

# El aprendizaje de máquina como herramienta de evaluación de políticas públicas implementadas a nivel nacional. Caso de la seguridad de tránsito.

Francisco Vergara  
*Departamento de Electrónica*  
*Universidad Técnica Federico Santa María*  
Valparaíso, Chile  
francisco.vergara.12@sansano.usm.cl

**Resumen**—En este documento se utiliza el aprendizaje de máquina sobre el registro de siniestros viales ocurridos en Chile entre los años 2010 y 2017 con la finalidad de estudiar el efecto de la nueva política de seguridad de tránsito que comienza a ser trabajada en 2018, desde una perspectiva social, buscando acercar la tecnología al mundo de las organizaciones sociales.

**Index Terms**—aprendizaje de máquina, siniestros viales, seguridad vial, organizaciones, ciudadanía, inteligencia artificial.

## I. INTRODUCCIÓN

Las tecnologías de información día a día presentan nuevos avances, mejorando así la eficiencia en negocios, marketing, publicidad y otros rubros, generando también una brecha cada vez más significativa entre la población y los profesionales de las industrias asociadas a los datos. En este contexto, este documento propone enfocar el Aprendizaje de Máquina como una herramienta para las organizaciones y la sociedad civil en las distintas áreas de expertiz de cada una, haciendo uso de la información que, por transparencia, debe ser accesible por cualquier persona que desee acceder a ella.

En base a esto, se estudiará el impacto que genera la implementación de una nueva política pública de seguridad de tránsito a nivel nacional, buscando comparar los resultados de los siniestros viales registrados luego de su implementación, con una predicción de lo que hubiese ocurrido al no implementarse.

Este documento busca ser una referencia para las organizaciones de la sociedad civil y la ciudadanía en general, por lo cual se pretende incluir el conocimiento adquirido en las distintas organizaciones que convergen en la movilidad urbana como un camino para mejorar la calidad de vida de las personas y la vivencia de la ciudad en el proceso de análisis de la información, además de registrar el origen de los datos y la comprensión de los mismos para poder supervisar el aprendizaje computacional. Sin embargo, existen secciones del documento con información técnica que explica lo que hay

Se agradece a la Fundación Andes Chile ONG por permitir el trabajo desde el aspecto tecnológico sobre la movilidad urbana, a MAP8 ONG por permitir la utilización de recursos informáticos para publicar información y al Colectivo Muévete por generar una instancia de trabajo colaborativa en torno a los datos publicados.

detrás del estudio realizado, desde el ámbito de la ingeniería y tecnologías de información.

Los códigos que sustentan este documento se encuentran codificados en el lenguaje de programación Python, desde donde se utilizan librerías de Ciencias de Datos como *Pandas* (para el trabajo con conjuntos de datos organizados), *Scikit Learn* (como kit de preparación de datos y aprendizaje de máquina), *Matplotlib* (utilizada para generar los gráficos que incluye el documento) y otras ya incluidas para manejar funciones básicas.

## II. ORIGEN Y USO DE LOS DATOS

Un factor importante a considerar cada vez que se desea trabajar con datos es su origen y el uso que se les busca entregar, tanto desde el momento de su adquisición como durante el estudio a realizar. En muchos casos, puede existir diferencias, por lo cual la información se puede considerar excesiva o incompleta.

### II-A. Instituciones

El conjunto de datos que se utilizará provienen desde una solicitud de transparencia realizadas por un miembro del Colectivo Muévete y un miembro del Movimiento contra el Exceso de velocidad Letal, en forma paralela, a la Comisión Nacional de Seguridad de Tránsito. Se destaca que la información obtenida no tienen en común la solicitud y tampoco presentan una finalidad específica. Sin embargo, los datos obtenidos difieren en formato y contenido, lo cual genera una primera observación a tener en cuenta. Por otro lado, cabe destacar que en ambos casos, se obtiene como resultado documentos con extensión XLSX (Microsoft Excel), lo cual dificulta el trabajo desde un computador de gama media, considerando que el total de registros supera los setecientos mil.

