

- Programando múltiples ligas deportivas con equidad en distancias de viaje: 5
Una aplicación en el fútbol juvenil argentino
Guillermo A. Durán, Mario Guajardo, Agustina F. López, Javier Marengo, Gonzalo A. Zamorano.
- Predicción de readmisión de pacientes pediátricos mediante aprendizaje supervisado 37
Patricio Wolff, Sebastián Ríos.
- Implementación de una heurística para la programación automática de horarios de una escuela secundaria 51
Guillermo Durán, Nazareno Faillace.
- Análisis de texto para Gestión de marca con el uso de Instagram aplicado a una comunidad de mascotas en Chile 81
Cinthya Vergara, Valeria Scapini, Simón Maturana, Francisco Paz.
- Optimización del barrido y recolección de hojas en la ciudad argentina de Trenque Lauquen 101
Diego Delle Donne, Valeria Di Tomaso, Guillermo Durán.

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: Av. Beauchef 851, Santiago, Chile.
Web URL: www.dii.uchile.cl/ris
Diseño Portada: Gabriella Fabbri

Carta Editorial Volumen XXXIII

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, de investigadores del ISCI y de académicos y estudiantes de Universidades argentinas.

Nuestro objetivo a través de esta publicación es contribuir a la generación y difusión de las tecnologías modernas de investigación de operaciones, 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 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 sufixo del año de publicación para diferenciarlas.

Detalles en www.dii.uchile.cl/ris

PROGRAMANDO MÚLTIPLES LIGAS DEPORTIVAS CON EQUIDAD EN DISTANCIAS DE VIAJE: UNA APLICACIÓN EN EL FÚTBOL JUVENIL ARGENTINO

GUILLERMO A. DURÁN *
MARIO GUAJARDO **
AGUSTINA F. LÓPEZ ***
JAVIER MARENCO ****
GONZALO A. ZAMORANO *****

Resumen

La principal categoría del fútbol juvenil argentino es disputada por los clubes que integran la primera división profesional del país. Cada club cuenta con seis divisiones juveniles, tres divisiones *mayores* (sub-20, sub-18 y sub-17) y tres divisiones *menores* (sub-16, sub-15 y sub-14). Temporada a temporada hay entonces seis torneos de divisiones juveniles para los clubes de primera división, con la particularidad de que las tres divisiones mayores juegan con el mismo fixture, mientras que las divisiones menores utilizan esa misma programación pero con las localías invertidas. Al realizarse los fixtures de manera manual, solía existir una gran diferencia en kilómetros viajados entre las divisiones mayores y las menores de un mismo club, una situación que no era bien vista por dirigentes, entrenadores y jugadores. Además, se tornaba imposible considerar las innumerables condiciones de todas las ligas para una buena calendarización. En este trabajo desarrollamos un enfoque de programación

* Departamento de Matemática e Inst. de Cálculo FCEyN-UBA y CONICET, Argentina, Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Santiago, Chile.

** Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Noruega

*** Departamento de Matemática e Instituto de Cálculo FCEyN-UBA, Argentina

**** Instituto de Ciencias, UNGS y Dto. de Computación, FCEyN-UBA, Argentina

***** Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Santiago, Chile

matemática para la calendarización de las seis ligas en forma simultánea, considerando una serie de criterios importantes para todas ellas. Entre ellos, intentamos balancear las distancias de viajes para todos los clubes mediante dos enfoques de resolución, uno en base a clusters y el otro a través de la consideración explícita de las distancias. Las soluciones entregadas por estos métodos han sido empleadas desde la temporada 2018 hasta la actualidad por la Superliga Argentina de Fútbol, proveyendo una serie de beneficios para todos los sectores involucrados.

Palabras Clave: Programación lineal entera, Sports scheduling, Travelling tournament problem.

1. Introducción

El fútbol es indiscutidamente uno de los deportes más populares del mundo. Practicado por unos 200 millones de jugadores en más de 200 naciones a nivel mundial, las principales ligas movilizan a cientos de millones de seguidores y generan grandes sumas de dinero. La última edición del Mundial Fifa, en Rusia 2018, generó beneficios por USD 5.3 billones, el Mundial de mayores ganancias hasta la fecha, y convocó una audiencia televisiva de 3.5 billones de espectadores. Mientras la mayor atención naturalmente se vuelca sobre las divisiones profesionales, el buen desarrollo de ligas juveniles es esencial para el sustento futuro del deporte profesional. Una gran cantidad de talentos se forjan en ligas juveniles, a menudo bajo el alero de los mismos clubes que participan en los torneos profesionales. Este es el caso de la Argentina, que con dos títulos mundiales en mayores es uno de los escasos seis países en el mundo cuyo seleccionado nacional ha sido campeón del mundo en más de una ocasión. El éxito a nivel profesional está indudablemente atado a la gran formación de jugadores en las divisiones inferiores, reflejado en los seis títulos mundiales conseguidos por su selección juvenil. Un sinnúmero de grandes jugadores han salido de las divisiones juveniles e infantiles de los clubes argentinos. Lionel Messi, por ejemplo, cuatro veces ganador del FIFA Balón de Oro y subcampeón del Mundo y mejor jugador en Brasil 2014, se inició en las divisiones infantiles del club Newell's Old Boys, y su temprano traspaso a Barcelona ocurrió directamente desde la liga infantil argentina que disputaba. Diego Maradona, campeón del Mundo y mejor jugador en México 86, por otra parte, surgió de las divisiones infantiles y juveniles de Argentinos Juniors, y tras un fugaz paso por Boca Juniors también recaló en Barcelona con sólo 21 años.

La colosal importancia entonces del desarrollo del deporte juvenil, motiva que la incorporación del uso de analytics y de investigación de operaciones, cuyo desarrollo en los últimos años se ha incrementado notablemente en torneos deportivos profesionales, llegue también a estas ligas. Este trabajo reporta la aplicación de técnicas de programación matemática para la calendarización de las ligas juveniles del fútbol masculino argentino. Uno de los principales desafíos enfrentados por los organizadores de estas ligas, es que la programación debe considerar simultáneamente seis ligas: sub-20, sub-18, sub-17, sub-16, sub-15 y sub-14. Además de una serie de condiciones habituales que se requieren para armar buenos fixtures deportivos, la programación de estas ligas intenta equiparar las distancias viajadas por las distintas divisiones de un mismo club. Todo esto genera entonces un problema que a través de los métodos manuales tradicionales usados por los organizadores se tornaba imposible de abordar eficientemente. En contraste, la adopción de investigación de operaciones ha venido a aliviar y mejorar considerablemente su tarea.

El trabajo está organizado de la siguiente manera. En la Sección 2 se describe un breve estado del arte en Sports Scheduling. La Sección 3 hace una descripción del problema. La sección 4 muestra el enfoque de resolución abordado, mientras que la Sección 5 describe en detalle los modelos de programación lineal entera implementados. La Sección 6 exhibe los resultados alcanzados, mientras que la última sección aborda las conclusiones y el posible trabajo futuro.

2. Sports Scheduling

La disciplina conocida como *sports scheduling* estudia la existencia y el diseño de fixtures de torneos deportivos, teniendo en cuenta diferentes condiciones requeridas por los organizadores. La gran variedad de formatos de torneos deportivos conduce a una gran cantidad de interesantes problemas de optimización, como lo refleja la reciente clasificación de problemas de la vida real en sports scheduling e implementación de un repositorio de instancias reales mostrada en [15]. El gran número de aplicaciones está también claramente reflejado en la variedad de artículos recopilados en ([8]). Limitándonos a recabar algunas de las publicaciones más recientes, [1] exhiben 12 años de experiencias programando el fútbol chileno y la expansión de ese trabajo a la programación de las clasificatorias sudamericanas para el Mundial de fútbol de Rusia 2018, posteriormente detallada en [6]. Por su parte, en [4] se ha recientemente reportado el uso de programación matemática para la programación

del torneo de vóleibol italiano.

Mientras la mayor atención en la literatura la han tenido ligas profesionales adultas, una pequeña minoría de papers se han enfocado en ligas no profesionales o juveniles, tales como hockey sobre hielo en Finlandia ([11]), fútbol en Bélgica ([14, 16]), cricket en Inglaterra ([17]), y tenis de mesa en Alemania ([13, 10]). Asimismo, mientras la mayoría de las aplicaciones se enfoca en la programación de una sola liga deportiva, unos pocos papers abordan el problema de programar multiple leagues. Entre estos, softbol en Estados Unidos ([9]), rugby en Nueva Zelanda ([3]), y tenis de mesa en Alemania ([12]).

Entre una variedad de criterios de optimización considerados en la literatura de sports scheduling, las distancias de viajes es uno de los principales. De hecho, el problema más frecuentemente estudiado por la literatura, el *Traveling Tournament Problem* ([7]) tiene como objetivo la minimización de la suma de las distancias recorridas por los equipos a lo largo de un torneo ficticio. Si bien la dificultad de dicho problema ha motivado una gran cantidad de trabajos metodológicos, su implementación en problemas reales es más bien escasa ([2, 5]).

En vista de esta revisión de literatura, nuestro trabajo busca contribuir con una nueva implementación práctica de investigación de operaciones para la calendarización de múltiples torneos juveniles programados en forma simultánea. Además, dentro de las varias condiciones consideradas, las distancias de viajes constituyen una parte central de nuestro criterio de optimización. En esta aplicación los organizadores buscan equiparar las distancias recorridas por las diferentes divisiones de un mismo club y dentro de un mismo torneo por parte de los equipos que pertenecen a la misma región geográfica. Ambos objetivos son importantes para darle mayor equidad a los campeonatos y para no generar desigualdades en los desgastes físicos entre los distintos clubes y entre las distintas divisiones del mismo club.

Para abordar el problema, desarrollamos dos enfoques basadas en programación matemática. Mientras en uno incorporamos explícitamente la minimización de las diferencias entre distancias recorridas por las diferentes divisiones de cada club, en otro enfoque novedoso generamos clusters de equipos por regiones geográficas y nos abocamos a equiparar el número de veces que cada equipo de un mismo club visita estos clusters. Estos enfoques traen como consecuencia la equiparación de las distancias recorridas por los equipos de la misma zona geográfica dentro de una misma liga, lo que, como fue expresado, es percibido como equitativo por los organizadores y que no necesariamente se refleja en el criterio de minimización de la suma de las distancias recorridas comúnmente usada en la literatura.

3. Descripción del problema

La principal categoría del fútbol juvenil argentino es disputada por los clubes que integran la primera división profesional del país. Cada club cuenta con seis divisiones juveniles, de las cuales tres son divisiones *mayores* (sub-20, sub-18 y sub-17) y tres son divisiones *menores* (sub-16, sub-15 y sub-14). Estos torneos tienen la particularidad de que las tres divisiones mayores juegan con el mismo fixture, mientras que las tres divisiones menores utilizan esa misma programación pero con las localías invertidas fecha a fecha.

El torneo 2017 estuvo compuesto por 30 equipos, que disminuyeron a 28 en el primer semestre de 2018 y a 26 para la temporada 2018-2019. En este trabajo mostraremos los modelos desarrollados y las soluciones entregadas para los torneos de divisiones juveniles disputados en el primer semestre de 2018 y para la temporada 2018-2019.

Los 30 equipos del Torneo 2017 se enfrentaron en un single round-robin, consagrándose campeón en cada categoría el equipo con más puntos.

En cambio, los 28 equipos que participaron en el Torneo de Transición 2018 fueron divididos en dos zonas de 14, denominadas *Zona A* y *Zona B* (ver Tabla 1). La Superliga Argentina de Fútbol (SAF), entidad que organiza los campeonatos, realizó un sorteo para determinar los integrantes de cada zona antes del inicio del torneo. Se establecieron 14 parejas de rivales clásicos, de modo tal que si uno de ellos quedaba sorteado en la Zona A entonces su rival era automáticamente asignado a la Zona B. Dentro de cada zona se jugó un *single round robin*, generando un total de 13 partidos por equipo (uno por cada fecha). Sin embargo, la SAF decidió añadir una fecha extra, que denominamos “fecha de clásicos”, llevando el torneo a 14 fechas. En esta fecha extra, cada equipo de la Zona A se enfrentaba con su rival clásico, que necesariamente pertenecía a la Zona B. La SAF estableció que esta fecha especial de clásicos fuera la número 7, relativamente a la mitad del torneo.

Para definir un ganador en cada división, los dos equipos que terminaron primeros en la Zona A se enfrentaron en una semifinal con los dos primeros de la Zona B, pero cruzados: es decir, el equipo mejor posicionado de la Zona A jugó contra el segundo de la Zona B, y viceversa. De estas semifinales surgieron dos equipos que finalmente disputaron la final en cada división, para así proclamar al ganador de la final en cada división como “Campeón del Torneo Transición de Juveniles 2018”.

Para la temporada 2018-2019 el esquema fue similar, pero con 26 equipos.

Tabla 1: Equipos del Torneo de Transición 2018.

| | Zona A | Zona B |
|----|--------------------|--------------------|
| 1 | River | Boca |
| 2 | San Lorenzo | Huracán |
| 3 | Vélez | Argentinos Jrs. |
| 4 | Tigre | Chacarita |
| 5 | Independiente | Racing |
| 6 | Banfield | Lanús |
| 7 | Defensa y Justicia | Arsenal |
| 8 | Estudiantes | Gimnasia y Esgrima |
| 9 | Rosario Central | Newell's |
| 10 | Colón | Unión |
| 11 | Talleres | Belgrano |
| 12 | San Martín SJ | Godoy Cruz |
| 13 | Atlético Tucumán | Patronato |
| 14 | Temperley | Olimpo |

De este modo, la Zona A quedó constituida por 13 equipos (nuevamente definidos por sorteo), mientras que en la Zona B quedaron sus 13 rivales clásicos. En esta oportunidad cada zona jugó nuevamente un single round-robin con el agregado de que el equipo libre en cada fecha enfrentó en dicha fecha a su rival clásico (que obligatoriamente debía quedar libre en la misma fecha). El Apertura de cada categoría se disputó con este formato de agosto a noviembre de 2018, mientras que el Clausura (disputado de marzo a junio de 2019) tuvo el mismo formato con las localías invertidas.

Al realizarse los fixtures en forma manual hasta 2017 solía existir una gran diferencia en kilómetros viajados entre las divisiones mayores y las menores de un mismo club (en algún caso esos kilómetros se triplicaban de una división a otra del mismo club), para la mayoría de los clubes, y entre equipos de la misma zona geográfica en un mismo campeonato.

El objetivo principal del presente trabajo fue entonces armar fixtures en los cuales se equipararan las distancias recorridas a lo largo de cada torneo entre las divisiones mayores y las menores de cada equipo, y balancear los kilómetros viajados por cada equipo de una misma zona geográfica en cada campeonato, sumado a cumplir con una serie de restricciones impuestas por los organizadores.

4. Enfoque de resolución

Analizamos el problema con 28 equipos, la modelación y resolución para el caso de 26 equipos sólo requiere de algunos ajustes. Para modelar el problema, usamos el enfoque basado en programación lineal entera que se describe en detalle en la próxima sección. La dificultad computacional de armar el fixture considerando a los 28 equipos juntos puede ser muy grande, por lo que decidimos utilizar las zonas ya sorteadas para descomponer el problema en dos. Resolvemos la programación para las divisiones mayores, dado que una vez resuelto ésta el fixture para las divisiones menores es el mismo invirtiendo las localías. Notar que si hubiera condiciones específicas para las divisiones menores, éstas pueden ser incorporadas trivialmente en la modelación de las divisiones mayores.

Para obtener la programación de las divisiones mayores, modelamos primero a la Zona A y luego a la Zona B. Notemos que estos fixtures no son independientes: una vez que tenemos los resultados de la Zona A, las localías de la “fecha de clásicos” (donde sólo hay partidos interzonales) ya quedan fijas. Esta información necesita ser incorporada en nuestro modelo para la Zona B. Notemos también que basta con programar sólo uno de los dos fixtures en cada temporada, por ejemplo el de las divisiones mayores, dado que el de las divisiones menores quedará completamente determinado al invertir las localías.

La necesidad de reducir las diferencias de viajes entre las divisiones mayores y menores de cada equipo admite más de una opción de modelamiento. En este trabajo consideramos las siguientes dos alternativas:

- La primera opción para garantizar que la diferencia de viajes entre divisiones mayores y menores de cada club sea reducida, es dividir a todos los equipos en grupos o “clusters” geográficos, de forma tal que queden agrupados por cercanía. La idea global es que cada equipo juegue la mitad de sus partidos de local y la mitad de visitante contra los equipos de los otros clusters y también contra los del suyo propio. De esta manera, el equipo I del cluster A juega dos partidos de local y dos de visita contra los equipos del cluster B (que imaginemos está conformado por cuatro equipos) en sus divisiones mayores. Como las menores de ese mismo equipo juegan con las localías invertidas, también jugarán dos partidos de local y dos de visita contra el cluster B. Al ser cercanos todos los equipos de un mismo cluster, las distancias recorridas entre las diferentes divisiones del equipo I serán razonablemente similares. A priori, no tenemos

forma de garantizar que exista una solución en la cual las localías se vean distribuidas tan equitativamente, por lo que la función objetivo se encargará de encontrar el mejor resultado posible en ese sentido.

- La segunda opción para armar un fixture que equipare los viajes realizados entre las divisiones mayores y las menores no utiliza los clusters geográficos e incorpora explícitamente la distancia (en kilómetros) entre los estadios de los clubes. Generamos una matriz de distancias entre los diferentes estadios, con la intención de plantear una función objetivo que minimice las diferencias de viajes entre las divisiones mayores y menores de cada equipo.

De antemano, se intuye que este último camino será más preciso pero también es esperable que sea más costoso en términos computacionales. Por ello proponemos avanzar con ambas alternativas y analizar los resultados que las mismas generan. Veremos más adelante además que estos objetivos traerán como consecuencia la equiparación de las distancias viajadas por equipos de la misma zona geográfica, dentro de cada uno de los campeonatos.

4.1. Zona A

El modelo incorpora condiciones que hacen a la estructura del torneo, así como también condiciones que fueron requeridas por la SAF para conseguir un fixture más justo y atractivo:

- El número máximo de partidos consecutivos que un equipo puede jugar con la misma localía es 2, lo que llamamos un break.
- No hay equipo que tenga más de un break de cada tipo a lo largo de todo el torneo: a lo sumo uno de visitante y uno de local.
- La segunda y la última fecha quedan libres de breaks.
- Los partidos entre los equipos grandes (River, San Lorenzo e Independiente en esta zona) no pueden darse ni en las 2 primeras ni en las 2 últimas fechas.
- No hay equipo que juegue dos partidos consecutivos contra un equipo grande.

Además, cada club tuvo la oportunidad de pedir sus propias condiciones, por ejemplo no tener viajes largos en determinadas fechas. La SAF, por su parte, también quiso tener en cuenta este aspecto al pedir que en una fecha entre semana ningún equipo se viera afectado por un viaje muy largo, ya

Tabla 2: Clusters geográficos para el Torneo de Transición 2018.

| | |
|--------------------------------|---|
| <u>Cluster CABA</u> | <u>Cluster Gran Buenos Aires</u> |
| River - Boca | Independiente - Racing |
| San Lorenzo - Huracán | Banfield - Lanús |
| Vélez - Argentinos Jrs. | Defensa y Justicia - Arsenal |
| Tigre - Chacarita | Estudiantes - Gimnasia y Esgrima |
| | |
| <u>Cluster Santa Fe</u> | <u>Cluster Córdoba/Cuyo</u> |
| Rosario Central - Newell's | Talleres - Belgrano |
| Colón - Unión | San Martín SJ - Godoy Cruz |
| | |
| <u>Cluster Extra</u> | |
| Atlético Tucumán - Patronato | |
| Temperley - Olimpo | |

que los jugadores son en su mayoría menores de edad y deben asistir a sus respectivas instituciones escolares.

Otra de las condiciones solicitadas fue no tener dos viajes largos ante un break de visita. Es decir, si un equipo juega dos partidos seguidos fuera de su estadio, al menos uno de ellos debe ser contra un adversario que esté geográficamente cerca.

4.1.1. Clusters geográficos

Como ya se mencionó, la primera opción para manejar las diferencias de viaje entre mayores y menores es dividir a los equipos en grupos o “clusters” geográficos, de forma tal que queden agrupados por cercanía. El armado de estos clusters fue realizado de manera manual, retomamos el tema sobre otras posibles composiciones de clusters en las conclusiones del trabajo. La situación ideal consiste en que cada equipo juegue la mitad de sus partidos de local y la mitad de visita contra los equipos de los otros clusters y también contra los del suyo propio.

En vista de los clubes participantes, los clusters quedaron conformados como especifica la Tabla 2. El mapa de la Figura 1 muestra la ubicación de cada uno.

Todos los clusters estuvieron conformados en cada zona por un número par de equipos, estableciéndose que si un equipo se asignaba a un determinado cluster en la zona A, su rival clásico estaría en el mismo cluster, pero en la zona B. Los equipos de la izquierda en la Tabla 2 son aquellos pertenecientes

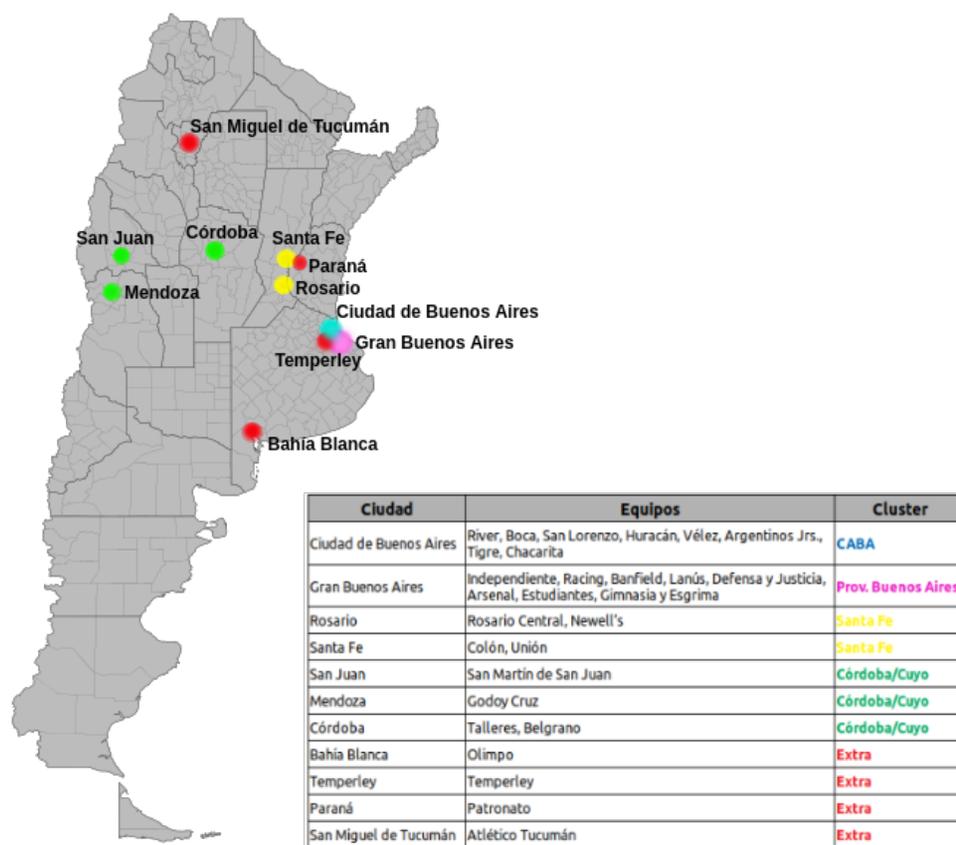


Figura 1: Clusters geográficos para el Torneo de Transición 2018.

a la Zona A mientras que los de la derecha son los de la Zona B.

Como se puede ver, los equipos dentro de un mismo cluster están cerca entre sí excepto los pertenecientes al “Cluster Extra”. Lo acordado en éste es que se sortearían en bloque: en una zona quedarían Patronato y Olimpo y en la otra Atlético Tucumán y Temperley. Se decidió esto pensando en los equipos de Ciudad de Buenos Aires (CABA) y del gran Buenos Aires (que son amplia mayoría): en una zona (finalmente fue la A) habría un viaje largo y uno corto (Atlético Tucumán, Temperley) para estos equipos, y en la otra zona, para compensar, habría dos viajes de distancia intermedia (Patronato, Olimpo), para estos mismos equipos.

Agrupados los equipos de esta forma, se busca que cada equipo juegue contra la mitad de los equipos de cada cluster con una localía y contra la otra mitad con la localía contraria. Como cada cluster está formado por un número par de equipos, se deben manejar de manera especial los partidos contra equipos del mismo cluster (que es una cantidad impar de partidos).

Para esto, el partido clásico (contra el rival de la otra zona) será el encargado de emparejar las localías. Por ejemplo, River debe jugar contra San Lorenzo, Vélez, Tigre y Boca (equipos de su propio cluster), siendo este último su contrincante clásico de la Zona B. Sería bueno que contra los tres equipos del cluster CABA jugara dos partidos de visitante (local) y uno de local (visitante), jugando entonces contra Boca también de local (visitante).

Hay un detalle para esta zona que no debemos pasar por alto: algunos equipos podrían tener que jugar de visitante en las mismas divisiones (las mayores por ejemplo) contra Atlético Tucumán y contra San Martín de San Juan, los dos clubes más alejados de la mayor parte de sus rivales (que a pesar de estar relativamente cerca entre sí, pertenecen a distintos clusters). Esto podría causar una diferencia considerable en la distancia recorrida entre las divisiones, especialmente para los equipos céntricos, ya que las otras divisiones (las menores) jugarían de local contra estos dos equipos. Imponemos entonces una restricción adicional solicitando que ninguno de estos equipos juegue con la misma localía contra estos dos clubes. Esta última observación da la idea de que se podrían haber considerado diferentes constituciones de clusters, tema que retomamos en las conclusiones del trabajo.

4.1.2. Matriz de distancias

La segunda opción para manejar las distancias de viaje es minimizar las diferencias de viaje entre las divisiones mayores y menores de un mismo club. Para esto, definimos dos posibles funciones objetivo: (a) minimizar la suma sobre todos los equipos de la diferencia de viaje entre sus divisiones mayores y menores y (b) minimizar la máxima de estas diferencias para todos los equipos.

Introducimos una matriz D de distancias entre los equipos. Claramente D es simétrica y los elementos de la diagonal valen 0.. Además, ubicamos a los equipos de forma tal que en las filas/columnas impares figuran las distancias de los equipos de la Zona A a los restantes, y en las pares las de los equipos de la Zona B. Más aún, el club de la fila/columna siguiente a cada club de la Zona A es su rival clásico.

4.2. Zona B

El modelo para diseñar el fixture de la Zona B es esencialmente igual que para la Zona A, teniendo en cuenta sólo algunas modificaciones dadas por el carácter secuencial del procedimiento propuesto.

Se modifica, por ejemplo, el conjunto de los equipos grandes, ya que en esta zona hay dos y no tres: Boca y Racing. Se mantienen los clusters, dado que hay

restricciones que los necesitan. Las funciones objetivo también se mantienen sin modificaciones.

Entre las restricciones que sufren algunos cambios podemos incluir a las de clubes y las específicas de la SAF, que básicamente se reformulan sobre la base de las necesidades de los equipos de esta zona. Las únicas restricciones completamente nuevas que debemos agregar en esta zona son las que expresan las localías de cada equipo en la fecha de clásicos (fecha 7), que como ya mencionamos, quedan determinadas por el fixture de la Zona A.

Considerando las modificaciones anteriormente presentadas, tendríamos un modelo completo para la Zona B sin tener en cuenta que debemos emparejar el kilometraje recorrido entre las divisiones mayores y menores de cada equipo. Para considerar esto, nuevamente se generarán tres modelos por separado, del mismo modo que se hizo para la Zona A.

5. Modelos de programación entera

Presentamos en esta sección los modelos planteados para la elaboración del fixture del torneo Transición 2018, con 28 equipos. Para representar los partidos contra los respectivos rivales clásicos de la Zona B, emplearemos en el modelo un equipo extra: el “equipo número 15”, que representará al adversario clásico.

Describimos en primer lugar el *modelo base* que nos servirá para atacar ambos enfoques. Para esto, introducimos los siguientes conjuntos:

- Fechas: $R = \{1, \dots, 14\}$.
- Equipos: $T = \{1, \dots, 15\}$.
- Equipos reales: $T^* = \{1, \dots, 14\}$.
- Equipos grandes: $BT = \{1, 2, 5\}$.
- CABA: $CABA = \{1, \dots, 4\}$.
- Buenos Aires: $BA = \{5, \dots, 8\}$.
- Santa Fe: $SF = \{9, 10\}$.
- Cordoba/Cuyo: $CC = \{11, 12\}$.
- Extra: $E = \{13, 14\}$.

El modelo requiere de las siguientes variables. Para todo equipo $i \in \mathbb{T}$, $j \in \mathbb{T}$ y para toda fecha $k \in \mathbb{R}$, se definen las variables

$$\begin{aligned} x_{ijk} &= \begin{cases} 1 & \text{si el equipo } i \text{ juega de local contra el equipo } j \text{ en la fecha } k, \\ 0 & \text{si no,} \end{cases} \\ y_{ik} &= \begin{cases} 1 & \text{si el equipo } i \text{ juega de local en las fechas } k \text{ y } k + 1, \\ 0 & \text{si no,} \end{cases} \\ w_{ik} &= \begin{cases} 1 & \text{si el equipo } i \text{ juega de visitante en las fechas } k \text{ y } k + 1, \\ 0 & \text{si no,} \end{cases} \\ z_{ij} &= \begin{cases} 1 & \text{si el equipo } i \text{ juega de local contra el equipo } j, \\ 0 & \text{si no.} \end{cases} \end{aligned}$$

Presentamos ahora las restricciones del modelo. En primer lugar, tenemos restricciones que relacionan las variables entre sí.

$$z_{ij} = \sum_{k \in \mathbb{R}} x_{ijk} \quad \forall i, j \in \mathbb{T} \quad (1)$$

$$\sum_{j \in \mathbb{T}} x_{ijk} + \sum_{l \in \mathbb{T}} x_{ilk+1} \leq 2(1 - w_{ik}) \quad \forall i \in \mathbb{T}^*, k \in \mathbb{R}, k \neq 14 \quad (2)$$

$$\sum_{j \in \mathbb{T}} x_{ijk} + \sum_{l \in \mathbb{T}} x_{ilk+1} \geq 1 - w_{ik} \quad \forall i \in \mathbb{T}^*, k \in \mathbb{R}, k \neq 14 \quad (3)$$

$$\sum_{j \in \mathbb{T}} x_{jik} + \sum_{l \in \mathbb{T}} x_{lik+1} \leq 2 \cdot (1 - y_{ik}) \quad \forall i \in \mathbb{T}^*, k \in \mathbb{R}, k \neq 14 \quad (4)$$

$$\sum_{j \in \mathbb{T}} x_{jik} + \sum_{l \in \mathbb{T}} x_{lik+1} \geq 1 - y_{ik} \quad \forall i \in \mathbb{T}^*, k \in \mathbb{R}, k \neq 14 \quad (5)$$

La variable z_{ij} se define en (1) y tiene por único objetivo simplificar la notación. Si el equipo i juega de local contra el equipo j en la fecha k , entonces obviamente el equipo i juega de local contra el equipo j .

Recordemos que la variable w_{ik} vale 1 si el equipo i tiene un break de visita en la fecha k y 0 si no. Si $w_{ik} = 1$, entonces la restricción (2) asegura que el equipo i no puede jugar de local ni en la fecha k ni en la $k+1$. Observemos que si $w_{ik} = 0$, la ecuación no nos dice nada que no supiéramos, con lo cual surge la necesidad de incorporar una nueva restricción para aclarar el rol de x_{ijk} en este caso. Por otra parte, si $w_{ik} = 0$, quiere decir que el equipo i no tiene break de visitante en la fecha k , o lo que es lo mismo, alguno de los dos partidos jugados en la fecha k o $k+1$ debe ser de local, condición asegurada por la

restricción (3). Las restricciones (4) y (5) son análogas pero ahora teniendo en cuenta los breaks de local.

Las restricciones (2)-(5) fueron pedidas para los equipos $i \in T^*$ y para las fechas $k \in R$ con $k \neq 14$. Esto se debe a que no hay breaks que comiencen en la fecha k (pues es la última) y a que el equipo ficticio sólo juega en una fecha en todo el torneo, con lo cual no es necesario hablar de sus breaks.

Presentamos a continuación las restricciones que definen la estructura global del torneo.

$$x_{iik} = 0 \quad \forall i \in T, k \in R \quad (6)$$

$$z_{ij} + z_{ji} = 1 \quad \forall i, j \in T, i \neq j \quad (7)$$

$$\sum_{\substack{j \in T \\ i \neq j}} x_{ijk} + x_{jik} = 1 \quad \forall i \in T^*, k \in R \quad (8)$$

$$\sum_{i \in T^*} \sum_{\substack{j \in T^* \\ j \neq i}} x_{ijk} = 7 \quad \forall k \in R, k \neq 7 \quad (9)$$

$$\sum_{\substack{j \in T \\ j \neq i}} z_{ij} = 7 \quad \forall i \in T \quad (10)$$

$$\sum_{j \in T} x_{15,j,7} = 7 \quad (11)$$

$$\sum_{j \in T} x_{j,15,7} = 7 \quad (12)$$

$$\sum_{j \in T} \sum_{\substack{k \in R \\ k \neq 7}} x_{j,15,k} + x_{15,j,k} = 0 \quad (13)$$

Las restricciones (6) especifican que ningún equipo debe jugar contra sí mismo. Las restricciones (7) piden que todos los equipos jueguen contra todos una única vez (recordemos que el modelo que estamos presentando es para la Zona A). Dado un equipo i (que no es el ficticio) y una fecha k , la restricción (8) especifica que i debe tener un único rival con quien enfrentarse, ya sea de local o de visita. La restricción (9) dice que en cada fecha deben jugarse 7 partidos (pues recordemos que son 14 equipos) y obviamente uno de los dos clubes que se enfrentan debe jugar de local. Para la fecha de clásicos y el equipo ficticio, la restricción será tratada de manera diferente, según veremos un poco más adelante.

La restricción (10) especifica que cada equipo $i \in T$ jugará 7 partidos de local y 7 de visitante a lo largo del torneo. Como además ya sabemos que

debe jugar en todas las fechas (por (8)) y que el total de fechas es 14, sería redundante pedir dos restricciones separadas para esto, así que, sin pérdida de generalidad, basta con pedir que cada equipo juegue 7 partidos de local.

Como se mencionó anteriormente, tenemos un tratamiento especial para el equipo número 15. Sabemos que debe jugar partidos sólo en la fecha 7 (que la SAF eligió como la fecha de clásicos), y para que resulte parejo para el resto de los equipos reales, las restricciones (11) y (12) piden que juegue 7 de visitante y 7 de local. Las restricciones (7), (11) y (12) implican que el equipo número 15 juega todos sus partidos en la fecha 7. Sin embargo, pedimos en la restricción (13) que no juegue en ninguna otra fecha, a los únicos efectos de evitar valores inválidos para esta variable.

