 5 % evaluada en términos de la función objetivo, esto debido fundamentalmente a la disminución de los inventarios a lo largo del horizonte de planificación.

A pesar de que en la Tabla 1 se muestran usos de capacidad del 100 % en algunos meses, esto se debe entender, cómo el modelo hace uso del parámetro de capacidad al resolver la instancia evaluada. Dada la existencia de costos de inventario en la función objetivo, el modelo evitará el uso de inventarios y anticipará sólo la producción estrictamente necesaria que permita satisfacer alta demanda de meses futuros, lo que puede llevar a que el modelo determine niveles de producción al límite de la capacidad en algunos meses.

En la práctica esto no necesariamente representa 100 % de uso de la capacidad instalada, sino que el uso de la capacidad traspasada al modelo como parámetro. De esta forma se puede controlar el nivel de ocupación real de la planta. Por otro lado, el modelo provee actualmente al tomador de decisión de una herramienta que le permite redistribuir, en caso necesario, la producción de manera de balancear el uso de la capacidad instalada en el tiempo.

La demanda de producto es relativamente estable en el mediano plazo (plazo considerado en el modelo), y en gran medida obedece a contratos suscritos con anterioridad, sin perjuicio de que existe variabilidad mes a mes, la que es controlada mediante sucesivas planificaciones.

---

## 5. Conclusiones

---

En esta investigación se aborda el problema de planificación de producción para una empresa de fabricación de medios de molienda para la minería con dos plantas productivas ubicadas en Chile, a través de un modelo de programación lineal para la solución del problema. El modelo resultó ser una herramienta útil en la administración de producción, obteniendo planes de producción factibles y que no contradicen las buenas prácticas.

Del análisis de los resultados obtenidos a partir del modelo, se observó a través de la comparación de éstos con datos reales, que el sistema es capaz de incrementar la productividad operativa en un 3.6 %, con una reducción de costos en un 5 %, esto debido fundamentalmente al mejor uso de la capacidad y de la disminución de los inventarios a lo largo del horizonte de planificación.

Dada la relativa estabilidad de la demanda, en el modelo desarrollado ésta se trata como un dato de entrada (parámetro), sin embargo, en el modelo se consideran stocks de seguridad al final del horizonte de planificación. En caso necesario, el tratamiento de la incertidumbre en la demanda puede hacerse a

través del análisis de diferentes escenarios que originan diferentes planes de producción.

El modelo se ha desarrollado con un enfoque flexible, adaptable a cambios que puedan producirse en las instalaciones (por ejemplo, incorporación de una nueva línea de producción), y a cambios de escenarios (por ejemplo, cambios en el patrón de demanda).

En esta investigación se ha asumido un comportamiento lineal en los costos de fabricación, supuesto considerado como aceptable, dada la estructura del proceso productivo y del nivel de uso de las instalaciones. En un nivel bajo de utilización de la capacidad, los costos fijos deben ser absorbidos por pocas unidades, aumentando el costo por unidad producida, mientras que ha medida que el nivel de utilización de la capacidad se incrementa, más unidades absorben el costo fijo, disminuyendo el costo por unidad producida. Otros factores como la fatiga de los trabajadores, fallas de equipos, la pérdida de precisión, y en general, una mayor dificultad en la coordinación de las operaciones, son elementos que influyen en la linealidad de los costos de fabricación, sin embargo, estos factores generalmente son tratados como factores que no afectan la linealidad de los costos de fabricación.

Debido a que la programación de la producción se realiza posterior al proceso de planificación, el modelo no sugiere las secuencias en que los diferentes productos se producirán en las líneas de producción. A partir del modelo propuesto, se recomienda continuar la desagregación en el proceso de planificación, desarrollando la programación de producción, junto a la determinación de los lotes de producción. Esto permitiría un mayor grado de integración de las operaciones.

Es posible ampliar el modelo propuesto, para integrar toda la cadena de suministro, buscando optimizar la producción de la demanda en las diferentes plantas, minimizando el costo de producción y de transporte de materias primas desde los proveedores y de productos terminados hacia los clientes.

## Referencias

- [1] Aghezzaf, E.-H. Production Planning and Warehouse Management in Supply Networks with Inter-Facility Mold Transfers. *European Journal of Operational Research*, Vol. 182 (3):1122-1139, 2007.
- [2] Alborno, V., Contesse, L. Modelos de Optimización Robusta para un Problema de Planificación Agregada de la Producción bajo Incertidumbre en las Demandas. *Investigación Operativa*, Vol. 7(3): 1-15, 1999.
- [3] Chase, R., Aquilano, N., Jacobs, R. Administración de producción y operaciones - Manufactura y servicios. *Irwin/McGraw-Hill*, 8<sup>va</sup> Ed. Págs. 4-19, 548-569.
- [4] Domínguez-Machuca, J., Alvarez, Ma.J., García,S., Domínguez-Machuca, M. y Ruíz,A. (1995a): Dirección de Operaciones – Aspectos Estratégicos en la Producción y los Servicios. *McGraw-Hill*, Págs. 63-90, 141-146. 1995.
- [5] Domínguez-Machuca, J., Alvarez, Ma.J., García, S., Domínguez-Machuca, M. y Ruíz, A. (1995b): Dirección de operaciones: Aspectos tácticos en la producción y los servicios. *McGraw-Hill*, Págs. 3-28, 31-54, 63-115. 1995.
- [6] Gazmuri, P., Arrate, I. Modeling and Visualization for a Production Planning Decision Support System. *International Transactions in Operational Research*, Vol. 2(3): 249-258, 1995.
- [7] Gomes da Silva, C.; Figueira, J.; Lisboa, J.; Barman, S. An interactive decision support system for an aggregate production planning model based on multiple criteria mixed integer linear programming. *Omega*, Vol. 34(2): 167-177, 2006.
- [8] Hax, A., Candea, D. Production and Inventory Management. *Prentice-Hall*. Págs. 69-101. 1984.
- [9] Heizer, J., Render, B. Dirección de la producción y operaciones: Decisiones estratégicas. *Pearson Educación*, 8<sup>va</sup> Ed., Págs. 3-24, 36-61. 2007.
- [10] Heizer, J., Render, B. Dirección de la producción y operaciones: Decisiones tácticas. *Pearson Educación*, 8<sup>va</sup> Ed., Págs. 109-134. 2008.
- [11] Jolayemi, J.; Olorunniwo, F. A Deterministic Model for Planning Production Quantities in a Multi-Plant, Multi-Warehouse Environment with Extensible Capacities. *International Journal of Production Economics*, Vol. 87(2): 99-113, 2004.
- [12] Nam, S., Logendran, R. Aggregate Production Planning - A Survey of Models and Methodologies. *European Journal of Operational Research*, Vol. 61(3): 255-272, 1992.

- [13] Pochet, Y., Wolsey, L.A. Production Planning by Mixed Integer Programming. *Springer*, 2006.
- [14] Tadei, R., Trubian, M., Avendaño, J. L., Della Groce, F., Menga, G. Aggregate planning and scheduling in the food industry - A case study. *European Journal of Operational Research*, Vol. 87(3): 564-573, 1995.
- [15] Takei, F., Mesquita, M. Aggregate Planning for a Large Food Manufacturer with High Seasonal Demand. *Brazilian Journal of Operations & Production Management*, Vol. 3(1): 5-20, 2006.
- [16] Timpe, C.; Kallrath, J. Optimal Planning in Large Multi-Site Production Networks. *European Journal of Operational Research*, Vol. 126(2): 422-435, 2000.
- [17] Vercellis, C. Multi-Plant Production Planning in Capacitated Self-Configuring Two-Stage Serial Systems. *European Journal of Operational Research*, Vol. 119(2): 451-460, 1999.