Los primeros datos obtenidos se concentran en un único archivo con el formato ya mencionado, mientras que el segundo conjunto se encuentra repartido en 10 archivos, cada uno correspondiente a un año (desde 2010 hasta 2019, inclusive)

de registros. Además, en estos últimos hay columnas con información no disponible en el primer caso<sup>1</sup>.

## II-B. Adquisición y finalidad

Los datos a trabajar se han registrado mediante el Formulario 2 Registro de Accidentes de Tránsito y Ferroviario (SIEC 2) de Carabineros de Chile, por medio del Departamento de Servicios de Tránsito y Carreteras. En este formulario, se informa una única causa basal probable (CBP). Aquellos otros elementos que pudieron haber contribuido a la ocurrencia del accidente<sup>2</sup>, pero que a criterio del Carabinero *in situ* no constituyen una CBP, no quedan registrados en el Formulario 2 [1]. Cabe destacar que esta información no está procesada y tiene como finalidad únicamente quedar registrada en los expedientes asociados a los juicios penales y de responsabilidad civil que se derivan los siniestros registrados por la institución.

El procesamiento y tabulación de la información es realizada por la Comisión Nacional de Seguridad de Tránsito, sin embargo, es posible encontrar datos incompletos para la finalidad de entidades ajenas a Carabineros de Chile o la ya mencionada comisión. Cabe destacar que por medio de los informes anuales publicados en los medios oficiales es posible notar la existencia de un mayor detalle del que se dispone en los conjuntos de datos recibidos.

## II-C. Objetivos del documento

Una vez detallado el origen y uso común para los datos, es necesario especificar el objetivo para el cual serán utilizados y, de esta forma, establecer los pasos necesarios para lograrlos. En este sentido, dado el enfoque ciudadano de este documento, hay dos elementos que pueden ser de interés: la propuesta de políticas públicas o regulaciones, en este caso, en el ámbito del transporte, y la evaluación de las implementaciones ya realizadas.

Dada la corta extensión de este estudio, se considera como objetivo específico la evaluación de la política nacional de seguridad de tránsito actual que fue publicada en 2017 y está basada en cinco ejes estratégicos: (1) gestión de la seguridad de tránsito, (2) vías y movilidad más seguras, (3) vehículos más seguros, (4) usuarios de las vías más seguros y (5) respuestas tras los siniestros de tránsito. El enfoque de dicha política es el Sistema Seguro, significando que el país, el cual ha sido promovido, especialmente, por países líderes en seguridad de tránsito y que establece como imperativo ético que las muertes y lesiones graves en el tránsito son inaceptables y que, por lo tanto, el sistema de tránsito debe diseñarse y utilizarse de manera tal que nadie pierda la vida o quede gravemente herido como resultado de un siniestro de tránsito [2]. De esta forma, considerando los ejes mencionados, cabe señalar que el conjunto de datos a utilizar comprende

<sup>1</sup>Esto permite destacar arbitrariedad en la información entregada mediante solicitud ciudadana, lo cual pone en duda directamente la transparencia de la institución y entorpece fuertemente el trabajo ciudadano.

<sup>2</sup>En el artículo citado se consideran los sucesos como “accidentes”, sin embargo, en este documento se utiliza la palabra “siniestro” debido a la no aleatoriedad en su ocurrencia. Se considera que siempre existe una causa humana asociada.

la evaluación para los ejes 2, 3 y 4, ya que los demás ejes disponen de una cantidad nula o insuficiente de registros asociados en el conjunto de datos disponible, lo cual puede perjudicar los resultados finales.

## II-D. Metodología

Para lograr el objetivo propuesto, es necesario entender que una evaluación corresponde a la determinación de valor, en este caso, de la política publicada en 2017 y que comienza a regir en 2018. Para esto, se considerará que un objetivo de la implementación de la política pública con visión cero es anular la ocurrencia de siniestros viales con resultados graves o fatales. Sin embargo, basta revisar el conjunto de datos para notar que esto no se cumple. Por tanto, es necesario verificar si el hito a evaluar ha influido en los resultados de siniestros viales registrados en 2018 y 2019, en comparación con los registros que podrían haber existido al no implementar la visión cero.