Hemos visto hasta aquí todas las restricciones necesarias para el formato básico del torneo. Presentamos a continuación restricciones sobre los breaks y los equipos grandes, requeridas por la SAF.

$$y_{ik} + y_{ik+1} \leq 1 \quad \forall i \in T^*, k \in R, k \neq 14 \tag{14}$$

$$w_{ik} + w_{ik+1} \leq 1 \quad \forall i \in T^*, k \in R, k \neq 14 \tag{15}$$

$$\sum_{k \in R} y_{ik} \leq 1 \quad \forall i \in T^* \tag{16}$$

$$\sum_{k \in R} w_{ik} \leq 1 \quad \forall i \in T^* \tag{17}$$

$$y_{i1} + w_{i1} + y_{i13} + w_{i13} = 0 \quad \forall i \in T^* \tag{18}$$

$$x_{ij1} + x_{ji1} + x_{ij2} + x_{ji2} + x_{ij13} + x_{ji13} + x_{ij14} + x_{ji14} = 0 \quad \forall i, j \in BT \tag{19}$$

$$1 - x_{ijk+1} - x_{jik+1} \geq x_{ilk} + x_{lik} \quad \forall i \in T^*, j, l \in BT, k \in R \tag{20}$$

$$\sum_{j \in CABAUBA} x_{1j3} + x_{j1,3} = 1 \tag{21}$$

$$\sum_{j \in CABAUBA} x_{3j3} + x_{j3,3} = 1 \tag{22}$$

$$\sum_{j \in CABAUBA} x_{2j1} + x_{j2,1} = 1 \tag{23}$$

$$\sum_{j \in CABAUBA} x_{2j6} + x_{j2,6} = 1 \tag{24}$$

$$\sum_{j \in CABAUBA} x_{14j1} + x_{j14,1} = 1 \tag{25}$$

$$\sum_{j \in CABAUBA} x_{14j4} + x_{j14,4} = 1 \tag{26}$$

$$\sum_{j \in \text{CABAUBA}} x_{14j6} + x_{j14,6} = 1 \quad (27)$$

$$x_{13,12,11} + x_{12,13,11} = 1 \quad (28)$$

$$x_{9,4,11} + x_{4,9,11} = 1 \quad (29)$$

$$x_{11,10,11} + x_{10,11,11} = 1 \quad (30)$$

Nuevamente, no tiene sentido hablar de breaks para el equipo ficticio ya que sabemos que juega sólo en una fecha; por lo que estas restricciones son sólo para los equipos que pertenezcan al conjunto $T^* = \{1, \dots, 14\}$.

Como se impuso que ningún equipo juegue más de dos partidos consecutivos con la misma localía, las restricciones (14) y (15) impiden que existan equipos con el mismo tipo de break en dos fechas consecutivas, ya que esto implicaría jugar al menos tres partidos seguidos de local o tres seguidos de visitante. Las restricciones (16) y (17) piden que no haya más de un break de cada tipo por equipo a lo largo del torneo. Es decir, ningún equipo puede jugar dos veces seguidas de local más de una vez en todo el torneo, ni tampoco de visitante. Por su parte, las restricciones (18) especifican que no hay breaks en la primer ni en la última fecha; es decir, ningún equipo juega con la misma localía en las fechas 1 y 2 ni en las fechas 13 y 14.

Las restricciones (19) solicitan que no haya enfrentamientos entre los equipos considerados grandes (River, San Lorenzo e Independiente en esta zona) ni en las dos primeras ni en las dos últimas fechas, con el fin de reservar estos partidos “atractivos” para otra etapa del torneo. Las restricciones (20) dicen que ningún equipo real debe jugar dos partidos seguidos contra un equipo grande, sin tener en cuenta las localías.

Finalmente, se tienen restricciones solicitadas por los equipos:

- Restricción (21): River solicitó no tener un viaje largo en la tercera fecha, es decir, jugar contra un equipo de CABA o de BA.
- Restricciones (22) – (27). Análogas a (21) pero para Vélez en la tercera fecha, para San Lorenzo en la primera y en la sexta, y para Temperley en la primera, la cuarta y la sexta, respectivamente.
- Restricción (28): Atlético Tucumán debe jugar contra San Martín de San Juan en la fecha 11.
- Restricción (29) y (30): Análogas pero para Rosario Central vs. Tigre y para Colón vs. Talleres.

Para terminar, presentamos restricciones adicionales que solicitan que los equipos no tengan dos viajes largos en caso de que les toque un break de visita.

$$w_{ik} \leq \left(\sum_{j \in \text{CABA} \cup \text{BA}} x_{jik} + x_{j,i,k+1} \right) + x_{15,i,k} + x_{15,i,k+1} \quad \forall i \in \text{CABA} \cup \text{BA},$$

$$\forall k \in R, k \neq 14 \quad (31)$$

$$w_{ik} \leq \left(\sum_{j \in \text{SF}} x_{jik} + x_{j,i,k+1} \right) + x_{15,i,k} + x_{15,i,k+1} \quad \forall i \in \text{SF}, \quad \forall k \in R, k \neq 14$$

$$(32)$$

$$w_{11k} \leq \left(\sum_{j \in \text{CCUSF}} x_{j,11,k} + x_{j,11,k+1} \right) + x_{15,11,k} + x_{15,11,k+1} \quad \forall k \in R, k \neq 14$$

$$(33)$$

$$w_{14k} \leq \left(\sum_{j \in \text{CABA} \cup \text{BA}} x_{j,14,k} + x_{j,14,k+1} \right) \quad \forall k \in R, k \neq 14 \quad (34)$$

En el caso de (34) ya no permitimos que se juegue el partido clásico en una de las dos fechas involucradas en el break, ya que el rival en este partido para Temperley es Olimpo (en Bahía Blanca), que no lo consideramos cercano. Notemos que debemos pedir también estas restricciones para breaks de local, ya que el fixture que estamos modelando es el de las mayores pero las menores juegan con las localías invertidas, por lo que un break de local para nuestro fixture es uno de visitante para el otro:

$$y_{ik} \leq \left(\sum_{j \in \text{CABA} \cup \text{BA}} x_{ijk} + x_{i,j,k+1} \right) + x_{i,15,k} + x_{i,15,k+1} \quad \forall i \in \text{CABA} \cup \text{BA},$$

$$\forall k \in R, k \neq 14 \quad (35)$$

$$y_{ik} \leq \left(\sum_{j \in \text{SF}} x_{ijk} + x_{i,j,k+1} \right) + x_{i,15,k} + x_{i,15,k+1} \quad \forall i \in \text{SF}, \quad \forall k \in R, k \neq 14$$

$$(36)$$

$$y_{11k} \leq \left(\sum_{j \in \text{CCUSF}} x_{11,j,k} + x_{11,j,k+1} \right) + x_{11,15,k} + x_{11,15,k+1} \quad \forall k \in R, k \neq 14 \quad (37)$$

$$y_{14k} \leq \left(\sum_{j \in \text{CABA} \cup \text{BA}} x_{14,j,k} + x_{14,j,k+1} \right) \quad \forall k \in R, k \neq 14 \quad (38)$$

Una última restricción adicional (sólo para el modelo de los clusters) es que la localía con que los equipos de CABA y de BA jueguen contra San Martín de San Juan y Atlético Tucumán sea distinta, ya que estos dos equipos están alejados de dichos clubes pero pertenecen a diferentes clusters. En el caso de permitir esto, se podrían presentar diferencias grandes entre divisiones de los equipos más céntricos.

$$z_{12i} = z_{i13} \quad \forall i \in \text{CABA} \cup \text{BA} \quad (39)$$

5.1. Modelo por clusters

Como ya mencionamos, el objetivo en este modelo es que cada equipo juegue la mitad de sus partidos de local y la mitad de visita contra los equipos de cada cluster (utilizando el partido interzonal para equiparar los partidos dentro de su propio cluster). Buscamos que la función objetivo reduzca la diferencia de cantidad de partidos con cada localía contra su propio cluster y contra los otros. En el caso ideal, el valor de la función objetivo sería 0.

Para esto, incorporamos al modelo las siguientes variables:

$$\begin{aligned} & Caba_i^+, Caba_i^- \quad \forall i \in T^*. \\ & intraCaba_i^+, intraCaba_i^- \quad \forall i \in \text{CABA}. \\ & BA_i^+, BA_i^- \quad \forall i \in T^*. \\ & intraBA_i^+, intraBA_i^- \quad \forall i \in \text{BA}. \\ & SF_i^+, SF_i^- \quad \forall i \in T^*. \\ & intraSF_i^+, intraSF_i^- \quad \forall i \in \text{SF}. \\ & CC_i^+, CC_i^- \quad \forall i \in T^*. \\ & intraCC_i^+, intraCC_i^- \quad \forall i \in \text{CC}. \\ & E_i^+, E_i^- \quad \forall i \in T^*. \\ & intraE_i^+, intraE_i^- \quad \forall i \in \text{E}. \end{aligned}$$

$Caba_i^+ + Caba_i^-$ representará la diferencia entre la cantidad de partidos que el equipo i (no perteneciente a CABA) jugará de visitante y los que jugará de local contra los equipos de CABA. Por su parte, $intraCaba_i^+ + intraCaba_i^-$ será la diferencia entre la cantidad de partidos que el equipo i (perteneciente

a CABA) jugará de visitante y los que jugará de local contra su propio cluster (contabilizando al interzonal como uno de estos partidos). Las variables sobre los equipos del resto de los clusters se definen de manera similar.

Ligamos estas nuevas variables con las anteriores por medio de las siguientes restricciones:

$$Caba_i^+ - Caba_i^- = \sum_{j \in CABA} z_{ij} - z_{ji} \quad \forall i \in T^*$$

$$intraCaba_i^+ - intraCaba_i^- = \sum_{\substack{j \in CABA \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \quad \forall i \in CABA$$

$$BA_i^+ - BA_i^- = \sum_{j \in BA} z_{ij} - z_{ji} \quad \forall i \in T^*$$

$$intraBA_i^+ - intraBA_i^- = \sum_{\substack{j \in BA \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \quad \forall i \in BA$$

$$SF_i^+ - SF_i^- = \sum_{j \in SF} z_{ij} - z_{ji} \quad \forall i \in T^*$$

$$intraSF_i^+ - intraSF_i^- = \sum_{\substack{j \in SF \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \quad \forall i \in SF$$

$$CC_i^+ - CC_i^- = \sum_{j \in CC} z_{ij} - z_{ji} \quad \forall i \in T^*$$

$$intraCC_i^+ - intraCC_i^- = \sum_{\substack{j \in CC \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \quad \forall i \in CC$$

$$E_i^+ - E_i^- = \sum_{j \in E} z_{ij} - z_{ji} \quad \forall i \in T^*$$

$$intraE_i^+ - intraE_i^- = \sum_{\substack{j \in E \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \quad \forall i \in E$$

Como podemos observar, el lado derecho de estas restricciones representa un valor que puede ser negativo, positivo o cero (es la diferencia entre los partidos de local y los de visitante que juega el equipo i contra el conjunto formado

por los equipos j correspondientes). La necesidad de incorporar variables con supraíndices “+” y “-” surge de la necesidad de representar el módulo de estos valores en la función objetivo:

$$\begin{aligned}
\text{mín} \quad & \sum_{i \in T^* \setminus CABA} \left| \sum_{j \in CABA} z_{ij} - z_{ji} \right| + \sum_{i \in T^* \setminus BA} \left| \sum_{j \in BA} z_{ij} - z_{ji} \right| \\
& + \sum_{i \in T^* \setminus SF} \left| \sum_{j \in SF} z_{ij} - z_{ji} \right| + \sum_{i \in T^* \setminus CC} \left| \sum_{j \in CC} z_{ij} - z_{ji} \right| \\
& + \sum_{i \in T^* \setminus E} \left| \sum_{j \in E} z_{ij} - z_{ji} \right| + \sum_{i \in CABA} \left| \sum_{\substack{j \in CABA \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \right| \\
& + \sum_{i \in BA} \left| \sum_{\substack{j \in BA \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \right| + \sum_{i \in SF} \left| \sum_{\substack{j \in SF \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \right| \\
& + \sum_{i \in CC} \left| \sum_{\substack{j \in CC \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \right| + \sum_{i \in E} \left| \sum_{\substack{j \in E \cup \{15\} \\ j \neq i}} z_{ij} - z_{ji} \right|.
\end{aligned}$$

De acuerdo con las definiciones anteriores, podemos reescribir la función objetivo del siguiente modo:

$$\begin{aligned}
\text{mín} \quad & \sum_{i \in T^* \setminus CABA} |Caba_i^+ - Caba_i^-| + \sum_{i \in T^* \setminus BA} |BA_i^+ - BA_i^-| \\
& + \sum_{i \in T^* \setminus SF} |SF_i^+ - SF_i^-| + \sum_{i \in T^* \setminus CC} |CC_i^+ - CC_i^-| \\
& + \sum_{i \in T^* \setminus E} |E_i^+ - E_i^-| + \sum_{i \in CABA} |intraCABA_i^+ - intraCABA_i^-| \\
& + \sum_{i \in BA} |intraBA_i^+ - intraBA_i^-| + \sum_{i \in SF} |intraSF_i^+ - intraSF_i^-| \\
& + \sum_{i \in CC} |intraCC_i^+ - intraCC_i^-| + \sum_{i \in E} |intraE_i^+ - intraE_i^-|.
\end{aligned}$$

Estas funciones no son lineales debido a que utilizan el valor absoluto. Para poder linealizarlas, utilizaremos que

$$\text{mín} \left\{ \sum_{i=1}^n |x_i| : Ax \leq b, x \in \mathbb{R}^n \right\}$$

es equivalente a

$$\text{mín} \left\{ \sum_{i=1}^n x_i^+ + x_i^- : A(x^+ - x^-) \leq b, x_i^+, x_i^- \geq 0 \text{ para } i = 1, \dots, n \right\}.$$

Luego, podemos reescribir la función objetivo como:

$$\begin{aligned}
\text{mín} \quad & \sum_{i \in T^* \setminus \text{CABA}} Caba_i^+ + Caba_i^- + \sum_{i \in T^* \setminus \text{BA}} BA_i^+ + BA_i^- \\
& + \sum_{i \in T^* \setminus \text{SF}} SF_i^+ + SF_i^- + \sum_{i \in T^* \setminus \text{CC}} CC_i^+ + CC_i^- \\
& + \sum_{i \in T^* \setminus \text{E}} E_i^+ + E_i^- + \sum_{i \in \text{CABA}} \text{intraCABA}_i^+ + \text{intraCABA}_i^- \\
& + \sum_{i \in \text{BA}} \text{intraBA}_i^+ + \text{intraBA}_i^- + \sum_{i \in \text{SF}} \text{intraSF}_i^+ + \text{intraSF}_i^- \\
& + \sum_{i \in \text{CC}} \text{intraCC}_i^+ + \text{intraCC}_i^- + \sum_{i \in \text{E}} \text{intraE}_i^+ + \text{intraE}_i^-.
\end{aligned}$$

5.2. Minimización de diferencias de viaje

Además de las variables previamente definidas (sin contar las que surgieron para el modelo de los clusters), debemos incorporar una variable d_i por cada equipo i , que represente la diferencia entre lo recorrido (sólo viaje de ida) por las divisiones mayores y por las menores:

$$d_i = \sum_{j \in T^*} D_{2i-1, 2j-1}(2z_{ji} - 1) + D_{2i-1, 2i}(2z_{15, i} - 1)$$

En la definición de d_i contemplamos sólo el kilometraje de ida porque considerar la vuelta sería multiplicar por una constante nuestra función objetivo, lo cual no afecta la solución óptima. Si el equipo i juega de visitante contra el equipo j , entonces sumaremos la distancia entre ellos (pues $z_{ji} = 1$, lo que hace que $2z_{ji} - 1$ también tome el valor 1 y entonces se suma el valor de la matriz $D_{2i-1, 2j-1}$, que al pertenecer i y j a la Zona A y al conjunto $\{1, \dots, 14\}$ corresponde a la distancia entre los estadios de estos dos equipos). Por el contrario, si i juega de local contra j , entonces $2z_{ji} - 1 = -1$ y esto hará que la distancia antes mencionada se reste. Esto es equivalente a sumar los viajes realizados por una división y a restar los realizados por la contraria, ya que como habíamos dicho, juegan con las localías invertidas. Para el caso del partido especial de la fecha 7, si i lo juega de visitante entonces $2z_{15, i} - 1 = 1$ y se sumará la distancia entre el equipo i y su rival clásico ($D_{2i-1, 2i}$), mientras que se restará en caso de que aquel partido i lo juegue de local. Luego, d_i es efectivamente la resta entre los kilómetros realizados por ambas divisiones (ida) para un equipo $i \in T^*$. Es preciso observar que d_i no es necesariamente positiva.

Se agregan también las variables d_i^+ y d_i^- para cada equipo i , que se definen como sigue.

$$d_i^+ - d_i^- = \sum_{j \in T^*} D_{2i-1, 2j-1} \cdot (2z_{ji} - 1) + D_{2i-1, 2i} \cdot (2z_{15i} - 1) \quad \forall i \in T^*$$

$$d_i^+, d_i^- \geq 0 \quad \forall i \in T^*$$

Con estas definiciones, podemos plantear las dos funciones objetivo con las que trabajamos. Para el primer modelo (suma de diferencias), tenemos la función objetivo

$$\min \sum_{i \in T^*} |d_i|,$$

mientras que para el segundo modelo (minimizar la máxima diferencia), tenemos la función objetivo

$$\min \max_{i \in T^*} |d_i|.$$

Notemos que estas funciones no son lineales debido a la presencia del valor absoluto en ambas y del máximo en el caso de la segunda. En consecuencia, debemos reformularlas para poder cumplir con los requisitos necesarios de este tipo de modelado. Para linealizar la primera función objetivo, sólo debemos reescribir el módulo como ya detallamos en el modelo anterior. Para linealizar la segunda función objetivo, utilizaremos el hecho de que el modelo

$$\min \left\{ \max\{|x_1|, \dots, |x_n|\} : Ax \geq b, x \in \mathbb{R}^n \right\}$$

es equivalente al modelo

$$\min \left\{ M : Ax \leq b, x_i \leq M, -M \leq x_i \text{ para } i = 1, \dots, n \right\},$$

que a su vez es equivalente a

$$\min \left\{ M : A(x^+ - x^-) \leq b, x_i^+ + x_i^- \leq M, x_i^+, x_i^- \geq 0 \text{ para } i = 1, \dots, n \right\}.$$

De esta forma, se pueden plantear ambas funciones objetivo de manera lineal para los modelos que tienen en cuenta las distancias precisas de viaje.

6. Resultados

Todos los modelos se resolvieron con Cplex 12.4, en un computador con procesador Intel CORE i7 y 8 GB de memoria RAM. Las instancias de los modelos de clusters fueron resueltas en todos los casos a optimalidad, siempre con función objetivo igual a 0 (es decir, consiguiendo equiparar exactamente los partidos de local y visita de cada equipo frente a sus rivales de cada cluster), en tiempos de corridas que van de 30 minutos a 10 horas. Mientras que las instancias de los modelos de matriz de distancias (Suma y Máximo) fueron corridas hasta 24 hs en cada caso, guardando la mejor solución obtenida hasta ese momento. En la mayor parte de los casos estos modelos no fueron resueltos a optimalidad, aunque se obtuvieron soluciones altamente satisfactorias a ojos de nuestra contraparte.

Las Tablas 3 y 4 muestran los resultados para el torneo Transición 2018 (que tuvo 28 equipos), mientras que las Tablas 5 y 6 muestran los resultados para el torneo Apertura 2018-2019 (que tuvo 26 equipos). Para cada uno de los tres modelos propuestos y para cada zona de los torneos, reportamos la distancia (ida y vuelta en kilómetros) que deben recorrer tanto las divisiones mayores como las menores de cada club. Especificamos además la suma de diferencias en los viajes entre divisiones y la máxima de ellas (siendo estas últimas dos las funciones objetivo consideradas).

Tabla 3: Resultados para el Torneo Transición 2018, Zona A.

| Equipos | Clusters | | | Matriz (suma) | | | Matriz (máximo) | | |
|--------------------|---------------------------|----------|----------------|----------------------------|----------|----------------|-------------------|----------|----------------|
| | Men. | May. | Dif. | Men. | May. | Dif. | Men. | May. | Dif. |
| River | 5.014,2 | 3.242,6 | 54,64 % | 4.125,8 | 4.131,0 | 0,13 % | 4.080,4 | 4.176,4 | 2,35 % |
| San Lorenzo | 3.063,6 | 5.179,4 | 69,06 % | 4.094,2 | 4.148,8 | 1,33 % | 4.244,6 | 3.998,4 | 6,16 % |
| Vélez | 3.225,6 | 5.179,4 | 55,08 % | 4.072,2 | 4.155,8 | 2,05 % | 4.592,2 | 3.635,8 | 26,31 % |
| Tigre | 5.155,8 | 3.255,8 | 58,36 % | 4.225,0 | 4.186,6 | 0,92 % | 4.233,6 | 4.178,0 | 1,33 % |
| Independiente | 5.073,8 | 3.208,4 | 58,14 % | 4.147,8 | 4.134,4 | 0,32 % | 4.234,0 | 4.048,2 | 4,59 % |
| Banfield | 3.272,8 | 5.100,4 | 55,84 % | 4.186,2 | 4.187,0 | 0,02 % | 4.298,8 | 4.074,4 | 5,51 % |
| Defensa y Justicia | 3.303,4 | 5.360,2 | 62,26 % | 4.345,6 | 4.318,0 | 0,64 % | 4.132,8 | 4.530,8 | 9,63 % |
| Estudiantes | 5.652,8 | 3.759,8 | 50,35 % | 4.704,8 | 4.707,8 | 0,06 % | 4.729,4 | 4.683,2 | 0,99 % |
| Rosario Central | 5.450,0 | 5.336,4 | 2,13 % | 5.076,4 | 5.710,0 | 12,48 % | 5.320,4 | 5.466,0 | 2,74 % |
| Colón | 6.129,2 | 7.11,08 | 16,13 % | 6.597,2 | 6.650,0 | 0,80 % | 6.436,0 | 6.811,0 | 5,83 % |
| Talleres | 9.202,0 | 7.596,2 | 21,14 % | 8.308,2 | 8.490,0 | 2,19 % | 8.406,2 | 8.392,0 | 0,17 % |
| San Martín SJ | 13.400,0 | 14.704,0 | 9,73 % | 13.330,0 | 14.774,0 | 10,83 % | 13.944,0 | 14.160,0 | 1,55 % |
| Atlético Tucumán | 15.722,0 | 14.918,0 | 5,39 % | 15.268,0 | 15.372,0 | 0,68 % | 14.896,0 | 15.744,0 | 5,69 % |
| Temperley | 4.899,8 | 4.776,4 | 2,58 % | 4.826,6 | 4.849,6 | 0,48 % | 5.034,0 | 4.642,2 | 8,44 % |
| Suma de difs. | 20.146,6 km. | | | 2.665,8 km. | | | 4.199,4 km. | | |
| Max. dif. absoluta | 2.115,8 km. (San Lorenzo) | | | 1.444 km. (S. M. San Juan) | | | 956,4 km. (Vélez) | | |

Tabla 4: Resultados para el Torneo Transición 2018, Zona B.

| Equipos | Clusters | | | Matriz (suma) | | | Matriz (máximo) | | |
|--------------------|-------------------------|----------|---------|-------------------|----------|---------|----------------------|----------|---------|
| | Men. | May. | Dif. | Men. | May. | Dif. | Men. | May. | Dif. |
| Boca | 4.277,4 | 3.496,0 | 23,30 % | 3.900,6 | 3.845,8 | 1,42 % | 4.009,8 | 3.736,6 | 7,31 % |
| Huracán | 3.949,2 | 3.717,6 | 6,23 % | 3.604,4 | 4.062,4 | 12,71 % | 3.351,0 | 4.315,8 | 28,79 % |
| Argentinos Jrs. | 3.715,4 | 3.893,2 | 4,79 % | 3.922,4 | 3.686,2 | 6,41 % | 3.371,0 | 4.237,6 | 25,71 % |
| Chacarita | 3.506,8 | 4.271,4 | 21,80 % | 3.883,4 | 3.894,8 | 0,29 % | 4.312,4 | 3.465,8 | 24,43 % |
| Racing | 3.908,8 | 3.811,2 | 2,56 % | 3.724,0 | 3.996,0 | 7,30 % | 3.832,8 | 3.887,2 | 1,42 % |
| Lanús | 4.201,4 | 3.538,6 | 18,73 % | 3.857,8 | 3.882,2 | 0,63 % | 3.929,9 | 3.811,0 | 3,10 % |
| Arsenal | 3.627,0 | 4.250,8 | 17,20 % | 4.304,8 | 3.573,0 | 20,48 % | 3.803,8 | 4.074,0 | 7,10 % |
| Gimnasia y Esgrima | 4.248,6 | 4.487,2 | 5,62 % | 4.326,8 | 4.409,0 | 1,90 % | 4.186,8 | 4.549,0 | 8,65 % |
| Newell's | 4.732,4 | 5.182,0 | 9,50 % | 5.172,0 | 4.742,4 | 9,06 % | 5.118,0 | 4.796,4 | 6,71 % |
| Unión | 6.686,0 | 5.780,2 | 15,67 % | 6.680,0 | 5.786,2 | 15,45 % | 6.389,2 | 6.077,0 | 5,14 % |
| Belgrano | 8.392,2 | 8.584,0 | 2,29 % | 8.626,0 | 8.350,2 | 3,30 % | 8.646,0 | 8.330,2 | 3,79 % |
| Godoy Cruz | 14.008,0 | 13.140,0 | 6,61 % | 13.602,0 | 13.546,0 | 0,41 % | 13.984,0 | 13.164,0 | 6,23 % |
| Patronato | 6.213,0 | 8.022,0 | 29,12 % | 7.445,0 | 6.790,0 | 9,65 % | 7.459,0 | 6.776,0 | 10,08 % |
| Olimpo | 10.866,0 | 10.192,0 | 6,61 % | 10.540,0 | 10.518,0 | 0,21 % | 9.922,0 | 11.136,0 | 12,42 % |
| Suma de difs. | 8.503,4 km. | | | 4.203,0 km. | | | 7.422,6 km. | | |
| Max. dif. absoluta | 1.809,0 km. (Patronato) | | | 893,8 km. (Unión) | | | 1.214,0 km. (Olimpo) | | |

Tabla 5: Resultados para el Torneo Apertura 2018-2019, Zona A.

| Equipos | Clusters | | | Matriz (suma) | | | Matriz (máximo) | | |
|--------------------|--------------------------------|----------|---------|------------------------|----------|---------|----------------------------------|----------|---------|
| | Men. | May. | Dif. | Men. | May. | Dif. | Men. | May. | Dif. |
| Boca | 4.602,8 | 4.230,4 | 8,80 % | 4.747,2 | 4.086,0 | 16,18 % | 4.634,8 | 4.198,4 | 10,39 % |
| Argentinos Jrs. | 4.232,4 | 4.572,4 | 8,03 % | 4.165,6 | 4.639,2 | 11,37 % | 4.464,0 | 4.340,8 | 2,84 % |
| Huracán | 4.212,0 | 4.606,8 | 9,37 % | 4.372,8 | 4.446,0 | 1,67 % | 4.410,4 | 4.408,4 | 0,05 % |
| Racing | 4.715,6 | 4.104,2 | 14,90 % | 4.965,6 | 3.854,2 | 28,84 % | 4.134,6 | 4.685,2 | 13,32 % |
| Lanús | 4.102,6 | 4.756,2 | 15,93 % | 4.167,0 | 4.691,8 | 12,59 % | 4.221,2 | 4.637,6 | 9,86 % |
| Defensa y Justicia | 5.016,4 | 4.176,2 | 20,12 % | 4.374,0 | 4.818,6 | 10,16 % | 4.037,8 | 5.154,8 | 27,66 % |
| Gimnasia y Esgrima | 4.867,8 | 4.787,0 | 1,69 % | 5.074,6 | 4.580,2 | 10,79 % | 5.021,6 | 4.633,2 | 8,38 % |
| Aldosivi | 9.140,0 | 9.662,0 | 5,71 % | 8.352,0 | 10.450,0 | 25,12 % | 9.432,0 | 9.370,0 | 0,66 % |
| Rosario Central | 4.860,0 | 5.948,4 | 22,40 % | 5.212,0 | 5.596,4 | 7,38 % | 5.842,4 | 4.966,0 | 17,65 % |
| Colón | 7.018,0 | 6.029,2 | 16,40 % | 6.580,0 | 6.467,2 | 1,74 % | 6.379,2 | 6.668,0 | 4,53 % |
| Belgrano | 8.414,2 | 7.912,0 | 6,35 % | 8.716,0 | 7.610,2 | 14,53 % | 7.724,0 | 8.602,2 | 11,37 % |
| Godoy Cruz | 11.948,0 | 13.568,0 | 13,56 % | 12.760,0 | 12.756,0 | 0,03 % | 12.856,0 | 12.660,0 | 1,55 % |
| S. M. de Tucumán | 15.392,0 | 12.003,2 | 28,23 % | 13.290,0 | 14.105,2 | 6,13 % | 13.648,0 | 13.747,2 | 0,73 % |
| Suma de difs. | 11.403,4 km. | | | 8.303,4 km. | | | 5.434,6 km. | | |
| Max. dif. absoluta | 3.388,8 km. (S. M. de Tucumán) | | | 2.098,0 km. (Aldosivi) | | | 1.117,0 km. (Defensa y Justicia) | | |

Tabla 6: Resultados para el Torneo Apertura 2018-2019, Zona B.

| Equipos | Clusters | | | Matriz (suma) | | | Matriz (máximo) | | |
|--------------------|--------------------------------|----------|---------|---------------------|----------|---------|-----------------|----------|---------|
| | Men. | May. | Dif. | Men. | May. | Dif. | Men. | May. | Dif. |
| River | 4.084,2 | 4.890,2 | 19,73 % | 3.525,2 | 5.449,2 | 54,58 % | 4.528,2 | 4.446,2 | 1,84 % |
| Vélez | 4.923,2 | 4.042,0 | 21,80 % | 4.546,8 | 4.418,4 | 2,91 % | 4.428,6 | 4.536,6 | 2,44 % |
| San Lorenzo | 4.885,8 | 4.148,8 | 17,76 % | 5.533,6 | 3.501,0 | 58,06 % | 4.407,2 | 4.627,4 | 5,00 % |
| Independiente | 4.094,6 | 5.013,2 | 22,43 % | 4.553,2 | 4.554,6 | 0,03 % | 4.895,4 | 4.212,4 | 16,21 % |
| Banfield | 4.279,6 | 4.969,6 | 16,12 % | 4.600,0 | 4.649,2 | 1,07 % | 4.670,6 | 4.578,6 | 2,01 % |
| Tigre | 4.108,4 | 5.015,6 | 22,08 % | 5.948,2 | 3.175,8 | 87,30 % | 4.981,6 | 4.142,4 | 20,26 % |
| Estudiantes | 5.471,6 | 4.738,4 | 15,47 % | 5.061,2 | 5.148,8 | 1,73 % | 5.169,0 | 5.041,0 | 2,54 % |
| Patronato | 6.202,0 | 6.929,0 | 11,72 % | 6.732,0 | 6.399,0 | 5,20 % | 6.570,0 | 6.561,0 | 0,14 % |
| Newell's | 4.992,4 | 4.690,0 | 6,45 % | 4.586,4 | 5.096,0 | 11,11 % | 4.988,0 | 4.694,4 | 6,25 % |
| Unión | 5.112,2 | 6.162,0 | 20,54 % | 5.236,2 | 6.038,0 | 15,31 % | 5.759,0 | 5.515,2 | 4,42 % |
| Talleres | 7.564,0 | 7.092,2 | 6,65 % | 7.612,2 | 7.044,0 | 8,07 % | 7.064,2 | 7.592,0 | 7,47 % |
| San Martín SJ | 11.948,0 | 13.460,0 | 12,65 % | 12.580,0 | 12.828,0 | 1,97 % | 12.608,0 | 12.800,0 | 1,52 % |
| Atlético Tucumán | 13.423,2 | 12.104,0 | 10,90 % | 12.319,2 | 13.208,0 | 7,21 % | 12.735,2 | 12.792,0 | 0,45 % |
| Suma de difs. | 11.055,4 km | | | 10.345,0 km. | | | 3.475,4 km. | | |
| Max. dif. absoluta | 1.512,0 km. (S.M. de San Juan) | | | 2.772,4 km. (Tigre) | | | 839,2 (Tigre) | | |

A continuación mostramos las diferencias entre los viajes de las divisiones mayores y las divisiones menores en el Torneo de Fútbol Juvenil de la Superliga

de 2017, el campeonato previo a los aquí estudiados, que fue programado por la SAF de manera manual.

Es preciso observar que la modalidad de juego era distinta: se enfrentaron todos los equipos contra todos en un single round robin, pero ya no divididos en dos zonas. Los clubes participantes eran 30. Las divisiones mayores jugaban con el fixture invertido de las menores (en términos de localía). Es decir, más allá de la estructura del torneo, sigue siendo un tópico interesante la reducción de las distancias de viaje realizadas por las divisiones mayores y las menores para todos los equipos.

El fixture utilizado en ese entonces fue el canónico, empleado en muchos torneos de fútbol en el mundo, que consiste en una plantilla de enfrentamientos entre distintos números en cada fecha, por lo que la SAF simplemente sorteaba qué equipo se asignaría a cada número. No se tenían en cuenta consideraciones de balance de distancias al momento de realizar este sorteo, ni ninguna de las restricciones adicionales incorporadas por la SAF a partir de la adopción de nuestros modelos. La Tabla 7 muestra la diferencia (en porcentaje) entre las distancias viajadas por las divisiones mayores y menores de cada equipo.

Tabla 7: Diferencia en porcentaje entre las distancias viajadas por las divisiones mayores y menores de cada equipo.

| | | | |
|--------------------|----------|--------------------|-----------------|
| River | 126,07 % | Belgrano | 44,26 % |
| Boca | 119,19 % | Estudiantes | 167,43 % |
| Vélez | 135,81 % | Gimnasia y Esgrima | 128,55 % |
| San Martín SJ | 10,95 % | Independiente | 113,81 % |
| Godoy Cruz | 1,92 % | Racing | 158,34 % |
| Banfield | 81,97 % | San Lorenzo | 201,16 % |
| Lanús | 27,25 % | Huracán | 45,66 % |
| Tigre | 157,81 % | Colón | 44,49 % |
| Defensa y Justicia | 175,39 % | Unión | 30,92 % |
| Arsenal | 34,22 % | Rosario Central | 2,47 % |
| Atlético Tucumán | 39,81 % | Newell's | 5,98 % |
| Patronato | 28,83 % | Aldosivi | 1,90 % |
| Temperley | 4,36 % | Atlético Rafaela | 36,66 % |
| Olimpo | 26,55 % | Quilmes | 147,61 % |
| Talleres | 29,97 % | Sarmiento | 15,67 % |

Como podemos observar, había grandes diferencias entre los viajes de las divisiones mayores y menores de los distintos clubes, teniendo diferencias que van desde 1,90 % (Aldosivi) hasta 201,16 % (San Lorenzo). En síntesis, había equipos para los cuales el kilometraje realizado por sus divisiones mayores y menores era prácticamente el mismo, y equipos donde la distancia recorrida

por unas de ellas triplicaba a la recorrida por las otras. En los fixtures generados por las técnicas presentadas en este trabajo, se obtienen para la gran mayoría de los equipos diferencias sensiblemente menores.

Como ya hemos comentado, como consecuencia de la implementación de nuestros modelos que buscan equiparar los viajes de las distintas divisiones de cada club, se consigue también un equilibrio de viajes entre los equipos de una misma región geográfica dentro de cada una de las divisiones. Es decir, aquellos clubes que están cerca entre ellos tienden a recorrer una cantidad de kilómetros similar a lo largo de todo el torneo, lo que constituye un logro en términos de la equidad deportiva de cada campeonato. La explicación es la siguiente: dados dos equipos con estadios geográficamente muy cercanos, la suma total de los kilómetros recorridos si ambos enfrentaran a todos sus rivales de visitante sería muy pareja. Como los modelos aquí implementados tienden a llevar lo que cada equipo viaja en sus divisiones mayores y menores a cifras cercanas al 50% de dicho total, entonces ambos equipos viajarán un kilometraje parejo dentro de la misma división.

Para medir esto, analicemos la media y el desvío estándar de los viajes realizados por las mayores y las menores de los equipos por región, en el último torneo programado de manera manual y en los torneos programados por nuestros modelos.

