





- Desentrañando *Twitter* para replicar los indicadores de prevalencia y percepción de riesgo de la marihuana. 5  
*Víctor D. Cortés, Felipe E. Vildoso, Juan D. Velásquez, Carlos F. Ibáñez.*
- Programación matemática para los fixtures de los torneos profesionales del básquet de la Argentina en un formato NBA. 35  
*Guillermo Durán, Santiago Durán, Javier Marengo, Federico Mascialino, Pablo A. Rey.*
- Logística marítima y terrestre de una empresa salmonera en Chile mediante programación matemática. 63  
*Diego Delle Donne, Guillermo Durán, Guido Fuentes, Javier Marengo, Juan Ignacio Villasante, Andrés Weintraub.*
- Un nuevo enfoque para la programación horaria en universidades. 91  
*Isabel Méndez-Díaz, Juan José Miranda-Bront, Paula Zabala.*
- Process Mining para gestión de incidencias en Telefónica Chile. 115  
*Néstor Contreras, Cinthya Vergara, Yerko Montenegro.*

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

ISSN 0716 - 1174

---

EDITOR

**Guillermo Durán**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

EDITOR ASOCIADO

**Richard Weber**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

AYUDANTE DE EDICIÓN

**Cinthya Vergara**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

COMITÉ EDITORIAL

**René Caldentey**

*University of Chicago, Estados Unidos*

**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**

*Facebook Core Data Science, Estados Unidos*

Financiado parcialmente por el Instituto Sistemas Complejos de Ingeniería.

---

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.

---

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

---

Representante legal: Fernando Ordóñez  
Correo electrónico: ris@dii.uchile.cl  
Diagramación: Cinthya Vergara

Dirección: República 701, Santiago, Chile.  
Web URL: [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)  
Portada: Gabriella Fabbri

---

---

## Carta Editorial Volumen XXX

---

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), de investigadores del ISCI y de académicos de la Universidad de Buenos Aires.

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 estas áreas, orientado tanto a problemáticas nacionales como a la realidad de países de características similares de la región.

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*

Richard Weber  
*Editor Asociado*

---

# Llamado a Presentar Trabajos

---

La Revista Ingeniería de Sistemas (RIS) busca constituir un canal de divulgación de los avances en las áreas de Gestión de Operaciones, Tecnologías de Información e Investigación Operativa, que incluya los mundos académico y empresarial. Son particularmente apropiados artículos orientados a la práctica de estas disciplinas, que estimulen su uso o den cuenta de aplicaciones innovadoras de ellas, especialmente en América Latina.

También son bienvenidos artículos con análisis del estado del arte en un campo particular y de la forma en que los avances en dicho campo se han utilizado en la práctica.

Se espera que los artículos estén escritos de manera que puedan ser leídos por personas no especialistas en el tema tratado. Se recomienda incluir una lista de lecturas sugeridas para que los lectores no especialistas puedan profundizar en el tema.

---

## Formato del Manuscrito

Los autores deben enviar un archivo en formato PDF del manuscrito que desean someter a referato a:

*Comité Editorial Revista Ingeniería de Sistemas,  
Departamento de Ingeniería Industrial,  
Universidad de Chile.  
Santiago, Chile.  
Email: ris@dii.uchile.cl*

Los manuscritos deben estar formateados para hojas tamaño carta, a doble espacio, márgenes de 2,5 centímetros en todos los lados, deben incluir un resumen de no más de 150 palabras y su extensión no debe exceder las 20 hojas.

La primera hoja debe contener el título del trabajo, nombre y dirección de los autores (teléfono y correo electrónico del autor de contacto), y un resumen de no más de 150 palabras.

---

## Referencias

Las referencias se deben citar en el cuerpo del texto usando el nombre del autor y el año de publicación, e.g., Morton (1998). Al final del artículo se debe incluir la lista en orden alfabético de las referencias citadas en el texto. Para referencias de revistas científicas el formato es el siguiente: Autor(es), Año de publicación. Título. Nombre completo de la revista , Volumen e.g.:

Kodialam, M. y H. Luss, 1998. Algorithms for Separable Nonlinear Resource Allocation Problems. *Operations Research* , 44(2), 272-284.

Para referencias de libros el formato es el siguiente: autor(es), año de publicación. Título. Editorial, Ciudad; e.g.:

Kleinrock, L., 1975. *Queueing Systems* . John Wiley, New York.

En caso de haber más de una referencia con el mismo autor y año de publicación, se debe usar "a", "b", etc. como sufijo del año de publicación para diferenciarlas.

Detalles en [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)

---

---

# DESENTRAÑANDO *Twitter* PARA REPLICAR LOS INDICADORES DE PREVALENCIA Y PERCEPCIÓN DE RIESGO DE LA MARIHUANA

---

VÍCTOR D. CORTÉS \*  
FELIPE E. VILDOSO \*  
JUAN D. VELÁSQUEZ \*  
CARLOS F. IBÁÑEZ \*\*

## Resumen

El objetivo de este trabajo es diseñar y desarrollar un sistema que recoja información de *Twitter* con el fin de monitorear el consumo de marihuana y su percepción de riesgo. La clasificación de texto y sentimientos emitida por los usuarios, y las conexiones entre usuarios (entorno social) son utilizados para construir indicadores de consumo y percepciones con respecto a la marihuana. Se obtuvo un conjunto de 1,361,285 usuarios chilenos. Los clasificadores de texto y consumo de marihuana individual tuvieron medidas de precisión superiores a 0,7. Se concluye que es posible construir un sistema que utilice a *Twitter* como fuente de datos para reproducir tendencias con respecto a la marihuana a nivel individual y agregado. Esta información permitiría complementar los resultados de los estudios nacionales de drogas y contribuir a diseñar las políticas sobre este tema en el país.

**Palabras Clave:** Marihuana, Prevalencia, Percepción de riesgo, Aprendizaje de Máquinas, Redes Sociales.

---

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

\*\*Departamento de Psiquiatría y Salud Mental Norte, Facultad de Medicina, Universidad de Chile, Santiago, Chile.

---

## 1. Introducción

---

Desde inicios del siglo XXI se ha evidenciado un incremento abrupto en el uso y penetración de Internet. La introducción de la Web 2.0 hizo posible la generación de contenido por parte de usuarios que antes estaban restringidos a la lectura, facilitando el flujo de ideas y conocimiento por medio de contenido informal. Este evento fue acompañado de un crecimiento sostenido por varios años del número de usuarios. La irrupción de las redes sociales no vino más que a acentuar esta evolución, ya que hizo aún más evidente el cambio de paradigma.

Cada día, gran cantidad de información es generada mediante plataformas donde los usuarios asumen el poder de creación de contenido. El usuario da a conocer datos personales, intereses, actividades, relaciones e interacciones con otros usuarios. Las personas usan estos escenarios para expresarse y comportarse de manera natural, y destinan gran parte de su tiempo a sumergirse en estos ambientes. Por esta razón, estos escenarios son interesantes desde el punto de vista de recolección de información que puede representar a las personas y sus comportamientos [15].

Por otro lado, a lo largo de la historia, el consumo de drogas ha sido asociado a efectos negativos en la vida de las personas. La consecuencia más directa es que provoca dependencia en algún nivel, independientemente de la droga, y a su vez, esta dependencia produce otro tipo de problemas relacionados con la calidad de vida y pérdidas para la sociedad. Esto lo convierte en objetivo de estudio e intervenciones, que están acompañados de grandes esfuerzos y desembolso de dinero por parte del Estado [7] y organismos privados.

En Chile, la marihuana es un caso especial de estudio. Esta droga tiene atención particular, debido al amplio debate y relevancia que se le ha dado. En los últimos años, la percepción de riesgo de la droga ha disminuido, reflejando una norma social a favor del consumo y pudiendo ser origen del alza significativa del consumo nacional. En efecto, para el año 2012, la tasa anual de prevalencia fue de 7,1 %, una de las más altas a nivel Latinoamericano en ese periodo. Cifra que aumentó aún más en el año 2014 (11,3 %). Por lo tanto, las preocupaciones son justificadas.

Desde los encargados del diseño y monitoreo de políticas públicas con respecto a la marihuana, surge la necesidad de hacer seguimiento al consumo agregado de la droga y mejorar la comprensión acerca de los mecanismos que incentivan tal comportamiento. Por esta razón, el Servicio Nacional para la

Prevención y Rehabilitación del Consumo de Drogas y Alcohol (SENDA) realiza estudios para recolectar información sobre la magnitud del consumo y percepción de riesgo, entre otras variables.

El costo del Estudio Nacional de Drogas en sus versiones de población general y escolar hace que el periodo entre estudios (cada dos años) sea mayor al recomendado. La frecuencia de los estudios dificulta el seguimiento continuo de la evolución de la prevalencia e impide la detección de cambios abruptos de manera temprana. Además no captura variables específicas que podrían explicar de mejor manera los niveles observados de consumo. Por lo tanto, la oportunidad se centra en la creación de fuentes complementarias, que puedan mejorar y enriquecer la calidad de información. Además de aumentar la frecuencia de recolección de datos.

En este contexto, la aplicación de técnicas de minado de datos pueden traducirse en ventajas significativas al momento de procesar información estructurada y en forma de textos, para brindar datos complementarios a los mecanismos usados actualmente. La dificultad de esta iniciativa reside en estructurar gran cantidad de texto e información relativa a los usuarios y sus interacciones. Dichos datos tienen que ser transformados en información que apoye a la toma de decisiones [21].

En síntesis, este estudio busca replicar los indicadores de prevalencia y percepción de riesgo de la marihuana del Estudio Nacional de Drogas. La similitud de las redes sociales, específicamente *Twitter*, con la forma natural en que se relacionan las personas y la variedad de información vertida en ellas, permite plantear la siguiente hipótesis: “*Es posible extraer y procesar información de Twitter para representar un fenómeno complejo como el consumo y opinión sobre la marihuana de la población chilena*”. Para validarla, fue recolectado un conjunto de usuarios chilenos de *Twitter*, su información personal, sus relaciones de seguimiento y sus *tweets*. En base a estos datos fueron calculadas métricas de Análisis de Redes Sociales, fueron aplicadas técnicas de minería de datos para clasificar textos, se utilizaron algoritmos de *Opinion Mining* para determinar sentimientos en los textos, y posteriormente se obtuvieron patrones con relación al consumo de marihuana.

La Sección 2 hace revisión de algunos estudios de métricas utilizadas para explicar el consumo de marihuana y otros, explorando las redes sociales online en relación a ciertas drogas. La Sección 3 muestra detalladamente nuestra propuesta, estableciendo las métricas y los modelos que fueron utilizados para validar la hipótesis. En la Sección 4 se describe la metodología de recolección de datos. Luego, la Sección 5 contiene detalles correspondiente al tratamiento de datos y a los resultados obtenidos. Finalmente, la Sección 6 presenta las conclusiones finales y posibles líneas de trabajo futuro.

---

## 2. Trabajo relacionado

---

En esta sección serán mencionadas varias líneas de investigación con respecto al consumo y opinión sobre la marihuana. En primer lugar, serán descritos estudios que evaluaron factores protectores y de riesgo con respecto al consumo. Estos estudios generalmente exploran la relación directa o indirecta de un conjunto de variables explicativas sobre el comportamiento observado. En segundo lugar, serán detallados estudios relacionados con Análisis de Redes Sociales para explicar el consumo de marihuana, ya que son los más acordes con la propuesta de investigación. En tercer lugar, serán nombrados algunos estudios relacionados con la exploración de drogas en entornos web. En último lugar, se hará una revisión de estudios aplicando Opinion Mining.

Varios han sido los intentos por establecer una teoría que explique el consumo de sustancias. Todas ellas, al igual que las variables que utilizan, tienen elementos en común, algunos elementos diferenciadores, y diferentes niveles de respaldo empírico. [16] es un intento por entender las similitudes, diferencias, intersecciones y vacíos de las distintas teorías más prominentes. En él son reconocidos varios niveles de influencia en el consumo de marihuana. Estos niveles son tres: intrapersonal, cultural e interpersonal. Este último está enfocado en el efecto que tiene el contorno social sobre el consumo individual de sustancias.

La Teoría de Aprendizaje Social fue una de las primeras en considerar los efectos sociales en el desarrollo de comportamientos y quizás la más conocida. Esta teoría define un proceso de adopción de conductas y tal como muestra [14], las mayoría de las variables propuestas del modelo se sostienen en un modelo estable de predicción de consumo de marihuana. Estas variables presentaron efectos directos y mediaron otro tipo de variables estructurales como el género, clase o edad.

En [19] y [5] estudiaron factores de riesgo que incorporaban factores sociales y los consideraron como los predictores más fuertes en todas las etapas de consumo. En otro estudio donde se comparaban varias teorías y se indagaba el efecto conjunto de variables grupales y psicológicas concluyeron que la orientación hacia un grupo de referencia usuario de marihuana es el predictor más sustancial de uso de marihuana ([9]).

En [22] se señaló que jóvenes con más usuarios de sustancias en sus redes sociales reportaron mayor consumo. Más precisamente, en [1], controlando por características de los padres y otros parámetros, encontraron que un incre-

mento en el 10 % de amigos cercanos y compañeros de curso quienes usaban marihuana incrementaba la probabilidad de uso en un 5 %. [20] agregó que la influencia de los pares se sostiene en el periodo de crecimiento, pero la influencia de los padres disminuye con el paso del tiempo. Esto lo confirma [18], señalando que aquellos que se relacionaron con usuarios entre la adolescencia tardía y la adultez temprana eran 1.6 veces más propensos a iniciar el uso de marihuana. Además confirma la relación entre usuarios cercanos y uso propio para el inicio y continuación del fenómeno.

El Análisis de Redes Sociales consiste en construir un conjunto de métricas a partir de las conexiones entre las personas. Es posible construir redes complejas en base a cinco nominaciones de amigos por cada uno de los sujetos encuestados. A partir de esto, la posición dentro de la estructura de la red social también influye en la conducta. Efectivamente, [8] sostiene que adolescentes menos incrustados en la red, mayor estatus y mayor proximidad a pares usuarios de sustancias eran más propensos al mismo comportamiento. En [12] se halló que el consumo por parte del grupo, e interacciones entre la posición en la red y el uso de pares predicen el consumo. En particular, personas que conectan grupos son especialmente afectados por el consumo.

Las normas sociales son ligadas con la percepción del individuo de la aprobación de los pares sobre algún comportamiento. Aplicado a la marihuana, en [11] fueron evidenciadas variaciones de desaprobación de la marihuana en distintos cortes generacionales y que estas diferencias afectaron directamente al consumo. Los cortes con menos de la mitad de desaprobación evidenciaron probabilidades de consumo 3.53 veces mayor que en cortes con 90 % aprox. de desaprobación.

A nivel individual, [13] muestra que el nivel de aprobación personal es similar al nivel de aprobación de amigos cercanos, y que todos los grupos tienen una percepción similar de la aprobación del estudiante típico. Además un mayor uso de marihuana tiende a producir mayor aprobación personal, mayor aprobación percibida de los amigos cercanos y mayor aprobación por parte de los padres.

Las comunidades online también fueron exploradas en relación al consumo de drogas. En [17] fue examinada la conexión entre las características de la red online y el consumo de sustancias de adultos jóvenes. El uso de drogas fue asociado con un elevado número de conexiones, y una elevada proporción de la red que discute y acepta el consumo de drogas. También se halló que la densidad de la red y el número total de conexiones fueron asociados a mayor consumo personal en hombres.

En [6] se evaluó la asociación entre la presencia del contenido de consumo de alcohol y otras drogas, las normas percibidas, y el consumo de marihuana

en adultos jóvenes. El remordimiento anticipado fue negativamente asociado con el consumo de marihuana. Al igual que en estudios mencionados antes, el consumo por parte de pares fue positivamente asociado con el consumo individual. Esto quiere decir que los resultados son similares tanto para efectos sociales online como efectos sociales normales.

La idea de explorar las redes sociales con el fin de complementar medios tradicionales de recolección de información con respecto a drogas no es nueva. En [4] se introducen herramientas para usar datos desde redes sociales. Se sugiere un enfoque estructurado para capturar tendencias emergentes en el abuso de drogas aplicando métodos de inteligencia artificial, computación lingüística, teoría de grafos y modelamiento basado en agentes. Es más, se sugiere a *Twitter* como una red social disponible públicamente para obtener datos.

Otra forma usada para obtener información de las redes sociales es la aplicación de técnicas de Opinion Mining. Este es un sub-campo de Text Mining que permite extraer las opiniones desde documentos. Específicamente, puede ser aplicado en el contenido emitido por los usuarios de la Web. [3] hace una revisión bibliográfica reciente con respecto al sub-campo. En él, se plantea el problema de extracción de opiniones y se nombran los principales enfoques para resolverlo. [2] detalla la utilización del enfoque no supervisado basado en lexicones. Este enfoque explota reglas y heurísticas obtenidas del lenguaje, basándose en la polaridad individual de las palabras y la aplicación de reglas que pueden cambiar o intensificar la polaridad del conjunto de palabras. Particularmente, este trabajo tuvo como resultado una API de Opinion Mining.

---

### 3. Propuesta de Investigación

---

El objetivo de este estudio es replicar los indicadores de prevalencia y percepción de riesgo de la marihuana del Estudio Nacional de Drogas, utilizando datos recolectados únicamente desde *Twitter*. La idea inicial es construir una red social de usuarios chilenos de *Twitter* con información suficiente para reconocer patrones en el comportamiento de las personas.

#### 3.1. Hipótesis de Investigación

La forma en que los usuarios de las redes sociales generan información y el modo en que éstos se relacionan, incentivan a entidades a sumergirse dentro de esos escenarios para obtener conocimiento. Especialmente, la semejanza entre la estructura de las redes reales y online, y el tamaño de las comunidades promueven el pensamiento de que cualquier descubrimiento obtenido allí es

generalizable para la población. A todo esto es sumado el estado actual de algoritmos que facilitan el cálculo de sentimientos y clasificación de un texto, un tipo de dato sin estructura.

Basado en lo expuesto antes, se declara la siguiente hipótesis de investigación: *“Es posible extraer y procesar información de Twitter para representar un fenómeno complejo como el consumo y opinión sobre la marihuana de la población chilena”*

En este sentido, se quiere clasificar a los textos (*tweets*) y calcular su polaridad de sentimientos para luego transformarlos en variables asociadas a cada usuario. En base a éstas y a otras métricas derivadas de Análisis de Redes Sociales se pretende clasificar el consumo de marihuana a nivel individual. Además se quiere evaluar la opinión con respecto a la droga. Estos resultados pueden utilizarse en la elaboración de índices agregados que ayuden a explicar el consumo de marihuana a nivel nacional.

### 3.2. Modelos

En primer lugar, se requiere calcular la polaridad de sentimientos para los *tweets*, y clasificar a los mismos con respecto a tres acciones:

- Consumo de marihuana (binaria).
- Mención de política de control de marihuana (binaria).
- Venta de marihuana (binaria).

En segundo lugar, se pretende utilizar los *tweets*, la información del usuario y análisis de redes sociales para determinar dos características en el usuario:

- Consumo de marihuana (binaria).
- Rango etario (categórica).

Cada punto anterior se llevará a cabo mediante algoritmos de aprendizaje supervisado, con excepción del cálculo de polaridad de sentimientos, ya que éste se implementará mediante algoritmos de Opinion Mining, utilizando el enfoque no supervisado basado en lexicón. En los textos, las variables generalmente nacen desde las mismas palabras representadas de forma matricial. Esta representación se utilizará para las tres clasificaciones de *tweets* y la edad en usuarios. Luego de aplicar esta transformación, el problema se convierte en uno típico de minería de datos. Por lo tanto, junto con el consumo en usuarios, serán evaluados los algoritmos que suelen obtener mejor rendimiento. Por ejemplo, Redes Neuronales, Support Vector Machines, Naïve Bayes, etc.

En la evaluación de rendimiento para cada modelo de clasificación, serán utilizadas algunas métricas derivadas de la matriz de confusión: *Precision*, *Recall* y *F-Measure*. Priorizando la primera por sobre las demás. Además será utilizada la técnica de validación cruzada en el entrenamiento de algoritmos.

---

## 4. Recolección de Datos

---

En esta sección se describirá el tipo de información que fue utilizada como datos de entrada para el estudio. En primer lugar se profundizará en la información extraíble por medio de la API de *Twitter*, moldeada por la naturaleza del servicio. Luego, se establecerá la estructura de datos requerida para el entrenamiento de los algoritmos de aprendizaje. Más tarde, se precisará la forma para obtener los datos desde su fuente. Finalmente, se definirán por separado los mecanismos de etiquetado de textos y de usuarios para sus respectivos análisis y algoritmos de clasificación.

### 4.1. Datos disponibles

La información disponible en *Twitter* está moldeada por las funcionalidades que ofrece su servicio de *microblogging*. Para efectos de este estudio, la información útil se puede dividir en tres tipos:

- Información acerca del usuario.
- Red del usuario formada por sus conexiones con otros usuarios.
- *Tweets* publicados por el usuario.

### 4.2. Estructura Necesaria

Para llevar a cabo las clasificaciones antes mencionadas se utilizarán algoritmos de aprendizaje supervisado, donde es necesario tener un conjunto de casos previamente etiquetados (manualmente). Además en el caso del texto, se requiere la construcción de un conjunto de variables que representen a cada documento. Las Tablas 1 y 2 muestran ejemplos de los etiquetados necesarios para *tweets* y para usuarios, respectivamente.

Tweet	Variables					Etiquetas		
	1	...	k	...	m	Consumo	Política	Venta
1	1	...	1	...	0	1	0	0
2	0	...	1	...	0	1	1	0
⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮	⋮
j	0	...	0	...	1	0	0	1
⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮	⋮
n	1	...	1	...	0	0	0	1

Tabla 1: Etiquetado de *tweets*

Usuario	Variables					Etiquetas	
	1	...	k	...	m	Consumo	Edad
1	1.3	...	11	...	0	1	18
2	0.2	...	20	...	0	1	35
⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮
i	2.2	...	5	...	1	0	22
⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋮
n	-0.1	...	4	...	0	0	15

Tabla 2: Etiquetado de usuarios

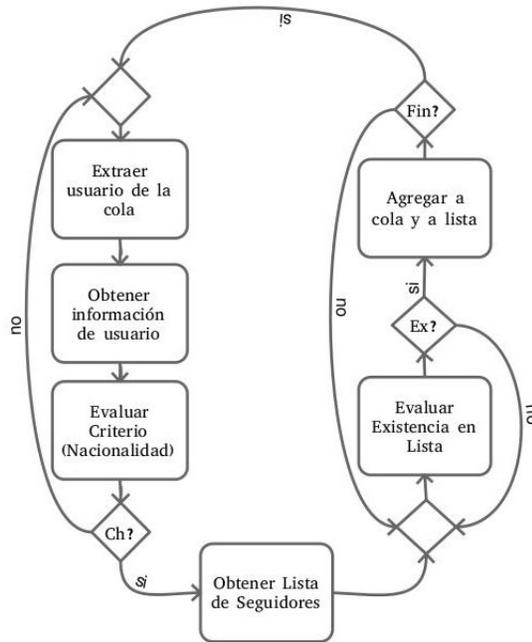
### 4.3. Recolección de Datos

El enfoque de *Twitter* es la difusión de mensajes cortos a través de la red de usuarios, conectados a través de las relaciones de seguimiento. Estos mensajes no son más que enlaces direccionados de un usuario (nodo) a otro. El objetivo fue obtener la red de usuarios chilenos, es decir, el conjunto de usuarios chilenos y sus conexiones, y los *tweets* publicados por cada uno de ellos.

El algoritmo de obtención de usuarios operó como uno de los algoritmos clásicos en recorrido de grafos, denominado Búsqueda en Anchura. La noción detrás del algoritmo es la siguiente: para cada elemento en la red se agregan todos los elementos adyacentes a él. Este procedimiento fue adaptado ligeramente para asemejar a un *Web Crawler*, llamado Crawler Focalizado. El cual recorre el grafo de la misma manera, pero sólo son agregados los nodos adyacentes de aquellos elementos que cumplan con cierto criterio. En el caso de este estudio, el criterio consiste en que los usuarios sean chilenos, cuya información está contenida en los datos del usuario.

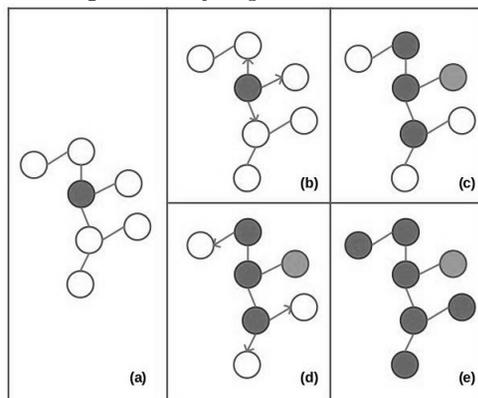
En la Figura 1 se muestra el proceso iterativo de extracción de usuarios y en la Figura 2 se muestra un ejemplo de dos iteraciones. El algoritmo comienza con una semilla en la Figura 2.a, en las figuras 2.b y 2.d se incorporan los nodos adyacentes que cumplen el criterio (nodos oscuros) y en las figuras 2.c y 2.e se muestra el estado final de cada iteración.

Figura 1: Crawler de Usuarios



Es importante mencionar que existe un porcentaje de usuarios de *Twitter* que bloquean el acceso a sus *tweets*. Para efectos del diseño, no fueron considerados los usuarios con esta condición.

Figura 2: Ejemplo de Iteración



Una vez establecida la base de usuarios se procedió a extraer el conjunto de *tweets* publicado por cada uno. El modo de clasificar los *tweets* relacionados con marihuana consistió en identificar palabras clave. El listado de palabras clave tuvo como origen a tres fuentes diferentes: conocimiento experto, bibliografía y una encuesta de uso de palabras actuales.

El conjunto total de palabras fue filtrado para confirmar su uso en *Twitter* y desambiguar el contexto de uso. Para hacer esto se extrajo un grupo *tweets* que contiene las palabras. Fueron aplicados algunos algoritmos de pre-procesamiento de texto para emplear *Topic Modeling* con el fin de identificar contextos diferentes de uso de las palabras (clusters) y verificar su empleo relacionado con marihuana. Luego, las palabras ambiguas fueron filtradas, considerando sólo aquellos *tweets* que contuvieran la cadena de caracteres “fum”. Esta regla parece ser muy restrictiva, pero un análisis exploratorio la arrojó como la palabra común más utilizada en el contexto.

Una vez que un *tweet* ha pasado el filtro, se calcula su polaridad con una API desarrollada en [2] que aplica *Sentiment Analysis* con el enfoque no supervisado basado en un lexicón etiquetado por emoción y un puntaje para cada término. El lexicón permite obtener un puntaje que va desde -50 a +50 donde el positivo indica que un texto habla positivamente, mientras que un puntaje negativo indica que la emoción es negativa. El lexicón utilizado fue desarrollado en [10]. Esta API aplica tres reglas gramaticales del lenguaje natural del español. La primera regla consiste en los intensificadores, es decir, palabras que amplifican el significado de una palabra que se encuentre en el lexicón. La segunda regla, es la de la negación, la cual invierte el valor que se obtiene en el lexicón. La tercera, y última regla, corresponde a las clausulas adversativas, esto es el uso de “pero” o palabras similares, en donde la primera parte antes de la clausula recibe una ponderación menor a la que viene después.

#### 4.4. Etiquetado de *Tweets*

Con el fin de obtener un etiquetado consistente de los documentos, se diseñó un conjunto de reglas de etiquetado de textos, las cuales son enumeradas a continuación:

1. El evaluador debe etiquetar el documento en respuesta a una pregunta definida claramente para cada categoría.
2. Cada una de las categorías del documento será evaluada en la misma instancia (por *tweet*).
3. Cada documento debe ser etiquetado por sólo una persona que califique como experto (usuario de *Twitter*).

4. Se seleccionará un porcentaje de casos que será etiquetado por todas las personas. Para ese conjunto se determinará el índice Kappa de Cohen, el cual indica la concordancia entre etiquetadores.

Dada la experiencia del grupo investigador, se buscó un número de *tweets* para la muestra que cumpliera con tener un error del 3 % al 98 % de confianza. Teniendo esos requerimientos se hace necesario contar con una colección de *tweets* de al menos 1.500. Este conjunto de *tweets* fue clasificado por un grupo de 12 personas. Es importante mencionar que no es sencillo contar con personas que pueda etiquetar manualmente todos los textos, por lo que se utilizó la siguiente metodología para que no tuvieran que etiquetar todos los datos. Primero, se separaron 50 *tweets* de los 1.500, los cuales fueron etiquetados por las 12 personas y que sirve para poder ver la concordancia entre los participantes. Luego, de los 1.450 restantes, se dividieron en 12 grupos para que cada evaluador se encargara solamente de 1 de ellos. Por último, la división y la distribución de textos fue hecha al azar.

#### 4.5. Etiquetado de Usuarios

El etiquetado se desarrolló mediante una encuesta directa a los usuarios. Una vez construida una base de datos de usuarios chilenos, se escogieron casos al azar para enviarles la encuesta. La misma fue publicada en un *tweet* para cada usuario, mencionándole directamente. La encuesta contuvo preguntas para determinar el consumo de marihuana, la edad y el sexo de cada usuario.

---

## 5. Resultados y Discusión

---

En esta sección será presentado el cúmulo de resultados derivados del estudio. Todos los datos brindan información relevante para comprender el fenómeno de *Twitter* y el consumo de marihuana dentro de ese contexto.

En primer lugar, se abordarán los resultados arrojados por la selección de palabras clave, que fueron utilizadas en la recolección de *tweets*. A continuación de esto, se hará referencia a información originada en la recolección de datos. Luego, serán mencionados algunos datos que fueron obtenidos en el etiquetado de *tweets* y usuarios. La evaluación de algoritmos también tendrá destinada un segmento. Finalmente, se pondrá enfoque en las métricas elaboradas a partir de los resultados anteriores.

### 5.1. Palabras Clave

El proceso de selección de palabras arrojó un total de 46 cadenas de caracteres, variando entre simples, bigramas y trigramas. Están divididos en dos grupos: cadenas únicamente relacionadas con marihuana y otras cadenas ambiguas, en las cuales se agrega el criterio mencionado anteriormente. La lista final de palabras clave se muestra en la Tabla 3.

Lista de Palabras Clave		
marihuana	cannabis	weed
mariguana	marijuana	prensada
porro (f)	thc	pito (f)
caño (f)	yerba (f)	sativa
sacate uno	canabis	macoña
de la buena (f)	hierba (f)	mota (f)
ganjah	cuete (f)	prensao
ganja	faso (f)	paraguaya (f)
de la wena (f)	cogollo (f)	bongazo
ganya	hachis	pitito (f)
matacola	hierva (f)	paragua (f)
marihuanita	troncho (f)	la verde (f)
canabica	cogollito (f)	pitits
cogoyo (f)	marimba (f)	paraguayo (f)
huiro (f)	bless (f)	yerva (f)
sacateuno		

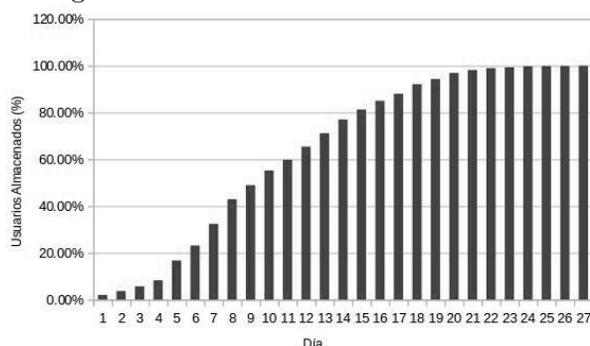
Tabla 3: Palabras Clave

### 5.2. Recolección de Datos de *Twitter*

La velocidad de recolección de *tweets* y usuarios fue una medida crucial durante el estudio. Ella depende de varios factores, tales como la velocidad de procesamiento de los recursos, la velocidad de Internet, el número total de nodos que fueron analizados, entre otros. Todo esto influyó en la cantidad de usuarios que fueron evaluados y almacenados cada día.

La Figura 3 muestra el porcentaje de usuarios acumulados desde el día que inició la extracción. El total de usuarios al final del periodo de extracción fue de 1.505.367, aunque el número de usuarios válidos para el análisis fue de 1.361.285, debido al bloqueo de información por parte de ellos. Se pueden apreciar tres fases diferentes en la Figura 3.

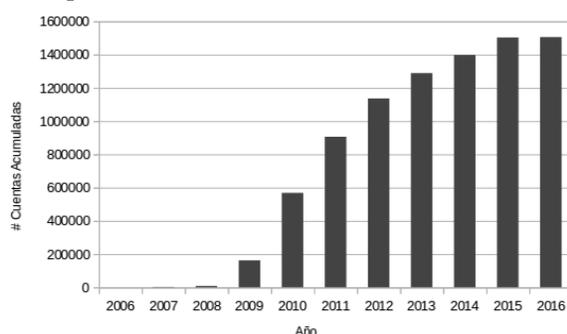
Figura 3: Gráfico de usuarios acumulados



La base de usuarios determina la información que puede ser extraída, debido a la fecha en que fueron consignados los datos. Por ejemplo, es imposible obtener métricas para fechas en donde no existían usuarios chilenos en *Twitter*. Lo anterior se ve reflejado en las Figuras 4 y 5, en donde la primera muestra el número acumulado de cuentas chilenas en *Twitter* para cada año y la segunda, revela el número de *tweets* creados para cada año.

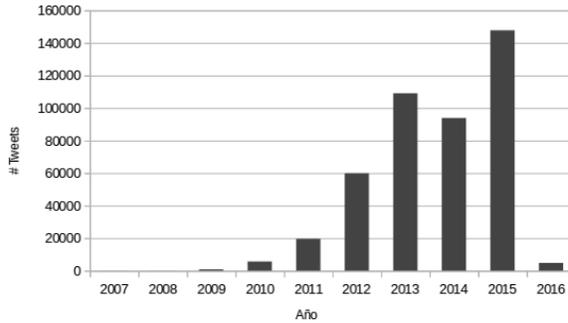
La Figura 4 revela que para años anteriores al 2009 existían poco usuarios, por lo que desacredita resultados que puedan ser originados para esos años. La Figura 5 muestra la composición de *tweets* en la base de datos, cuya forma está determinada por la restricción de *Twitter* de los últimos 3.200 *tweets* por usuario y el número de usuarios por año. En efecto, el grueso de los *tweets* se encuentra entre los años 2010 y 2016. Cabe recordar que sólo son almacenados los *tweets* relacionados con marihuana.

Figura 4: Gráfico de cuentas acumuladas



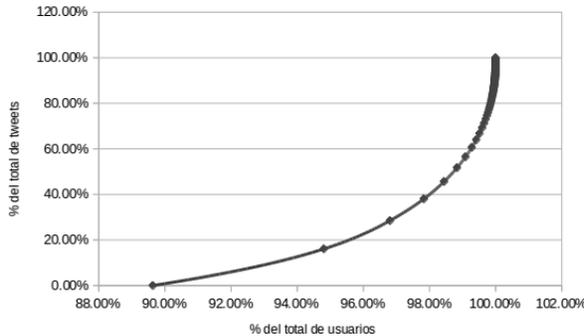
Los *tweets* no sólo son analizados desde el punto de vista de su distribución de tiempo, sino que también desde el potencial de generación por parte de los usuarios. En otras palabras, es interesante determinar cuántos usuarios están involucrados en la generación de la mayoría de los *tweets* ligados con

Figura 5: Número de *tweets* por año



marihuana. La curva de *Lorenz* de la Figura 6 explora esta idea, determinando que cerca del 10 % de los usuarios han producido el total de *tweets* almacenados en la base de datos, representando un total de 141.063 usuarios. Es aún más impresionante observar que cerca del 2 % de los usuarios generaron un 60 % de los datos, lo cual exhibe la desigualdad en la producción de textos de esta naturaleza.

Figura 6: Curva de *Lorenz* de *Tweets*



### 5.3. Etiquetado de *Tweets*

A lo largo de los capítulos se han mencionado las categorías en que era necesario etiquetar los *tweets*. Tres de éstas están dedicadas para el entrenamiento de algoritmos y otra para determinar la precisión de las palabras clave. El proceso de etiquetado de *tweets* arrojó un total de 1.450 únicamente etiquetados y 50 etiquetados por cada una de las 12 personas. Es lógico empezar por los resultados obtenidos desde este último grupo, es decir, las medidas de acuerdo.

La Tabla 4 resume las medidas de acuerdo entre las 12 personas. Se observa un amplio nivel de acuerdo relativo para todas las categorías, todas superan-

do el 0,95. Esta medida bruta es corregida para incorporar los efectos de la aleatoriedad en el proceso de etiquetado. Esto da como resultado el coeficiente *Kappa* de Fleiss, dedicado a reflejar el nivel de acuerdo entre más de dos personas. Tal como se visualiza en la Tabla 4, todas las categorías muestran un coeficiente cercano al 60 %, a excepción de la categoría de venta. Esto se debe a que pesar de tener el nivel de acuerdo relativo más alto, los datos no tiene mayor variabilidad, por lo que el coeficiente es castigado directamente. Sin considerar esta categoría, las etiquetas cuentan con fuerza moderada de acuerdo.

Categoría	Acuerdo Relativo	Kappa de Fleiss
Ligado a marihuana	0.95	0.60
Consumo	0.89	0.56
Políticas	0.87	0.56
Venta	0.99	0.09

Tabla 4: Medidas de Acuerdo

Los 50 casos producen un dilema al momento de completar los 1.500 casos, ya que algunos reflejan contradicción entre las personas. Para solucionar esto se aproximó al promedio en cada uno de los 50. Los porcentajes para los casos positivamente clasificados se muestran en la Tabla 5. En ella se ve que el porcentaje realmente relacionado con marihuana es de 94,73 %, reflejando la precisión del procedimiento de búsqueda de palabras clave. Las categorías de consumo y políticas tienen heterogeneidad suficiente para el correcto entrenamiento de algoritmos. No así la categoría de venta, ya que sólo tiene heterogeneidad del 0,20 %, cantidad insuficiente para el entrenamiento de cualquier clasificador confiable. Por esta razón y su coeficiente *Kappa*, esta categoría es apartada de cualquier manipulación en etapas siguientes.