---

# GENERACIÓN DE EQUIPOS DE TRABAJO MEDIANTE ANÁLISIS DE REDES SOCIALES E IDENTIFICACIÓN DE ATRIBUTOS PERSONALES

---

FRANCISCO MOLINA J.\*

PABLO LOYOLA H.\*

JUAN D. VELÁSQUEZ\*

## Resumen

*En la actualidad, gran parte de las organizaciones cuentan con sistemas de información que permiten supervisar procesos productivos, generar indicadores económicos y optimizar procesos administrativos, dejando en un segundo plano la gestión de las personas que componen la institución y los aprendizajes que han adquirido a lo largo de sus vidas. En particular, diversas instituciones de educación superior cuentan con plataformas virtuales que registran calificaciones, almacenan archivos y proveen de foros a los alumnos. Sin embargo, estos sistemas no permiten apreciar la capacidad real de los estudiantes, ya que su foco se centra principalmente en optimizar labores administrativas y no en recoger y procesar información clave de los alumnos, con la cual se podría generar el conocimiento necesario para gestionar el desarrollo personal y profesional del estudiantado. El presente artículo detalla la construcción de un sistema de información capaz de recolectar una serie de datos específicos de las personas (sus intereses, las redes sociales que poseen, sus experiencias laborales y la capacidad de adquirir roles dentro de un equipo) con la finalidad de procesarlos y generar estadísticas, análisis de redes sociales y equipos de trabajo cuyos integrantes sean compatibles entre sí.*

**Palabras Clave:** *Equipo de Trabajo, Análisis de Redes Sociales, Generación de Conocimiento.*

---

\*Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

---

## 1. Introducción

---

Varios autores han demostrado que la capacidad real y el éxito de un individuo no pueden ser pronosticados sólo por su coeficiente intelectual (CI) ni por su rendimiento académico [5, 23]. Sin embargo, la mayoría de los establecimientos de educación superior continúan basando su sistema educativo en la aprobación de exámenes de conocimientos, obviando el ejemplo de universidades reconocidas a nivel mundial (como el Massachusetts Institute of Technology o MIT), donde la formación de profesionales exitosos se basa en potenciar de forma armónica la capacidad cognitiva y las habilidades individuales e interpersonales de los alumnos<sup>1</sup>.

En particular, Amartya Sen[22] sostiene que al observar la capacidad de una persona, lo que interesa es evaluarla en términos de su *habilidad real* para lograr funcionamientos valiosos como parte de la vida. Por lo tanto, si uno de los fines de la educación es permitir que los individuos logren integrarse plenamente en la sociedad, debiese incorporarse una mirada más sistémica a la problemática de la enseñanza, avanzando hacia modelos educacionales que busquen el pleno desarrollo de la personalidad humana y no sólo una parte de ella.

Por esto, es de suma importancia profundizar en asuntos como el observado por la Comisión Nacional de Acreditación (CNA-Chile)<sup>2</sup> al momento de certificar la carrera de Ingeniería Civil de la Universidad de Chile a mediados de 2007:

*Se destaca la calidad del profesional formado en la carrera, en especial su capacidad analítica, su formación en ciencias básicas y su capacidad técnica. Sin embargo, se plantea la necesidad de reforzar herramientas de gestión, de relaciones interpersonales, liderazgo, emprendimiento y comunicación.*<sup>3</sup>

A continuación se expone una metodología que trata de hacerse cargo, a

---

<sup>1</sup>Basado en la charla de Woody Flowers (Profesor y Ph.D. del MIT, distinguido como *MacVicar Faculty Fellow* por sus contribuciones sobresalientes a la enseñanza de pregrado) realizada el 23 de octubre de 2007 en la Escuela de Ingeniería de la Universidad de Chile.

<sup>2</sup>Para velar por la excelencia de la educación superior en Chile, en octubre de 2006 se promulgó la *Ley de Aseguramiento de la Calidad de la Educación Superior N°20,129*, la cual crea la Comisión Nacional de Acreditación *CNA – Chile* cuya función es “verificar y promover la calidad de las Universidades, Institutos Profesionales y Centros de Formación Técnica”. Fuente: Sitio web de la Comisión Nacional de Acreditación. [Consulta: 21 de junio de 2009]. Disponible en: [http://www.cnachile.cl/portada/port\\_pres\\_general.html](http://www.cnachile.cl/portada/port_pres_general.html)

<sup>3</sup>Fuente: Sitio web de la Comisión Nacional de Acreditación. Documento: *Acuerdo de Acreditación N° 443. Carrera de Ingeniería Civil, Universidad de Chile*. Junio de 2007. [Consulta: 21 de junio de 2009]. Disponible en: [http://www.cnachile.cl/acreditacion/ac\\_carreras/ACUERDO\\_N443.pdf](http://www.cnachile.cl/acreditacion/ac_carreras/ACUERDO_N443.pdf)

través del uso de TIC's, del problema observado en la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile, el cual se puede extrapolar a diversas organizaciones de otras índoles.

El presente artículo comienza con la sección 2, la cual contiene un breve marco teórico sobre los principios básicos del trabajo en equipo y las herramientas principales del análisis de redes sociales. A continuación, en la sección 3, se estudian las variables que influyen en el desempeño de los equipos de trabajo dentro de la FCFM, se establecen los requerimientos de información del sistema y se comenta brevemente sobre la construcción del sitio web que alberga el sistema. Luego, en la sección 4, se detalla la experiencia piloto realizada en el Departamento de Ingeniería Industrial (DII) de la FCFM y se discuten los resultados obtenidos. Finalmente, en la sección 5 se presentan las principales conclusiones que se obtuvieron a lo largo del trabajo.

---

## 2. Conceptos Preliminares

---

### 2.1. El Equipo de Trabajo: Principios Básicos

Para entender el funcionamiento de un equipo, es necesario conocer algunas nociones básicas como los pasos para diseñar un grupo de trabajo efectivo, los roles posibles dentro de un equipo, los factores que influyen en el desempeño de un grupo de trabajo y los beneficios y desventajas de formar equipos de trabajo.

#### 2.1.1. Pasos para Diseñar Equipos de Trabajo

En uno de sus trabajos más reconocidos, Hackman (1987) definió cuatro etapas a seguir para la creación de un grupo de trabajo efectivo: trabajo previo, creación de condiciones de desempeño, formación y construcción del equipo, y prestar asistencia al equipo durante la marcha. (Ver Fig. 1)

Con respecto al número ideal de integrantes, varios autores aconsejan un rango que varía entre cuatro y ocho personas[13, 9, 25], ya que un número mayor presenta más conflictos internos[11], se enfrenta a problemas logísticos como encontrar suficiente espacio físico y tiempo para reunirse[13] y la tendencia a colaborar de sus miembros disminuye naturalmente a medida que el equipo crece[6]).

Por otro lado, se ha demostrado que considerar las habilidades sociales de los miembros[18], la capacidad de liderazgo y el nivel de responsabilidad y la capacidad de compromiso[13] influyen altamente en el desempeño del equipo.

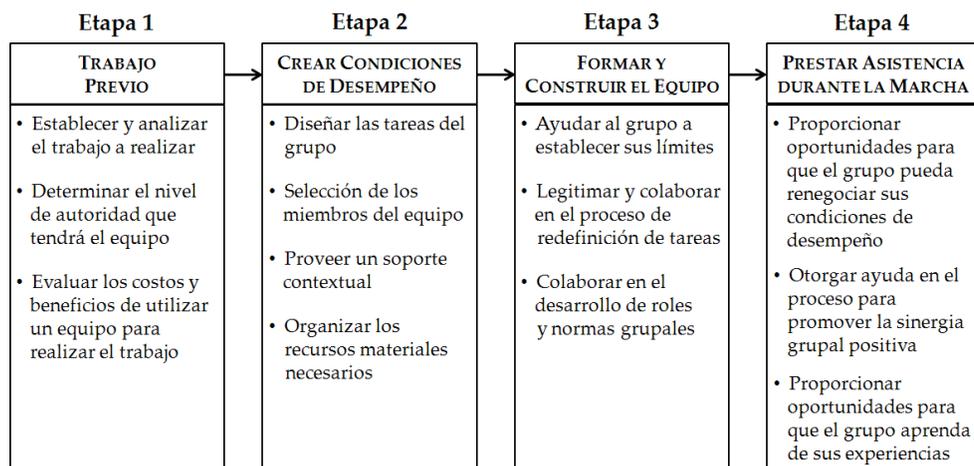


Figura 1: Etapas a seguir para la creación de un grupo efectivo

### 2.1.2. Roles dentro del Equipo

Por otro lado, en [14] se investigaron diversos equipos de producción, servicios y procesos, y observaron que los equipos eficientes concentran una serie de roles que desempeñan sus miembros en base a sus habilidades, capacidades e intereses, donde la ausencia de alguna de estas actividades impacta significativamente el desempeño grupal. A continuación, se presenta un resumen de los ocho roles propuestos como esenciales por lo autores mencionados:

- **Asesorar:** Recolectar y resguardar que la información este disponible para que el equipo pueda tomar decisiones correctas en función del cumplimiento de las metas acordadas.
- **Innovar:** Generar y experimentar con nuevas ideas, preguntándose a menudo si las tareas del equipo se desarrollan de la mejor forma posible.
- **Promover:** Explorar y presentar oportunidades a ejecutivos (y/o terceros) con el fin de captar recursos para que el equipo pueda realizar sus labores.
- **Desarrollar:** Escuchar a los clientes (internos y externos) para detectar sus necesidades, y luego verificar la aplicabilidad de incorporar nuevos enfoques en los esfuerzos de planificación.
- **Organizar:** Involucra ordenar al equipo para que sus miembros sepan qué hacer, cómo y cuándo, fijando objetivos y plazos claros para hacer funcionar las cosas.