Tabla 8: Comparación de viajes (km.) entre torneos: región CABA y GBA

| Torneo | Mayores | | Menores | |
|------------------------------|----------|---------|----------|---------|
| | Promedio | Desvío | Promedio | Desvío |
| Torneo 2017: Manual | 9.903,3 | 3.578,5 | 9.022,7 | 3.557,4 |
| Transición 2018: Clusters | 4.152,1 | 697,4 | 4.107,9 | 590,4 |
| Transición 2018: Suma | 4.127,6 | 316,2 | 4.132,5 | 308,7 |
| Transición 2018: Máximo | 4.120,3 | 341,7 | 4.139,8 | 414,7 |
| Apertura 2018-2019: Clusters | 4.575,0 | 349,7 | 4.542,6 | 427,4 |
| Apertura 2018-2019: Suma | 4.429,5 | 585,0 | 4.688,2 | 586,3 |
| Apertura 2018-2019: Máximo | 4.545,9 | 283,5 | 4.571,8 | 330,8 |

En el torneo 2017, programado de manera manual, agrupando a los equipos participantes de la Ciudad de Buenos Aires y del Gran Buenos Aires, la media de kilómetros recorridos entre ida y vuelta por las divisiones mayores fue más del doble que en cualquiera de los otros torneos del 2018 independientemente del modelo utilizado (esto es esperable porque se jugó el doble de partidos). El desvío, a su vez, resulta ser más de cinco veces mayor, lo que nos indica que el kilometraje recorrido por los equipos de una misma región geográfica

Tabla 9: Comparación de viajes entre torneos: región Santa Fe

| Torneo | Mayores | | Menores | |
|------------------------------|----------|---------|----------|---------|
| | Promedio | Desvío | Promedio | Desvío |
| Torneo 2017: Manual | 14.744,1 | 2.796,9 | 13.255,1 | 1.788,2 |
| Transición 2018: Clusters | 6.100,1 | 1.078,8 | 6.029,7 | 736,4 |
| Transición 2018: Suma | 5.935,7 | 739,9 | 6.194,1 | 922,7 |
| Transición 2018: Máximo | 5.985,3 | 774,5 | 6.144,5 | 849,2 |
| Apertura 2018-2019: Clusters | 5.707,4 | 592,3 | 5.495,7 | 883,4 |
| Apertura 2018-2019: Suma | 5.799,4 | 509,6 | 5.403,7 | 727,4 |
| Apertura 2018-2019: Máximo | 5.460,9 | 757,0 | 5.742,2 | 496,2 |

Tabla 10: Comparación de viajes entre torneos: región Córdoba/Cuyo

| Torneo | Mayores | | Menores | |
|------------------------------|----------|---------|----------|---------|
| | Promedio | Desvío | Promedio | Desvío |
| Torneo 2017: Manual | 27.705,1 | 5.450,2 | 23.863,1 | 7.048,2 |
| Transición 2018: Clusters | 11.081,6 | 3.004,4 | 11.175,0 | 2.476,5 |
| Transición 2018: Suma | 11.290,1 | 2.903,0 | 10.966,6 | 2.503,8 |
| Transición 2018: Máximo | 11.011,6 | 2.673,8 | 11.245,1 | 2.720,3 |
| Apertura 2018-2019: Clusters | 10.508,1 | 3.020,1 | 9.968,6 | 2.002,1 |
| Apertura 2018-2019: Suma | 10.059,6 | 2.739,9 | 10.417,1 | 2.287,4 |
| Apertura 2018-2019: Máximo | 10.413,6 | 2.344,3 | 10.063,1 | 2.680,6 |

logró equipararse bastante a partir de la implementación de nuestros modelos. Para las divisiones menores, el promedio de viaje en 2017 es de alrededor del doble de cualquiera de los modelos de 2018 que construimos, mientras que el desvío es más de seis veces mayor.

Otro grupo geográfico que podemos distinguir en 2017 es el conformado por los equipos de Santa Fe (provincia ubicada al noroeste de la ciudad de Buenos Aires), para los cuales la media de distancia recorrida para las mayores en 2017 fue más del doble que en cualquier modelo y cualquier torneo de 2018. El desvío, mientras tanto, se ve triplicado en ese año. En lo que respecta a las divisiones menores, el promedio de kilómetros realizados en 2017 vuelve a doblar a los otros seis promedios analizados, mientras que el desvío es más que el doble en casi todos los casos. Como antes, esto nos habla del poco equilibrio entre viajes que había entre equipos de esta región cuando el fixture era realizado de modo manual y cómo se balanceó a partir de 2018.

Finalmente, considerando a los equipos de la región de Cuyo (San Juan

y Mendoza) y Córdoba, zonas en el centro-oeste del país, las diferencias son bien marcadas. En 2017, el kilometraje recorrido fue más que el doble que en cualquiera de los torneos y modelos del 2018. En lo que al desvío respecta, el del 2017 para las menores ronda el triple comparado con todos los modelos, mientras que para las mayores lo duplica en la mayoría de los casos.

Más allá de las diferencias en balances de viajes, existen otras mejoras derivadas de las técnicas aplicadas en este trabajo, como consecuencia de la incorporación de restricciones adicionales. Por ejemplo, en 2017 Defensa y Justicia jugó como local en la primera y en la segunda fecha, algo que fue restringido en los modelos presentados en este trabajo.

Otra de las condiciones incorporadas fue que no hubiera partidos entre equipos grandes ni en las dos primeras ni en las dos últimas fechas. Esto pudo ser cumplido por todos nuestros calendarios, mientras que en 2017 Racing y River se enfrentaron en la segunda fecha y Boca e Independiente en la penúltima.

Los modelos aquí presentados incluyen restricciones para tener viajes cortos en caso de tener breaks de visitante, situación que no podía ser considerada en el procedimiento manual. Por ejemplo, en 2017 las divisiones mayores de Colón (Santa Fe) jugaron de visitante en la tercera fecha contra Aldosivi (Mar del Plata, Bs. As.) y en la cuarta contra Boca (CABA), teniendo que realizar un total de 2730 km. en dos fines de semana consecutivos. A las divisiones menores de Patronato (Entre Ríos) les tocó jugar de visita contra San Martín de San Juan (San Juan) en la quinta fecha y contra Aldosivi en la sexta, contabilizando 3768 km. de viaje.

7. Conclusiones y trabajo futuro

En este trabajo se presentaron las técnicas empleadas para programar los campeonatos juveniles de los equipos de primera división de la Superliga Argentina de Fútbol durante 2018 y 2019, con el principal objetivo de balancear las distancias de viaje recorridas por las divisiones mayores y menores de cada uno de los clubes, además de conseguir un balance en viajes entre equipos de una misma zona geográfica dentro de cada campeonato y cumplir con una serie de restricciones adicionales impuestas por la SAF.

Si bien el esquema particular de estos torneos es poco convencional, las ideas que empleamos en este trabajo podrían ser extrapoladas a un torneo doble round robin (como se suele jugar la mayoría de los torneos de fútbol en el mundo) en el cual se deseen balancear los viajes de cada equipo en las

dos etapas del torneo. Esta situación es de interés dado que en esta clase de campeonatos todos los clubes se enfrentan dos veces entre ellos pero con las localías invertidas, permitiéndonos relacionar este esquema tradicional con las dos divisiones de los torneos aquí estudiados.

Con relación a posibles trabajos futuros, podríamos combinar ambas alternativas de resolución y construir modelos en los cuales minimicemos la suma de las diferencias de las cantidades de viajes a cada cluster de todos los equipos, teniendo en cuenta que la suma sobre todos los equipos de diferencias de distancias esté acotada, o que esté acotada la máxima de ellas; o incluso que ambas estén acotadas. De esta manera, contemplaríamos la reducción en diferencia de viajes desde las restricciones, tal como en el modelo de los clusters, pero usando explícitamente la matriz de distancias, lo que nuevamente liberaría la función objetivo. Fijar valores para acotar las diferencias puede ser un trabajo arduo, pero podríamos basarnos en los resultados ya obtenidos para encontrar algunos que se acomoden a lo esperado.

También parece una buena idea, a priori, dejar de armar los clusters en ambas zonas teniendo en cuenta los rivales clásicos de los partidos interzonales, permitiendo incluso que en la Zona A y en la Zona B los clusters no sean análogos, o que tengan diferentes cantidades de equipos. Esto obligaría a un cierto retoque en la modelación pero podría aportar muy buenos resultados.

Otra idea podría ser combinar modelos al diseñar las distintas zonas. Por ejemplo, utilizar el modelo de los clusters para diseñar la Zona A de algún torneo, pero utilizar alguno de los dos modelos que surgen a partir de la matriz de distancias para buscar un fixture para la Zona B. Más aún, podrían estudiarse muchas de las combinaciones posibles y elegir aquel fixture que resulte mejor, permitiendo incluso correr la Zona B y a posteriori la Zona A.

Los resultados de nuestros modelos fueron presentados a la Superliga Argentina de Fútbol para constituir los fixtures oficiales para el Torneo Transición de Juveniles 2018 (Marzo-Junio) y para el Torneo de Juveniles 2018-2019 (Agosto 2018-Junio 2019). Esta institución fue la que decidió cuál de ellos sería el utilizado finalmente en cada caso. Al momento de la escritura de este artículo hemos entregado una serie de propuestas para la Temporada 2019-2020, que se disputará con 24 equipos.

La satisfacción de la SAF con los resultados de este proyecto hizo ampliar los alcances del mismo. Según Mariano Elizondo, presidente de la entidad, "gracias al éxito del proyecto con las ligas juveniles, el último año hemos extendido la colaboración con el grupo de la Universidad de Buenos Aires (UBA) también a la programación de nuestra Primera División y a la asignación de horarios a los partidos en cada fecha. Estamos muy satisfechos con la colaboración iniciada hace casi 2 años con la UBA, y esperamos que la misma

se mantenga en el tiempo e incluso podamos ampliarla a otras áreas de nuestra organización.”

Agradecimientos: Este trabajo forma parte de un convenio de cooperación entre el Instituto de Cálculo de la Facultad de Ciencias Exactas y Naturales de la UBA y la SAF, para asesorar a esta última institución en las programaciones de los diferentes campeonatos que organiza. Agradecemos la permanente colaboración para concretar este trabajo de las autoridades de la SAF, en particular de Enrique Sacco y Juan Pablo Paterniti, responsables de la programación de los campeonatos. El primer autor está parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería, ISCI, Chile (CONICYT PIA/BASAL AFB180003), y los subsidios UBACyT 20020170100495BA (Argentina) y ANPCyT PICT 2015-2218 (Argentina).

Referencias

- [1] F. Alarcón, G. Durán, M. Guajardo, J. Miranda, H. Muñoz, L. Ramírez, M. Ramírez, D. Sauré, M. Siebert, S. Souyris, et al. Operations research transforms the scheduling of chilean soccer leagues and south american world cup qualifiers. *Interfaces*, 47(1):52–69, 2017.
- [2] 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.
- [3] W. Burrows y C. Tuffley. Maximising common fixtures in a round robin tournament with two divisions. *arXiv preprint arXiv:1502.06664*, 2015.
- [4] G. Cocchi, A. Galligari, F. P. Nicolino, V. Piccialli, F. Schoen, y M. Scianrone. Scheduling the italian national volleyball tournament. *Interfaces*, 48(3):271–284, 2018.
- [5] G. Durán, S. Durán, J. Marengo, F. Mascialino, y P. A. Rey. Scheduling argentina’s professional basketball leagues: A variation on the travelling tournament problem. *European Journal of Operational Research*, 275(3):1126–1138, 2019.
- [6] G. Durán, M. Guajardo, y D. Sauré. Scheduling the south american qualifiers to the 2018 fifa world cup by integer programming. *European Journal of Operational Research*, 262(3):1109–1115, 2017.

- [7] K. Easton, G. Nemhauser, y M. Trick. The traveling tournament problem: Description and benchmarks. In *Principles and Practice of Constraint Programming*, páginas 580–585. Springer, 2001.
- [8] M. J. Fry y J. W. Ohlmann. Introduction to the special issue on analytics in sports, part ii: Sports scheduling applications, 2012.
- [9] M. Grabau. Softball scheduling as easy as 1-2-3 (strikes you're out). *Interfaces*, 42(3):310–319, 2012.
- [10] S. Knust. Scheduling non-professional table-tennis leagues. *European Journal of Operational Research*, 200(2):358–367, 2010.
- [11] K. Nurmi, D. Goossens, y J. Kyngäs. Scheduling a triple round robin tournament with minitournaments for the finnish national youth ice hockey league. *Journal of the Operational Research Society*, 65(11):1770–1779, 2014.
- [12] J. Schönberger. The championship timetabling problem-construction and justification of test cases. In *Proceedings of MathSport International 2017 Conference*, página 330, 2017.
- [13] J. Schönberger, D. C. Mattfeld, y H. Kopfer. Memetic algorithm timetabling for non-commercial sport leagues. *European Journal of Operational Research*, 153(1):102–116, 2004.
- [14] T. A. Toffolo, J. Christiaens, F. C. Spieksma, y G. V. Berghe. The sport teams grouping problem. *Annals of Operations Research*, 275(1):223–243, 2019.
- [15] D. Van Bulck, D. Goossens, J. Schönberger, y M. Guajardo. Robinx: A three-field classification and unified data format for round-robin sports timetabling. *European Journal of Operational Research*, 2019.
- [16] D. Van Bulck, D. R. Goossens, y F. C. Spieksma. Scheduling a non-professional indoor football league: a tabu search based approach. *Annals of Operations Research*, 275(2):715–730, 2019.
- [17] M. Wright. Scheduling an amateur cricket league over a nine-year period. *Journal of the Operational Research Society*, 69(11):1854–1862, 2018.

PREDICCIÓN DE READMISIÓN DE PACIENTES PEDIÁTRICOS MEDIANTE APRENDIZAJE SUPERVISADO

PATRICIO WOLFF *
SEBASTIÁN A. RÍOS *

Resumen

La readmisión hospitalaria es un problema frecuente y costoso. El problema de predecir la readmisión pediátrica es difícil de abordar debido a que depende de múltiples factores y los datos requeridos pueden no estar disponibles. En esta investigación se abordó un problema que día a día adquiere mayor importancia, la readmisión pediátrica evitable. Para lograr el objetivo, se deben implementar una serie de operaciones, como: limpieza de datos erróneos, etiquetar casos, aplicar técnicas de balanceo de clases, probar diferentes modelos de clasificación y evaluar métricas para determinar mejores resultados. Diferentes métodos de clasificación fueron comparados en este estudio, dos métodos basados en árboles de decisión y uno en Regresión Logística (RL). Random Forest (RF) es la técnica que muestra mejores resultados con el Área bajo la curva ROC (AUC) de 0,683 y un Valor Predictivo Positivo (PPV) de 7,94. Para su evaluación, se utilizó 5-fold Cross-Validation y curvas ROC.

Palabras Clave: Readmisión Hospitalaria, Machine learning, Pediatría.

* Centro de Investigación en Inteligencia de Negocios, Departamento de Ingeniería Industrial, Universidad de Chile

1. Introducción

La readmisión hospitalaria se define como la admisión no programada de un paciente a un centro de atención después de un egreso o alta, dentro de un período de tiempo, que generalmente es de 7, 30 a 90 días. La readmisión hospitalaria es un problema frecuente y costoso [11]. Este problema ha sido ampliamente estudiado en el mundo [12]. No todas las readmisiones son evitables, sin embargo, aquellas que sí lo son están relacionadas con un cuidado subóptimo y con una mala gestión del proceso de dado de alta [21]. El problema del riesgo de readmisión en la actualidad ha tomado mayor relevancia en Chile. En especial cuando se discute la posibilidad de basar el financiamiento de los hospitales públicos en base a su producción de Grupos Relacionados por el Diagnóstico (GRD)¹. La experiencia de Estados Unidos es que el financiamiento hospitalario con esta característica debe ser corregido por un índice de readmisión [15] o como en el caso de Reino Unido, estableciendo penalidades financieras a los hospitales [13]. Debido a las nuevas estructuras de financiamiento de la salud en algunos países, las readmisiones hospitalarias se han convertido en un indicador central de la calidad de la atención [2]. Las readmisiones como una medida de calidad de la atención médica han sido objeto de un fuerte debate tanto en entornos hospitalarios para adultos como en pediatría [16].

El problema de la predicción de las readmisiones hospitalarias es difícil de abordar debido a que su origen se puede deber a múltiples factores y la data disponible puede ser insuficiente. La literatura científica reporta una gran cantidad de análisis cuantitativos, principalmente para pacientes adultos, incluidos enfoques predictivos basados en aprendizaje automático. Sin embargo, las readmisiones en hospitales pediátricos han recibido poca atención hasta hace poco. En el caso de la readmisión pediátrica la literatura reconoce que parece ser más difícil de resolver que en el caso de los adultos [10].

El Hospital Dr. Exequiel González Cortés (HEGC) es un centro de alta complejidad asistencial y docente que atiende a la población infantil del área sur de Santiago de Chile. En el periodo 2011 a 2017 el hospital contó con una dotación aproximada de 126 camas, pudiendo aumentar a 150 en campaña de invierno. En este periodo el hospital categorizó 56.558 egresos en su sistema

¹Los Grupos Relacionados por el Diagnóstico (GRD) son un sistemas de clasificación de pacientes, utilizados para medir la producción hospitalaria. Se obtienen a partir de la información que se registra en la historia clínica de cada paciente, y relacionan las características clínicas de los pacientes con el uso de recursos.

de codificación GRD. El uso de sistemas de estimación de la complejidad de las hospitalizaciones como GRD que actualmente ha tenido una importante penetración en Chile ha permitido contar con un gran número de datos estandarizados y específicos. Los datos contenidos en el GRD no poseen un campo específico que detalle si un determinado egreso produjo una readmisión en un periodo de 30 días a contar del alta. Por este motivo la etiqueta específica que permite el aprendizaje supervisado debe ser generada y revisada por un equipo experto, que entre otras cosas determina: si la segunda admisión es asociable clínicamente a la primera; si es una readmisión evitable; o si es una readmisión programada.

La predicción de readmisión en pacientes adultos se ha abordado con diversos enfoques [12, 1], entre estos por medio del aprendizaje supervisado [9, 26, 7, 19, 25, 22, 20]. A menudo, estos modelos exhiben un rendimiento predictivo deficiente y no serían adecuados para su uso en un entorno clínico [9]. Una revisión sistemática [12] de las publicaciones en esta área informa Área bajo la curva ROC (AUC) entre 0,56 y 0,72 (ver sección 2.5). Sin embargo, este resultado puede mejorar si se considera un mayor volumen de datos, tales como: actividades de asistencia diaria post-alta, discapacidad visual y estado funcional [6]; o data longitudinal [18].

La idea de esta investigación es extender los resultados de la investigación presentada en [23], probando otros métodos de aprendizaje supervisado bien estudiados en la literatura. En esa investigación se presentaron los resultados de la predicción de readmisión hospitalaria en pacientes pediátricos mediante el uso de *Support Vector Machines* (SVM), *Naive Bayes* (NB) y Redes Neuronales Artificiales (ANN, por su sigla en inglés). En dicho estudio se utilizaron técnicas de balanceo de clases y *5-fold Cross-validation*, encontrándose que el mejor AUC ($p < 0,001$) se obtuvo con el enfoque de NB (0,655). En esta investigación se busca mejorar los resultados mostrados en esa investigación por la incorporación de otros modelos no considerados en el estudio, pero conservando el diseño del estudio y aplicando la solución sobre la misma data. La utilización conjunta de las técnicas de balanceo de clases y *Cross-validation*, junto con métodos de aprendizaje supervisado como *Random Forest* han mostrado buenos resultados, en especial en problemas de riesgo clínico con data desbalanceada [24].

2. Materiales y Métodos

A continuación se describen los métodos utilizados los cuales se presentan en el esquema de la Figura 1.

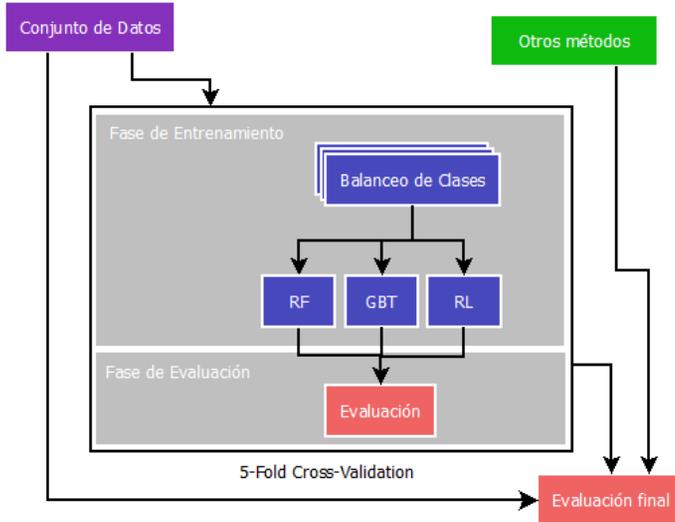


Figura 1: Diseño del estudio

2.1. El conjunto de datos

La data utilizada proviene del codificador GRD-IR v.3 del Ministerio de Salud, el que contiene 56.558 egresos de 35.063 pacientes del HEGC, de estos egresos 2.106 provocaron una readmisión. La información disponible corresponde a información del paciente como la fecha de nacimiento/edad, sexo, etnia, país de origen, domicilio, previsión y peso GRD al egreso, el cual contiene de forma intrínseca la complejidad del diagnóstico principal, secundario y procedimiento principal realizado. Además, esta data contiene información del ingreso del paciente al hospital como la fecha de ingreso, el servicio en el cual ingresó y su procedencia. Por otro lado, se cuenta con información de traslados internos como fecha/hora, servicio de origen y destino interno. Esta data también contiene información del alta del paciente como fecha/hora, servicio que da el alta y destino del paciente.

2.2. Etiquetas

No todo egreso hospitalario con reingreso posterior es considerado como un egreso que generó una readmisión (etiqueta positiva en el problema de aprendizaje supervisado) en este problema particular. Esto se debe a que existen casos que corresponden a reingresos programados, fuera del periodo de 30 días o no asociados al egreso anterior. Para obtener las etiquetas se analizaron: (1) readmisiones en periodos menores a 30 días desde el alta anterior; (2) se eliminaron tratamientos programados como quimioterapias; (3) se eliminaron ingresos al hospital por otros servicios que no son la urgencia o cuya estadía fue menor a un día; finalmente, (4) se revisaron las etiquetas por medio de un comité experto para corroborar que efectivamente correspondieran a un reingreso. La data y las etiquetas utilizadas en este estudio son exactamente las mismas que la utilizadas en el estudio anterior [23], por lo que los resultados obtenidos en estas dos publicaciones son comparables.

2.3. Métodos de Clasificación

En este estudio se entrenaron y probaron dos modelos basados en árboles de decisión (*Random Forest* y *Gradient Boosted Trees*) y uno basado en Regresión Logística.

Los árboles de decisión fueron introducidos en la comunidad científica por primera vez en en 1984 [4] y reconocidos en la literatura de Aprendizaje de Maquinas en 1986 [17]. *Random Forest* (RF) [3] son un tipo de métodos basados en el ensamblaje de un conjunto de árboles de decisión entrenados en base a subconjuntos aleatorios del conjunto de datos original, cada uno de los cuales se entrena de forma independiente. Cada nodo de un árbol representa una regla de división para un atributo específico donde la cantidad de variables de cada árbol es menor al numero total de variables del modelo completo, que se seleccionan de forma aleatoria. Luego, estos arboles se combinan con la finalidad de reducir la varianza a través de la combinación de los resultados de varios clasificadores débiles (Bagging).

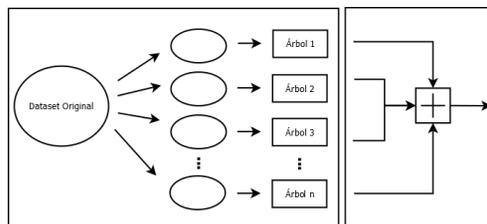


Figura 2: *Bagging* en RF

La profundidad máxima de los árboles utilizada en el caso de RF fue de 8 y se probaron bosques de 10, 100 y 1000 árboles, encontrándose un mejor resultado utilizando 100 árboles.

Los *Gradient Boosting Trees* (GBT) son considerados unos de los modelos más poderosos y mayormente utilizados para problemas de aprendizaje supervisado. Al igual que en el caso de los RF este modelo se basa en árboles de decisión débiles, que son agregados al modelo, con la finalidad de construir un modelo más preciso en cada iteración. Los parámetros de los árboles de decisión que son agregados al modelo se determinan mediante el cálculo del gradiente descendiente que minimiza la función de pérdida (que es predefinida). La principal diferencia con *Bagging* es que la secuencia de construcción de árboles de *Boosting* se adapta el predictor a distintos segmentos del conjunto de entrenamiento de forma secuencial.

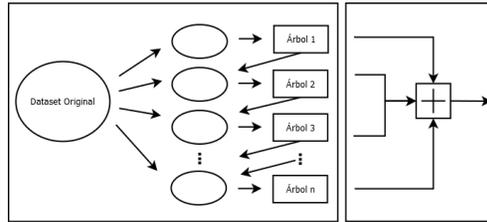


Figura 3: *Boosting*

La Regresión Logística es un método estadístico ampliamente utilizado en problemas de clasificación dicotómica. Este método pretende encontrar la probabilidad de ocurrencia $y \in [0, 1]$ de un evento en base a regresores $x \in \mathbb{R}^n$. El modelo utiliza la función logística (1) para estimar la probabilidad de ocurrencia de la readmisión, en base al vector de datos de entrada x_i con $i \in \{1, \dots, N\}$ con sus respectivas etiquetas $y_i \in \{1, 0\}$.

$$p(x_i) = \frac{1}{1 + e^{-\beta_0 - \sum_{j=1}^V \beta_j x_{ij}}} \quad (1)$$

El proceso de estimación de los parámetros utiliza la función de máxima verosimilitud [8], el que busca maximizar la probabilidad de obtener los resultados etiquetados en base a y_i . Esto permite obtener estimadores asintóticamente eficientes, insesgados y distribuidos normalmente.

2.4. Balanceo de clases

Una de las primeras dificultades al enfrentar problemas de este tipo es que el conjunto de datos están intrínsecamente muy desbalanceados. Esto se debe

principalmente a que los eventos de alto riesgo no son tan frecuentes. Esto genera conjuntos de datos con clases de bajo riesgo formadas por numerosos casos y clases de alto riesgo con muy pocos casos. En la actualidad existen una serie de técnicas que permiten en enfrentar el problema del desbalanceo de clases [14]. La literatura muestra una amplia utilización de estas técnicas en diferentes problemas. Específicamente, se utilizó *Synthetic Minority Over-sampling Technique* (SMOTE) [5] para aumentar las clases minoritarias. En el algoritmo SMOTE, las nuevas muestras de la clase minoritaria se generan por interpolación aleatoria entre k vecinos seleccionados al azar de la clase minoritaria. La interpolación aleatoria se realiza dentro de la envolvente convexa de las muestras de referencia, como se observa en la Figura 4.

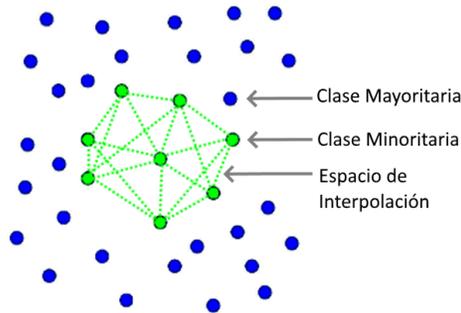


Figura 4: *Algoritmo SMOTE*

2.5. Métricas de Evaluación

En cada iteración de la validación cruzada se calculó la matriz de confusión y las métricas de rendimiento derivadas de él, finalmente reportando la media aritmética de estos resultados y su desviación estándar (SD). Definamos TP, TN, FP y FN como verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos respectivamente. Entonces calculamos el *recall* o sensibilidad como:

$$recall = \frac{TP}{TP + FN}, \quad (2)$$

el Valor Predictivo Positivo (PPV) como:

$$PPV = \frac{TP}{TP + FP}, \quad (3)$$

el $f - score$ se define como:

$$f - score = \frac{2}{1/R + 1/PPV} \quad (4)$$

Cuando las clases están fuertemente desbalanceadas estas métricas son más informativas de la detección exitosa de la clase minoritaria en comparación al Ajuste (Acc) definido como:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

El análisis utilizando las curvas *Receiver Operating Characteristic* (ROC) se han utilizado ampliamente para comparar diferentes clasificadores del estado del arte en clasificación binaria. La curva ROC es una representación gráfica de la sensibilidad frente a 1-especificidad, según se varía el umbral de discriminación. La curva se gráfica para diferentes resultados de modelos y permite comparar de forma gráfica la respuesta de estos. El Área Bajo la Curva ROC (AUC), o también conocido como estadístico C una prueba no paramétrica aplicada a dos muestras independientes. El AUC representa numéricamente el área entre la curva ROC y la horizontal y es equivalente a la prueba de Mann-Whitney. El AUC toma valores entre 0 y 1 y a medida que el AUC de un clasificador se acerca al valor 1 (clasificador perfecto), mayor será su capacidad discriminativa. Además, se debe notar que valores menores o iguales a 0,5 corresponden a clasificadores con nula o mala capacidad discriminativa. El AUC se reporta a menudo para medir la calidad del resultado en estudios sobre la predicción de readmisión en pacientes adultos [1].

Cross-Validation es una técnica ampliamente utilizada para evitar el sobreajuste de los modelos y garantizar que los resultados son independientes de una partición arbitraria entre datos de entrenamiento y prueba. La técnica consiste en repetir y calcular la media aritmética de los resultados obtenidos por el modelo en diferentes particiones de los datos, en este caso cinco particiones (5-fold). Los valores obtenidos en este conjunto de pruebas son los que se utilizan para la comparación entre los diferentes modelos.

3. Resultados

En la Tabla 1 se presentan los resultados obtenidos con los 3 modelos utilizados en este trabajo e incluyendo el mejor de los resultados mostrado en el trabajo anterior.

Tabla 1: Resultados

| | AUC (SD) | recall (SD) [%] | f-score (SD) [%] | PPV (SD) [%] |
|-------------------------------|---------------|-----------------|------------------|--------------|
| Trabajo Anterior [23] | 0,653 (0,014) | 69,80 (4,97) | 9,83 (0,53) | 5,29 (0,31) |
| <i>Random Forest</i> | 0,683 (0,009) | 24,07 (1,67) | 11,94 (0,90) | 7,94 (0,62) |
| <i>Gradient Boosted Trees</i> | 0,682 (0,013) | 57,65 (5,52) | 11,56 (0,60) | 6,43 (0,31) |
| Regresión Logística | 0,668 (0,010) | 68,90 (1,65) | 10,27 (0,17) | 5,55 (0,09) |

En el Figura 5 se presenta la curva ROC obtenida por los 3 modelos presentados en este trabajo, en conjunto con el mejor resultado obtenido en el trabajo anterior, el cual corresponde a *Naive Bayes*.

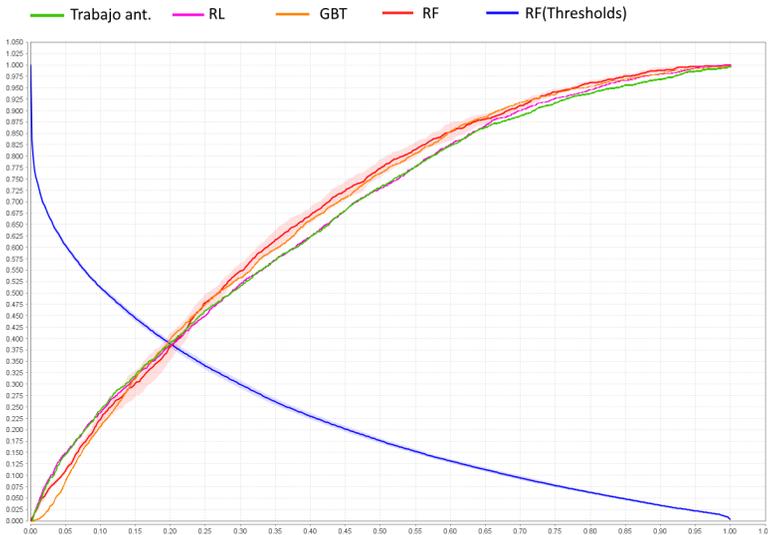


Figura 5: Curva ROC

4. Discusión

Los valores de AUC obtenidos en este trabajo para RF muestran un comportamiento de clasificación superior (*Pairwise t-test* $p < 0,004$) al resultado obtenido en el trabajo anterior. En el caso de GBT el resultado en AUC es superior, pero menos significativo (*Pairwise t-test* $p < 0,008$) debido a la variabilidad del resultado expresado en una desviación estándar mayor que en el caso de RF. Tanto RF como GBT muestran mejores resultados ($p < 0,025$ y $p < 0,07$) en AUC respecto de LR. Una métrica que es presentada en [23] y que resulta interesante de analizar en este caso es la sensibilidad (o *recall*), esta métrica muestra que el resultado del trabajo anterior es superior ($p < 0,005$)

a los resultados obtenidos con los modelos basados en árboles de decisión. Por otro lado, el resultado obtenido en PPV muestra que el modelo RF es significativamente superior al resultado mostrado por NB ($p < 0,0002$) y GBT ($p < 0,003$).

La curva ROC muestra gráficamente un resultado levemente superior de los modelos basados en árboles de decisión con respecto a la curva ROC anteriormente obtenida para la predicción de readmisiones. Al presentarse en este trabajo curvas ROC por sobre la mejor curva ROC presentada anteriormente, intrínsecamente se observa que los resultados basados en árboles de decisión, en este problema particular muestran mejor comportamiento que SVM y ANN.

5. Conclusiones

Por lo que sabemos, este es el segundo estudio de este tipo en Chile, y entre los primeros a nivel mundial, centrados en la predicción readmisiones pediátricas. Además, los resultados obtenidos con los modelos utilizados en este trabajo sobrepasan a los resultados obtenidos en el único trabajo anterior de predicción de readmisión hospitalaria pediátrica en 30 días conocido por los autores (utilizando la misma data).

En este documento, se presentan los resultados de un estudio que utilizó un conjunto de datos anonimizados recopilados durante seis años de operación en un hospital pediátrico público en Santiago, Chile. La cantidad de datos recopilados es grande para un estudio de un solo sitio (56.558 altas y 2.106 readmisiones). Hemos aplicado adicionalmente tres métodos predictivos sobre los datos administrativos utilizados para la estimación del costo del paciente por medio del codificador GRD. Los resultados son buenos, logrando un alto rendimiento predictivo ($AUC = 0,683$) que es comparable a otros estudios sobre datos de pacientes adultos.

La aplicación de estos modelos en ambientes clínicos reales puede variar. La predicción de las readmisiones tiene como objetivo central evitar que estas ocurran, por medio del soporte a la decisión clínica. En el proceso de alta de un paciente se cuenta con toda la información que es requerida para la predicción de estos modelos. Es en ese instante cuando la predicción puede apoyar al personal del hospital, recomendando por ejemplo: cuidados especiales post alta, educación al cuidador o postergación del alta.

Existen diferentes estrategias para evitar una readmisión no planificada como Telemonitoreo y visitas programadas a los pacientes, además de estrategias como educación para los tutores respecto de los cuidados del paciente, etc.

Si bien, estas estrategias han mostrado ser efectivas tienen un costo elevado para el hospital. En este sentido, se requieren modelos que permitan predecir los pacientes que tienen un mayor riesgo de ser readmitidos y focalizar los recursos en estos pacientes. Mejorar la capacidad predictiva permitirá centrar los recursos en los pacientes de mayor riesgo, lo que mejorará el resultado sanitario, sin incrementar significativamente los costos.