Categoría	Heterogeneidad
Ligado a marihuana	94,73 %
Consumo	13,80 %
Políticas	18,87 %
Venta	0,20 %

Tabla 5: Heterogeneidad en las etiquetas

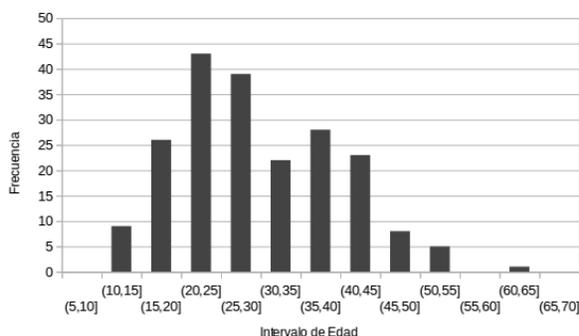
#### 5.4. Etiquetado de Usuarios

La encuesta a usuarios de *Twitter* fue realizada en el periodo comprendido entre el 9 de Febrero del año 2016 y el 6 de Marzo del mismo año. En ese

periodo fue contestada por un total de 209 personas. Luego del cruce con la base de datos fueron obtenidos 204 casos factibles de uso, reflejando una tasa de respuesta del 0,3%. Algunos casos descartados no fueron hallados en la base de datos y otros ni siquiera en una búsqueda manual por *Twitter*. Utilizando la tasa de consumo anual de marihuana (11,3%) del año 2014 como tasa de heterogeneidad de los datos, se obtiene un error de 4,35% y un nivel de confianza de 95%.

Las 3 preguntas sencillas abordadas en la encuesta arrojaron algunas estadísticas para el análisis. En primer lugar, la edad de los casos recogidos brindan una visión corta de la distribución de edad de los usuarios chilenos de *Twitter*. La Figura 7 muestra dicha distribución, se aprecia claramente la ausencia de edades a los extremos. El intervalo con más presencia es entre los 20 y 30 años, y la edad media de la muestra es de 30,1 años. Por otro lado, 42,1% de los usuarios reconocieron el consumo de marihuana en el último año y un 34,3% en el último mes. Además un 43,6% de los casos corresponden al sexo femenino. Lo antes mencionado es comparado con la edad media de 34,9 y el 51,4% de mujeres obtenidos en el CENSO del año 2012.

Figura 7: Distribución de edad de la muestra



Los datos obtenidos por la encuesta desprenden varios puntos interesantes para el estudio. El primero va de acuerdo con la creencia de que los usuarios de *Twitter* son más jóvenes que la población general. El segundo sugiere la presencia de más hombres que mujeres, al contrario que los datos del CENSO. Cabe destacar que es muy probable que el tipo de encuesta esté presentando distorsiones. Sin ir más lejos es de esperarse que el alto porcentaje de prevalencia anual sea reflejo de la percepción que la encuesta estaba enfocada para consumidores de marihuana.

## 5.5. Evaluación de Algoritmos

En esta sección fueron elegidos los algoritmos que contaron con las mejores medidas de rendimiento para las tareas de clasificación y regresión. En cada tarea fueron evaluados varios algoritmos, fue seleccionado aquel que contó con mayor poder predictivo.

Antes de presentar los números es necesario aclarar algunos puntos con respecto al procedimiento de evaluación y elección:

- La representación matricial de un grupo de textos produce una cantidad enorme de atributos. *Information Gain* fue la única técnica de reducción de atributos que fue empleada, ya que ha mostrado dar buenos resultados en texto. A pesar de esto, los algoritmos con mejor rendimiento no la incorporan, por lo que no será mencionada.
- Los algoritmos no son los únicos que varían en el proceso de prueba. Hay una serie de parámetros que pueden ser modificados, pero su procedimiento de evaluación no será detallado. El rendimiento de cada algoritmo incorpora intrínsecamente estas modificaciones, siendo sólo nombradas junto al algoritmo.
- Varios algoritmos de entrenamiento fueron descartados por sus costos de empleo (tiempo de entrenamiento y exigencia computacional). Esto es aplicable a los textos, debido a la gran cantidad de atributos que generan.
- Si bien todas las medidas de rendimiento brindan información relevante acerca de la aplicación del algoritmo, en la elección se priorizó aquellos que tuvieron mayor precisión para la clase de interés. Esto refleja la necesidad de recuperar casos en que efectivamente se evidencie el comportamiento.

### 5.5.1. Consumo en *tweets*

La evidencia de consumo de marihuana en *tweets* es abordado como un problema de categorización binaria, es decir, un problema de clasificación clásico. Por esto, existe una gran cantidad de algoritmos capaces de realizar la tarea. Aquí fue evaluada la utilidad de tres algoritmos: *Naive Bayes* con monogramas y vectores de atributos binarios, *Voted Perceptron* con monogramas y vectores log-normalizados, y *Support Vector Machines* con monogramas a trigramas y vectores binarios.

Las Tablas 6, 7 y 8 muestran las medidas de rendimiento para cada uno de los algoritmos. En ellos se aprecian los valores de *Precision*, *Recall* y *F-Measure*. Los valores ponderados de todas las medidas se ven beneficiados de las altas cifras y la gran cantidad de casos para la clase cero. Como fue puntualizado anteriormente, se priorizó la *Precision* de la clase de consumo, por lo tanto fue elegido el modelo de *Support Vector Machines*. El valor de *Recall* puede parecer poco, pero es compensado por su *Precision*, que si bien no es muy alto, es el mejor entre todos.

Clase	Precision	Recall	F-Measure
No consumo (0)	0,923	0,838	0,878
Consumo (1)	0,358	0,565	0,438
Ponderado	0,845	0,8	0,818

Tabla 6: Rendimiento de *Naive Bayes* para el consumo en *tweets*

Clase	Precision	Recall	F-Measure
No consumo (0)	0,88	0,971	0,923
Consumo (1)	0,486	0,174	0,256
Ponderado	0,826	0,861	0,831

Tabla 7: Rendimiento de *Voted Perceptron* para el consumo en *tweets*

Clase	Precision	Recall	F-Measure
No consumo (0)	0,883	0,978	0,928
Consumo (1)	0,588	0,193	0,291
Ponderado	0,843	0,87	0,84

Tabla 8: Rendimiento de SVM para el consumo en *tweets*

### 5.5.2. Políticas en *tweets*

Al igual que en la parte anterior, la presencia de políticas relacionadas con marihuana en los *tweets* es un problema clásico de clasificación binaria. A pesar de ello, en esta oportunidad el conjunto de algoritmos es diferente: SVM con monogramas a trigramas y vector log-normalizado, *Voted Perceptron* con monogramas y vector log-normalizado, y Árbol de Decisión C4.5 con monogramas y vector binario.

Las Tablas 9, 10 y 11 muestran las cuatro medidas de rendimiento para cada uno de los algoritmos bajo mira. Las tres alternativas tienen valores parecidos en las medidas de la clase cero y la ponderación para ambas clases.

Por ende, el factor diferenciador está en las métricas de la clase 1. Nuevamente la decisión residió en la mayor *Precision*, es decir, *Voted Perceptron*. Cabe destacar que los otros dos algoritmos tienen asociados mayores valores de *Recall*, pero son menospreciadas a cambio del valor antes mencionado.

El mejor modelo para la clasificación de políticas resulta ser considerablemente mejor que su par de consumo. De hecho, es 0,23 veces mejor en *Precision*. Esto implica que la presencia de elementos que permitan determinar si un *tweet* corresponde a políticas relacionadas con marihuana es más clara. Pudiendo ser necesario más contexto para determinar de manera certera si un *tweet* menciona consumo de marihuana.

Clase	Precision	Recall	F-Measure
No políticas (0)	0,865	0,967	0,913
Políticas (1)	0,714	0,353	0,473
Ponderado	0,837	0,851	0,83

Tabla 9: Rendimiento de SVM para políticas en *tweets*

Clase	Precision	Recall	F-Measure
No políticas (0)	0,86	0,971	0,912
Políticas (1)	0,722	0,322	0,445
Ponderado	0,834	0,849	0,824

Tabla 10: Rendimiento de *Voted Perceptron* para políticas en *tweets*

Clase	Precision	Recall	F-Measure
No políticas (0)	0,874	0,946	0,908
Políticas (1)	0,639	0,413	0,502
Ponderado	0,83	0,845	0,832

Tabla 11: Rendimiento de C4.5 para políticas en *tweets*

### 5.5.3. Edad de Usuarios

La predicción de edad comparte la misma base de las clasificaciones anteriores. En el sentido de que utiliza elementos del lenguaje para reconocer parámetros ocultos que determinen la edad de las personas. Se apoya en la percepción de que los individuos cambian el conjunto de palabras que ocupan a lo largo de su vida y que generaciones enteras comparten elementos léxicos.

El elemento novedoso de esta parte radica en que ya no se trata de encapsular los textos dentro de categorías, sino que se intenta asociar a los casos

dentro de un rango de valores. Esta diferencia también se ve reflejada en las métricas de rendimiento que se ocupan. En esta ocasión son utilizadas medidas de relación lineal y diferencias agregadas entre los datos reales y los predichos. Específicamente se utilizan la correlación de *Pearson* y otros errores.

En esta instancia fueron evaluados tres algoritmos diseñados para realizar regresiones de datos: Regresión Lineal, M5P y la versión de *Support Vector Machines* para regresiones. Todas fueron entrenados con monogramas y vectores de frecuencias log-normalizados. Aunque el mejor modelo tiene una pequeña variación. Las medidas de rendimiento se muestran en la Tabla 12, ahí figuran todas las opciones más una versión de SVM con vectores binarios. Se puede apreciar claramente que las medidas mejoran estrictamente de arriba hacia abajo. El mejor modelo es la última versión de SVM, teniendo una correlación de *Pearson* de 0,583 y error absoluto medio de 6,28. En otras palabras, el modelo se equivoca en promedio cerca de 6 años.

Modelo	Correlación	MAE	RMSE
Regresión Lineal	0,248	7,913	9,792
M5P	0,469	7,286	9,234
SVMreg log-normalizado	0,526	6,573	8,503
SVMreg binario	0,583	6,280	8,151

Tabla 12: Rendimiento de algoritmos de edad

#### 5.5.4. Consumo de Usuarios

La clasificación de consumo de marihuana por parte de los usuarios se incluye junto a los típicos modelos de Minería de Datos, debido a que aquí no se hará tratamiento de textos para conseguir un conjunto de atributos. El grupo de 13 atributos, compuesto por medidas derivadas de los *tweets* y el entorno social del usuario, será utilizado para predecir su consumo de marihuana.

Fueron utilizados tres algoritmos para comparar sus rendimiento en la clasificación: *Support Vector Machines*, *Multilayer Perceptron* y *Voted Perceptron*. Los tres en sus versiones optimizadas arrojaron medidas casi idénticas, sólo variando en la medida *ROC Area*, sugiriendo que los tres algoritmos expresen casi todo el poder predictivo del conjunto de variables. La Tabla 13 muestra el conjunto de medidas de rendimiento para SVM, utilizado por defecto como modelo final, porque permite apreciar la influencia de las variables en la clase. Estos valores se diferencian ampliamente a modelos anteriores, ya que las medidas están balanceadas. Esto se cumple para las dos clases y para los valores de *Precision* y *Recall*. En síntesis, individualmente será recuperado el

68,4 % de los consumidores de marihuana y 62,8 % del total de predichos serán efectivamente consumidores en el último año.

Clase	Precision	Recall	F-Measure
No consumo (0)	0,757	0,709	0,732
Consumo (1)	0,628	0,684	0,665
Ponderado	0,703	0,698	0,7

Tabla 13: Rendimiento de SVM para consumo en usuarios

La Tabla 14 muestra los pesos normalizados para cada variable. Los datos indican que las variables con mayor poder predictivo son la edad, la emisión de *tweets* relacionados con marihuana, *tweets* sobre políticas de marihuana, el porcentaje de consumidores en el vecindario personal y las nominaciones fuera de la red social. Específicamente, la primera y la quinta disminuyen el riesgo de consumo, y la segunda, la tercera y la cuarta lo aumentan. En contraposición, la polaridad y la densidad son las variables más débiles.

Atributo	Peso Normalizado
Edad	-1,58
<i>Tweets</i> de marihuana	2,67
Consumo en <i>tweets</i>	0,51
Políticas en <i>tweets</i>	1,50
Polaridad	-0,14
Polaridad de Políticas	0,68
Seguidores	0,27
Densidad	0,05
<i>Reach Centrality</i>	0,95
Uso en vecindario	1,37
Polaridad en Vecindario	0,31
Distancia a consumidores	0,30
Nominaciones externas	-1,80
Intercepto	0,13

Tabla 14: Influencia de variables en el consumo de marihuana

Se confirman varias creencias y se replican algunos resultados obtenidos en la literatura. En primer lugar, los consumidores se concentran en segmentos de edad más jóvenes. La popularidad de una persona aumenta su riesgo a consumir marihuana. El comportamiento del entorno influencia directamente al comportamiento de las personas. Bajo este contexto, la emisión de *tweets* de consumo de los amigos predice con mayor fuerza que la emisión propia.

La publicación de cualquier mensaje relacionado con marihuana dice mucho del consumo. Además, todas las medidas de cercanía a otros consumidores aumentan el riesgo de consumo.

## 5.6. Indicadores

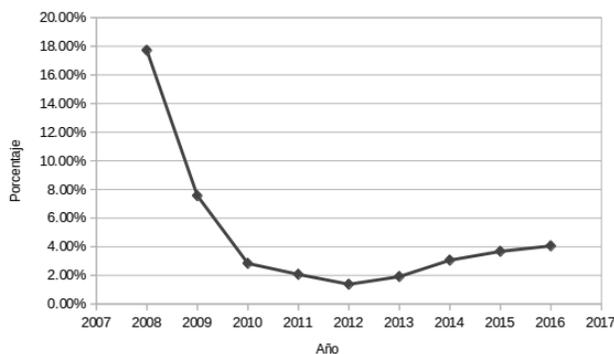
En esta sección será mostrado el producto final de esta memoria: las réplicas inicialmente prometidas y resultados. En esta ocasión se procederá de la siguiente manera: prevalencia, frecuencia de consumo, polaridad, polaridad de políticas, porcentaje de consumidores entre amigos, oferta de marihuana, y palabras utilizadas en *tweets* de consumo.

### 5.6.1. Prevalencia

En epidemiología, la prevalencia representa el porcentaje de la población que evidencia cierta característica en un periodo de tiempo. En este caso se trata de perseguir la misma definición, pero desde datos consignados en *Twitter*. Para obtener esta métrica es necesaria toda la estructura de la aplicación, desde los recolectores de información de *Twitter* hasta el clasificador de consumo. Este último es utilizado para determinar el consumo de marihuana en el último año para una muestra de usuarios.

La Figura 8 muestra el cálculo de prevalencia para cada año, la cual revela un alto porcentaje para los años anteriores al 2010. Los valores para esos años pueden estar sobrestimados debido a los pocos datos de usuarios y *tweets* para ese periodo, y la utilización de un supuesto clave: las relaciones entre usuarios creados en ese tiempo no han cambiado drásticamente al avanzar los años. Los años posteriores al 2009 muestran una evolución paulatina del consumo de marihuana entre los usuarios chilenos de *Twitter*.

Figura 8: Prevalencia Anual



Un análisis necesario, para determinar la representatividad de los datos con respecto a la población chilena, es realizar una comparación entre los datos mostrados en la Figura 8 y los datos recogidos por la Encuesta Nacional de Drogas. La prevalencia histórica arrojada por esta encuesta es expuesta en la Figura 9. Se puede apreciar una similitud entre la tendencia de los años 2010 y 2016 de la curva producida por el predictor y la tendencia entre los años 2008 y 2014 del estudio nacional, aunque se tiene un desfase de dos años. La Figura 10 grafica la comparación de curvas para el periodo entre 2008 y 2014, corrigiendo a la curva predicha por un ponderador y desplazándola dos años atrás.

Figura 9: Prevalencia Nacional

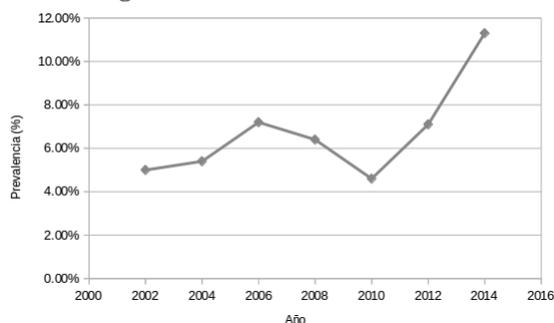
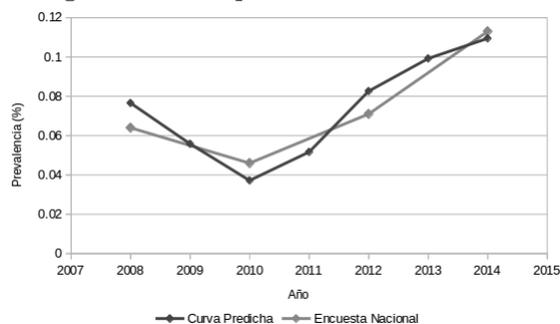


Figura 10: Comparación de Prevalencia

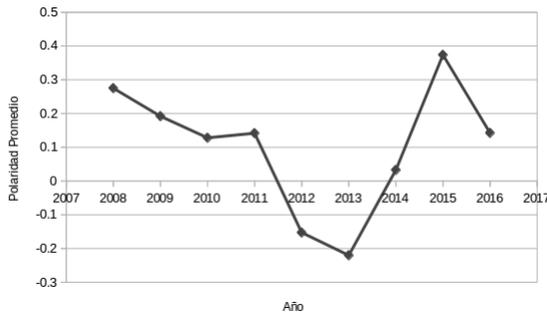


La gran similitud entre curvas es innegable, presentando un coeficiente de correlación de *Pearson* de 0,933. Aunque hay que destacar que la curva predicha fue retrasada en dos años. Este desfase podría estar producido por las variables utilizadas en el predictor, es decir, el resultado está totalmente condicionado a elementos presentes en *Twitter*. Esto quiere decir que el consumo de marihuana no es reflejado inmediatamente en el contexto de *Twitter*, ya que requiere que los usuarios presenten el comportamiento y luego generen contenido que esté relacionado con él. El desfase será replicado para el análisis de otras métricas.

### 5.6.2. Polaridad

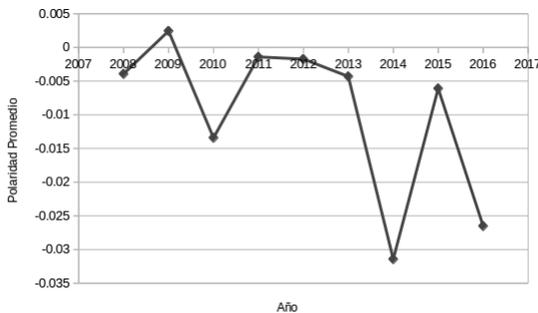
La polaridad de *tweets* refleja que tan positivos o negativos son los textos emitidos por los usuarios de *Twitter*. Esta métrica trata de incorporar las opiniones vertidas en el *tweet* promedio para cada periodo y así, realizar seguimiento al efecto en la opinión de las personas a partir de ciertos eventos. La polaridad es calculada para cada año, mes y día para los cuales se poseen datos. La Figura 11 exhibe la evolución anual para esta métrica. En una primera instancia sólo se mencionará su forma, evidenciando una baja desde el año 2008 y recuperándose desde el año 2013.

Figura 11: Polaridad de *Tweets*



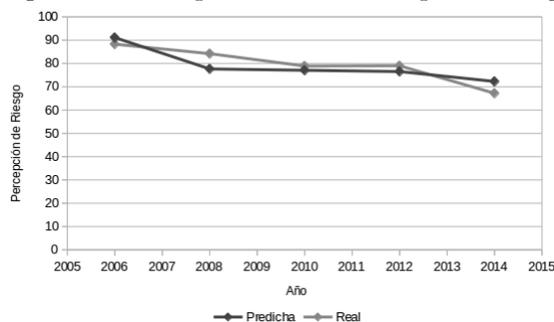
La polaridad también es calculada en base a los usuarios. Cada usuario tiene un número de *tweets* asociados y ellos, una polaridad. Se calcula tomando el promedio entre los *tweets* del usuario y luego, el promedio de los usuarios. Esto implica que gran cantidad de usuarios tendrán polaridad igual a cero. La Figura 12 muestra la evolución de la polaridad de usuarios a través de los años, detectándose una baja sostenida. Es importante destacar la diferencia de forma entre los dos gráficos de polaridad, señalando que la evolución diaria de los *tweets* no es incorporada directamente en la polaridad de usuarios.

Figura 12: Polaridad de Usuarios



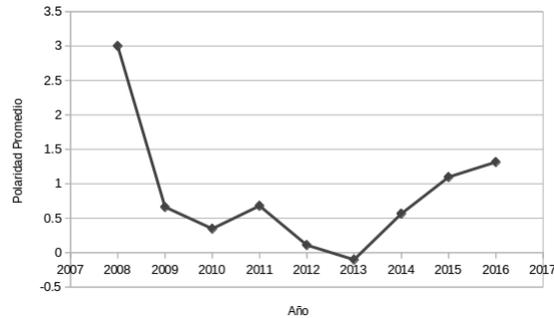
En un momento se planteó que la polaridad de marihuana podría estar relacionada con la percepción de riesgo de la droga. Obedeciendo la definición de percepción de riesgo que corresponde al porcentaje de la población que considera riesgoso el consumo experimental o frecuente de marihuana. Para efectos del análisis se considerará sólo el segundo. Por otro lado, es lógico pensar que esta medida tiene similitud con el promedio de polaridad para *tweets* negativos. En otras palabras, las mismas personas que opinan negativamente de la droga también la consideran riesgosa. La Figura 13 la explora de esta idea, comparando la percepción de riesgo con el promedio de polaridad negativa de *tweets*. Esta última transformada mediante una reflexión con respecto al eje horizontal, retrasada en dos años y escalada. El coeficiente de correlación de *Pearson* es de 0,819, evidenciando un gran parecido tanto gráfico como numérico y apoyando nuevamente la teoría del desfase.

Figura 13: Comparación de Percep. de Riesgo



### 5.6.3. Polaridad de Políticas

La polaridad de *tweets* de políticas relacionadas con marihuana comparte el mismo principio que su par mencionado anteriormente, pero esta vez es aplicado sólo a *tweets* clasificados como políticas. La Figura 14 muestra la curva de esta métricas a lo largo de los años. Se hayan diferencias claras con respecto al gráfico de polaridad general de *tweets* de marihuana. Por otro lado, es inevitable notar la relación entre el aumento de polaridad de los últimos años y el de toda la atención mediática que ha sufrido la marihuana en casi el mismo periodo. También es importante notar la similitud de la curva con la apreciada para la prevalencia. Esta relación es apoyada por el modelo predictor de consumo en usuarios.

Figura 14: Polaridad en *Tweets* de Políticas


---

## 6. Conclusiones y Trabajo Futuro

---

Este estudio propone la utilización de la información generada en *Twitter* para replicar un comportamiento en la población general. El funcionamiento contempla la combinación de varios algoritmos y procedimientos para obtener los resultados deseados. La aplicación permite extraer información de los usuarios de *Twitter* y el contenido relacionado con marihuana que ellos mismos crearon. Asimismo, faculta la clasificación de los *tweets* con respecto a varias categorías y el cálculo de polaridad. Además de esto, implementa un modelo de predicción individual de consumo de marihuana.

Uno de los mayores valores de la aplicación es que brinda la posibilidad de extraer información útil desde *tweets*, que directamente son textos, el ejemplo clásico de información no estructurada. El rendimiento de los clasificadores sobre texto es medianamente bueno, bordeando el 65% de *Precision* para la clase buscada y 84% ponderada. Pero se pueden apreciar diferencias con respecto a cada clasificación. La clasificación de políticas en *tweets* es claramente mejor, indicando que dependiendo del tema, la división entre clases es más ambigua o requiere más información del contexto.

En este trabajo se reconoce el valor de las relaciones entre usuarios de *Twitter*, ya que sin ellas disminuiría en gran medida el poder predictivo del clasificador de consumo de marihuana. Además reproduce resultados obtenidos en otros estudios realizados con redes sociales fuera del contexto virtual, implicando que el tipo de relación pasa desapercibido o que las relaciones en *Twitter* son reflejo de las relaciones de contacto directo. Se destaca que el consumo de marihuana es mayormente predicho por declaraciones de consumo por parte de amigos que las propias, dándole respaldo a los estudios que señalan

que el comportamiento de un individuo es fuertemente afectado por los pares.

Fue evidenciado un desfase de dos años entre los valores predichos por la aplicación y los recolectados por la Encuesta Nacional de Drogas. Se señala que el comportamiento se ve reflejado de forma retardada en las redes sociales, porque requiere que los individuos viertan esta información en sus cuentas. Aún así los datos son generados frecuentemente, ya que la polaridad es reportada diariamente, y el predictor tiene capacidad de determinar consumo a nivel individual.

Todo esto no sería posible sin los permisos concedidos por *Twitter* para acceder a la información. Si bien los casos de bloqueo de información por parte de los usuarios no son menores, el porcentaje que no lo hace brinda una gran cantidad de información para realizar el análisis. Con el tiempo *Twitter* podría implementar políticas tan restrictivas como las de Facebook.

Como trabajo futuro se plantean dos líneas de desarrollo. Primero, mejorar el rendimiento del clasificador de consumo de personas. Esto se puede hacer mediante la incorporación de variables que puedan explicar de mejor manera la varianza del comportamiento. Por ejemplo, se puede utilizar una técnica más refinada de conexiones, reflejando la intensidad de la relación. Finalmente, se propone replicar el estudio a otras drogas, especialmente las lícitas, como el alcohol y el tabaco. Es probable que tengan mayor presencia en las redes sociales y la metodología no requiere modificaciones.

**Agradecimientos:** Los autores agradecen al Servicio Nacional para la Prevención y Rehabilitación del Consumo de Drogas y Alcohol (SENDA), por su apoyo en la calibración de los modelos desarrollados. Este trabajo fue parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816).

## Referencias

- [1] M.M. Ali, A. Amialchuk, y D.S. Dwyer. The social contagion effect of marijuana use among adolescents. *PloS one*, 6(1):e16183, 2011.
- [2] J. Balazs. Diseño, desarrollo e implementación de una aplicación de web opinion mining para identificar el sentimiento de usuarios de twitter con respecto a una compañía de retail. Disponible en <http://wic.uchile.cl/wp-content/uploads/2016/01/Thesis-Jorge-Balazs.pdf>, 2015. Universidad de Chile, Tesis de Pregrado.

- [3] J.A. Balazs y J.D. Velásquez. Opinion mining and information fusion: A survey. *Information Fusion*, 27:95–110, 2016.
- [4] M. Chary, N. Genes, A. McKenzie, y A.F. Manini. Leveraging social networks for toxicovigilance. *Journal of Medical Toxicology*, 9(2):184–191, 2013.
- [5] M.J. Cleveland, M.E. Feinberg, D.E. Bontempo, y M.T. Greenberg. The role of risk and protective factors in substance use across adolescence. *Journal of Adolescent Health*, 43(2):157–164, 2008.
- [6] S.H. Cook, J.A. Bauermeister, D. Gordon-Messer, y M.A. Zimmerman. Online network influences on emerging adults' alcohol and drug use. *Journal of youth and adolescence*, 42(11):1674–1686, 2013.
- [7] Sistema de Información Regional de México y Fundación Chile 21. *Políticas de drogas en México y Chile: Estimación de costos económicos y sociales y de escenarios alternativos*. SIREM, Fundación Chile 21 y Foundation Open Society Institute (OSI-ZUG), 2013.
- [8] S.T. Ennett, K.E. Bauman, A. Hussong, R. Faris, V.A. Foshee, L. Cai, y R.H. DuRant. The peer context of adolescent substance use: Findings from social network analysis. *Journal of research on adolescence*, 16(2):159–186, 2006.
- [9] I.J. Ginsberg y J.R. Greenley. Competing theories of marijuana use: A longitudinal study. *Journal of Health and Social Behavior*, páginas 22–34, 1978.
- [10] N. Hernández. Metodología para el diseño y construcción de un lexicón de opinion mining, basado en comentarios de twitter aplicado al proyecto “opinionzoom”. Disponible en [http://wic.uchile.cl/wp-content/uploads/2016/05/Memoria\\_IN6909\\_NHernandez\\_cybertesis.pdf](http://wic.uchile.cl/wp-content/uploads/2016/05/Memoria_IN6909_NHernandez_cybertesis.pdf), 2016.
- [11] K.M. Keyes, J.E. Schulenberg, P.M. O'Malley, L.D. Johnston, J.G. Bachman, G. Li, y D. Hasin. The social norms of birth cohorts and adolescent marijuana use in the united states, 1976–2007. *Addiction*, 106(10):1790–1800, 2011.
- [12] K. Kobus y D.B. Henry. Interplay of network position and peer substance use in early adolescent cigarette, alcohol, and marijuana use. *The Journal of Early Adolescence*, 2009.

- [13] J.W. LaBrie, J.F. Hummer, y A. Lac. Comparing injunctive marijuana use norms of salient reference groups among college student marijuana users and nonusers. *Addictive behaviors*, 36(7):717–720, 2011.
- [14] G. Lee, R.L. Akers, y M.J. Borg. Social learning and structural factors in adolescent substance use. *W. Criminology Rev.*, 5:17, 2004.
- [15] E. Marrese-Taylor, J.D. Velásquez, F. Bravo-Marquez, y Y. Matsuo. Identifying customer preferences about tourism products using an aspect-based opinion mining approach. *Procedia Computer Science*, 22:182–191, 2013.
- [16] J. Petraitis, B.R. Flay, y T.Q. Miller. Reviewing theories of adolescent substance use: organizing pieces in the puzzle. *Psychological bulletin*, 117(1):67, 1995.
- [17] S.A. Stoddard, J.A. Bauermeister, D. Gordon-Messer, M. Johns, y M.A. Zimmerman. Permissive norms and young adults' alcohol and marijuana use: The role of online communities. *Journal of Studies on Alcohol and Drugs*, 73(6):968–975, 2012.
- [18] A.L. Stone, L.G. Becker, A.M. Huber, y R.F. Catalano. Review of risk and protective factors of substance use and problem use in emerging adulthood. *Addictive behaviors*, 37(7):747–775, 2012.
- [19] M. van den Bree y W.B. Pickworth. Risk factors predicting changes in marijuana involvement in teenagers. *Archives of general psychiatry*, 62(3):311–319, 2005.
- [20] M.J. Van Ryzin, G.M. Fosco, y T.J. Dishion. Family and peer predictors of substance use from early adolescence to early adulthood: An 11-year prospective analysis. *Addictive behaviors*, 37(12):1314–1324, 2012.
- [21] J.D. Velasquez, A. Bassi, H. Yasuda, y T. Aoki. Mining web data to create online navigation recommendations. In *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*, páginas 551–554. IEEE, 2004.
- [22] S.L. Wenzel, J.S. Tucker, D. Golinelli, H.D. Green, y A. Zhou. Personal network correlates of alcohol, cigarette, and marijuana use among homeless youth. *Drug and alcohol dependence*, 112(1):140–149, 2010.

---

# PROGRAMACIÓN MATEMÁTICA PARA LOS FIXTURES DE LOS TORNEOS PROFESIONALES DEL BÁSQUET DE LA ARGENTINA EN UN FORMATO NBA

---

GUILLERMO DURÁN<sup>\*</sup>SANTIAGO DURÁN<sup>\*\*</sup>JAVIER MARENCO<sup>\*\*\*</sup>FEDERICO MASCIALINO<sup>\*\*\*\*</sup>PABLO A. REY<sup>\*\*\*\*\*</sup>

## Resumen

Hasta la temporada 2013-2014 la Liga Nacional de Básquet de la Argentina se jugaba solamente durante los fines de semana. A partir de la temporada 2014-2015, y en base a una propuesta presentada por los autores de este trabajo, el torneo adquirió un formato similar al de la NBA, la liga profesional de básquet de los Estados Unidos de América. Los partidos se juegan en cualquier día de la semana y los juegos de visita se programan de manera consecutiva en giras de uno a cuatro partidos solicitadas por cada equipo, con el objetivo de reducir los kilómetros viajados y respetar las preferencias de cada equipo. De esta manera, el problema a resolver consiste en una variación no estudiada aún en la literatura del conocido *Traveling Tournament Problem*. Presentamos en este trabajo los modelos de programación matemática utilizados para armar el fixture en las temporadas 2014-2015 y 2015-2016 de la Liga. La

---

<sup>\*</sup>Departamento de Ingeniería Industrial, Universidad de Chile. Departamento de Matemática e Instituto de Cálculo, FCEyN, Universidad de Buenos Aires y CONICET, Argentina. e-mail: gduran@dm.uba.ar.

<sup>\*\*</sup>Departamento de Matemática, FCEyN, Universidad de Buenos Aires y CONICET, Argentina.

<sup>\*\*\*</sup>Instituto de Ciencias, Universidad Nacional de General Sarmiento y Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina.

<sup>\*\*\*\*</sup>Departamento de Matemática, FCEyN, Universidad de Buenos Aires, Argentina.

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

modelación consiste de dos etapas, primero se define la secuencia de los partidos de cada equipo y después un segundo modelo asigna los días en que cada partido se tiene que disputar, considerando en ambas etapas una serie de restricciones solicitadas por los equipos participantes, la televisión y la asociación que aglutina a todos los equipos. Con un esquema similar también se ha programado durante estas mismas dos temporadas la segunda división de la Liga, el Torneo Nacional de Ascenso, que es disputado actualmente por 26 equipos. La implementación de estos modelos significó un ahorro superior al 30 % en el promedio de kilómetros viajados por partido jugado como visita, con el consiguiente beneficio económico y el menor desgaste para los jugadores.

PALABRAS CLAVE: Programación en deportes, Fixture, Básquet, Programación entera

---

## 1. Introducción

---

El básquet es uno de los deportes más populares de la Argentina. La creación de la Liga Nacional en 1985, a propuesta de uno de los más famosos entrenadores del momento, León Najnudel, significó un gran impulso para este deporte en todo el país. La creación de la Liga, junto a la aparición a fines de los 90 de una camada de jugadores notables encabezada por Emanuel “Manu” Ginóbili (estrella aún vigente de los San Antonio Spurs en la NBA de América), convirtió a la Argentina en una potencia mundial de este deporte. La medalla de oro olímpica en Atenas 2004, sumada al subcampeonato del mundo en Indianápolis 2002 y a la medalla de bronce olímpica en Beijing 2008, llevaron al país a la cima del básquet mundial durante la década pasada.

Hoy la Liga Nacional de Básquet (LNB), la primera división del básquet de la Argentina, es altamente profesional y está muy expandida a lo largo y a lo ancho de un país muy extenso. A diferencia del fútbol, que es el principal deporte nacional y tiene su principal desarrollo en las grandes urbes (Buenos Aires y su radio de influencia, Rosario, Santa Fe, Córdoba), el básquet tiene un fuerte impacto regional. Por ejemplo, la final de la temporada 2014-2015 de la Liga fue disputada entre Quimsa de Santiago del Estero (que fue finalmente el campeón) y Gimnasia Indalo de Comodoro Rivadavia, mientras que los dos equipos derrotados en las finales de Conferencia (la instancia previa a la gran final de la Liga) fueron San Martín de Corrientes y Quilmes de Mar del Plata. En la temporada 2015-2016 los finalistas fueron San Lorenzo de Almagro (de Buenos Aires), a la postre campeón, y La Unión de Formosa, mientras que los otros dos finalistas de Conferencias fueron Bahía Básquet (de Bahía Blanca)

y Olímpico de Santiago del Estero. Por otra parte, en la temporada 2016-2017 de la Liga los 20 equipos provendrán de 10 provincias distintas (mientras que en la Primera División del fútbol los 30 equipos participantes son de ocho provincias diferentes, y 26 de ellos provienen de sólo cuatro distritos). Estos ejemplos dan una muestra cabal de la expansión y la federalización del básquet en la Argentina.

Los principales equipos de básquet del país tienen altos presupuestos y son apoyados en muchos casos por los gobiernos provinciales y municipales. Estos presupuestos son incluso comparables con los de algunos equipos de fútbol de Primera División. El impacto mediático del básquet argentino también es muy importante. Los periódicos nacionales cubren diariamente las noticias del básquet, mientras que son televisados semanalmente tres partidos de la Liga y un partido del Torneo Nacional de Ascenso (TNA), la segunda división.

A pesar de la alta profesionalización del deporte, hasta la temporada 2013-2014, tanto la Liga como el TNA eran programados de manera manual, con el consecuente perjuicio para los clubes, producto de una programación deficiente. Las grandes distancias a recorrer a lo largo de la temporada hacían más necesaria la aplicación de técnicas modernas de *sports scheduling*.

En los últimos 20 años diversas aplicaciones de *sports scheduling* a problemas del mundo real en diferentes deportes han sido reportadas en la literatura, incluyendo el fútbol [1, 4, 7, 12, 13, 15], el básquet [10, 18, 20], el hockey sobre hielo [6, 11] y el cricket [19]. Interesantes recopilaciones sobre *sports scheduling* han sido publicadas en los últimos años [9, 14].

Un excelente banco de pruebas para diferentes modelos, algoritmos y herramientas metodológicas en *sports scheduling* es el *Traveling Tournament Problem* (TTP) [5]. Dado un conjunto de  $n$  equipos y las distancias entre los estadios donde hace de local cada uno de ellos, el TTP consiste en armar una programación de un torneo ficticio *double round-robin* (cada equipo juega contra cada otro equipo dos veces, una de local y una de visita), con  $2(n - 1)$  rondas, de modo que ningún equipo juegue menos que  $L$  ni más que  $U$  partidos consecutivos como local ni como visita (típicamente  $L = 1$  y  $U = 3$ ), ningún equipo juegue contra otro en dos rondas consecutivas y se minimice la distancia total viajada. Se considera que ningún equipo retorna a su casa en medio de una secuencia de partidos como visita (y esta secuencia se denomina una *gira*). El TTP es un problema de optimización combinatorial muy difícil de resolver y además su complejidad computacional tampoco está totalmente resuelta. Se sabe que el problema es NP-hard para  $L = 1$  y  $U = 3$  [16] y para  $L = 1$  y  $U = \infty$  [2], pero para otros valores de  $L$  y  $U$  aún es un problema abierto. La primera aplicación del TTP a un problema real fue reportada hace unos años en el vóley argentino [3].

