

## Predicción geográfica de delitos contra la propiedad mediante Redes Neuronales y SARIMA

**Fredy Humberto Troncoso Espinosa**

<https://orcid.org/0000-0002-9972-3123>

[froncos@ubiobio.cl](mailto:froncos@ubiobio.cl)

Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad del Bío-Bío Concepción, Chile

**Danilo Alberto Gómez Correa**

<https://orcid.org/0000-0002-8735-7832>

[dagomez@ubiobio.cl](mailto:dagomez@ubiobio.cl)

Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad del Bío-Bío Concepción, Chile

Recibido (18/03/22 ) Aceptado (05/04/22)

**Resumen:** Lograr estimar la cantidad de delitos que se cometerán en una determinada zona geográfica es de importante para la gestión de recursos destinados a la prevención de delitos. Esta investigación desarrolla dos modelos predictivos de series de tiempo para la predicción geográfica de delitos contra la propiedad en dos comunas de Chile, Talcahuano y Hualpén. Los modelos investigados fueron Redes Neuronales y SARIMA. Ambos modelos fueron entrenados y probados con información proporcionada por la Fiscalía Regional del BioBío, Chile. La información contiene los delitos contra la propiedad cometidos en ambas comunas entre los años 2009 y 2019. Los modelos fueron comparados mediante las medidas de error de pronóstico MAPE, MAD y MSE. La comparación de los resultados no evidencian diferencias estadísticamente significativas entre los resultados de ambos modelos.

**Palabras Clave:** Modelo predictivo, Metodología KDD, Series de Tiempo, Redes neuronales, Análisis Criminal.

### Geographic prediction of crimes against property using Neural Networks and the SARIMA model

**Abstract:** Predicting the number of crimes that will be committed in a certain geographical area is important for the management of resources destined for crime prevention. This research develops two predictive models of time series for the geographic prediction of property crimes in two districts of Chile Talcahuano and Hualpén. The models investigated were Neural Networks and SARIMA. Both models were trained and tested with the information provided by the Regional Prosecutor's Office of BioBío, Chile. The information contains the crimes against property committed in both districts between 2009 and 2019. The models were compared using the MAPE, MAD, and MSE forecast error measures. The comparison of the results does not show statistically significant differences between the results of both models.

**Keywords:** Predictive Model , Time Series, Neural Networks, Crime Analytics.



## I. INTRODUCCION

Según la última Encuesta Nacional Urbana de Seguridad (Enusc) la población chilena cree que la delincuencia ha aumentado en el país alcanzando una percepción de inseguridad de un 84,3%. También, informa que durante el año 2020 el 19,2% de los hogares declara haber sido víctima de delitos, siendo esta la menor cifra desde el año 2012 [1]. A pesar de la disminución durante el año 2020 de los delitos, las autoridades confirman que muchos de estos son más violentos que los ocurridos en décadas pasadas. Por otro lado, en la región del Biobío, región en la que se centra esta investigación, los delitos definidos como de “mayor connotación social” (delitos violentos y delitos contra la propiedad) [2], disminuyeron en 4,8 puntos con respecto al 2019, siendo actualmente de un 15,7%, mientras que, la percepción de inseguridad en la población aumentó a un 85,2%.

Para fortalecer la investigación de los delitos de mayor connotación social, la Fiscalía de Chile creó en el marco de la ley N°20.861 “El Sistema de Análisis Criminal y Focos Investigativos (SACFI)”, integrada por fiscales especializados en focos investigativos y analistas criminales, con conocimiento en metodologías de investigación cuantitativa y cualitativa, y en el sistema de justicia criminal [3]. El análisis criminal, para SACFI, es el conjunto de procesos orientados a entregar información oportuna y pertinente en relación a patrones y tendencias de hechos criminales [4].

Para la búsqueda de patrones y tendencia de hechos criminales que apoyen a la toma de decisiones y el análisis de la tasa de criminalidad luego de aplicar medidas preventivas, se ha usado series temporales como ARIMA [5] [6].