La readmisión hospitalaria evitable es un evento específico objetivo, que revela un resultado clínico no deseado. Esta tiene la característica de ser un punto de intersección de dos áreas de la investigación, como la medicina basada en evidencia y la ciencia de los datos. En este sentido los trabajos que basan en resultados clínicos permiten avanzar en técnicas de aprendizaje supervisado, así como discutir el potencial clínico de las aplicaciones realizadas. La etiqueta experta limita el número de casos en estudio, principalmente porque el proceso de etiquetado puede tomar mucho tiempo para su realización y/o estar sujeto a errores. Se requiere avanzar en determinar un mayor número de resultados clínicos con estas características y el estudio de modelos de aprendizaje supervisado que robustezcan los Sistemas de Apoyo a las Decisiones Clínicas (CDSS, por su sigla en inglés).

Agradecimientos: Los autores desean agradecer a la Dra. María Begoña Yarza y al Profesor PhD. Manuel Graña por su cooperación en esta investigación. Esta investigación es parcialmente financiada por CONICYT, Programa de Formación de Capital Humano avanzado (CONICYT-PCHA/Doctorado Nacional/2015-21150115).

Referencias

- [1] A. Artetxe, A. Beristain, y M. Graña. Predictive models for hospital readmission risk: A systematic review of methods. *Computer Methods and Programs in Biomedicine*, 164:49 – 64, 2018.
- [2] N. S. Bardach, E. Vittinghoff, R. Asteria-Peñaloza, J. D. Edwards, J. Yazdany, H. C. Lee, W. J. Boscardin, M. D. Cabana, y R. A. Dudley. Measuring hospital quality using pediatric readmission and revisit rates. *Pediatrics*, 132(3):429–436, 2013.
- [3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
- [4] L. Breiman, J. H. Friedman, R. A. Olshen, y C. J. Stone. Classification and regression trees. *Cole Publishing, Monterey*, 535:358, 1984.

- [5] N. V. Chawla, K. W. Bowyer, L. O. Hall, y W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.
- [6] E. A. Coleman, S.-j. Min, A. Chomiak, y A. M. Kramer. Posthospital care transitions: patterns, complications, and risk identification. *Health Serv Res*, 39(5):1449–1465, Oct 2004. 15333117[pmid].
- [7] S. Cui, D. Wang, Y. Wang, P.-W. Yu, y Y. Jin. An improved support vector machine-based diabetic readmission prediction. *Computer Methods and Programs in Biomedicine*, 166:123 – 135, 2018.
- [8] S. R. Eliason. Maximum likelihood estimation. 1993.
- [9] J. Futoma, J. Morris, y J. Lucas. A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, 56:229–238, 2015.
- [10] A. Garmendia, M. Graña, J. M. Lopez-Guede, y S. Rios. Predicting patient hospitalization after emergency readmission. *Cybernetics and Systems*, 48(3):182–192, 2017.
- [11] S. F. Jencks, M. V. Williams, y E. A. Coleman. Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428, 2009. PMID: 19339721.
- [12] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, y S. Kripalani. Risk prediction models for hospital readmission. *JAMA*, 306(15):1688, Oct 2011.
- [13] Z. Kmietowicz. Hospitals will be fined for emergency readmissions, says lansley. *BMJ*, 340, 2010.
- [14] V. López, A. Fernández, S. García, V. Palade, y F. Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, 2013.
- [15] C. K. McIlvennan, Z. J. Eapen, y L. A. Allen. Hospital readmissions reduction program. *Circulation*, 131(20):1796–1803, May 2015.
- [16] M. M. Nakamura, S. L. Toomey, A. M. Zaslavsky, J. G. Berry, S. A. Lorch, A. K. Jha, M. C. Bryant, A. T. Geanacopoulos, S. S. Loren, D. Pain, y M. A. Schuster. Measuring pediatric hospital readmission rates to drive quality improvement. *Academic Pediatrics*, 14(5, Supplement):S39 – S46,

2014. Advances in Children's Healthcare Quality: The Pediatric Quality Measures Program.
- [17] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.
- [18] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G. E. Duggan, J. Irvine, Q. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenbom, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. D. Howell, C. Cui, G. S. Corrado, y J. Dean. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1(1):18, 2018.
- [19] B. K. Reddy y D. Delen. Predicting hospital readmission for lupus patients: An rnn-lstm-based deep-learning methodology. *Computers in Biology and Medicine*, 101:199 – 209, 2018.
- [20] K. Shameer, K. W. Johnson, A. Yahi, R. Miotto, L. Li, D. Ricks, J. Jebarakaran, P. Kovatch, P. P. Sengupta, S. Gelijns, A. Moskovitz, B. Darrow, D. L. David, A. Kasarkis, N. P. Tatonetti, S. Pinney, y J. T. Dudley. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using mount sinai heart failure cohort. In *Pacific Symposium on Biocomputing 2017*, páginas 276–287, 2017.
- [21] M. J. Swain y H. Kharrazi. Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data. *International Journal of Medical Informatics*, 84(12):1048 – 1056, 2015.
- [22] M. Vukicevic, S. Radovanovic, A. Kovacevic, G. Stiglic, y Z. Obradovic. Improving hospital readmission prediction using domain knowledge based virtual examples. In L. Uden, M. Heričko, y I.-H. Ting, editors, *Knowledge Management in Organizations*, páginas 695–706, Cham, 2015. Springer International Publishing.
- [23] P. Wolff, M. Graña, S. A. Ríos, y M. B. Yarza. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. *BioMed Research International*, 2019:9, 2019.
- [24] P. Wolff, S. A. Ríos, y M. Graña. Setting up standards: A methodological proposal for pediatric triage machine learning model construction based on clinical outcomes. *Expert Systems with Applications*, 138:112788, 2019.

- [25] C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, y F. Wang. Readmission prediction via deep contextual embedding of clinical concepts. *PLOS ONE*, 13(4):1–15, 04 2018.
- [26] B. Zheng, J. Zhang, S. W. Yoon, S. S. Lam, M. Khasawneh, y S. Poranki. Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications*, 42(20):7110 – 7120, 2015.

IMPLEMENTACIÓN DE UNA HEURÍSTICA PARA LA PROGRAMACIÓN AUTOMÁTICA DE HORARIOS DE UNA ESCUELA SECUNDARIA

GUILLERMO A. DURÁN * **
NAZARENO FAILLACE **

Resumen

Desde 2009 en Argentina se está llevando a cabo el proyecto Nueva Escuela Secundaria (NES) con el objetivo de reestructurar el nivel medio de la educación pública modificando los planes de estudio. Esto supone, entre otros, el desafío de elaborar horarios escolares que incluyan nuevas asignaturas, docentes y proyectos así como también un uso eficiente de los recursos edilicios de la escuela. A diferencia de la universidad, en la escuela secundaria los alumnos son agrupados según el año que estén cursando y no tienen libertad de elegir a qué asignaturas asistir durante el ciclo lectivo, dado que están preestablecidas por la currícula. El problema consiste en designar una asignatura a cada hora cátedra de cada curso, de manera tal que se cumplan los requerimientos del sistema educativo y de la disponibilidad docente. En la actualidad, los horarios escolares en las escuelas secundarias son elaborados manualmente, dando lugar a un amplio margen de mejora. Para desarrollar una herramienta que permita automatizar este procedimiento garantizando la elaboración de horarios de buena calidad, hemos implementado una heurística que divide el problema en etapas que combinan Programación Lineal Entera (PLE) y un algoritmo de búsqueda local (*stochastic hill-climbing*). El objetivo es aumentar el cumplimiento de características deseables del horario, como la compactación de horas de una misma asignatura o la reducción de horas inactivas de los docentes, respetando las condiciones que garantizan la factibilidad.

* Departamento de Matemática, FCEyN-UBA, Argentina, Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Santiago, Chile.

** Instituto de Cálculo FCEyN-UBA y CONICET, Argentina

Al aplicar la herramienta desarrollada a los horarios de una escuela secundaria pública de la Ciudad de Buenos Aires, según un sistema de penalizaciones elaborado para medir cualitativamente un horario, el resultado de la automatización mostró una mejora del 55,82 % y del 73,28 % con respecto a los horarios confeccionados manualmente en 2017 y 2018, respectivamente. El incremento de la calidad se vio reflejado, entre otros, en una mejor distribución de los recursos edilicios, en una disposición temporal equilibrada de las asignaturas y en horarios de trabajo más favorables para los docentes.

Palabras Clave: Programación lineal entera, Horarios escolares, Scheduling Problems

1. Introducción

El problema de la planificación de horarios de universidades y escuelas ha sido estudiado en detalle y la cantidad de trabajos publicados sobre ese tema ha ido aumentando a lo largo de los años (véase, por ejemplo, [2, 5, 4, 6, 8, 9]). La importancia de la programación de un buen horario radica en el impacto que tiene sobre la organización laboral de los docentes y el uso eficiente de recursos pedagógicos para la enseñanza que reciben los estudiantes, como, por ejemplo, salas de informática, laboratorios de ciencias o clases de apoyo escolar.

En particular, en este trabajo estudiamos una subclase de este tipo de problemas: la elaboración de horarios para escuelas secundarias (conocido en la literatura como *High School Scheduling Problem*, HSSP). La naturaleza del HSSP puede ser resumida de la siguiente manera: las clases de distintos conjuntos de alumnos deben ser programadas durante el horario laboral de los docentes teniendo en cuenta su disponibilidad y su especialización. El HSSP puede ser considerado como un problema menos restrictivo que el de programación de horarios de una universidad, puesto que en el caso de las escuelas secundarias, por ejemplo, no es necesario minimizar la superposición de clases que involucran a los mismos estudiantes.

En la confección del horario escolar se tienen en cuenta dos tipos de restricciones. Por un lado, las restricciones duras son aquellas que un horario debe cumplir para considerarse factible y poder ser aplicado. Entre ellas se encuentran las relacionadas a la disponibilidad horaria de los docentes, a la cantidad de horas cátedra semanales de las asignaturas y que los alumnos no tengan horas cátedra libres. Por otro lado, las restricciones blandas son aquellas relacionadas a peticiones de alumnos, docentes y directivos, que deben ser

satisfechas tanto como sea posible pero que no interfieren con la factibilidad del horario. Por ejemplo, es deseable que no hayan tres horas cátedra consecutivas de la misma asignatura o que los docentes no tengan horas cátedra inactivas en sus días laborales. En general, debido a la complejidad del mundo real, es muy difícil encontrar una solución que no viole ninguna de las restricciones blandas. Por esta razón, se utiliza una función de penalidad para evaluar cuán buena es una solución. Diremos que un horario factible es de *buen calidad* si minimiza, en la medida de lo posible, la violación de las restricciones blandas.

Dado que los condicionamientos propios de la vida real hacen que el problema pertenezca a la clase \mathcal{NP} -completo [11], a lo largo del tiempo se ha encarado desde numerosos enfoques. Se han utilizado técnicas basadas en *constraint programming* [3, 6, 8] y meta-heurísticas como búsqueda tabú [7], recocido simulado [1] y algoritmos evolutivos [4, 9], entre otros. Por otra parte, las aplicaciones prácticas de soluciones al HSSP están íntimamente relacionadas al sistema educativo (y por ende al país) donde se llevan a cabo. Actualmente, en la gran mayoría de las escuelas secundarias argentinas, los horarios son armados manualmente. Los directivos o docentes encargados de tal tarea concuerdan casi unánimemente que es compleja, tediosa, consume gran cantidad de tiempo y en ciertos casos no es transparente, por lo que ven con entusiasmo el desarrollo de una herramienta matemático-computacional que permita automatizarla. Por esta razón, decidimos diseñar una heurística que emplea modelos de Programación Lineal Entera (PLE), un algoritmo de búsqueda local y la búsqueda en entornos variados, de manera tal que pueda emplearse para elaborar horarios escolares en el marco del sistema educativo argentino y que sea fácilmente adaptable a países con sistemas educativos similares.

Como se verá en la siguiente sección, el horario escolar del sistema educativo argentino se puede dividir en dos partes: horario de clases y horario extraclase. Debido a que su conjunto de restricciones y complejidad difieren notablemente, optamos por dividir la resolución del problema en dos etapas, una para cada horario. De esta manera, el armado del horario escolar se lleva a cabo de manera incremental: horarios de clase factibles son obtenidos a partir de un modelo de PLE, luego son mejorados a través de un algoritmo de búsqueda local y finalmente se elige al mejor de ellos para completarlo con la asignación de las horas extraclase mediante dos modelos de PLE.

El artículo se estructura de la siguiente manera. En la sección 2 se describe el problema que tratamos, dando algunos detalles del sistema educativo. En la sección 3 se describe el procedimiento de la heurística y las etapas en las que se divide el problema. En la sección 4 se comparan los resultados obtenidos con los horarios elaborados manualmente para el 2017 y 2018. Finalmente, en la sección 5 se exponen las conclusiones y el trabajo a futuro.

2. Descripción del problema

Las escuelas secundarias públicas de la Ciudad de Buenos Aires generalmente cuentan con dos turnos: mañana y tarde. El plan de estudios se divide en cinco años y en cada uno de esos años los estudiantes son distribuidos en dos o tres divisiones, con el fin de reducir la cantidad de estudiantes por aula. De aquí en adelante denominaremos *curso* a cada uno de esos conjuntos de alumnos. Por ejemplo, si se tienen dos divisiones por año, se tendrá un total de diez cursos. Cada curso tiene asignada un aula y en cada hora cátedra de la jornada escolar tiene designada una materia. En el sistema educativo argentino los estudiantes no pueden elegir a qué asignaturas asistir, por lo que permanecen en la misma aula todo el día y los profesores son los que rotan. Las aulas son destinadas a los mismos cursos año a año, por lo tanto su distribución no es parte del problema.

Generalmente cada curso tiene un total de 35 horas cátedra de clase dispuestas de Lunes a Viernes. Están distribuidas homogéneamente para el turno mañana, y, para el turno tarde, con un mínimo de 6 y un máximo de 8 horas cátedra diarias. Cada hora cátedra tiene una duración de 40 minutos. Diremos que un curso del turno tarde tiene prehora un determinado día si ingresa durante el horario correspondiente a la séptima hora del turno mañana. Es decir, la séptima hora del turno mañana se superpone con la prehora del turno tarde. Por esta razón, al momento de asignar qué curso tiene prehora un determinado día, no sólo es importante tener en cuenta la cantidad de aulas libres con las que cuenta la escuela, sino también con la disponibilidad de los docentes que enseñan en ambos turnos. Por otro lado, los estudiantes no pueden tener horas cátedra inactivas entre dos horas de clase. En general, las clases del turno mañana se desarrollan de 7 : 40 a 12 : 40 hs. y las clases del turno tarde de 12 : 20 a 18 hs., pues algunos cursos tienen prehora. Cada turno cuenta con dos recesos de 10 minutos.

En ciertas ocasiones, puede ocurrir que los estudiantes de un curso puedan elegir qué asignatura cursar en una determinada franja horaria (por ejemplo, Inglés o Francés). Incluso puede suceder que los alumnos que elijan una asignatura sean agrupados con alumnos de otro curso que hayan elegido la misma durante dichas franjas horarias. Por ejemplo, si los dos cursos de segundo año pueden elegir entre Francés e Inglés, se agruparán a aquellos alumnos que hayan elegido el mismo idioma independientemente de su división. De aquí en adelante nos referiremos a esta situación como una simultaneidad, dado que

se debe dictar esa asignatura al mismo tiempo en los cursos implicados. En nuestro modelo, para cada simultaneidad consideramos que las asignaturas a elegir son una sola (en el ejemplo, Inglés-Francés) que debe ser programada en las mismas horas cátedra para los cursos correspondientes.

Por otro lado, existe una serie de preferencias en lo referido a la distribución horaria de las materias, que serán incluidas al modelo como restricciones blandas y cuya violación será sancionada por la función de penalidad en la fase de búsqueda local. En general, es deseable que las asignaturas no ocupen más de tres horas cátedra diarias. Más aún, se espera que su carga horaria semanal sea distribuida en bloques de dos horas adyacentes. Por ejemplo, si Matemática en primer año tiene una carga horaria semanal de 5 horas cátedra, la distribución ideal consiste en dos bloques en dos días y una hora cátedra en otro día. Además, es preferible que no hayan bloques de tres horas cátedra consecutivas (que llamaremos *triples*) de la misma asignatura. Vale la pena mencionar que no pueden haber más de tres horas cátedra de una asignatura un mismo día, lo cual será tenido en cuenta al momento de elaborar horarios de clase factibles.

Con respecto a los docentes, se debe tener en cuenta que no todos tienen disponibilidad completa, pues tienen cargos en distintas instituciones. Por esta razón, antes de que comience el armado de horarios del siguiente ciclo lectivo, cada docente debe presentar a las autoridades del colegio una declaración jurada con su disponibilidad horaria. Además, existe la restricción de que un docente no puede estar enseñando en dos cursos al mismo tiempo así como también los cursos no pueden estar transitando dos asignaturas a la vez (siendo las simultaneidades excepciones de ambas restricciones).

Otro factor a tener en cuenta es la situación actual de cambio en las escuelas secundarias en Argentina y, en particular, en la Ciudad de Buenos Aires llevado a cabo por el proyecto Nueva Escuela Secundaria (NES). Una de las modificaciones que implica el proyecto es dividir los cinco años de la secundaria en dos ciclos: un ciclo de formación general común a las escuelas de todas las orientaciones (1^{ero} y 2^{do} año) y un ciclo de formación orientada (3^{ero} a 5^{to} año). Al mismo tiempo, la propuesta aspira a implementar nuevos métodos como, por ejemplo, las parejas pedagógicas, que consiste en que dos profesores den clase de manera conjunta, articulando los conocimientos de ambas asignaturas. Naturalmente, la puesta en marcha de todos estos cambios presentan nuevos desafíos en la confección de los horarios de las escuelas secundarias, como conjugar la disponibilidad simultánea de distintos docentes, distribuir la carga semanal de una nueva asignatura de manera tal que sea atractiva para los docentes que quieran tomar el cargo y el uso de recursos esenciales para ciertas orientaciones (por ejemplo, Informática).

Por último, cada turno tiene actividades en contraturno, sean clases de Educación Física, apoyo escolar o proyectos extracurriculares. Las dos últimas pertenecen a lo denominado *éxtraclase* y la asistencia a ellas es opcional y no existen divisiones de cursos. En lo que respecta a Educación Física, su horario es confeccionado por sus profesores, y se adapta al horario de clases; razón por la cual no contemplamos su elaboración.

De esta forma, el horario escolar se divide en dos partes. Por un lado, el horario de clases, compuesto por las horas cátedra a los que los alumnos, divididos en cursos, deben asistir obligatoriamente. Por el otro, el horario *extraclase*, conformado por las clases de apoyo y los proyectos, en los cuales no existe distinción por curso y cuya asistencia es opcional. Debido a que la elaboración del horario *extraclase* resulta un problema de tamaño mucho menor (menos materias y docentes, no hay división por cursos) y que tiene menor jerarquía que el horario de clases, decidimos seguir la metodología que actualmente llevan a cabo las autoridades escolares. Esto consiste en confeccionar primero el horario de clases y luego adaptar el de *extraclase*, de manera tal que la cursada obligatoria pueda estructurarse sin tener en cuenta las actividades opcionales.

3. Resolución del problema

Diseñamos la heurística teniendo en cuenta la prioridad de cada una de las partes que conforman al horario escolar. En primer lugar, el horario de clases es el que tiene mayor jerarquía. Por esta razón, el objetivo de la Etapa I es encontrar un horario de clases factible que viole tan pocas restricciones blandas como sea posible. Para lograrlo, utilizamos un modelo de PLE para generar horarios de clase factibles. A continuación, empleamos cada uno de ellos como punto de partida para el algoritmo de búsqueda local, con el fin de hallar horarios de clase de mejor calidad. El objetivo del algoritmo de búsqueda local es explorar el espacio de soluciones candidatas hasta encontrar un óptimo o hasta alcanzar un límite de tiempo, utilizando muy poca memoria [10]. Otra ventaja es que permite hallar soluciones razonables en un espacio de búsqueda muy extenso, donde los algoritmos exactos resultan inapropiados. Para finalizar con la Etapa I, nos quedamos con el horario de clases que presente el menor valor de la función de penalidad.

Posteriormente, para completar el horario escolar, llevamos a cabo la Etapa II: confeccionar el horario *extraclase* adaptándolo al resultado obtenido de la Etapa I. Dado que las clases de apoyo repercuten en el desempeño de los

estudiantes y los proyectos son actividades complementarias a la enseñanza, las primeras tienen mayor prioridad y presentan más restricciones. Por lo tanto, en un primer momento se asignan las clases de apoyo mediante un modelo de PLE. Por último, se asignan las horas de proyecto con otro modelo de PLE para concluir con la elaboración del horario escolar.

Al dividir el problema en etapas, no existe garantía de que la solución obtenida sea la óptima. Sin embargo, dado que se trata de un problema \mathcal{NP} -completo, nuestro objetivo es hallar una buena solución en un término de tiempo aceptable. Los modelos de PLE han sido implementados con la API de CPLEX 12.8.1 para Python en una computadora con procesador Intel Core i5 7400 a 3.5 GHz y con 8-GB RAM.

3.1. Modelo de PLE para creación de horarios iniciales

El objetivo de este modelo es generar horarios de clase que cumplan con la mayoría de las restricciones que aseguran su factibilidad. La experiencia proporcionada por la investigación en el campo del HSSP indica que, en general, lo más eficiente es tener la menor cantidad de restricciones duras posible, debido a que obstaculizan a los algoritmos de optimización local. Al partir desde un horario factible, el vecindario de soluciones válidas es más amplio con menos restricciones duras [12]. Esto significa que dentro del modelo habrán restricciones blandas cuya violación no permitirá que un horario sea aplicable en la práctica. Este tipo particular de restricciones reciben el nombre de *restricciones duras suavizadas*. Por ende, el modelo de PLE no contempla este tipo de restricciones y, posteriormente, en la definición de la función de penalidad les otorgamos un valor de penalidad por ser violadas muy superior al del resto de las restricciones blandas.

En nuestro caso, el conjunto de restricciones duras suavizadas son las relacionadas a la disponibilidad de los docentes. En primer lugar, observamos que suavizar dichas restricciones permite que el algoritmo de búsqueda local tenga más libertad al momento de recorrer el vecindario de una determinada solución. Además, permite que el modelo de PLE sea resuelto en menor cantidad de tiempo. Sin embargo, consideraremos como restricción dura la disponibilidad de los docentes en el caso de las simultaneidades. Por lo tanto, los horarios obtenidos con el modelo ubican a las asignaturas con simultaneidades en horarios en los que los docentes que deben dictarlas están disponibles. Además el modelo maximizará la cantidad de bloques de dos horas que formen dichas asignaturas.

A continuación introducimos algunos elementos de la notación utilizada en el modelo de PLE, resumida en la Tabla 1.

Tabla 1: Descripción de los conjuntos utilizados en el modelo de PLE de la Etapa I y sus características en el caso de estudio

| Notación | Descripción | Característica en el caso de estudio |
|-------------|---|---|
| M | Conjunto de las asignaturas | $M = \{0, \dots, 29\}$ |
| Δ | Conjunto de días de la semana | $\Delta = \{0, \dots, 4\}$ |
| H | Conjunto de horas cátedra | $H = \{0, \dots, 79\}$ |
| H_{TM} | Conjunto de horas cátedra del turno mañana | $H_{TM} = \{h \in H : \left\lfloor \frac{h}{8} \right\rfloor \equiv 0 \pmod{2}\}$ |
| H_{TT} | Conjunto de horas cátedra del turno tarde | $H_{TT} = \{h \in H : \left\lfloor \frac{h}{8} \right\rfloor \equiv 1 \pmod{2}\}$ |
| H_d | Conjunto de horas cátedra correspondientes al día $d \in \Delta$ | $H_d = \{h \in H : h \equiv d \pmod{16}\}$ |
| P | Conjunto de profesores | $P = \{0, \dots, 49\}$ |
| C | Conjunto de cursos | $C = \{0, \dots, 21\}$ |
| C_{TM} | Conjunto de cursos del turno mañana | $C_{TM} = \{0, \dots, 11\}$ |
| C_{TT} | Conjunto de cursos del turno tarde | $C_{TT} = \{12, \dots, 21\}$ |
| $cgs_{i,c}$ | Constante que representa la carga horaria semanal de horas cátedra de la asignatura i en el curso c . Si la asignatura i no se dicta en el curso c , entonces $cgs_{i,c} = 0$. | |
| K | Conjunto conflictivo | |
| CK_{TM} | Conjunto de conjuntos conflictivos del turno mañana | $ CK_{TM} = 40$ |
| CK_{TT} | Conjunto de conjuntos conflictivos del turno tarde | $ CK_{TT} = 42$ |
| CK_{PH} | Conjunto de conflictos por prehora | $ CK_{PH} = 379$ |
| S | Conjunto de simultaneidades | $ S = 4$ |

En primer lugar, se considera $H = \{0, \dots, 79\}$ al conjunto de horas cátedra. Consideramos que hay 16 horas cátedra diarias (8 para cada turno). Así, por ejemplo, la hora 0 representa la prehora del Lunes para el turno mañana y la hora 79 representa la última hora cátedra del Viernes para el turno tarde. También se define C el conjunto de cursos de la escuela. En el caso de estudio, $C = \{0, \dots, 21\}$.

Llamamos M al conjunto de asignaturas que se dictan en la escuela. En el caso de estudio, $M = \{0, \dots, 29\}$. Como se debe tener en cuenta que un docente no puede estar en dos cursos al mismo tiempo (excepto en el caso de las simultaneidades), se denominará como *conjunto conflictivo* a un conjunto de tuplas de asignaturas y cursos que comparten docentes. Por ejemplo, si la asignatura $m_1 \in M$ es dictada en el curso 3 por el mismo docente que dicta la asignatura $m_{12} \in M$ en el curso 8, luego el conjunto conflictivo que representa esta situación es $\{(m_1, 3), (m_{12}, 8)\}$. Por esta razón, para cierta hora cátedra h y conjunto conflictivo K , no puede ocurrir que más de una tupla $(asignatura, curso) \in K$ sea asignada a h ; de lo contrario, un docente estaría

siendo asignado a dos cursos al mismo tiempo. Las simultaneidades son consideradas excepciones y, por lo tanto, se incluye en el conjunto conflictivo una sola tupla que representa a los cursos que forman parte de la simultaneidad. Notamos como CK_{TM} al conjunto de conjuntos de conflictos del turno mañana y CK_{TT} al del turno tarde.

Por otro lado, como hay docentes que enseñan durante los dos turnos, se debe evitar que estén dando clases simultáneamente durante la séptima hora del turno mañana y la prehora del turno tarde, pues ambas ocurren durante el mismo periodo de tiempo. Por esta razón, se construye un conjunto de tuplas que agrupa asignaturas que son dictadas por un docente que enseña en ambos turnos siguiendo la siguiente idea: sea CK_{PH} el conjunto en cuestión, supongamos que un mismo docente dicta la asignatura m_4 en el curso 0 (que corresponde al turno mañana) y a los cursos 12 y 13 (que corresponden al turno tarde); entonces, se agregará a CK_{PH} las tuplas $(m_4, 0, m_4, 12)$ y $(m_4, 0, m_4, 13)$. De esta manera, se podrá indicar con el modelo que si la asignatura m_4 en el curso 0 es asignada en una séptima hora, entonces en la prehora de ese día no puede asignarse m_4 en los cursos 12 ni 13, o viceversa. Notamos al conjunto de estas tuplas como CK_{PH} y se denominará *conjunto de conflictos por prehora*.

Con respecto a las simultaneidades, se notará cada una como s . La simultaneidad está representada por una tupla cuya primera componente es la asignatura a dictar y la segunda es el conjunto de cursos donde debe ser dictada simultáneamente. Por ejemplo, si la asignatura $m \in M$ debe ser dictada simultáneamente en los cursos 6 y 7, entonces $s = (m, \{6, 7\})$ representa este requerimiento. Se nota como S al conjunto de simultaneidades.

Las variables de decisión son las siguientes, con $i \in M$, $h \in H$, $c \in C$:

$$x_{ihc} = \begin{cases} 1 & \text{si la asignatura } i \text{ se dicta en la hora } h \text{ en el curso } c \\ 0 & \text{c.c.} \end{cases}$$

$$u_{ihc} = \begin{cases} 1 & \text{si la asignatura } i \text{ tiene un bloque en el curso } c \text{ que comienza en la hora } h \\ 0 & \text{c.c.} \end{cases}$$

A continuación, presentamos los conjuntos de restricciones del modelo:

R1: Un curso no puede tener asignadas materias en horas cátedra que no corresponden a su turno:

$$\sum_{i \in M} \sum_{h \in H_{TT}} x_{ihc} = 0 \quad \forall c \in C_{TM} \quad (1)$$

$$\sum_{i \in M} \sum_{h \in H_{TM}} x_{ihc} = 0 \quad \forall c \in C_{TT} \quad (2)$$

R2: Cantidad de horas cátedra máximas y mínimas por día. En el caso de estudio, la cantidad máxima y mínima para el turno mañana es 7 y para el turno tarde 6 y 8 respectivamente:

$$\sum_{i \in M} \sum_{h \in H_d} x_{ihc} \leq 7 \quad \forall c \in C_{TM}, \forall d \in \Delta \quad (3)$$

$$\sum_{i \in M} \sum_{h \in H_d} x_{ihc} \geq 7 \quad \forall c \in C_{TM}, \forall d \in \Delta \quad (4)$$

$$\sum_{i \in M} \sum_{h \in H_d} x_{ihc} \leq 8 \quad \forall c \in C_{TT}, \forall d \in \Delta \quad (5)$$

$$\sum_{i \in M} \sum_{h \in H_d} x_{ihc} \geq 6 \quad \forall c \in C_{TT}, \forall d \in \Delta \quad (6)$$

R4: A lo sumo una asignatura por hora cátedra

$$\sum_{i \in M} x_{ihc} \leq 1 \quad \forall c \in C, \forall k \in H \quad (7)$$

R5: Se debe satisfacer la carga horaria semanal de la asignatura

$$\sum_{h \in H} x_{ihc} = cgs_{i,c} \quad \forall c \in C, \forall i \in M \quad (8)$$

R6: si no hay prehora, no se deben dictar asignaturas en la prehora. En el caso de estudio, el turno mañana no tiene prehoras

$$\sum_{h \in \{0,16,32,48,64\}} x_{ihc} = 0 \quad \forall i \in M, \forall c \in C \quad (9)$$

R7: los alumnos no pueden tener horas libres. Para esto, se pide que se asigne una asignatura a todas las horas entre la primera y la sexta, inclusive. De esa manera, las únicas horas en las que puede no asignarse una asignatura son la prehora y la séptima

$$\sum_{i \in M} x_{ihc} = 1 \quad \forall c \in C_{TM}, \forall h \in \{h \in H_{TM} : h \not\equiv 0 \pmod{8} \wedge h \not\equiv 7 \pmod{8}\} \quad (10)$$

$$\sum_{i \in M} x_{ihc} = 1 \quad \forall c \in C_{TT}, \forall h \in \{h \in H_{TT} : h \not\equiv 0 \pmod{8} \wedge h \not\equiv 7 \pmod{8}\} \quad (11)$$

R8: un docente no puede estar en distintos cursos al mismo tiempo. Para esto utilizaremos los conjuntos de conjuntos de conflictos de ambos turnos y

los conflictos por prehora

$$\sum_{(i,c) \in K} x_{ihc} \leq 1 \quad \forall h \in H, \forall K \in CK_{TM} \quad (12)$$

$$\sum_{(i,c) \in K} x_{ihc} \leq 1 \quad \forall h \in H, \forall K \in CK_{TT} \quad (13)$$

$$x_{k_1,h,k_2} + x_{k_3,h+1,k_4} \leq 1 \quad \forall h \in \{h \in H_{TM} : h \equiv 7 \pmod{8}\}, \\ \forall k = (k_1, k_2, k_3, k_4) \in CK_{PH} \quad (14)$$

R9: se fija la simultaneidad del dictado de asignaturas cuando es requerido. Además, se pide que las horas de cada una de las simultaneidades sean ubicadas teniendo en cuenta la disponibilidad de todos los docentes implicados en ella. De define H_s el conjunto de horas cátedra donde están disponibles todos los docentes implicados en la simultaneidad s :

$$x_{s_1,h,s_2_k} = x_{s_1,h,s_2_j} \quad \forall h \in H, \forall s_2_k, s_2_j \in s_2, s_2_k \neq s_2_j, \forall s \in S \quad (15)$$

$$\sum_{h \in H_s} x_{s_1,h,s_2_j} = cgs_{(s_1,s_2_j)} \quad \forall s_2_j \in s_2 \forall s \in S \quad (16)$$

R10: se determina la formación de bloques para las simultaneidades. Dada $s = (s_1, s_2)$ una simultaneidad, se tomará como representante del conjunto de cursos de s a $s_r = \text{mín } s_2$. Definimos el conjunto $B_s = \{h \in H_s : h + 1 \in H_s\}$ el conjunto de horas que pueden encabezar bloques. Se define como β_s la cantidad de bloques que puede formar s_1 . Así, (17) establece que, si h no puede encabezar un bloque, u_{ihc} debe valer cero; (18) obliga que se formen todos los bloques posibles; (19) determina que si un bloque empieza en h , otro no puede empezar en $h + 1$ (es decir, los triples se cuentan como un solo bloque) y, finalmente, (20) relaciona a las variables u con las variables x .

$$u_{s_1,h,s_r} = 0 \quad \forall h \in H \setminus B_s, \forall s \in S \quad (17)$$

$$\sum_{h \in B_s} u_{s_1,h,s_r} = \beta_s \quad \forall s \in S \quad (18)$$

$$u_{s_1,h,s_r} + u_{s_1,h+1,s_r} \leq 1 \quad \forall h \in B_s, \forall s \in S \quad (19)$$

$$x_{s_1,h,s_r} + x_{s_1,h+1,s_r} \geq 2 - (1 - u_{s_1,k,s_r}) \cdot 30 \quad \forall h \in B_s, \forall s \in S \quad (20)$$

La función objetivo a minimizar es:

$$\sum_{i \in M} \sum_{c \in C} \sum_{h \in H} x_{ihc}$$

Se eligió esta función objetivo dado que el conjunto de restricciones $R5$ hacen que tenga el mismo valor para toda solución factible. Recordar que el objetivo

del modelo es hallar horarios de clase que cumplan con las restricciones duras ($R1$ a $R10$). Por lo tanto, cualquier solución factible debe ser óptima.

Para el caso de estudio, el *solver* resolvió el modelo en 0,22 segundos y en 0,35 segundos encontró otras 25 soluciones factibles, sumando así un total de 26 horarios de clases iniciales. En la siguiente subsección se describe qué algoritmo de búsqueda local se aplicará a cada uno de ellos con el objetivo de transformarlos en horarios de clase factibles y minimizar, en la medida de lo posible, la violación de las restricciones blandas.

3.2. Aplicación del Algoritmo de Búsqueda Local

Para poder aplicar un algoritmo de búsqueda local, es necesario especificar la representación del problema, el objetivo y la función de evaluación [13]. Estos tres elementos son esenciales para manipular las soluciones candidatas a óptimos, definir el propósito del problema y establecer una forma de comparar la calidad de dos soluciones. El horario de clases es representado como lo que se conoce en el ámbito de las escuelas secundarias con el nombre de grilla: básicamente tiene la estructura de una matriz con tantas columnas como cursos y tantas filas como horas de clase semanales. En la fila i columna j se encuentra el nombre de la asignatura que se imparte en el curso j durante la hora i . En el caso de estudio, al haber 22 cursos y un máximo de 8 horas diarias de clase, la grilla tiene una dimensión de 40×22 (la hora 1 es la prehora del Lunes mientras que la hora 40 es la última hora cátedra del Viernes).