- **Producir:** Considerando las metas fijadas, se relaciona con la entrega de productos o resultados, cumpliendo con altas normas de efectividad y eficiencia.
- **Inspeccionar:** Controlar y efectuar auditorías de calidad, de finanzas, de seguridad y de personal, entre otras para asegurar el cumplimiento de las normas y metas establecidas.
- **Mantener:** Revisar periódicamente la efectividad del equipo, preocupándose de preservar las normas de calidad y de proteger la continuidad de los procesos operacionales.

**2.1.3. Factores que Influyen en el Desempeño del Equipo**

En la Fig. 2 se exhiben los tres tipos de factores que repercuten en el rendimiento grupal [16], presentándose para cada categoría una lista de factores que se repiten constantemente en la literatura que aborda esta temática.<sup>4</sup>

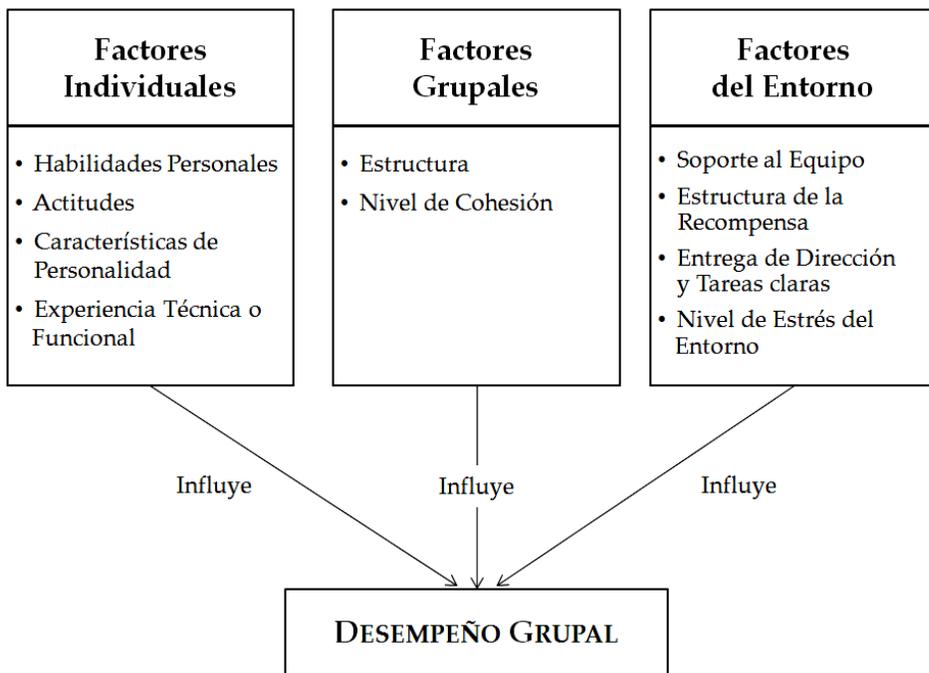


Figura 2: Factores que afectan el desempeño grupal

<sup>4</sup>Dentro de los trabajos consultados para estudiar los factores que repercuten en el desempeño grupal se encuentran: [8, 12, 3, 7, 11, 25, 2, 10, 18, 6]

#### 2.1.4. Beneficios y Desventajas del Trabajo en Equipo

Algunas de las razones que han posicionado al trabajo en equipo como una manera eficiente de realizar labores complejas se pueden resumir de la siguiente manera:

- Los equipos consiguen desarrollar habilidades y capacidades complementarias que, por definición, son superiores a las que podrían aportar los miembros de forma aislada.
- Gracias a su capacidad de diálogo e interpretación, los equipos pueden ajustar sus enfoques a la nueva información y a los nuevos retos con mayor velocidad, exactitud y eficacia que un individuo aislado con gran cantidad de conexiones organizacionales.
- Los equipos bien administrados pueden aprovechar mejor los recursos y obtener mejores resultados de desempeño.
- La pertenencia a un equipo de trabajo proporciona apoyo mental y fomenta el compañerismo, siendo estos factores de motivación.
- La experiencia de trabajar en equipo contribuye positivamente al aprendizaje y al bienestar de cada uno de sus miembros.

Por otro lado, algunas desventajas de crear y utilizar equipos se pueden resumir brevemente de la siguiente manera:

- Como se confunden los términos “grupo de trabajo” y “equipo” con facilidad, se tiende a administrar ambos sistemas de la misma forma, obteniéndose equipos que producen menos que las contribuciones individuales que podrían hacer sus miembros.
- La formación de un equipo toma más tiempo y gasta más recursos que la generación de un grupo de trabajo.

### 2.2. Breve Introducción al Análisis de Redes Sociales

A continuación se presentan los conceptos y herramientas básicas del Análisis de Redes Sociales (ARS): definiciones principales, representación de una red social, y algunas propiedades de las redes y los actores.

#### 2.2.1. Conceptos Básicos del Análisis de Redes

A continuación, se presentan algunos términos necesarios de manejar para entender y utilizar el análisis de redes.

- **Actor:** Entidad social que interactúa con otras entidades. Esta puede ser una persona discreta, un grupo de individuos, una organización, una ciudad, una nación, etc.
- **Conexión o Vínculo:** Corresponde al tipo de relación que “conectan” par de actores distintos. Estos vínculos pueden ser de naturaleza muy diferente: relaciones transaccionales, de comunicación, instrumentales, sentimentales, de poder o autoridad, de parentesco, etc.
- **Grupo:** Conjunto finito de actores sobre los cuales son efectuadas las mediciones.
- **Red Social:** Conjunto finito de actores que contienen un determinado número de relaciones definidas entre ellos.

### 2.2.2. Representación de una Red Social

Una de las nociones matemáticas que mejor permite representar una red de manera sistemática y concisa es el **grafo** (ver Fig. 3 ), ya que es un modelo simplificado de una red social donde los actores son representados por nodos (*vértices o puntos*) y los vínculos son trazados mediante líneas (*bordes o aristas*).

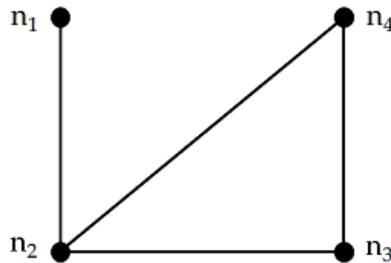


Figura 3: Ejemplo de un grafo

En un grafo, la ubicación de puntos en el espacio y el largo de las líneas no tienen relevancia. La única información importante es el conjunto de nodos y la presencia o ausencia de líneas entre diferentes pares de puntos [24].

### 2.2.3. Propiedades Básicas de Redes y Actores

Si  $g$  es el número de nodos y  $L$  el número de líneas existentes en un grafo, entonces se tienen las siguientes propiedades:

- **Grado de un nodo:** Corresponde al número de líneas que inciden sobre un determinado nodo  $i$ , denotando este valor por  $d(n_i)$ .<sup>5</sup>

<sup>5</sup>El uso de la letra “d” para denotar el grado de un nodo proviene del término en inglés *degree*.

- **Densidad:** Si un grafo tiene  $g$  nodos, entonces el máximo de líneas que puede tener es  $g(g-1)/2$ . Entonces, la densidad viene siendo la relación entre las líneas existentes y el total de líneas posibles que puede tener la red. En términos matemáticos, la densidad se puede expresar como 
$$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)}.$$
- **Diámetro:** El *diámetro de una red* es la distancia -geodésica- más larga existente dentro de la red.
- **Agujero Estructural:** Dado que difícilmente se tienen redes con densidad máxima, se llama “agujero estructural” a la falta de una conexión entre dos nodos cualquiera.