Técnicas de Machine Learning también han sido utilizadas para el descubrimiento de patrones y tendencias en bases de datos relacionadas a hechos delictuales [7]. En específico, las redes neuronales han demostrado un buen desempeño en la tarea de predecir las ubicaciones de hechos delictuales [8] [9], siendo necesario para esto la correcta georreferenciación de los hechos criminales [10].

El objetivo de esta investigación fue la búsqueda de un modelo que permitiera predecir la ocurrencia de hechos criminales en forma sectorial en dos comunas de interés de la Rregión del Biobío, Talcahuano y Hualpen. Dado su desempeño predictivo y amplia aplicación al crimen, los modelos considerados fueron el modelo estadístico basado en series

temporales ARIMA y el modelo predictivo basado en redes neuronales aplicado a series de tiempo. La mejor alternativa permitirá tener un pronóstico mensual de delitos en forma sectorizada para las comunas de Talcahuano y Hualpén. Esto permitirá a la Fiscalía tomar decisiones de prevención basadas en la evidencia estadística

Para el entrenamiento y prueba de los modelos, la Fiscalía de la Región del BioBío aportó con una base de datos de los delitos contra la propiedad ocurridos entre el año 2009 y 2019. Esta base de datos contiene el registro de 32.392 crímenes donde se registra, entre otros atributos, fecha del delito, tipo de delito y su georreferenciación.

## II. METODOLOGÍA

Para dar tratamiento a la base de datos y obtener los modelos predictivos, se utilizó la metodología Knowledge Discovery in Databases KDD, definido como: el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos. Esta metodología está compuesta por cinco etapas [11]. La figura 1 muestra el proceso KDD aplicado a la base de datos de delitos entregada por la Fiscalía Regional del BioBío, cuyas etapas son:

- Selección de datos, donde se determinan las fuentes de datos y el tipo de información a utilizar.
- Pre procesamiento de la base de datos, con el fin de tener información más confiable, la que aporte mayor valor a la predicción. Esta etapa incorpora el análisis de datos faltantes, de datos inconsistentes, y el análisis de datos fuera de rango.
- Transformación y selección de variables, que engloba cualquier proceso que modifique la forma de los datos para generar nuevas variables, que enriquezcan la información y obtener un mejor patrón. Luego se procede a seleccionar aquellas variables más importantes.
- Minería de datos, en la cual se aplican las técnicas que permitirán extraer el patrón relevante desde los datos, como las técnicas de machine learning o algoritmos de clasificación. Las técnicas de machine learning aprenden el patrón general oculto en los datos y luego lo utilizan para generar una predicción. La predicción consiste en asignar un registro u observación a una clase o grupo previamente definido.
- Interpretación y evaluación, donde se interpretan los patrones de datos que se descubrieron y se evalúa el impacto del modelo en su futura implementación.