El objetivo del algoritmo de búsqueda local será minimizar el valor de la función de penalidad Ψ . Sea \mathcal{G} el espacio de grillas que cumplen con las restricciones duras (pero no necesariamente con las restricciones duras suavizadas), se define $\Psi: \mathcal{G} \rightarrow \mathbb{R}$ como la función de penalidad que, cuantificando la violación de las restricciones blandas y duras suavizadas, permite comparar la calidad entre distintos horarios de clases. Es decir, sean G_1 y G_2 grillas que representan distintos horarios de clases, diremos que G_1 es de mejor calidad que G_2 si $\Psi(G_1) < \Psi(G_2)$. Para definir Ψ mantuvimos una estrecha comunicación con las autoridades de la escuela, de manera tal que la función de penalidad reflejara de manera fiel la prioridad que se le asigna al cumplimiento de cada restricción blanda y a las relaciones entre ellas. Por ejemplo, es muy importante que una asignatura con una carga horaria semanal de dos horas forme un bloque, incluso si eso implica que un docente tenga una o dos horas inactivas más. Por otro lado, Ψ penaliza fuertemente la violación de las restricciones duras suavizadas, puesto que deseamos que el algoritmo de búsqueda local proporcione soluciones que las respeten y, por ende, que sean aplicables.

Como hemos mencionado anteriormente, también resulta necesario definir

la noción de vecindad en nuestro problema. Dada una grilla G que representa un horario de clases, una grilla G' es vecina de G si y sólo si G' respeta las restricciones duras y se obtiene realizando un solo intercambio en G de las asignaturas de dos horas de clase de un mismo curso. Dichas asignaturas deben ser distintas. En la Figura 1 se muestra un ejemplo.

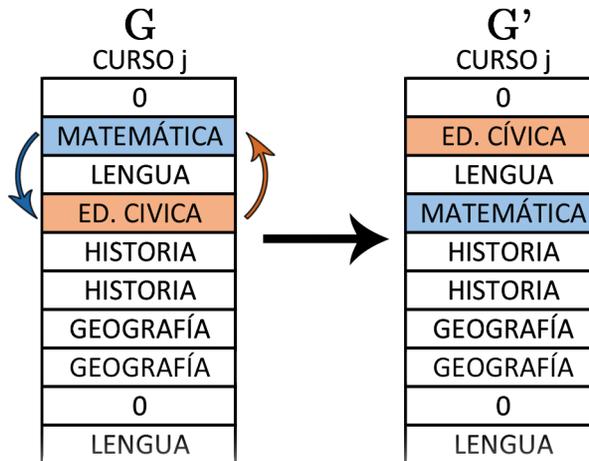


Figura 1: Ejemplo de generación de un vecino a partir del intercambio de horas entre dos materias en un curso.

El algoritmo de búsqueda local que empleamos está basado en el procedimiento de *hill-climbing*. El mismo parte de un estado inicial s y recorre su vecindario hasta hallar un vecino s' con menor valor de la función objetivo; en ese caso, el estado actual pasa a ser s' y se repite el procedimiento hasta no hallar vecinos con menor valor en la función objetivo o hasta alcanzar un límite de iteraciones o tiempo. La sencillez de este método es la causa de varias desventajas, como que el resultado final está fuertemente ligado al estado inicial o que es propenso a estancarse rápidamente en óptimos locales. En consecuencia, a lo largo de los años se han añadido modificaciones al *hill-climbing*, con el objetivo de sortear algunas de sus dificultades.

Como hemos anticipado en la sección anterior, uno de los cambios introducidos es el número de estados iniciales: con la primera parte de la Etapa I se generaron horarios de clase que respetan las restricciones duras a los que se les aplicarán el algoritmo de búsqueda local. De esta manera, esperamos recorrer más ampliamente el espacio de búsqueda.

Otra modificación que introducimos al método de *hill-climbing* es la componente probabilística que lo convierte en el método conocido en la literatura como *hill-climbing* estocástico [13]. En *hill-climbing*, el criterio para pasar de una grilla a una vecina es que la penalidad decrezca. Sin embargo, resulta in-

interesante experimentar qué sucede si, con cierta probabilidad, se permitiesen movimientos que empeoren la penalidad. La probabilidad de elegir al nuevo vecino v_n para ser el nuevo estado actual tiene una distribución Bernoulli con parámetro p , donde p depende de la diferencia de penalidad entre el estado actual v_c y el vecino en cuestión v_n y viene dado de la siguiente manera [13]:

$$p(v_c, v_n) = \frac{1}{1 + \exp\left(\frac{\Psi(v_n) - \Psi(v_c)}{T}\right)}$$

Notar la inclusión del parámetro T , que permanece constante a lo largo de la ejecución del algoritmo. Mientras más grande sea el valor de T , menor será la importancia de la mejora relativa entre los puntos v_n y v_c . En particular, si $T \rightarrow \infty$, la probabilidad de aceptación se aproxima a 0,5, transformando al *hill-climbing* estocástico en un paseo al azar. Por otro lado, si $T \rightarrow 0$, el *hill-climbing* estocástico pasa a comportarse como el *hill-climbing* ordinario. Por tanto, hay que encontrar T que no sea demasiado grande ni demasiado pequeño.

Para hallarlo, consideramos la escala que manejan las penalidades totales, que se encuentran en el orden de 10^4 , y estudiamos la función que define a p . Se debe tener en cuenta cuánto se le permitirá empeorar el valor de la función de penalidad a un vecino. Por ejemplo, si se considerara T tal que hay probabilidades no nulas de elegir un vecino que empeore la penalidad por un valor lo suficientemente grande, podría ocurrir que, en este vecino, se asigne una hora cátedra a un docente en la que no está disponible. Teniendo en cuenta este factor y con un poco de experimentación, concluimos que el intervalo $[50, 100]$ proporciona valores apropiados para T . Más precisamente, elegimos $T = 50$ pues demostró descender más rápido y al mismo tiempo aceptar vecinos peores (en promedio, el 33 % empeora la penalidad). Notar la probabilidad de elegir un vecino con la misma penalidad que la actual es de 0,5, sin importar T .

Por otra parte, consideramos importante introducir una leve modificación en el recorrido del vecindario que permite ir mejorando gradualmente el horario de clases de cada curso. La iteración comienza recorriendo el vecindario de la grilla actual efectuando intercambios en el horario del Curso 1. Si alguno de los vecinos es elegido, se efectúa el intercambio de horas en el horario del Curso 1. Si ningún vecino es elegido, no se realizan cambios. Independientemente de si se ha ejecutado o no una variación, se repite el procedimiento en el Curso 2. De esta manera, la iteración concluye al recorrer el vecindario del último curso (en el caso de estudio, el Curso 22). En el Algoritmo 1 se muestra el pseudocódigo de *hill-climbing* estocástico aplicado a cada grilla obtenida con el modelo de PLE de la sección anterior.

Para cada uno de los horarios de clases iniciales, ejecutamos el método de *hill-climbing* estocástico con 1000 iteraciones. En la Tabla 2 se muestran los datos sobre los tiempos de ejecución para el caso de estudio, con el hardware mencionado anteriormente.

Tabla 2: Tiempos de ejecución de *hill-climbing* estocástico para el caso de estudio

| | Tiempo (h:mm:ss) |
|-----------------|------------------|
| Promedio | 0: 30: 46 |
| Desvío estándar | 0: 03: 48 |
| Máximo | 0: 40: 35 |
| Mínimo | 0: 23: 03 |

Luego de haberles aplicado el algoritmo a las 26 grillas iniciales, 21 superaron la penalidad de la **grilla** elaborada manualmente por las autoridades de la escuela (penalidad estándar). De esas 21, 18 cumplieron con todas las restricciones duras suavizadas.

Con el desarrollo de un método que permita hallar buenos horarios de cursada en poco tiempo culmina el trabajo dedicado a resolver la Etapa I del problema. A continuación, se llevará a cabo la resolución de la Etapa II: distribuir las horas extracurriculares.

3.3. Modelo de PLE para la asignación de clases de apoyo

La notación empleada para plantear el problema de Programación Lineal Entera que permitirá asignar las clases de apoyo, se presenta en la Tabla 3.

Algoritmo 1 *Hill-climbing* estocástico aplicado a cada grilla obtenida a partir del modelo de PLE de la Sección 3.1

Entrada: G grilla, $iter_max$ cantidad máxima de iteraciones

```

1: procedure HILLCLIMBINGESTOCASTICO( $G, iter\_max$ )
2:    $mejor\_grilla \leftarrow G$ 
3:    $penalidad\_actual \leftarrow \Psi(G)$ 
4:    $minima\_penalidad \leftarrow \Psi(G)$ 
5:    $iter \leftarrow 0$ 
6:   mientras  $iter \leq iter\_max$  hacer                                ▷ loop 1
7:     para  $curso \in \mathcal{C}$  hacer                                       ▷ loop 2
8:        $S \leftarrow$  intercambios posibles de horas
9:       para  $s \in S$  hacer                                           ▷ loop 3
10:         $J \leftarrow$  vecino de  $G$  intercambiando las materias de las horas
      en  $s$ 
11:        si  $J$  cumple las restricciones duras entonces
12:           $nueva\_penalidad \leftarrow \Psi(J)$ 
13:           $u \leftarrow$  número aleatorio con distribución uniforme en  $[0, 1]$ 
14:          si  $u < p(G, J)$  entonces
15:             $penalidad\_actual \leftarrow nueva\_penalidad$ 
16:             $G \leftarrow J$ 
17:            si  $nueva\_penalidad < minima\_penalidad$  entonces
18:               $mejor\_grilla \leftarrow G$ 
19:               $minima\_penalidad \leftarrow nueva\_penalidad$ 
20:            fin si
21:            break loop 3                                ▷ Pasar al próximo curso
          fin si
        fin si
      fin para
    fin para
     $iter \leftarrow iter + 1$ 
  fin mientras
  devolver  $mejor\_grilla$ 
fin procedure

```

Tabla 3: Descripción de los conjuntos utilizados en el modelo de PLE para la asignación de clases de apoyo y sus características en el caso de estudio

| Notación | Descripción | Característica en el caso de estudio |
|--------------|--|---|
| M | Conjunto de las asignaturas sobre las que se dan clases de apoyo | $M = \{0, \dots, 19\}$ |
| M_{TM} | Conjunto de las asignaturas sobre las que se dan clases de apoyo para el turno mañana | $ M_{TM} = 16$ |
| M_{TT} | Conjunto de las asignaturas sobre las que se dan clases de apoyo para el turno tarde | $ M_{TT} = 14$ |
| Δ | Conjunto de días de la semana | $\Delta = \{0, \dots, 4\}$ |
| H | Conjunto de horas cátedra | $H = \{0, \dots, 79\}$ |
| H_{TM} | Conjunto de horas cátedra de la mañana | $H_{TM} = \{h \in H : \lfloor \frac{h}{8} \rfloor \equiv 0 \pmod{2}\}$ |
| H_{TT} | Conjunto de horas cátedra de la tarde, salvo las prehoras, pues el turno mañana siempre tiene séptima | $H_{TT} = \{h \in H : \lfloor \frac{h}{8} \rfloor \equiv 1 \pmod{2} \wedge h \not\equiv 0 \pmod{8}\}$ |
| H_d | Conjunto de horas cátedra correspondientes al día $d \in \Delta$ | $H_d = \{h \in H : h \equiv d \pmod{16}\}$ |
| $H_d^{(TM)}$ | Conjunto de horas cátedra correspondientes a la mañana del día $d \in \Delta$ | $H_d \cap H_{TM}$ |
| $H_d^{(TT)}$ | Conjunto de horas cátedra correspondientes a la tarde del día $d \in \Delta$ | $H_d \cap H_{TT}$ |
| P | Conjunto de profesores que dictan clases de apoyo | $P = \{0, \dots, 26\}$ |
| K_{TM} | Conjunto de horas que resultan inconvenientes para el turno mañana | $K_{TM} = \{h \in H_{TT} : h \bmod 16 \geq 12\}$ |
| K_{TT} | Conjunto de horas que resultan inconvenientes para el turno tarde | $K_{TM} = \{h \in H_{TM} : h \bmod 16 \leq 4\}$ |
| lim | Constante que representa el límite preferible de clases de apoyo en la misma hora cátedra, por cuestiones de espacio. | $lim = 3$ |
| T | Diccionario cuya clave es una asignatura y un turno y su valor es la cantidad de horas diarias máximas de apoyo preferibles. | |
| D | Diccionario cuya clave es un docente y su valor son las horas en las que puede dar clases de apoyo | |
| F | Diccionario cuya clave es un docente y su valor es un conjunto de horas favorables para dar clases de apoyo | $F_{(p)} \subseteq D_{(p)}$ |
| C | Diccionario con el cargo de clases de apoyo de cada docente | |

Por un lado se introduce el diccionario T donde se establece un tope para la cantidad de horas de clases de apoyo diarias de una asignatura con el fin de

lograr una mejor distribución semanal. Esto es relevante ya que los horarios de gimnasia ocupan a los estudiantes en distintos horarios y días, por lo que la mejor manera de maximizar su accesibilidad a las clases de apoyo es procurar que estas estén bien distribuidas. El tope se determina según la cantidad de horas de apoyo que se brindan de la asignatura en ese turno: a menor oferta, menor es el tope. En el caso de estudio se utilizaron los siguientes topes en función de la oferta $O_{(i,t)}$ de clases de apoyo de la asignatura i para el turno t :

$$T_{(i,t)}(O_{(i,t)}) = \begin{cases} 0 & \text{si } O_{(i,t)} < 2 \\ 1 & \text{si } O_{(i,t)} = 2 \\ 2 & \text{si } O_{(i,t)} \in \{3, 4, 5\} \\ 3 & \text{si } O_{(i,t)} \in \{6, 7, 8\} \\ 4 & \text{c.c.} \end{cases}$$

Naturalmente, al tratarse de una restricción blanda, se aplicarán penalidades a la violación de los topes, pero no determinarán la factibilidad de una solución.

Asimismo, se introducen dos diccionarios que incumben a los docentes. Por un lado, un diccionario D con las horas en las que cada docente puede dictar clases de apoyo; es decir, las horas de disponibilidad del docente menos las horas que ya tiene ocupadas con clases. Por otro lado, un diccionario F con las horas favorables para cada docente: las horas favorables son horas en las que el docente está disponible y que son horas inactivas u horas próximas a a horas activas. Por tanto, para cada docente p , $F_{(p)} \subseteq D_{(p)}$.

Por último, se introduce un diccionario C con el cargo de cada docente en cuanto a clases de apoyo. En C , la clave es un docente y el valor es un conjunto de ternas (*asignatura, turno, carga*) que representa que ese docente debe dar *carga* horas de clase de apoyo de *asignatura* para *turno*.

Sean $p \in P$, $i \in M$, $h \in H$, $d \in \Delta$ y $t \in \{0, 1\}$, se definen las variables del problema:

$$x_{pjh} = \begin{cases} 1 & \text{si el profesor } p \text{ da apoyo de } i \text{ en la hora } h \\ 0 & \text{c.c.} \end{cases}$$

y_{it} : cantidad de clases de apoyo de la asignatura i en horas poco convenientes para el turno t , $y_{it} \in \mathbb{Z}$

v_{itd} : cantidad de horas excesivas de clase de apoyo de la asignatura i del turno t en el día d , para (i, t) tales que $O_{(i,t)} \geq 2$, $v_{itd} \in \mathbb{Z}$

u_{ip} : cantidad de horas de clase de apoyo de la asignatura i que no son convenientes para el profesor p , $u_{ip} \in \mathbb{Z}$

w_h : sobrecarga de horas de clase de apoyo en la hora h , $w_h \in \mathbb{Z}$

A continuación se presentan los conjuntos de restricciones del problema:

R1: un profesor no puede estar en dos lugares al mismo tiempo:

$$\sum_{i \in M} x_{pjh} \leq 1 \quad \forall p \in P, \forall h \in H$$

R2: la séptima hora del turno mañana se superpone con la prehora del turno tarde:

$$\sum_{i \in M} x_{p,i,h} + \sum_{i \in M} x_{p,i,h+1} \leq 1 \quad \forall p \in P, \forall h \in \{h \in H : h \equiv 7 \pmod{16}\}$$

R3: no se pueden dictar clases de apoyo durante la séptima hora del lunes del turno tarde:

$$x_{p,i,15} = 0 \quad \forall p \in P, \forall i \in M$$

R4: debe cumplirse la cantidad de horas de apoyo semanales. Las clases de apoyo para el turno mañana deben darse durante las horas de la tarde y viceversa. Las clases de apoyo solo pueden darse en las horas en las que el profesor no esté dando clase. Recordar que los elementos de $C_{(p)}$ son de la forma $c = (c_1, c_2, c_3)$, $c_1 \in M$, $c_2 \in \{0, 1\}$ y c_3 es la carga horaria:

$$\sum_{h \in H_{TT} \cap D_{(p)}} x_{p,c_1,h} = c_3 \quad \forall p \in P, \forall c \in \{c = (c_1, c_2, c_3) \in C_{(p)} : c_2 = 0\}$$

$$\sum_{h \in H_{TM} \cap D_{(p)}} x_{p,c_1,h} = c_3 \quad \forall p \in P, \forall c \in \{c = (c_1, c_2, c_3) \in C_{(p)} : c_2 = 1\}$$

R5: se desea ubicar las clases de apoyo en horarios convenientes para los alumnos de ambos turnos:

$$\sum_{p \in P} \sum_{h \in K_{TM}} x_{pjh} \leq y_{i0} \quad \forall i \in M_{TM}$$

$$\sum_{p \in P} \sum_{h \in K_{TT}} x_{pih} \leq y_{i1} \quad \forall i \in M_{TT}$$

R6: se penaliza la concentración de clases de apoyo de una asignatura en el mismo día:

$$\sum_{p \in P} \sum_{h \in H_d^{(TT)}} x_{pih} \leq T_{(i,0)} + v_{i0d} \quad \forall d \in \Delta \forall i \in M, O_{(i,0)} > 1$$

$$\sum_{p \in P} \sum_{h \in H_d^{(TM)}} x_{pih} \leq T_{(i,1)} + v_{i1d} \quad \forall d \in \Delta \forall i \in M, O_{(i,1)} > 1$$

R7: se desea que las horas de clase de apoyo se integren bien a los horarios de los docentes, preferentemente llenando horas inactivas que hayan quedado de la confección del horario de clases:

$$\sum_{h \in H \setminus F_{(p)}} x_{pih} = u_{ip} \quad \forall i \in M \forall p \in P$$

R8: es preferible que no hayan demasiadas clases de apoyo dictándose al mismo tiempo:

$$\sum_{p \in P} \sum_{i \in M} x_{pih} \leq \text{lim} + w_h \quad \forall h \in H$$

La función objetivo a minimizar para el caso de estudio es la siguiente:

$$\sum_{p \in P} \sum_{i \in M} \sum_{h \in H} x_{pih} + 25 \cdot \sum_{i \in M} \sum_{t \in \{0,1\}} y_{it} + 25 \cdot \sum_{\substack{i \in M \\ O_{(i,t)} > 1}} \sum_{t \in \{0,1\}} \sum_{d \in \Delta} v_{itd} + 20 \cdot \sum_{i \in M} \sum_{p \in P} u_{ip} + 15 \cdot \sum_{h \in H} w_h$$

Los coeficientes utilizados para representar a la penalidad por infringir a las restricciones blandas fueron obtenidos mediante un ajuste experimental: las autoridades de la escuela otorgaron un orden de prioridad y se ajustaron los coeficientes hasta que el resultado de la asignación les resultó satisfactorio. Este es un problema de menores dimensiones que el resuelto para generar los horarios de clases iniciales. En efecto, con el hardware mencionado al comienzo de la sección, la asignación de clases de apoyo es resuelta en 0,08 segundos. Continuamos entonces con la asignación de los proyectos.

3.4. Modelo de PLE para la asignación de proyectos extraescolares

Hay dos tipos de actividades que se consideraban proyectos: en las que participan los estudiantes y en las que no. Luego, como ocurría en el caso de las clases de apoyo, se penalizará que la asignación de proyectos en los que

participan alumnos a horarios que les resultan inconvenientes. También se le otorgará prioridad a ocupar horas inactivas de los docentes que hayan quedado luego de la distribución de las horas de clase y de las horas de apoyo.

Con respecto a los proyectos, en el caso de estudio no se impusieron límites a cuántos pueden haber por día o por hora cátedra. Si existieran tales límites, las restricciones del problema de PLE que las representan serían similares a las de la sección anterior. Para resolver este problema, se introducen algunos diccionarios y conjuntos.

Se define como R al diccionario donde las claves son los docentes y los valores es el conjunto de proyectos que tiene a cargo cada uno. Así pues, $R_{(p)}$ es un conjunto de tuplas $r = (r_1, r_2, r_3, r_4)$ donde r_1 es el proyecto, r_2 es el turno, r_3 es la carga horaria semanal y r_4 vale 1 si participan los alumnos y 0 en caso contrario.

Además, se introducen distintas penalidades por asignar proyectos donde deben participar alumnos a horarios poco favorables para los ellos. Para ambos turnos se definen las horas: κ_{TM} a partir de las cuales es demasiado tarde para el turno mañana y κ_{TT} antes de la cual es demasiado temprano para el turno tarde. En el caso de estudio, se estableció que κ_{TM} es la tercera hora de la tarde y que κ_{TT} es la cuarta hora del turno mañana. En consecuencia, la función de penalidad que depende de la hora h se define como:

$$\rho(h) = \begin{cases} (\kappa_{TT} + 1 - (h \bmod 16))^2 & \text{si } h \in K_{TT} \\ ((h \bmod 16) - \kappa_{TM} + 1)^2 & \text{si } h \in K_{TM} \\ 0 & \text{c.c.} \end{cases}$$

Así, por ejemplo, la penalidad por asignar un proyecto en la segunda hora del turno mañana del martes se calcula como:

$$\rho(18) = (4 + 1 - (18 \bmod 16))^2 = 9$$

La notación utilizada para este modelo se encuentra en la Tabla 4.

Sean $p \in P$, $i \in M$, $h \in H$ y $t \in \{0, 1\}$ se definen las variables de decisión del problema:

$$x_{pih} = \begin{cases} 1 & \text{si el profesor } p \text{ da el proyecto } i \text{ en la hora } h \\ 0 & \text{c.c.} \end{cases}$$

y_{itk} : cantidad de veces que el proyecto i ocurre en la hora inconveniente h para el turno t (definido para $i \in M_A$), $y_{itk} \in \mathbb{Z}$

u_{ip} : cantidad de horas de proyecto i que no son convenientes para el profesor p , $u_{ip} \in \mathbb{Z}$

Tabla 4: Descripción de los conjuntos utilizados en el modelo de PLE para la distribución de proyectos extraescolares y sus características en el caso de estudio

| Notación | Descripción | Característica en el caso de estudio |
|----------|--|--|
| M | Conjunto de proyectos | $M = \{0, \dots, 20\}$ |
| M_{TM} | Conjunto de proyectos para el turno mañana | $ M_{TM} = 14$ |
| M_{TT} | Conjunto de proyectos para el turno tarde | $ M_{TT} = 13$ |
| M_A | Conjunto de proyectos donde participan alumnos | $ M_A = 11$ |
| H | Conjunto de horas cátedra | $H = \{0, \dots, 79\}$ |
| H_{TM} | Conjunto de horas cátedra de la mañana | $H_{TM} = \{h \in H : \left\lfloor \frac{h}{8} \right\rfloor \equiv 0 \pmod{2}\}$ |
| H_{TT} | Conjunto de horas cátedra de la tarde , salvo las prehoras, pues el turno mañana siempre tiene séptima | $H_{TT} = \{h \in H : \left\lfloor \frac{h}{8} \right\rfloor \equiv 1 \pmod{2} \wedge h \not\equiv 0 \pmod{8}\}$ |
| P | Conjunto de profesores que llevan a cabo proyectos | $P = \{0, \dots, 27\}$ |
| K_{TM} | Conjunto de horas que resultan inconvenientes para el turno mañana | $K_{TM} = \{h \in H_{TT} : h \bmod 16 \geq 12\}$ |
| K_{TT} | Conjunto de horas que resultan inconvenientes para el turno tarde | $K_{TM} = \{h \in H_{TM} : h \bmod 16 \leq 4\}$ |
| D | Diccionario cuya clave es un docente y su valor son las horas en las que puede llevar a cabo proyectos | |
| F | Diccionario cuya clave es un docente y su valor es un conjunto de horas favorables para llevar a cabo proyectos | $F_{(p)} \subseteq D_{(p)}$ |
| C | Diccionario con el cargo de clases de apoyo de cada docente | |
| R | Diccionario donde las claves son los docentes y los valores es el conjunto de proyectos que tiene a cargo cada uno | |

Los conjuntos de restricciones para el problema son los siguientes:

R1: un profesor no puede estar en dos lugares al mismo tiempo:

$$\sum_{i \in M} x_{pih} \leq 1 \quad \forall p \in P, \forall h \in H$$

R2: la séptima hora del turno mañana se superpone con la prehora del turno tarde:

$$\sum_{i \in M} x_{p,i,h} + \sum_{i \in M} x_{p,i,h+1} \leq 1 \quad \forall p \in P, \forall h \in \{h \in H : h \equiv 7 \pmod{16}\}$$

R3: no se pueden dictar proyectos durante la séptima hora del Lunes del

turno tarde:

$$x_{p,i,15} = 0 \quad \forall p \in P, \forall i \in M$$

R4: debe cumplirse la cantidad de horas de proyecto. Los proyectos para el turno mañana deben darse en el turno tarde y viceversa. Los proyectos solo pueden darse en las horas en las que el profesor no esté dando clase:

$$\sum_{h \in D_{(p)} \cap H_{TT}} x_{p,r_1,h} = r_3 \quad \forall p \in P, \forall (r_1, 0, r_3, r_4) \in R_{(p)}$$

$$\sum_{h \in D_{(p)} \cap H_{TM}} x_{p,r_1,h} = r_3 \quad \forall p \in P, \forall (r_1, 1, r_3, r_4) \in R_{(p)}$$

R5: se desea ubicar los proyectos en los que participan alumnos en horarios convenientes para los alumnos de ambos turnos:

$$\sum_{p \in P} x_{pih} \leq y_{i0h} \quad \forall h \in K_{TM} \forall i \in M_A$$

$$\sum_{p \in P} x_{pih} \leq y_{i1h} \quad \forall h \in K_{TT} \forall i \in M_A$$

R6: se desea que las horas de proyectos se integren bien a las fichas de los docentes, preferentemente ocupando horas inactivas:

$$\sum_{h \in H \setminus F_{(p)}} x_{pih} = u_{ip} \quad \forall i \in M \forall p \in P$$

La función a minimizar es la siguiente:

$$\sum_{p \in P} \sum_{i \in M} \sum_{h \in H} x_{pih} + 10 \cdot \sum_{i \in M} \sum_{t \in \{0,1\}} \sum_{h \in H} \rho(h) y_{ith} + 20 \cdot \sum_{i \in M} \sum_{p \in P} u_{ip}$$

Al igual que ocurría con el modelo para asignar las clases de apoyo, los pesos fueron ajustados experimentalmente y este problema es de rápida resolución: el solver necesitó sólo 0,1 segundos. La asignación de horas para los proyectos es el último paso en la confección del horario escolar, por lo tanto el procedimiento de construcción termina una vez obtenido el resultado de este problema de PLE.

4. Comparación con la elaboración manual

4.1. Caso de estudio

Hemos tomado como caso de prueba a una escuela pública de la Ciudad de Buenos Aires. Contamos con los datos utilizados para elaborar los horarios que rigieron durante el ciclo lectivo de 2017 y de 2018. En el caso de prueba, el turno mañana está compuesto por doce cursos y el turno tarde por diez, sumando un total de veintidós cursos y 770 horas de clase que deben ser programadas. La escuela cuenta con un plantel de 50 docentes y con la existencia de prehora para los cursos del turno tarde. Debido a que los docentes y directivos ocasionalmente deben reunirse en un Taller Docente los Lunes a las 17:20 horas, ese día no pueden programarse clases en la séptima hora para el turno tarde. En ambos turnos, los alumnos tienen dos recesos de diez minutos: el primero, entre la segunda y tercera hora y el segundo, entre la cuarta y la quinta.

Por otro lado, la escuela cuenta con los siguientes recursos edilicios: dos salas de informática, cuyo uso debe ser distribuido entre seis asignaturas; un laboratorio de biología; un laboratorio de físico-química; un salón para actividades artísticas, requerido por cuatro asignaturas; y un salón de usos múltiples. Debido a que la oferta de recursos excede la demanda, suele suceder que ciertas asignaturas se ven obligadas a prescindir de ellos o a turnarse para utilizarlos. Lo preferible es que la superposición horaria en el requerimiento de los mismos sea la mínima posible.

A contraturno, la escuela ofrece clases de apoyo escolar de 16 materias para el turno mañana y de 14 materias para el turno tarde, dictadas por 27 profesores del plantel docente. Por cuestiones de espacio físico, las autoridades de la escuela establecen que puedan dictarse a lo sumo tres clases de apoyo simultáneamente. Por otra parte, se ofrecen once proyectos en los que pueden participar los alumnos.

4.2. Resultados de la aplicación de la heurística desarrollada

Para evaluar el desempeño de los horarios elaborados utilizando la herramienta que desarrollamos en este trabajo, utilizamos como indicadores las penalizaciones por violar las restricciones blandas y las restricciones duras suavizadas. La suma de ellas es el valor de la función de penalidad Ψ para el

horario de clases. Describimos entonces los indicadores de desempeño:

- *Horas separadas de asignatura*: dada una asignatura que se dicta en un día durante dos o más horas, se prefiere que dichas horas sean consecutivas.
- *Bloques de asignatura*: se entiende por bloque a un conjunto de dos o más horas consecutivas en las que se imparte la misma materia. Se desea que las materias conformen la mayor cantidad de bloques de dos horas posible.
- *Cantidad de triples*: mientras que es posible dictar tres horas de una materia en un mismo día, se prefiere que esto ocurra la menor cantidad de veces posible.
- *Horas inactivas de los docentes*: es preferible que los docentes tengan la menor cantidad de horas inactivas durante su horario laboral. Como el trabajo del docente es remunerado según el cargo, una hora inactiva es tiempo perdido para el docente.
- *Excesiva demanda de recursos*: la escuela cuenta con un conjunto de recursos y cada uno cuenta con una demanda distinta. Se prefiere que la superposición horaria del requerimiento de recursos sea la menor posible.
- *Cantidad excesiva de prehoras*: se desea que, por una cuestión de espacio físico, no hayan más de 6 cursos del turno tarde con prehora por día.
- *Intervalos interruptores*: es preferente que los intervalos de los alumnos no interrumpan los bloques de las materias.

Además, incluimos la penalidad por no respetar las restricciones duras suavizadas para mostrar que los horarios generados automáticamente también las cumplen. Agregamos el requerimiento de no dictar clases en el turno tarde durante la séptima hora del Lunes como otra restricción dura suavizada.

En el la Tabla 5 se encuentra la comparación entre las penalidades del horario de clases elaborado manualmente para el año 2017 y del generado con la heurística que desarrollamos, así como también la diferencia relativa entre ambos, calculada como:

$$\left(\frac{\text{penalidad_grilla_automatica}}{\text{penalidad_grilla_manual}} - 1 \right) \cdot 100$$

Como se puede observar, el horario de cursada generado automáticamente supera al elaborado manualmente en todas las categorías, salvo en las que el horario estándar no incurría en ninguna penalidad. Por ejemplo, en el horario

Tabla 5: Comparación entre las penalidades del horario de clases elaborado manualmente para el año 2017 y del generado con la heurística que desarrollamos

| | Grilla automática | Grilla manual | Diferencia relativa |
|--|-------------------|------------------|---------------------|
| Penalidad por no respetar disponibilidad docente | 0 | 0 | 0 % |
| Penalidad por séptima hora del Lunes TT | 0 | 0 | 0 % |
| Penalidad por horas inactivas de los docentes | 1.877,73 | 5.267,55 | -64,35 % |
| Penalidad por intervalos interruptores | 950 | 1.500 | -36,67 % |
| Penalidad por falta de bloques de asignaturas | 4.485 | 6.325 | -29,09 % |
| Penalidad por cantidad de <i>triples</i> | 600 | 4.200 | -85,71 % |
| Penalidad por horas separadas de asignaturas | 0 | 2.600 | -100 % |
| Penalidad por cantidad excesiva de prehoras | 0 | 0 | 0 % |
| Penalidad por excesiva demanda de recursos | 1.760 | 2.000 | -12 % |
| Penalidad total | 9.672,73 | 21.892,55 | -55,82 % |

de cursada estándar ocurría siete veces que se dictaban tres horas de la misma asignatura en un sólo día, mientras que esto ocurre una sola vez en el horario generado automáticamente. Asimismo, se logró reducir completamente la ocurrencia de asignaturas que se dictaban en horas separadas a lo largo de un mismo día y se mejoró la eficiencia del uso de los recursos. Con respecto a las horas extracurriculares, en el horario generado automáticamente sólo dos horas de apoyo y cinco horas de proyectos fueron asignadas a horarios poco convenientes para los alumnos.

Al aplicar la heurística a la elaboración automática del horario escolar para el 2018, el solver logró resolver el problema y poblar el *pool* de soluciones en 0,63 segundos. Se obtuvieron así 28 horarios iniciales a los que se les aplicó paralelamente el método de *hill-climbing* estocástico. Como era de esperar, la etapa de optimización fue la que consumió la mayor cantidad de tiempo: una hora y cuarenta y ocho minutos. Al horario con menor penalización que cumpliera con las restricciones duras suavizadas se le aplicó la distribución de horas de extraclase que, como ocurrió en el caso del horario de 2017, demoró menos de un segundo. En total, la generación del horario escolar tardó casi dos horas. En la Tabla 6 se encuentra la comparación entre las penalidades de los horarios elaborados manual y automáticamente. Es importante aclarar que entre los dos años hubieron varios cambios. Por un lado, hubieron modificaciones en el plantel docente debido a renunciadas y a jubilaciones. Los cargos de los docentes que no están más en la escuela se presentan a concurso. Si para Febrero todavía ningún docente reclama las horas correspondientes a esos cargos, se confecciona el horario escolar sin tener en cuenta su posible disponibilidad horaria. Una vez confeccionado el horario, el docente que desee concursar por esas horas deberá adaptar su disponibilidad horaria a ellas. Por esta razón,

Tabla 6: Comparación entre las penalidades del horario de clases elaborado manualmente para el año 2018 y del generado con la heurística que desarrollamos

| | Grilla automática | Grilla manual | Diferencia relativa |
|--|-------------------|---------------|---------------------|
| Penalidad por no respetar disponibilidad docente | 0 | 0 | 0 % |
| Penalidad por séptima hora del Lunes TT | 0 | 0 | 0 % |
| Penalidad por horas inactivas de los docentes | 1.688,10 | 4.602,62 | -63,32 % |
| Penalidad por intervalos interruptores | 1.150 | 1.700 | -32,35 % |
| Penalidad por falta de bloques de asignaturas | 3.105 | 6.095 | -49,06 % |
| Penalidad por cantidad de <i>triples</i> | 1.200 | 7.200 | -83,33 % |
| Penalidad por horas separadas de asignaturas | 650 | 5.200 | -87,50 % |
| Penalidad por cantidad excesiva de prehoras | 0 | 0 | 0 % |
| Penalidad por excesiva demanda de recursos | 480 | 6.160 | -92,21 % |
| Penalidad total | 8.273,10 | 30.957,62 | -73,28 % |

consideramos que las asignaturas que no tienen asignados profesores todavía, son dictadas por docentes con total disponibilidad horaria. De esta manera, en el proceso de optimización hay mayor libertad para intercambiar horarios y, al mismo tiempo, queda conformado para el futuro docente un horario que no posea muchas horas inactivas. Debido a la cantidad de cargos que debe ser cubiertos y la ductilidad que esto le proporciona al proceso de elaboración del horario, se han podido formar más bloques de asignaturas.

Otra modificación que vale la pena mencionar es la introducción de nuevas asignaturas debido a la disposición de la NES. Muchas de esas asignaturas requieren el uso de las salas de computación. A pesar de esto, la escuela no ha sido adecuada a la nueva demanda, lo que significa un mayor requerimiento de los mismos recursos con los que se contaba el año anterior. La mayor diferencia en cuanto a penalidad respecto al horario estándar se encuentra en esta categoría, por lo que se demuestra que el algoritmo ha contribuido al uso eficiente de los recursos.