El resto del artículo está organizado de la siguiente manera. En la Sección 2 presentamos una descripción de los formatos de los torneos, antiguos y actuales, tanto de la Liga como del TNA, y describimos el problema que vamos a resolver. En la Sección 3 mostramos los modelos de programación entera implementados para resolver los diferentes problemas. En la Sección 4 analizamos el impacto del ahorro en viajes gracias al nuevo formato y la aplicación de nuestros modelos. Por último, en la Sección 5 exhibimos las conclusiones y el posible trabajo futuro.

---

## 2. Contexto y descripción del problema

---

La Asociación de Clubes de Básquet de la Argentina (AdC) es la institución que organiza los campeonatos de la LNB y del TNA. Hasta la temporada 2013-2014 la Liga la jugaban 16 equipos, en un formato de parejas. Disputaban primero una fase regional en dos zonas de ocho equipos cada una, jugando todos contra todos, ida y vuelta, y después una fase nacional con los 16 equipos, de nuevo todos contra todos, ida y vuelta. En total cada equipo jugaba 44 partidos a lo largo de la serie regular (fase regional más fase nacional). Para la fase nacional se acumulaba el 50 % de los puntos obtenidos en la fase regional, en una tabla de posiciones que agrupaba a los 16 equipos.

El torneo se organizaba en un formato por parejas y los partidos se disputaban solamente viernes y domingos. Cada fin de semana una pareja de equipos  $A = (A_1, A_2)$  visitaba a otra pareja  $B = (B_1, B_2)$ . Se enfrentaban el viernes  $A_1$  con  $B_2$  y  $A_2$  con  $B_1$ , invirtiendo el domingo los rivales. Había un fin de semana en la fase regional y un fin de semana en la fase nacional destinados exclusivamente a los enfrentamientos entre equipos de la misma pareja. Las parejas se armaban por cercanía geográfica, con la idea de que el equipo visitante viajara poco de un viernes a un domingo, aunque existía una condición extra de que equipos de una misma ciudad no fueran pareja para evitar que se superpusieran siempre los días de sus partidos de local. Al finalizar la serie regular los 12 mejores disputaban el título en un sistema de *play-offs* al mejor de cinco partidos (excepto la final que era al mejor de siete partidos).

No hubo descensos de la Liga al TNA en la temporada 2013-2014. El torneo empezó en octubre de 2013 y finalizó en junio de 2014, con un receso para las fiestas de Navidad y Año Nuevo.

Por otra parte, el TNA en la temporada 2013-2014 lo disputaron 20 equipos. Se jugó también en 2 fases, pero en este caso una zonal y otra regional. La fase zonal eran 4 grupos de 5 equipos cada uno, jugando todos contra todos,

ida y vuelta, a un partido por fin de semana. La fase regional estuvo compuesta por dos grupos (norte y sur) de 10 equipos cada uno, jugando todos contra todos, ida y vuelta, nuevamente a un partido por fin de semana. En total cada equipo jugó 26 partidos a lo largo de la serie regular. Para la fase regional se acumularon el 50 % de los puntos obtenidos en la fase zonal, en dos tablas de posiciones (grupo norte y grupo sur) con 10 equipos cada una. Al finalizar la serie regular los 8 mejores de cada grupo se enfrentaron en un sistema de *play-offs* al mejor de cinco partidos, en búsqueda de los dos ascensos a la Liga. Los décimos de cada grupo descendieron al Torneo Federal, la tercera división del básquet argentino.

El formato en parejas de la Liga Nacional y el de un partido por semana del TNA eran sencillos de programar de manera manual pero mostraban una misma falencia: en un país muy extenso había poco aprovechamiento de los viajes para disputar varios partidos juntos, lo que implicaba mayor número de kilómetros viajados a lo largo de la temporada y mayor desgaste para los jugadores. Por ello, y a propuesta de los autores de este trabajo, a partir de la temporada 2014-2015 se decidió cambiar el sistema de disputa y pasar a un formato similar al de la NBA, por giras, donde se pudieran aprovechar bien los viajes. La NBA, la liga profesional de básquet de los Estados Unidos, es la más importante del mundo. A diferencia de las grandes ligas de Europa, la NBA presenta las mismas características que la liga argentina en cuanto a las grandes distancias de viajes y a la gran cantidad de partidos que se disputan en la serie regular. Hasta nuestro conocimiento, no se ha reportado en la literatura cómo se efectúa la programación de la NBA.

Con este nuevo formato en la liga argentina, por ejemplo, el equipo de Formosa (ciudad situada 1200 km al norte de Buenos Aires) podría viajar a Buenos Aires y jugar tres partidos consecutivos (lunes-miércoles-viernes), antes de volver a su ciudad. Este cambio, además de provocar fuertes ahorros en viajes, implicaba abandonar las parejas y los partidos sólo en fines de semana. Las giras serían propuestas por los equipos, de modo de no exigir necesariamente las que implicaran el menor recorrido global, sino que cada equipo pudiera conjugar ahorros en viajes con decisiones de índole más deportivas. También se propuso para mejorar el atractivo del torneo programar buena parte de los partidos de mayor convocatoria al principio y al final de la serie regular.

Con el ascenso de dos equipos del TNA a la Liga en la temporada 2013-2014 y la ausencia de descensos en la Liga, la LNB pasó a tener 18 equipos en 2014-2015. El sistema de disputa fue por Conferencias (Norte y Sur), de nueve equipos cada una. Se organizaron nuevamente una fase regional y una fase nacional. En la fase regional se enfrentaron todos contra todos ida y vuelta,

repetiendo dos veces más cada “clásico” (partido entre dos equipos con alta rivalidad). Por ser impar el número de equipos en cada Conferencia hubo un clásico que cruzaba a un equipo del Norte, Atenas de Córdoba, con uno del Sur, Boca Juniors, de la ciudad de Buenos Aires. De este modo cada equipo disputaba en la fase regional 18 partidos. La fase nacional fue un todos contra todos, ida y vuelta, de los 18 equipos, lo que agregaba 34 partidos más por equipo, llegando a 52 en toda la serie regular. Las tablas de posiciones eran por Conferencia y se acumulaban todos los puntos de la fase regional y de la fase nacional. Al finalizar la serie regular los siete primeros de cada Conferencia pasaron a los cuartos de final del *play-off* de su respectiva Conferencia, mientras que el octavo y el noveno disputaron un *play-off* al mejor de tres partidos para determinar el último clasificado a cuartos de final. A partir de allí jugaron en un sistema de *play-offs* al mejor de cinco partidos hasta determinar al campeón de cada Conferencia. Los dos campeones de Conferencias se enfrentaron en la gran final de la Liga Nacional en un *play-off* al mejor de siete partidos. Nuevamente se determinó que no hubiera descensos al TNA.

Para la temporada 2015-2016 los equipos en la Liga pasaron a ser 20. El esquema fue similar al de la temporada anterior: por Conferencias, de 10 equipos cada una, y en dos fases (regional y nacional), ambas de todos contra todos, ida y vuelta. Cada equipo disputó entonces en la serie regular 56 partidos (18 en la fase regional, 38 en la fase nacional). Las tablas de posiciones fueron por Conferencia y se acumularon todos los puntos de la fase regional y de la fase nacional. Al finalizar la serie regular, los dos primeros de cada Conferencia estuvieron clasificados directamente a los *play-off* de semifinales de su respectiva Conferencia, mientras que del tercero al sexto se enfrentaron en *play-offs* de cuartos de final al mejor de cinco partidos. Del séptimo al noveno de cada Conferencia finalizaron su temporada, mientras que los décimos de cada Conferencia se enfrentaron entre sí, en un *play-off* de cinco partidos, para evitar el descenso al TNA (serie en la que Lanús, de Buenos Aires, superó a Sionista de Paraná). Las semifinales y finales de Conferencia se disputaron en *play-offs* a cinco partidos, mientras que los dos campeones de cada Conferencia se enfrentaron en un *play-off* a siete partidos, para determinar al ganador de la Liga Nacional.

En la Figura 1 podemos ver la ubicación en el mapa de la Argentina de los equipos de la Liga Nacional durante la temporada 2015-2016.

El TNA para la temporada 2014-2015 también pasó a un formato con giras. Lo disputaron 24 equipos. La fase zonal fueron seis grupos de cuatro equipos cada uno, jugando a cuádruple vuelta. La fase regional fueron dos Conferencias (Norte y Sur) de 12 equipos cada una, jugando todos contra todos, ida y vuelta, dentro de cada Conferencia. En total cada equipo jugó 34 partidos a lo largo

Figura 1: Equipos participantes de la LNB, temporada 2015-2016.



de la serie regular. Para la fase regional se acumulaba el 50% de los puntos obtenidos en la fase zonal, en dos tablas de posiciones (Conferencia Norte y Conferencia Sur) con 12 equipos cada una. Al finalizar la serie regular los dos ascensos a la Liga se definieron en un sistema de *play-offs* al mejor de cinco partidos por serie. No hubo descensos al Torneo Federal.

La temporada 2015-2016 del TNA fue disputada en un formato similar, por 26 equipos. En este caso hubo 4 zonas en la fase zonal: norte con 7 equipos, centro-norte con 6 equipos, centro-sur con 6 equipos y sur con 7 equipos. Para que en esta fase a doble vuelta todos jugaran 12 partidos se incluyó un interzonal entre equipos de la centro-norte con equipos de la centro-sur (y por ello se programaron estas 2 zonas como una única, que llamamos zona centro). La fase regional fue de nuevo en 2 Conferencias (Norte y Sur), pero ahora de 13 equipos cada una, y los equipos acumularon el 100% de los puntos obtenidos en la fase zonal. Cada equipo disputó en este caso 36 partidos en la serie regular. El único ascenso a la Liga, obtenido por Hispano Americano de Río Gallegos, se definió en un sistema de *play-offs* al mejor de cinco partidos cada serie. Nuevamente no hubo descensos al Torneo Federal.

En la Figura 2 podemos ver la ubicación en el mapa de la Argentina de los equipos del TNA durante la temporada 2015-2016.

Con la implementación del nuevo formato en las dos categorías se decidió colocar semanas completas de descanso a lo largo de la competencia. En el caso de la Liga, una de ellas es en la fase regular y dos más en la fase nacional. En el caso del TNA, es ubicada entre ambas fases. Estas semanas se usan a posteriori para reprogramaciones, producto de suspensiones de partidos ya sea por competencias internacionales cuyas fechas no se conocían al programar el fixture, u otros motivos de fuerza mayor. En general, en caso de suspenderse algún partido en el marco de una gira de visita de algún equipo, se intenta mover toda la gira, para no desperdiciar tiempos de viaje. Las reprogramaciones suelen ser sencillas por lo que se hacen de manera manual.

El objetivo principal de nuestra programación es asignar a cada equipo las giras de visita que sus directivos solicitaron, y esta medida es la que vamos a incorporar en la función objetivo del modelo que asigna la secuencia de partidos de cada equipo. Estas giras son de habitualmente dos o tres partidos consecutivos fuera de casa, antes de volver a la propia. En algún caso puede haber giras de un partido solo (cuando se viaja a un lugar cercano, o a un lugar lejano que no tenga cerca otros posibles rivales), o hasta de cuatro partidos, en algunos casos aislados. Otras restricciones usualmente empleadas tienen que ver con tener en cuenta no disponibilidades de ciertos estadios en ciertas fechas; no tener largas secuencias de partidos consecutivos de local o de visita; considerar la participación de algunos equipos en competencias internacionales.

Figura 2: Equipos participantes del TNA, temporada 2015-2016.



les para evitar superposiciones con sus partidos en el torneo local; tener en cuenta ciertos requerimientos de los canales de televisión que transmiten semanalmente los partidos; ubicar partidos atractivos en momentos estratégicos del torneo.

Dada la dificultad computacional de resolver en un único modelo el orden de los partidos y los días específicos de cada juego, se resolvió atacar el problema en dos etapas: la primera asigna la secuencia de partidos de cada equipo, y la segunda indica el día de cada juego, tomando en cuenta las preferencias de días de local que cada equipo expresa. Los modelos empleados para cada etapa son presentados en la próxima sección.

---

### 3. Enfoque de solución

---

Presentamos en esta sección el enfoque de solución a través de modelos de programación entera empleados para la asignación del orden de los partidos de cada equipo, en una primera etapa, y de asignación de días a cada juego (manteniendo el orden de juegos dado por la resolución de la primera etapa del problema), en una segunda etapa.

Los modelos fueron implementados con el lenguaje de modelado zimpl y resueltos con Cplex 12.5 en un computador con dos procesadores corriendo a 1.5 GHz y con una memoria RAM de 2 GB.

#### 3.1. Modelo para asignación de juegos

La asignación de la secuencia de juegos de cada equipo se realiza por medio de un modelo de programación entera que toma como entrada principal las giras de visita solicitadas. Como en este primer modelo no se tienen en cuenta los días del calendario en los que se juega cada partido, entonces se recurre a una aproximación para programar los partidos a lo largo del calendario. Para esto, se definen tres *fechas* por semana, de modo tal que cada equipo puede jugar hasta tres partidos por semana. Se puede considerar que estas tres fechas corresponden al lunes, miércoles y viernes de la semana en cuestión, y que entonces este primer modelo asigna juegos solamente en estos días. El segundo modelo (a ser descrito en la Sección 3.2) se encargará luego de distribuir los partidos a lo largo de la semana.

Los datos para el primer modelo son los siguientes:

- Conjunto  $E$  de equipos.
- Para cada equipo  $e \in E$ , un conjunto  $G(e)$  de giras posibles. Cada gira

es una secuencia de uno a cuatro equipos distintos de  $e$  y distintos entre sí. Definimos  $G = \cup_{e \in E} G(e)$  como el conjunto de todas las giras.

- Para cada equipo  $e \in E$ , un conjunto  $GP(e) \subseteq G(e)$  de *giras preferidas* para el equipo  $e$ . Estas giras, con su correspondiente orden estricto de partidos, son las seleccionadas por el equipo, y conforman un cubrimiento de todos los partidos de visita de ese equipo. Por su parte, las giras de  $G(e) \setminus GP(e)$  son giras agregadas por los autores para ayudar a la factibilidad del modelo. Se considera que los partidos en una gira se juegan en fechas consecutivas.
- Cantidad  $n$  total de fechas. Definimos además  $F = \{1, \dots, n\}$  como el conjunto de todas las fechas.
- Conjunto  $C$  de *fechas de corte*. Una gira no puede comenzar antes y terminar después de una fecha de corte. Este conjunto (definido por la AdC) está originado por la existencia de semanas de descanso dentro del calendario, y entonces no se debe comenzar una gira si no hay tiempo suficiente antes de una semana de descanso. En el contexto del modelo, esto se traduce en estas fechas de corte, y suponemos que corresponden al viernes inmediatamente anterior a una semana de descanso.

Es importante mencionar que, con estos datos, los equipos no jugarán en todas las fechas dado que al considerar tres fechas por semana, la cantidad de fechas es aproximadamente un 50% superior a la cantidad de partidos que cada equipo debe jugar (se planifica el torneo para que cada equipo dispute en promedio 2 partidos por semana). Por este motivo, el modelo debe contemplar la existencia de una cierta cantidad de *byes* (fechas donde el equipo está libre, o sea, sin partido) para cada equipo. Esta característica, junto con la existencia de giras predefinidas, hace que el problema modelado corresponda a una variación no estudiada en la literatura del TTP. El objetivo será maximizar el número de partidos asignados al fixture que pertenezcan a giras de visita elegidas por los equipos.

Para la formulación del modelo, introducimos una variable binaria  $z_{tk}$  para cada gira  $t \in G$  y cada fecha  $k \in F$  con  $k \leq n - |t| + 1$  (siendo  $|t|$  la cantidad de partidos en la gira  $t$ ), de modo tal que  $z_{tk} = 1$  si y sólo si la gira  $t$  comienza a partir de la fecha  $k$ . Además, introducimos la variable auxiliar  $x_{ijk}$  para cada par de equipos  $i, j \in E$ ,  $i \neq j$  y cada fecha  $k \in F$ , de modo tal que  $x_{ijk} = 1$  si y sólo si el equipo  $i$  juega de local contra el equipo  $j$  en la fecha  $k$ . Aunque estas últimas variables se pueden definir a partir de las variables  $z$  (y, por lo tanto, no son indispensables para el modelo), su introducción permite simplificar la

formulación del modelo. Con estos elementos, podemos plantear el siguiente modelo para el primer problema:

1. La función objetivo busca maximizar la cantidad de partidos seleccionados que provengan de giras preferidas:

$$\text{máx} \sum_{e \in E} \sum_{t \in GP(e)} \sum_{k \in F} |t| z_{tk}.$$

2. Se juegan todos los partidos:

$$\sum_{k \in F} x_{ijk} = 1 \quad \forall i, j \in E, i \neq j. \quad (1)$$

3. Cada equipo juega a lo sumo un partido por fecha:

$$\sum_{j \in E} x_{ijk} + x_{jik} \leq 1 \quad \forall i \in E, \forall k \in F. \quad (2)$$

4. Con la definición, de la variable  $x_{ijk}$  dada más arriba, ésta debe tomar el valor 1 si y sólo si el equipo  $j$  realiza una gira que incluye al equipo  $i$  en la fecha  $k$ . Dado un equipo  $i \in E$  y una gira  $t \in G$  que incluye al equipo  $i$ , definimos  $\text{pos}(t, i) \in \{0, \dots, |t| - 1\}$  como la posición del equipo  $i$  dentro de la secuencia de partidos dada por  $t$ :

$$x_{ijk} = \sum_{t \in G(j)} z_{t, k - \text{pos}(t, i)} \quad \forall i, j \in E, i \neq j, \forall k \in F. \quad (3)$$

5. Se juegan al menos dos partidos de local dentro de las cuatro fechas posteriores a cada gira, para no tener secuencias de partidos de visita muy largas para un equipo dado:

$$\sum_{j \in E} \sum_{s=0}^4 x_{ij, k+|t|+s} \geq 2z_{tk} \quad \forall i \in E, \forall t \in G(i), \quad (4)$$

$$\forall k \in F, k + |t| + 4 \leq n.$$

6. Ninguna gira puede comenzar antes y terminar después de una fecha de corte:

$$z_{tk} = 0 \quad \forall c \in C, \forall t \in G, \forall k \in F, k \leq c < k + |t| - 1. \quad (5)$$

7. Se tiene un *bye* antes o después de cada gira, para favorecer el descanso de los jugadores:

$$\sum_{j \in E} (x_{ij, k+|t|} + x_{ji, k+|t|}) + \sum_{j \in E} (x_{ij, k-1} + x_{ji, k-1}) \leq 2 - z_{tk}$$

$$\forall i \in E, \forall t \in G(i), \forall k \in F, 1 < k \leq n - |t|. \quad (6)$$

8. No se deben tener más de dos *byes* seguidos, para no estar largos tiempos sin jugar:

$$\sum_{j \in E} \sum_{s=0}^2 x_{ij,k+s} + x_{ji,k+s} \geq 1 \quad \forall i \in E, \forall k \in F, k+2 \leq n. \quad (7)$$

9. Se debe tener al menos un partido de visita cada  $\ell$  fechas. Esta restricción está originada en la necesidad de no tener demasiados partidos seguidos de local, para no saturar a los aficionados. Se trabajó habitualmente con  $\ell = 6$ :

$$\sum_{j \in E} \sum_{s=0}^{\ell-1} x_{ji,k+s} \geq 1 \quad \forall i \in E, \forall k \in F, k + \ell - 1 \leq n. \quad (8)$$

10. Naturaleza de las variables:

$$z_{tk} \in \{0, 1\} \quad \forall t \in G, \forall k \in F, \quad (9)$$

$$x_{ijk} \in \{0, 1\} \quad \forall i, j \in E, i \neq j, \forall k \in F. \quad (10)$$

Se podría utilizar una función objetivo alternativa que represente la distancia total recorrida por los equipos (multiplicando cada variable  $z_{tk}$  por la distancia total de la gira  $t$ ), y que penalice las distancias de las giras no preferidas (por ejemplo, duplicando la distancia de estas giras). Esta posibilidad fue explorada para las instancias de la temporada 2014-2015, y los resultados fueron muy similares a los obtenidos con la función objetivo presentada aquí. Esto se debe a que las giras propuestas por los equipos ya representan buenos recorridos en términos de las distancias totales. Dado el comportamiento similar de ambos modelos preferimos la versión explicada en este trabajo que nos permite independizarnos de las distancias de viajes.

La estructura del problema hace que el mismo sea muy difícil de resolver para instancias de más de 13 equipos, como muestra la Tabla 1. Esta conclusión es acorde con lo conocido para el TTP donde existen aún instancias abiertas para  $n = 12$  [17].

En la primera parte de la Tabla 1 (primeras 13 filas) se reportan todas las instancias consideradas en la sección anterior (excepto la fase zonal del TNA temporada 2014-2015 que al ser grupos de 4 equipos se resuelven en todos los casos en menos de 1 segundo), mostrando sus respectivos tamaños, la cantidad de variables y restricciones del modelo para estos casos, el tiempo de resolución en segundos y el *gap* de optimalidad final.

Puede verse que el modelo permite resolver adecuadamente las instancias zonales y regionales de ambos campeonatos alcanzando optimalidad en todos

ellos excepto en un caso del TNA donde existe un *gap* de 2,85%. En cambio, no puede resolver las fases nacionales de la Liga, en la que participan todos los equipos durante una cantidad prolongada de fechas. En estos casos, el *solver* no puede encontrar siquiera soluciones factibles luego de varias horas de ejecución, y entonces es necesario recurrir a técnicas alternativas para hallar una solución satisfactoria, como se describe a continuación.

La existencia de semanas de descanso permite dividir el diseño del fixture en *etapas* separadas (aunque no independientes). Claramente no puede haber giras que comiencen antes y terminen después de cada fecha de corte y los partidos programados en una determinada etapa ya no deben considerarse como pendientes para etapas posteriores. Sea  $C = \{f_1, \dots, f_p\}$  el conjunto de fechas de corte, y definimos además  $f_0 = 0$  y  $f_{p+1} = n$ . Las consideraciones anteriores sugieren el siguiente algoritmo heurístico, que se utilizó en este trabajo para las instancias más grandes y permitió obtener soluciones satisfactorias en todos los casos:

1. Dividir el conjunto  $F$  de fechas en subconjuntos  $F_0, \dots, F_p$ , de modo tal que  $F_q = \{f_q + 1, f_q + 2, \dots, f_{q+1}\}$ , para  $q = 0, \dots, p$ . Es decir, las fechas de  $F_q$  son consecutivas y cubren un intervalo entre dos fechas de corte consecutivas, correspondiendo así a la  $(q + 1)$ -ésima etapa del campeonato.
2. Definir  $P := \emptyset$ . Este conjunto representa los partidos ya jugados en las etapas anteriores.
3. Para  $q$  desde 0 hasta  $p$ , realizar los siguientes pasos:
  - a) Resolver el modelo (2)-(10) donde  $F$  tiene ahora  $\lfloor \alpha |F_q| \rfloor$  fechas (siendo  $0 < \alpha < 1$ , el  $\alpha$  lo elegimos de modo de balancear los *byes* de cada equipo a lo largo del torneo), reemplazando la restricción (1) por

$$\sum_{k \in F} x_{ijk} \leq 1 \quad \forall i, j \in E, i \neq j,$$

(es decir, ahora no se deben jugar todos los partidos), eliminando las giras que incluyen partidos de  $P$ , y reemplazando la función objetivo por la siguiente expresión:

$$\text{máx} \sum_{e \in E} \left[ \sum_{t \in GP(e)} \sum_{k \in F} 2|t|z_{tk} + \sum_{t \in G(e) \setminus GP(e)} \sum_{k \in F} |t|z_{tk} \right].$$

Es decir, se busca maximizar la cantidad de partidos seleccionados para el intervalo de fechas, contando “doble” a los partidos que participan en las giras seleccionadas por los equipos.

- b) Agregar a  $P$  los partidos jugados en la solución del paso anterior.
4. Combinar los fixtures generados para cada etapa y retornar la solución obtenida.

Este procedimiento busca diseñar el fixture por etapas, tomando las fechas de corte como divisiones de cada una de ellas. El modelo que se resuelve para cada etapa (paso 3a en el algoritmo) busca maximizar el número de partidos incluidos en la misma, con el objetivo de facilitar la tarea de las etapas siguientes. Para  $q = 0, \dots, p$ , en la etapa  $q$  no se consideran las  $|F_q|$  fechas de la etapa, sino que se resuelve el modelo con  $\lfloor \alpha |F_q| \rfloor$  fechas, siendo  $0 < \alpha < 1$  un parámetro del algoritmo (nos ha dado buenos resultados fijar ese  $\alpha$  en 0,8 o 0,9). Esta modificación permite que la maximización de la cantidad de partidos en la etapa no genere demasiados partidos en las primeras etapas, y en consecuencia demasiados *byes* en las últimas etapas.

El resultado final de este algoritmo puede no incluir a todos los partidos, y en ese caso se podría iterar el procedimiento repitiendo los pasos de las etapas anteriores con restricciones adicionales que soliciten una solución distinta a las obtenidas en iteraciones anteriores para esa etapa. Esto genera un algoritmo de tipo *backtracking* hasta que el fixture completo contenga todos los partidos. Para las instancias consideradas en este trabajo, el resultado final contuvo a todos los partidos y entonces no fue necesario recurrir a esta técnica.

Las últimas seis filas de la Tabla 1 muestran el comportamiento del modelo resuelto en el paso 3a del algoritmo sobre los dos campeonatos de la Liga Nacional considerados en este trabajo. Se alcanzó optimalidad dentro del tiempo límite de una hora seleccionado en todos los casos excepto en una de las fases nacionales de la Liga. Los fixtures obtenidos resultaron altamente satisfactorios en todos los casos. Más del 90 % de los partidos de visita programados en los fixtures finales de todas las instancias resueltas corresponden a partidos de visita pertenecientes a giras preferidas por los equipos.

### 3.2. Modelo para asignación de días

Una vez obtenida una solución mediante el modelo de la sección anterior, resta asignar los días del calendario en que se llevará a cabo cada uno de los juegos. Para ello se utiliza el modelo que formulamos en esta subsección. Hasta nuestro conocimiento no hay en la literatura de *sports scheduling* utilización de modelos matemáticos para asignar los días de los partidos en una competencia deportiva.

Instancia	$ E $	$ G $	$ F $	Vars.	Restr.	Tiempo	Gap
2014/2015 TNA Sur	12	125	36	9684	10535	298,30	0,00 %
2014/2015 TNA Norte	12	131	36	9900	10732	298,74	0,00 %
2014/2015 LNB Sur	9	*312	21	7617	5586	61,99	0,00 %
2014/2015 LNB Norte	9	*312	21	6457	4493	15,44	0,00 %
2014/2015 LNB	18	630	56	43884	32706	—	—
2015/2016 TNA Sur	7	*248	17	4981	2056	59,48	0,00 %
2015/2016 TNA Centro	12	*248	17	6630	4857	27,06	0,00 %
2015/2016 TNA Norte	7	*248	17	5015	2737	208,54	0,00 %
2015/2016 TNA Reg. Sur	13	160	40	13160	14372	3600	2,85 %
2015/2016 TNA Reg. Norte	13	270	40	17560	23622	52,69	0,00 %
2015/2016 LNB Sur	10	*1543	24	39342	8486	9,43	0,00 %
2015/2016 LNB Norte	10	*1543	24	35904	8486	480,20	0,00 %
2015/2016 LNB	20	1543	53	102979	36404	—	—
2014/2015 LNB 1/3	18	*630	14	13258	28299	24,45	0,00 %
2014/2015 LNB 2/3	18	*630	16	15264	24889	320,74	0,00 %
2014/2015 LNB 3/3	18	*630	19	18126	31317	1072,07	0,00 %
2015/2016 LNB 1/3	20	*1543	23	44689	21505	1767,01	0,00 %
2015/2016 LNB 2/3	20	*1543	12	23304	9733	3600	5,98 %
2015/2016 LNB 3/3	20	*1543	12	23304	8652	3349,63	0,00 %

Tabla 1: Características y tiempos de resolución del primer modelo para las instancias consideradas. Los tiempos están expresados en segundos, y para las instancias marcadas con “—” no se pudo encontrar solución factible luego de una hora de ejecución. Las cantidades de giras marcadas con asteriscos corresponden al total de las giras para todos los equipos, no sólo a las giras para la zona en consideración.

Este segundo modelo recibe como entrada el orden de los juegos de cada equipo a lo largo del torneo, los días designados para el desarrollo del mismo y una lista de preferencias para cada equipo con días del calendario en que prefieren jugar de local.

Quedan entonces determinados los siguientes conjuntos:

- Conjunto  $E$  descrito en la sección anterior.
- Conjunto  $D = \{1, \dots, m\}$ , los días calendario tomados de corrido a partir de un día 1 fijado previamente.
- Cantidad  $l$  total de partidos que juega cada equipo, tomando el máximo en caso de que no coincida la cantidad de partidos que juegan los distintos equipos. Notar que si bien en un torneo completo la cantidad de partidos

que juega cada equipo es la misma, hay casos en los que se hace necesario considerar que esto no sucede, como explicaremos más adelante.

- Conjunto  $J = \{1, \dots, l\}$  de índices de la secuencia de los juegos a ser disputados por cada equipo, de acuerdo al orden determinado por el modelo anterior.

La aplicación del modelo se hará para los torneos completos cuando el modelo inicial haya resuelto todo el torneo junto, o para cada bloque, cuando así lo haya hecho el primer modelo. Claramente la solución de un bloque no afecta a la solución de los demás.

Para aplicar el modelo en un determinado bloque, cuando no se exige que la cantidad de partidos jugados por cada equipo sea la misma, se consideran partidos ficticios para trabajar con un único valor  $l$ .

Se introduce la variable binaria  $x_{ipt}$  para cada equipo  $i \in E$ , cada juego  $p \in J$  y cada día  $t \in D$ , la cual cumple que  $x_{ipt} = 1$  si y sólo si el equipo  $i$  juega su juego  $p$  el día  $t$ . Introducimos además las variables binarias auxiliares  $da_{ipt}$  y  $db_{ipt}$ . Estas variables son necesarias debido a que no siempre es posible cumplir con las restricciones pedidas por la AdC de tener dos días de descanso antes y después de cada gira. De este modo,  $db_{ipt} = 1$  si y sólo si el equipo  $i$  tiene un solo día de descanso antes de  $t$ , día en el que juega su primer partido de gira  $p$ ; análogamente queda definida la variable  $da_{ipt}$ , para el caso de un solo día de descanso posterior a la gira  $p$ . En la función objetivo hay una penalización cada vez que alguna de estas variables vale 1.

Teniendo en cuenta las preferencias de días de cada equipo, se define un parámetro  $pref(i, t)$  para cada equipo  $i \in E$  y para cada día  $t \in D$ , que vale 1 si el equipo  $i$  prefiere jugar de local el día  $t$  y 0, en caso contrario. Por último, se introduce un parámetro  $gira(i, p)$  para cada equipo  $i \in E$  y cada juego  $p \in J$ , que cumple que  $gira(i, p) = 1$  si y sólo si el equipo  $i$  juega su juego  $p$  en una gira y 0, en caso contrario.

De esta manera, formulamos el modelo para esta segunda etapa:

1. La función objetivo busca maximizar la cantidad de juegos desarrollados en días preferidos por el equipo local, penalizando además la cantidad de veces que los equipos empiezan o terminan giras sin la cantidad de días de descanso pedidos. El valor 1000 para el coeficiente de penalización fue definido de manera de lograr el efecto deseado sin provocar problemas numéricos en la resolución computacional del modelo.

$$\max \sum_{i \in E, p \in J, t \in D} x_{ipt} \cdot pref(i, t) - 1000 \cdot \sum_{i \in E, p \in J, t \in D} (db_{ipt} + da_{ipt}).$$

2. No se invierte el orden de los partidos:

$$\begin{aligned} x_{ipt} + x_{iqs} &\leq 1 && \forall i \in E, \forall p, q \in J, \\ &&& \forall s, t \in D, p < q, s \leq t. \end{aligned} \quad (11)$$

3. Se juegan todos los partidos:

$$\sum_{t \in D} x_{ipt} = 1 \quad \forall i \in E, \forall p \in J. \quad (12)$$

4. Ningún equipo juega dos partidos en días consecutivos:

$$\begin{aligned} x_{ipt} + x_{i,p+1,t+1} &\leq 1 && \forall i \in E, \forall p \in J, p < |J| \\ &&& \forall t \in D, t < |D|. \end{aligned} \quad (13)$$

5. Se estipula un día de descanso entre dos partidos consecutivos de una misma gira:

$$\begin{aligned} x_{ipt} &= x_{i,p+1,t+2} && \forall i \in E, \forall p \in J, p < |J| \\ &&& \forall t \in D, t < |D| - 1, \end{aligned} \quad (14)$$

con  $p$  y  $p + 1$  dos partidos de la misma gira.

6. Se intenta dejar al menos dos días de descanso antes y después de cada gira:

a) si  $p$  es el primer partido de la gira,

$$\begin{aligned} x_{i,p-1,t-2} + x_{ipt} &\leq 1 + db_{ipt} && \forall i \in E, \forall p \in J, p > 1 \\ &&& \forall t \in D, t > 2. \end{aligned} \quad (15)$$

b) si  $p$  es el último partido de la gira:

$$\begin{aligned} x_{ipt} + x_{i,p+1,t+2} &\leq 1 + da_{ipt} && \forall i \in E, \forall p \in J, p < |J| \\ &&& \forall t \in D, t < |D| - 1. \end{aligned} \quad (16)$$

7. Se asigna el mismo día a ambos equipos para los partidos en que se enfrentan entre ellos:

$$x_{ipt} = x_{jqt} \quad \forall i, j \in E, \forall t \in D$$

si el partido  $p$  para  $i$  y el partido  $q$  para  $j$  corresponden al partido en que se enfrentan  $i$  y  $j$ .

## 8. Naturaleza de las variables:

$$x_{ipt} \in \{0, 1\} \quad \forall i \in E, \forall p \in J, \forall t \in D. \quad (17)$$

$$da_{ipt} \in \{0, 1\} \quad \forall i \in E, \forall p \in J, \forall t \in D \quad (18)$$

$$db_{ipt} \in \{0, 1\} \quad \forall i \in E, \forall p \in J, \forall t \in D. \quad (19)$$

A diferencia de lo ocurrido con el modelo de la sección anterior, en este caso la estructura del problema no le impide al *solver* encontrar una solución óptima en pocos segundos para todas las instancias consideradas. Por ejemplo, para la Fase Regional del TNA Norte en la temporada 2015-2016, uno de los casos más grandes resueltos, el modelo tiene 21168 variables y 161284 restricciones, demanda un tiempo de resolución de 6,32 segundos y encuentra una solución óptima. Cabe aclarar que en este caso la variable  $da_{ipt}$  toma el valor 1 en dos ocasiones y la variable  $db_{ipt}$  lo hace en una, esto nos muestra que no existe una solución factible que respete en todos los casos los días de descanso pedidos antes y después de cada gira.

---

## 4. Impacto

---

Los modelos aquí presentados fueron utilizados para armar las programaciones de las temporadas 2014-2015 y 2015-2016, tanto de la Liga Nacional, como del TNA, en sus fases regionales y nacionales. Mostramos a continuación un análisis de los kilómetros viajados por los diferentes equipos con el formato actual (y nuestra programación), y con el formato anterior de parejas de los torneos. Veremos que aunque la función objetivo del modelo inicial no minimiza los kilómetros viajados sino que maximiza los partidos de visita seleccionados en la programación provenientes de giras preferidas, por la forma en que estas giras fueron diseñadas se consigue una fuerte disminución en los kilómetros recorridos por cada equipo.

La Tabla 2 compara para el caso de la Liga Nacional las temporadas 2013-2014 (la última disputada con el formato anterior y programación manual), con la 2014-2015 y la 2015-2016, ambas programadas por nosotros y con el sistema de giras. Como puede observarse, en la temporada 2014-2015 la totalidad de los 16 equipos que participaron en ambas temporadas viajó menos kilómetros que en la temporada 2013-2014, a pesar de que todos disputaron 4 partidos más de visita. El número de kilómetros recorridos de manera global por todos los equipos es un 8 % menor (353.365 contra 385.810), a pesar de que en la temporada 2014-2015 no sólo se juegan más partidos de visita sino que

también hay 2 equipos más. Si calculamos el promedio de kilómetros viajados por equipo, por cada partido de visita, dicho valor en la temporada 2014-2015 es un 31 % más bajo que en la 2013-2014 (755 km contra 1096 km).

En lo que se refiere a la comparación con la temporada 2015-2016, los resultados son similares. 13 de los 15 equipos que se repiten en esta última temporada y en la 2013-2014, viajan menos en la 2015-2016, a pesar de que tienen 6 partidos más de visita cada uno (en sólo 2 casos, Argentino y Libertad, se invierte este resultado, pero con valores que dan un virtual empate entre ambas temporadas). En lo que hace al promedio de kilómetros viajados por equipo por cada partido de visita, dicho valor en la temporada 2015-2016 es casi un 30 % más bajo que en la 2013-2014 (772 km contra 1096 km).