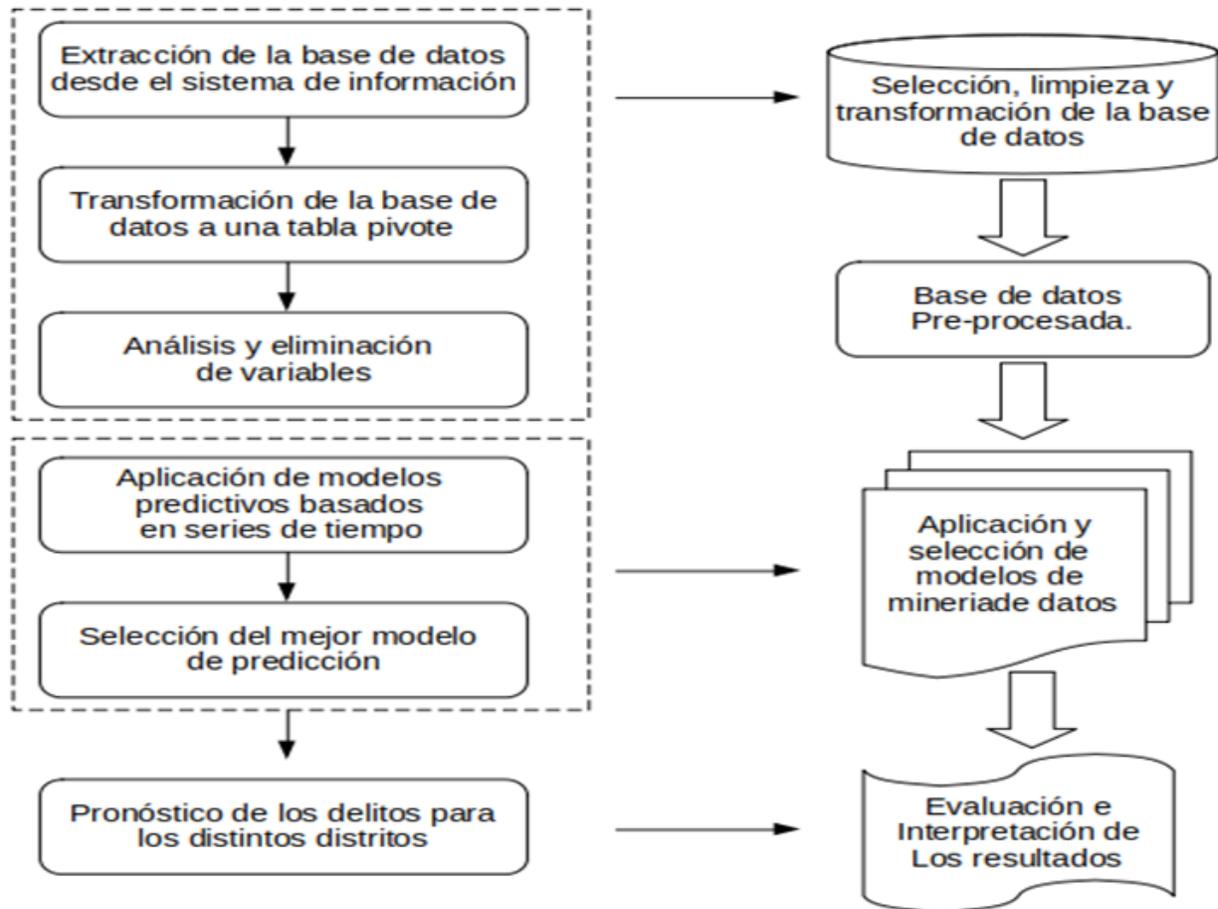


Fig. 1. Procesos de datos dentro de la metodología Knowledge Discovery in Databases KDD.

### III.DESARROLLO

#### A.Selección de datos

Como se mencionó anteriormente, los datos proporcionados por la fiscalía de la Regional del Biobío, corresponden a los delitos registrados en Hualpén y Talcahuano en un periodo de 11 años (2009-2019).

Para la predicción, se consideró una partición del área geográfica correspondientes a las comunas aleda-

ñas Hualpen y Talcahuano según distritos censales como se muestra la Figura 2. El distrito censal constituye la división del territorio comunal que permite organizar el trabajo censal. Esta división en el área urbana se realiza según la cantidad de viviendas, tratando de mantener un número homogéneo por distrito. El total de distritos en la comuna de Hualpen es de 6 y el de la comuna de Talcahuano es de 22.