Entre otros logros, se ha logrado disminuir la cantidad de bloques de tres horas consecutivas de la misma asignatura: en el horario estándar habían doce, mientras que en el horario obtenido con el programa hay sólo dos. Con respecto a las horas extraclase, sólo han quedado asignadas a horarios poco convenientes para los alumnos dos horas de clases de apoyo y cuatro horas de proyectos.

5. Conclusiones y trabajo futuro

En general, la elaboración manual de horarios es una tarea que comienza alrededor de Noviembre y debe estar terminada para mediados de Febrero, cuando comienza el nuevo año escolar y debe comunicarse a los docentes cómo quedan conformados sus horarios de trabajo. No es que la tarea requiera la dedicación exclusiva de todo ese tiempo, sino que, como debe efectuarla un directivo o un docente, quienes a su vez tienen otras ocupaciones más urgentes, la confección del horario se lleva a cabo espaciada y progresivamente. En comparación, el programa desarrollado en este trabajo permite confeccionar un horario en poco menos de dos horas. Por lo tanto, si debieran modificarse algunas características como la disponibilidad de ciertos docentes, se podría volver a ejecutar la herramienta para que elabore nuevos horarios acorde a las nuevas exigencias; siempre y cuando el ciclo lectivo no haya comenzado aún. Como se ha mostrado en las secciones anteriores, el programa no sólo efectúa la tarea rápidamente sino que también genera horarios que son mejores que los elaborados manualmente, según las prioridades que establecieron los directivos. Por tanto, puede concluirse que la heurística cumple con el objetivo propuesto: armar horarios de calidad rápidamente.

Concluimos que la utilización de métodos de búsqueda local permiten proponer una función de penalidad que represente mejor a los criterios que guían la elaboración de horarios en la vida real y su implementación es relativamente sencilla. Además, la aplicación de los algoritmos de búsqueda local permitió plantear ciertas hipótesis sobre el *espacio* de los horarios de cursada: todo parece apuntar a que existen extensos lomos y que los valles son planos, dado que todas las ejecuciones de *hill-climbing* básico culminaron al agotar la cantidad máxima de pasos laterales y no por la ausencia de vecinos de igual o menor penalidad (mínimo local). En este contexto, la variante estocástica del *hill-climbing* es la más próspera, puesto que ofrece una probabilidad de aceptar un empeoramiento razonable de la penalidad, permitiendo el escape de las mesetas, amén de tener un tiempo de ejecución que se ajusta a la meta de este trabajo.

Como ocurre en muchos problemas de Investigación Operativa, es esencial tener presente la opinión y el procedimiento que realiza la persona encargada de realizar la tarea manualmente. Ya que debe llevar a cabo el trabajo regularmente, ha adquirido mucha destreza en esta tarea y muchas de las ideas o técnicas que desarrolló pueden servir de inspiración al momento de automati-

zar el proceso. Tal es el caso de este trabajo. La ubicación prioritaria de las asignaturas que se dictan con simultaneidad y la división de la elaboración del horario en dos etapas, así como también los valores elegidos para las penalizaciones, fueron decisiones basadas en técnicas que utilizaban los directivos y la profesora que armaba manualmente el horario. Mantuvimos un diálogo con ellos a lo largo de la elaboración del trabajo y quedaron muy satisfechos con los resultados que brindó este programa.

Tomando la herramienta desarrollada en este trabajo como base, se pueden añadir ciertas mejoras. Por un lado, implementar un algoritmo que permita ajustar un horario ya armado a un cambio repentino (como el cambio en la disponibilidad de algún docente) realizando la menor cantidad de alteraciones posible. Otra mejora consiste en incluir en el modelo la preferencia que tienen los docentes por dictar clases en ciertas horas del turno mañana o del turno tarde. Esto supone también la implementación de una función que calcule el balance de cuántas preferencias de cada docente están siendo satisfechas. Asimismo, se pueden agregar más capas de complejidad a la función de penalidad para reflejar aún mejor la realidad. Por ejemplo, la función que penaliza horas inactivas de los docentes podría tener en cuenta no sólo la cantidad, sino también qué horas tienen libres: las horas libres más cercanas al mediodía podrán ser ocupadas en la Etapa II con clases de apoyo o proyectos, por lo que no deberían recibir igual penalización que las que se encuentran alejadas.

Más en general, si bien el orden de prioridades establecido por los directivos de la escuela con la que se trabajó es razonable, también sería interesante encontrar una escala de preferencias para el cumplimiento de las restricciones blandas más homogéneo u objetivo, que pudiese aplicarse a todas las escuelas y que no necesariamente dependa del criterio de cada autoridad.

Referencias

- [1] Abramson, D., Krishnamoorthy, M., Dang, H.,. Simulated annealing coloring schedules for the school timetabling problem. *Asia-Pacific Journal of Operational Research*, (16), 1999.
- [2] Bondy J.A. and Murty U.S.R. *Graph theory with applications*. North-Holland, 1976.
- [3] Chorbev I., Loskovska S., Dimitrovski I., Mihajlov D. Solving the high school scheduling problem modelled with constraints satisfaction using hybrid heuristic algorithms. En W. Bednorz, editor, *Greedy Algorithms*. InTech, 2008.

- [4] Côte , P., Wong, T., Sabourin, R. Application of a hybrid multi-objective evolutionary algorithm to the uncapacitated exam proximity problem. En Burke, E.K., Trick, M., editor, *Selected Papers from the 5th International Conference on the Practice and Theory of Automated Timetabling, Lecture Notes in Computer Science*, volume 3616. Springer, 2005.
- [5] Csima J. *Investigations on a Time-Table Problem*. PhD thesis, Institute of Computer Science, University of Toronto, 1965.
- [6] Deris, B., Omatu, S., Ohta, H., Samat, D. University timetabling by constraint-based reasoning: A case study. *Journal of Operational Research Society*, 12(48):1178–1190, 1997.
- [7] Di Gaspero, L., Schaerf, A. Tabu search techniques for examination timetabling. En Burke, E.K., Erben, W., editor, *Selected Papers from the 3rd International Conference on the Practice and Theory of Automated Timetabling, Lecture Notes in Computer Science*, volume 2079. Springer, 2001.
- [8] Nonobe, K., Ibaraki, T. A tabu search approach to the constraint satisfaction problem as a general problem solver. *European Journal of Operational Research*.
- [9] N. Saptarini, I. W. Suasnawa, y P. Ciptayani. Senior high school course scheduling using genetic algorithm. *Journal of Physics: Conference Series*, 953:012067, 01 2018.
- [10] Stuart Russell y Peter Norvig. *Artificial Intelligence: a Modern Approach*. Pearson Education, 2003.
- [11] Valouxis C., Gogos C., Alefragis P., Housos H. Decomposing the high school timetable problem. *Practice and Theory of Automated Timetabling*.
- [12] van der Kooy N. J. The high school scheduling problem: Improving local search fairness evaluation. Master’s thesis, Mathematical Institute, University of Leiden, 2017.
- [13] Zbigniew Michalewicz y David B. Fogel. *How to Solve It: Modern Heuristics*. Springer, 2004.

ANÁLISIS DE TEXTO PARA GESTIÓN DE MARCA CON EL USO DE INSTAGRAM APLICADO A UNA COMUNIDAD DE MASCOTAS EN CHILE

CINTHYA VERGARA *
VALERIA SCAPINI **
SIMÓN MATURANA *
FRANCISCO PAZ ***

Resumen

Las redes sociales se han convertido en un lugar de gran importancia para el análisis de clientes y posicionamiento de marca. El presente trabajo busca evaluar las interacciones de una cuenta de Instagram con 22.000 seguidores vinculada al mundo de los animales, principalmente domésticos, e identificar qué percepción hay del contenido mediante estadísticas y Análisis de Sentimientos de los comentarios. Para ello se recopilieron datos de 1.758 publicaciones y 6.116 comentarios. Sobre ellos se obtuvieron estadísticas mediante el uso de la API de Instagram y Análisis de Sentimientos con el modelo VADER que permite analizar texto y *emoticonos*. El análisis muestra que es posible hacer gestión de marca con Instagram de acuerdo a los datos otorgados en las cuentas mediante su API e identificar la polaridad de comentarios en base a texto y símbolos. De acuerdo con la técnica utilizada, la polaridad muestra que un 36% de los comentarios son positivos, 7% negativos y 57% neutral. Finalmente, el trabajo muestra que es posible analizar una comunidad en línea que interactúa a través redes sociales, que es de gran relevancia observar cómo se usan los iconos y las imágenes en la construcción de representaciones colectivas y que el Análisis de Redes Sociales contribuye al desarrollo de una marca.

Palabras Clave: Análisis de Redes Sociales, Instagram, Emoticonos, Análisis de Texto, Análisis de Sentimientos, Mascotas.

*Departamento Ingeniería Industrial, Universidad de Chile, Santiago, Chile

**Escuela de Ingeniería Comercial, Universidad de Valparaíso, Santiago, Chile

***Fundador de ZoooSalud, Santiago, Chile

1. Motivación y contexto

En nuestros días, hemos sido testigo de la aparición de una serie de herramientas informáticas basadas en tecnología web que han cambiando la forma de interacción social, han dado origen a nuevas formas de organización y han abierto oportunidades económicas y sociales. En particular, los Sitios de Redes Sociales (SNS, por su nombre en inglés Social Network Site) han permitido conectar personas y organizaciones para compartir información, conocimiento y experiencias de manera dinámica [5, 11, 26, 13].

Es así como, para distintas organizaciones, es posible obtener información, mejorar la productividad, gestionar recursos y transparentar actividades [6, 18, 30] permitiendo desarrollar estrategias de posicionamiento de marca y comprensión de mercado.

El procesamiento de información basado en la web, sin embargo, no está libre de dificultades. Por un lado, permite obtener estadísticas basadas en datos estructurados como lo son la cantidad de “Me Gusta”, seguimientos a la cuenta, número de comentarios, número de veces que es mirada una publicación, cantidad de cuentas que ven la publicación, entre otros, y por otro lado genera gran cantidad de información, basada textos o contenido audiovisual, que necesita de procesamiento previo para poder ser utilizada y generar valor.

En particular, organizaciones del mundo de las mascotas han visto en las redes sociales un espacio idóneo para el desarrollo de sus actividades gracias a la posibilidad de publicar campañas de adopción, promoción de actividades, difusión de consejos para tenencia responsable de mascotas e incluso la búsqueda de mascotas perdidas. En Chile, se estima que al 2019 un 73 % de los hogares posee una mascota[3] y el mercado tiene un crecimiento sostenido en los productos y servicios, el monto de venta en de comida premium creció un 221 % entre los años 2013-2018, ha tenido un crecimiento de alrededor del 92 % en ese periodo con una proyección monto en ventas para el 2022 de 1.152,6 millones de dólares [14].

Frente al impacto que tienen las redes sociales, en relación con la entrega de información y la creciente preocupación por las actividades relacionadas con el mundo de los animales domésticos, este artículo busca mostrar el comportamiento de usuarios en una cuenta de Instagram perteneciente a ZoooSalud que, con sus 22.000 seguidores, se encuentra dentro de las cuentas chilenas con más seguidores dentro del grupo de perfiles de comunidades de mascotas. El análisis del perfil se realizará mediante herramientas de Análisis de Texto y

la información que es posible capturar con la API de Instagram en su versión 2019.

El trabajo se organiza de la siguiente manera. En la primera sección se presenta la motivación del tema a estudiar, en la segunda sección se desarrolla el marco teórico y trabajos relacionados. En la tercera sección se presenta la metodología utilizada en el estudio, los resultados y finalmente se concluye al respecto.

2. Marco de referencia y trabajos relacionados

El Análisis de Redes Sociales (SNA, por su nombre en inglés Social Network Analysis) tiene sus orígenes tanto en las ciencias sociales como en los campos más amplios del análisis de redes y la teoría de grafos. Permite visualizar y modelar las entidades de una red, identificando relaciones entre estas entidades a través de algoritmos, estadísticas y herramientas informáticas [31, 32, 35]. En esta sección se muestran los conceptos básicos relacionados con el SNA.

2.1. Redes Sociales

Según Lozares [22] una red social es una conexión de un conjunto de individuos (actores, personas, empresas, grupos, comunidades, entre otras) por medio de reacciones sociales definidas (que puede ser una relación o un conjunto de relaciones sociales). Generalmente estos individuos comparten atributos o intereses, razón para que quieran interactuar entre ellos o con otros individuos de la red. Existen diversas formas para que los individuos puedan interactuar, entre ellas se encuentran los correos electrónicos, la mensajería, publicación en blogs o en páginas de perfil [25].

En términos generales, el análisis de redes sociales se puede realizar desde dos perspectivas. La primera, desde el punto de vista del actor en la red de acuerdo con cuántos vínculos tiene con otros actores, qué tipo de vínculos mantienen y qué clase de información fluye entre ellos. La segunda, en relación a la estructura completa de la red y se analiza principalmente los vínculos que todos los miembros mantienen dentro de un contexto [12].

Las redes sociales pueden ser analizadas a través de la teoría de grafos donde cada nodo o vértice representa un individuo y los arcos representan las relaciones o interacciones entre ellos. Siendo esta estructura una de las más utilizadas para ordenar una red social en las distintas disciplinas [2, 10, 23].

Por su parte, los Sitios de Redes Sociales (SNS) son plataformas web donde personas de todo el mundo, pueden conectarse con diferentes usuarios u orga-

nizaciones y establecer una red para compartir y expresar sus ideas, experiencias, opiniones, sentimientos, documentos, material audiovisual entre otros. El volumen de personas que interactúan a través de Internet en tales plataformas ha generado una gran cantidad de datos con contenido social [16, 35]. En julio de 2019 existían más de 4,33 mil millones de personas/usuarios activos de Internet, equivalente a un 56 % de la población mundial. Dentro de los usuarios activos de Internet existen más de 2.000 millones de usuarios que utilizan las redes sociales y se espera que sigan creciendo a medida que los dispositivos móviles vayan aumentando [8].

Dentro de las aplicaciones que encontramos en el Análisis de Redes Sociales se encuentran la gestión de percepción de marca con análisis de sentimientos, los resúmenes de flujos de contenido, estimación de importancia de contribuidores de contenido, métodos para recomendar anuncios, sistemas y métodos para etiquetado de mensajes sensibles al contexto, análisis de sitios web y publicidad basado en el contenido de redes sociales, optimización de la participación en las redes sociales, analizar y responder al contenido generado por el usuario, análisis en tiempo real de sentimientos en redes sociales, análisis de *Hashtags* entre muchas otras [4, 19, 33, 34].

Con todo, el uso de las redes sociales es muy diverso. En plataformas como *Facebook* o *Google+* se identifica un mayor preocupación en las relaciones entre amigos y familiares, *Tumblr* o *Twitter* tienen un perfil principalmente comunicacional con temas rápidos y contingentes [9] o *Instagram* que se centra en la comunidad e interacciones sociales cotidianas a través de historias, vídeos y fotos que muestran las actividades personales de sus usuarios [20].

2.2. Instagram

Instagram es una plataforma que se encuentra activa desde el año 2010, que permite compartir el estado actual mediante fotografías y vídeos, los cuales pueden ser ajustadas con diversos filtros. Es posible publicar en un hilo permanente o en historias que muestran contenido de manera temporal. En el año 2012 fue adquirida por *Facebook* y al año siguiente ya contaba con más de 100 millones de usuarios [30]. Actualmente la plataforma tiene más de 1.000 millones de usuarios [9, 29].

El medidor básico de una cuenta es la cantidad de seguidores y se conocen como: *Microinfluencers* a los usuarios que tienen un alcance reducido con un rango de seguidores entre 500 y 10.000, como *Influencers* a los que poseen entre 10.000 y 100.000 seguidores, *Macroinfluencers* a los que poseen entre 100.000 y 500.000 seguidores, *Megainfluencers*: entre 500.000 y 1.500.000 seguidores y como *Famosos* a los que poseen más de 1.500.000 seguidores.

Instagram es una plataforma que se ocupa extensivamente para el desarrollo de marcas y promociones a través de su modalidad de empresas. Es posible identificar cuentas que logran un gran alcance de personas de acuerdo con la cantidad de seguidores, cantidad de comentarios, reacción sobre los comentarios y capacidad de convertir la visualización del contenido publicado en alguna acción posterior, como por ejemplo comprar un producto o servicio, asistir a actividades, visitar ciertos lugares, promover campañas e ideas, entre otros [24].

2.3. Minería de Texto y Análisis de Sentimientos

La Minería de Texto o, en inglés, *Text Mining* es parte de los análisis provenientes de la minería de datos y del campo de la inteligencia artificial. Se entiende como un conjunto de procesos informáticos que buscan extraer conocimiento de acuerdo con un criterio de novedad o similitud en textos, para lo cual se utilizan modelos simplificados de teorías lingüísticas, estadísticas, y tecnologías de comprensión del lenguaje natural [1, 15].

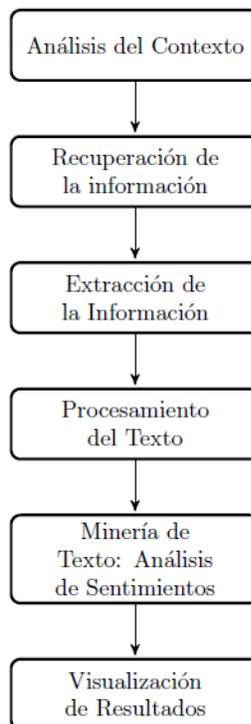


Figura 1: Etapas Minería de Texto

Dentro del análisis de textos se encuentra el Análisis de Sentimientos (en inglés, *Sentiment Analysis*) cuyo propósito es determinar si el sentimiento ge-

nerado por una oración es, en general, positivo o negativo. También se conoce como Análisis de Opinión (en inglés, Opinion Mining) dado que dentro de sus principales aplicaciones se encuentra identificar la emocionalidad de las opiniones vertidas sobre algún tema en ambientes web. La Figura 1 resume los pasos de la metodología.

El Análisis de Sentimientos es bastante complejo, donde una de sus principales dificultades se encuentra en la misma complejidad del lenguaje y sus usos. El sentimiento generado por una oración depende directamente del contexto, nivel cultural, costumbres, connotaciones y símbolos y una multitud de factores de mayor o menor importancia que pueden alterar el sentimiento generado por una declaración escrita.

Existen varias técnicas para realizar Análisis de Sentimientos [21, 27, 28] así como distintos diccionarios y herramientas de análisis de lenguaje natural. En términos informáticos también existen distintas herramientas que pueden tener mejor o peor rendimiento tanto computacional como de aproximación [17].

2.4. Valence Aware Dictionary and sEntiment Reasoner

Dentro de las herramientas para realizar Análisis de Sentimientos nos encontramos con la librería Valence Aware Dictionary and sEntiment Reasoner (VADER), que utiliza una combinación de métodos cualitativos y métodos cuantitativos para producir y validar un léxico de sentimiento estándar. Las características léxicas se combinan de acuerdo a reglas que incorporan convenciones gramaticales y sintácticas que las personas usan al expresar o enfatizar un sentimiento [17] incluyendo símbolos, como lo son los *emoticonos* o signos de exclamación. Además VADER puede trabajar simultáneamente con textos en inglés y español.

La valoración del sentimiento que genera cada palabra se basa en el concepto de la *sabiduría de la multitud* donde se asume que opinión colectiva es más confiable que la opinión individual y el diccionario en conjunto con los puntajes fueron estimados en base a las valoraciones del servicio Mechanical Turk operado por Amazon Web Services que permite a empresas contratar trabajadores ubicados de forma remota para realizar tareas a pedido.

Las clasificaciones de léxico de una oración se estiman sumando los puntajes de opinión donde cada palabra incluida en el diccionario VADER dentro de la oración se le asigna un puntaje dentro de la escala de $[-4]$ o Extremadamente negativo a $[4]$ o Extremadamente positivo, con asignación $[0]$ para un valor Neutro. El puntaje es una métrica que calcula la suma de todas las clasificaciones del léxico que se han normalizado entre -1 (más extremo negativo)

y +1 (más extremo positivo) de acuerdo a la siguiente formula:

$$\frac{x}{\sqrt{x^2 + \alpha}}$$

donde x es la suma de los puntajes de las palabras de que componen oración y α es un parámetro de normalización relacionado con el máximo valor esperado que, por defecto, toma el valor de 15.

Los puntajes positivos, neutros y negativos de cada palabra se suman y generan el puntaje de la oración o texto. A este indicador se le llama compuesto (en inglés compound) y los valores de cada rango determinan:

- sentimiento positivo: puntaje compuesto $\geq 0,05$
- sentimiento neutral : puntaje compuesto $> -0,05$ y $< 0,05$
- sentimiento negativo: puntaje compuesto $\leq -0,05$

En resumen, el algoritmo se construye gracias a un combinación de métodos cualitativos y cuantitativos que permite generar una lista estándar de características léxicas junto con una medida de intensidad de sentimiento. Incluye cinco heurísticas simples basadas en la puntuación, capitalización o uso de mayúsculas, modificadores de intensidad de acuerdo a un diccionario de palabras que pueden aumentar o disminuir la intensidad de otras palabras, el uso de la palabra “pero” para cambiar la polaridad y el análisis de palabras que podrían generar negación del significado de otra palabra. Su uso muestra un buen rendimiento en textos de poca longitud como los generados en micro-blogs.

3. Metodología

Para iniciar el estudio se llevó a cabo un análisis del contexto revisando la situación estratégica de ZoooSalud mediante la revisión de problemas y posibles causas mediante un árbol de problemas [7] para luego generar objetivos estratégicos y definir acciones para mejorar cumplirlos.

Dentro del análisis estratégico se identifica la necesidad de realizar gestión de usuarios mejorando el análisis de comportamiento en las distintas plataformas. Se revisó la cantidad de interacciones por plataforma y se seleccionó Instagram dado su gran cantidad de seguidores y su potencial para generar visitas al sitio web. La Figura 2 resumen el análisis realizado.

Con ello, para realizar el estudio se accedió directamente a los datos obtenidos desde la plataforma *Instagram* a través de su API y el lenguaje de



MAPA ESTRATÉGICO ZOOOSALUD



(a) Mapa de Objetivos Estratégico ZoooSalud



(b) Árbol de Diagnóstico

Figura 2: Análisis de Contexto ZoooSalud

programación Python 3.7.4. Se obtuvieron los datos generales de la cuenta y los comentarios realizados para cada una de las publicaciones.

Sobre los datos se aplicó la metodología de extracción de conocimiento a partir de datos [15] para análisis de texto. Se ordenaron los comentarios y se realizó una limpieza de texto que consistió principalmente en separar dentro del texto las palabras de los *emoticonos* para luego filtrar palabras mediante un diccionario de “*stopwords*” o palabras vacías, quitar dobles espacios, sacar sólo puntos y convertir las palabras a letras minúsculas. Se mantuvieron los signos de exclamación e interrogación además del signo utilizado para menciones o # utilizado para marcar temas.

Se aplicó sobre cada comentario el algoritmo de VADER para realizar el análisis de sentimientos y clasificación de acuerdo a la polaridad positiva, negativa, neutra y compuesta.

Finalmente se revisaron los resultados y se utilizó la información obtenida para evaluar la estrategia de contenidos de la cuenta bajo análisis.

4. Análisis de caso: ZoooSalud

ZoooSalud es una comunidad de amantes de los animales que busca el bienestar animal mediante la entrega de información y herramientas útiles para proteger, mejorar la calidad de vida, promover el bienestar animal y la tenencia responsable de mascotas. En sus orígenes, surgió la idea de mejorar la calidad de vida de los animales desde una mirada integral, considerando a todos los actores y los distintos aspectos que afectan el desarrollo de los animales, no sólo frente a problemas de salud. Se buscaba además darle importancia a las relaciones sociales, actividades y la entrega de beneficios para así generar un círculo virtuoso entre las personas que tengan bajo su responsabilidad animales y los proveedores de productos y servicios necesarios para su cuidado.

ZoooSalud tiene presencia en medios virtuales. Consta de un sitio web, una cuenta de Twitter, una de Facebook y una de Instagram. El sitio web funciona como centro de actividades y es posible encontrar en él mascotas en adopción, mascotas perdidas, consejos para aprender cómo cuidar correctamente a una mascota, eventos realizados por organizaciones y fundaciones de animales certificadas. Mientras que las cuentas de redes sociales se incluyeron en el año 2017 para difundir contenido del sitio web e información para mantener una comunidad activa. Espera convertirse en un referente para la comunidad habla hispana, no sólo de Chile, y mejorar el contenido ingresando a nuevas plataformas como un canal de *Youtube*.

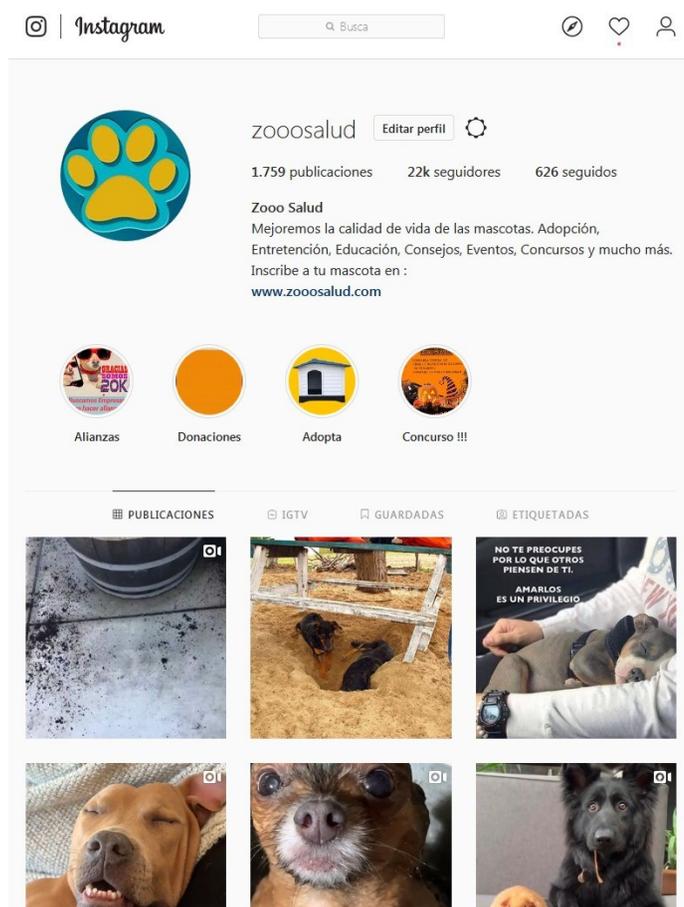


Figura 3: Cuenta de Instagram de ZooSalud

En particular la cuenta de Instagram es la que posee mayor número de seguidores e interacciones diarias. Con 22.000 seguidores, dentro del mundo de las redes sociales es posible considerarla como una cuenta de *influenciador* o, en inglés, *influencer*. La Tabla 1 resume las principales interacciones de la cuenta.

Tabla 1: Estadísticas de Interacciones

| Item | Cantidad |
|--|----------|
| Publicaciones | 1.758 |
| “Me Gusta” | 405.339 |
| Distintos usuarios que han dado “Me Gusta” | 182.555 |
| Distintos usuarios que han comentado | 4.146 |
| Comentarios | 6.116 |

Al ordenar los comentarios se realizó un análisis descriptivo. En primer lugar se identifica que la principal forma de comentar es a través de *emoticonos*. Al obtener las palabras y/o símbolos más frecuentes, nos encontramos con que los 14 primeros términos más frecuentes fueron *emoticonos* seguidos en el lugar 15 por la palabra *amor* y, en el lugar 16, la palabra *perro*. En la Tabla 2 se muestran los 10 emoticonos con más frecuencia dentro de los comentarios.

Tabla 2: Principales 10 emoticonos por frecuencia

| Emoticono | Frecuencia |
|-----------|------------|
| 😊 | 2.445 |
| 😁 | 1.238 |
| ♥ | 1.108 |
| 😊 | 1.056 |
| 👉 | 503 |
| 👩 | 310 |
| 😏 | 305 |
| 👉👉 | 218 |
| ♥♥ | 205 |
| 😏 | 194 |

Quitando los símbolos y *emoticonos* se obtuvo una nube de palabras que muestra los términos más frecuentes. Como es posible ver en la Figura 4 las palabras que más se repiten son *amor*, *perro/perros*, *jajaja*, *hermoso*, *dios*, *bello*, *perrito*, *animales*, *hermosa*, *quiero*, *belleza*.



Figura 4: Mapa de Palabras ZooSalud

Las palabras muestran de manera cualitativa que la comunidad de Zooo-Salud tiene una tendencia a comentar de manera positiva de acuerdo a los adjetivos utilizados en los comentarios. También es posible identificar que hay más comentarios relacionados con “perros” lo que podría indicar la cuenta comparte principalmente contenido de esta especie en particular.

Es de interés conocer en qué momento del día se realizan las interacciones a través de comentarios para poder mantener conversaciones con la comunidad. Luego, el horario en que más se realizan comentarios es entre las 19:00 y 23:00 horas, horario de Chile continental, como es posible ver en la Figura 5.

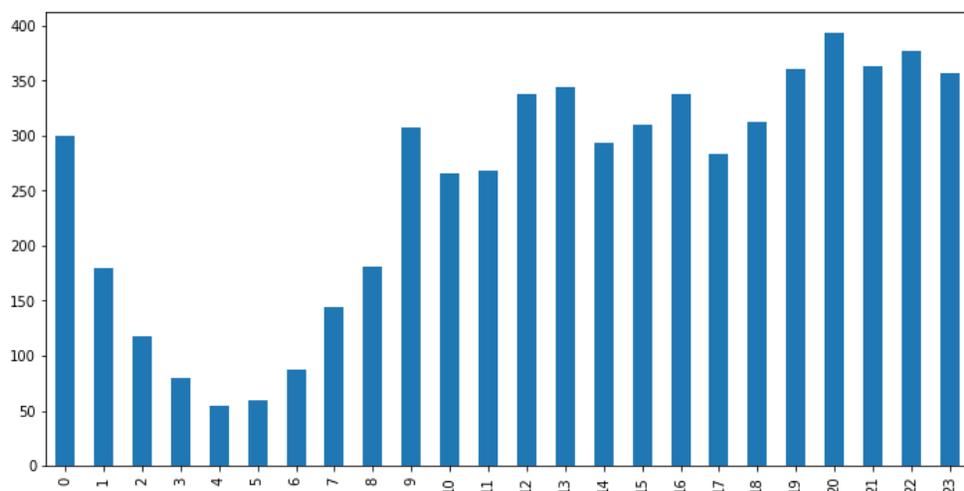


Figura 5: Comentarios por hora

Otro aspecto a evaluar son los usuarios que más participan dentro del sitio. En relación a las cantidad de comentarios, se identifica un usuario que ha realizado 201 comentarios mientras que en general los usuarios tienen en promedio 1,5 comentarios sobre el total de publicaciones a la fecha. La Tabla 3 muestra a los 10 usuarios que más comentan.

Tabla 3: Top 10 usuarios que más comentan

| Usuario | Comentarios |
|------------|-------------|
| Usuario 1 | 201 |
| Usuario 2 | 47 |
| Usuario 3 | 41 |
| Usuario 4 | 37 |
| Usuario 5 | 33 |
| Usuario 6 | 26 |
| Usuario 7 | 24 |
| Usuario 8 | 18 |
| Usuario 9 | 18 |
| Usuario 10 | 14 |

También es posible identificar a aquellos usuarios que más han interactuado con “Me Gusta”, donde el promedio de interacciones de este tipo, sobre los usuarios que han indicado les gusta alguna publicación, es de 2.2 “Me Gusta”. En la Tabla 4 es posible ver la frecuencia de los 10 usuarios con más interacciones del tipo “Me Gusta”.

Tabla 4: Top 10 usuarios con mayor cantidad de “Me Gusta”

| Usuario | N° “Me Gusta” |
|----------|---------------|
| Liker 1 | 1.542 |
| Liker 2 | 1.152 |
| Liker 4 | 1.045 |
| Liker 4 | 880 |
| Liker 5 | 615 |
| Liker 6 | 559 |
| Liker 7 | 511 |
| Liker 8 | 507 |
| Liker 9 | 452 |
| Liker 10 | 443 |

Luego de limpiar el texto de manera apropiada para el algoritmo de VADER se obtuvo que los comentarios son principalmente de polaridad *neutra* con una tendencia *positiva*, mientras que la minoría de los comentarios posee una polaridad *negativos*. La Tabla 5 muestra el resumen de la clasificación de los comentarios de acuerdo a VADER.

Tabla 5: Polaridad de Comentarios

| Calificación | Cantidad | Porcentaje |
|--------------------|----------|------------|
| Precisión Positiva | 2.209 | 36.1 % |
| Precisión Negativa | 405 | 6.6 % |
| Precisión Neutra | 3.502 | 57.3 % |
| Total | 6.116 | 100 % |

Con ello, es posible identificar que el algoritmo en permite procesar distintos tipos de texto, jergas y símbolos. Sin embargo la polaridad de algunos texto no corresponde a la connotación que tiene dentro de la comunidad. La Tabla 6 muestra 15 de los principales *emoticonos* y el puntaje que entrega sobre ellos el algoritmo de VADER. A modo de ejemplo, si tomamos el *emoticono* que representa un corazón, el puntaje indica que es un emoticono neutral, sin embargo para el equipo de ZoooSalud y su comunidad tiene una connotación totalmente positiva.

Tabla 6: Polaridad principales *emoticonos*

| Emoticono | Negativo | Neutro | Positivo | Compuesto |
|-----------|----------|--------|----------|-----------|
| 😊 | 0,000 | 0,500 | 0,500 | 0,459 |
| ❤️ | 0,000 | 1,000 | 0,000 | 0,000 |
| 😄 | 0,218 | 0,345 | 0,437 | 0,440 |
| 😁 | 0,000 | 0,353 | 0,647 | 0,671 |
| 😂 | 0,000 | 0,333 | 0,667 | 0,718 |
| 👏 | 0,000 | 1,000 | 0,000 | 0,000 |
| 💕 | 0,000 | 1,000 | 0,000 | 0,000 |
| 🐾 | 0,000 | 1,000 | 0,000 | 0,000 |
| 👩 | 0,000 | 1,000 | 0,000 | 0,000 |
| 😬 | 0,000 | 0,417 | 0,583 | 0,422 |
| 😞 | 0,756 | 0,244 | 0,000 | -0,477 |
| 👍 | 0,000 | 1,000 | 0,000 | 0,000 |
| 😈 | 0,000 | 0,571 | 0,429 | 0,459 |
| ☀️ | 0,000 | 1,000 | 0,000 | 0,000 |
| 👉 | 0,000 | 0,000 | 1,000 | 0,730 |

Otro aspecto a considerar es que texto con polaridad negativa no significan necesariamente reacciones negativas al contenido. Al revisar las publicaciones con comentarios que expresan sentimientos como *tristeza* o *molestia* se identifica que tienen relación a alguna temática que va en contra del cuidado de los animales, por lo que, una reacción de polaridad negativa podría ser en esta comunidad una reacción que demuestra compromiso con los temas tratados y, por lo tanto, no sería considerado como algo negativo o perjudicial para la imagen de marca.

5. Discusión y conclusiones

De acuerdo con las tendencias de mercado ligados a productos y servicios para animales domésticos y la penetración de Internet en Chile, las actividades relacionadas a los animales domésticos muestran ser un tema de gran interés en la sociedad actual.

Para el caso particular de ZoooSalud el análisis de comportamiento de sus usuarios permite estimar el *compromiso* (engagement) con la marca, identificar los usuarios más leales y analizar la proyección de la marca. Esto de gran utilidad para generar vínculos con empresas o fundaciones que se puedan ver beneficiadas por una comunidad robusta de amantes de los animales dispuesta a participar de actividades o elegir sus productos y servicios.