#### 2.2.4. Software de Análisis y Representación de Redes Sociales: UCINET

Durante años, varios autores [4, 17, 20, 21] coinciden en destacar a UCINET, de *Analytic Technologies*<sup>6</sup>, como uno de los programas más estandarizados para el análisis de redes.

Algunas de sus principales virtudes son la facilidad para exportar datos provenientes desde hojas de cálculos, como Excel y otros programas ofimáticos, las posibilidades de gestión de matrices y grafos, y la disponibilidad de variadas herramientas para trabajar con técnicas como *Cliques*, *N-Clan*, *K-Plex*, MDS o CONCOR, entre otras.

---

### 3. Levantamiento de Requerimientos

---

En este capítulo se propone la construcción de un sistema de información que permite recolectar y procesar datos de alumnos, para luego formar equipos de trabajo que maximicen la compatibilidad entre sus miembros.

Para lograr esto, primero se estudia el funcionamiento de los grupos de alumnos que trabajan (o que han trabajado) juntos en la FCFM y se determinan las variables principales que afectan (o que afectaron) su funcionamiento. A continuación, se establecen los requerimientos generales del sistema de información a construir. Finalmente, se presenta el modelo matemático diseñado para evaluar y clasificar grupos.

#### 3.1. Estudio: Caracterización de los Equipos de Trabajo en la FCFM

El presente apartado tiene como propósito describir el conjunto de métodos y técnicas utilizadas para caracterizar los equipos de trabajo en la FCFM,

<sup>6</sup><http://www.analytictech.com/>

especificando el enfoque metodológico de la investigación, las técnicas utilizadas para la recolección de los datos y los procedimientos realizados para el análisis e interpretación de los resultados.

**Enfoque metodológico del estudio.** El estudio integró en su diseño un enfoque metodológico mixto para lograr un alto grado de integración entre los enfoques cuantitativos y cualitativos [19]. De este modo, la investigación se dividió en dos partes esenciales: una enfocada desde lo cualitativo, con la aplicación de entrevistas abiertas y semiestructuradas, y otra orientada a lo cuantitativo, a través de la aplicación de encuestas basadas en un cuestionario estructurado.

Más en detalle, a través de informantes clave se buscaron “grupos destacados” y “grupos de bajo desempeño” en la FCFM. Luego se entrevistaron estos grupos para saber más de su funcionamiento. A continuación, se consolidó la información recopilada y se realizó una encuesta a los grupos para levantar datos cuantitativos precisos. Por último, se compararon las variables que influyeron al conjunto de grupos exitosos con las variables que afectaron a los grupos no exitosos y se concluyó.

En resumen, mediante encuestas y entrevistas semiestructuradas se buscó información sobre los grupos de trabajo en dos ejes principales: características personales de los integrantes y aspectos de la conformación del grupo y sus dinámicas recurrentes.

**Entrevistas abiertas a informantes clave.** Se realizaron entrevistas abiertas a los siguientes informantes clave: Profesores de 8 carreras diferentes (cuyas asignaturas exigen a los alumnos aprobar un trabajo grupal), funcionarios de distintas áreas de la FCFM (como la Subdirección de la Escuela, el Área de Deportes y el Área de Infotecnologías, entre otras) y Estudiantes que tienen (o que han tenido) cargos relevantes dentro de la FCFM (miembros de Centros de Estudiantes y Centros Deportivos de años anteriores.)

A cada entrevistado se le pidió definir, desde su perspectiva, lo que debía tener un grupo de alumnos para llegar a ser exitoso. Además, a cada uno se le pidió mencionar dos grupos pertenecientes a la FCFM: uno que haya obtenido resultados sobresalientes y otro con resultados mediocres o insuficientes. Además, se les solicitó una breve comparación entre los grupos y los motivos por los cuales, desde su perspectiva, uno obtuvo buenos resultados y el otro no.

**Selección de equipos a estudiar.** Para obtener una cantidad manejable de grupos a investigar, de la lista sugerida por los informantes clave a través de las entrevistas abiertas, se eliminaron los que no presentaban claros resultados sobresalientes (o deficientes), los que trabajaron menos de un semestre juntos y los que no presentaban una rápida accesibilidad a sus miembros.

Considerando lo anterior, del conjunto sugerido de 27 grupos se seleccionaron 17 para ser estudiados con mayor profundidad: 10 que presentaron desem-

peños sobresalientes y 7 que destacaron por su mal funcionamiento o sus pobres resultados.

**Entrevistas a grupos de trabajo.** Se realizaron 33 entrevistas semiestructuradas a estudiantes de los grupos seleccionados. A todos se les pidió comentar lo más posible sobre los siguientes ejes temáticos:

- Aspectos de la conformación del grupo y sus dinámicas recurrentes.
- Los factores personales con los que definiría a un equipo exitoso
- La magnitud del éxito alcanzado con el grupo y los factores que influyeron.
- La existencia de liderazgo dentro del grupo y la forma de administrarlo.
- La manera de resolver conflictos.

A continuación, se llevó a cabo el proceso de análisis de datos, a través del cual se desarrollaron las tareas de elaboración de unidades de análisis, categorización, codificación, síntesis y agrupamiento de la información, dependiendo de los elementos más relevantes de cada entrevista.<sup>7</sup>

**Confeción y aplicación de la encuesta.** Tras el análisis de las entrevistas se observaron diferencias significativas entre los equipos exitosos y los no exitosos. Luego, para comprobar cuantitativamente estos contrastes (y para recoger las opiniones de todos sobre cada punto), a los entrevistados se realizó una encuesta con escalas de Likert<sup>8</sup>. (Algunos ejemplos de preguntas son: “En mi grupo hay integrantes que destacan por su alta productividad” o “En mi grupo se asignan las tareas según intereses y habilidades de cada integrante”.)

**Análisis de los resultados obtenidos.** Se realizó un análisis crítico del proceso vivido por los grupos de trabajo entrevistados. Este análisis se centró en identificar aquellos factores que influyen en el desempeño grupal, separándose estas variables en tres líneas principales:

- Efectos de las características individuales de los integrantes.
- Las repercusiones en el rendimiento provocadas por variables grupales o estructurales.
- La influencia de factores externos o exógenos al grupo.

---

<sup>7</sup>Para tales efectos se utilizó el método de análisis de contenido propuesto por Sampieri et al. (2003), el cual es definido como una técnica apta para estudiar la comunicación de una manera objetiva, sistemática y que cuantifica los contenidos en categorías, aportando de este modo a realizar inferencias válidas y confiables de datos respecto a su contexto.

<sup>8</sup>Para más detalles de la metodología creada por Rensis Likert, se recomienda consultar algunos textos prácticos como [1] o [15].

A continuación, para cada uno de los ítems mencionados, se formulan algunos comentarios en base a la información cualitativa y cuantitativa revelada en el proceso:

- Factores Individuales En primer lugar, se destaca el rol de la experiencia al momento de liderar un grupo, ya que más del 50% de los equipos exitosos contaron con líderes que tenían al menos una **experiencia pasada en manejo de grupos** mientras que un tercio de los grupos no exitosos acusó a sus líderes de incompetentes en el cargo, o simplemente no reconoció liderazgo alguno a lo largo del proyecto.

Además, el 50% de los equipos destacados contaron con **miembros altamente productivos** (o con **integrantes brillantes**) que permitían *sacar* trabajos difíciles o tediosos en poco tiempo, descongestionando las labores del grupo. Por el contrario, los grupos no exitosos reconocieron, en un 67% de las veces, haber tenido problemas con algunos conocimientos técnicos, lo cual los retrasaba.

Cabe mencionar que lo anterior podría ser una de las razones que explica la gran diferencia entre los niveles de motivación existentes entre ambos tipos grupos: cerca del 70% de los miembros de grupos exitosos reconoció como un factor *energizante* la motivación existente intragrupo, mientras que el 52% de los grupos fallidos reconoció la existencia de algún grado de desmotivación durante el proyecto debido a causas como el descontento con la productividad grupal o con la calidad del trabajo realizado por algunos compañeros.