En primera instancia, lo que se busca encontrar puede parecer imposible, ya que se requiere comparar datos registrados con datos que no existen debido a la imposibilidad de obtener siniestros en 2018 y 2019 que no estén bajo la implementación de la nueva política pública de seguridad de tránsito en 2017, sin embargo, esa es justamente la finalidad del aprendizaje de máquina: estimar el comportamiento de un conjunto de datos basados en un conjunto conocido. En este caso, se implementará esta herramienta para predecir la cantidad de muertes o lesionados graves asociados al periodo a evaluar, considerando el conjunto de datos disponibles previos a 2017. De esta forma, existirán dos conjuntos disponibles para el estudio: los resultados estimados y realmente registrados. Por tanto, la evaluación de la nueva política pública se basará en la comparación de estos, concluyéndose positiva si es que se encuentra una reducción importante de resultados graves o fatales en los siniestros viales ocurridos.

## III. ANÁLISIS

En este punto, es necesario tener más detalles de los datos disponibles, por lo cual se hace necesaria una comprensión del conjunto. A continuación se mostrará el listado de columnas disponibles y una breve descripción en caso de ser necesario.

1. *Año*<sup>3</sup>.
2. *IdAccidente*: registro en la base de datos desde la que es extraída la información para ser compartida.
3. *Fecha*<sup>3</sup>.
4. *Hora*<sup>3</sup>.
5. *Región*<sup>4</sup>.
6. *Comuna*<sup>4</sup>.
7. *Tipo Accidente*: corresponde a la clasificación específica del siniestro.

<sup>3</sup>Las columnas con variable temporal se refieren al momento en que se registró el siniestro.

<sup>4</sup>Las columnas con datos de ubicación se refieren al lugar en que se registró el siniestro.

8. *Tipo (CONASET)*: se refiere a la clasificación principal del siniestro. Los tipos posibles son volcadura, colisión, choque, atropello, otro tipo, caída o incendio.
9. *Zona*: identifica si el siniestro ocurrió en zona urbana o rural.
10. *Ubicación Relativa*<sup>5</sup>: identifica el espacio físico en que se produjo el siniestro. Algunas ubicaciones relativas son puente, rotonda, túnel, tramo de vía recta, entre otros.
11. *Causa (CONASET)*<sup>5</sup>: corresponde a la causa registrada del siniestro sin mayor detalle. Entre estas se encuentran las fallas mecánicas, pérdida de control del vehículo, deficiencias viales y otras causas.
12. *Causa Accidente*<sup>5</sup>: indica mayor detalle sobre la columna anterior, como por ejemplo el tipo de falla mecánica, el motivo de la pérdida del control del vehículo, el tipo de deficiencia vial y otros.
13. *Calle Uno*<sup>6</sup>.
14. *Calle Dos*<sup>6</sup>.
15. *Número*: en caso de no haber una intersección suficientemente cercana, se considerará el número asociado a la propiedad más cercana al lugar del siniestro.
16. *Ruta*: en caso de que el lugar sea una carretera, se indicará el nombre de la ruta.
17. *Ubicación KM*: indica el kilómetro de la ruta en que se registró el siniestro.
18. *Pistas de Ida*: referida a la cantidad de pistas de 'ida' que hay en el lugar.
19. *Pistas de Regreso*: referida a la cantidad de pistas de 'regreso' que hay en el lugar.
20. *Calzada*: indica si la calzada en que ocurrió el siniestro es unidireccional o bidireccional. También indica la existencia de bandejón central cuando es necesario.
21. *Tipo Calzada*: puede ser de pavimento o asfalto.
22. *Estado Calzada*: se refiere a la evaluación en el momento del registro sobre la calzada. Puede ser bueno, malo o regular.
23. *Condición Calzada*<sup>5</sup>: indica si la calzada se encuentra seca, húmeda o en otro estado.
24. *Estado Atmosférico*<sup>5</sup>: referido a si, al momento del registro, se encontraba despejado, nublado, con lluvia u otro.
25. *Fallecidos*<sup>7</sup>.
26. *Graves*<sup>7</sup>.
27. *Menos Graves*<sup>7</sup>.
28. *Leves*<sup>7</sup>.