El análisis de sentimientos sobre los comentarios permite evaluar qué tipo lenguaje predomina dentro de la comunidad y cómo relacionarse con ella de manera apropiada, además de evaluar qué reacción genera cada tipo de contenido.

El modelo utilizado VADER para análisis de sentimientos muestra ser apropiado para este tipo de comunidad donde prevalece el usos de emoticones por sobre el texto y donde, además, existen respuestas en diferentes idiomas (inglés y español).

Analizar el significado del lenguaje a través de símbolos y si su polaridad corresponde con la denotación que la comunidad le entrega es complejo y necesario para un buen análisis. Muchos símbolos que eran considerado como neutrales por el modelo, en la práctica podrían ser considerados como positivos dado el tipo de contenido que se publica. Además, las reacciones negativas o positivas no deben ser analizadas con un juicio de valor directo, es decir, una reacción identificada positiva no necesariamente se puede atribuir como una respuesta beneficiosa para la comunidad, del mismo modo que una opinión de polaridad negativa puede ser algo beneficioso. Por ejemplo, si el conteni-

do habla sobre abandono de mascotas es esperable que las reacciones a ese contenido tengan polaridad negativa.

Los resultados obtenidos, desde la obtención de los datos para evaluar el dinamismo de la cuenta como el análisis de los comentarios mediante Análisis de Sentimientos, permitieron entender mejor a la comunidad y dar paso a discusiones estratégicas sobre la línea editorial del sitio, tipo o variedad de contenido y buscar estrategias para aumentar las interacciones con las publicaciones por un lado y la cantidad de seguidores por otro.

Con todo, el uso de redes sociales para organizaciones se muestra como un espacio muy enriquecedor para comprender a una audiencia y posicionar una marca. Las herramientas de análisis de redes sociales además permiten monitorear de manera permanente la relación entre los usuarios interesados y los productos o servicios que se van ofertando.

El trabajo permitió abrir una serie de discusiones y queda como desafío seguir explorando otro tipo de modelos del mundo del Análisis de las Redes Sociales para potenciar la marca y tener una comprensión en profundidad de los usuarios vinculados a ella.

Agradecimientos: A la Escuela de Ingeniería Comercial (Santiago) de la Universidad de Valparaíso por haber apoyado el desarrollo del estudio y al equipo de ZoooSalud por contribuir con los datos de sus publicaciones y análisis estratégico.

Referencias

- [1] C. C. Aggarwal y C. Zhai. *Mining text data*. Springer Science & Business Media, 2012.
- [2] C. T. Butts et al. Social network analysis with SNA. *Journal of statistical software*, 24(6):1–51, 2008.
- [3] CADEM. El Chile que viene. Reporte técnico, Mayo 2019. Recuperado de <https://www.cadem.cl/encuestas/el-chile-que-viene-mascotas/>.
- [4] P. J. Carrington, J. Scott, y S. Wasserman. *Models and methods in social network analysis*, volume 28. Cambridge university press, 2005.
- [5] M. E. P. Cázares. La producción del conocimiento. *Enlce: Revista Venezolana de Información, Tecnología y Conocimiento*, 10(1):21–30, 2013.
- [6] J. Celaya. *La empresa en la Web 2.0*. Ediciones Gestión 2000, 2000.

- [7] A. Chevallier. *Strategic thinking in complex problem solving*. Oxford University Press, 2016.
- [8] J. Clement. Global digital population as of July 2019 (in millions) - statista. Reporte técnico, Septiembre 2019. Recuperado de <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
- [9] J. Clement. Most famous social network sites worldwide as of July 2019. Reporte técnico, Septiembre 2019. Recuperado de <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
- [10] T. Coffman, S. Greenblatt, y S. Marcus. Graph-based technologies for intelligence analysis. *Communications of the ACM*, 47(3):45–47, 2004.
- [11] COMSCORE. Jerarquía de necesidades móviles. evolución de móvil como la herramienta principal del hombre digital. Reporte técnico, 2017. Recuperado de <https://www.comscore.com/lat/layout/set/popup/Request/Presentations/2017/Jerarquia-de-Necesidades-Moviles?>
- [12] G. Edwards. Mixed-method approaches to social network analysis. 2010.
- [13] N. B. Ellison, J. L. Gibbs, y M. S. Weber. The use of enterprise social network sites for knowledge sharing in distributed organizations: The role of organizational affordances. *American Behavioral Scientist*, 59(1):103–123, 2015.
- [14] Euromonitor. Pet care in Chile. Reporte técnico, 2018. Recuperado de <https://www.euromonitor.com/pet-care-in-chile/report>.
- [15] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, et al. *Advances in knowledge discovery and data mining*, volume 21. AAAI press Menlo Park, 1996.
- [16] K. Hampton, L. S. Goulet, L. Rainie, y K. Purcell. Social networking sites and our lives. *Pew Internet & American Life Project*, 16:1–85, 2011.
- [17] C. J. Hutto y E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth international AAAI conference on weblogs and social media*, 2014.
- [18] IProspect. Comercio basado en datos: Cómo ganar la carrera del comercio en la nueva economía digital. Reporte técnico, 2019. Recuperado de <https://www.iprospect.com/es/ar/insights/whitepapers/data-driven-commerce/>.

- [19] C. Kruegel, R. McCoy, H. Sherman, y A. Daw. System for and method of analyzing and responding to user generated content. Reporte técnico, Noviembre 2013. US Patent App. 13/618,072.
- [20] E. Lee, J.-A. Lee, J. H. Moon, y Y. Sung. Pictures speak louder than words: Motivations for using instagram. *Cyberpsychology, Behavior, and Social Networking*, 18(9):552–556, 2015.
- [21] B. Liu et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666, 2010.
- [22] C. Lozares Colina. La teoría de redes sociales. *Papers: revista de sociología*, (48):103–126, 1996.
- [23] P. V. Marsden y N. Lin. *Social structure and network analysis*. Sage Beverly Hills, 1982.
- [24] J. Miles. *Instagram power*. McGraw-Hill Publishing, 2013.
- [25] J. e. a. Morley. Method and apparatus for authenticating a magnetic fingerprint signal using an adaptive analog to digital converter. Reporte Técnico Patent N: US 7,210,627 B2, May 2007.
- [26] N. Nic, R. Fletcher, A. Kalogeropoulos, D. A. Levy, y R. K. Nielsen. *Reuters Institute Digital News Report 2018*. Reuters Institute for the Study of Journalism, 2018.
- [27] B. Pang, L. Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.
- [28] P. R. Prashanth. Fine-grained sentiment analysis in python (part 1). Reporte técnico, Septiembre 2019. Recuperado de <https://towardsdatascience.com/fine-grained-sentiment-analysis-in-python-part-1-2697bb111ed4>.
- [29] K. Pte. Ltd., digital 2019: Global digital overview, january 2019. Reporte técnico.
- [30] J. Ramos. *Instagram para empresas*. XinXii, 2015.
- [31] J. Scott. Social network analysis. *Sociology*, 22(1):109–127, 1988.
- [32] J. Scott. Social network analysis, overview of. *Computational Complexity: Theory, Techniques, and Applications*, páginas 2898–2911, 2012.

- [33] K. G. Smith y W. B. George. Social media contributor weight. Reporte técnico, Septiembre 2016. US Patent 9,442,984.
- [34] N. Spivack y D. Ter Heide. Systems and methods for recommending advertisement placement based on in network and cross network online activity analysis. Reporte técnico, Noviembre 2012. US Patent App. 13/403,962.
- [35] C. Vergara. Health topics with data science: Exploratory research with social network analysis. En E. Rodríguez, editor, *Data Analytics Applications in Latin America and Emerging Economies*, 1, capítulo 13. Auerbach Publications, 1 edition, Julio 2017.

OPTIMIZACIÓN DEL BARRIDO Y RECOLECCIÓN DE HOJAS EN LA CIUDAD ARGENTINA DE TRENQUE LAUQUEN

DIEGO DELLE DONNE^{*}
VALERIA DI TOMASO^{**}
GUILLERMO DURÁN^{***}

Resumen

En el presente trabajo se propone un método para asignar eficientemente las rutas de los barrenderos en la ciudad argentina de Trenque Lauquen, mediante programación lineal entera. La solución además tiene en cuenta la determinación de puntos fijos para el depósito de los residuos generados por el barrido, resuelto también mediante programación lineal entera (PLE); y el ruteo de camiones que recolectan a los mismos, modelado como un traveling salesman problem (TSP) asimétrico y resuelto con el software Concorde. Anteriormente al presente trabajo, el municipio realizaba estas tareas mediante una programación manual que, a juicio de los responsables de las mismas, era poco eficiente, dejando en muchas oportunidades la ciudad plagada de hojas y no permitiendo estimar con precisión los operarios que se necesitan en cada zona de la ciudad. La implementación aquí expuesta optimiza la distribución de barrenderos en los distintos sectores, permitiendo mantener limpia la ciudad, y determinando qué cantidad de operarios es necesario contratar. Además, logra reducir en un 53.75 % la cantidad de puntos de depósito de montículos, así como también disminuye la longitud de las rutas de los camiones recolectores en un 12 %. Esto se traduce en ahorro económico de combustible, así como también en la cantidad de tiempo que insume realizar la tarea.

Palabras Clave: Recolección de residuos, Ruteo de vehículos, Programación Matemática, Problema del viajero de comercio.

^{*}Instituto de Ciencias de la Universidad Nacional de General Sarmiento

^{**}Instituto de Cálculo FCEyN-UBA, Argentina

^{***}Departamento de Matemática e Instituto de Cálculo FCEyN-UBA y CONICET, Argentina, Departamento de Ingeniería Industrial, FCFM-Universidad de Chile, Chile

1. Introducción y descripción del problema

Trenque Lauquen es una ciudad ubicada 445 km al oeste de la ciudad de Buenos Aires, la capital de la Argentina. Según el Censo Poblacional Nacional de 2010, Trenque Lauquen contaba en ese entonces con un total de 33.442 habitantes distribuidos en sus 400 hectáreas de superficie.

Una característica distintiva de la ciudad es que la mayoría de sus arterias cuentan con amplias *ramblas* o boulevards plantadas con diversas especies de árboles. En total pueden contabilizarse 616 de estas ramblas, constituyendo una ciudad que tiene una alta proporción de espacio verde por habitante. Esta característica obliga a mantener un intenso barrido de las calles de la ciudad para evitar principalmente que el follaje caído se acumule en los cordones y tapone los desagües. Idealmente, cada calle debería barrerse con una cierta frecuencia determinada (e.g., todos los días, día por medio, etc.) según la parte de la ciudad en la que se encuentre. A este respecto, existen dos tipos de barrido. El barrido *mecánico* realizado con vehículos barredores queda reservado para los cordones de ramblas, ya que al no haber vehículos estacionados se puede circular con mayor facilidad que sobre los cordones de las aceras. Este tipo de barrido no será parte de este estudio. Por otro lado, el barrido *manual*, realizado por operarios con herramientas simples, se realiza en el resto de los cordones de la ciudad; aproximadamente unas 1800 cuadras. Al momento de comenzar nuestro trabajo, esta tarea era realizada por 84 empleados distribuidos en 3 zonas. Cada una de estas zonas era supervisada por un capataz que determinaba, muchas veces a demanda, qué cuadras debía barrer cada operario cada día. El número de operarios en la zona 1 era 21, quienes barrían un total de 629 cuadras; mientras que la zona 2 contaba con 35 operarios y 678 cuadras, y la zona 3, con 28 operarios y 507 cuadras. La Figura 1 ilustra la división de la ciudad en las 3 zonas mencionadas. La composición de las mismas está definida por criterios geográficos y administrativos y no será parte de reformulación en este estudio. Se muestran también las 4 *bases* desde donde los operarios comienzan sus recorridos de barrido.

Anteriormente al presente trabajo, existía una asignación manual de los barrenderos a la zonas y el recorrido de cada barrendero era definido por él mismo “a ojo”. Mientras que las hojas eran colocadas en *montículos* en todas las intersecciones de cuadras. Posteriormente, durante el mismo rango horario, camiones de recolección se encargaban de recoger los montículos y llevar lo recolectado a un basural ubicado 7 kilómetros al oeste de la ciudad.

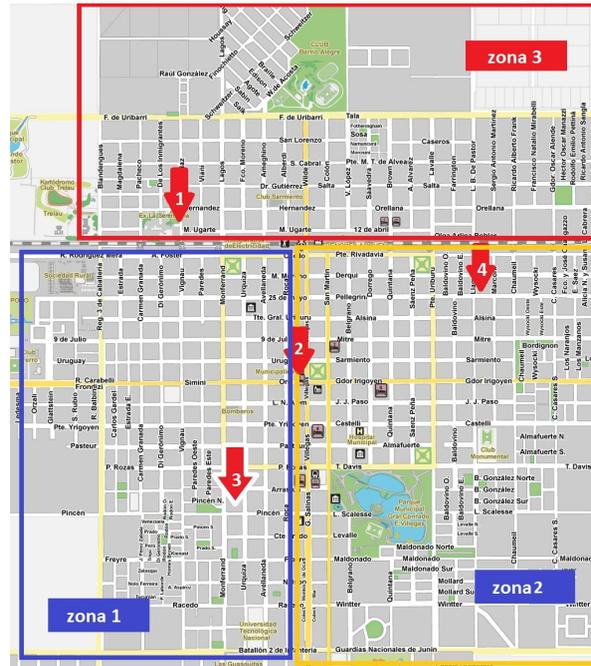


Figura 1: División de la ciudad en tres zonas y ubicación de las cuatro bases de inicio de recorridos.

Dicha tarea era llevada a cabo por dos camiones por zona, cada uno con dos operarios distintos, además del conductor del camión.

Entre algunos de los inconvenientes que se presentaban antes del desarrollo del presente trabajo se pueden mencionar los siguientes:

- **Incumplimiento de los recorridos preestablecidos:** Muchas de las calles no se barrían con la frecuencia mínima esperada. Esto sugería una posible falta de personal aunque también podría ser causa de una administración subóptima del personal existente.
- **Falta de un recorrido bien definido:** Debido al punto anterior, los recorridos se determinaban a demanda cada día, lo cual hacía complicada la asignación de recorridos y conllevaba a otros problemas (e.g., desbalanceo en las cargas de trabajo de los operarios).
- **Mala distribución de operarios por zona:** En la práctica la cantidad de cuadras barridas por operario cada día debía estar en promedio entre los 20 y 24 cordones. Sin embargo, la proporcionalidad de la asignación manual era difícil de explicar: mientras en la Zona 1 la asignación llevaba a que cada operario tuviera que barrer en promedio casi 30 cuadras, en

las Zonas 2 y 3 ese promedio bajaba a alrededor de 20 cuadras por operario por día.

- **Falta de coordinación entre los camiones recolectores y los barrenos:** Principalmente debido a la superposición de horarios de trabajo, lo cual impide una organización secuencial de estas tareas. En la práctica solía ocurrir que los camiones pasaban por algunas esquinas antes de que los montículos de hojas fueran depositados.

Teniendo en cuenta los problemas descritos, los objetivos del presente trabajo fueron los siguientes:

1. Evaluar si la cantidad de operarios actual es la adecuada (objetivo principal de la Municipalidad de Trenque Lauquen). Generar para ello planes de barrido y asignación de recorridos factibles que cumplan con las frecuencias deseadas de barrido para cada calle. Además, que los recorridos generados cumplan con un cierto nivel de *elegancia* o simpleza (lo cual veremos en detalle en la Sección 3.1).
2. Determinar las esquinas dónde los operarios depositarían los montículos, donde el municipio instalaría contenedores con tal fin. La cantidad de esquinas debe ser suficiente para que los operarios puedan descargar frecuentemente lo recogido, pero idealmente no demasiadas para lograr recorridos eficientes de los camiones recolectores.
3. Definir rutas eficientes para los camiones recolectores de forma tal de optimizar su uso y tiempo de trabajo.

Algunas consideraciones preliminares importantes son las siguientes:

- **Unidad de trabajo:** Para disminuir el tamaño del problema se decidió establecer a la manzana (un bloque de 4 cuadras) como la unidad de trabajo.
- **Tiempo de barrido por manzana:** Por sugerencia de la municipalidad establecimos el tiempo de barrido en 12 minutos cada 100 mts. Con esta definición, el tiempo de barrido estimado para una manzana se calculó en proporción a su longitud (notar que no necesariamente todas las cuadras miden exactamente 100 mts.). Una manzana usual, de 4 cuadras de 100 mts. cada una, se barrerá entonces en 48 minutos.
- **Horario de trabajo diario de cada barrenos:** La jornada de trabajo de cada barrenos es de 6hs, de lunes a viernes. Dado que se pretende

una asignación pareja de manzanas por barrendero y que se busca que todas las manzanas asignadas puedan ser barridas a lo largo de la jornada de trabajo, establecimos que la carga de trabajo asignada a cada barrendero debe estar entre 5 y 6 hs diarias.

- **Frecuencia de barrido:** Debido a distintas características geográficas, como la cantidad de árboles, el hecho de si una calle es una avenida, o la cercanía a la frontera de la ciudad, todas cuestiones que generan mayor suciedad, algunas cuadras necesitan un barrido más frecuente. Luego de discutirlo con la municipalidad, se decidió dividir el conjunto de manzanas de la ciudad en dos categorías: manzanas de barrido diario y manzanas de barrido día por medio. Con respecto a la categoría de manzanas que se barren día por medio, para la resolución del modelo el grupo de manzanas se dividió en dos mitades, y se determinó un ciclo de barrido alternado, a completar en dos semanas. Es decir, si por ejemplo consideramos una manzana de la primera mitad, en la primera semana del ciclo la misma se barre los días lunes, miércoles y viernes, y en la segunda semana, martes y jueves; mientras que una manzana perteneciente a la otra mitad, se barre en la primera semana martes y jueves, y a la semana siguiente lunes, miércoles y viernes.

El manuscrito está organizado de la siguiente manera. En la próxima sección se presenta una breve revisión bibliográfica. En la Sección 3 se exhiben las estrategias de resolución de cada una de las etapas del proyecto. En la Sección 4 se muestran los resultados, mientras que en la sección final se destacan las conclusiones y el posible trabajo futuro.

2. Revisión bibliográfica

Entre la literatura dedicada al tema (y a temas similares), se pueden mencionar los trabajos que veremos a continuación. Algunos de ellos comparten ciertas similitudes con nuestro problema, por lo que constituyeron un gran aporte para la concreción del proyecto.

En [9] se describe una implementación de recorridos de barredoras del área rural de Lancashire, Inglaterra. Allí cada distrito cuenta con un vehículo que comienza su recorrido desde un depósito. Una de las características del problema es la capacidad limitada de cada barredora, lo cual lo convierte en una versión NP-difícil del conocido problema del Cartero Chino (o *Chinese Postman*, en inglés). A diferencia de otros trabajos, al ser un área rural, tiene

múltiples puntos para vaciar su contenido; las calles pueden considerarse de doble sentido porque el poco tráfico permite que las barredoras violen las reglas de tránsito, y no existen restricciones en cuanto a momentos de barrido disponibles. Esto último sí ocurre en algunas ciudades urbanas donde hay horarios de estacionamiento, que impiden que las barredoras realicen su trabajo en esas franjas temporales. Para resolver el problema, se utiliza un algoritmo heurístico que define las rutas de barrido que minimizan la distancia recorrida sin barrer, así también como asegurando que los dos lados de cada calle se barran el mismo día.

En [3] se desarrolla un sistema para múltiples barredoras. El trabajo está centrado en las ciudades de New York y Washington D.C. A diferencia de [9], se enfocan en un área urbana con mayoría de calles de sentido único y múltiples restricciones de estacionamiento. Estas últimas son tratadas como ventanas de tiempo. Los resultados empíricos obtenidos en este trabajo, obtienen mejores resultados al determinar un recorrido y luego dividirlo para cada barrendero que al realizar la inversa, ya que de esta manera se obtiene un ahorro mayor en cantidad de kilómetros recorridos.

En [6] se desarrollan técnicas de programación lineal entera (PLE) para la optimización de la recolección de residuos reciclables en el municipio de Morón. Allí se resuelve el problema del cartero chino en un grafo mixto, pero con restricciones adicionales de tránsito, dado que dicha recolección se realiza con un camión. Dado la complejidad que representa el hecho de ser un grafo mixto, también se implementan algoritmos de aceleración de búsqueda, que unen los subciclos generados en la solución del modelo. Junto con esto, también se utiliza un algoritmo de búsqueda tabú que mejora la distribución del área de recolección de cada uno de los camiones.

En [2] se estudian algoritmos heurísticos de zonificación para la recolección de residuos, también para camiones. Allí, además de fijar el área que recorre cada camión, a través de la identificación de ciclos, se determina cómo debe ser recorrida cada área de manera tal que se respeten las reglas de tránsito y se minimice el desgaste del vehículo.

Finalmente, en [5] se analiza la optimización en la recolección de residuos en contenedores de la zona sur de la Ciudad de Buenos Aires. El problema se reduce al clásico TSP y se utiliza el software CONCORDE [7] para resolverlo. Aprovechando la multiplicidad de soluciones que propone el software, se reduce el desgaste producido por los camiones, medido como el concepto físico de trabajo mecánico. Exceptuando la minimización del trabajo de los camiones, este último problema se asemeja bastante a la tercera etapa del problema estudiado en este trabajo.

3. Estrategia de resolución

En un primer momento se intentó resolver el problema completo planteando un modelo de PLE. Lamentablemente, el tamaño del modelo en cuanto a cantidad de variables y restricciones se tornó irresoluble en la práctica.

La estrategia de resolución finalmente implementada consistió entonces en dividir al problema en tres etapas, cada una implementada sobre cada una de las tres zonas en las que se divide la ciudad. Como expresamos en la descripción del problema, por una cuestión operativa sugerida por la Municipalidad se decidió no modificar el armado de dichas zonas. La primera etapa determina qué manzanas barrerá cada uno de los operarios. Denominamos *segmento*, al conjunto de manzanas contiguas asignadas a un mismo barrendero. Cada segmento tiene que poder ser barrido en su totalidad en la jornada diaria de un barrendero. En la segunda etapa, se determina en qué orden se recorrerá cada una de las cuadras del segmento asignado a cada barrendero y cuáles serán las esquinas predeterminadas para dejar los montículos de hojas, lugar donde el municipio instalará contenedores a tal fin. Por último, en la tercera etapa, se diagraman las rutas de los camiones recolectores.

3.1. Etapa I: asignación de manzanas a barrenderos

El objetivo de esta etapa es asignar un *segmento* (i.e., un conjunto de manzanas contiguas) a cada uno de los operarios de forma tal de cubrir todas las manzanas a barrer en cada zona. Además, se pretende que la distribución del trabajo sea balanceada entre los operarios (i.e., que la carga de cada uno esté dentro de los límites preestablecidos, o sea que pueda realizarse en un tiempo que esté entre las 5 y 6 horas), y que los segmentos utilizados respeten una cierta simpleza o elegancia, esto último para evitar que los operarios se desorienten (lo cual llevaría a pérdidas de eficiencia y/o barridos incompletos).

Para resolver esta etapa del problema, se tomó como base el trabajo realizado en [4], en el cual se resuelve el problema de la asignación de viviendas a censistas para el Censo Poblacional Nacional 2010 en la Argentina. El trabajo propone un modelo de PLE para este problema y, dado que este modelo utiliza una cantidad exponencial de variables, se propone resolver el mismo iterativamente utilizando subconjuntos de estas variables, hasta obtener una solución. Estos subconjuntos son generados por un algoritmo que va construyendo los segmentos en forma incremental en cuanto al tamaño. Si bien ambos problemas comparten la misma esencia, hay diferencias principalmente en cuanto a

la función objetivo a optimizar y a la manera de generar los segmentos.

Sea M el conjunto de manzanas y \mathcal{S} el conjunto de posibles segmentos (explicaremos en la siguiente subsección cómo se generan), el modelo de PLE adaptado a nuestro problema utiliza una variable binaria x_s para cada segmento $s \in \mathcal{S}$ que indica si s se asignará a un operario o no. Con estas definiciones, el modelo propuesto es:

$$\min \sum_{s \in \mathcal{S}} val_s \cdot x_s \quad (1)$$

$$\sum_{\substack{s \in \mathcal{S} \\ m \in s}} x_s = 1 \quad \forall m \in M \quad (2)$$

$$\sum_{s \in \mathcal{S}} x_s \leq b_{max} \quad (3)$$

$$x_s \in \{0, 1\} \quad (4)$$

donde b_{max} es una cota superior para la cantidad de operarios disponibles (la función objetivo tenderá a encontrar un valor cercano al mínimo necesario para poder cubrir toda la zona) y val_s es el coeficiente de valuación para el segmento s .

En este coeficiente se debe capturar la simpleza del segmento (con valores más bajos para segmentos más simples o elegantes) y para ello se analizaron dos criterios distintos de calidad de un segmento dado, utilizando ideas desarrolladas para medir la compacidad de distritos electorales (ver por ejemplo [12] y [14]).

- **Valuación 1: *Compacidad*** En esta valuación intentamos reflejar la compacidad del segmento, intentando así evitar segmentos dispersos o de formas complejas. La medición está inspirada en el momento de inercia de un distrito (en nuestro caso el distrito sería el segmento), expuesto en [13]. Allí, para cada uno de los tramos censales (aquí el equivalente sería una manzana), se calcula el cuadrado de la distancia entre el centro de gravedad del distrito y el centro de gravedad de cada tramo censal, multiplicado por la cantidad de población del mismo. La suma de todos estos números es el centro de inercia del distrito.

En nuestro caso se calcula el centro de gravedad del segmento completo y de cada una de las manzanas del mismo. Definimos la valuación de un segmento s como la distancia máxima entre el centro de gravedad de s y el de una manzana de s . Es decir, $val_s = \max_{m \in s} \{dist(centro(s), centro(m))\}$.

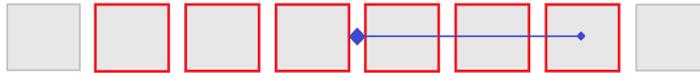
A continuación se tienen dos segmentos con la misma cantidad de manzanas. La valuación de cada uno está dada por la longitud del segmento

más largo y puede observarse que el segmento de la derecha tiene una mejor (i.e., menor) valuación.



■ **Valuación 2: Convexidad**

Si bien la valuación anterior privilegia algunos segmentos compactos, la misma perjudica a segmentos que también representan opciones muy interesantes por su simpleza, como es el caso de una hilera de manzanas:



Como alternativa a este problema, proponemos una segunda valuación que calcula el área real del segmento $area(s)$ y la compara con el área de su cápsula convexa $area(conv(s))$. Dado que los segmentos se forman por manzanas contiguas, mientras más convexo sea un segmento, mayor será su simpleza. La comparación entre ambas áreas se midió de dos formas distintas:

1. Valuación 2a: $val_s = area(conv(s)) - area(s)$
2. Valuación 2b: $val_s = \frac{area(conv(s))}{area(s)}$

Dada que $S \subseteq conv(S) \forall$ conjunto S , la primera opción es siempre mayor o igual que 0 y alcanza el óptimo en ese valor, indicando que el segmento es convexo. En la segunda opción ocurre esto mismo pero con el valor 1.

En nuestra implementación, el cálculo de la cápsula convexa se realiza mediante el algoritmo de Jarvis March y de la fórmula del determinante de Gauss para calcular el área. En [8] y [1] se pueden encontrar detalles sobre los mismos, respectivamente.

A continuación se puede ver el mismo ejemplo visto en la primera valuación. El segmento de la derecha en este caso es óptimo, mientras que en el de la izquierda, se puede ver una diferencia entre las dos áreas, lo cual impacta directamente en la simpleza del segmento.



3.1.1. Resolución heurística del modelo de PLE

Al igual que lo ocurrido en [4], la cantidad de variables del modelo de PLE es demasiado elevada como para resolverlo directamente (incluso generar todos los segmentos factibles resulta impracticable en instancias reales). Adaptando las ideas desarrolladas en [4], proponemos un algoritmo que genera incrementalmente subconjuntos de segmentos de \mathcal{S} cada vez más grandes e iterativamente ejecuta el modelo de PLE restringido a estos subconjuntos de variables. El algoritmo finaliza en cuanto se encuentre una solución óptima del modelo restringido. Vale notar que si bien la solución hallada es óptima para el modelo restringido, ésta podría no serlo para el modelo general utilizando todos los segmentos posibles de \mathcal{S} . Es decir, este algoritmo plantea una resolución heurística del modelo.

El Algoritmo 1 ilustra el procedimiento implementado para la generación de los segmentos en cada iteración. Comenzando desde un conjunto *base* B_1 de segmentos formados por una sola manzana, en cada iteración se pretende extender cada segmento previamente generado agregándole alguna de las manzanas adyacentes al mismo. Esto se realiza exhaustivamente, es decir, si un segmento s tiene k manzanas adyacentes, se obtienen entonces k segmentos nuevos a partir de s , uno por cada manzana adyacente. Luego de haber aplicado esto a todos los segmentos previamente generados, se ejecuta el modelo de PLE con todos los segmentos generados que sean factibles (i.e., que requieran entre un mínimo y un máximo tiempo de barrido, que como fue dicho, lo establecimos entre 5 y 6 horas).

Con respecto a la generación de segmentos, dado que la cantidad de ellos era muy grande, se impuso una restricción al momento de aceptar un segmento factible, descartando aquellos cuyo cociente entre el área de la cápsula convexa y el área real del segmento era mayor a 1,5 (es decir que el descarte de segmentos se realiza a través de la Valuación 2b). Esto nos permite reducir el número de segmentos de manera considerable; retomaremos este punto en la sección de Resultados.

Algoritmo 1 Generación incremental de segmentos

```

1:  $B_1 = M$  (conjunto de manzanas)
2:  $i = 2$ 
3: mientras no se encontró solución hacer
4:    $B_i = B_{i-1}$ 
5:   para  $s \in B_{i-1}$  hacer
6:     para  $m \in \{\text{manzanas adyacentes a } s\}$  hacer
7:        $B_i = B_i \cup \{s + m\}$ 
8:     fin para
9:   fin para
10:  Resolver modelo de PLE con los segmentos válidos de  $B_i$ 
11: fin mientras
12: Devolver la solución hallada

```

3.2. Etapa II: Recorridos de los barrenderos y ubicación de montículos

Habiendo determinado ya las manzanas asignadas a cada operario, esta segunda etapa se enfoca en la definición de los recorridos para cada uno de ellos, en conjunto con la determinación de las esquinas en donde se ubicarán los contenedores para depositar allí los montículos de hojas.

Inicialmente se diseñó un modelo de PLE para determinar en simultáneo tanto los recorridos y las esquinas donde se instalará un contenedor, ya que hay una gran dependencia entre estas decisiones. Lamentablemente, el tamaño del modelo planteado hizo imposible una resolución práctica del mismo, debido al gran número de variables. Finalmente se optó por una segunda estrategia en la cual se determina primero el orden de recorrido de las cuadras para cada barrendero y *a posteriori* se decide en qué esquinas ubicar contenedores.

3.2.1. Definición de los recorridos

El hecho de definir los recorridos por un lado y determinar la ubicación de los contenedores por otro, permite además definir el recorrido de cada barrendero independientemente del resto, lo cual también reduce los tiempos de cómputo significativamente .

Para cada operario, el recorrido se define hallando un *circuito Euleriano* en el grafo asociado al conjunto de manzanas que dicho operario tiene asignado. En primera instancia se consigue un *grafo Euleriano* resolviendo el Problema del Cartero Chino por medio del clásico modelo de PLE para tal problema, que puede encontrarse en [10]. Este método consiste en agregar el menor número de aristas para transformar el grafo en Euleriano. Finalmente, mediante la

utilización del algoritmo de Hierholzer [11], se determina en tiempo lineal el orden de recorrido de las cuadras dentro del segmento asignado.

3.2.2. Ubicación de los contenedores

Habiendo definido los recorridos para cada barrendero, presentamos a continuación un modelo de PLE para determinar la ubicación de los contenedores, con el objetivo de minimizar la cantidad de los mismos en cada zona. Para ello, se decidió realizar una discretización temporal y definir los siguientes parámetros:

- T : Conjunto de intervalos de tiempo.
- V : Conjunto de esquinas de la zona.
- B : Conjunto de barrenderos de la zona.
- $carga(i, t_1, t_2)$: Cantidad de cuadras barridas por el operario $i \in B$ desde el momento $t_1 \in T$ hasta el momento $t_2 \in T$.
- C_{max} : Cantidad máxima de cuadras que un operario puede barrer antes de necesitar depositar en un contenedor.
- $esquina(i, t)$: Esquina en donde se encuentra el operario $i \in B$ en el tiempo $t \in T$.

El modelo utiliza una variable de decisión binaria x_j para cada esquina $j \in V$ que indica si en j se ubicará un contenedor o no. Además, utiliza una variable binaria w_{ijt} para decidir si el barrendero $i \in B$ deposita en el contenedor $j \in V$ en el tiempo $t \in T$; vale aclarar que la variable w_{ijt} se anula si la esquina j no aparece en el recorrido de i a tiempo t . Con estas definiciones, el modelo es el siguiente:

$$\text{mín} \sum_{j \in V} x_j \quad (5)$$

$$\text{carga}(i, t_1, t_2) \leq C_{max} + C_{max} \sum_{t \in [t_1, t_2]} \sum_{j \in V} w_{ijt} \quad (6)$$

$$\forall i \in B, \forall t_1, t_2 \in T, t_1 < t_2$$

$$w_{ijt} \leq x_j \quad (7)$$

$$\forall j \in V, \forall i \in B, \forall t \in T$$

$$w_{ijt} \in \{0, 1\} \quad (8)$$

$$\forall j \in \{j \in V / \exists i \in B, \exists t \in T, j = \text{esquina}(i, t)\}, \forall i \in B, \forall t \in T$$

$$x_j \in \{0, 1\} \quad (9)$$

$$\forall j \in V, \forall i \in B, \forall t \in T$$

La función objetivo (5) minimiza la cantidad de contenedores a ser instalados. Las restricciones (6) aseguran que un operario dispone de suficientes contenedores en su camino como para nunca sobrepasar la carga máxima, mientras que las restricciones (7) impiden que un operario deposite en una esquina que no fue asignada para contenedores. Finalmente, las restricciones (8) anulan las variables w_{ijt} que no se corresponden con el recorrido de los operarios.

Lamentablemente, el tamaño de este modelo para cada una de las zonas no es resoluble en la práctica. Para superar este inconveniente, se decidió particionar el conjunto de barrenderos, definiendo así subzonas más pequeñas de cada zona, y se aplicó el modelo de PLE independientemente en cada una de estas subzonas para luego unificar las soluciones en una solución global para la zona completa.

Esta decisión puede llevar obviamente a soluciones subóptimas (en la Sección 3.3 analizamos este aspecto con más detalle). Un claro ejemplo, es el de encontrar dos contenedores (de dos ejecuciones distintas del modelo) en esquinas distintas de una misma intersección, lo cual puede entorpecer muchísimo el recorrido de los camiones recolectores en la práctica. Por este motivo, se diseñó una etapa de *postprocesamiento* de la solución para intentar heurísticamente mejorar la solución global obtenida. El algoritmo propuesto recorre cada par de contenedores en la solución y en caso de encontrar dos de ellos que se encuentren a muy poca distancia entre sí, los unifica en uno sólo y actualiza el recorrido de los barrenderos que usen esos contenedores para mantener la coherencia de este cambio. esta actualización puede hacerse siempre de manera bastante directa. El Algoritmo 2 ilustra todo este procedimiento.