Equipo	Temporada 2013-2014			Temporada 2014-2015			Temporada 2015-2016		
	km Viajados	Partidos	Promedio	km Viajados	Partidos	Promedio	km Viajados	Partidos	Promedio
Argentino	18561	22	843,68	15255	26	586,73	20014	28	714,79
Ciclista	-	-	-	16826	26	647,15	-	-	-
Quilmes	23546	22	1070,27	19335	26	743,65	22764	28	813,00
Peñarol	25455	22	1157,05	19485	26	749,42	22764	28	813,00
Bahía Basket	28347	22	1288,50	17426	26	670,23	20372	28	727,57
Gimnasia Indalo	47684	22	2167,45	37218	26	1431,46	43215	28	1543,39
Boca Juniors	20278	22	921,73	15632	26	601,23	18332	28	654,71
Obras Sanitarias	20278	22	921,73	17360	26	667,69	18332	28	654,71
Lanús	20278	22	921,73	18814	26	723,62	18814	28	671,93
La Unión	27069	22	1230,41	22013	26	846,65	26362	28	941,50
Regatas	24362	22	1107,36	20121	26	773,88	21626	28	772,36
San Martín	-	-	-	20484	26	787,85	22752	28	812,57
Quimsa	24380	22	1108,18	19367	26	744,88	20722	28	740,07
Olimpico	24305	22	1104,77	19649	26	755,73	21397	28	764,18
Atenas	22087	22	1003,95	22008	26	846,46	21790	28	778,21
Libertad	18401	22	836,41	18127	26	697,19	18685	28	667,32
Sionista	19648	22	893,09	15419	26	593,04	17967	28	641,68
Estudiantes	21131	22	960,50	18826	26	724,08	18818	28	672,07
Instituto	-	-	-	-	-	-	20390	28	728,21
Ferro	-	-	-	-	-	-	18992	28	678,29
San Lorenzo	-	-	-	-	-	-	18332	28	654,71
Total	385810		1096,05	353365		755,05	432440		772,21

Tabla 2: Comparación de kilómetros viajados por los equipos de la LNB durante las temporadas 2013-2014, 2014-2015 y 2015-2016.

En la Tabla 3 mostramos una comparación entre lo viajado en la temporada 2015-2016 en la Liga, programada con nuestros modelos, contra lo que se hubiera viajado de haberse aplicado en esta última temporada el anterior formato de parejas (con parejas armadas con el mismo criterio con el que se armaban antes). Allí se puede apreciar una reducción del 22 % a favor del nuevo sistema (772 km contra 991 km), en el promedio de viaje por equipo, por partido de visita.

En lo que hace al TNA, hacemos las comparaciones de lo viajado en la realidad en las temporadas 2014-2015 y 2015-2016 contra lo que hubieran viajado de haberse disputado los torneos con el formato antiguo que tenía este campeonato, de un partido por semana y sin giras de visita. Notar que no

Equipo	Temporada 2015-2016					
	Formato Parejas			Formato Usado (con giras)		
	km Viajados	Cantidad de Partidos	Promedio	km Viajados	Cantidad de Partidos	Promedio
Argentino	24613	28	879,04	20014	28	714,79
Quilmes	29026	28	1036,64	22764	28	813,00
Peñarol	29026	28	1036,64	22764	28	813,00
Bahía Basket	31451	28	1123,25	20372	28	727,57
Gimnasia Indalo	59947	28	2140,96	43215	28	1543,39
Boca Juniors	22622	28	807,93	18332	28	654,71
Obras Sanitarias	22622	28	807,93	18332	28	654,71
Lanús	22622	28	807,93	18814	28	671,93
La Unión	33447	28	1194,54	26362	28	941,50
Regatas	24362	28	870,07	21626	28	772,36
San Martín	28293	28	1010,46	22752	28	812,57
Quimsa	29129	28	1040,32	20722	28	740,07
Olimpico	29129	28	1040,32	21397	28	764,18
Atenas	25807	28	921,68	21790	28	778,21
Libertad	23170	28	827,50	18685	28	667,32
Sionista	23365	28	834,46	17967	28	641,68
Estudiantes	25247	28	901,68	18818	28	672,07
Instituto	25807	28	921,68	20390	28	728,21
Ferro	22622	28	807,93	18992	28	678,29
San Lorenzo	22622	28	807,93	18332	28	654,71
Total	554929		990,94	432440		772,21

Tabla 3: Comparación entre los kilómetros viajados en el formato de giras utilizado y el formato de parejas alternativo, para la temporada 2015-2016.

comparamos con la última temporada disputada con el formato anterior (la 2013-2014) debido a que la composición de los equipos participantes varió mucho (hubo varios equipos que se retiraron y fueron reemplazados por equipos nuevos), por lo que la comparación no tendría sentido.

En la Tabla 4 hacemos la comparación para la temporada 2014-2015. Para el formato anterior vemos que hay 3 partidos menos de visita por equipo dado que en el esquema previo ambas fases eran a doble vuelta (y en el caso real la fase zonal fue a cuádruple vuelta). Como puede apreciarse hay una reducción global de los kilómetros viajados por todos los equipos de casi un 30 % a favor del nuevo formato, mientras que en el caso del promedio por partido de visita la reducción supera el 42 %.

En la Tabla 5 hacemos la comparación para la temporada 2015-2016. Vemos una reducción de los kilómetros totales viajados y del promedio de kilómetros viajados por partido de visita superior a un 38 %, siempre a favor del nuevo sistema.

El caso más significativo de ahorro de viajes en el TNA es el de Hispano Americano de Río Gallegos, equipo con sede en una de las ciudades más australes del país y del mundo. Este equipo, por ejemplo, debe viajar más de 2500 km para enfrentar a Platense en la ciudad de Buenos Aires. Como puede apreciarse en las Tablas 3 y 4, Hispano Americano redujo sus viajes en un 48 % en la temporada 2014-2015 y un 53 % en la temporada 2015-2016, en ambos

Equipo	Temporada 2014-2015					
	Formato anterior (sin giras)			Formato usado (con giras)		
	km Viajados	Cantidad de partidos	Promedio	km Viajados	Cantidad de partidos	Promedio
9 de Julio	20522	14	1465,86	13711	17	806,53
Anzorena	26840	14	1917,14	20159	17	1185,82
Alianza San Luis	20804	14	1486,00	16180	17	951,76
Banda Norte	18306	14	1307,57	14029	17	825,24
Deportivo Viedma	25880	14	1848,57	18470	17	1086,47
Estudiantes de Olavarría	19974	14	1426,71	13705	17	806,18
Ferro	20264	14	1447,43	20449	17	1202,88
Hispano Americano	58918	14	4208,43	36915	17	2171,47
Huracán de Trelew	31006	14	2214,71	21385	17	1257,94
Monte Hermoso	25342	14	1810,14	17756	17	1044,47
San Lorenzo	20264	14	1447,43	15400	17	905,88
Sport Club Cañadense	21364	14	1526,00	15712	17	924,24
<hr/>						
Norte	km Viajados	Cantidad de Partidos	Promedio	km Viajados	Cantidad de Partidos	Promedio
Tomás de Rocamora	13308	14	950,57	8083	17	475,47
Atlético Echague	9748	14	696,29	7099	17	417,59
Instituto de Córdoba	13186	14	941,86	9092	17	534,82
Barrio Parque	13186	14	941,86	9092	17	534,82
La Unión de Colón	13466	14	961,86	8690	17	511,18
Oberá TC	22160	14	1582,86	12797	17	752,76
Club Atlético San Isidro	9944	14	710,29	7163	17	526,12
UNCAus	14736	14	1052,57	8831	17	519,47
Unión de Santa Fe	9740	14	695,71	6586	17	387,41
Villa Ángela Basket	14562	14	1040,14	8877	17	522,18
Sarmiento	14534	14	1038,14	9905	17	582,65
Tiro Federal Morteros	10846	14	774,71	8944	17	421,35
Total (ambas Conferencias)	468900		1395,54	329030		860,45

Tabla 4: Comparación entre los kilómetros viajados por los equipos del TNA en la temporada 2014-2015 según el formato anterior y el actual.

casos en el promedio de kilómetros recorridos por partido jugado de visita. Cabe destacar que este equipo se consagró campeón del TNA en esta última temporada por lo que jugará la Liga Nacional en la temporada 2016-2017.

Como puede apreciarse los ahorros son aún más significativos para el TNA, debido a que el formato de un partido por semana era más perjudicial con respecto a ahorro de viajes que el formato de parejas que utilizaba la Liga.

Si convertimos los ahorros de kilómetros a platas ahorradas, considerando un costo estimado de 2 dólares por kilómetro viajado (que es el costo de viaje en bus), el ahorro global para la Liga Nacional en la temporada 2015-2016 es de alrededor de US\$ 245.000 (122.489 km, ver Tabla 3), mientras que el ahorro global para el TNA en la misma temporada es de más de US\$ 486.000 (243.340 km, ver Tabla 5). Esto da un ahorro para toda la temporada de ambas divisiones sumadas de una cifra cercana a los US\$ 731.000. Esta cifra es un piso para los ahorros obtenidos dado que estamos tomando el costo de los viajes en bus como base para la estimación y hay varios viajes que los equipos hacen en avión, que es un medio de transporte más caro.

Equipo	Temporada 2015-2016					
	Formato anterior (sin giras)			Formato Usado (con giras)		
Sur	km Viajados	Cantidad de Partidos	Promedio	km Viajados	Cantidad de Partidos	Promedio
Deportivo Viedina	24956	18	1386,44	17255	18	958,61
Estudiantes de Olavarría	22202	18	1233,44	13163	18	731,28
Hispano Americano	71598	18	3977,67	33198	18	1844,33
Huracán de Trelew	34382	18	1910,11	21751	18	1208,39
Monte Hermoso	25484	18	1415,78	15241	18	846,72
Ciclista	22896	18	1272,00	13776	18	765,33
Gimnasia de La Plata	23722	18	1317,89	16804	18	933,56
Platense	22770	18	1265,00	15576	18	865,33
Atenas de Carmen de Patagones	24956	18	1386,44	16722	18	929,00
Tomás de Rocamora	26670	18	1481,67	15516	18	862,00
Parque Sur	27888	18	1549,33	17235	18	957,50
Olimpo	23416	18	1300,89	15880	18	882,22
Petrolero	37048	18	2058,22	21432	18	1190,67
<hr/>						
Norte	km Viajados	Cantidad de Partidos	Promedio	km Viajados	Cantidad de Partidos	Promedio
Atlético Echagüe	14628	18	812,67	11097	18	616,50
Barrio Parque	21118	18	1173,22	12924	18	718,00
La Unión de Colón	20240	18	1124,44	12624	18	701,33
Oberá TC	26950	18	1497,22	15341	18	852,28
Club Atlético San Isidro	15236	18	846,44	11735	18	651,94
Unión de Santa Fe	14942	18	830,11	10414	18	578,56
Villa Ángela Basket	17624	18	979,11	12515	18	695,28
Sarmiento	15664	18	870,22	11683	18	649,06
Tiro Federal Morteros	16696	18	927,56	12071	18	670,61
Comunicaciones	18966	18	1053,67	12442	18	691,22
Hindú	15664	18	870,22	10100	18	561,11
Salta Basket	34616	18	1923,11	16923	18	940,17
UNCAus	16532	18	918,44	10106	18	561,44
Total (ambas Conferencias)	636864		1360,82	393524		840,86

Tabla 5: Comparación entre los kilómetros viajados por los equipos del TNA en la temporada 2015-2016 según el formato anterior y el actual.

Como se desprende del análisis realizado la principal causa de los ahorros es el haber pasado a un formato de torneos con giras, al estilo de lo que se hace en la NBA. Claramente, este nuevo sistema no podría haberse implementado sin la ayuda para su programación de las herramientas matemático-computacionales aquí presentadas (mientras que los formatos utilizados anteriormente se programaban fácilmente de manera manual).

---

## 5. Conclusiones y trabajo futuro

---

Presentamos en este trabajo los modelos de programación matemática utilizados para armar el fixture en las últimas 2 temporadas de la Liga Nacional y el Torneo Nacional de Ascenso, Primera y Segunda División del básquet profesional de la Argentina. La modelación consiste en ambos casos de 2 etapas, el primer modelo permite armar el orden de los partidos de cada equipo, para lo cual se resuelve una nueva variación del TTP, mientras que el segundo modelo asigna los días en que cada partido se tiene que disputar, considerando en ambos modelos una serie de restricciones solicitadas por los equipos

participantes, la televisión y la AdC. Se priorizaron dentro de las condiciones establecidas las cuestiones que tienen que ver con ahorros en viajes, así como también el mejorar el atractivo de los torneos para el público, a través de la ubicación de los partidos más importantes en fechas estratégicas dentro de la programación y de ciertos retoques en el sistema de disputa que llevaron a que hasta último momento los equipos participantes tuvieran algo importante en disputa.

Hasta nuestro conocimiento es la primera vez que se reporta en la literatura de *sports scheduling* la programación de un campeonato con un formato similar al de la NBA.

La implementación de estos modelos significó ahorros superiores al 30 % en el promedio de kilómetros viajados por partido jugado como visita, con el consiguiente beneficio económico (superior a los 700.000 US\$ por temporada) y el menor desgaste para los jugadores. Hay casos donde los ahorros fueron muy superiores al promedio, debido a la situación geográfica del equipo en cuestión y a la decisión dirigenal de aprovechar al máximo el sistema de giras. Uno de tales casos es el de Hispano Americano mencionado en la sección anterior. Otro es el de Bahía B́asquet, en la Liga Nacional, que muestra un ahorro superior al 40 % en el promedio de kilómetros viajados por partido jugado como visita en ambas temporadas. Su presidente, Juan Ignacio “Pepe” Sánchez, medalla de oro olímpica con la selección argentina en Atenas 2004, opinó lo siguiente: “Gracias al sistema desarrollado por el grupo de Investigación de Operaciones nuestro equipo ha llevado sus costos y tiempos de viaje a poco más de la mitad, lo que ha permitido mayor tiempo de descanso y entrenamiento, y una disminución en las lesiones con respecto a otras temporadas” [8].

Una decisión que han tomado varios equipos es reinvertir estos ahorros en reemplazar viajes en bus por viajes en avión para largas distancias (hasta la temporada 2013-2014 muchos de estos viajes, que a veces superaban los 1000 km, se realizaban en bus), lo que ha implicado una fuerte mejora en la calidad de viaje de los jugadores.

Algunos sectores habían expresado que el nuevo sistema podía tener como consecuencia una menor cantidad de victorias de visita por efecto de giras más largas, y menor afluencia de público por no tener días fijos para los partidos y por la “irregularidad” del fixture. Sobre el primer punto, los resultados de las series regulares no confirman esa predicción, el porcentaje de victorias de visita se ha mantenido prácticamente constante. En la última temporada con el sistema anterior hubo 32 % de victorias de visita, cifra que pasó a 33 % en la 2014-2015 y a 31 % en la 2015-2016. Con respecto a la concurrencia de público a los estadios, no contamos con cifras oficiales, pero la percepción de los dirigentes de la AdC es que la asistencia también se ha mantenido en

valores similares a los que había en el pasado.

El nuevo sistema ha tenido una aceptación mayoritaria en la comunidad del básquet argentino aunque también ha recibido algunas críticas de sectores más aferrados al sistema tradicional de parejas y partidos solamente los fines de semana, que se había mantenido en los últimos 30 años de la Liga. En ese sentido, cobran interés las declaraciones a medios periodísticos del entrenador de la selección nacional argentina y de Peñarol de Mar del Plata, Sergio Hernández, antes de la disputa de la última fecha de la serie regular de la temporada 2015-2016: “Aún los que somos críticos del nuevo formato debemos reconocer que el atractivo del torneo ha sido increíble. En esta última fecha en todos los partidos se juega por algo importante”.

Quien también expresó su opinión muy positiva sobre el nuevo formato de la Liga y el uso de técnicas matemáticas para su programación, es el mayor referente de la historia del básquet argentino, Emanuel “Manu” Ginóbili, estrella del básquet mundial y también campeón olímpico con Argentina en 2004: “Considerando el gran tamaño del país y la falta de rutas y vuelos interconectando las ciudades, algo había que hacer con el sistema del campeonato. Los matemáticos de la UBA hicieron un gran trabajo optimizando el tiempo de viaje de los equipos, lo que ha permitido un importante ahorro de dinero para toda la Liga. Seguramente hay ajustes para hacer, pero estoy muy entusiasmado con el cambio implementado” [8].

La conformidad de los directivos de la AdC con este proyecto ha llevado a extender el vínculo con el equipo académico para las próximas temporadas y también a explorar nuevos campos de colaboración. Actualmente estamos empezando a trabajar en un modelo que permita asignar los árbitros a los partidos de la Liga. El proceso se hace hoy de manera manual y existe el convencimiento que de aplicarse también en este caso técnicas modernas de Investigación de Operaciones, la asignación podría hacerse mucho más eficientemente.

***Agradecimientos:*** A la directiva y los clubes de la AdC por su compromiso con la implementación de este proyecto, especialmente a Fabián Borro, su presidente, y a Sergio Guerrero, su Secretario Técnico, quien apoya nuestro trabajo día a día. A “Pato” De Benedetti y todo el departamento de Competiciones de la AdC, por su invaluable ayuda en el proceso de preparación del fixture. A Carlos y Ricardo Prunes, por haber confiado en este grupo de académicos y habernos recomendado a la AdC. A los periodistas Matías Traversa, Fabián Pérez, Fabián García, Alejandro Sonich y Ricardo Sbrana por su colaboración en distintas etapas de este proyecto. A Mario Guajardo, Jaime Miranda, Richard Weber y Andrés Weintraub por sus múltiples sugerencias que mejoraron

la versión final de este trabajo. El primer autor está parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería, ISCI, Chile (ICM-FIC: P05-004-F, CONICYT: FB0816) y los subsidios UBACyT 20020130100808BA (Argentina), ANPCyT PICT 2012-1324 (Argentina) y FONDECyT 1140787 (Chile).

## Referencias

- [1] T. Bartsch, A. Drexler, y S. Kröger. Scheduling the professional soccer leagues of Austria and Germany. *Computers and Operations Research*, 33(7):1907–1937, 2006.
- [2] R. Bhattacharyya. A note on complexity of traveling tournament problem. *Optimization Online*, 2480, 2009. Consultado en junio 6, 2016, [http://www.optimization-online.org/DB\\_FILE/2009/12/2480.pdf](http://www.optimization-online.org/DB_FILE/2009/12/2480.pdf).
- [3] F. Bonomo, A. Cardemil, G. Durán, J. Marengo, y D. Sabán. An application of the traveling tournament problem: The Argentine volleyball league. *Interfaces*, 42(3):245–259, 2012.
- [4] G. Durán, M. Guajardo, J. Miranda, D. Sauré, S. Souyris, A. Weintraub, y R. Wolf. Scheduling the Chilean soccer league by integer programming. *Interfaces*, 37(6):539–552, 2007.
- [5] K. Easton, G. Nemhauser, y M. Trick. The traveling tournament problem description and benchmarks. In *International Conference on Principles and Practice of Constraint Programming*, páginas 580–584. Springer, 2001.
- [6] C. Fleurent y J.A. Ferland. Allocating games for the NHL using integer programming. *Operations Research*, 41(4):649–654, 1993.
- [7] D. Goossens y F. Spieksma. Scheduling the belgian soccer league. *Interfaces*, 39(2):109–118, 2009.
- [8] INFORMS. O.R. transforms scheduling of Chilean Soccer Leagues and South American World cup qualifiers. Consultado en Mayo 31, 2016, [https://www.pathlms.com/informs/events/533/thumbnail\\_video\\_presentations/26171](https://www.pathlms.com/informs/events/533/thumbnail_video_presentations/26171), Mayo 2016.
- [9] G. Kendall, S. Knust, C.C. Ribeiro, y S. Urrutia. Scheduling in sports: An annotated bibliography. *Computers and Operations Research*, 37(1):1–19, 2010.

- [10] G.L. Nemhauser y M.A. Trick. Scheduling a major college basketball conference. *Operations Research*, 46(1):1–8, 1998.
- [11] K. Nurmi y J. Kyngäs. Improving the schedule of the Finnish Major Ice Hockey League. In *Proceedings of the 2nd International Conference on the Mathematics in Sport. Groningen, Netherlands*, 2009.
- [12] R.V. Rasmussen. Scheduling a triple round robin tournament for the best Danish soccer league. *European Journal of Operational Research*, 185(2):795–810, 2008.
- [13] D. Recalde, R. Torres, y P. Vaca. Scheduling the professional Ecuadorian football league by integer programming. *Computers and Operations Research*, 40(10):2478–2484, 2013.
- [14] C.C. Ribeiro. Sports scheduling: Problems and applications. *International Transactions in Operational Research*, 19(1-2):201–226, 2012.
- [15] C.C. Ribeiro y S. Urrutia. Scheduling the Brazilian soccer tournament: Solution approach and practice. *Interfaces*, 42(3):260–272, 2012.
- [16] C. Thielen y S. Westphal. Complexity of the traveling tournament problem. *Theoretical Computer Science*, 412(4):345–351, 2011.
- [17] M. Trick. Challenge traveling tournament instances. *Online reference at <http://mat.gsia.cmu.edu/TOURN/>*, Consultado el 7 de junio de 2016, 2016.
- [18] S. Westphal. Scheduling the German basketball league. *Interfaces*, 44(5):498–508, 2014.
- [19] M.B. Wright. Scheduling fixtures for New Zealand cricket. *IMA Journal of Management Mathematics*, 16(2):99–112, 2005.
- [20] M.B. Wright. Scheduling fixtures for basketball in New Zealand. *Computers and Operations Research*, 33(7):1875–1893, 2006.



---

# LOGÍSTICA MARÍTIMA Y TERRESTRE DE UNA EMPRESA SALMONERA EN CHILE MEDIANTE PROGRAMACIÓN MATEMÁTICA

---

DIEGO DELLE DONNE \*

GUILLERMO DURÁN \*\* \*\*\*

GUIDO FUENTES \*\*\*\*

JAVIER MARENCO \*

JUAN IGNACIO VILLASANTE \*\*

ANDRÉS WEINTRAUB \*\*

## Resumen

La industria salmonera es uno de los principales sectores de exportación de Chile. En la etapa de engorda en agua salada, los peces son cultivados en grandes jaulas flotantes agrupadas en *centros de cultivo* a lo largo de toda la Décima Región del sur de Chile. Al momento de la cosecha, cuando alcanzan un peso y volumen de venta, los peces deben ser transportados en barcos hasta los denominados *centros de acopio*, ubicados cerca de las plantas de matanza. Llegados al centro de acopio los peces son depositados en grandes jaulas en espera a ser enviados a las *plantas de matanza*. En estas plantas se sacrifican y limpian los peces que luego son enviados hacia las *plantas de proceso* mediante transporte terrestre. Uno de los principales desafíos de las empresas salmoneras con este tipo de cadena de proceso es la planificación global de la logística marítima y terrestre, combinada con la planificación del inventario en las jaulas en los centros de acopio y de las líneas faenas en las plantas de matanza. En este trabajo estudiamos este problema y presentamos algoritmos para su resolución. El problema es dividido en dos etapas: la

---

\* Departamento de Computación, FCEN, UBA, Argentina e Instituto de Ciencias, Universidad Nacional de General Sarmiento, Argentina

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

\*\*\* Instituto de Cálculo y Departamento de Matemática, FCEN, UBA, Argentina y CONICET, Argentina

\*\*\*\* Empresas AquaChile, Chile

logística marítima y planificación de las plantas de matanza por un lado, y la logística terrestre por otro. Para la primera etapa presentamos un modelo de programación lineal entera mixta y un algoritmo heurístico y argumentamos el por qué del uso de la heurística en el caso práctico. La segunda etapa se resuelve heurísticamente mediante un algoritmo goloso. Detallamos los métodos implementados y presentamos resultados obtenidos en la aplicación de los mismos a problemas reales en una de las principales empresas salmoneras de Chile.

**Keywords:** Salmonicultura, Programación Entera, Heurísticas, Cadena de procesos.

---

## 1. Introducción

---

Los salmones son especies exóticas en Chile. Fueron introducidos al país entre 1850 y 1920. En 1921 los primeros salmones Coho fueron traídos y hasta 1973 diversos estudios e investigación fueron desarrollados por el instituto de fomento pesquero (IFOP) para ver el comportamiento y crecimiento de los salmones en Chile. Entre 1978 y 1980 nacieron las primeras compañías privadas dedicadas exclusivamente al cultivo de salmones y ya en 1985 había 36 centros de cultivo operando en Chile con una producción total que excedía las 1200 toneladas anuales. En 1990 las primeras ovas de salmón Coho fueron cultivadas permitiendo tener todo el ciclo de producción en el país. Este hecho fue el punto de despegue de la salmonicultura a gran escala en Chile. Para el año 2006 la producción chilena había alcanzado las 647.000 toneladas, llegando al segundo lugar de la producción mundial de salmones. En el año 2007 el virus ISA golpeó la industria, lo que generó una severa crisis afectando fuertemente los niveles productivos de la salmonicultura chilena. La industria se recuperó de la crisis del ISA y ya para el año 2014 la producción total había alcanzado las 955.000 toneladas de salmón (ver [21] para más detalles). Hoy en día la industria se enfrenta a problemas como el bloom de algas que han hecho disminuir la producción, sin embargo se espera que la producción total para el año 2016 sea mayor a las 700.000 toneladas.

La salmonicultura puede ser dividida en dos etapas principales, agua dulce y agua salada. En agua dulce los salmones son incubados y llevados a un tamaño cercano a los 200 gramos de peso para luego ser transportados a los *centros de cultivo* en agua salada. Al momento de la cosecha, cuando alcanzan un peso y volumen de venta, los peces deben ser transportados en barcos hasta los denominados *centros de acopio*, ubicados cerca de las plantas de matanza.

Llegados al centro de acopio los peces son depositados en grandes jaulas en espera a ser enviados a las *plantas de matanza*. En estas plantas se sacrifican y limpian los peces que luego son enviados hacia las *plantas de proceso* mediante transporte terrestre.

Los centros de cultivo están ubicados entre fiordos, islotes y bahías en zonas remotas de la costa del sur de Chile, lugares que en general tienen muy difícil acceso, por lo que el transporte hacia y desde los centros de cultivo en la mayoría de los casos se realiza en barco. Uno de los principales desafíos de la industria es la planificación de la logística y la cadena de proceso general. Esta es una tarea difícil porque tiene muchos eslabones, los cuales están sujetos a gran variabilidad. AquaChile es la compañía salmonera más grande de Chile, su producción en el año 2013 fue de alrededor de 102.000 toneladas brutas de cosecha [1]. El presente trabajo busca resolver el problema de la planificación de la logística marítima, de las faenas en las plantas de matanza y de la logística terrestre en la cadena de procesos de la empresa.

La literatura en acuicultura es extensa y ha sido estudiada por varios autores. Trabajos en áreas de recursos naturales relacionadas, como la agricultura y las pesquerías, han ayudado al desarrollo de esta área. Bjørndal et al. [3] hacen una revisión de la investigación de operaciones en el campo de la acuicultura. Los enfoques más estudiados son dos: el primero corresponde a modelos biológicos y el segundo corresponde a modelos económicos. Los modelos biológicos describen cómo se relacionan los sistemas de producción con el medio ambiente. Estos modelos abordan problemas como el de factores ambientales y crecimiento [6], las condiciones de cultivo de los peces [4], la relación entre alimentación y enfermedades [10] y el *trade-off* entre crecimiento y mortalidad [19]. Los modelos económicos relacionan los modelos biológicos con los precios de mercado, inventario, requerimientos de capital, restricciones de recursos, etc.

Bjørndal [2] desarrolla un modelo para describir las condiciones óptimas de cosecha para peces de acuicultura. Gran parte de la investigación desarrollada en esta área corresponde a modelos de crecimiento, alimentación y tiempo de cosecha. Otros trabajos en esta línea desarrollan modelos bioeconómicos para el crecimiento de los peces en función de la alimentación [9] y aplican programación dinámica en el proceso de crecimiento de camarones cultivados en granjas, para la toma de decisiones relacionadas con inventario, periodos de cultivo y momento de cosecha [17, 18].

Entre las aplicaciones recientes de investigación de operaciones en acuicultura, Cisternas et al. [12] presentan un modelo de optimización para planificar el mantenimiento y cambio de las mallas de los centros de cultivo. En cuanto a la cadena de producción, Bravo et al. [5] desarrollan dos modelos de pro-

gramación lineal entera mixta, uno para la fase de agua dulce y otro para la fase de agua salada, donde el primero tiene por objetivo satisfacer la demanda por peces que tienen los centros de cultivo en agua salada al mínimo costo y el segundo tiene por objetivo la planificación de las cosechas y siembras en los centros de cultivo. Existen trabajos de ruteo relacionados a la industria, como el de Romero et al. [20], que trata de la planificación del despacho de alimentos en barcos y inventario en los centros de cultivo. Sin embargo, no pareciera haber trabajos relacionados al transporte de peces y a la planificación del proceso de las plantas de matanza. Ya en un espectro más amplio, Christiansen et al. [11] realizan una revisión del campo de ruteo y programación de barcos, enfatizando las diferencias entre transporte terrestre y marítimo. Fagerholt y Lindstad [13] presentan un desarrollo para un problema de multiproductos y múltiples puertos, y Bronmo et al. [8] proponen una heurística multipartida para la programación de barcos. En el área más aplicada, Bronmo et al. [7] presentan un problema de planificación de barcos vagabundos. Los barcos vagabundos son barcos que venden su capacidad de carga que no tienen comprometida o contratada en mercados spot, por lo que no tienen una programación de itinerarios de viajes previamente definida. Esta debe ser planificada al momento de vender la capacidad restante de los barcos.

---

## 2. Descripción del problema

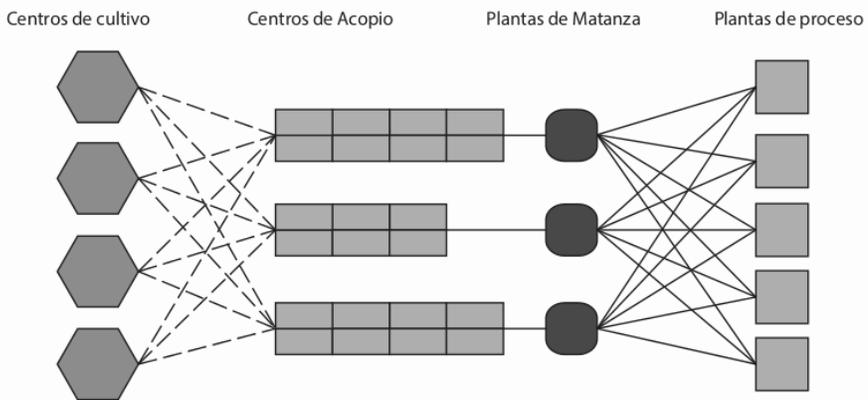
---

AquaChile posee una gran cantidad de centros de cultivo, tres centros de acopio, tres plantas de matanza y cinco plantas de proceso. La cadena de proceso se inicia en los centros de cultivo desde donde los peces a cosechar son transportados en barco hacia los centros de acopio. Luego de reposar en dichos centros, los peces se faenan en las plantas de matanza para luego ser transportados por tierra en camiones refrigerados hasta las plantas de proceso, las cuales representan el final de la cadena. La Figura 1 muestra un diagrama de la cadena de proceso.

En AquaChile, cada semana se confecciona un programa de demandas a partir de los requerimientos del área comercial. Este programa establece las cantidades de peces de cada centro de cultivo que se deben procesar en cada planta de proceso cada día de la semana. El programa especifica también las cantidades requeridas por especie, pero no especifica la planta de matanza que debe utilizarse para cada cosecha. De esta manera, la empresa debe planificar en conjunto la logística marítima, el manejo del inventario en los centros de acopio, los lotes de faena en las plantas de matanza y la logística terrestre

hacia las plantas de proceso, con el fin de satisfacer el programa de demandas de cada semana. Estas decisiones deben tener en cuenta los costos y tiempos de navegación, las capacidades de los barcos, las capacidades en jaulas de los centros de acopio, las capacidades de faena y horarios de funcionamiento de las plantas de matanza y los costos y tiempos de viaje terrestres entre las plantas de matanza y las de proceso. El objetivo principal de la empresa es satisfacer en tiempo y forma el programa de demanda de cada semana, tratando de hacerlo al menor costo posible. Hasta el momento de comenzar este trabajo, esta planificación solía ser realizada manualmente por el planificador de la cadena en una simple planilla de cálculo.

Figura 1: Cadena de proceso de AquaChile. Línea punteada: transporte marítimo.



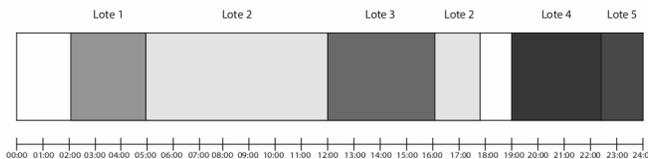
## 2.1. Características de la cadena de proceso

A continuación definimos algunos detalles de los diferentes aspectos y etapas de la cadena de proceso de AquaChile:

- **Centros de cultivo:** Los centros de cultivo consisten en un número de jaulas localizadas entre fiordos, islotes y bahías protegidas. En estos lugares es donde los salmones son engordados para luego poder ser cosechados.
- **Wellboats:** Los *wellboats* son barcos especialmente diseñados para el transporte de peces vivos. Hay varios tipos y tamaños. Tienen tuberías y bombas especiales para succionar los peces hacia y desde los estanques.

- Centros de Acopio:** Los centros de acopio disponen de jaulas que sirven para mantener un inventario vivo de salmones para las plantas de matanza. Cada centro de acopio está ubicado al lado de una planta de matanza y se conecta a la misma por medio de tuberías a través de las cuales se envían los peces para las faenas.
- Especies:** AquaChile tiene tres especies distintas de salmón: Coho, Salar y Trucha. Después de cada viaje los peces deben reposar en las jaulas del centro de acopio entre 16 y 24 horas (según la especie) antes de ser faenados en las plantas de matanza para reducir el estrés del viaje y poder tener una mejor calidad de producto.
- Plantas de matanza:** En las plantas de matanza es donde se sacrifica, eviscera y limpia a los salmones. Estas están conectadas a los centros de acopio vía grandes tuberías por las cuales los salmones son succionados para ser faenados en la planta. La Figura 2 muestra un diagrama ejemplificando la utilización a lo largo del día de la línea de faena en una planta de matanza. Cada *lote* representa una cantidad de salmón a faenar que será luego enviado a una determinada planta de proceso. Es importante notar que dos lotes de salmón distintos no pueden ser faenados al mismo tiempo y que un lote puede ser faenado en distintos momentos, siempre y cuando cumpla con la fecha de llegada a la planta de proceso que lo demanda. Las plantas de matanza tienen horarios de trabajo determinados que deben ser respetados.

Figura 2: Faena de diferentes lotes de salmón en planta de matanza



- Camiones:** Luego de ser faenados, los peces son transportados en camiones refrigerados hasta las plantas de proceso. AquaChile posee un contrato con una compañía de transporte para el traslado de los peces. En la sección 4 describimos en mayor detalle los aspectos relacionados al problema de la planificación del transporte terrestre.
- Plantas de proceso:** En las plantas de proceso se elaboran y empaquetan los productos finales y representan el fin de la cadena de proceso. Estas plantas tienen horarios de trabajo determinados y los camiones

solo pueden descargar los peces, provenientes de las plantas de matanza, durante dichos horarios. Si un camión llega a la planta fuera de estos horarios deberá esperar a que comience el siguiente turno para poder descargar los peces. Algunas plantas de proceso comparten su locación con una planta de matanza, con lo cual en estos casos no se requiere el transporte en camión de los peces.

- **Programa de las plantas de proceso:** Las plantas de proceso demandan peces todos los días. Estas demandas son incluidas en un programa semanal el cual especifica la fecha en que tienen que llegar los salmones a las plantas de proceso, la especie de salmón, el número de unidades, el peso unitario promedio y desde qué centro de cultivo tienen que ser cosechados.

La planificación conjunta de la logística marítima, del manejo del inventario en los centros de acopio, de los lotes de faena en las plantas de matanza y de la logística terrestre hacia las plantas de proceso es un problema de difícil resolución debido a la gran cantidad de decisiones y consideraciones que deben tomarse en cuenta y es especialmente difícil realizarla a mano en una planilla de cálculo. Esta planificación debe ser realizada al menos una vez a la semana y cada vez que haya cambios en las condiciones del problema. En algunos casos debe ser planificada dos y hasta tres veces en un mismo día dados los diferentes factores que afectan a la cadena de proceso, especialmente las condiciones climáticas.

Para una empresa salmonera como AquaChile, una buena planificación de la cadena de proceso es muy importante porque asegura un buen suministro de salmón a las plantas de proceso y disminuye los costos de transporte marítimo y terrestre. Claramente, apoyar la toma de decisiones con herramientas computacionales basadas en algoritmos de optimización para la planificación diaria sería una gran ventaja para las operaciones de AquaChile, reforzando así su competitividad estratégica en la industria.

En este trabajo presentamos algoritmos para resolver el problema planteado dividiéndolo en dos etapas: la logística marítima y planificación de las plantas de matanza por un lado y la logística terrestre por otro. Para la primera etapa presentamos, en la sección 3, un modelo de programación lineal entera mixta y un algoritmo heurístico y argumentamos el por qué del uso de la heurística en el caso práctico. Para la segunda etapa proponemos, en la sección 4, una heurística de tipo *golosa* para resolver el problema de la logística terrestre.

---

### 3. Logística marítima y planificación de faenas

---

La primera parte del problema corresponde a la planificación de la logística marítima, combinada con la planificación de la ocupación de las jaulas de los centros de acopio y la coordinación de los procesos de faena en las plantas de matanza. Vale notar que como resultado de esta etapa se obtienen además los viajes por tierra que deben realizarse mediante camiones para el transporte de los peces desde las plantas de matanza hasta las plantas de proceso. En esta etapa no se realiza una optimización sobre la logística terrestre, quedando este aspecto por resolver en una segunda etapa (la cual se detalla en la sección 5). La planificación de la logística marítima debe ser realizada semanalmente por el programador y corresponde a tomar una serie de decisiones que determinarán, por ejemplo, en qué barcos se transportarán los salmones desde los centros de cultivo hacia los centros de acopio. Estas decisiones no son simples y deben resolver el problema optimizando el funcionamiento de toda la cadena de proceso de AquaChile y aprovechando de la mejor manera posible los viajes de los barcos. Las decisiones a considerar durante la planificación de la logística marítima involucran los siguientes aspectos:

- **Transporte marítimo:** Qué barco utilizar para transportar cada orden desde los centros de cultivo hasta los centros de acopio, cuánto volumen llevar, qué especie y cuándo zarpar.
- **Selección de centro de acopio y jaula:** A qué centro de acopio llevar los peces y en qué jaula de dicho centro descargarlos.
- **Programación de las faenas:** Cuándo faenar cada lote de peces para poder cumplir con la demanda de las plantas de proceso en tiempo y forma respetando las restricciones de capacidad de las plantas de matanza.