### III-A. Preparación

Comprendidos los datos, es necesario realizar la limpieza correspondiente para trabajar únicamente en aquellos elementos que son de interés para realizar este estudio, por lo cual es

<sup>5</sup>Dado que el conjunto de posibilidades en esta columna son varias, sólo se mencionan algunos elementos.

<sup>6</sup>En algunos casos, se indica el lugar del siniestro considerando la intersección en que ocurrió. Esta está dada por dos calles: 'Calle Uno' y 'Calle Dos'

<sup>7</sup>Cantidad registrada en el siniestro.

necesario definir las columnas de interés, que corresponde a las siguientes: *Año, Fecha, Tipo (CONASET), Ubicación Relativa, Causa (CONASET), Calzada, Tipo Calzada, Estado Calzada, Condición Calzada, Estado Atmosférico, Fallecidos, Graves.*

Como se puede notar, el análisis a realizar considerará sólo resultados a nivel nacional. Este estudio se puede realizar a nivel regional para conocer el impacto de las medidas adoptadas en los distintos sectores de la división política del país, sin embargo, al ser menor la cantidad de datos que a nivel nacional, la precisión no será la misma. En este aspecto, es importante considerar que el análisis a nivel comunal puede ser de alta imprecisión debido a la limitación en la cantidad de registros disponibles.

El siguiente paso en la preparación de los datos es interpretar y, en caso de ser necesario, eliminar los registros con valores nulos. Para esto es necesario convenir que si las columnas *Fallecidos* y *Graves* poseen valor nulo, serán interpretadas como cero. En el resto, se eliminarán los registros con al menos un elemento nulo debido a que no existe una interpretación válida para todos los casos en que pueda faltar información.

Para terminar la preparación de los datos, es necesario definir tres subconjuntos:

1. Conjunto de entrenamiento: considerará los datos con los cuales se realizará el entrenamiento en el aprendizaje de máquina. Es el conjunto más grande y se compone por los datos obtenidos entre 2010 y 2017.
2. Conjunto de pruebas: utilizado para evaluar el estado de aprendizaje. En base a la comparación entre los resultados obtenidos y los datos disponibles aquí se puede decidir si el sistema está listo o no para ser utilizado en una predicción. Este conjunto sólo contiene los datos de 2016, sin existir un motivo particular para esta selección.
3. Conjunto para evaluación: es el conjunto de registros realizados entre 2018 y 2019 y se debe comparar con los resultados precedidos por el sistema de aprendizaje, que representan los registros realizados en el periodo mencionado sin la aplicación de una nueva política de seguridad de tránsito.

Para comprender mejor el modelo a utilizar, si bien el proceso que sigue se realizará a nivel computacional, se requiere conocer el conjunto de entrenamiento para tener una mayor precisión en las siguientes etapas. Por otro lado, esta es una sección de bastante interés para el área en la que se realiza el análisis, ya que permite escoger los valores de *k* y *weight* en el siguiente ítem, junto con supervisar el trabajo que realiza el computador, representando un momento de gran colaboración entre el mundo de los datos y, en este caso, el mundo del transporte.

Con la finalidad de entender el conjunto de datos, se realizará un análisis descriptivo de los datos a nivel superficial. Para esto, se considera el método *describe()* disponible en la librería *Pandas* de Python, obteniendo los resultados mostrados en el Cuadro I. Por otro lado, los datos cuantitativos se encuentran resumidos estadísticamente en el Cuadro II. Sin embargo, cabe

Cuadro I  
INFORMACIÓN CUALITATIVA<sup>a</sup>

Columna	Más frecuente	Frecuencia
Fecha	2017-06-16	12,98 %
Tipo (CONASET)	COLISION	54,11 %
Ubicación Relativa	TRAMO DE VIA RECTA	53,28 %
Causa (CONASET)	IMPRUDENCIA CONDUCTOR	52,59 %
Calzada	BIDIRECCIONAL	55,14 %
Tipo Calzada	ASFALTO	53,69 %
Estado Calzada	BUENO	91,78 %
Condición Calzada	SECO	88,47 %
Estado Atmosférico	DESPEJADO	83,12 %

<sup>a</sup>El conjunto de entrenamiento posee 421454 datos.

destacar que esta información corresponde al total de siniestros registrados, lo cual explica la pequeña magnitud de los valores, ya que normalmente se espera que no sea la mayoría de los siniestros registrados los que tengan resultados graves o fatales.