- Factores Grupales. El 82% de los grupos exitosos manifestó que los objetivos del proyecto estaban claros desde un comienzo y /o que sus intereses estaban alineados entre sí. En contraste, más de un tercio de los grupos no exitosos reconoció tener miembros con diferentes objetivos e intereses, los cuales constantemente discutían o se increpaban entre sí.
- Factores del Entorno. Se encontraron dos variables que diferenciaron los tipos de grupos estudiados: recurrir a gente externa para solucionar problemas puntuales y la importancia de contar con un espacio adecuado para trabajar. Con respecto al primer factor, se detectó que un número significativo de grupos exitosos recurrieron a personas externas para buscar consejo en algún momento del proyecto desarrollado, mientras que los *no exitosos* casi no contaron con ayuda externa, reconociendo en las entrevistas que habría sido de gran ayuda en momentos complicados. Por otro lado, a pesar de que la segunda variable no aparece tan clara en la encuesta, se decidió incluir debido a la información recogida en varias entrevistas a miembros de grupos no exitosos, los cuales manifes-

taron que no tenían un lugar fijo para trabajar y que eso disminuía su capacidad productiva.

### 3.2. Requerimientos Generales del Sistema de Información

Considerando la recopilación bibliográfica sobre el trabajo en equipo y la investigación realizada en la FCFM, se decidió que el sistema de información debía contemplar, como mínimo, los siguientes requerimientos de información y perfiles de usuario:

- **Requerimientos de Información.** Se deben realizar consultas sobre cuatro tipos de datos que influyen significativamente en la conformación de los grupos de trabajo: los intereses personales de los alumnos (en cuanto a trabajo y tiempos de ocio), su nivel de experiencia laboral, sus redes sociales dentro de la FCFM y sus preferencias para asumir un rol dentro de un grupo de trabajo.
- **Perfiles de Usuario.** Se distinguen dos perfiles distintos que interactuarán con el sistema: **alumno** (quien puede ingresar, modificar y eliminar los datos antes mencionados) y **analista de datos** (quien realiza los requerimientos de información al sistema para obtener, dado un perfil del trabajo a realizar<sup>9</sup>, un set de grupos potenciales, los cuales son dispuestos en un ranking construido por un modelo matemático<sup>10</sup> que recoge y pondera los datos ingresados por los alumnos.)

### 3.3. Clasificación de los Grupos de Trabajo

Para evaluar el posible funcionamiento de un equipo, se diseñó un modelo matemático con 4 factores que se pueden manejar en el proceso de formación de un grupo y que influyen significativamente en su desempeño. Además, cada factor se puede subdividir en 2 variables que aportan información al modelo, dándole mayor robustez:

- **Intereses de los alumnos:** “sector de interés para trabajar” y “actividades que se realizan en el tiempo libre”.
- **Experiencias que aporten habilidades técnicas:** “experiencias laborales” y “Experiencias en voluntariados”.
- **Las redes sociales:** “vínculos existentes entre los miembros antes de formarse el grupo” y “Vínculos que poseen los miembros con alumnos externos al grupo”.

---

<sup>9</sup>Por ejemplo, podría estar orientado hacia la creación de negocios, hacia la acción social, hacia temas medioambientales, etc.

<sup>10</sup>En el Anexo se presenta el modelo matemático diseñado, junto con una breve explicación de sus parámetros.

- **Preferencias para trabajar en grupo:** “Cantidad de roles presentes en un grupo de n alumnos” y “efectividad de los miembros en sus respectivos roles”.

Por lo tanto, considerando estas 8 variables se construyó el modelo que permite asignarle un puntaje a cada grupo en virtud de las características individuales y grupales de sus miembros:

$$\alpha \left[ \alpha_1 \begin{pmatrix} \text{alineamiento} \\ \text{sectores de} \\ \text{interés} \end{pmatrix} + \alpha_2 \begin{pmatrix} \text{alineamiento} \\ \text{de actividades} \\ \text{personales} \end{pmatrix} \right] + \beta \left[ \beta_1 \begin{pmatrix} \text{experiencia} \\ \text{laboral} \end{pmatrix} + \beta_2 \begin{pmatrix} \text{experiencia} \\ \text{en} \\ \text{voluntariados} \end{pmatrix} \right] \\ + \gamma \left[ \gamma_1 \begin{pmatrix} \text{redes internas} \\ \text{antes de} \\ \text{formar el grupo} \end{pmatrix} + \gamma_2 \begin{pmatrix} \text{redes} \\ \text{externas de} \\ \text{los miembros} \end{pmatrix} \right] + \delta \left[ \delta_1 \begin{pmatrix} \text{cantidad de} \\ \text{roles en} \\ \text{el grupo} \end{pmatrix} + \delta_2 \begin{pmatrix} \text{efectividad} \\ \text{de los roles} \\ \text{en el grupo} \end{pmatrix} \right]$$

Los parámetros  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$  son constantes que, para el experimento realizado, se definieron en base al peso relativo que tuvo cada una de las 4 áreas (intereses, experiencia, redes y roles) en el éxito de los equipos estudiados dentro de la FCFM.

---

## 4. Construcción del Sistema de Información y Experimentación

---

Para construir el sistema de información propuesto, se diseñó un sitio web con todas las funcionalidades descritas anteriormente (recolección de datos provenientes de alumnos, procesamiento de los datos y presentación de la información en un tiempo razonable). Posteriormente, se testeó su funcionalidad y su usabilidad con un grupo de alumnos del Departamento de Ingeniería Industrial de la Universidad de Chile, obteniendo resultados que permitieron analizar la viabilidad de utilizar la metodología propuesta.

### 4.1. Arquitectura

Para el desarrollo se utilizaron herramientas de código abierto, las cuales permitieron la reutilización de módulos pertenecientes a otras aplicaciones sin costo alguno.

En resumen, se utilizó el motor de bases de datos relacional MySQL (versión 4.1.22) y el lenguaje PHP (versión 5.1.4) para el software. Además, para la interacción del sistema con el usuario, ya sea al ingresar datos o al hacer requerimientos de información, se decidió utilizar navegadores web de amplia aceptación (como Firefox, Internet Explorer, Google Chrome y Opera), dejando el procesamiento de los datos completamente en el servidor web.

## 4.2. Módulos de la Interfaz

Para que el sistema fuese adoptado de la mejor forma por los usuarios, se le bautizó con el nombre de *U-Proyectos*, lo que sigue una lógica institucional derivada de otros sistemas de información utilizados por los alumnos de la FCFM: U-Cursos y U-Campus.

Por otro lado, se diseñaron módulos orientados a dos tipos de perfil: **alumno** (los cuales permiten el ingreso de intereses, experiencias laborales, las redes sociales y las preferencias para trabajar en grupo) y **analista de datos** (donde se puede filtrar a los alumnos según sus intereses y experiencias, pudiendo agruparlos en equipos de trabajo cuyos miembros se complementen entre sí).

## 4.3. Caso de Prueba y Resultados Obtenidos

Para testear el funcionamiento del sistema, se envió, mediante correo electrónico, una invitación personalizada con la explicación de la iniciativa, un nombre de usuario y un password, a 183 estudiantes del DII (equivalente al 5% del alumnado total de la FCFM), permitiéndoles entrar libremente durante una semana al sitio web.

Luego de finalizado el plazo, se obtuvo una tasa de ingreso cercana al 30% (53 alumnos), con más de 2400 datos ingresados. (Ver Tabla 1)

Indicadores	Resultados
Nº total de alumnos que ingresaron al sistema	53
Tiempo promedio de permanencia en el sitio web** (mm:ss)	13:02
Módulo con menor tiempo promedio de ingreso de datos (mm:ss)	Perfil de Trabajo (01:18)
Módulo con mayor tiempo promedio de ingreso de datos (mm:ss)	Redes Sociales (05:12)
Cantidad total de datos ingresados	2411
Promedio de datos ingresados por alumno	45,5
Módulo con menor cantidad de alumnos que ingresaron datos	Redes Sociales (37 alumnos)
Módulo con mayor cantidad de alumnos que ingresaron datos	Intereses Personales (53 alumnos)
Consultas o comentarios sobre el funcionamiento del sitio web	3

Tabla 1: Resultados de la recolección de datos

Una vez concluida la fase de recolección de datos, se procedió al análisis y

extracción de conocimiento del sistema. Para esto, se definieron tres áreas principales a considerar: estadísticas generales sobre los alumnos, descubrimiento de redes sociales y generación de grupos de trabajo.