Las decisiones anteriores están condicionadas por una serie de restricciones que deben ser tomadas en cuenta a la hora de planificar. Los centros de acopio tienen un número determinado de jaulas, las cuales tienen capacidades establecidas. Se debe respetar los horarios de funcionamiento y capacidad de faena de las plantas de matanza. Se tiene que mantener la trazabilidad de los peces, por lo que no se puede descargar peces en una jaula de un centro de acopio que esté previamente ocupada y no se puede cargar peces de distintas jaulas en un mismo estanque de un wellboat (no se pueden mezclar). Por último se deben respetar las fechas de las demandas de las plantas de proceso, es decir, llegar con los peces a las plantas de proceso en el día en que fueron requeridos.

### 3.1. Modelo de programación lineal entera mixta

En esta sección presentamos el modelo de programación lineal entera mixta desarrollado para la resolución del problema. Se modela el tiempo por medio de períodos discretos de igual duración, siendo la duración un parámetro del modelo. Para simplificar el modelo, asumimos que todos los barcos están disponibles desde el inicio del horizonte de planificación, ya que cualquier otro caso es fácilmente modelable y no agrega valor al modelo desarrollado. En la tabla 1 se describen los parámetros del problema utilizados. El modelo desarrollado hace uso de las siguientes variables de decisión:

- $x_{bact}^I \in \{0, 1\}$ : 1 si el barco  $b \in B$  comienza un viaje de *ida* desde  $a \in A$  hacia  $c \in C$  en el período  $t \in T$ . 0 en caso contrario.
- $x_{bcate}^V \in \{0, 1\}$ : 1 si el barco  $b \in B$  comienza un viaje de *vuelta* desde  $a \in A$  hacia  $c \in C$  en el período  $t \in T$ , transportando peces de la especie  $e \in E$ . 0 en caso contrario.
- $x_{bat}^S \in \{0, 1\}$ : 1 si el barco  $b \in B$  se queda detenido en  $a \in A$  en el período  $t \in T$ . 0 en caso contrario.
- $w_{bcatej} \in \mathbb{R}_+$ : cantidad de toneladas de la especie  $e \in E$  que lleva el barco  $b \in B$  desde  $c \in C$  hasta la jaula  $j \in J(a)$  de  $a \in A$ , empezando el viaje en  $t \in T$ .
- $s_{ajcet} \in \mathbb{R}_+$ : cantidad de toneladas de la especie  $e \in E$  del cultivo  $c \in C$  en la jaula  $j \in J(a)$  del acopio  $a \in A$  en  $t \in T$  disponibles para faena (es decir, ya reposados).
- $s_{ajcet}^R \in \mathbb{R}_+$ : cantidad de toneladas de la especie  $e \in E$  del cultivo  $c \in C$  que se descargan en la jaula  $j \in J(a)$  del acopio  $a \in A$  en el período  $t \in T$  (estos peces pasarán al inventario real de la jaula cuando termine su período de reposo).
- $y_{ajcet} \in \{0, 1\}$ : 1 si la jaula  $j \in J(a)$  del acopio  $a \in A$  está ocupada con peces de especie  $e \in E$  del cultivo  $c$  en  $t \in T$  (ya sea con peces en reposo o disponibles para faena). 0 en caso contrario.
- $f_{ajcept} \in \mathbb{R}_+$ : toneladas de peces de especie  $e \in E$  del cultivo  $c \in C$  que se faenan de la jaula  $j \in J(a)$  del acopio  $a$ , comenzando el proceso en  $t \in T$ , para ser enviados a  $p$ .

Parámetro	Descripción
$A$	Conjunto de centros de acopio.
$J(a)$	Conjunto jaulas del centro de acopio $a \in A$ .
$B$	Conjunto de barcos.
$C$	Conjunto de centros de cultivo.
$D$	Conjunto de días (consecutivos) de la planificación.
$E$	Conjunto de especies.
$P$	Conjunto de plantas secundarias.
$T \subset \mathbb{N}$	Conjunto de períodos de la planificación ( $T = \{1, \dots,  T \}$ ).
$dia(t)$	Indica a qué día $d \in D$ pertenece el período $t \in T$ .
$dem(c, p, d, e)$	Toneladas de salmón demandadas por la planta $p \in P$ del centro de cultivo $c \in C$ para el día $d \in D$ de la especie $e \in E$ .
$cap(b)$	Capacidad de carga de peces del barco $b \in B$ .
$costo_m(b, a, c)$	Costo de transporte marítimo para el barco $b \in B$ entre el acopio $a \in A$ y el cultivo $c \in C$ .
$costo_t(a, p)$	Costo de transporte terrestre entre el acopio $a \in A$ y la planta de proceso $p \in P$ .
$t_m(a, c)$	Tiempo de navegación entre $a \in A$ y $c \in C$ .
$t_t(a, p)$	Tiempo de viaje por tierra entre el acopio $a \in A$ y la planta de proceso $p \in P$ .
$t_r(e)$	Tiempo mínimo de reposo en acopio para la especie $e \in E$ .
$cap_j(a, j)$	Capacidad de la jaula $j$ del centro de acopio $a \in A$ .
$capp(a)$	Capacidad de proceso de la planta de matanza asociada al centro de acopio $a \in A$ .
$ini(b)$	Centro de acopio inicial para el barco $b \in B$ . Es el centro de acopio del cual el barco parte la navegación.
$s^I(a, j, c, e)$	Inventario inicial de salmones del centro de cultivo $c \in C$ y especie $e \in E$ en la jaula $j$ del centro de acopio $a \in A$ .

Tabla 1: Parámetros utilizados por el modelo de programación lineal entera mixta.

Con estas definiciones, el modelo de programación lineal entera mixta desarrollado queda como:

$$\begin{aligned} \text{Minimizar } & \sum_{b \in B} \sum_{a \in A} \sum_{c \in C} \sum_{t \in T} \left( \text{costo}_m(b, a, c) \left( x_{bact}^I + \sum_{e \in E} x_{bcate}^V \right) \right) + \\ & \sum_{a \in A} \sum_{j \in J(a)} \sum_{c \in C} \sum_{e \in E} \sum_{p \in P} \sum_{t \in T} \text{costo}_t(a, p) \cdot f_{ajcept} \end{aligned}$$

sujeito a las siguientes restricciones:

- $\forall (b, a) \in B \times A$  y  $t = 1$

$$\sum_{c \in C} x_{bact}^I + x_{bat}^S = \delta_a \tag{1}$$

donde  $\delta_a = 1$  si  $a = ai(b)$  y  $\delta_a = 0$  si no.

- $\forall (b, a, t) \in B \times A \times T, t > 1,$

$$\sum_{\substack{c \in C \\ t > t_m(a,c)}} \sum_{e \in E} x_{bca(t-t_m(a,c))e}^V + x_{ba(t-1)}^S = \sum_{c \in C} x_{bact}^I + x_{bat}^S \tag{2}$$

- $\forall (b, c, t) \in B \times C \times T,$

$$\sum_{\substack{a \in A \\ t > t_m(a,c)}} x_{bac(t-t_m(a,c))}^I = \sum_{a \in A} \sum_{e \in E} x_{bcate}^V \tag{3}$$

- $\forall (b, c, a, t, e) \in B \times C \times A \times T \times E,$

$$\sum_{j \in J(a)} w_{bcatej} \leq x_{bcate}^V \cdot \text{cap}(b) \tag{4}$$

- $\forall (a, t) \in A \times T, j \in J(a),$

$$\sum_{b \in B} \sum_{\substack{c \in C \\ t > t_m(a,c)}} \sum_{e \in E} w_{bca(t-t_m(a,c))ej} \leq \text{cap}j(a, j) \cdot \left( 1 - \sum_{c \in C} \sum_{e \in E} y_{ajcet} \right) \tag{5}$$

- $\forall (a, c, e) \in A \times C \times E, j \in J(a),$

$$s_{ajce1} + \sum_{p \in P} f_{ajcep1} = s^I(a, j, c, e) \tag{6}$$

- $\forall (a, c, e, t) \in A \times C \times E \times T, j \in J(a),$

$$\sum_{b \in B} w_{bca(t-t_m(a,c))ej} = s_{ajcet}^R \tag{7}$$

- $\forall (a, c, e, t) \in A \times C \times E \times T, j \in J(a),$

$$s_{ajce(t-1)} + s_{ajce(t-t_r(e))}^R = s_{ajcet} + \sum_{p \in P} f_{ajcept} \tag{8}$$

$$\blacksquare \forall(a, c, e, t) \in A \times C \times E \times T, j \in J(a),$$

$$s_{ajcet} + \sum_{t-t_r(e) < t' < t} s_{ajcet'}^R \leq \text{cap}j(a, j) \cdot y_{ajcet} \quad (9)$$

$$\blacksquare \forall(a, t) \in A \times T, j \in J(a),$$

$$\sum_{c \in C} \sum_{e \in E} y_{ajcet} \leq 1 \quad (10)$$

$$\blacksquare \forall(a, t) \in A \times T,$$

$$\sum_{j \in J(a)} \sum_{c \in C} \sum_{e \in E} \sum_{p \in P} f_{ajcept} \leq \text{capp}(a) \quad (11)$$

$$\blacksquare \forall(c, p, d, e) \in C \times P \times D \times E,$$

$$\sum_{a \in A} \sum_{j \in J(a)} \sum_{\substack{t \in T \\ \text{dia}(t+t_i(a,p))=d}} f_{ajcept} \geq \text{dem}(c, p, d, e) \quad (12)$$

Las restricciones (1), (2) y (3) aseguran que los viajes realizados por los barcos sigan un flujo factible. La restricción (4) asegura que la carga de un barco no exceda su capacidad y la restricción (5) asegura que las descargas en una jaula no excedan la capacidad de la misma. Esta última impide además la descarga de peces en una jaula si la misma no está vacía. La restricción (6) registra el inventario inicial para el primer período y la restricción (7) registra el inventario de peces en reposo en una jaula (i.e, para el período en que un barco realiza su descarga en tal jaula). La restricción (8) calcula el inventario de peces de una jaula en un período dado en función del inventario en el período anterior y de la cantidad de peces faenados en este período (y eventualmente de los peces que terminan su reposo en tal período). La restricción (9) asegura que si el inventario de una jaula en un período es positivo, entonces la variable que indica si la jaula tiene peces debe ser igual a 1. Por su parte, la restricción (10) impide que en una jaula existan peces de distinta especie y/o centro de cultivo (por requisito de trazabilidad). Finalmente, las restricciones (11) impiden que se exceda la capacidad de faenado de la planta y las restricciones (12) aseguran que los lotes de faena satisfagan las demandas de las plantas de proceso en tiempo y forma.

Si bien el tamaño de los bloques para la discretización del tiempo puede hacerse tan pequeño como se quiera, los tiempos de resolución del modelo en instancias reales fueron prohibitivos para la utilización de bloques más pequeños que 1 hora. Así, las soluciones halladas, aun siendo soluciones óptimas para el modelo, no representan en general soluciones buenas en la práctica. De hecho, en una experimentación preliminar pudimos verificar que las soluciones

halladas por la heurística (detallada en la siguiente sección) resultaban muchas veces mejores soluciones que las soluciones óptimas del modelo (debido principalmente al tamaño del bloque de discretización). Por este motivo, se decidió profundizar el desarrollo en el método heurístico para poder así resolver el problema real acorde a las necesidades de la empresa.

### **3.2. Heurística para la logística marítima y planificación de faenas**

Presentamos en esta sección una heurística que sigue principalmente un esquema *goloso*, con algunas particularidades (que detallaremos en la sección 3.2.2) que buscan evadir los clásicos problemas de este tipo de esquemas (por ejemplo, relacionados a la obtención de óptimos locales). La heurística desarrollada se divide en dos etapas. La primera busca utilizar el inventario inicial (es decir, remanente de la semana anterior) para satisfacer la mayor cantidad de demanda que pueda. La segunda etapa se encarga de planificar los viajes de barco necesarios para traer peces para las demandas que no han podido satisfacerse mediante el inventario inicial.

Ambas etapas contemplan también la planificación completa de la línea de faena en las plantas de matanza, indicando los tiempos de inicio y fin de cada lote a faenar. La segunda etapa contempla también el uso de las jaulas en los centros de acopio y define qué jaulas se utilizarán para la descarga de los peces de los viajes que se planifiquen. Lo que busca la heurística es en primer lugar satisfacer la mayor cantidad de peces demandados por las plantas de proceso y en segundo lugar minimizar los costos operacionales (es decir, de navegación y de transporte terrestre).

Una solución provista por la heurística se compone básicamente de una serie de *actividades*, como por ejemplo los viajes de los barcos, la descarga de peces en una jaula, la faena de un lote de peces, etc. Durante la ejecución de la heurística, la solución se va construyendo gradualmente agregando actividades a la misma. Es importante aclarar que el orden en que se van agregando las actividades a la solución puede no respetar el orden cronológico de las actividades mismas.

Por ejemplo, en un momento de la ejecución puede planificarse un lote de faena para el día 4 a las 16:00 hs y más adelante puede planificarse otro lote a faenar para el día 4 a las 10:00 hs (aun en una misma planta de matanza). Obviamente, en tal caso la heurística se asegura que la faena de las 10:00 hs no interfiera con ninguna otra faena ya planificada para dicha planta (por ejemplo, deberá terminar antes de las 16:00 hs). Algo similar puede ocurrir con el uso de las jaulas. Por ejemplo, puede planificarse que un barco descargue

peces en una jaula el día 5 a las 10:00 hs, y más adelante en la ejecución puede planificarse que otro barco descargue peces en esta misma jaula el día 3 a las 14:30 hs. En tal caso, la heurística se asegura que esta última descarga sea completamente faenada antes de las 10:00 hs del día 5, momento en el cual la jaula debería estar ya vacía para la descarga de peces en ella.

### 3.2.1. Algoritmo General

El primer objetivo de la heurística desarrollada consiste en satisfacer parte de la demanda semanal con el inventario que se tiene en los centros de acopio al inicio de la programación. Para ello se planifican con este inventario los lotes de faena necesarios con el objetivo de satisfacer la mayor cantidad de demanda posible. Una vez utilizado todo el inventario inicial, se busca satisfacer todas las demandas que no pudieron ser satisfechas con inventario, para las cuales será necesario el uso de los barcos para transportar los peces desde los centros de cultivo hacia los centros de acopio.

Con tal objetivo, se confeccionan los itinerarios de viaje para los barcos disponibles. Un *viaje* se compone de un *trayecto de ida* desde un centro de acopio hasta uno de cultivo y un *trayecto de vuelta* desde este último hasta un centro de acopio (que puede ser el mismo que el origen del viaje o no). En el trayecto de ida el barco viaja vacío y en el de vuelta transporta los peces cosechados en el centro de cultivo. El *itinerario* de un barco se compone de un conjunto de viajes consecutivos indicando para cada viaje los días y horarios de inicio y fin (contemplando los tiempos de viaje, y de carga y descarga de peces).

Al momento de agregar un viaje al itinerario de un barco, el algoritmo toma también las siguientes decisiones respecto a los peces que se traerán en tal viaje:

- Se decide qué cantidad de peces se transportará y en qué jaula del centro de acopio se los depositará. Para ello debe asegurarse que la jaula esté vacía en el momento en que el barco arriba al acopio.
- Se determina qué demandas se cubrirán (total o parcialmente) con los peces traídos. Se define también si una cantidad de estos peces quedarán como remanente de inventario al final de la semana.
- Se define en qué momento (o momentos) se realizará la faena de los peces traídos, teniendo en cuenta no exceder la capacidad de matanza de la planta. Eventualmente, si ya se tenía planificada otra descarga de peces en la misma jaula para una fecha futura, el algoritmo debe asegurarse

también que la jaula quede vacía para ese momento (es decir, que la carga traída se faene por completo antes de esa fecha).

En cada iteración del algoritmo, se agrega un nuevo viaje al itinerario de algún barco junto con todas las actividades mencionadas más arriba, relacionadas con los peces transportados. El Algoritmo 1 detalla el proceso general de asignación de inventario inicial (Fase I) y planificación de viajes (Fase II).

---

**Algoritmo 1:** Algoritmo general

---

- 1: FASE I: *Asignación del inventario inicial*
  - 2:  $J_s \leftarrow$  conjunto de jaulas con salmones al inicio del programa semanal
  - 3: **para cada**  $j \in J_s$ , en orden creciente según cantidad de peces de  $j$   
**hacer**
  - 4:   **para cada** demanda  $d$  compatible con  $j$ , en orden temporal **hacer**
  - 5:     **si** se puede faenar de  $j$  para  $d$  **entonces**
  - 6:       Asignar peces de  $j$  para cubrir demanda  $d$
  - 7:       Actualizar demanda restante de  $d$  y stock en la jaula  $j$
  - 8:     **fin si**
  - 9:   **fin para**
  - 10: **fin para**
  
  - 11: FASE II: *Planificación de viajes para satisfacer demandas restantes*
  - 12: **para cada** demanda restante  $d$  **hacer**
  - 13:    $viajes \leftarrow \emptyset$
  - 14:   **para** cada barco  $b$ , acopio  $a$  **hacer**
  - 15:      $v \leftarrow$  confeccionar viaje y faenas para  $(b, a, d)$  [ver detalle en el texto]
  - 16:     Si se obtuvo un viaje  $v$  y faenas asociadas, agregar  $v$  a  $viajes$ .
  - 17:   **fin para**
  - 18:   Elegir el “mejor” viaje (y faenas asociadas) del conjunto  $viajes$  y agregarlo al itinerario del barco correspondiente.
  - 19: **fin para**
- 

Entre las líneas 2 y 10 el algoritmo asigna el stock inicial. En la línea 2 se selecciona el conjunto de jaulas  $J_s$  que contienen salmones al inicio de la programación (es decir., el inventario inicial). Recordemos que para que un barco descargue salmones en un centro de acopio, éste debe poseer suficientes jaulas vacías disponibles. Es por esto que las jaulas de  $J_s$  son recorridas en orden creciente según la cantidad de peces que tenga (línea 3), con el objetivo de vaciar las jaulas lo antes posible, liberándolas para posibles futuras descargas de peces.

En la línea 4, se recorre el conjunto de demandas en busca de demandas que puedan ser satisfechas con inventario inicial.

Decimos que los peces de una jaula son *compatibles* con una demanda si cumplen con los requerimientos de ésta (es decir., centro de cultivo y especie requerida).

Se prioriza satisfacer primero las demandas de los primeros días de la semana, ya que para el resto de los días pueden utilizarse los barcos (notemos que para las demandas de los primeros días puede no haber tiempo suficiente para ir a buscar los peces a los centros de cultivo requeridos).

Por tanto, en la línea 4 se ordenan las demandas por fecha, desde la más temprana a la más tardía.

En el caso de que dos demandas tengan la misma fecha, para minimizar costos de transporte, se eligen las demandas asociadas a la planta de proceso más cercanas al centro de acopio del cual se asignan los peces.

Para cubrir una demanda con inventario inicial asociado a una jaula, la planta de matanza correspondiente al centro de acopio de la jaula al cual pertenecen los peces debe tener capacidad de faena y además deben ser capaces de llegar a la planta de proceso respectiva en la fecha establecida por la demanda.

Si se quiere cubrir la demanda  $d$  con peces asociados a la jaula  $j$  (línea 6), el algoritmo debe verificar que la planta de matanza del acopio de  $j$  sea capaz de faenar la cantidad a cubrir de la demanda  $d$  a tiempo (línea 5).

Entre las líneas 12 y 19 la heurística planifica los viajes marítimos y las faenas asociadas a los peces traídos en los mismos. El algoritmo itera por cada demanda que no haya sido satisfecha aún (línea 12).

Al igual que para el inventario inicial, esta lista de demandas se recorre en orden creciente según las fechas de las mismas, pero en este caso, si dos demandas coinciden en su fecha se elige primero la demanda que requiera mayor cantidad de peces. Dada una demanda  $d$ , en las líneas 13 a 17 se intenta seleccionar el mejor viaje posible para satisfacer dicha demanda.

Para ello, se debe seleccionar un barco con el cual transportar los peces requeridos por  $d$  y un centro de acopio en el cual depositarlos para su matanza futura.

Con tal objetivo, se prueban todas las combinaciones de barcos  $b$  y acopios  $a$  (línea 13) y para cada una de ellas se confecciona un viaje  $(b, a, d)$  y se planifican todas las actividades relacionadas con los peces que se transportan en el mismo, como se explicó anteriormente.

Es así que la confección del viaje  $v$  en la línea 15 no es un proceso trivial y eventualmente podría no ser factible (por ejemplo, si el acopio en cuestión no posee jaulas disponibles donde depositar los peces o bien si la planta de matanza está ya saturada en su capacidad). Si la cantidad de peces traída (es

decir., la capacidad del barco) supera la cantidad requerida por  $d$ , el excedente de peces se utilizará para cubrir demandas futuras que requieran estos peces y para ello se planificarán también nuevos lotes de faena correspondientes a estas otras demandas. Finalmente, en la línea 18 se elige el “mejor” viaje del conjunto de viajes confeccionados y se lo agrega al itinerario del barco correspondiente. Al dejar en firme este viaje se registran en la solución las siguientes modificaciones:

- Se agenda la llegada de los peces a la jaula correspondiente.
- Se programan las faenas de los peces para los momentos que se hayan determinado y se registra el momento en que la jaula queda vacía, si es el caso.
- Se actualizan las demandas que serán satisfechas (total o parcialmente) con los peces traídos en el viaje, es decir, se resta de cada una de éstas, las cantidades cubiertas por los lotes a faenar registrados.

Para determinar cuál es el “mejor” viaje (línea 18), se utilizaron dos criterios. El primero selecciona el viaje que utilice lo mejor posible la capacidad del barco. Por ejemplo, un viaje que transporta 120 toneladas de peces en un barco de capacidad 130 toneladas será mejor que otro que transporta 120 toneladas en un barco de capacidad 180 toneladas, pues el primero utiliza un 92,3% de su capacidad mientras que el segundo sólo un 66,7%.

El segundo prioriza el viaje que tenga el menor costo operacional. Este costo incluye los costos de viaje del barco en los trayectos de ida y vuelta y un estimado de los costos de transporte por tierra desde la planta de matanza hasta la planta de proceso final. El algoritmo itera con los dos criterios y selecciona la mejor solución global.

### 3.2.2. El algoritmo general y diferentes puntos de partida

Es evidente que los algoritmos desarrollados utilizan un esquema *goloso* para construir las soluciones al problema y es sabido que en este tipo de algoritmos se toman decisiones basadas en un criterio de optimalidad local, sin medir muchas veces el impacto que éstas tendrán en las decisiones que se tomarán a futuro, durante la ejecución. En particular, en la segunda etapa del Algoritmo 1, la planificación de un viaje para una determinada demanda  $d$  tiene un fuerte impacto también en demandas compatibles con  $d$ , ya que usualmente la carga transportada en un barco suele alcanzar para satisfacer varias demandas del programa. Es así, que los primeros viajes planificados en este algoritmo, imponen considerables decisiones a futuro.

Para mitigar los efectos adversos de los algoritmos de tipo *goloso*, es común combinar esta técnica con alguna técnica de mejora gradual de las soluciones obtenidas (es decir, búsqueda local) como por ejemplo las metaheurísticas GRASP [14] y *Tabú Search* [15, 16], entre otras. Lamentablemente, desarrollar un procedimiento de búsqueda local para el problema tratado en este trabajo no pareciera ser una tarea sencilla, pues hay muchos factores a considerar. Por ejemplo si se quiere hacer un intercambio de barco para dos soluciones, se deben verificar muchas cosas, tales como los horarios de llegada, las capacidades de los barcos, las capacidades de faena de las plantas de matanza y horarios en que serán faenados los lotes, y las capacidades de las jaulas en los centros de acopio.

Como resultado, en la mayoría de los casos, el intercambio resulta infactible. Por este motivo, omitimos la utilización de técnicas dependientes de una búsqueda local, como las mencionadas más arriba. Por otro lado, mediante una breve experimentación preliminar, pudimos ver que las soluciones obtenidas al variar arbitrariamente los primeros viajes planificados se diferencian en gran medida.

Con el objetivo de cubrir todas las alternativas para los primeros viajes, presentamos en el Algoritmo 2 un procedimiento similar al Algoritmo 1 pero en el cual la segunda etapa se ejecuta varias veces fijando en cada caso un viaje arbitrario para el primer viaje planificado.

Éste es el algoritmo desarrollado finalmente para AquaChile.

---

**Algoritmo 2:** Algoritmo general con iteraciones en la Fase II

---

- 1:  $solucion\_base \leftarrow$  solución (incompleta) de la Fase I del Algoritmo 1
  - 2:  $S \leftarrow \emptyset$
  - 3:  $D_1 \leftarrow$  Demandas del primer día del programa
  - 4: **para cada** barco  $b$ , acopio  $a$  y demanda  $d \in D_1$  **hacer**
  - 5:      $sol \leftarrow solucion\_base$
  - 6:      $v \leftarrow$  confeccionar viaje para  $(b, a, d)$  en  $sol$
  - 7:     **si** se obtuvo un viaje  $v$  **entonces**
  - 8:         Agregar  $v$  al itinerario correspondiente en  $sol$
  - 9:         Completar  $sol$  con la Fase II del Algoritmo 1
  - 10:        Si se obtuvo una solución, agregarla a  $S$
  - 11:     **fin si**
  - 12: **fin para**
  - 13: Elegir la mejor solución de  $S$  y devolverla (si  $S = \emptyset$ , terminar sin solución).
-

---

## 4. Programación de la logística terrestre

---

La segunda etapa de la cadena de procesos de AquaChile corresponde al transporte de los lotes de peces faenados por las plantas de matanza hacia las plantas de proceso. Los peces son cargados en contenedores debidamente refrigerados y estos contenedores son transportados en camiones hacia las plantas de proceso. Semanalmente el planificador recibe el programa de faenas de las plantas de matanza, el cual le indica las cantidades de peces que deben ser transportadas desde las plantas de matanza hacia las plantas de proceso. En base al programa, él debe decidir con qué contenedores va a transportar los peces desde las distintas plantas. AquaChile posee dos tipos de contenedores para mover lotes de peces:

- **Bins:** Los bins son contenedores individuales. En cada bin se puede cargar alrededor de 450 Kg de pescado con hielo. Cada camión cargado puede llevar hasta 28 bins (alrededor de 12,6 toneladas). Al ser individuales, los bins permiten el transporte de varios lotes diferentes en un sólo camión, lo que flexibiliza el transporte.
- **Estanques:** Los estanques corresponden a un sólo gran contenedor. Cada uno puede cargar alrededor de 18 toneladas de pescado con hielo. La ventaja de los estanques es que se puede transportar más cantidades de pescado que en bins por viaje, pero sólo pueden llevar pescado de un lote en particular (los lotes no se pueden mezclar).

Para la programación existen varios factores y restricciones que condicionan la asignación de contenedores. Cada contenedor debe ser cargado con hielo antes de ser cargado con peces para asegurar un buen transporte y que los peces lleguen en buenas condiciones a la planta de proceso. Algunas de las plantas de matanza sólo pueden cargar peces en estanques y algunas plantas de proceso sólo pueden recibir los peces en bins. También se deben considerar los tiempos de traslado entre las plantas y el tiempo que toma cargar el hielo en los estanques para calcular si es que un estanque puede llegar a tiempo para un determinado proceso. Este cálculo del tiempo sólo afecta a los estanques porque estos son limitados. Con los bins no es necesario porque las plantas de matanza siempre tienen bins disponibles y son cargados con hielo en la misma planta.

Para trasladar los contenedores de un lugar a otro, AquaChile posee un contrato con una empresa de transporte externa que se encarga de proveer los

camiones a las plantas de matanza de acuerdo a la planificación de contenedores de AquaChile. Las plantas de matanza siempre tienen bins a disposición, por lo que los camiones sólo deben presentarse en la planta a la hora establecida por el planificador de AquaChile.

En el caso de los estanques, los camiones deben pasar a buscarlos vacíos por donde sea que estén, llevarlos a las plantas de matanza y luego transportarlos cargados a las plantas de proceso.

Por ejemplo, si se necesita trasladar tres estanques desde una planta de matanza hacia una planta de proceso y en la planta de matanza no se dispone de estanques, los camiones deberán buscar los estanques vacíos, llevarlos a la planta de matanza, esperar a que sean cargados y transportarlos a las plantas de proceso correspondientes.

La asignación de los camiones a las tareas es realizada independientemente por la empresa de transporte, la cual se encarga de decidir qué camiones utilizar para cada tarea.

AquaChile paga simplemente una tarifa por viaje realizado entre plantas de matanza y plantas de proceso. La tarea del planificador de AquaChile es programar la asignación de contenedores a los lotes de peces faenados tomando en cuenta las restricciones anteriormente señaladas, garantizando que todos los peces sean transportados a destino y minimizando los costos de transporte (por ejemplo, aprovechando de la mejor manera posible los estanques para disminuir la cantidad de viajes de camiones).

#### 4.1. Heurística de asignación de contenedores

El problema de asignación de contenedores no es una tarea fácil para ser realizada manualmente. En algunas semanas del año el programador debe asignar más de 200 contenedores a las distintas faenas de las plantas de matanza. Este proceso es altamente demandante y le lleva más de un día en completarlo, generalmente durante el fin de semana. Para ayudarlo con la planificación, se desarrolló una heurística golosa que asigna contenedores a las distintas faenas de la semana disminuyendo los costos de transporte.

La heurística se encarga de asignar contenedores a cada uno de los lotes faenados en las plantas de matanza. Para cada lote verifica qué tipo de contenedor puede ser asignado y cuántos contenedores de ese tipo deberían ser asignados. Debido a que la cantidad de estanques que posee la empresa es limitada, es necesario un correcto uso de ellos para poder hacer una buena planificación y evitar que las plantas que no pueden utilizar bins se queden sin estanques disponibles. Para eso, la heurística crea *itinerarios* para cada uno de los estanques, donde guarda todos sus movimientos y mantiene actualiza-

da la posición final de cada estanque y la hora a la que estará nuevamente disponible. En el Algoritmo 3 se muestra en detalle el funcionamiento de la heurística.

---

**Algoritmo 3:** Planificación de contenedores

---

```

1: para  $t \in$  Días hacer
2:    $L \leftarrow$  Conjunto de lotes a transportar en el día  $t$ 
3:   para lote  $l \in L$  hacer
4:     mientras  $l > 0$  hacer
5:       si Hay contenedores compatibles con  $l$  entonces
6:         si  $l >$  camión cargado con bins entonces
7:           Asignar estanque disponible más cercano  $e$ 
8:           Actualizar cantidad remanente de  $l$ 
9:           Actualizar itinerario de  $e$ 
10:        si no
11:          Asignar Bins
12:          Actualizar cantidad remanente de  $l$ 
13:        fin si
14:      si no
15:        Moverse a próximo  $l$ 
16:      fin si
17:    fin mientras
18:  fin para
19: fin para

```

---

Teniendo en cuenta que algunas plantas de matanza sólo pueden cargar en estanque y que la empresa cuenta con un número limitado de ellos, la heurística privilegia los lotes asociados a estas plantas de manera que no les vaya a faltar estanque al momento de la carga. Por esta razón los lotes son ordenados primero por prioridad de estanque y luego por distancia (desde la planta de matanza a la de proceso) en orden decreciente (línea 3). Estos criterios se basan en los utilizados por el programador. Como la compañía de transporte cobra por viaje en base a las distancias entre las plantas de matanza y plantas de proceso, se intenta minimizar los costos de transporte priorizando el uso de estanque para el traslado de lotes que tengan que viajar mayores distancias por sobre el resto. De esa manera se intenta garantizar que por lo menos los viajes más largos (es decir, los más caros) sean la menor cantidad posible y minimizando así el costo de transporte (los estanques son más eficientes, porque pueden cargar más peces que los bins).

La heurística busca asignar el contenedor más conveniente a cada lote de

peces. Esto depende de la cantidad de peces del lote que quede por transportar y de la cantidad de contenedores disponibles para ese entonces. En la línea 5 se verifica si hay algún contenedor compatible para el lote  $l$ . Si no hay, el remanente de  $l$  no puede ser transportado y se continúa con el siguiente lote. Por otro lado, si hay contenedores compatibles el algoritmo verifica la cantidad de peces del lote  $l$  que queda por ser asignada. Si la cantidad es mayor que la capacidad de un camión cargado con bins (línea 6), carga los peces en el estanque disponible más cercano (línea 7). Si no hay estanque disponible o la cantidad remanente de peces del lote  $l$  no supera la capacidad de un camión cargado con bins, asigna el lote a bins. Con esto se busca no desaprovechar un estanque para poca cantidad de peces. Es decir, si el remanente de  $l$  no supera la capacidad de un camión con bins, entonces es asignado a bins para no desaprovechar el estanque (al cual le sobraría capacidad de carga y podría ser utilizado para otro lote más grande).

Al asignar un estanque a un lote se actualiza el itinerario del estanque (línea 9) para registrar el recorrido del estanque y las horas en las que llega a sus destinos. A partir de esta información, se calcula si los estanques estarán disponibles para transportar determinados lotes futuros, tomando en cuenta los tiempos de viaje hacia las plantas de matanza y de carga de hielo.

Finalmente, se calculan los costos en base a las tarifas cobradas por la empresa de transporte. Es importante mencionar que si bien el algoritmo intenta disminuir los costos de contrato con la compañía que provee los camiones, esta última se ve también beneficiada por las decisiones que toma la heurística en sus operaciones. En la práctica, cada estanque es transportado la mayor parte del tiempo por el mismo camión. Esto se traduce en que la heurística al programar los contenedores y priorizar la asignación de los estanques más cercanos, estará programando gran parte de la flota de camiones. Implicará entonces que la empresa de transporte facture menos, pero sus asignaciones serán más eficientes y reducirá sus costos operacionales.

---

## 5. Resultados

---

Las dos aplicaciones desarrolladas fueron puestas a prueba en AquaChile y al día de la fecha son utilizadas como herramientas de apoyo en la programación de las actividades diarias de la empresa o bien de simulación de escenarios para apoyar la toma de decisiones gerenciales. Presentamos en esta sección un análisis de la puesta en práctica de las herramientas implementadas en AquaChile.

### 5.1. Logística marítima y planificación de faenas

Entre diciembre de 2014 y marzo de 2015 se realizó la puesta en práctica de la herramienta de planificación marítima y de faenas. Durante estos meses, se utilizó la herramienta en paralelo con la planificación manual con el objetivo de analizar y comparar las soluciones para medir las diferencias en eficiencia y determinar aspectos que el software pudiera no estar tomando en cuenta. Luego de varias pruebas y ajustes, se logró finalmente que las soluciones entregadas por la herramienta cumplan los requerimientos prácticos de la empresa. El software funcionó sin fallas y cumplió con las expectativas de los usuarios y la gerencia, quienes estimaron que las soluciones entregadas por la herramienta son de muy buena calidad para el proceso de la empresa.

Durante los primeros meses luego de la puesta en marcha definitiva, la planificación se realizó semanalmente con la nueva herramienta, aunque luego los cambios durante la semana eran implementados manualmente sobre la marcha en el día a día en terreno. Además de esta baja en la producción, la industria se vio fuertemente afectada por el *bloom de algas* de principios del año 2016, el cuál generó la muerte de alrededor de 43.000 toneladas de salmones en el sur de Chile. Todo esto se tradujo en una baja importante en las cantidades de peces transportados, una mayor capacidad relativa en los centros de acopio, barcos, plantas de matanza y proceso, lo cual por su parte simplificó mucho el problema de planificación al punto de poder resolverlo sin necesidad del uso de la herramienta, acaso utilizando recursos en exceso. Con todo esto, durante el año 2015 solo cerca de 20 semanas fueron planificadas íntegramente con la herramienta desarrollada. De todas maneras, la gerencia de la empresa sigue teniendo en cuenta los potenciales beneficios de la herramienta desarrollada y espera en un futuro reactivar en mayor medida su utilización.

Si bien los cambios de las condiciones en la industria redujeron la utilización de la herramienta para la planificación semanal, ésta es actualmente utilizada para la simulación de escenarios. Las simulaciones permiten evaluar las posibles planificaciones ante eventuales cambios en los recursos disponibles (barcos, acopios, plantas, etc.) y considerar de esta manera cambios de gran impacto en la empresa. Un ejemplo claro de esto es la estrategia de la empresa con respecto a los barcos. AquaChile negocia contratos para el arriendo de barcos por períodos de tiempo fijo, los cuales se negocian dependiendo de los planes de cosecha que se tienen para la temporada. Durante el segundo trimestre del 2015, los contratos de dos de sus cuatro barcos estaban a punto de expirar y la empresa debía decidir si renovarlos, cambiarlos por otros, o ampliar o reducir la flota actual. En particular, la empresa tenía dos opciones: quedarse con los dos barcos de 85 toneladas o cambiarlos por uno sólo de 130 toneladas.

La toma de esta decisión debía tener en cuenta muchas consideraciones, tales como la demanda de peces que iba a tener la empresa para ese trimestre, qué tan alejados estaban los centros de cultivo a cosechar, la disponibilidad de centros de acopio y plantas de matanza, etc. Frente a tal importante decisión, se utilizó la herramienta desarrollada para realizar simulaciones en distintos escenarios y evaluar los costos de las planificaciones provistas con el fin de elegir la flota más conveniente. Como resultado pudo observarse que si bien el contrato de dos barcos de 85 toneladas era más caro que el de un barco de 130, el primero permitía desarrollar planificaciones más económicas que el segundo dejando así una mayor rentabilidad final a la empresa. Es así que luego de muchas simulaciones, se tomó la decisión de mantener los dos barcos de 85 toneladas.

## 5.2. Logística Terrestre

El software de planificación de contenedores ha tenido un alto impacto en la logística terrestre de AquaChile y la empresa se ha visto considerablemente mejorada mediante el uso de la herramienta, tanto de forma cuantitativa como cualitativa. La herramienta desarrollada fue puesta en marcha en diciembre del año 2014 y se continúa utilizando hasta la fecha. El programador utiliza la heurística para planificar los estanques utilizados cada semana; los cambios menores son modificados manualmente y los mayores son introducidos a la aplicación para la reprogramación automática de la planificación.