### III-B. Modelo y Entrenamiento

Antes de continuar con el análisis, es necesario sustentar técnicamente el trabajo, lo que requiere entrar en algunos detalles del Aprendizaje de Máquina, que puede ser considerado un subcampo de la Inteligencia Artificial (IA) [3]. Dentro de esta disciplina, es posible encontrar el aprendizaje supervisado, en que los algoritmos “aprenden” desde conjuntos de entrenamientos etiquetados, es decir, estos estudian desde datos que contienen gran parte de las posibles entradas que serán requeridas posteriormente para generar predicciones o salidas, similar a uno de los métodos comunes de estudios entre los universitarios que revisan las evaluaciones de semestres anteriores para estimar la dificultad de la prueba que deben enfrentar en una asignatura.

De la idea anterior, se sigue que los algoritmos utilizados para el aprendizaje de máquina pueden tratarse como sistemas o cajas negras en las que se ingresa un conjunto de entradas y, luego de cierto procesamiento, se puede obtener una o más salidas. Sin embargo, es necesario profundizar.

Para este estudio, se utiliza el método *k-Nearest Neighbor* (en español, *k-Vecinos más Cercanos*) que, en terminos sencillos, se basa en buscar, para un elemento desconocido (lo que se desea predecir), aquel en conjunto de entrenamiento que es más parecido para estimar el resultado de este nuevo dato. Una analogía para poder entender mejor esta idea parte con la pregunta ¿Cuál es la forma correcta de predecir el precio de una casa? Una forma es describir cada propiedad en terminos de sus características y asignar a cada una de estas un valor monetario que se ponderará con el resto de precios

Cuadro II  
INFORMACIÓN CUANTITATIVA<sup>b</sup>

	Fallecidos	Graves
Media	0.020074	0.093188
Desviacion Estandar	0.161589	0.345732

<sup>b</sup>El conjunto de entrenamiento posee 421454 datos.

asociados para obtener el costo estimado total. Otra opción es averiguar con personas conocidas que viven en sectores similares y estimar un precio similar a estos [4]. Naturalmente, la segunda opción es más cómoda y es precisamente de lo que trata el método utilizado en este documento.

Específicamente, se utilizará *k-Neighbor Regressor* (Regresión del *k*-ésimo vecino más cercano, en español) que provee la librería *Sci-kit Learn* de Python. Esta elección se debe a la característica principalmente continua de los resultados buscados. Si bien, en este caso, las víctimas fatales o graves de un siniestro vial se contabilizan con números naturales y el cero, la gran variabilidad que pueden tener los resultados y la cantidad de datos disponibles permiten que este método funcione correctamente. Sin embargo, cabe destacar que para no generar distancia con la naturaleza de los datos, los resultados de las predicciones que se realizan en este documento son aproximados a un número entero positivo o a cero.

Continuando con el detalle, es necesario detallar los siguientes elementos necesarios para utilizar *k-Neighbor Regressor*:

- Número de vecinos cercanos (*k*): indica la cantidad de datos o “vecinos cercanos” a considerar para cada etiqueta. La forma correcta para elegir *k* es escoger una fracción de las muestras a entrenar como conjunto de pruebas y experimentar con diferentes valores del parámetro para ver dónde se obtiene el mejor desempeño [4]. Como se estableció previamente, el conjunto de pruebas corresponde a los registros realizados durante 2016 y, para considerar el valor óptimo de *k*, basta comparar los valores de la media y desviación estándar para los fallecidos y lesionados graves, tanto estimados como reales. Luego de esto, se establece  $k = 35$ .
- Peso (*weight*): se refiere a cómo contribuye cada dato clasificado en la “comunidad” o conjunto de vecinos cercanos a la estimación de la clasificación de un dato que se desea predecir. Para la librería de Python utilizada, se dispone de *uniform*, en que cada elemento contribuye de la misma forma al resultado, y de *distance*, donde cada elemento cercano tendrá un peso proporcional a su distancia con el valor que representa la entrada de lo que se desea predecir. La forma de escoger el valor a utilizar es similar al caso anterior y, para este caso, ha sido  $weight = distance$ .
- Algoritmo: para generar y encontrar el “vecino más cercano” hay diversas formas, sin embargo, se mencionarán 4 opciones a grandes rasgos. El primero es *brute*, es decir, fuerza bruta, que es útil para conjuntos de datos pequeños, sin embargo, no se profundizará debido a que la cantidad de datos utilizados en este trabajo no calza con la condición mencionada, lo cual implica un costo computacional para la búsqueda de complejidad  $O(DN)$ <sup>8</sup>.