### Estadísticas Generales

La información recopilada por el sistema permitió conocer rápidamente tendencias y preferencias de los alumnos que participaron en la prueba, las cuales pueden servir de apoyo a la toma de decisiones por parte de profesores, centros de alumnos y/o personal administrativo de la FCFM. A modo de ejemplo, se presentan algunos de las estadísticas obtenidas:

- Sectores de trabajo que concentran el interés de los alumnos.
- Actividades que realizan los alumnos en su tiempo libre.
- Cantidad de empresas (por rubro) donde han trabajado los estudiantes.
- Áreas donde han trabajado los alumnos. (Por ejemplo, dentro del grupo de prueba, el área donde más han trabajado los alumnos es “Ventas”, con un 24 %, seguida de “Computación e Informática” y “Administración”, ambas con un 12 %. Y, por el contrario, no reciben menciones “Auditoría”, “Contabilidad”, “Control de Gestión”, “Facturación”, “Producción y Manufactura” y “Recursos Humanos”.)

### Descubrimiento de Redes Sociales

Con los datos recogidos en el módulo “Redes Sociales”, se armó un archivo de texto plano compatible con el programa *Ucinet 6*. Luego, con este software se analizó el grupo de prueba, identificándose diversas características del entretrejo social de los alumnos. (Ver Fig. 4)

Además, siendo  $g$  el número de alumnos y  $L$  la cantidad de vínculos entre ellos, se obtuvo algunos indicadores interesantes de analizar posteriormente:

- Densidad de la red<sup>11</sup>:  $\Delta = \frac{L}{g(g-1)} = \frac{43}{27(27-1)} = 6,13\%$ .
- Mayor grado de entrada:  $d(\text{Alumno}_{11}) = 5$ .
- Mayor grado de salida:  $d(\text{Alumno}_8) = 6$ .
- Accesibilidad: No todos los nodos son accesibles dentro de la red  $\text{Alumno}_{18}$  y  $\text{Alumno}_{19}$  no presentan lazos de amistad con resto de sus compañeros.

### Generación de Grupos de Trabajo

Para probar el sistema y encontrar oportunidades de mejoras, se diseñó un experimento (o proyecto piloto) que tuvo las siguientes etapas y resultados:

<sup>11</sup>Al ser un grafo dirigido se deben considerar los vínculos recíprocos. Por lo tanto, el número de arcos no se divide por la mitad y es equivalente a:  $g(g-1)$ .

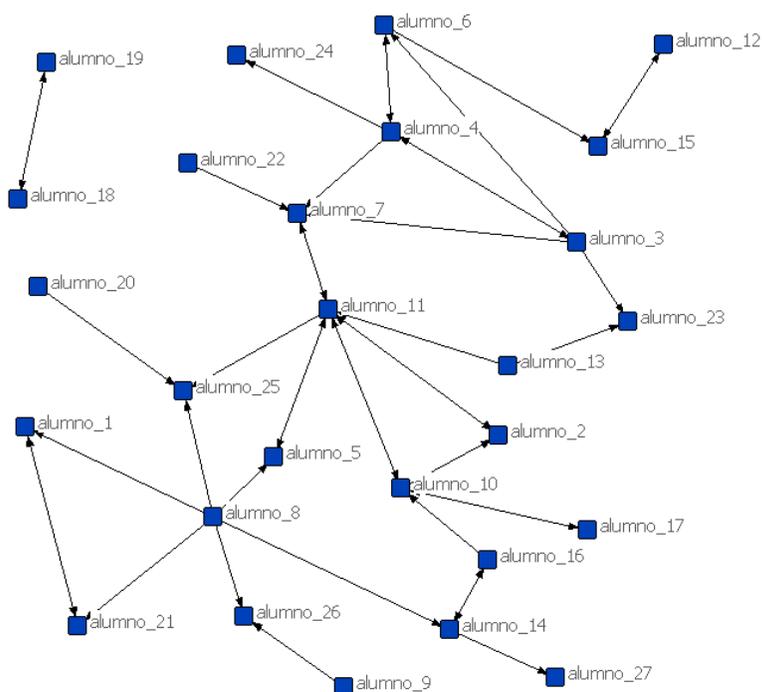


Figura 4: Grafo “Vínculos de amistad dentro del grupo de prueba”

- Proyecto: Bosquejar una campaña publicitaria para U-Proyectos en la FCFM, con el fin de lograr la mayor cantidad de alumnos inscritos en un determinado tiempo. (Fecha de ejecución: 31 de marzo a 14 de abril de 2010.)
- Capacidades e intereses que debían tener los miembros del grupo: Interés por la publicidad, por la creación de negocios y por trabajar haciendo consultorías.
- Buscar un grupo en el sistema que cumpla los requisitos: Se ingresaron los intereses mencionados y se buscaron grupos de 4 personas (el mínimo recomendado en la literatura que habla del tema). El sistema evaluó 3060 posibles grupos en un tiempo razonable y entregó un ranking con los 20 primeros. Se seleccionó el grupo ubicado en la primera posición.
- Convocar a las personas que propuso el sistema: Se realizó una reunión donde 3 de los 4 alumnos pudieron asistir y todos aceptaron realizar el proyecto. El cuarto miembro fue contactado por teléfono y también aceptó participar.
- Entrega de recursos físicos y monetarios: No se pudo proveer de un espacio físico de trabajo. Sin embargo, se entregó una pequeña cantidad de dinero para que el grupo dispusiera de él como estimase conveniente.

- Registro del proceso: Dado lo acotado del proyecto, el contacto fue preferentemente vía email y teléfono, dejándose registro de los comentarios y sugerencias del grupo de prueba.
- Resultados obtenidos: La principal herramienta de comunicación fue el correo electrónico y el celular, dejando en segundo plano las reuniones personales. Con respecto a los recursos monetarios, el grupo no gastó el dinero que le fue asignado y quiso devolverlo al terminar el proyecto. En cuanto a la convivencia, los miembros del grupo no tuvieron discusiones ni grandes desacuerdos. Al final de la experiencia, la mitad del grupo manifestó interés por seguir participando de U-Proyectos, particularmente explorando nuevas formas de hacer crecer la iniciativa.

#### 4.4. Discusión de los resultados obtenidos

El experimento realizado entregó interesantes luces sobre mejoras que podrían ir perfeccionando el proceso de formación de nuevos grupos de trabajo:

- Ya que los grupos no deberían tener menos de 4 integrantes (idealmente), parece necesario formar grupos de 6 ó más personas considerando que una o dos podrían retirarse del proyecto eventualmente.
- Si bien el sistema predice la “capacidad esperada” de un determinado grupo, no predice el comportamiento individual de los miembros (no se sabe a priori si participarán de toda la experiencia o la dejarán en algún punto intermedio). Por ende, parece interesante estudiar e introducir en el modelo posibles factores de “fuga” de los alumnos, como la carga horaria que estos tienen (dentro y fuera de la universidad), por ejemplo.
- Luego de terminada la experiencia, los alumnos que llegaron hasta el final reconocieron que se conocían con anterioridad, pero que no habían colocado ninguna vinculación entre ellos en el sistema ya que no eran “mejores amigos” ni habían trabajado juntos anteriormente. Por lo tanto, parece interesante agregar en las consultas del módulo Redes Sociales el “grado” de amistad que se tiene con otras personas. Por ejemplo: “somos conocidos”, “existe una leve amistad”, “tenemos una enemistad”, etc.
- Como el grupo de prueba no contó con un espacio fijo para trabajar (ya sea físico o virtual), no se pudo participar de las conversaciones que mantuvieron sus miembros y, por ende, no se pudo conocer cabalmente los problemas cotidianos que surgieron durante el proyecto. Por lo tanto, se visualiza como necesario (a futuro) proveer de un espacio físico que permita centralizar las acciones y acompañar a los grupos durante su desarrollo, además de disponer de herramientas vía web para trabajar colaborativamente (como foros, chats y espacio para guardar archivos).

- Al ver la motivación y el entusiasmo generados en los miembros del grupo (quienes manifestaron su interés de seguir trabajando en el proyecto), se pudo apreciar tangiblemente el cumplimiento de uno de los objetivos principales del proyecto: aumentar la tasa de alumnos que están participando de iniciativas grupales en la FCFM, generando sinergias donde antes no las había.

---

## 5. Conclusiones y Trabajo Futuro

---

La construcción de un sistema de información centrado en la recopilación y el procesamiento de los intereses que poseen las personas puede generar diversas formas de conocimiento aún no dimensionadas por completo.

Por ejemplo, un problema recurrente en el Depto. de Ingeniería Industrial de la Universidad de Chile es la petición de ramos electivos, ya que un número significativo de alumnos queda en cursos cuyo tema principal no responde a sus intereses profesionales o personales, provocando baja asistencia, desmotivación y bajo rendimiento durante el semestre. Por ende, usar un sistema de información como el propuesto serviría para caracterizar de forma eficiente las preferencias de los alumnos que tomarán cursos electivos cada semestre, disminuyendo la asimetría de información entre oferta (universidad) y demanda (alumnos).

Por otro lado, al recoger experiencias y trabajos realizados con anterioridad, un sistema como el descrito permitiría develar parte del conocimiento almacenado en los miembros de una organización o de una comunidad (ya sean empresas, instituciones benéficas o educacionales, organismos gubernamentales, etc.), el cual, en general, permanece invisible y no es aprovechado.

En resumen, un sistema de información orientado a la recolección de 4 datos específicos de las personas: intereses, experiencias, redes sociales y perfil para trabajar en grupo; es capaz de entregar información rica en contenido, permitiendo tomar decisiones de manera rápida y eficiente al momento de querer potenciar todo tipo de organizaciones, en particular las de corte educacional.

### 5.1. Trabajo Futuro

Con el fin de profundizar y mejorar el trabajo aquí presentado, es posible definir algunas líneas de investigación a realizar en el futuro:

- Adaptar e implementar el sistema en distintas organizaciones del país (ya sean privadas o gubernamentales), con el fin de ajustar y comprobar el real alcance del modelo.
- Buscar nuevas variables que enriquezcan el modelo.

- Si bien las percepciones son útiles en primera instancia, es necesario buscar formas más exactas de obtener información de las personas. (Por ejemplo, utilizando encuestas ya probadas para medir la personalidad.)

### Agradecimientos

Los autores agradecen el aporte del Instituto Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16)

### Referencias

- [1] L. AIKEN. *Test Psicológicos y Evaluación*. Prentice Hall, México, 2003.
- [2] W. BENNIS. *Secretos de los grandes equipos*. Granica, 2002.
- [3] M. CAMPION, G. MEDSKER, and A. HIGGS. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, 46(1):823–850, 1993.
- [4] E. CHURCHILL and C. HALVERSON. Guest editorsíntroduction: Social networks and social networking. *IEEE Internet Computing*. *IEEE Internet Computing*, 78(5):14–19, 2005.
- [5] D. GOLEMAN. *La Inteligencia Emocional*. Barcelona: Kairós, 1996.
- [6] L. GRATTON and T. J. ERICKSON. Eight ways to build collaborative teams. *Revista Harvard Business Review*, 1(1):101–109, 2007.
- [7] R. GUZZO and M. DICKSON. Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology*, 3(47):101–109, 1996.
- [8] J. R. HACKMAN. *The design of work teams*. LORSCH, J. W. (editor). *Handbook of Organizational Behavior*. Englewood Cliffs: Prentice-Hall, 1987.
- [9] J. R. HACKMAN. *Por qué hay equipos que no funcionan*. En HESSELBEIN, F. y COHEN, P. (compiladores). *De Líder a Líder*. Granica, 2000.
- [10] J. R. HACKMAN. *Leading teams: Setting the stage for great performances*. Boston: Harvard Business School Press, 2002.
- [11] J. R. HACKMAN, R. WAGEMAN, and T. RUDDY. *Team effectiveness in theory and practice*. En COOPER, C. y LOCKE, E. A. (editores). *Industrial and organizational psychology: Theory and practice*. Oxford, UK: Backwell, 2000.

- [12] D. KATZENBACH, J.; SMITH. The discipline of teams. *Revista Harvard Business Review*, 1(1):111–120, 1993.
- [13] J. KATZENBACH and D. SMITH. *The Wisdom of Teams*. Boston: Harvard Business School Press, 1993.
- [14] D. MARGERISON, CH. McCANN and R. DAVIES. Focus on team appraisal. *Team Performance Management an International Journal*, 1(4):13–18, 1995.
- [15] R. McDANIEL, C. y GATES. *Investigación de Mercados Contemporánea*. México: Thomson Editores, 1999.
- [16] J. E. McGRATH. *Social psychology: A brief introduction*. New York: Holt, Rinehart and Winston, 1964.
- [17] J. L. MOLINA. *El Análisis de Redes Sociales. Una introducción*. Barcelona: Bellaterra, 2001.
- [18] M. MORGESON, F.; REIDER and M. CAMPION. Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, 1(58):583–611, 2005.
- [19] C. SAMPIERI, R.; FERNÁNDEZ and P. BAPTISTA. *Metodología de la Investigación*. México: McGraw-Hill Interamericana S.A., 2003.
- [20] L. SANZ MENÉNDEZ. Análisis de redes sociales: o como representar las estructuras sociales subyacentes. *Apuntes de Ciencia Tecnología y Sociedad*, 1(7):21–29, 2003.
- [21] J. SCOTT. *Social Network Analysis: A Handbook. 2ª Edición*. SAGE Publications, 2000.
- [22] A. SEN. *Capacidad y bienestar*. En NUSSBAUM, M. y SEN, A. (compiladores). *La calidad de vida*. Fondo de Cultura Económica, México, 1996.
- [23] P. SENGE. *La Quinta Disciplina*. Barcelona: Granica, 1992.
- [24] S. WASSERMAN and K. FAUST. *Social Network Analysis: Methods and Applications*. New York and Cambridge, ENG: Cambridge University Press, 1994.
- [25] R. S. WINTER. *Manual de Trabajo en Equipo*. Madrid: Díaz de Santos, 2000.





# Programas de Postgrado y Postítulos DII

## DOCTORADO



La carencia de capital humano avanzado es una preocupación nacional. Por ello, la Universidad de Chile ha creado el doctorado en sistemas de ingeniería (DSI), el cual forma profesionales de excelencia en el área de sistemas de ingeniería. Su foco es el desarrollo de habilidades de resolución de problemas con técnicas avanzadas y metodologías multidisciplinarias.



Los egresados cuentan con herramientas y habilidades para abordar problemas complejos, que combinan gran tamaño, aleatoriedad y aspectos dinámicos y/o externalidades importantes, como consecuencia de la interacción sistémica entre agentes y procesos. Son capaces de formular soluciones con una perspectiva que integre el conocimiento técnico con metodologías en el estado del arte, apoyando al desarrollo productivo y social de las organizaciones o permitiendo la formación de nuevos enfoques y teorías en el ámbito científico.

El DSI es impartido por los Departamentos de Ingeniería Industrial, Ingeniería Matemática, Ingeniería Eléctrica y la división Transporte de Ingeniería Civil de la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile.

Más información en [www.sistemasdeingenieria.cl/doctorado/](http://www.sistemasdeingenieria.cl/doctorado/)

## MAGÍSTERES

### Magíster en Gestión de Operaciones MGO

El Magíster en Gestión de Operaciones (MGO) busca formar profesionales de excelencia en esquemas de gestión, uso de modelos y tecnologías de información, con capacidad de resolver problemas complejos en gestión de operaciones.

El Magíster en Gestión de Operaciones es impartido por el Centro de Gestión de Operaciones (CGO) del Departamento de Ingeniería Industrial de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, que recoge la vasta experiencia de sus integrantes



en las áreas forestal, minera, de servicios y manufactura, tanto a través de proyectos de investigación como de consultoría.

Los egresados del Magíster en Gestión de Operaciones se desempeñan en cargos de primer nivel en empresas de servicios y manufactura nacionales e internacionales. También trabajan en universidades como académicos e investigadores, y algunos siguen doctorados en prestigiosas universidades como MIT y Columbia.

#### Requisitos de Admisión

Poseer el grado de licenciado en Ciencias de la Ingeniería o su equivalente.

#### Calendario de Postulaciones

##### Semestre Otoño

Período de postulaciones: de octubre a 15 de diciembre.

Inicio de clases: marzo de cada año.