Para cuantificar el desempeño del software, se corrieron simultáneamente a mano y con la aplicación cuatro semanas de un mismo mes y se compararon los resultados obtenidos. La tabla 2 muestra los costos incurridos por la planificación manual y por la planificación con la heurística, y la diferencia entre ambas. Se muestran además los valores promedio teniendo en cuenta las cuatro semanas. Se puede verificar que en las cuatro semanas de la muestra, el software entregó resultados significativamente mejores que la programación manual. En promedio, la planificación con el software es cerca de un 9% más eficiente que la manual, implicando ahorros de alrededor de US\$ 5500 semanales.

SEMANA	COSTO (US\$)		TONELADAS		COSTO (US\$/TON)		DIFERENCIA	
	MAN	HEUR	MAN	HEUR	MAN	HEUR	US\$	%
1	50.957	45.194	1.269	1.269	39,86	35,62	5.403	10,65 %
2	54.806	49.634	1.363	1.363	40,21	36,41	5.172	9,44 %
3	72.828	65.388	2.036	2.036	35,77	32,12	7.440	10,22 %
4	70.604	66.918	2.117	2.117	33,35	31,61	3.687	5,22 %
PROMEDIO	62.298	56.783	1.696	1.696	37,30	33,94	<b>5.425</b>	<b>8,88 %</b>

Tabla 2: Comparación entre las planificaciones manuales y las automáticas.

En cuanto a los beneficios cualitativos que la herramienta ha aportado a la empresa, es preciso mencionar la considerable disminución en el tiempo requerido por el programador para realizar la planificación. Manualmente, debía dedicar un día completo o a veces más para planificar la semana. Hoy en día, mediante el uso de la herramienta, la planificación se realiza en alrededor de una hora y media. Esto permite destinar el resto del tiempo a evaluar la calidad de las soluciones y probar distintas alternativas y potenciales escenarios, cotejando así posibilidades de inversión para el mejoramiento de la empresa. Otro factor importante a considerar es la información detallada que entrega la aplicación. Anteriormente el programador tenía una vaga idea de los horarios de llegada y salida de los camiones y de los estanques que se utilizarían durante la semana. Actualmente el software entrega el detalle completo de las soluciones, informando de los itinerarios de todos los estanques, especificando orígenes, destinos, horarios y cantidades de peces a cargar. Además de esto, cada planificación queda guardada en la base de datos de la empresa, las que pueden ser copiadas para correr semanas similares y ser reanalizadas para eventuales mejoras en el funcionamiento de la logística de AquaChile.

---

## 6. Conclusiones

---

Este trabajo presentó el desarrollo e implementación de herramientas de optimización para resolver el problema de planificación de la logística marítima, faenas en plantas de matanza y logística terrestre de una de las principales empresas salmoneras de Chile.

Una de las grandes dificultades de la planificación de la cadena de proceso de AquaChile es la gran variabilidad de sus componentes (demandas, inventarios, transporte, etc). Esta se debe principalmente a los constantes cambios de las condiciones climáticas del sur de Chile, los que obligan a replanificar los programas de la empresa, en muchos casos, varias veces a la semana. Para poder abordar esta variabilidad, las herramientas fueron desarrolladas de modo tal que cualquier cambio en la cadena de proceso de la empresa pudiera ser replanificado en cortos periodos de tiempo y de manera sencilla. Se prefirió desarrollar un enfoque determinístico por sobre uno estocástico, debido a que este último no habría podido resolver el problema real de planificación semanal de la empresa.

Tanto para la planificación de la logística marítima y faenas en plantas de matanza, como para la planificación de la logística terrestre se utilizaron enfoques de resolución heurísticos, obteniendo resultados de muy buena calidad.

Si bien el problema de la logística marítima se intentó resolver mediante un enfoque de programación lineal entera mixta, las soluciones entregadas eran de peor calidad que las proporcionadas por la heurística, debido al tamaño de los bloques de tiempo del modelo.

La implementación y uso de las herramientas desarrolladas ha tenido un gran impacto en las operaciones de AquaChile. El software de la logística marítima ha demostrado entregar buenas soluciones a la hora de realizar las planificaciones semanales y ha permitido a la empresa poder simular escenarios para poder elegir las flotas de barco más adecuadas para los periodos futuros. El software de la logística terrestre ha permitido planificar los viajes con mejoras que rondan el 9% con respecto a la planificación manual, y los tiempos de planificación han disminuido considerablemente, desde más de un día de planificación a un par de horas.

Es importante destacar también que el exitoso desarrollo del proyecto se debe en gran medida al trabajo conjunto y coordinado entre el grupo de académicos, gerencia y todo el personal de la empresa. El buen entendimiento entre las partes, la facilidad de acceder a la información y correcta organización del trabajo fueron claves para un buen desarrollo.

Como trabajo futuro se propone resolver el problema de la logística marítima y la planificación de faenas en las plantas de matanza con un método mixto, el cuál combine la actual heurística con un modelo de programación lineal. Consideramos que un método de estas características podría mejorar aún más las soluciones que se obtienen actualmente.

**Agradecimientos:** A la empresa AquaChile por toda su colaboración y apoyo económico para la concreción de este proyecto. Este trabajo contó también con el soporte económico del Instituto Sistemas Complejos de Ingeniería (ICM:P-05-004-F, CONICYT: FBO816). Guillermo Durán fue parcialmente financiado por los subsidios FONDECyT 1140787, UBACyT 20020130100808BA (Argentina) y ANPCyT PICT 2012-1324 (Argentina). Andrés Weintraub recibió soporte económico de FONDECyT 1120318.

## Referencias

- [1] AquaChile. Sitio web: <http://www.aquachile.com>.
- [2] T. Bjørndal. Optimal harvesting of farmed fish. *Marine Resource Economics*, páginas 139–159, 1988.

- [3] T. Bjørndal, D.E. Lane, y A. Weintraub. Operational research models and the management of fisheries and aquaculture: a review. *European Journal of Operational Research*, 156(3):533–540, 2004.
- [4] B. Björnsson. Effects of stocking density on growth rate of halibut (*hipoglossus hippoglossus* l.) reared in large circular tanks for three years. *Aquaculture*, 123(3):259 – 270, 1994.
- [5] F. Bravo, G. Durán, A. Lucena, J. Marengo, D. Morán, y A. Weintraub. Mathematical models for optimizing production chain planning in salmon farming. *International Transactions in Operational Research*, 20(5):731–766, 2013.
- [6] J.R. Brett. 10 - environmental factors and growth. In D.J. Randall W.S. Hoar y J.R. Brett, editors, *Bioenergetics and Growth*, volume 8 of *Fish Physiology*, páginas 599 – 675. Academic Press, 1979.
- [7] G. Brønmo, M. Christiansen, y B. Nygreen. Ship routing and scheduling with flexible cargo sizes. *The Journal of the Operational Research Society*, 58(9):1167–1177, 2007.
- [8] G. Brønmo, M. Christiansen, K. Fagerholt, y B. Nygreen. A multi-start local search heuristic for ship scheduling—a computational study. *Computers & Operations Research*, 34(3):900–917, 2007.
- [9] O.J. Cacho, U. Hatch, y H. Kinnucan. Bioeconomic analysis of fish growth: effects of dietary protein and ration size. *Aquaculture*, 88(3):223 – 238, 1990.
- [10] C.Y. Cho y D.P. Bureau. Development of bioenergetic models and the fish-prfeq software to estimate production, feeding ration and waste output in aquaculture. *Aquatic Living Resources*, 11(04):199–210, 1998.
- [11] Marielle Christiansen, Kjetil Fagerholt, Bjørn Nygreen, y David Ronen. Chapter 4 maritime transportation. In C. Barnhart y G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, páginas 189 – 284. Elsevier, 2007.
- [12] F. Cisternas, D. Delle Donne, G. Durán, C. Polgatz, y A. Weintraub. Optimizing salmon farm cage net management using integer programming. *Journal of the Operational Research Society*, 64(5):735–747, 2013.
- [13] K. Fagerholt y H. Lindstad. Turborouter: An interactive optimisation-based decision support system for ship routing and scheduling. *Maritime Economics & Logistics*, 9(3):214–233, 2007.

- [14] T.A. Feo y M.G. Resende. Greedy randomized adaptive search procedures. *Journal of Global Optimization*, 6:109–134, 1995.
- [15] F. Glover. Tabu search - part 1. *ORSA Journal on Computing*, 1(3):190–206, 1989.
- [16] F. Glover. Tabu search - part 2. *ORSA Journal on Computing*, 2(1):4–32, 1990.
- [17] P. Leung, E. Hochman, K. Wanitprapha, Y.C. Shang, y J. Wang. Optimal harvest schedule for maricultured shrimp: a stochastic sequential decision model. 1989.
- [18] PingSun Leung y Yung C. Shang. Modeling prawn production management system: A dynamic markov decision approach. *Agricultural Systems*, 29(1):5 – 20, 1989.
- [19] M. Mangel y J. Stamps. Trade-offs between growth and mortality and the maintenance of individual variation in growth. *Evolutionary Ecology Research*, 3(5), 2001.
- [20] G. Romero, G. Durán, J. Marenco, y A. Weintraub. An approach for efficient ship routing. *International Transactions in Operational Research*, 20(6):767–794, 2013.
- [21] SalmonChile. Sitio web: <http://www.salmonchile.cl>.

---

# UN NUEVO ENFOQUE PARA LA PROGRAMACIÓN HORARIA EN UNIVERSIDADES

---

ISABEL MÉNDEZ-DÍAZ \*  
JUAN JOSÉ MIRANDA-BRONT \* \*\*  
PAULA ZABALA \* \*\*

## Resumen

En este trabajo consideramos un problema de programación horaria proveniente de una aplicación real de una universidad privada de Argentina. El objetivo es diseñar horarios que contemplen simultáneamente los recursos disponibles de la universidad (docentes, aulas, horarios, etc.) y las preferencias de los alumnos sobre las materias que desean cursar. Nuestra propuesta es un algoritmo heurístico basado en programación lineal entera que, de acuerdo a la experiencia obtenida con datos reales provenientes de la universidad, mostró una excelente performance en la calidad de las soluciones obtenidas medida conforme a la satisfacción estudiantil alcanzada en base a sus preferencias.

**Palabras Clave:** Programación lineal entera, Metaheurística, Programación de horarios

---

## 1. Introducción

---

En el contexto de entidades educativas, la programación horaria suele ser una problemática difícil de resolver debido a la multiplicidad de requerimientos que debe satisfacer una solución. De acuerdo al escenario que se desee abordar, los trabajos existentes en la literatura se enmarcan fundamentalmente en tres tipos de problemas. Por un lado, la **Programación de exámenes (ETP)**

---

\* Departamento de Computación, FCEyN, Universidad de Buenos Aires

\*\* Consejo Nacional de Investigaciones Científicas y Técnicas

[4, 8, 16], cuyo objetivo es programar horarios y asignar aulas a exámenes en un período de tiempo, satisfaciendo un conjunto de restricciones, por ejemplo que no haya superposición horaria entre exámenes que debe rendir un mismo alumno. Un segundo enfoque es la **Programación de horarios de materias post inscripción (PECTP)** [16] donde el fin es programar horarios y asignar aulas a un conjunto de materias dentro de un período de tiempo, satisfaciendo la selección previamente realizada por los alumnos. Hay diferentes restricciones a tener en cuenta, en particular que las clases correspondientes a las materias elegidas por un mismo alumno no tengan superposición horaria. Finalmente, la **Programación de currícula (CCTP)** [3, 12, 14, 16] busca diagramar horarios y asignar aulas respetando la currícula de cada carrera, por ejemplo las clases de las materias que deben cursarse en el mismo período no deben superponerse en horario.

Una de las principales diferencias entre estos dos últimos problemas es que en el PECTP la asignación se realiza considerando una pre-inscripción de los alumnos a materias. En cambio, en el CCTP solamente se considera información referida a correlatividades y precedencias curriculares entre las materias.

Existen variantes de los tres enfoques mencionados que fundamentalmente difieren en qué tipo de restricciones consideran o desde qué óptica se busca una solución factible y/o óptima. Bettinelli et al. [2], Kristiansen et al. [11] y Babaei et al. [1] realizan un muy detallado relevamiento de los problemas de diagramación horaria y Lübbecke [15] analiza las dificultades en la implementación práctica de los diferentes enfoques de la literatura.

En su mayoría, la metodología para este tipo de problemas está enmarcada en propuestas metaheurísticas, fundamentalmente debido a que el tamaño de las instancias y la complejidad del problema hacen imposible un abordaje con métodos exactos. Entre los muchos trabajos podemos mencionar a Lewis [13] que usa la técnica de *Simulated Annealing*, Jat et al. [9] proponen un algoritmo genético con búsqueda tabú y Nothegger et al. [17] un algoritmo colonia de hormigas. Cambazard et al. [5] analizan diferentes algoritmos, incluyendo *Constraint Programming*. También pueden encontrarse en la literatura varios trabajos que utilizan modelos de programación lineal entera para desarrollar heurísticas, entre los cuales podemos mencionar a van del Broek et al. [20], Sørensen et al. [18], Kristiansen et al. [10] y Daskalaki et al. [7].

Cabe señalar que la mayoría de las propuestas aplican una fase previa de preprocesamiento en la cual, a partir de la derivación de conflictos, pueden disminuir significativamente la cantidad de variables y/o restricciones del problema, el espacio de búsqueda, etc. Este proceso es clave y de esta manera logran muy buenos resultados computacionales tanto en instancias de la

literatura como en instancias reales.

La perspectiva adoptada por van den Broek et al. [19] para el problema de horarios de Eindhoven University of Technology considera que los horarios de las clases ya están establecidos y que los alumnos tienen una lista de preferencias de materias a partir de la cual se le debe asignar un número dado de materias. El objetivo es lograr una asignación que satisfaga restricciones de recursos (por ejemplo capacidad de aulas), se ajuste a la programación horaria previamente diagramada y maximice la satisfacción del conjunto de alumnos de acuerdo a sus preferencias.

En este trabajo, nos enfocamos en una aplicación proveniente de una problemática real de una universidad privada de Argentina donde el objetivo es diseñar horarios que contemplen simultáneamente los recursos disponibles y las preferencias de los alumnos sobre las materias que desean cursar. El problema comparte muchas características con PECTP y con el trabajo de van den Broek et al. [19], pero la diferencia esencial es que en PECTP las materias en las cuales los alumnos están inscriptos están definidas a priori y en [19] los horarios de las clases están preestablecidos. En nuestro caso, tanto la asignación de materias a alumnos en función de sus preferencias como la asignación horaria son contempladas y definidas en forma conjunta.

Hasta donde llega nuestro conocimiento, no existe en la literatura este enfoque simultáneo. La contribución de nuestro trabajo es abordar un problema práctico que integra y combina problemas de la literatura. Proponemos una metodología para resolverlo basada en programación lineal entera y reportamos resultados en instancias reales. Además analizamos el beneficio obtenido por la universidad con el nuevo proceso aplicado en la diagramación horaria de las materias.

El trabajo está organizado de la siguiente manera. En la sección 2 introducimos el contexto y descripción de las características de la problemática de la universidad. En la sección 3 presentamos un modelo de programación lineal entera que formaliza el problema. En la sección 4 describimos la metodología propuesta para resolverlo y en la sección 5 analizamos la performance del algoritmo en instancias reales. Finalmente, la sección 6 muestra las conclusiones del trabajo y líneas futuras.

---

## 2. Contexto y características del problema

---

La universidad ofrece 12 carreras de grado, con aproximadamente 120 materias dictadas por alrededor de 220 docentes para 1200 alumnos. El año académico está dividido en dos semestres cuyos horarios se programan en forma independiente. Cada día de la semana está dividido en bloques horarios (en general de una hora) y el horario semanal se repite durante todo el semestre.

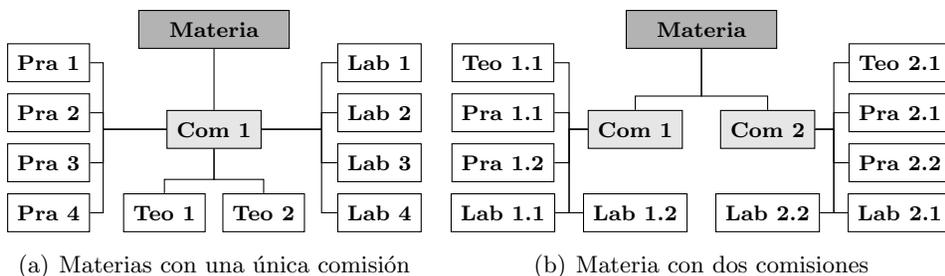
Las materias correspondientes a la currícula de una carrera de grado no tienen mucha relación de correlatividades. Esta característica otorga cierta libertad en la combinación de materias que puede cursar un alumno cada semestre.

Para cada carrera, las materias están clasificadas en mandatorias y electivas. En el primer caso, corresponden a materias que el alumno debe cursar obligatoriamente para obtener su título. En el segundo caso, son materias que están agrupadas por temáticas. De cada grupo temático los alumnos pueden elegir a lo sumo cierta cantidad que está establecida a priori en la currícula de la carrera correspondiente.

Las materias están conformadas por tres tipos de clases: teórica, práctica y laboratorio, combinadas en tres posibles esquemas: teórica, teórica-práctica o teórica-práctica-laboratorio. Cada una de estas clases tiene un patrón horario definido por una frecuencia semanal y una cantidad de bloques horarios continuos por vez. A cada unidad de este patrón lo llamamos *evento*.

Además, las materias están divididas en comisiones. Cada comisión tiene sus propias clases que siguen el mismo esquema. Los alumnos pueden estar inscriptos en una única comisión, no permitiéndose el cruce entre clases de diferentes comisiones y cada clase tiene un cupo determinado por la universidad.

Figura 1: Ejemplo de esquemas



La figura 1 ilustra posibles esquemas. En figura 2(a) la materia consiste de una única comisión y un alumno puede estar inscripto en cualquier combinación de teórica, práctica y laboratorio. La figura 2(b) muestra una materia con dos comisiones. Cada alumno puede estar inscripto en comisión 1 o comisión 2 (excluyente) y dentro de cada comisión, en cualquier combinación de clases que respete el esquema. Por ejemplo Teo. 1.1, Prac. 1.2 y Lab. 1.1 es una combinación factible, pero Teo. 1.1, Prac. 1.1, Lab. 2.2 no lo es.

Previo al comienzo de cada semestre, cada departamento de la universidad decide la oferta de materias y qué docentes estarán a cargo de cada una de sus clases. Junto a esta información, los docentes informan su disponibilidad horaria así como también el máximo de veces por semana que desean dedicar a la enseñanza.

Respecto a la infraestructura, la universidad dispone de 30 aulas con capacidad que varía entre 8 y 110 asientos. Las aulas están clasificadas según sus características en *tipos*, por ejemplo disponibilidad de equipamiento multimedia, instrumental de laboratorio. etc. Dependiendo de los requerimientos, las clases deben ser asignadas a ciertos tipos de aulas.

El proceso habitual de programación horaria seguido por la universidad está bajo la responsabilidad de una única persona quien, usando su conocimiento y datos estadísticos acerca de potenciales alumnos de cada materia, en forma manual procura armar una asignación horaria. Durante un tedioso proceso iterativo, contempla la disponibilidad de recursos y evita conflictos horarios con el objetivo de asegurar una cantidad mínima de materias a cada alumno. Este último punto es particularmente importante debido a que, siendo una facultad privada, los alumnos deben tener la posibilidad de terminar su carrera en tiempo y forma sin verse obligados a afrontar gastos extras.

A partir de una diagramación inicial, si hubiera algún conflicto o algún alumno no pudiera cursar el número requerido de materias, la persona responsable de la asignación, consulta a docentes acerca de la posibilidad de una mayor flexibilidad horaria y repite el proceso hasta lograr una programación horaria que la satisfaga. Una vez logrado el objetivo, los alumnos son informados sobre las materias que le fueron asignadas. Como resultado de este proceso, y fundamentalmente por causa de una asignación compulsiva que no tiene en cuenta los intereses de los alumnos, al comienzo del semestre numerosos alumnos piden al responsable de la programación cambios en la asignación. En algunos casos, las demandas de cambios son simples de satisfacer. En otros, alteran significativamente la programación y el responsable se encuentra ante la disyuntiva de negar el pedido o tener que rehacer el engorroso proceso buscando otras alternativas. En cualquiera de los dos casos se presenta una situación que sería deseable evitar.

Este proceso tiene varios puntos débiles. En primer lugar, es ineficiente e insume mucho tiempo ya que no es humanamente posible considerar todas las alternativas. Por otro lado, el trabajo depende exclusivamente de una persona quien está expuesta a una gran presión para llegar en tiempo y conformar a docentes y alumnos respetando los recursos disponibles. Finalmente, la asignación de materias a los alumnos es compulsiva y no les da posibilidad de elegir cómo planificar su carrera universitaria.

Debido a esta situación compleja, la universidad decide implementar un sistema informático que ayude en el proceso de programación horaria y contacta al grupo de Optimización Combinatoria del Departamento de Computación de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires del cual formamos parte. Este trabajo de consultoría y desarrollo se enmarcó dentro de las actividades de transferencia de tecnología realizadas por nuestro grupo.

Después de un relevamiento exhaustivo de los diferentes agentes involucrados en la diagramación horaria, nuestra sugerencia fue desarrollar un sistema de programación horaria donde, fundamentalmente, se cambia la óptica sobre la cual se toman decisiones incluyendo las preferencias de los alumnos en el proceso de asignación.

Para esto, el proceso sugerido es que se ofrezca a cada alumno una lista de materias que debe ordenar de acuerdo a sus preferencias. De esa lista se le asignará un número máximo de materias y se le garantiza un número mínimo de asignaciones para evitar retrasos en su carrera. Ambos números deben ser establecidos por la universidad. Por otro lado, la universidad clasifica a los alumnos de acuerdo a su desempeño académico con el objetivo de priorizar las preferencias de aquellos con mayor dedicación y además puede indicar si alguna materia de la lista debe ser asignada obligatoriamente (para contemplar casos en que no cursar cierta materia les cierre oportunidades en el próximo semestre).

El sistema entonces tendrá como objetivo realizar una asignación a materias para cada alumno y una programación horaria factible contemplando los recursos disponibles (aulas, docentes, etc) que maximice la preferencia global de los alumnos ponderada por el desempeño académico.

La propuesta, además de considerar en forma simultánea la asignación a materias y la diagramación horaria, debe contemplar que, a diferencia de muchos de los problemas que aparecen en la literatura, la universidad requiere que las restricciones operacionales y de recursos sean estrictas y no pueden ser violadas. En el caso que no sea factible cumplirlas, el sistema podría ser usado en escenarios menos restrictivos definidos por la universidad, como por ejemplo aumentando la disponibilidad horaria de los docentes.

Como mencionamos anteriormente, el problema así planteado conjuga características del PECTP y del trabajo de van den Broek et al. [19] quien toma en cuenta listas de preferencias pero asume que los horarios ya están fijados. Por otro lado, si la lista de preferencias de los alumnos está compuesta por exactamente la cantidad de materias garantizadas como mínimo, entonces nuestro problema se convierte en PECTP más allá de algunas diferencias en restricciones particulares.

Sin embargo, en nuestro caso la asignación de las materias es parte del proceso de decisión lo cual dificulta el problema. El conocer a priori las materias que cursa un alumno permite realizar una etapa previa de preprocesamiento y deducción de conflictos que pueden ser usados en la definición de la programación horaria. Por ejemplo, clases correspondientes a un mismo alumno no pueden tener intersección horaria.

En nuestro caso, este análisis no es posible debido a que generalmente la lista de preferencias de los alumnos es mayor que la cantidad de materias que debe asignarse. Por lo tanto, los conflictos horarios derivados de la asignación de materias dependen de cuáles efectivamente le sean asignadas y no es algo que pueda deducirse previamente. Además, dada la estructura de las materias, esta misma situación se presenta a nivel comisión.

---

### 3. Modelo de programación lineal entera

---

En esta sección formalizamos el problema a través de un modelo de programación lineal entera. Comenzamos por introducir algunas definiciones y notación agrupadas por incumbencia:

- Días y bloques horarios

$\mathcal{D}$	Conjunto de días de la semana.
$S^d$	Conjunto de bloques horarios del día $d \in \mathcal{D}$ .
$\mathcal{S} = \bigcup_{d \in \mathcal{D}} S^d$	Conjunto de bloques horarios.

- Aulas

$\mathcal{R}$	Conjunto de aulas.
$Types$	Conjunto de tipos de aulas.
$cap^r$	Capacidad del aula $r \in \mathcal{R}$ .
$nr^i$	Cantidad de aulas del tipo $i$ para $i \in Types$ .
$Classes^i$	Clases que pueden ser asignadas a aulas del tipo $i$ .
$t^r$	Tipo del aula $r \in \mathcal{R}$ .

■ Materias

$\mathcal{M}$	Conjunto de materias.
$\mathcal{C}$	Conjunto de clases.
$\mathcal{G}$	Conjunto de grupos de materias electivas.
$Q^m$	Comisiones de la materia $m \in \mathcal{M}$ .
$T^m$	Clases teóricas de la materia $m \in \mathcal{M}$ .
$Cl^m$	Clases de la materia $m \in \mathcal{M}$ .
$Lec^q \subseteq \mathcal{C}$	Clases teóricas de la comisión $q$ .
$Ex^q \subseteq \mathcal{C}$	Clases prácticas de la comisión $q$ .
$Lab^q \subseteq \mathcal{C}$	Clases de laboratorio de la comisión $q$ .
$Slots^c \subseteq \mathcal{S}$	Bloques horarios factibles para la clase $c \in \mathcal{C}$ .
$V^c$	Eventos de la clase $c \in \mathcal{C}$ .
$nd^c$	Periodicidad semanal de la clase $c \in \mathcal{C}$ .
$nv^c$	Cantidad de eventos consecutivos en un mismo día de la clase $c \in \mathcal{C}$ .
$Init^c$	Primer evento de cada bloque de eventos consecutivos de la clase $c \in \mathcal{C}$ .
$R^c \subseteq \mathcal{R}$	Conjunto de aulas factibles para la clase $c \in \mathcal{C}$ .
$q^c$	Cupo de la clase $c \in \mathcal{C}$ .
$nreq^m$	Cantidad de alumnos que deben cursar la materia $m \in \mathcal{M}$ .

■ Alumnos

$\mathcal{A}$	Conjunto de alumnos.
$rk^a$	Desempeño académico del alumno $a \in \mathcal{A}$ .
$\bar{m}^a$	Cantidad máxima de materias a ser asignadas al alumno $a \in \mathcal{A}$ .
$m^a$	Cantidad mínima de materias a ser asignadas al alumno $a \in \mathcal{A}$ .
$M^a \subseteq \mathcal{M}$	Conjunto de materias elegidas por el alumno $a \in \mathcal{A}$ .
$G^a$	Conjunto de grupos de electivas elegidas por el alumno $a \in \mathcal{A}$ .
$w^{am}$	Valor de preferencia para la materia $m \in M^a$ otorgado por el alumno $a \in \mathcal{A}$ .
$\bar{m}^{ag}$	Cantidad máxima de materias del grupo $g \in G^a$ que puede cursar el alumno $a \in \mathcal{A}$ .

- Docentes

$\mathcal{F}$	Conjunto de docentes.
$S^f \subseteq \mathcal{S}$	Disponibilidad de bloques horarios del docente $f \in \mathcal{F}$ .
$C^f \subseteq \mathcal{C}$	Clases a cargo del docente $f \in \mathcal{F}$ .
$ndays^f$	Cantidad máxima de días por semana de enseñanza del docente $f \in \mathcal{F}$ .

Consideramos los siguientes 6 conjuntos de variables binarias:

- $x_{ac}$  toma valor 1 si el alumno  $a$  es asignado a la clase  $c$ . Esta variable está definida sólo para aquellas clases  $c$  que correspondan a materias que se encuentran en la lista de preferencias del alumno  $a$ .
- $xs_{acs}$  toma valor 1 si la clase  $c$  asignada al alumno  $a$  es programada en el bloque horario  $s$ . Esta variable está definida sólo para aquellas clases  $c$  que correspondan a materias que se encuentran en la lista de preferencias del alumno  $a$  y para bloques horarios  $s$  disponibles del docente a cargo de la clase  $c$ .
- $y_{cvs}$  toma valor 1 si el evento  $v$  de la clase  $c$  es programado en el bloque horario  $s$ . Esta variable está definida sólo para los bloques horarios  $s$  disponibles del docente a cargo de la clase  $c$ .
- $yr_{crs}$  toma valor 1 si la clase  $c$  es asignada al aula  $r$  en el bloque horario  $s$ . Esta variable está definida sólo para los bloques horarios  $s$  disponibles del docente a cargo de la clase  $c$  y para las aulas  $r$  que tengan el tipo requerido por la clase.
- $w_c$  toma valor 1 si la clase  $c$  es programada en el semestre.
- $z_{fd}$  toma valor 1 si el docente  $f$  tiene asignada una clase el día  $d$ . Esta variable está definida sólo para los días  $d$  en que haya bloques horarios disponibles para el docente  $f$ .

Las restricciones que describimos a continuación modelan muchas de las usuales que aparecen en la literatura así como también algunas particulares para nuestro problema.

### 3.1. Asignación alumno-clases

Según los esquemas de las materias, toda materia tiene clases teóricas por lo cual consideramos que un alumno está inscripto en una materia cuando está asignado a una de sus clases teóricas. Teniendo esto en cuenta, las siguientes restricciones modelan la asignación de materias para cada alumno:

$$\sum_{c \in T^m} x_{ac} \leq 1 \quad a \in \mathcal{A}, m \in M^a \quad (1)$$

$$\sum_{c \in Lec^q} x_{ac} = \sum_{c \in Ex^q} x_{ac} \quad a \in \mathcal{A}, m \in M^a, q \in Q^m, Ex^q \neq \emptyset \quad (2)$$

$$\sum_{c \in Lec^q} x_{ac} = \sum_{c \in Lab^q} x_{ac} \quad a \in \mathcal{A}, m \in M^a, q \in Q^m, Lab^q \neq \emptyset \quad (3)$$

$$\underline{m}^a \leq \sum_{\substack{c \in T^m \\ m \in M^a}} x_{ac} \leq \bar{m}^a \quad a \in \mathcal{A} \quad (4)$$

$$\sum_{\substack{c \in T^m \\ m \in g}} x_{ac} \leq \bar{m}^{ag} \quad a \in \mathcal{A}, g \in G^a \quad (5)$$

$$\sum_{a \in \mathcal{A}} x_{ac} \leq q^c w_c \quad c \in \mathcal{C} \quad (6)$$

Las restricciones (1) establecen que cada alumno puede estar asignado a lo sumo a una teórica de una única comisión de cada materia de su lista de preferencias. Si se requiere una asignación obligatoria, entonces la restricción vale por igualdad. Las restricciones (2) y (3) fuerzan la asignación a un esquema factible, es decir que si un alumno está asignado a una teórica de una comisión de una materia, entonces también deberá estar asignado a una clase práctica de la misma comisión y una de laboratorio, respectivamente. Las restricciones (4) aseguran un número mínimo y máximo de materias, y (5) es análogo para materias electivas que pertenecen a un mismo grupo. Finalmente, las restricciones (6) impiden que no se asigne más alumnos que el cupo permitido.

### 3.2. Factibilidad temporal

Las restricciones que describimos a continuación modelan la asignación de bloques horarios a clases impidiendo superposición horaria cuando corresponde.

$$\sum_{s \in Slots^c} y_{cvs} = w_c \quad c \in \mathcal{C}, v \in V^c \quad (7)$$

$$\sum_{s \in S^d} \sum_{v \in Init^c} y_{cvs} \leq 1 \quad d \in \mathcal{D}, c \in \mathcal{C} \quad (8)$$

$$\sum_{s=s_0+1}^{s_0+nv^c-1} \sum_{v=v_i+1}^{v_i+nv^c-1} y_{cvs} - (nv^c - 1)y_{cv_i s_0} = 0 \quad d \in \mathcal{D}, s_0 \in Slots^c,$$

$$s_0 \leq \bar{s}^d - nv^c,$$

$$c \in \mathcal{C}, v_i \in Init^c \quad (9)$$

$$\sum_{s \in S^d} \sum_{v \in Init^c} y_{cvs} + \sum_{s \in S^{d+1}} \sum_{v \in Init^c} y_{cvs} \leq 1 \quad d \in \mathcal{D}, d \neq \bar{d}, c \in \mathcal{C} \quad (10)$$

Las restricciones (7) programan los eventos de una clase en exactamente un bloque horario entre los factibles. Las restricciones (8) aseguran la frecuencia semanal y (9) que los eventos de un mismo día sean consecutivos. Por último, las restricciones (10) garantizan que las clases no se dicten en días consecutivos (para aquellas materias que así lo requieren).

### 3.3. Factibilidad de aulas

La asignación de aulas debe satisfacer restricciones estándares: no exceder la capacidad y no tener superposición horaria. Esto es establecido por las restricciones (12) y (13). La asignación de aula a una clase en un bloque horario es determinada por las restricciones (11) y (14) asegura que no haya cambio de aula dentro de un mismo día.

$$\sum_{r \in R^c} yr_{crs} = \sum_{v \in V^c} y_{cvs} \quad c \in \mathcal{C}, s \in Slots^c \quad (11)$$

$$\sum_{c \in \mathcal{C}, r \in R^c} yr_{crs} \leq 1 \quad r \in \mathcal{R}, s \in \mathcal{S} \quad (12)$$

$$\sum_{a \in \mathcal{A}} xs_{acs} \leq \sum_{r \in R^c} cap^r yr_{crs} \quad c \in \mathcal{C}, s \in Slots^c \quad (13)$$

$$(nv^c - 1)(y_{cv_i s_0} + yr_{crs_0} - 1) \leq \sum_{s=s_0+1}^{s_0+nv^c-1} yr_{crs} \quad d \in \mathcal{D}, s_0 \in Slots^c,$$

$$s_0 \leq \bar{s}^d - nv^c, c \in \mathcal{C},$$

$$v_i \in Init^c, r \in R^c \quad (14)$$

### 3.4. Factibilidad de cursada para alumnos

Cuando a un alumno se le asignan materias, las clases correspondientes deben estar programadas en bloques horarios que no estén superpuestos. Esto garantiza que el alumno pueda asistir a todas las clases.

$$y_{cvs} + x_{ac} - xs_{acs} \leq 1 \quad a \in \mathcal{A}, m \in M^a, c \in Cl^m \\ v \in V^c, s \in Slots^c \quad (15)$$

$$2xs_{acs} - \sum_{v \in V^c} y_{cvs} - x_{ac} \leq 0 \quad a \in \mathcal{A}, m \in M^a, c \in Cl^m \\ s \in Slots^c \quad (16)$$

$$\sum_{\substack{c \in Cl^m \\ m \in M^a}} xs_{acs} \leq 1 \quad a \in \mathcal{A}, s \in \mathcal{S} \quad (17)$$

$$\sum_{c \in Cl^m} \sum_{s \in S^d} xs_{acs} \leq 2 \quad a \in \mathcal{A}, m \in M^a, d \in \mathcal{D} \quad (18)$$

$$\sum_{m \in M^a} \sum_{c \in Cl^m} xs_{acs^d} + \sum_{m \in M^a} \sum_{c \in Cl^m} xs_{acs^{\bar{d}}} \\ - \sum_{\substack{s \in S^d \\ s \neq \bar{s}^d, \bar{s}^d}} \sum_{m \in M^a} \sum_{c \in Cl^m} xs_{acs} \leq 1 \quad a \in \mathcal{A}, d \in \mathcal{D} \quad (19)$$

Las restricciones (15) y (16) relacionan las variables  $x_{ac}$ ,  $y_{cvs}$  y  $xs_{acs}$ . Las restricciones (17) fuerzan que un alumno no tenga superposición horaria y (18) establecen que dentro de un día no tenga más de dos eventos por materia (restricción impuesta por cuestiones pedagógicas por la universidad).

Finalmente, para evitar muchas horas ociosas en la universidad, las restricciones (19) exigen que si un alumno tiene clases el primer y último bloque horario de un día, entonces al menos tenga algún otro bloque horario ocupado el mismo día.

### 3.5. Factibilidad docente

Basado en la información sobre disponibilidad horaria de cada docente y teniendo en cuenta que sus clases no deben solaparse, las restricciones (20) aseguran factibilidad horaria. Además, las restricciones (21) y (22) imponen cotas superiores a la cantidad de días dedicados a la docencia.

$$\sum_{c \in C^f} \sum_{v \in V^c} y_{cvs} \leq 1 \quad f \in \mathcal{F}, s \in S^f \quad (20)$$

$$\sum_{c \in C^f} \sum_{v \in \text{Init}^c} \sum_{s \in S^d} y_{cvs} \leq |S^d| z_{fd} \quad f \in \mathcal{F}, d \in \mathcal{D} \quad (21)$$

$$\sum_{d \in \mathcal{D}} z_{fd} \leq n \bar{d} a y s^f \quad f \in \mathcal{F} \quad (22)$$

### 3.6. Función objetivo

El objetivo a maximizar es la satisfacción global del alumnado. Para esto, para cada alumno  $a \in \mathcal{A}$ , consideramos una suma ponderada entre su preferencia ( $w_c^{ac} = w^{am}$ ) por la materia  $m \in M^a$  y su desempeño académico ( $r k^a$ ) definido por la universidad. Dado que la asignación de un alumno a una materia está definida por su asignación a una de sus clases teóricas, las variables  $x_{ac}$  y parámetros no negativos  $\alpha$  y  $\beta$ , nos permiten modelar una función para maximizar la *calidad* de las asignaciones.

$$\text{(GPECTP-ILP)} \quad \text{máx } z_{ILP} = \sum_{a \in \mathcal{A}} \sum_{\substack{c \in T^m \\ m \in M^a}} (\alpha r k^a + \beta w_c^{ac}) x_{ac} \quad (23)$$

$$\text{s.t.: (1) - (22)}$$

$$x_{ac}, x_{sacs}, y_{cvs}, y_{rcrs}, z_{fd}, w_c \in \{0, 1\}$$

La formulación tiene un gran número de variables y restricciones que hace difícil su resolución usando software de propósito general como CPLEX. Nuestra experiencia computacional nos mostró que CPLEX tiene dificultades tanto en encontrar soluciones factibles como en resolver las relajaciones. Por otro lado, las técnicas de preprocesamiento no logran una reducción significativa del tamaño del modelo. Como mencionamos anteriormente, debido a que la asignación a materias y a horarios se realiza en forma conjunta, hay muy pocas incompatibilidades entre variables que pueda ser deducida a priori, como ocurre en otros problemas de asignación horaria.