En segundo lugar, se tiene el algoritmo *k-d tree* (versión corta la expresión, en español, árbol *k*-dimensional), en

<sup>8</sup>Se considera *D* como la cantidad de dimensiones o columnas de la entrada y *N* la cantidad de muestras para el entrenamiento.

que cada elemento del conjunto de entrenamiento es guardado como un nodo, utilizando árboles binarios como estructura de datos [5]. Además de la distribución de la información, se tiene que cada nodo almacena la información de su distancia a sus vecinos, lo cual disminuye el tiempo de cálculos. La complejidad algorítmica para la búsqueda, considerando menos de veinte dimensiones o columnas, es de  $O[D\log(N)]^8$  [6].

El tercer algoritmo es *ball-tree*, basado en un árbol binario completo en que una “bola” (hiper-esfera en el espacio Euclidiano de  $k$ -dimensiones) se encuentra asociado con cada nodo de manera tal que la bola interior de cada nodo contiene a todas las bolas de sus hijos [7]. Esta configuración permite aplicar *k-d tree* en una mayor cantidad de dimensiones, con una complejidad algorítmica de búsqueda  $O(D\log(N))$  para cualquier cantidad de columnas en el conjunto de entrada [6].

Finalmente, dada la descripción previa, es necesario destacar la selección automática del algoritmo, que corresponde a la cuarta opción que se encuentra disponible en el método utilizado. Esta considerará *brute* sólo cuando  $k > \frac{N}{2}$ <sup>8</sup>, mientras que en el resto de los casos, según el conjunto y sus características, escogerá *k-d tree* o *ball-tree* [6]. Es por esto que se utiliza esta opción en este trabajo.

Por último, se debe mencionar el uso del codificador *OrdinalEncoder* contenido en el paquete de preprocesamiento de *Sci-kit Learn* en Python. Su uso se justifica debido a que las etiquetas<sup>9</sup> utilizadas tienen formato de texto, pero el método *KNeighborsRegressor* sólo acepta etiquetas numéricas.

Finalmente, la sección de código en Python que respalda el fundamento es el siguiente (considerando  $x$  como el conjunto de entradas de entrenamiento):

```
# Importando librerías necesarias
from sklearn.neighbors import *
from sklearn.preprocessing import *

# Según lo definido
weights = 'distance'
n_neighbors = 35

# Instancia del método
knn = KNeighborsRegressor(n_neighbors, \
                          weights=weights)

# Codificando conjunto de entrenamiento
enc_train = OrdinalEncoder()
enc_train.fit(x)
x_t = enc_train.transform(x)

# Codificando conjunto de prueba
enc_test = OrdinalEncoder()
enc_test.fit(x_test)
to_test = enc_test.transform(x_test)

# Predicción de prueba
testing = knn.fit(x_t, y).predict(to_test)
```

<sup>9</sup>Una etiqueta corresponde a uno de los valores únicos para cada columna en el conjunto de datos.

Es posible notar en la última línea de código que se ha generado una prueba de predicción utilizando un conjunto de datos de entrada llamado *to\_test* que, en base a lo descrito en la sección de Preparación, corresponde a un subconjunto del set de entrenamiento con datos del año 2016. Cabe destacar que la predicción realizada en este caso, utilizando el método *score(input\_dataset, expected\_output)* de *KNeighborsRegressor*, tiene una precisión aproximada de 74%, lo cual implica que hay puntos en los resultados son mayores o menores, en forma significativa, a los valores esperados, sin embargo, el modelo tiene un nivel aceptable en cuanto a la veracidad de una predicción.

### III-C. Predicción

Ya realizada la parte más extensa, queda estimar la cantidad de víctimas graves y fatales para los años 2018 y 2019. En este punto, cabe destacar que la implementación de una nueva política pública de seguridad de tránsito se basa en la visión cero, por lo cual, el modelo predictivo ya entrenado considerará exactamente los mismos siniestros, con sus respectivas características, que ocurrieron realmente en los años señalados, siendo el resultado a comparar la cantidad de personas afectadas, ya que dicha política busca disminuir considerablemente la cantidad de muertes y lesionados graves (no especificando una importante reducción en los siniestros). En base a esto, es bueno explicitar que, si la implementación mencionada tiene resultados, un siniestro ocurrido por imprudencia del conductor debiese implicar, a lo más, lesiones leves (no consideradas en este estudio), ya que este ha sido más precavido de lo que hubiese sido en una situación en que dicha política no se hubiese implementado.

Explicado el supuesto, al revisar las imágenes se puede notar en 1 hay al menos 8 *peaks* que sobresalen de lo normal, mientras que sólo entre los primeros dos meses hay valores mucho más bajos de lo que refleja la realidad, sin embargo, dado que el modelo posee únicamente esas grandes diferencias, es posible atribuirlos a información dentro del conjunto de entrenamiento que generó etiquetas o vecinos bajo situaciones anormales (notar, por ejemplo, que el máximo ocurre en el año 2019, en el mismo periodo que ocurre el mínimo en 2018). Dado lo anterior, fuera de estos puntos, es posible notar que, en general, la predicción es bastante similar a la realidad, a excepción de las fechas en que se registraron muertes por siniestros viales no consideradas en la predicción.

Al observar el gráfico en 2, es posible notar un mejor seguimiento, en general, de la predicción a lo realmente ocurrido. Sin embargo, nuevamente es posible notar un sector en que la estimación falla, indicando valores considerablemente bajos respecto de la realidad y tal como se vio en el caso anterior, esto ocurre en los primeros dos meses de 2018. Por otro lado, en este gráfico es posible notar que la predicción indica mayor cantidad de *peaks* respecto de los datos registrados en la realidad.

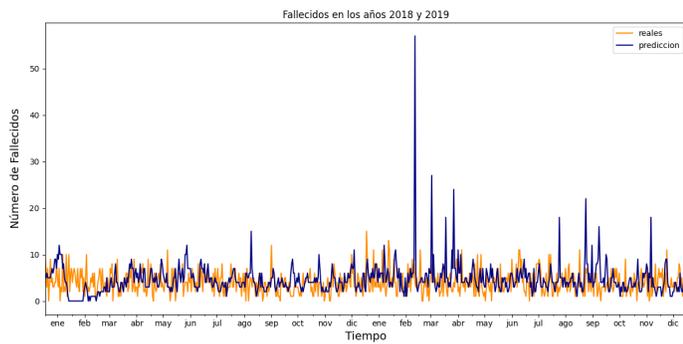


Figura 1. Personas fallecidas estimadas (azul) y reales (naranja) en el periodo 2018-2019. Elaboración propia con datos de CONASET y predicción generada mediante Aprendizaje de Máquina.

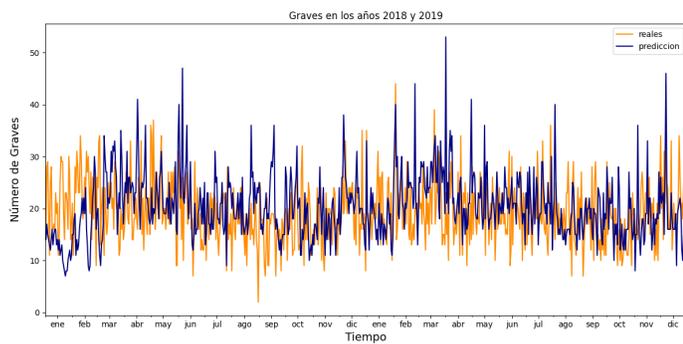


Figura 2. Personas con lesiones graves estimadas (azul) y reales (naranja) en el periodo 2018-2019. Elaboración propia con datos de CONASET y predicción generada mediante Aprendizaje de Máquina.