##### Semestre Primavera

Período de postulaciones: de abril a 15 de junio.

Inicio de clases: julio de cada año.

Mayor información: [julie@dii.uchile.cl](mailto:julie@dii.uchile.cl) | [www.dii.uchile.cl/mgo](http://www.dii.uchile.cl/mgo) | Tel: (56 2) 978 4017 | (56 2) 978 4073

## Magíster en Economía Aplicada MAGCEA

El Magíster en Economía Aplicada (MAGCEA) busca formar profesionales de gran competencia analítica y una sólida base en economía.

Es impartido por el Centro de Economía Aplicada (CEA) del Departamento de Ingeniería Industrial, de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. El CEA es líder en investigación en economía en Chile y en el desarrollo de propuestas para las políticas públicas.



Muchos graduados del programa han seguido estudios de doctorado en universidades como Harvard, Stanford, MIT, Princeton y Yale, entre otras.

Nuestros egresados son altamente requeridos en el mercado laboral, se desempeñan en empresas líderes a nivel nacional e internacional, en destacados organismos estatales y en importantes organismos internacionales.

### Requisitos de Admisión

Poseer un título profesional, nacional o extranjero, que exija al menos cinco años de estudios, o el grado de licenciado en campos disciplinarios afines a la especialidad.

### Calendario de Postulaciones

Semestre Otoño

Período de postulaciones: de octubre a 15 de diciembre.

Inicio de clases: marzo de cada año.

Semestre Primavera

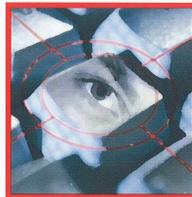
Período de postulaciones: de abril a 15 de junio.

Inicio de clases: julio de cada año.

Mayor información: [magcea@dii.uchile.cl](mailto:magcea@dii.uchile.cl) | [www.magcea-uchile.cl](http://www.magcea-uchile.cl) | Tel: (56 2) 978 4084 | (56 2) 978 4073

## Magíster en Ingeniería de Negocios con Tecnologías de Información MBE

El Magíster en Ingeniería de Negocios con Tecnologías de Información (MBE: Master in Business Engineering) tiene como objetivo formar a un tipo de profesional especialista, que integre gestión y tecnología en el diseño de los negocios y que tenga, además, las habilidades necesarias para iniciar y facilitar innovaciones en la empresa. Todas éstas, herramientas imprescindibles dentro del mundo emprendedor en el que se inserta.



Los egresados de este Magíster, por lo tanto, estarán en condiciones no sólo de generar, dirigir y ejecutar proyectos innovadores de transformación, sino además de rediseñar empresas tradicionales usando las TI y de desarrollar nuevas empresas basadas en tales tecnologías.

El programa incluye, como parte integral del mismo, el desarrollo de un proyecto de este tipo.

### Requisitos de Admisión

Poseer un título profesional con al menos cinco años de estudio, o licenciatura en una disciplina afín a la especialidad. También, buenos antecedentes académicos y/o laborales.

### Calendario de Postulaciones

Semestre Otoño

Hasta el 15 de enero de cada año.

Inicio de clases: marzo de cada año.

Semestre Primavera

Hasta el 15 de junio de cada año.

Inicio de clases: julio de cada año.

Mayor información: [mbe@dii.uchile.cl](mailto:mbe@dii.uchile.cl) | [www.mbe-uchile.cl](http://www.mbe-uchile.cl) | Tel: (56 2) 978 4835 | (56 2) 978 4935



## Magíster en Gestión para la Globalización

A partir del reconocimiento del factor humano de alta calificación como elemento diferenciador de los países exitosos, el programa se orienta a la formación de jóvenes profesionales chilenos con antecedentes de excelencia académica para desempeñarse eficazmente en la empresa global.

Se persigue con ello contribuir a abordar los desafíos en materia de capital humano y social que actualmente enfrenta Chile en esta etapa de su desarrollo.

El riguroso proceso de selección busca promover la excelencia, la meritocracia y la diversidad.



El Magíster en Gestión para la Globalización (MGPG) es impartido por la Universidad de Chile a través del Departamento de Ingeniería Industrial, tiene duración de 19 meses y dedicación full time.

Nueve de ellos se realizan en el extranjero, incluyendo estudios en prestigiosas universidades de EE.UU., Australia e Inglaterra, y un *study tour*.

Todos los estudiantes del MGPG han recibido la Beca Minera Escondida (otorgada por BHP Billiton), que cubre la totalidad los principales gastos académicos y de manutención relacionados.

**Mayor información: [www.magisterglobalizacion.cl](http://www.magisterglobalizacion.cl)**

## Magíster en Gestión y Políticas Públicas (MGPP)

El Magíster en Gestión y Políticas Públicas, tiene como propósito la formación avanzada de profesionales interesados en la formulación y ejecución de políticas públicas.

El MGPP forma líderes y servidores públicos del más alto nivel, capaces de conceptualizar, pensar y discutir sus visiones e ideas sobre el futuro de América Latina.

Se imparte en dos modalidades: Diurna y Ejecutiva.



### Características Distintivas

- \* Excelencia Académica
- \* Cuerpo docente de primer nivel
- \* Orientado a profesionales de formación diversa
- \* Alta tasa de graduación (83%)
- \* Reconocido entre los mejores en su área en América Latina.
- \* Acreditado por el CNA
- \* 15 años formando líderes

### Requisitos de Admisión

- \* Poseer el grado de licenciado o título universitario

### Versión en Horario Diurno:

Inicio: junio de cada año  
Duración: 19 meses

### Versión en Horario Ejecutivo:

Inicio: julio de cada año  
Duración: 24 meses

### Postulaciones:

Hasta el **15 de octubre** para personas que postulan a becas de instituciones

Hasta el **15 de abril** para personas que cuentan con fondos propios

Mayor información: [mgpp@dii.uchile.cl](mailto:mgpp@dii.uchile.cl) | [www.mgpp.cl](http://www.mgpp.cl) | Tel.: (562) 978 4067 | Fax 689 4987

Ven a estudiar con los mejores



# MBA

MAGÍSTER EN GESTIÓN Y DIRECCIÓN DE EMPRESAS

MBA FULL TIME | MBA PART TIME | MBA ICI-IC

Programa top ten en América Latina  
1er lugar carrera Ingeniería Industrial  
1ra en ranking calidad universitaria  
fuente AméricaEconomía & QuéPasa, 2009

Domeyko 2313, Santiago  
teléfono: 978 4002  
mba@dii.uchile.cl  
www.mbauchile.cl

## EDUCACIÓN EJECUTIVA

Educación ejecutiva 2010



Claves de diferenciación profesional



gestión · retail · finanzas  
marketing · operaciones  
tecnologías · innovación  
habilidades directivas  
estrategias · control de gestión  
inteligencia de negocios

1er lugar carrera Ingeniería Industrial  
ranking escuelas más valoradas 2009, revista QuéPasa

1ra en calidad  
ranking calidad universidades chilenas 2009,  
AméricaEconomía

Domeyko 2369, Santiago  
teléfono 978 40 02  
diplomas@dii.uchile.cl  
www.udiploma.cl

diplomas de postítulo  
cursos de especialización  
cursos in company  
seminarios y conferencias



El Programa de Desarrollo de Casos del **Instituto Sistemas Complejos de Ingeniería** busca generar material docente innovador y herramientas que ayuden a mejorar la enseñanza de la ingeniería en Chile.

Los **Casos de Estudio** son documentos pedagógicos que presentan problemas basados en desafíos reales que han sido abordados por académicos del Instituto.

La aplicación de un **Caso de Estudio** en clases se basa en la idea de discutir el problema, identificar sus distintas aristas y posibles estrategias de solución. A diferencia de los problemas clásicos, los Casos tienen varias posibles soluciones y motivan pensamiento creativo en el alumno.

En nuestro sitio web ([www.sistemasdeingenieria.cl](http://www.sistemasdeingenieria.cl)), sección “Casos”, usted podrá encontrar todos los Casos de Estudio generados hasta hoy. Si usted es docente y le gustaría tener la guía docente de algún caso, basta con que nos escriba a [casos@sistemasdeingenieria.cl](mailto:casos@sistemasdeingenieria.cl) y se la haremos llegar sin costo.

[www.sistemasdeingenieria.cl](http://www.sistemasdeingenieria.cl) sección “Casos”