---

## 4. Metodología

---

Debido a las dificultades mencionadas, desarrollamos un proceso en dos fases basado en el modelo GPECTP-ILP de programación entera. La idea general del proceso es, en una primera etapa, considerar una relajación GPECTP-RED del modelo que incluye un subconjunto de alumnos y relaja algunas de las restricciones. En una segunda etapa, a partir de la asignación horaria obtenida en la primera etapa, se considera el modelo original GPECTP-ILP fijando las variables correspondientes a asignación de bloque horario a las clases y se busca una asignación de materias a alumnos que maximice la preferencia global.

Debido a que en la primera etapa consideramos una relajación del problema y sólo un subconjunto de alumnos, la segunda etapa puede brindar soluciones con un valor bajo en la función objetivo o, aún peor, resultar infactible. Para aumentar las posibilidades de obtener una buena asignación, en forma aleatoria generamos  $l$  subconjuntos de alumnos  $\bar{A}_1, \dots, \bar{A}_l$  usados en forma independiente para generar distintas programaciones horarias en la primera etapa. Además, cada vez que se resuelve GPECTP-RED para cada subconjunto, generamos más de una posible solución. Esto lo realizamos utilizando la rutina *populate* de CPLEX que permite generar óptimos alternativos o soluciones cercanas al óptimo dentro de un gap determinado. Los detalles del algoritmo utilizado por CPLEX pueden ser consultados en [6].

Para la construcción de los subconjuntos  $\bar{A}_1, \dots, \bar{A}_l$  seguimos la siguiente estrategia. Inicialmente todos los subconjuntos tienen a los alumnos:

1. cuyo su desempeño académico se encuentra dentro del  $K\%$  superior del alumnado
2. que no fueron incluidos en 1 pero que tienen dentro de su lista de preferencia materias de asignación obligatoria. Para ellos, incluimos en su lugar un alumno *ficticio* con sólo las materias obligatorias.

Finalmente, los alumnos no considerados en 1 ó 2 son asignados equitativamente en forma aleatoria a uno de los  $l$  subconjuntos. En caso que el alumno ficticio de 2 pertenezca al mismo subconjunto que el alumno original, se elimina del subconjunto.

En Algoritmo 1 mostramos un bosquejo del proceso. Como parámetros de entrada, además de los datos que definen la instancia, llamamos  $l$  al número de subconjuntos,  $N$  al número máximo de soluciones del proceso *populate*,  $TILIM$  al tiempo máximo de ejecución de la segunda etapa, y  $K$  al porcentaje usado para la construcción de los subconjuntos.

---

**Algoritmo 1** BOSQUEJO DEL ALGORITMO

---

**Entrada:** Instancia,  $l$ ,  $N$ ,  $TILIM$ ,  $K$

**Salida:** Conjunto  $\Delta$  de programaciones horarias factibles de GPECTP.

1. Formular el modelo GPECTP-ILP. Computar  $UB$  como el valor óptimo de la relajación lineal.
  2. Construir  $\bar{A}_1, \dots, \bar{A}_l$  de acuerdo a  $K$ .
  3. Sea  $BESTLB = -\infty$  y  $\Delta = \emptyset$  el *pool* de soluciones factibles.
  4. **Para**  $j = 1, \dots, l$  **hacer:**
    5. Formular el modelo ILP de la fase 1 para el subconjunto  $\bar{A}_j$ , resolver y generar  $N$  asignaciones horarias a clases usando la rutina *populate* de CPLEX. Sea  $\Gamma_j$  el *pool* de soluciones.
    6. **Para**  $i \in \Gamma_j$  **hacer:**
      7. Sea  $x^{(i)}$  la  $i$ -ésima solución de la primera fase. Construir el modelo ILP fijando las variables de asignación de horario a clases. Sea  $UB^{(i)}$  el valor objetivo de la relajación.
      8. Si  $UB^{(i)} \leq BESTLB$ , entonces omitir esta solución y retornar a 7 con la solución  $i + 1$  de  $\Gamma$ .
      9. Si no, resolver ILP por a lo sumo  $TILIM$  segundos. Sea  $z^{(i)}$  el valor de la función objetivo de la mejor solución encontrada durante la etapa 2. Si  $z^{(i)} > BESTLB$ , entonces fijar  $BESTLB = z^{(i)}$ , guardar la solución en  $\Delta$  y continuar con la solución  $i + 1$  de  $\Gamma_j$ .
  10. **fin Para**
  11. **fin Para**
  12. Reportar todas las soluciones de  $\Delta$ .
- 

A continuación describimos más detalles de cada etapa.

▪ **Etapas 1**

El propósito de esta etapa es generar programaciones horarias para alimentar la segunda etapa utilizando una relajación del problema, que llamaremos GPECTP-RED, que sea tratable computacionalmente. Con este fin, como ya mencionamos, consideramos un subconjunto de alum-

nos  $\bar{A} \subseteq \mathcal{A}$ . Además descartamos las variables y restricciones de asignación de aulas, reemplazándolas por la restricción descrita a continuación. Si  $nr^i$  es la cantidad de aulas disponibles de tipo  $i \in Types$  y  $Classes^i \subseteq \mathcal{C}$  el conjunto de clases a las que debe asignarse un aula de tipo  $i$ , entonces, para cada bloque horario, restringimos que la cantidad de clases programadas en el bloque no puede exceder  $nr^i$ . Claramente esto es una relajación del problema y no garantiza que pueda realizarse una asignación de aulas, sin embargo reduce sustancialmente el tamaño del modelo y genera soluciones parciales de buena calidad.

Por otro lado, dado que el conjunto de alumnos es reducido, no hay seguridad que todas las clases sean programadas y esto puede influir negativamente en las características de la solución que alimentará a la segunda etapa. Para salvar esta dificultad, consideramos una función objetivo (24) alternativa que combina en forma ponderada la función objetivo del modelo general restringida al conjunto de alumnos  $\bar{A}$  y el número de clases programadas.

$$\begin{aligned}
 \text{(GPECTP-RED)} \quad & \text{máx } \gamma \sum_{c \in T^m} w_c + \delta z_{ILP} & (24) \\
 \text{s.t.:} & (1) - (10), (15) - (22), \\
 & \text{restringido al conjunto } \bar{A} \\
 & \sum_{c \in Classes^i} y_{cvs} \leq nr^i \quad \forall i \in Types, s \in \mathcal{S} \\
 & x_{ac}, x_{sacs}, y_{cvs}, z_{fd}, w_c \in \{0, 1\}
 \end{aligned}$$

## ■ Etapa 2

La segunda etapa está enfocada en obtener una asignación de materias a alumnos y una asignación de aulas a materias a partir de una diagramación horaria obtenida en la etapa 1.

Formalmente, sea  $\pi^* = (x^*, xs^*, y^*, z^*, w^*)$  una solución de la primera etapa que representa una asignación factible de bloques horarios a clases. Entonces, la segunda etapa consiste en formular y resolver el modelo GPECTP-ILP fijando las variables  $y = y^*$ ,  $w = w^*$  y  $z = z^*$ .

Esta segunda etapa tiene dos condiciones posibles de parada. Por un lado, imponemos un tiempo límite de ejecución por cada diagramación de la etapa 1 (*TILIM*) establecido por el usuario. La segunda condición determina que, si la relajación del modelo tiene un valor objetivo inferior a la mejor diagramación obtenida hasta el momento, el proceso para.

Esto evita perder tiempo en la búsqueda de asignaciones que no pueden mejorar la solución actual.

Cabe mencionar que en la etapa 2 podría ocurrir que el modelo resulte no factible debido a que no se pudiera satisfacer la cantidad mínima de asignaciones para cada alumno. En este caso se descarta la asignación de bloques horarios a clases y se procede con la próxima solución obtenida en la etapa 1. Es claro que la elección de  $\bar{A}$  es un punto clave para la calidad de la solución que alimenta la etapa 2. A mayor cantidad de alumnos que conformen el subconjunto  $\bar{A}$ , mejor la calidad de la solución. Sin embargo, esto va acompañado de un mayor incremento en los requerimientos del tiempo computacional de la etapa 1.

---

## 5. Experiencia computacional

---

Con el objetivo de evaluar la efectividad de la propuesta, consideramos 3 instancias reales correspondientes al primer semestre del 2013 (1-2013), segundo semestre del 2014 (2-2014) y primer semestre del 2015 (1-2015). Los experimentos fueron realizados en una workstation Intel(R) Core(TM) i7 CPU (3.40GHz) y 16 Gb de RAM. El algoritmo fue programado en C++ usando CPLEX 12.1 Callable Library como LP y MIP solver. Los parámetros  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$  los fijamos en 1. Cabe señalar que realizamos experimentos con otros valores pero no los incluimos en este reporte ya que el algoritmo no presenta diferencias en su performance. Estos parámetros cumplen la función de ser una herramienta que el usuario tiene para ponderar los diferentes términos de la función objetivo para considerar distintos escenarios.

En la Tabla 1 mostramos un resumen de las características de las instancias consideradas, tales como cantidad de alumnos, promedio de cantidad de materias en la lista de preferencias, cantidad total de materias, etc. En las últimas dos columnas, informamos la cantidad de variables y de restricciones del modelo GPECTP-ILP.

El desempeño académico de cada alumno, así como también la preferencia de materias, es un valor descendente entre 10 y 1 y se consideran 5 días divididos en 5 bloques horarios.

Para los experimentos utilizamos la configuración por defecto que trae CPLEX, imponiendo un tiempo límite de 1000s para cada subconjunto. Para evaluar la calidad de las soluciones encontradas, ejecutamos el algoritmo con  $l = 0, 5, 10, 20$  y  $K = 0\%, 5\%, 10\%$ . La calidad de la solución es medida conforme a la satisfacción global del alumnado con respecto a sus preferencias.

Inst	$ \mathcal{A} $	$\text{av}\bar{m}^a$	$\text{av} M^a $	$ \mathcal{M} $	$\sum_{m \in \mathcal{M}}  Q^m $	$ \mathcal{C} $	$\sum_{c \in \mathcal{C}}  V^c $	$ \mathcal{S} $	$ \mathcal{F} $	$\text{av} S^f $	$ \mathcal{R} $	vars	const
1-2013	981	4.15	7	118	131	341	473	30	220	7.55	31	310063	751086
2-2014	915	4.24	6.88	122	136	335	461	36	211	8.66	31	259395	769712
1-2015	1065	4.12	6.31	111	118	329	450	36	216	9.5	31	303229	804693

Tabla 1: Descripción de las instancias 1-2013, 2-2014 y 1-2015.

Los resultados los presentamos en las tablas 2, 3 y 4. Para cada combinación, informamos la cantidad de soluciones factibles encontradas, ( $|\Delta|$ ), el valor de la función objetivo y la cantidad de asignaciones de la mejor solución encontrada con sus correspondientes *gaps* y tiempo total de cómputo (en segundos). Para el cálculo del *gap* utilizamos el valor de la relajación de GPECTP-ILP, adecuando la función objetivo en cada caso. Si se excede la disponibilidad de memoria para cierta combinación de parámetros, en la tabla aparece *ml*.

Los resultados obtenidos son los esperados. Cuanto más grande sean  $K$  y  $l$ , mejor calidad de resultado. Por ejemplo, los resultados correspondientes al primer semestre del 2013, muestran que para  $l = 5, 10$  se generan subconjuntos  $\bar{A}$  de tamaño grande y GPECTP-RED resulta difícil de resolver, excediendo la capacidad de memoria. Cuando  $l = 0$  (un único conjunto, sólo alumnos con materias obligatorias y los que tengan desempeño académico dentro del  $K\%$  superior del alumnado), la etapa 1 encuentra programaciones horarias para alimentar a la etapa 2. Sin embargo, la calidad de la solución es inferior a la obtenida con  $l = 20$ . Esto era previsible ya que los subconjuntos para  $l = 20$  incluyen más alumnos. Respecto a los valores de  $K$ , también puede notarse que se incrementa la calidad cuando incrementamos el valor  $K$ . Claramente, la mejora en la calidad es a expensas de un mayor tiempo computacional. Sin embargo, dado el contexto de la aplicación, son tiempos de cómputo muy aceptables.

Conclusiones similares pueden hacerse para las instancias del segundo semestre de 2014 y primer semestre de 2015, si bien para estos casos el algoritmo pudo encontrar soluciones para más combinaciones de parámetros. Analizando los datos, notamos que para estas instancias la disponibilidad horaria de los docentes era mayor, facilitando la factibilidad de la programación horaria.

$l/K$	0%						5%					
	$ \Delta $	Obj	Gap	Asig	Gap	Tm	$ \Delta $	Obj	Gap	Asig	Gap	Tm
0	2	50514	6.61	3898	4.01	1680	3	51421	4.95	3929	3.25	1500
5	ml	-	-	-	-	-	ml	-	-	-	-	-
10	ml	-	-	-	-	-	1	52449	3.03	3999	1.53	63902
20	8	52232	3.44	3979	2.01	8580	9	52457	3.02	3997	1.58	67500

Tabla 2: Función objetivo y tiempo computacional para la instancia real 1-2013 para diferentes valores de  $K$  y  $l$ .

$l/K$	0%						5%						10%					
	$ \Delta $	Obj	Gap	Asig	Gap	Tm	$ \Delta $	Obj	Gap	Asig	Gap	Tm	$ \Delta $	Obj	Gap	Asig	Gap	Tm
0	3	46199	9.44	3635	6.25	1245	4	48588	4.75	3750	3.02	1457	5	48657	4.62	3766	2.61	1977
5	ml	-	-	-	-	-	ml	-	-	-	-	-	ml	-	-	-	-	-
10	3	48926	4.09	3773	2.43	19324	ml	-	-	-	-	-	ml	-	-	-	-	-
20	13	49059	3.83	3769	2.53	11694	8	49281	3.4	3792	1.93	35575	8	49356	3.25	3790	1.99	81934

Tabla 3: Función objetivo y tiempo computacional para la instancia real 2-2014 para diferentes valores de  $K$  y  $l$ .

$l/K$	0%						5%						10%					
	$ \Delta $	Obj	Gap	Asig	Gap	Tm	$ \Delta $	Obj	Gap	Asig	Gap	Tm	$ \Delta $	Obj	Gap	Asig	Gap	Tm
0	2	45321	10.84	3935	8.68	9159	3	46391	8.74	4039	6.27	3099	4	46021	9.47	3994	7.31	9661
5	ml	-	-	-	-	-	ml	-	-	-	-	-	ml	-	-	-	-	-
10	3	47040	7.46	4065	5.66	56214	2	46946	7.65	4052	5.96	66772	ml	-	-	-	-	-
20	5	47366	6.82	4104	4.76	77646	11	47482	6.59	4119	4.41	77575	8	47833	5.9	4140	3.92	102042

Tabla 4: Función objetivo y tiempo computacional para la instancia real 1-2015 para diferentes valores de  $K$  y  $l$ .

Con el objetivo de hacer un análisis más detallado del comportamiento del algoritmo, experimentamos con el primer semestre del 2013 con  $l = 0$  y considerando disponibilidad de todos los docentes en todos los bloques horarios.

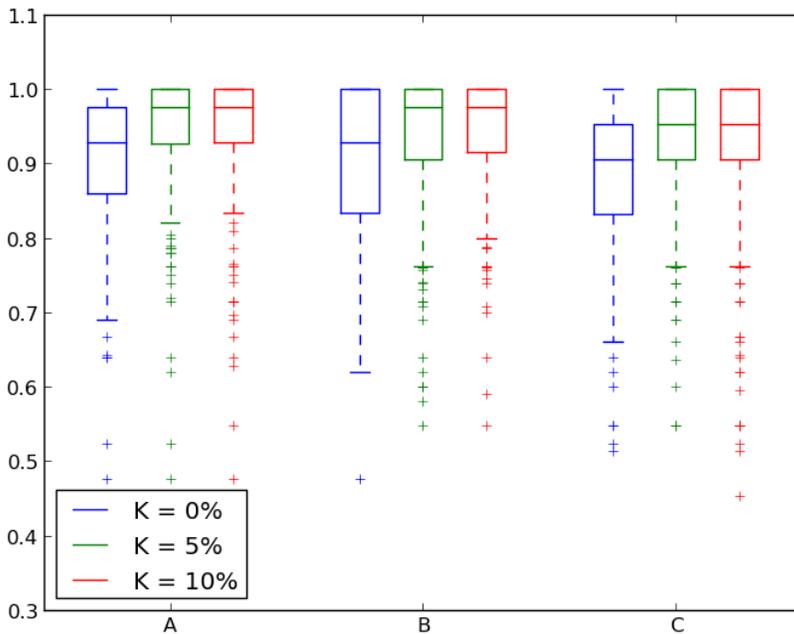
El próximo experimento está enfocado en analizar la satisfacción del alumno. Para esto, para cada alumno  $a \in \mathcal{A}$ , calculamos su valor de satisfacción como

$$sv_a = \sum_{m \in M^a} w^{am} x_{am}. \tag{25}$$

y además, calculamos una cota superior de la satisfacción por alumno considerando el valor de la relajación lineal de GPECTP-ILP con función objetivo (25). El cociente entre estos dos valores lo tomaremos como medida de satisfacción del alumno. En la figura 2 presentamos un diagrama de cajas donde particionamos a los alumnos en tres conjuntos (A,B,C) de igual cardinal ordenados por desempeño académico. Una primera observación es que el nivel de

satisfacción se incrementa cuando  $K$  crece, con mayor valor de la media y menos dispersión. Esto es consistente con los resultados mostrados previamente. Por otro lado, también puede observarse que el nivel de satisfacción es mayor para los alumnos con mejor desempeño académico.

Figura 2: Satisfacción del alumnado según desempeño.



Finalmente, realizamos un último experimento para evaluar el impacto de haber tenido en cuenta las preferencias de alumnos y su desempeño académico. Para esto, generamos un escenario donde la función objetivo considera que todos los alumnos tienen igual desempeño y todas las materias tienen igual preferencia. A posteriori, calculamos cuál es la satisfacción global de estas soluciones considerando las reales preferencias y los desempeños académicos y nos quedamos con la mejor.

En la tabla 5 presentamos los resultados para las 3 instancias y el tiempo de cómputo.

Los resultados muestran que haber considerado las preferencias tiene un impacto positivo en la calidad de las soluciones. En todos los casos, la satisfacción se mejora alrededor de un 10 %, no obstante a un costo computacional mayor pero dentro de niveles razonables para la aplicación.

Inst	$l$	$K$	sin prefs.		con prefs.	
			Obj	Tiempo	Obj	Tiempo
1-2013	20	5	48435	14700	52457	67500
2-2014	20	10	45597	18660	49356	81934
1-2015	20	10	44460	49020	47833	102042

Tabla 5: Análisis de satisfacción.

Por último queremos destacar que la metodología propuesta ha tenido un efecto positivo en la gestión de la universidad. Por un lado, la diagramación horaria que se llevaba a cabo en forma manual e insumía cerca de 2 meses, se redujo a unos pocos días y el responsable de realizar el proceso dispone de una herramienta que le permite analizar diferentes escenarios en poco tiempo. En general, a los 30 minutos de ejecución se obtiene la primera solución y, en promedio, después de 3 horas se cuenta con 10 soluciones. El tiempo total del proceso es de alrededor de 10 horas.

Por otro lado, los alumnos tienen una mayor libertad en elegir las materias y organizar sus estudios teniendo ingerencia en la asignación a partir de sus listas de preferencias. Como mencionamos al comienzo, la asignación se hacía en forma compulsiva y los alumnos reclamaban cambios al conocer las materias que debían cursar. A partir de la implementación del nuevo proceso, las demandas de cambios se redujeron notablemente. Esto influye no sólo en la satisfacción del alumnado sino también en que la universidad puede usar sus recursos en forma más eficiente.

Todo el proceso de relevamiento y definición del problema se llevó a cabo con la cooperación activa del responsable de la diagramación horaria, así como también con el departamento de sistemas de la universidad. El compromiso de las autoridades y personal de la universidad en llevar adelante un cambio tecnológico y de paradigma en la diagramación horaria fue fundamental para el éxito del proyecto.

---

## 6. Conclusiones y líneas de trabajo futuro

---

En este trabajo consideramos un problema de diagramación horaria proveniente de una aplicación real de una universidad privada de Argentina. Hasta donde llega nuestro conocimiento, el proceso propuesto está basado en un enfoque no contemplado en la literatura del área.

Desarrollamos un procedimiento heurístico en dos fases, basados en mode-

los de programación lineal entera que contemplan las restricciones operacionales así como también las preferencias de los alumnos.

La heurística fue analizada en instancias reales, obteniendo soluciones de muy buena calidad, con *gaps* del orden del 3-4% y resultó una herramienta muy útil para la gestión de la universidad.

Como trabajo futuro, además de incorporar algunas otras restricciones como disponibilidad horaria por aula y por alumno, podría ser de interés mejorar el comportamiento por defecto de CPLEX para reducir los tiempos de cómputo. Entre las posibilidades destacamos caracterizar desigualdades válidas e implementar rutinas de separación y heurísticas primales.

**Agradecimientos:** Este trabajo fue parcialmente financiado por la Universidad de Buenos Aires, Programación Científica 2014-2017, UBACyT 20020100100666.

## Referencias

- [1] H. Babaei, J. Karimpour, y A. Hadidi. A survey of approaches for university course timetabling problem. *Computers & Industrial Engineering*, 86:43–59, 2015.
- [2] A. Bettinelli, V. Cacchiani, R. Roberti, y P. Toth. An overview of curriculum-based course timetabling. *TOP*, 23(2):313–349, 2015.
- [3] E.K. Burke, J. Mareček, A.J. Parkes, y H. Rudová. A branch-and-cut procedure for the udine course timetabling problem. *Annals of Operations Research*, 194(1):71–87, 2012.
- [4] E.K. Burke, N. Pham, R. Qu, y J. Yellen. Linear combinations of heuristics for examination timetabling. *Annals of Operations Research*, 194(1):89–109, 2012.
- [5] H. Cambazard, E. Hebrard, B. O’Sullivan, y A. Papadopoulos. Local search and constraint programming for the post enrolment-based course timetabling problem. *Annals of Operations Research*, 194(1):111–135, 2012.
- [6] E. Danna, M. Fenelon, Z. Gu, y R. Wunderling. Generating multiple solutions for mixed integer programming problems. In *International Conference on Integer Programming and Combinatorial Optimization*, páginas 280–294. Springer, 2007.

- [7] S. Daskalaki y T. Birbas. Efficient solutions for a university timetabling problem through integer programming. *European Journal of Operational Research*, 160(1):106–120, 2005.
- [8] C. Gogos, P. Alefragis, y E. Housos. An improved multi-staged algorithmic process for the solution of the examination timetabling problem. *Annals of Operations Research*, 194(1):203–221, 2012.
- [9] S.N. Jat y S. Yang. A hybrid genetic algorithm and tabu search approach for post enrolment course timetabling. *Journal of Scheduling*, 14(6):617–637, 2011.
- [10] S. Kristiansen, M. Sørensen, y T.R. Stidsen. Integer programming for the generalized high school timetabling problem. *Journal of Scheduling*, 18(4):377–392, 2015.
- [11] S. Kristiansen y T. R. Stidsen. A comprehensive study of educational timetabling - a survey. Technical report, Department of Management Engineering, Technical University of Denmark, 2013.
- [12] G. Lach y M.E. Lübbecke. Curriculum based course timetabling: new solutions to udine benchmark instances. *Annals of Operations Research*, 194(1):255–272, 2012.
- [13] R. Lewis. A time-dependent metaheuristic algorithm for post enrolment-based course timetabling. *Annals of Operations Research*, 194(1):273–289, 2012.
- [14] Z. Lü y J. Hao. Adaptive tabu search for course timetabling. *European Journal of Operational Research*, 200(1):235–244, 2010.
- [15] M.E. Lübbecke et al. Comments on: An overview of curriculum-based course timetabling. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 23(2):359–361, 2015.
- [16] B. McCollum, A. Schaerf, B. Paechter, P. McMullan, R. Lewis, A.J. Parkes, L. Gaspero, R. Qu, y E.K. Burke. Setting the research agenda in automated timetabling: The second international timetabling competition. *INFORMS Journal on Computing*, 22(1):120–130, 2010.
- [17] C. Nothegger, A. Mayer, A. Chwatal, y G.R. Raidl. Solving the post enrolment course timetabling problem by ant colony optimization. *Annals of Operations Research*, 194(1):325–339, 2012.

- [18] M. Sørensen y F. Dahms. A two-stage decomposition of high school timetabling applied to cases in Denmark. *Computers and Operations Research*, 43:36–49, 2014.
- [19] J. Van den Broek, C. Hurkens, y G. Woeginger. Timetabling problems at the TU Eindhoven. *European Journal of Operational Research*, 196(3):877–885, 2009.
- [20] J.J.J. Van den Broek y C.A.J. Hurkens. An IP-based heuristic for the post enrolment course timetabling problem of the ITC2007. *Annals of Operations Research*, 194(1):439–454, 2012.

---

# PROCESS MINING PARA GESTIÓN DE INCIDENCIAS EN TELEFÓNICA CHILE

---

NÉSTOR CONTRERAS \*

CINTHYA VERGARA \*\*

YERKO MONTENEGRO \*\*

## Resumen

El Diseño de Procesos de Negocios con apoyo en técnicas de *Process Mining* es un área emergente en Chile. Las técnicas convencionales de modelamiento y Rediseño de Procesos conlleva un alto gasto de recursos y tiempo de desarrollo, además de grandes esfuerzos en gestión del cambio e implantación de mejoras. Telefónica Chile es una empresa de telecomunicaciones que ofrece servicios principalmente de redes fijas, móviles y de banda ancha. Dentro de su operación interna posee el Servicio de Atención de Incidencias donde se revisan y gestionan soluciones para problemas operativos de la compañía. En la búsqueda de una mejor operación general para Telefónica en su segmento Movistar se realizó un proyecto práctico, desde el planteamiento estratégico hasta el rediseño del proceso específico, donde se utilizaron técnicas y herramientas de *Process Mining*. Con ello, presentamos en este trabajo el modelamiento y Rediseño de Procesos basado en el uso de técnicas de *Process Mining*, donde se lograron mejoras cuantitativas relacionadas a una disminución en los tiempos de resolución en la gestión de incidencias, permitiendo además la identificación de problemas en el proceso y en la relación con usuarios en tiempo real. Las técnicas de *Process Mining* mostraron gran valor para la empresa, permitiendo acelerar los procesos de mejora del servicio, gestión de procesos y hacer efectivos los Service Level Agreement (SLA).

**Palabras Clave:** Process Mining, Gestión de Incidencias, Business Process Management, Mejora Continua, Data Science.

---

\*Telefónica Chile, Santiago, Chile

\*\*Departamento de Ingeniería Industrial ([www.dii.uchile.cl](http://www.dii.uchile.cl)), Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

---

## 1. Contexto y Motivación

---

Actualmente las empresas buscan soluciones que les permitan competir a escala global, adelantarse a nuevos eventos y adaptarse rápidamente a mercados cada vez más dinámicos. En particular, el sector de las telecomunicaciones continúa afrontando retos e incesantes cambios que han hecho que se haya redefinido completamente en un corto periodo de tiempo. Nuevos ecosistemas de negocio obligan a las empresas a adaptarse y transformarse ante las nuevas realidades. Esta necesidad imperiosa de transformación va unida a la de innovación. Las empresas tienen que poner en marcha ecosistemas que promuevan y favorezcan la innovación que les asegure el éxito en el largo plazo.

Telefónica es una empresa global de telecomunicaciones, que sustenta a dos marcas comerciales, Movistar y O2. Dentro de los servicios que entrega se encuentran soluciones de comunicación, información y entretenimiento, con presencia en Europa, África, Latinoamérica y, desde 2010, en Asia. Tiene presencia en 21 países y un promedio de 120.000 empleados. En Chile, tanto el mercado de telefonía móvil como fijo es dominado por Movistar-Telefónica con un 37,6 % del mercado en móvil y 44,6 % en fija. También mantiene la mayor participación de mercado en los servicios de Internet alcanzando un 37,6 % en móvil y un 38 % en fijo [13].

El uso de tecnología de información y comunicaciones (TICs) ha permitido encontrar métodos eficientes que forman parte de la estrategia de la compañía, esto ha logrado incrementos en la eficiencia, en la calidad de sus productos y servicios, y ha disminuido los tiempos de respuesta ante las necesidades de los clientes.

Para su operación cuenta con la Gerencia de Tecnología Operaciones y Sistemas, dependiente de la Dirección de Tecnología de Información, que tiene por principal función dar soporte a las aplicaciones que apalancan los procesos de negocio de la compañía, siendo varios de estos procesos los que generan mayores ingresos. Bajo esta gerencia se encuentra la Subgerencia de Soporte de Servicios TI, que tiene por función establecer procedimientos y protocolos para el manejo de situaciones de excepción que aseguren información oportuna a los niveles ejecutivos y técnicos, y coordinar la oportuna toma de acciones para restablecer los servicios.

Este proyecto se centró en implementar un modelo de control y seguimiento que evalúe las actividades de gestión de incidencias dependiente de la Subgerencia de Soporte de Servicios TI. El objetivo era que, la institución

logre mejores niveles de operación y de calidad de los servicios ofrecidos a los clientes internos. La motivación principal se fundamenta en el desconocimiento y descontrol de los tiempos de solución de incidencias, lo que conlleva un desmedro en la percepción del servicio interno. Además, esto se traduce en un alto costo debido a una mala planificación y coordinación de recursos.

El diseño del proceso de control y seguimiento de la gestión de incidencias se enfocó en optimizar y gestionar los problemas que se producen debido a la continuidad operacional que Telefónica-Movistar necesita asegurar. El desarrollo de la solución ligada al control y seguimiento de incidencias se hizo con la herramienta de minería de procesos Disco Miner [5] y ProM [20], herramientas con algoritmos y paneles de monitoreo especializados en análisis de procesos, las cuales se nutren en base a los *logs de eventos* del workflow que apoya al proceso de Gestión de Incidencias generados a partir de un proceso de extracción y carga estandarizado (ETL).

Finalmente, el rediseño del proceso y el uso de herramientas especializadas, permitió identificar de manera gráfica y numérica las interacciones entre los actores en la operación diaria del proceso y observar las relaciones entre actividades y recursos, sobrecargas en actores y cuellos de botellas en el proceso. Los usuarios vieron una propuesta de mejora coherente con la operación diaria que reflejó las necesidades de los distintos equipos de trabajo. Este enfoque ordena la búsqueda de soluciones, genera mejora continua en las actividades con un mayor dinamismo y permite gestionar los niveles de eficiencia y eficacia al momento de resolver una problemática dentro de la empresa. La metodología utilizada muestra ser efectiva tanto para el levantamiento y diagnóstico de procesos así como también para su rediseño y mejora continua.

---

## 2. Conceptos Clave

---

Este trabajo se realizó bajo la óptica de la Gestión de Procesos de Negocios (BPM, por su nombre en inglés Business Process Management) donde se entiende como un Proceso de Negocio al conjunto de tareas o actividades, lógicamente estructuradas, que ayudan a transformar varios insumos (input), en un producto final (output) por medio de la aplicación de varios factores, con el que se persigue obtener un valor para la organización. En ella se busca guiar a las organizaciones en el diseño, construcción e implementación de sus procesos, entendiendo a estos últimos como una pieza clave para el éxito en los negocios.

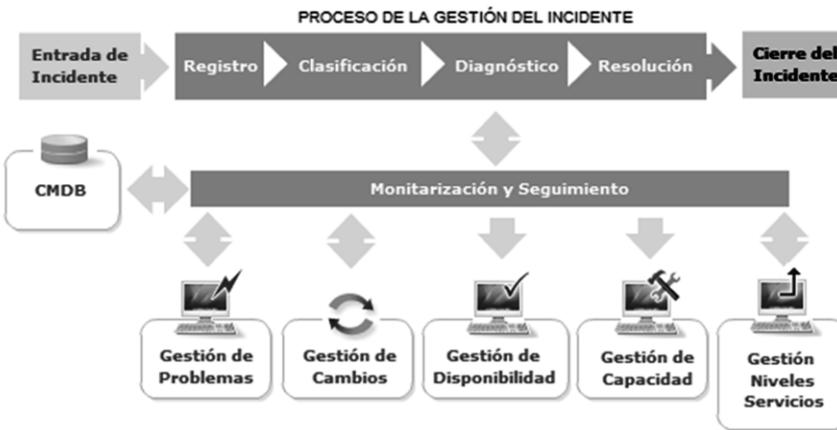
## 2.1. Gestión de Incidencias

La Gestión de Incidencias tiene como objetivo resolver cualquier incidente que cause una interrupción en el servicio de la manera más rápida y eficaz posible. Se preocupa principalmente de restaurar un servicio frente a cualquier situación que lo deje sin operar normalmente.

Para un área de Gestión de Incidencias es esencial generar información para, por ejemplo:

- La Gestión de Niveles de Servicio: es esencial que los clientes dispongan de información puntual sobre los niveles de cumplimiento de los SLAs y que se adopten medidas correctivas en caso de incumplimiento.
- Monitorizar el rendimiento del Centro de Servicios: conocer el grado de satisfacción del cliente por el servicio prestado y supervisar el correcto funcionamiento de la primera línea de soporte y atención al cliente.
- Optimizar la asignación de recursos: los gestores deben conocer si el proceso de escalado ha sido fiel a los protocolos preestablecidos y si se han evitado duplicidades en el proceso de gestión.
- Identificar errores: puede ocurrir que los protocolos especificados no se adecuen a la estructura de la organización o las necesidades del cliente por lo que se deban tomar medidas correctivas.
- Disponer de Información Estadística: que puede ser utilizada para hacer proyecciones futuras sobre asignación de recursos, costos asociados al servicio, etc.

Figura 1: Proceso de Gestión de Incidencias Telefónica Chile



Además, para el correcto seguimiento de todo el proceso es indispensable la utilización de métricas que permitan evaluar de la forma más objetiva posible el funcionamiento del servicio. Algunos de los aspectos clave a considerar son:

- Número de incidentes clasificados temporalmente y por prioridades.
- Tiempos de resolución clasificados en función del impacto y la urgencia de los incidentes.
- Nivel de cumplimiento de los *Niveles de Servicio (SLA)*.
- Costos asociados.
- Uso de los recursos disponibles en el Centro de Servicios.
- Porcentaje de incidentes, clasificados por prioridades, resueltos en primera instancia por el Centro de Servicios.
- Grado de satisfacción del cliente.

Con todo esto, llevar registro y realizar gestión sobre situaciones que impidan u obstaculicen el funcionamiento de actividades dentro de una empresa es clave para asegurar un buen servicio a sus clientes y disminuir los costos y pérdidas generados frente a una paralización de actividades.

## 2.2. Minería de Procesos

El Diseño y Modelamiento de los Procesos de Negocios [22] es primordial para la gestión y ejecución de los procesos [3, 6, 16]. La Minería de Procesos o, en inglés, *Process Mining*, es una disciplina que tiene como objetivo descubrir, monitorear y mejorar procesos de negocio a través de la extracción de conocimiento del registro de eventos de los sistemas de información [17].

*Process Mining* busca extraer conocimiento a partir de lo que pasa en la realidad utilizando la información de los registros de un sistema o *log de eventos* y, con ellos, generar mejoras mediante el uso de técnicas y herramientas ligadas al descubrimiento de procesos, control de actividades, indicadores de procesos y organización y estructuras sociales. Mediante un análisis sistemático, ayuda a definir mejoras operativas y entrega información de utilidad para la gestión de nuevas capacidades y la gestión de producción de una organización. Responde a preguntas tales como: ¿el proceso diseñado es efectivamente el que opera en la realidad?; ¿los roles asignados cumplen con lo definido?; ¿es correcta, suficiente y verídica toda la información recibida?; ¿se esconden malas prácticas dentro de los procesos?; ¿dónde están los problemas y qué los genera?; ¿qué recursos

son los con mayor carga?; ¿se cumplen los niveles de servicio?; ¿por qué se producen cuellos de botella?.

Está altamente ligado al área de Minería de Datos o, en inglés, Data Mining y el uso de algunas de sus técnicas, pero con una visión centrada en el análisis particular de los procesos de negocios [7, 15]. Algunas de las técnicas más tradicionales de Minería de Datos pueden ser utilizadas en el contexto del *Process Mining*, sin embargo, existen nuevas técnicas desarrolladas específicamente para realizar minería de procesos.

Un paso esencial para el desarrollo de proyectos de *Process Mining*, es la obtención de un *log de eventos* que se encuentre presente en los sistemas de información de la empresa. Estos registros temporales de la trazabilidad de un proceso son cada vez más comunes en cuanto la informatización en las empresas se ha expandido en diversas áreas. Este debe contener al menos un ID o identificador de Caso, una actividad relacionada al proceso bajo análisis, un evento y un registro de tiempo (timestamp) o identificador de secuencia de actividades por caso. Adicionalmente es posible agregar información como recursos utilizados en cada actividad, registro de tiempo de inicio y fin por actividad, costos, reglas, entre otros. De este modo, la base de análisis será un *log de eventos* el cual debe cumplir las siguientes características:

- Los datos provienen de una perspectiva de procesos.
- El *log de eventos* viene de un proceso completo.
- Cada evento corresponde a una actividad ejecutada dentro del proceso.
- Varios eventos son agrupados en una instancia o caso de proceso.
- Cada caso forma una secuencia de eventos ordenada por su timestamp.

### 2.2.1. Tipos y perspectivas de *Process Mining*

Se pueden distinguir tres tipos de Minería de Procesos [17, 21].

1. Descubrimiento (Discovery): Se usa un registro de eventos para producir un modelo sin usar información a-priori. Existen variados algoritmos para el Descubrimiento Alpha Mining, Heuristic Miner, Genetic Mining, Fuzzy Mining, entre otros.
2. Conformidad (Conformance): Se compara un proceso existente con un registro de eventos del mismo proceso, para verificar si la realidad, según

el registro, es equivalente al modelo y viceversa<sup>1</sup>.

3. Mejora (Enhancement): Se busca extender o mejorar un modelo de proceso existente con la información del proceso real almacenado en un registro de eventos.