#### IV. CONCLUSIÓN

Para cerrar el estudio realizado, es necesario responder al objetivo principal planteado, este es, estimar si la aplicación de una nueva política nacional de seguridad de tránsito con visión cero ha generado impactos en los resultados de los siniestros viales ocurridos en los años 2018 y 2019. Para esto, a modo de recuento, se ha generado una estimación de la cantidad de fallecidos y lesionados graves con los datos de los años comprendidos por el periodo 2010-2017, en que dicha política aún no se implementaba. En base a esto, la predicción realizada con una precisión aproximada del 74 %, muestra el resultado de fatalidad y lesiones graves en el último periodo esperados y se comparan con los siniestros registrados en el mismo lapso.

De lo anterior, y revisando los gráficos, es posible notar que la variación en los resultados, salvo los casos mencionados en el ítem anterior, es mínima e, incluso, en la cantidad de fallecidos es posible estimar un menor número de los registrados. Es por esto que, junto con la intuición ciudadana, es posible concluir que el impacto de la nueva política implementada a nivel nacional ha sido nulo en general. Si bien en la cantidad de lesionados graves se ve un resultado con tendencia a ser

positivo, la diferencia no es significativa y tampoco presenta una disminución con el pasar del tiempo. En forma aún más crítica, es posible notar que la cantidad de fallecidos es mayor a la que se podría estimar, lo cual, para una política con visión cero, implica un resultado negativo.

Finalmente, es necesario agregar como comentario que, tal como se ha mencionado en secciones previas, es importante utilizar el conocimiento de ingeniería y tecnología disponible para servir a nuestra sociedad, por lo cual, es imperativo que, con la misma energía que se utilizan los conocimientos para estudiar el comportamiento humano buscando fines comerciales, se desarrollen y liberen las tecnologías para mejorar la calidad de vida de las personas con las que se comparte, por lo menos, el territorio. En este aspecto, si bien no se espera que el Aprendizaje de Máquina o el Análisis de Datos sean un tópico popular en la sociedad, sí existen herramientas para generar medios tecnológicos de transparencia, entre las que se puede destacar *Jupyter Notebooks* para generar un estudio guiado a cualquier persona con conocimientos de computación básicos.

#### REFERENCIAS

- [1] P. F. Luis Ignacio Rizzi, "Alcohol en conducción y su incidencia en la ocurrencia de accidentes de tránsito con víctimas fatales en Chile: Fallencias en las estadísticas nacionales," *Ingeniería de Transporte*, vol. 18, no. 1, pp. 3–9, 2014.
- [2] C. N. de Seguridad de Tránsito, *Política Nacional de Seguridad de Tránsito 2017*. Ministerio de Transportes y Comunicaciones, Octubre 2017.
- [3] S. S. Laura Igual, *Introduction to Data Science: A Python approach to concepts, techniques and applications*. Springer Nature, 2017.
- [4] S. S. Skiena, *The Data Science Design Manual*. Springer Nature, 2017.
- [5] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, pp. 509–517, September 1975.
- [6] scikit-learn developers, "Nearest neighbors."
- [7] S. M. Omohundro, "Five balltree construction algorithms," tech. rep., 1989.

#### SOBRE EL AUTOR

Nací en Santiago el 30 de Septiembre de 1993. Viví en la capital hasta 2012, año en que comencé mis estudios en la Universidad Técnica Federico Santa María, donde ahora realizo mi último año de pregrado. En 2016 empiezo a participar activamente en instancias de innovación tecnológica, formando parte de varios equipos con distintos niveles de desarrollo y éxito. En 2017 inicio la vida del activismo a través de la creación de la Fundación Andes Chile ONG, una organización dedicada a promover el uso de la bicicleta como medio de transporte y la educación vial. Ese mismo año, ingreso al Colectivo Muévete y luego ingreso a la ONG Map8 como encargado del área de innovación.

Fuera de la formalidad, me he dedicado parcialmente a la reparación de computadores a domicilio en sectores populares de Valparaíso y a asesorar emprendimientos en las áreas de tecnologías de información y electrónica general. Además, en estos momentos estoy desarrollando un emprendimiento personal que busca facilitar la digitalización de las PyMES nacionales.