Algoritmo 2 Postproceso de unificación de montículos

- 1: $S =$ Conjunto de montículos de la solución
 - 2: **para** $(m_1, m_2) \in S$ **hacer**
 - 3: **si** m_1 y m_2 están muy cerca **entonces**
 - 4: Unificar m_1 y m_2
 - 5: Actualizar las rutas de los barrenderos que utilicen m_1 y/o m_2
 - 6: **fin si**
 - 7: **fin para**
-

3.3. Etapa III: Recorrido de los camiones recolectores

Una vez fijados los puntos de instalación de contenedores, en esta última etapa se determina en qué orden van a ser recolectados los mismos por los camiones de la zona, de manera tal que se reduzca la distancia recorrida. Luego de discutirlo con el personal de la Municipalidad involucrado en el proyecto, y a sugerencia del equipo académico, se decidió que el turno de los camiones quedaría completamente desfasado con el de los barrenderos. Esto simplifica enormemente esta última etapa, ya que no impone restricciones individuales de tiempo para la recogida de cada contenedor.

Naturalmente, al ser camiones y no personas, al problema se le suman las restricciones impuestas por las reglas de tránsito, que impiden que los camiones giren a la izquierda donde está prohibido, realicen giros en U o que se recorra una calle de contramano. Para la primera de las consideraciones, se construyó un grafo sin las aristas que representaban giros prohibidos, identificados a partir de la fórmula matemática que permite calcular el ángulo entre dos segmentos, en este caso dados por dos cuerdas (aristas) adyacentes.

Recordemos que en cada zona se cuenta con dos camiones disponibles para la recolección. Por una cuestión de simplicidad, se separaron los contenedores en dos mitades iguales divididas según la posición geográfica: una mitad dada por los contenedores ubicados más al norte y otra por los contenedores ubicados más al sur.

Las distancias entre un contenedor y otro se calcularon aplicando un algoritmo de *Camino Mínimo* (en nuestra implementación, utilizamos el algoritmo de Dijkstra [8]), contemplando reglas de tránsito. Notar que debido a estas reglas la distancia de ir de un contenedor i a otro j puede ser distinta que la de ir de j a i .

Teniendo las distancias entre cada par de contenedores, se generó un grafo dirigido completo donde los vértices representan los contenedores y cada arista tiene asociado como peso la distancia entre los dos contenedores que conecta. Finalmente, sobre este grafo se resolvió el problema del Viajante de Comercio

Asimétrico (ATSP) utilizando el software CONCORDE [7].

4. Resultados

La exposición de los resultados se realizará separándola en cada una de las etapas. Para el caso de la primera, que se encuentra en proceso de implementación en la ciudad, se mostrarán los resultados de todas las zonas con sus diversas parametrizaciones, mientras que las otras dos etapas se analizarán con un caso en particular, simplemente para mostrar el potencial de mejora que se puede alcanzar una vez que se lleve a la práctica la totalidad de la solución.

4.1. Detalles de implementación

Para la construcción del grafo representando la ciudad, se obtuvo la información desde OpenStreetMap. La misma está dada por el formato .osm, que cuenta con dos estructuras básicas. Por un lado los *nodos*, que pueden representar intersecciones de dos calles (esquinas) o edificios emblemáticos, como escuelas, museos, etc, y por otro lado los *ways*, definidos por un conjunto de nodos, que representan calles, avenidas, caminos, etc. A su vez, los *nodos* cuentan con la posición georreferenciada, lo cual resulta útil a la hora de construir el grafo de la ciudad, pues permite medir la distancia entre dos nodos. En este trabajo se utilizó la librería de Python `imposm.parser`, que permite trabajar con el formato mencionado. Más específicamente se usó el parser (también llamado analizador sintáctico) `OSMparser`.

4.2. Resultados de la Etapa I

Para determinar cuál de las 3 valuaciones sería utilizada en la entrega final se hicieron algunas pruebas preliminares y se les mostró los resultados a los responsables en el municipio. En la Figura 2 se observan los resultados obtenidos por el modelo con las tres valuaciones propuestas, para el fragmento de la Zona 3 correspondiente a las calles que se barren diariamente. Los cuadrados adyacentes pintados con el mismo color corresponden al segmento asignado a cada operario. Como podemos ver, la mayoría de los segmentos de la Valuación 1 (izquierda, abajo) están formados por un conjunto de manzanas que se ajustan bastante a la forma de un cuadrado (aunque incompleto en algunos casos), lo que sugiere que esta valuación logra captar la esencia de lo que queríamos, es decir, tener poca dispersión de los recorridos. Por otro lado, los resultados para las Valuaciones 2a (centro, arriba) y 2b (derecha, abajo) muestran que estas valuaciones tienden a obtener segmentos más “alargados” que

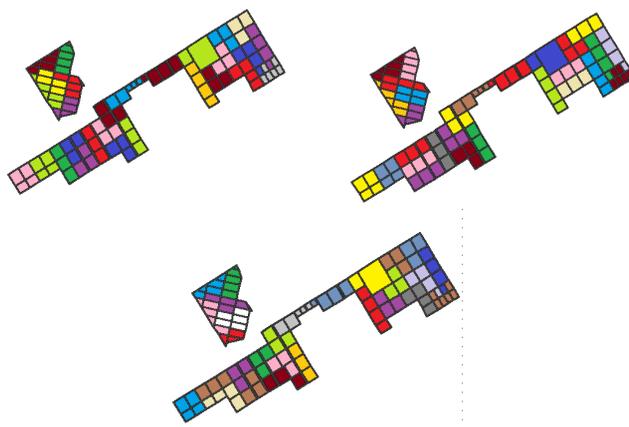


Figura 2: Asignación en un fragmento de la Zona 3 con la Valuación 1 (izquierda, abajo), la Valuación 2a (centro, arriba) y la Valuación 2b (derecha, abajo).

las anteriores. Para la Municipalidad, este tipo de solución era más deseable, ya que les resultaba más fácil monitorear que se cumpla con el barrido. Las soluciones obtenidas por la Valuación 2a tuvieron mayor aprobación de parte de los responsables del barrido en el municipio y por lo tanto, finalmente fue la elegida.

El proceso iterativo descrito en la Sección 3.1.1 se aplica de la siguiente manera: se generan primero todos los segmentos factibles de hasta 5 manzanas (en general es el primer entero donde se dan segmentos factibles). De no encontrar solución, se agregan también los de 6 manzanas. Así hasta conseguir una solución factible (si hace falta se llega hasta los de 9 manzanas, más de 9 no tiene sentido porque ya no se generan nuevos segmentos factibles). En la mayor parte de los casos se pudo encontrar solución después de generar los segmentos de 8 manzanas.

Recordemos que para el descarte de segmentos se tomó como constante de tolerancia de la relación entre el área de la cápsula convexa y el área real del segmento al valor 1,5. De esta forma, la zona 1, con 200 manzanas, pasó de generar alrededor de 803.000 segmentos a tan sólo 225.000. Por su parte, la zona 2, con 253 manzanas pasó de alrededor de 895.000 a 200.000, y finalmente la zona 3, con 166 manzanas, pasó de alrededor de 129.000 segmentos a tener 73.000. Todos estos valores están tomados en promedio a partir de las distintas corridas realizadas para los diferentes escenarios.

Dado que el tiempo de barrido de una cuadra o de una manzana puede ser variable y depender de la velocidad de barrido del barrendero o incluso de

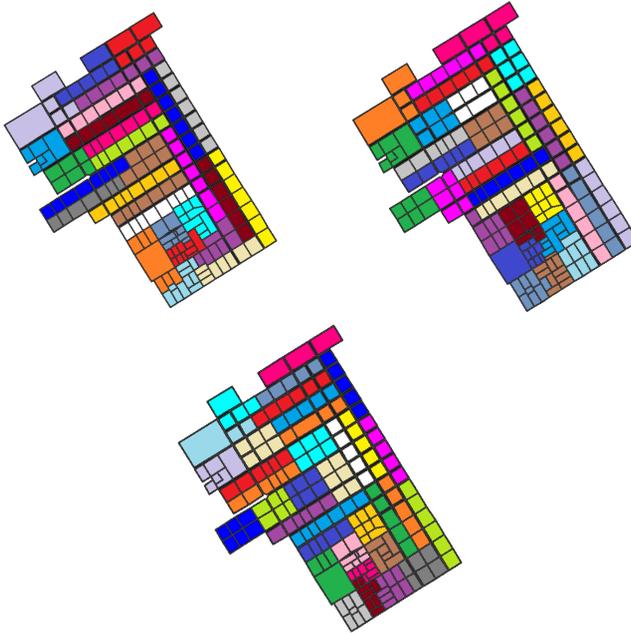


Figura 3: Asignación en la Zona 1 con 28, 31 y 35 barrenderos (izquierda, centro y derecha, respectivamente).

otros factores, decidimos en este caso correr diferentes escenarios, y al escenario “estándar” de 12 minutos cada 100 metros, le agregamos un escenario “rápido” de 10 minutos cada 100 metros y un escenario “lento” de 14 minutos cada 100 metros.

En la Figura 3 se muestran los resultados finales obtenidos para la Zona 1, con 28, 31 y 35 barrenderos (izquierda, centro y derecha de la figura, respectivamente), que representan las soluciones para los escenarios rápido, estándar y lento, respectivamente. Dado que por la formulación del modelo no se puede asegurar que estos valores constituyen el menor número de barrenderos para los cuales existe una solución (podría existir una solución con menos barrenderos que tiene valor peor de la función objetivo), hicimos en todos los casos el test de chequear si dichas soluciones existían (poniendo como b_{max} el entero inmediatamente inferior al valor obtenido). Tanto en las instancias de esta zona como en las zonas siguientes el resultado de este test fue que el problema quedaba infactible, es decir, no se obtenía solución corriendo el modelo con segmentos de hasta 9 manzanas.

A diferencia de la Zona 1, que previo a este trabajo era la más particular por la baja cantidad de barrenderos que tenía en proporción con la cantidad de manzanas a ser barridas; las zonas 2 y 3 se corrieron sólo para 2 instancias, la estándar y la lenta.

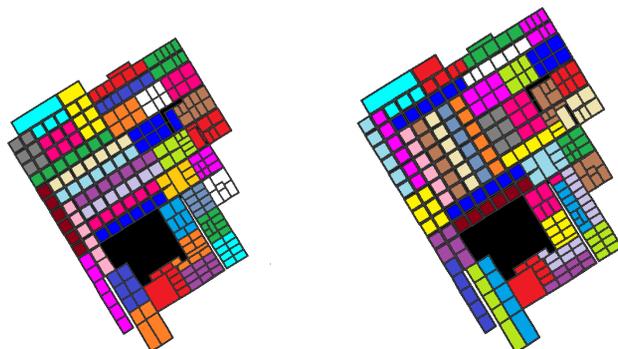


Figura 4: Asignación en la Zona 2 con 36 y 40 barrenderos (izquierda y derecha, respectivamente).

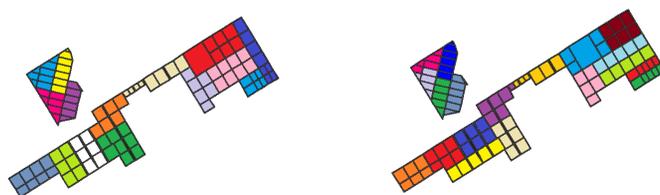


Figura 5: Asignación en las áreas de frecuencia diaria de la Zona 3 con 15 y 19 barrenderos (izquierda y derecha, respectivamente).

En la Figura 4 se observan las soluciones con 36 y 40 barrenderos (estándar y lenta, respectivamente) para la Zona 2. Por otra parte, en la Figura 5 están los resultados para las manzanas de barrido diario en la Zona 3, con 15 y 19 barrenderos (estándar y lenta, respectivamente), mientras que en la Figura 6 están los resultados de las manzanas de barrido día por medio, con 5 y 6 barrenderos (estándar y lenta, respectivamente). Es preciso notar que en esta última figura, la cantidad de segmentos es exactamente el doble de la de barrenderos mencionados y esto se debe a que el esquema de barrido se completa en dos semanas, por lo que un mismo barrendero barre segmentos distintos dependiendo del día de la semana. En este caso, la división de los segmentos en dos grupos se realizó tomando por un lado la mitad de segmentos más al oeste, y por otro los que están más al este. Con estos números, la cantidad de barrenderos total de esa zona es de 20 y 25 (estándar y lenta, respectivamente), según las 2 instancias consideradas.

Un detalle a mencionar, es que en general los conjuntos obtenidos, al ser bastante compactos, son fáciles de identificar para los barrenderos. Este hecho resulta bastante apreciado por parte de los responsables técnicos de Trenque Lauquen.

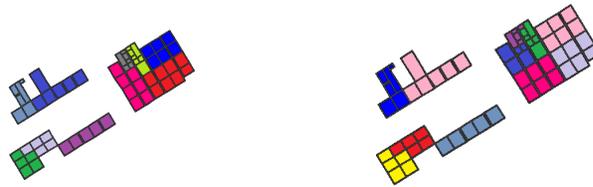


Figura 6: Asignación en las áreas de frecuencia media de la Zona 3 con 5 y 6 barrenderos (izquierda y derecha, respectivamente).

A continuación podemos observar una tabla con algunos resultados numéricos, que indican cómo varía la proporción de barrenderos antes y después del presente trabajo.

Tabla 1: Variación de la proporción de barrenderos

| | Zona 1 | Zona 2 | Zona 3 | Total |
|---------------------------------------|--------|--------|--------|-------|
| Cant. de operarios antes | 21 | 35 | 28 | 84 |
| Cant. de operarios después (promedio) | 31 | 38 | 23 | 92 |
| Variación en la cantidad de operarios | +48 % | + 8 % | -18 % | +9 % |

Como se ve, las estimaciones en la cantidad de operarios por zona no estaban realizadas de manera proporcional a la cantidad de manzanas de las mismas. Mediante estos resultados, se puede llegar a la conclusión de que en la zona 3 este número estaba sobreestimado, mientras que en la zona 1 estaba considerablemente subestimado y en la zona 2 levemente subestimado. Si bien esto se veía a simple vista, lo importante de este modelo, es que permite conocer qué tan alejado se está de un número más adecuado. La estimación promedio indicaría que son necesarios unos 8 barrenderos extra, lo que constituye un 9% más de empleados. Antes de este estudio el municipio cubría las zonas no barridas con horas extra de sus operarios, que son más costosas e ineficientes. A partir de este estudio se ha podido estimar con precisión la cantidad de operarios que hacen falta en cada zona para llegar a barrerla de manera completa y cuántos nuevos operarios tienen que ser incorporados con tal fin.

4.3. Resultados de la Etapa II

Dado que en la actualidad la Etapa I está en pleno proceso de implementación, y por ende aún no se eligió la instancia que mejor se ajusta a la realidad en cada zona, los resultados de esta etapa y de la siguiente se mostrarán solamente con el ejemplo de la imagen central de la Figura 3, que corresponde a

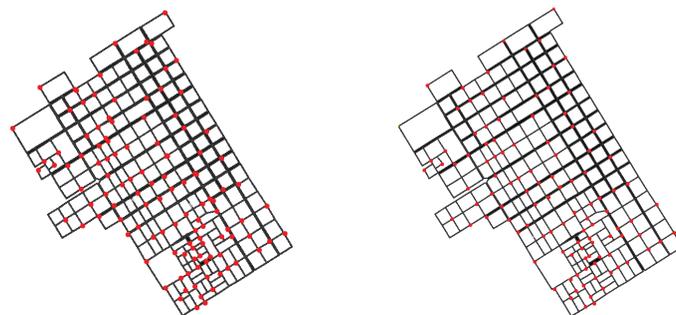


Figura 7: Ubicación de los montículos propuesta por el modelo de PLE (izquierda) y después del postprocesado (derecha) en la Zona 1.

la Zona 1 con 31 barrenderos.

En la imagen izquierda de la Figura 7 se observa cómo quedan determinados los 117 contenedores en la solución obtenida por el modelo de PLE, antes de realizarse el postproceso de unificación de aquellos que están muy cercanos. En la imagen derecha de la misma figura se ilustra la solución final, luego del proceso de unificación de contenedores, en la cual la cantidad final se reduce a 111 (es decir, se redujo en un 4,2%).

Anteriormente al presente trabajo, se depositaban montículos de hojas en cualquiera de las intersecciones, por lo que la tarea de recolección resultaba ser muy ineficiente. Por ejemplo en la Zona 1 había aproximadamente unos 240 montículos diarios a recoger, es decir más del doble que la cantidad obtenida en nuestros resultados.

El modelo planteado demostró ser en general muy eficiente. A modo de ejemplo, la Figura 8 muestra un segmento de 7 manzanas en la zona 2, en el cual sólo se utilizan cuatro contenedores distintos (aunque el barrendero deposita en cada uno de ellos más de una vez). Las cuadras se numeran por el orden que son barridas, los puntos destacados A,B,C y D son las ubicaciones de los contenedores y las cuadras que se recorren dos veces pero sólo se barren la primera vez, aparecen bajo la etiqueta “duplicada”. Una situación bastante similar ocurre en el resto de los recorridos; sólo se utilizan 3 o 4 contenedores por barrendero, en los cuales se deposita múltiples veces.

4.4. Resultados de la Etapa III

La Figura 9 muestra los resultados obtenidos para los recorridos de los dos camiones de la Zona 1. La imagen izquierda muestra el recorrido de la parte sur y la derecha el de la parte norte. Las instancias del ATSP se resuelven en pocos segundos ya que son instancias muy pequeñas. Dado que los camiones

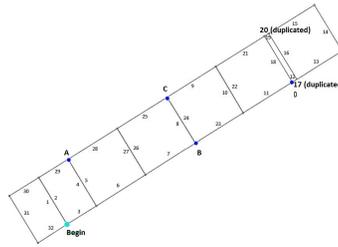


Figura 8: Ejemplo de ubicación eficiente de los contenedores propuesta por el modelo de PLE.

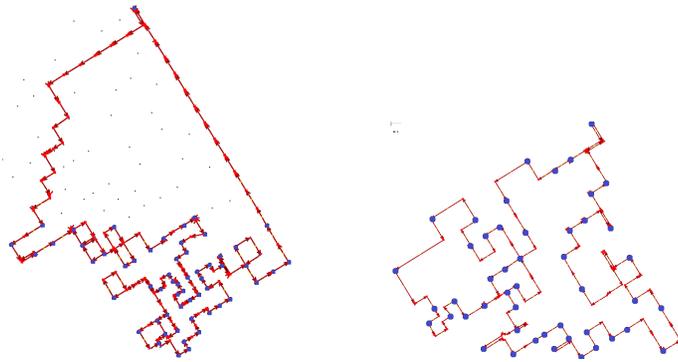


Figura 9: Recorridos obtenidos para los camiones recolectores para la Zona 1.

se guardan en un depósito alejado de la ciudad, ubicado al noreste de la región urbana, el tour comienza en un contenedor de la parte superior derecha del mapa. Lo mismo ocurre para la zona norte. En ambos casos, los contenedores a recoger están marcados con un punto resaltado.

Previamente a este trabajo, dado que todas las esquinas se utilizaban como depósitos de montículos de hojas, los camiones fijaban su recorrido pasando por todas las cuadras de la ciudad en forma zigzagueante. Este nuevo recorrido, a pesar de parecer un poco más complejo a la vista, supone un importante ahorro en la distancia recorrida.

En términos numéricos, anteriormente se recorría un total de aproximadamente 32,2 km para los dos camiones en conjunto. El nuevo recorrido sólo insume 28,4 km entre ambos (14,7 km y 13,7 km para la zona sur y norte respectivamente), lo que produce una reducción del 12 % en la distancia total.

5. Conclusiones y trabajo a futuro

En cuanto a los resultados obtenidos en este trabajo, la asignación eficiente de cuadradas de los barrenderos era el objetivo principal para la Municipalidad. Como ya se mencionó, en la actualidad la ciudad se encuentra en plena fase de implementación de la Etapa I. Mediante este trabajo se pudo reestimar la cantidad necesaria de operarios para cada una de las zonas, que previamente estaban desbalanceadas.

Paulatinamente se va incorporando personal al equipo de operarios, y se van implementando los nuevos recorridos presentados en este trabajo. Hasta el momento los resultados han sido satisfactorios teniendo en cuenta fundamentalmente la percepción de los responsables de organizar la tarea en la Municipalidad.

Un aspecto que probablemente se tenga en cuenta más adelante, es asociar la dificultad de la cuadra o manzana a cambios dados por motivos estacionales. Por ejemplo, una cuadra que tiene muchos árboles va a insumir más tiempo de barrido durante el otoño, época en se produce la caída de hojas. En estos casos, por ejemplo, una solución posible podría llegar a ser multiplicar el tiempo que insume barrer esa cuadra por un factor de corrección adecuado.

Una idea interesante desde el punto de vista teórico, pero quizás un poco difícil de realizar en la práctica, hubiera sido distinguir a cada barrendero por su velocidad de barrido. Con esta información, para cada manzana se podría haber cronometrado el tiempo que ese trabajador tarda en limpiarla, y luego, dividiendo por una constante de velocidad de barrido de dicho operario, podría haberse obtenido una estimación absoluta para la dificultad temporal de cada manzana. De esta forma, se podrían haber conseguido soluciones mucho más exactas al ajustar el parámetro por la constante de cada barrendero.

Otro elemento que también podría haberse considerado sería sumarle a cada segmento una penalidad temporal según que tan alejado esté el centro de masa del punto de salida del barrendero. En la práctica, algunos de los barrenderos guardan los carritos en sus propias casas, con lo cual el punto de salida puede llegar a ser variable, y por ende no tuvo mucho sentido incluirlo en el modelo que aplicamos. Pero si se realizara la distinción por barrendero sí podría hacerse.

Es intención del municipio extender este tipo de estudios a otras áreas de su funcionamiento cotidiano. Según Miguel Fernández, actual Intendente de la ciudad, “desde nuestro punto de vista, esta clase de análisis que nos habilitan a

tomar decisiones informadas es a lo que deberíamos apuntar también en otras áreas, ya que nos permiten optimizar los recursos disponibles”.

Agradecimientos: Este trabajo forma parte de un convenio de cooperación entre el Instituto de Cálculo de la Facultad de Ciencias Exactas y Naturales de la UBA y la Municipalidad de Trenque Lauquen. Agradecemos la permanente colaboración para concretar este proyecto de las autoridades del municipio, en particular de Miguel Fernández, Intendente de la ciudad; Marcelo Ferreyra, Subsecretario de Modernización, Ciencia y Tecnología; y Adhemar Enrietti, Subsecretario de Obras y Servicios Públicos. El tercer autor está parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería, ISCI, Chile (CONICYT PIA/BASAL AFB180003), y los subsidios UBACyT 20020170100495BA (Argentina) y ANPCyT PICT 2015-2218 (Argentina).

Referencias

- [1] W. H. Beyer. Crc standard mathematical tables and formulae. *28th ed. Boca Raton, FL: CRC Press*, páginas 123–124, 1987.
- [2] M. Bianchetti, G. Durán, I. Koch, y J. Marengo. Algoritmos de zonificación para el problema de la recolección de residuos urbanos: El caso de estudio de una ciudad argentina. *Revista Ingeniería de Sistemas*, 31:81–110,, 2017.
- [3] L. D. Bodin y S. J. Kursh. A computer-assisted system for the routing and scheduling of street sweepers. *Operations Research*, 26(4):525–537, 1978.
- [4] F. Bonomo, D. D. Donne, G. Durán, y J. Marengo. Automatic dwelling segmentation of the buenos aires province for the 2010 argentinian census. *Interfaces*, 43(4):373–384, 2013.
- [5] F. Bonomo, G. Durán, F. Larumbe, y J. Marengo. A method for optimizing waste collection using mathematical programming: a buenos aires case study. *Waste Management & Research*, 30(3):311–324, 2012.
- [6] G. Braier, G. Durán, J. Marengo, y F. Wesner. An integer programming approach to a real-world recyclable waste collection problem in argentina. *Waste Management & Research*, 35(5):525–533, 2017.
- [7] Concorde. Concorde TSP Solver, 2015. Recuperado de <http://www.math.uwaterloo.ca/tsp/concorde/index.html>.

- [8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, y C. Stein. *Introduction to algorithms*. MIT press, 2009.
- [9] R. W. Eglese y H. Murdock. Routeing road sweepers in a rural area. *Journal of the Operational Research Society*, 42(4):281–288, 1991.
- [10] H. A. Eiselt, M. Gendreau, y G. Laporte. Arc routing problems, part i: The chinese postman problem. *Operations Research*, 43(2):231–242, 1995.
- [11] C. Hierholzer y C. Wiener. Über die möglichkeit, einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. *Mathematische Annalen*, 6(1):30–32, 1873.
- [12] P. J. Taylor. A new shape measure for evaluating electoral district patterns. *American Political Science Review*, 67(3):947–950, 1973.
- [13] J. B. Weaver y S. W. Hess. A procedure for nonpartisan districting: Development of computer techiques. *The Yale Law Journal*, 73:288–308, 1963.
- [14] H. P. Young. Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, 13(1):105–115, 1988.

Instituto Sistemas Complejos de Ingeniería

INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA

ISCI
Instituto Sistemas
Complejos de Ingeniería

El Instituto Sistemas Complejos de Ingeniería (ISCI) trabaja en un continuo que va desde el desarrollo de metodologías de frontera hasta aplicaciones innovativas en Sistemas Complejos de Ingeniería, en los cuales interactúan infraestructura, comportamiento humano y gestión.

GRUPOS DE INVESTIGACIÓN

- Recursos Naturales: Minería, Forestal y Acuicultura
- Transporte
- Localización y Logística
- Consumer Analytics
- Smartcities
- Data Science
- Organización Industrial
- Energía

PATROCINA:



INSTITUCIONES ALBERGANTES:



ANEXO DEL ESTADIO
PUEBLO Y MARTIN GARCÍA
UNIVERSIDAD DE CHILE



INSTITUCIÓN ASOCIADA:



contacto@isci.cl
www.isci.cl

Programas de Postgrado y Postítulos DII



DOCTORADO
EN SISTEMAS DE INGENIERÍA
UNIVERSIDAD DE CHILE



FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Doctorado en Sistemas de Ingeniería



**Un Programa Interdisciplinario
en Investigación Avanzada**

Formamos especialistas con una sólida base tecnológica y un conocimiento profundo de las herramientas que permiten modelar, entender y optimizar sistemas complejos en que interactúan elementos físicos de comportamiento humano.

www.dsiuchile.cl
doctorado@sistemasdeingenieria.cl
56 29784017
Beauchef 851
Torre Oriente 2° Piso
Oficina 204

7
AÑOS

Universidad Acreditada
Diciembre de Pregrado
Gestión Institucional
Diciembre de Postgrado
Vinculación con el medio
Nov 2018 - Nov 2025

4
AÑOS

Doctorado Acreditado
Agosto 2016 a
Agosto 2020
CNA-Chile



Programas de Postgrado y Postítulos DII



MGO
INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE



FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Magíster en Gestión de Operaciones

El objetivo del Magíster en Gestión de Operaciones es formar graduados de excelencia en investigación de operaciones, quienes podrán enfrentar problemas complejos en gestión de operaciones. Integrando herramientas matemáticas, económicas y tecnológicas.

- **Articulación con Pregrado**
(Doble titulación y descuento de 50% del arancel)
- **Líder en Gestión de Operaciones**
- **Más de 200 Graduados en empresas destacadas y doctorados de clase mundial.**



Cuerpo Docente
de Excelencia



Posicionamiento de
sus Egresados

MAGÍSTER EN GESTIÓN
OPERACIONES

www.mgo.uchile.cl
56 29784017
Beauchef 851
Torre Oriente 2° Piso
Oficina 204

Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Industrial

Programas de Postgrado y Postítulos DII

MAGÍSTER EN ECONOMÍA APLICADA



OBJETIVO

El Magíster en Economía Aplicada forma profesionales y académicos de gran capacidad analítica y sólida base en economía.

4

LÍNEAS DE ESPECIALIZACIÓN

Macroeconomía
Microeconomía
Economía Financiera
Métodos Empíricos

CUERPO ACADÉMICO



Publicaciones en las principales revistas académicas y conferencias internacionales
Investigación internacional
Iniciativas que favorecen la investigación



GRADUADOS MAGCEA



"El MAGCEA es sin lugar a dudas un programa de élite, y se lo recomiendo a quienes no solo tengan el deseo de profundizar sus conocimientos económicos, sino también la constancia y perseverancia necesaria para superar desafíos".

Tomás Wilner,
Ingeniero Civil Industrial, Universidad de Chile. Estudiante,
PhD Economics, Northwestern University.



"Hacer el MAGCEA me permitió optar a experiencias laborales estrechamente ligadas con la formación que recibí y tengo la certeza, que lo aprendido en el MAGCEA será un impulso importante para seguir especializando mi carrera profesional".

Marisol Vargas,
Ingeniero Civil Industrial, Universidad de Chile. Analista de
riesgo de estrategias globales, Compass Group.



"Realizar el MAGCEA fue clave para hacer un doctorado en economía en una de las mejores universidades de Estados Unidos. La calidad del programa fue fundamental para realizar investigación de alto impacto".

Anaíve Pedraza,
Ingeniero Civil Matemático, Universidad de Chile. Estudiante,
PhD Economic Analysis & Policy, Stanford University.



"El MAGCEA me entregó las bases y herramientas para realizar un trabajo de calidad en el área de economía, la cual no era mi especialidad de pregrado. También, me ayudó a conseguir el trabajo que quería, poco tiempo después de haber egresado".

Paulina Arena,
Ingeniero Civil Matemático, Universidad de Chile. Economista,
División Antimonopolios de la Fiscalía Nacional Económica.



"El MAGCEA significó una herramienta fundamental para acceder a mi trabajo actual y para encontrar mi primer trabajo. El MAGCEA, me facilitó entrar al mundo de las políticas públicas y desde una mirada técnica, aportar un granito de arena a la sociedad".

Stefano Banff,
Ingeniero Civil Industrial, Universidad de Chile. Economista,
División de Prospectiva y Análisis de Impacto Regulatorio,
Subsecretaría de Energía.



"Los profesores del magíster me guiaron durante todo el proceso de postulación al doctorado, aconsejándome en cuanto a qué universidades elegir, cómo postular a las becas y por supuesto alentándome cuando fue necesario".

Valentina Costeiras,
Ingeniero Civil Industrial, Universidad de Chile. Estudiante, PhD
Economics, University College of London (UCL).

+200
GRADUADOS

90% TUVO MEJOR RENTA

83% ACCEDIÓ A MEJOR TRABAJO

96% RECOMENDARÍA MAGCEA

35%

PHD EN UNIVERSIDADES DE CLASE MUNDIAL

Northwestern University
Stanford University
UCLA
University of Minnesota
University of Pennsylvania
Yale University
Berkeley University
Princeton University

19%

ORGANISMOS REGULADORES Y SECTOR PÚBLICO

Fiscalía Nacional Económica
Superintendencia de Seguridad Social

41%

SECTOR PRIVADO Y EN LA BANCA

HSBC
AFP Modelo
Banco Penta
BCI
Banchile
Codelco
BAIN Company
Banco Central

7
AÑOS
Universidad Acreditada
Decreto de Programa
Educativo 2009
Decreto de Programa
de Postgrado
Nº 2018 - N° 2025

8
AÑOS
Magíster Acreditado
Octubre 2012 hasta
Octubre 2018
Aprobación CII
Nº 2018 - N° 2025

Comisión Nacional de Acreditación CNA-Chile

Descuento de arancel de hasta 100% para mujeres (2 cupos máximo por año)

INFRAESTRUCTURA DE PRIMER NIVEL

Informaciones: www.magcea-uchile.cl - +56 2 2978 4072 | +56 2 2978 4073

Programas de Postgrado y Postítulos DII



MAGÍSTER EN GESTIÓN Y POLÍTICA PÚBLICAS

PROGRAMA LÍDER EN CALIDAD Y COMPROMISO PÚBLICO

Referente en la formación de profesionales que cuentan con los conocimientos, las competencias y las habilidades para liderar los procesos de diagnóstico, comprensión, resolución y priorización de problemas públicos, vinculando el conocimiento con las necesidades de los países y resguardando los criterios de eficiencia, eficacia, impacto y legitimidad ciudadana de las acciones públicas.

REFERENTE EN AMÉRICA LATINA Y EL CARIBE EN LA FORMACIÓN EN GESTIÓN Y POLÍTICAS PÚBLICAS

1

SOMOS UNIVERSIDAD DE CHILE:
Liderazgo, excelencia y vocación pública.

2

FORMACIÓN INTERDISCIPLINARIA
para la comprensión y resolución de
problemas públicos complejos.

3

INCORPORAMOS LA GESTIÓN como
sello de nuestra propuesta formativa.

4

UN ESPACIO DE APRENDIZAJE
desafiante, diverso y global.

5

CUERPO ACADÉMICO DE EXCELENCIA
y destacada trayectoria en el ámbito
público.

6

UN PROGRAMA INTERNACIONAL:
formamos liderazgos para enfrentar
los desafíos globales y de la región en el
ámbito público.

7

una extensa red de graduados
presentes en más de 20 países.

Síguenos en
redes sociales

MGPP_UChile

@MGPPUChile

@MGPP_DII

Más
información

www.mgpp.cl

mgpp@dii.uchile.cl

+56 2 229784067

+56 2 229784043



Programas de Postgrado y Postítulos DII



MBA UNIVERSIDAD DE CHILE
INGENIERÍA INDUSTRIAL

PRIMEROS EN GESTIÓN

TALLERES DE HABILIDADES DIRECTIVAS
QUE TE PREPARAN PARA CARGOS DE ALTO LIDERAZGO

ENFOQUE DEL MANAGEMENT SCIENCE PARA
TOMAR MEJORES DECISIONES DE GESTIÓN

MBA EXECUTIVE

24 o 20 meses

Viernes (tarde) y sábado cada 2 semanas

MBA PART TIME

20 o 16 meses

3 a 4 veces a la semana (tardes)

8

ANOS

REGISTRO DE GESTIÓN
Y DIRECCIÓN DE EMPRESAS
MAYO 2012 A MAYO 2018



MBA UChile
Ingeniería Industrial | Universidad de Chile

Admisiones: +56 2 229784002

Programas de Postgrado y Postítulos DII



MBA

INDUSTRIA MINERA
INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE

FORMANDO LÍDERES EN GESTIÓN Y DIRECCIÓN DE EMPRESAS MINERAS

Magíster en Formato Weekend - Clases cada 2 semanas

- MBA único e innovador
- Selecto equipo docente
- Énfasis en habilidades directivas
- Importante networking

+ SEMINARIO ESPECIALIZACIÓN EN CHINA

Domina las herramientas necesarias para liderar un mundo globalizado y conoce la realidad de la industria minera en el mercado más relevante a nivel mundial.

*Curso opcional

fcfm
FACULTAD DE CIENCIAS
FÍSICO-MATEMÁTICAS
UNIVERSIDAD DE CHILE

7
AÑOS
Universidad Acreditada
Ciencia de Postgrado
Ciencia Industrial
Ciencia de Postgrado
Investigación
Vinculación con el medio

8
AÑOS
Magíster en Gestión y
Dirección de Empresas
Acreditado
Marzo 2017 a Mayo 2020

CNA
Comisión Nacional
de Acreditación
CNA-Chile

Más información en:

mbamin@dii.uchile.cl
Fono: (+56 2) 29780554
www.mbamin.cl

Programas de Postgrado y Postítulos DII



Desarrolla una mirada estratégica de tu organización

y lidera proyectos de
innovación tecnológica
en procesos de negocios

15 AÑOS DE
EXPERIENCIA



MBE

MASTER IN BUSINESS ENGINEERING
INGENIERÍA INDUSTRIAL | UNIVERSIDAD DE CHILE

Duración: 3 semestres académicos

www.mbe.cl

Contacto: mbe@dii.uchile.cl | +56 22 978 4835



Programas de Postgrado y Postítulos DII

EDUCACIÓN EJECUTIVA



EDUCACIÓN EJECUTIVA
UNIVERSIDAD DE CHILE
INGENIERÍA INDUSTRIAL

EXCELENCIA PARA IMPULSAR TU CARRERA

DIPLOMADOS

En diferentes áreas y formatos

CURSOS DE ESPECIALIZACIÓN

Más específicos y actualizados

PROGRAMAS CORPORATIVOS

Creados a la medida de tu empresa

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