Estos ítemes, a su vez, cubren las siguientes perspectivas [14, 18, 21]:

- Perspectiva de Control de Flujo (Control-Flow Perspective): Busca la caracterización de todos los caminos o rutas posibles.
- Perspectiva Organizacional (Organizational Perspective): Tiene como fin estructurar la organización ya sea clasificando a las personas en términos de roles y unidades organizacionales, o para mostrar la red social [4, 12].
- Perspectiva de Casos (Case Perspective): Se enfoca en las propiedades de los casos (caracterización por rutas, actores u otros elementos del *log de eventos*).
- Perspectiva de Tiempo (Performance Perspective): Se relaciona con la ocurrencia y la frecuencia de los eventos. Posibilita descubrir cuellos de botella, medir niveles de servicio, monitorear la utilización de recursos y predecir el tiempo restante de los casos en ejecución [19].

---

### 3. Diagnóstico para Análisis de Incidencias

---

La Gestión de Incidencias tiene la responsabilidad de resolver fallas y problemas para restablecer la operación normal de los servicios y minimizar el impacto en las operaciones del negocio, asegurando la disponibilidad y los niveles de calidad. En el caso de análisis, las incidencias se producen debido a problemas vinculados con los servicios contratados y pueden abarcar: software (Ej. ERP) y hardware (Ej. estaciones de trabajo).

Desde el punto de vista de la estructura de un proceso de gestión de incidencia, estos por lo general se componen de tres etapas que dependen secuencialmente una de otra: registro y clasificación, análisis y resolución del incidente.

---

<sup>1</sup>En [11] se plantea un marco teórico que permite verificar la conformidad y medirla usando métricas. En [10] se ven enfoques alternativos y se propone ETConformance, que es recomendado para procesos con alta cantidad de estados. En [1] se proponen Modelos Flexibles.

Figura 2: Macro Procesos de Atención de Incidencias



En términos funcionales si un usuario interno detecta una incidencia en un servicio, se comunica con el soporte de primer nivel mediante el sistema de gestión o por vía telefónica, donde se registra y clasifica el incidente, y luego se deriva al soporte de segundo nivel encargado del análisis y resolución, esta derivación se realiza a alguna de las 13 empresas relacionadas a los servicios analizados.

Desde el punto de vista del flujo de actividades de soporte, el 97,85 % lo realizan 6 empresas. A su vez, la principal empresa que registra actividades de incidencia concentra el 30,25 % de éstas, existiendo una alta concentración de fallas similares.

Es importante considerar que la compañía dispone de contratos de provisión de servicios y soporte con sus proveedores. Estos contratos definen los plazos máximos y los lineamientos a considerar en las incidencias.

Al momento del estudio la gestión de incidencias no posee un proceso formal de control que le permita detectar acciones correctivas a sus problemas en la gestión, esto ha sido desencadenado por el desconocimiento de los resultados del proceso. Desde esta posición, los tomadores de decisión refieren al retrabajo como la falencia principal.

En el análisis del proceso, se estudiaron dos períodos (diciembre del 2012 a marzo del 2013 y diciembre del 2013 a marzo del 2014), en ellos se detectaron importantes problemas de registro como omisiones en campos de diagnóstico y en la asignación de recursos.

Los datos correspondientes al año 2013 correspondían al periodo de prueba del proyecto realizado, por lo que fueron sacados del análisis al no mostrar información confiable de los procesos. Los datos obtenidos durante el 2014 reflejan un mejor registro gracias a los ajustes en el diseño de proceso y asignación de roles propuestos en el transcurso del proyecto. De esta manera, se extrajeron un total de 34.567 registros correspondientes a los años 2013 y 2014, de los cuales se seleccionaron 16.247 correspondientes al año 2014 (47 %).

Luego, al realizar un análisis de los casos del año 2014 se identificó que el 13 % de los datos (2.166 registros) no contaban con un responsable asignado para su resolución. Ello condujo para efectos de la selección de datos a sacarlos del análisis de mejora, sin embargo, la identificación de esta situación permitió identificar problemas de retraso, incumplimiento de SLA y retrabajos

debido a la falta de seguimiento y reasignaciones. Finalmente, la base de datos a utilizar contiene 14.081 registros (41 %) de la base original.

Más allá del proceso de limpieza y selección de datos, es importante recalcar que los errores en el registro tienen directa implicancia en el del proceso real, dado que es una de las causas que aumenta el retrabajo debido a la mala clasificación de éstos y los retrasos en la gestión del incidente al no tener responsables en su asignación.

Lo anterior transparenta la necesidad de realizar una evaluación del proceso de gestión de incidencias que permita identificar los principales problemas y que permita tomar acciones para corregirlos, lo que finalmente se traduce en la definición de un proceso de monitoreo, rediseño de procesos dentro del área con nuevos responsables y diseño de controles e indicadores.

---

#### 4. *Process Mining* para Gestión de Incidencias

---

La utilización de *Process Mining* para Gestión de Incidencias apunta a mejorar el servicio y la creación de un proceso de control y seguimiento que permita tener una constante evaluación del funcionamiento de la operación con sus respectivos procedimientos. Con ello se busca implantar actividades de control y seguimiento al proceso actual que permitan generar una constante evaluación del desempeño según lo indicado en el punto 2.1.

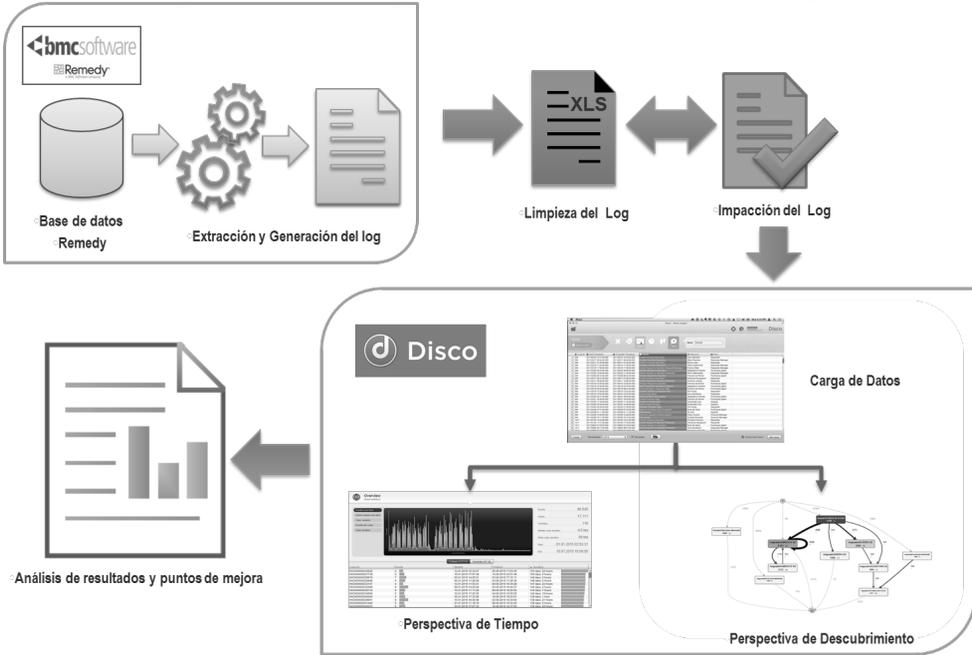
En el mercado existe una serie de herramientas para realizar *Process Mining*, como ProM (open source y gratuita) [20] y Disco Miner (comercial). En particular Disco Miner se basa en el framework Fuzzy Miner [8] con una completa gama de métricas y estrategias de modelamiento. Este trabajo utilizó Disco Miner con licencia académica para obtener el mapa de procesos, simulación e identificación de cuellos de botella, análisis de carga de recursos y análisis de tiempo, así como también para realizar la tipificación de casos y diagnóstico para el área. Para el análisis de redes sociales y grupos de trabajo se utilizó ProM.

Los datos procesados fueron importados en la herramienta con información de cada caso, tiempos de atención, encargado de resolverlo y la actividad correspondiente. Cada registro proporciona los hechos que materializan información relevante para generar los algoritmos para la identificación de puntos de mejoras.

Como muestra la figura 3, el proceso realizado comienza con la extracción, transformación y generación del *log del eventos* desde el workflow Remedy ®, con él se seleccionaron las columnas necesarias para inspeccionar el proceso

cargando los datos en la herramienta Disco Miner donde se realizó descubrimiento del proceso, simulación y revisión de los indicadores del proceso para, finalmente, realizar el análisis de resultados e identificación de los puntos de mejora.

Figura 3: Metodología utilizada para *Process Mining*



A continuación, se analiza el *log de eventos* y se dan a conocer los modelos de control de flujo obtenidos mediante aplicación de *Process Mining*.

#### 4.1. Generación del Log

Se definió como actividad única el número de incidencia, que indica cada caso o hilo ejecutado y almacenado en el log.

Por cada incidencia los elementos claves de registro son: la empresa de soporte asignada, el grupo asignado (Mainframe, Mesa de Ayuda, Swift, etc.), el recurso para resolver la incidencia (usuario) y la estadística de tiempo de inicio y fin de cada registro.

Estos registros del log permiten hacer un seguimiento completo de la incidencia en el proceso. A su vez, se hace necesario una clasificación e identificación del diagnóstico a nivel de sistema, para reducir los errores de asignación en función a análisis futuros de minería de procesos.

Figura 4: Log del Proceso

	A	B	C	D	E	F	
	Número de incidencia*	Fecha de envío	Tipo de asignación*	Empresa de soporte*	Organización de soporte*	Grupo asignado*	U
1	INC000000132270	26-12-2014 10:11:30	Asignado	TIS CHILE	TIS CHILE SOPORTE	TIS SOPORTE	
2	INC000000132270	26-12-2014 10:13:00	Asignado	TIS CHILE	TIS CHILE SOPORTE	PLATAFORMA TOKEN RSA	
4	INC000000132270	29-12-2014 12:58:43	Asignado	HP Chile	SERVICIO MESA DE AYUDA	HP MESA DE AYUDA	
5	INC000000132270	29-12-2014 12:58:43	Asignado	HP Chile	SERVICIO MESA DE AYUDA	HP MESA DE AYUDA	JK
6	INC000000149020	23-12-2014 20:48:34	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
7	INC000000149020	23-12-2014 20:48:34	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
8	INC000000149171	23-12-2014 20:52:25	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
9	INC000000149171	23-12-2014 20:52:25	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
10	INC000000150513	23-12-2014 20:55:39	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
11	INC000000150513	23-12-2014 20:55:40	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
12	INC000000151614	23-12-2014 21:02:59	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
13	INC000000151614	23-12-2014 21:02:59	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
14	INC000000151654	23-12-2014 21:05:58	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
15	INC000000151654	23-12-2014 21:05:58	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
16	INC000000151898	16-12-2014 15:12:02	Asignado	ACCENTURE - CHILE	SOPORTE DE SERVICIOS TI	ACCENTURE N2	J
17	INC000000151898	16-12-2014 15:12:02	Asignado	ACCENTURE - CHILE	SOPORTE DE SERVICIOS TI	ACCENTURE N2	F
18	INC000000152104	23-12-2014 21:09:26	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
19	INC000000152104	23-12-2014 21:09:26	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
20	INC000000152595	23-12-2014 21:14:16	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	
21	INC000000152595	23-12-2014 21:14:16	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	W
22	INC000000153084	23-12-2014 21:35:46	Asignado	TELEFONICA CHILE	SISTEMAS DE ATENCIÓN COMERCIAL	SOPORTE TERMINALES	

### 4.2. Limpieza de datos y selección de atributos

Para sistematizar el análisis se definió dentro del sistema que, mensualmente o por solicitud del encargado del área de gestión de incidencias, se ejecute un proceso automatizado de Limpieza, Transformación y Carga (ETL por su sigla en inglés de *Extract, Transform and Load*) de los datos de acuerdo a las reglas de negocio definidas en este proyecto. La limpieza y carga es básicamente la selección de los atributos de interés y el filtro de datos a partir de reglas que aseguran la completitud e integridad de los datos para el análisis. Una vez que los datos se encuentran depurados, se cargan en la herramienta de *Process Mining* para su análisis dentro de ella.

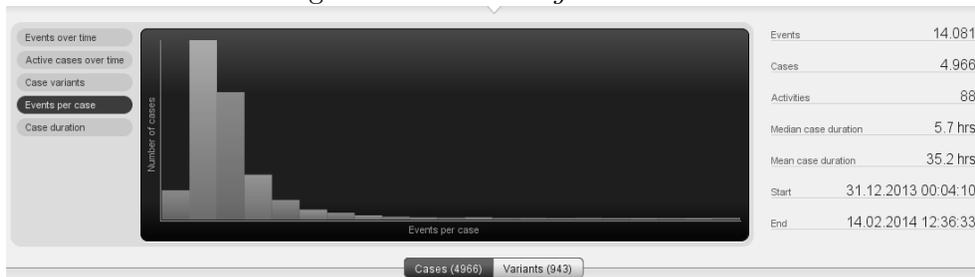
Figura 5: Vista de carga de datos en Disco Miner

Número de incidencia*	Tipo de asignación*	Empresa de soporte*	Organización de soporte*	Grupo asignado*	Usuario asignado
1	Asignado	EVERIS CHILE SA	SOPORTE DE SERVICIOS TI	EVERIS N2	
2	Asignado	EVERIS CHILE SA	SOPORTE DE SERVICIOS TI	EVERIS N2	CAMILLO E BERRIOS MARTINEZ
3	Iniciado	IBM CHILE S.A	Servicios Microsoft	Servicios Microsoft	RAUL MOL
4	Asignado	AMDOCS CHILE SPA	SOPORTE DE SERVICIOS TI	AMDOCS N2	
5	Asignado	INDRA SISTEMAS CHILE S.A	SOPORTE DE SERVICIOS TI	INDRA N2	
6	Asignado	INDRA SISTEMAS CHILE S.A	SOPORTE DE SERVICIOS TI	INDRA OTROS PROVEEDORES	ENRIQUE ZUÑIGA VILLALOBOS
7	Asignado	EVERIS CHILE SA	SOPORTE DE SERVICIOS TI	EVERIS N2	ANGGIE PICA FUENTES
8	Asignado	EVERIS CHILE SA	SOPORTE DE SERVICIOS TI	EVERIS N2	CAMILLO E BERRIOS MARTINEZ
9	Iniciado	IBM CHILE S.A	Servicios Microsoft	Servicios Microsoft	RAUL MOL
10	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	TELEFONICA CHILE	INCIDENCIAS_MOVIL	
11	Asignado	IBM CHILE S.A	EXPLOTACION	EXPLOTACION MCRANGE	
12	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	TELEFONICA CHILE	INCIDENCIAS_MOVIL	LUIS INJOSA ALVARADO
13	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	SOPORTE DE SERVICIOS TI	ACCENTURE N2	LUIS INJOSA ALVARADO
14	Iniciado	IBM CHILE S.A	Servicios Microsoft	Servicios Microsoft	RAUL MOL
15	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	TELEFONICA CHILE	INCIDENCIAS_MOVIL	LUIS INJOSA ALVARADO
16	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	TELEFONICA CHILE	INCIDENCIAS_MOVIL	LUIS INJOSA ALVARADO
17	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	TELEFONICA CHILE	INCIDENCIAS_MOVIL	DAVID GONZALEZ
18	Asignado	IBM CHILE S.A	EXPLOTACION	EXPLOTACION MCRANGE	RAUL HERNAN MOLINA AGUIRRE
19	Asignado	ACCENTURE CHILE ASESORIAS & SERVICIOS L.	SOPORTE DE SERVICIOS TI	ACCENTURE N2	LUIS INJOSA ALVARADO
20	Iniciado	IBM CHILE S.A	Servicios Microsoft	Servicios Microsoft	RAUL MOL
21	Asignado	AMDOCS CHILE SPA	SOPORTE DE SERVICIOS TI	AMDOCS N3	
22	Asignado	IBM CHILE S.A	EXPLOTACION	EXPLOTACION MCRANGE	NICOLAS P REYES ESCALONA
23	Asignado	INDRA SISTEMAS CHILE S.A	SOPORTE DE SERVICIOS TI	REQUERIMIENTOS SOPORTE	CHRISTIAN A MAGAÑA RUIZ
24	Asignado	AMDOCS CHILE SPA	SOPORTE DE SERVICIOS TI	AMDOCS N3	CARLOS PEREZ
25	Asignado	IBM CHILE S.A	EXPLOTACION	EXPLOTACION MCRANGE	RAUL MOL
26	Iniciado	IBM CHILE S.A	Servicios Microsoft	Servicios Microsoft	RAUL MOL
27	Asignado	IBM CHILE S.A	EXPLOTACION	EXPLOTACION MCRANGE	RAUL MOL

### 4.3. Análisis de datos desde la perspectiva descubrimiento

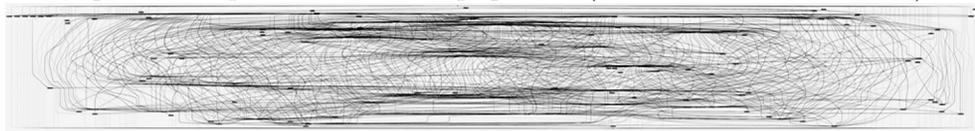
Para el descubrimiento del proceso se realiza análisis del *log de eventos* con Disco Miner para obtener una visualización de la situación actual, considerando los datos disponibles en la herramienta de gestión de incidencias se generan las estadísticas globales del *log de eventos* como muestra la figura 6.

Figura 6: Resumen *log de eventos*



Podemos ver entonces, que existen 14.081 eventos distribuidos en 4.966 casos de acuerdo a 88 actividades. Esto da origen a 943 variantes o posibles rutas dentro de la gestión de incidencias. Luego, este análisis inicial permite el descubrimiento del proceso completo y su correspondiente mapa.

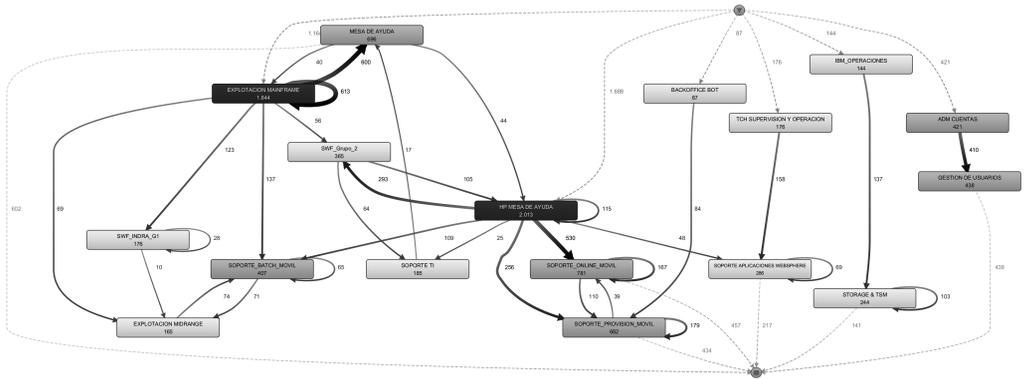
Figura 7: Mapa de Procesos “Spaghetti” (descubrimiento sin filtrar)



Normalmente el análisis de un *log de eventos* en procesos con gran cantidad de actividades y recursos tiende a ser caótico y visualmente es conocido como el “spaghetti process” [17], por ello es necesario analizar qué variantes, actividades y rutas del proceso son relevantes o interesantes al momento de realizar un análisis. Al realizar el filtro dentro de la herramienta es posible identificar rutas críticas y procesos principales, además de cuellos de botella, reprocesamientos, sobrecarga de trabajo, tiempos de ejecución, entre otros. La figura 8 muestra una posible vista de análisis para el proceso revisado.

Con este análisis es posible identificar variantes de proceso anómalas e idealmente tomar acciones para corregirlas, así como también identificar variantes de alta frecuencia para evaluarlas y mejorarlas en base a los indicadores de proceso. Si miramos la figura 8 y tomamos sólo la actividad de “Explotación Mainframe” vemos que gran parte de las solicitudes (13%) llega directo a esta actividad y dentro de las 1.844 ejecuciones de la actividad el 33% (600)

Figura 8: Mapa de procesos filtrado



se redirecciona a la “Mesa de Ayuda” y otro 33% (613) se reprocessa, mientras que sólo un 34% sigue una ruta orientada la resolución del problema.

El análisis de mapa de procesos permite identificar visualmente la ejecución y eficiencia del mismo. Se pueden identificar donde se encuentran las interrupciones de mayor tiempo y las mayores reiteraciones con algún incumplimiento del proceso, carga de trabajo y cuellos de botella, ruta crítica, anomalías en el proceso y validación de actividades diseñadas.

#### 4.4. Análisis de Redes Sociales

Con el fin de entender las relaciones de trabajo e identificar posibles anomalías en la distribución de carga se realizó un análisis de redes sociales tomando en consideración a los recursos asignados para resolver cada incidencia.

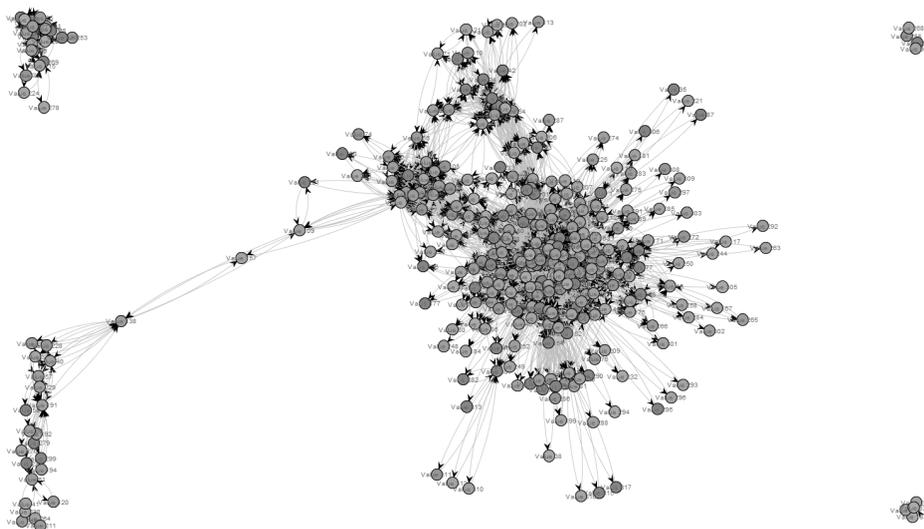
Para la configuración de la red se utilizó como factor el indicador de centralidad “betweenness” que refleja la cantidad de veces donde el nodo pertenece a las rutas más cortas respecto de un nodo a todos los demás. Se calcula

$$C(i) = \sum_{j < k} g_{jk}(i) / g_{jk}$$

Donde  $g_{jk}$  es el número de rutas más cortas conectando  $kj$  y  $g_{jk}(i)$  es el nodo de referencia. Para la distribución de los nodos se utilizó el algoritmo Force-Directed [2, 9] anonimizando los recursos.

Por un lado se analizó la red social de las personas (ver figura 9) de acuerdo al análisis de trabajo en conjunto. La red resultante muestra que existe una serie de grupos de trabajos, principalmente 5 grandes grupos, con tres grupos que están totalmente aislados.

Figura 9: Red social personas asignadas



El análisis de los grupos formados muestra que existe una operación principal ligada a una variante de proceso con una alta frecuencia, esta variante debería coincidir con el proceso principal bien definido. Sin embargo en este caso sólo muestra la variante más frecuente dentro de la forma de operar actual que tiene la empresa. Asimismo, el gran clúster formado muestra que existe una alta interdependencia entre los trabajadores de las distintas áreas, reflejando desde otra perspectiva la alta tasa de retrabajo y las asignaciones entre áreas existentes al momento de realizar el análisis.

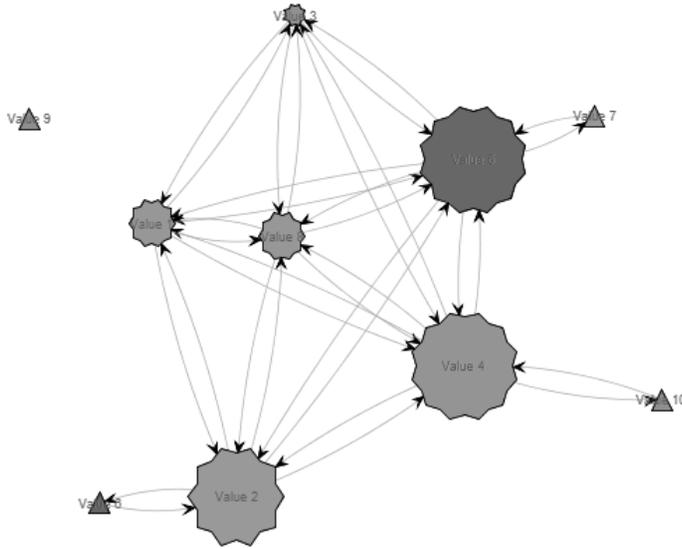
Por otro lado se analizó la red social de las empresas responsables de entregar el servicio y solucionar finalmente el incidente.

El análisis de las empresas muestra que existe un alto grado de trabajo conjunto con tres empresas que centran la recepción de las solicitudes de recepción de incidencias (en la figura 10 aquellos círculos de mayor tamaño). Ésto confirma el análisis de personas donde concluimos que existen procesos con malas asignaciones de trabajo y alta carga de reasignaciones de una a empresa a otra.

#### 4.5. Análisis de datos desde la perspectiva tiempo

Al analizar la estadística es posible revisar cuánto demora un caso actividad por actividad o por recurso. Para el caso presentado (proceso filtrado) en este trabajo el procesamiento de una incidencia alcanza en promedio 35.2 horas pero con una mediana de 5.7 horas.

Figura 10: Red social personas asignadas



Tomando nuevamente la actividad de “Explotación Mainframe” y tomando los eventos que se reprocesan, podemos ver que la duración media es de 99.2 minutos pero con un rango de duración de 12 segundos a 22 horas. Con lo que podemos identificar que las actividades tienen una alta varianza en su desarrollo y poca estandarización en su ejecución.

Al realizar este análisis, se llevó a cabo una revisión de las actividades dentro del log y se crearon varias hipótesis para identificar la causa de este alto tiempo. Los problemas identificados abordaron desde el poco compromiso del equipo con el registro de las incidencias y, por lo tanto, una alta tasa de incidencias pendientes en sistema así como también incidencias que efectivamente no se resolvieron y quedaron pendientes. Este análisis permitió identificar problemas en el proceso y revisar los acuerdos de nivel de servicio para hacer efectivos los contratos que, en algunos casos, gatillan multas para la empresa proveedora del servicio.



---

## 5. Conclusiones

---

El diseño del proceso de control y seguimiento de Gestión de Incidencias generó un gran aporte ya que se focalizó en la necesidad de optimizar y gestionar los incidentes que se generan por continuidad operacional de Telefónica, analizando eventos reflejados inicialmente en el estado de las distintas señales provistas por las distintas entidades que forman parte de la cadena de servicios ofrecidos por Telefónica Chile en su línea Movistar.

De esta manera, no sólo se obtuvieron herramientas permanentes de análisis y control de procesos, si no también se logró identificar cuál es la secuencia de actividades que recorre una incidencia desde el inicio hasta el fin y se visualizó cuáles son los roles que intervienen con sus respectivas interacciones. Estos análisis generaron resultados asociados a una gran cantidad de interacciones entre los mismos roles que al final no provocaban un beneficio al proceso sino retraso en la entrega de la solución.

Con todo, la Gestión de Incidencias es clave para el cumplimiento de acuerdos de una empresa y la calidad de un servicio por lo que invertir en herramientas que mejoren la visualización, seguimiento y capacidad de medición reporta en concreto grandes beneficios para toda empresa.

El correcto almacenamiento de información es esencial para poder realizar cualquier análisis y evaluación de cada actividad, por lo que, tener un buen levantamiento y diseño de procesos es la base para toda empresa que busque excelencia en la entrega de sus productos o servicios.

*Process Mining* como herramienta de análisis avanzada requiere de una buena fuente de información y, por lo tanto, su uso y aplicación es recomendable en casos donde exista cierto grado de madurez en la gestión y seguimiento de procesos.

El proyecto realizado permitió encontrar una metodología para buscar mejoras oportunas e identificar cómo se llevan a cabo las actividades. La inclusión de indicadores de procesos permitió además cuantificar el nivel real de servicio entregado por cada uno de los proveedores, evaluar su rendimiento y el cumplimiento de contratos. Además, conocer la frecuencia con que se lleva a cabo cada actividad y la posibilidad de analizarla tanto por la perspectiva de actividades como recursos utilizados ha permitido tomar acciones para mitigar riesgo de operación y distribuir de mejor manera la asignación de recursos.

**Agradecimientos:** Este trabajo fue realizado gracias a Telefónica Chile de acuerdo al trabajo de tesis “Control y Seguimiento de Atención de Incidencias Utilizando Minería de Procesos”, proyecto de Néstor Contreras para el grado de Magíster en Ingeniería de Negocios con Tecnologías de Información y patrocinado por el Grupo de Data Science del Instituto de Sistemas Complejos (ISCI) de la Universidad de Chile con el fin de probar técnicas de *Process Mining* en casos reales.

## Referencias

- [1] A. Adriansyah, B. Van Dongen, y W. Van Der Aalst. Conformance checking using cost-based fitness analysis. In *Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International*, páginas 55–64. IEEE, 2011.
- [2] R. Chernobelskiy, M. Cunningham, K. and Goodrich, S. Kobourov, y L. Trott. Force-directed lombardi-style graph drawing. In *International Symposium on Graph Drawing*, páginas 320–331. Springer, 2011.
- [3] M. Drachmann, S. Freisner, y N. Müller. Don’t mind the gap –the future CIO. *IT University of Copenhagen*, 2014.
- [4] D.R. Ferreira y C. Alves. Discovering user communities in large event logs. In *International Conference on Business Process Management*, páginas 123–134. Springer, 2011.
- [5] Fluxicon. Disco - Process Mining Tool. <https://fluxicon.com/disco/>, jul 2016.
- [6] Gartner. Hunting and harvesting in a digital world - Gartner CIO Agenda Report, 2013.
- [7] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, y M.C. Shan. Business process intelligence. *Computers in industry*, 53(3):321–343, 2004.
- [8] C. W Günther y W. Van Der Aalst. Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International Conference on Business Process Management*, páginas 328–343. Springer, 2007.
- [9] M. Lombardi y R. Hobbs. *Mark Lombardi: Global Networks*. Independent Curators, 2003.

- [10] J. Muñoz-Gama y J. Carmona. A fresh look at precision in process conformance. In *International Conference on Business Process Management*, páginas 211–226. Springer, 2010.
- [11] A. Rozinat y W. Van Der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.
- [12] M. Song, C. Günther, y W. Van Der Aalst. Trace clustering in process mining. In *International Conference on Business Process Management*, páginas 109–120. Springer, 2008.
- [13] SUBTEL. Sector telecomunicaciones primer semestre 2015. [http://www.subtel.gob.cl/wp-content/uploads/2015/03/PPT\\_Series\\_JUNIO\\_2015\\_VF.pdf](http://www.subtel.gob.cl/wp-content/uploads/2015/03/PPT_Series_JUNIO_2015_VF.pdf), sep 2015.
- [14] A. Tiwari, C.J. Turner, y B. Majeed. A review of business process mining: state-of-the-art and future trends. *Business Process Management Journal*, 14(1):5–22, 2008.
- [15] W. Van Der Aalst. Using process mining to bridge the gap between bi and bpm. *IEEE Computer*, 44(12):77–80, 2011.
- [16] W. Van Der Aalst. Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2):7, 2012.
- [17] W. Van Der Aalst, A. Adriansyah, A.K. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. Chandra Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, et al. Process mining manifesto. In *International Conference on Business Process Management*, páginas 169–194. Springer, 2011.
- [18] W. Van Der Aalst, H. A Reijers, A. Weijters, B. Van Dongen, A. De Medeiros, M. Song, y H. Verbeek. Business process mining: An industrial application. *Information Systems*, 32(5):713–732, 2007.
- [19] W. Van Der Aalst, M. Schonenberg, y M. Song. Time prediction based on process mining. *Information Systems*, 36(2):450–475, 2011.
- [20] B. Van Dongen, A. de Medeiros, H. Verbeek, A. Weijters, y W. van der Aalst. The prom framework: A new era in process mining tool support. In *Proceedings of the 26th International Conference on Applications and Theory of Petri Nets, ICATPN'05*, páginas 444–454, Berlin, Heidelberg, 2005. Springer-Verlag.

- [21] B. Van Dongen, A.K. De Medeiros, H. Verbeek, A. Weijters, y W. Van Der Aalst. The prom framework: A new era in process mining tool support. In *International Conference on Application and Theory of Petri Nets*, páginas 444–454. Springer, 2005.
- [22] M. Weske. Business process management architectures. In *Business Process Management*, páginas 333–371. Springer, 2012.





---

## Programas de Postgrado y Postítulos DII

---

### DOCTORADO

**Doctorado**  
en Sistemas de Ingeniería



**Sólida formación en herramientas metodológicas necesarias para identificar, analizar, modelar y resolver problemas complejos en sistemas de ingeniería**

Contacto:  
562-29784017 | [doctorado@sistemasdeingenieria.cl](mailto:doctorado@sistemasdeingenieria.cl)  
Informaciones y postulación en línea en:  
[www.dsi.uchile.cl](http://www.dsi.uchile.cl)

**fcfm** FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

### MAGÍSTERES

 INGENIERIA INDUSTRIAL  
UNIVERSIDAD DE CHILE



**MGO** | **Magíster  
Gestión de Operaciones**

Formar profesionales de excelencia en investigación de operaciones, quienes podrán enfrentar problemas complejos en gestión de operaciones, integrando herramientas matemáticas, económicas y tecnológicas.

**fcfm** FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Informaciones en: [www.mgo.uchile.cl](http://www.mgo.uchile.cl)  
Contacto: 562-29784073

# Programas de Postgrado y Postítulos DII



## MAGCEA

MAGÍSTER EN ECONOMÍA APLICADA

Busca formar profesionales y académicos de gran capacidad analítica y sólida base en economía

[www.cea-uchile.cl](http://www.cea-uchile.cl) | [infocea@dii.uchile.cl](mailto:infocea@dii.uchile.cl) | 562-29784073



**MBE**  
Master in Business Engineering

Magíster en Ingeniería de Negocios  
con Tecnologías de la Información

## Los líderes de hoy comprenden cómo la tecnología lleva a las empresas al éxito.

### A Quién está Dirigido

Ejecutivos y profesionales que deseen liderar o ejecutar proyectos innovadores de diseño integral y sistémico de los negocios orientados a mejorar su competitividad.

### Metodología

Este es un Magíster Integrador, conformado por un conjunto de cursos de gestión, modelos analíticos aplicados, diseño de negocios, arquitectura y procesos, tecnologías de información de base y diseño de aplicaciones, y de inducción de habilidades de innovación.

Además de las evaluaciones tradicionales por medio de controles y exámenes, una parte fundamental del trabajo de los alumnos será el desarrollo de un proyecto de innovación en el negocio de la empresa auspiciadora -donde ejecutará su residencia-, el cual se llevará a cabo durante todo el programa, en los cursos obligatorios del mismo.

### Duración:

3 semestres académicos más un semestre para dar término al Proyecto de Grado.

### Horario:

Martes o jueves vespertino, viernes de 14:30 a 18:00 horas  
y sábados de 8:30 a 11:45 horas.

### Informaciones:

Coordinadora: Ana María Valenzuela.  
(56 2) 978 4835 / [anamaria@dii.uchile.cl](mailto:anamaria@dii.uchile.cl)

[www.dii.uchile.cl](http://www.dii.uchile.cl)



# Programas de Postgrado y Postítulos DII



MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS-MGPP



LÍDERES DE EXCELENCIA PARA AMÉRICA LATINA

MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS

**ADMISIÓN 2017**

Postulaciones en línea en:  
[www.mgpp.cl](http://www.mgpp.cl)

(562) 2978 4067 - 2978 4043  
[mgpp@dii.uchile.cl](mailto:mgpp@dii.uchile.cl)

**Inicio Horario Diurno**

Fines de Mayo de 2017

**Inicio Horario Ejecutivo**

Julio de 2017



Magíster en Gestión y Políticas Públicas  
Acreditado 7 años, desde octubre de 2011  
y hasta octubre de 2018



Comisión Nacional  
de Acreditación

Universidad de Chile - Acreditada  
7 años, en todas las Áreas (Docencia de  
Pregrado, Docencia de Postgrado,  
Investigación, Gestión Institucional,  
Vinculación con el Medio), desde diciembre  
de 2011 y hasta diciembre 2018.

---

# Programas de Postgrado y Postítulos DII

---

 **MBA**  
UNIVERSIDAD DE CHILE  
INGENIERÍA INDUSTRIAL

**PRIMEROS EN GESTIÓN**

    
MARKETING FINANZAS OPERACIONES

**EN AMÉRICA LATINA**  
- RANKING AMÉRICA ECONOMÍA 2016 -

MBA PART TIME  
MBA EXECUTIVE  
MBA FULL TIME  
GLOBAL MBA

 INGENIERIA INDUSTRIAL  
UNIVERSIDAD DE CHILE

+56 2 2978 4002 | [mba@dii.uchile.cl](mailto:mba@dii.uchile.cl)  
[www.mbauchile.cl](http://www.mbauchile.cl)



INGENIERÍA INDUSTRIAL  
UNIVERSIDAD DE CHILE

# MBA

Versión Industria Minera  
7ª Versión - Santiago

Magíster en Gestión y  
Dirección de Empresas

[www.mbamin.cl](http://www.mbamin.cl)



[mbamin@dii.uchile.cl](mailto:mbamin@dii.uchile.cl)



+56 22978 4020



FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

---

## Programas de Postgrado y Postítulos DII

---

### EDUCACIÓN EJECUTIVA



INGENIERÍA INDUSTRIAL  
UNIVERSIDAD DE CHILE

**DESARROLLA  
TUS HABILIDADES  
Y MARCA LA  
DIFERENCIA**

### EDUCACIÓN EJECUTIVA

- DIPLOMADOS
- CURSOS DE ESPECIALIZACIÓN
- PROGRAMAS PARA EMPRESAS
- SEMINARIOS Y TALLERES

[www.eeuchile.cl](http://www.eeuchile.cl)

---

diplomas@dii.uchile.cl  
+56 2 2978 4002



FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE