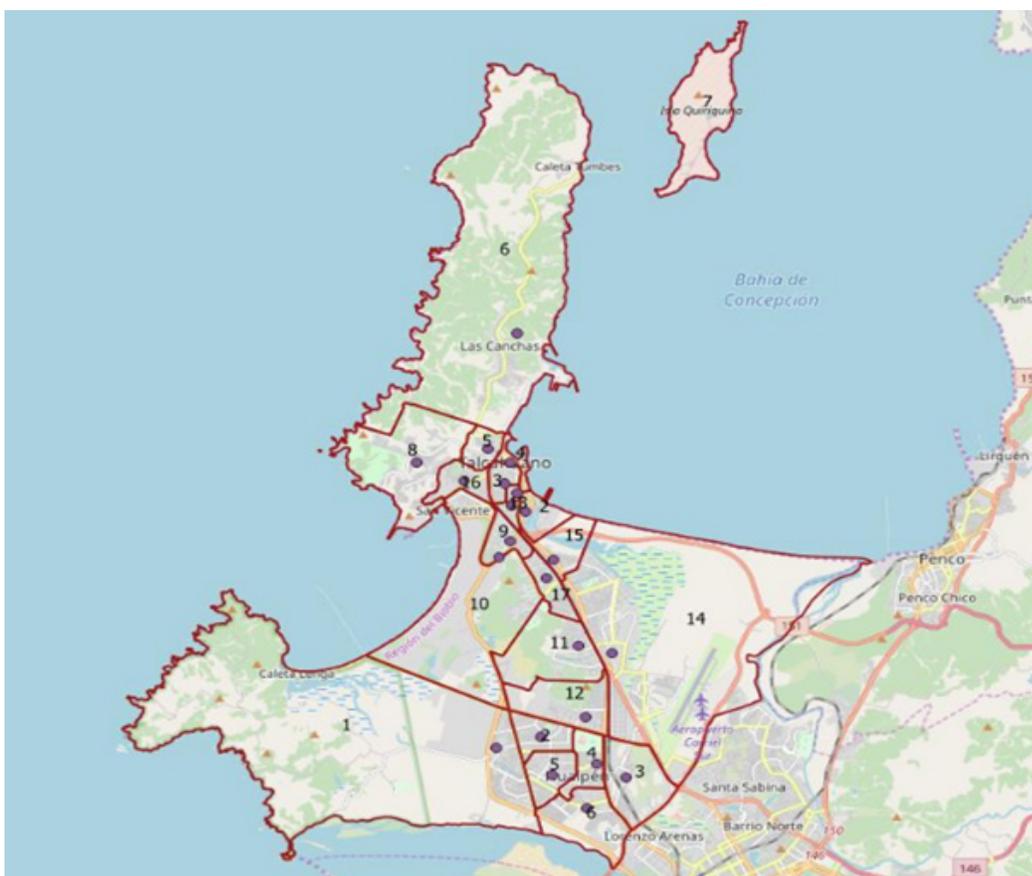


Fig. 2. Distritos censales comuna de Hualpén y Talcahuano.

Luego de la segmentación de los delitos por distrito censal las variables a considerar son las siguientes:

- FIELD 1: Numero de identificación correlativo del registro.

- Fecha del delito, la cual señala de forma numérica el día, mes y año del delito.

- Rol único de causa del delito y el año en el que ocurrió.

- Comuna donde ocurrió el delito.

- Delito: informa el tipo de delito ocurrido, estos pueden ser robo con intimidación, robo con violencia, robo en lugar no habitado, robo en lugar habitado, robo por sorpresa, robo en bienes nacionales de uso público o robo de vehículo motorizado.

- Latitud, que es la coordenada angular que señala la ubicación geográfica de la ocurrencia del delito en el polo sur.

- Longitud, que es la coordenada angular que señala la ubicación geográfica de la ocurrencia del delito en el hemisferio occidental.

- Identificador de distrito censal por comuna, HP significa Hualpén y TH simboliza Talcahuano.

### **B.Transformación y Pre procesamiento del conjunto de datos.**

A partir de la base de datos entregada por la Fiscalía, se creó la Tabla 1 (tabla pivote) que resume el número de delitos cometidos en cada distrito censal de las comunas Hualpén y Talcahuano por cada mes del año entre el 2009 y 2019. Esta tabla permitirá la aplicación de los modelos predictivos.

**TABLA 1. Tabla de datos resultante.**

Nº	MES DELITO	HP1	HP2	HP3	HP4	...	TH1	TH2	TH3	TH4	...	TH14
1	2009-01-01	16	11	10	5	...	2	1	5	8	...	22
2	2009-02-01	19	11	14	5	...	1	1	11	8	...	32
3	2009-03-01	15	16	5	5	...	0	10	5	2	...	32
...	...	...	...	...	...	...	...	...	...	...	...	...
130	2019-10-01	11	21	7	2	...	3	2	7	7	...	47
131	2019-11-01	17	14	12	4	...	1	2	11	6	...	62
132	2019-12-01	18	16	11	3	...	1	1	11	7	...	55

La Tabla 1 muestra un identificador (ID) correspondiente al mes del año y 22 columnas que hacen referencia a cada distrito. Esto permitirá aplicar los modelos a cada distrito censal.

En la etapa de preprocesamiento se realizó el análisis de datos faltantes. Se decidió eliminar la columna de "TH7" ya que no entrega información relevante. Cabe mencionar que el distrito "TH7" corresponde al sector de la Isla Quiriquina, la cual es administrada por la Armada de Chile y es sede de la escuela de grumetes Alejandro Navarrete Cisterna por lo cual se explica la baja ocurrencia de los distintos tipos de delitos.

Para el análisis de los valores atípicos, se utilizó el diagrama de caja y bigotes. Se comprobó visualmente la existencias de valores fuera de rango. Estos valores se identificaron y se imputaron por la técnica K-Nearest Neighbor [12].

#### IV.RESULTADOS

A continuación se muestran la aplicación de los modelos a los datos mostrados en la Tabla 1.

##### A.Minería de datos

Para obtener el pronóstico de los delitos en los distritos de Hualpén y Talcahuano, se entrenó y probó un modelo predictivo basado en redes neuronales aplicado a series de tiempo y un modelo estadístico basado en series de tiempo.

Una serie temporal es una secuencia de  $n$  observaciones ordenadas y equidistantes cronológicamente [13], en otras palabras, son un grupo de datos registra-

dos durante un periodo semanal, trimestral o anual, los cuales en conjunto se utilizan para predecir el futuro. Dentro de las series de tiempo se pueden considerar dos tipos, las series temporales estacionarias (su varianza y promedio no dependen del tiempo) y las no estacionarias (sus propiedades si dependen del tiempo).

Para trabajar las series de tiempo existen distintos modelos estadísticos que se pueden aplicar a ellas, en donde se destacan los modelos de suavizamiento exponencial, modelos autorregresivos integrados de promedios móviles (ARIMA) y modelos de vectores autorregresivos (VAR), en cada uno de estos modelos se debe considerar la posible presencia de componentes de las series temporales, los cuales son tendencia, ciclicidad, estacionalidad y un componente aleatorio. El modelo que se escogió para realizar los pronósticos de predicción es el modelo ARIMA.

Para estimar el modelo ARIMA se utilizó la metodología Box-Jenkins [14]. Esta metodología tiene como objetivo identificar y estimar un modelo estadístico que se interpreta como generador de datos muestrales. Este método consta de cuatro pasos que permiten la identificación de parámetros, estimación de parámetros, exámenes de diagnóstico y finalmente pronóstico. Estos en conjunto permiten conocer si una serie de tiempo sigue un proceso autorregresivo (AR), de medias móviles (MA), autorregresivo y medias móviles (ARMA) o autorregresivo integrado de promedios móviles (ARIMA). Para utilizar esta metodología es importante que la serie de tiempo sea estacionaria. Si no lo es, se debe aplicar las diferenciaciones para que lleguen a serlo y

así consolidar una base válida para un futuro pronóstico [15].

La aplicación de la metodología Box-Jenkins se realizó mediante el lenguaje de programación R [16]. La aplicación de la metodología se describe a continuación:

**Identificación:** En este paso se realizó un análisis exploratorio a las series temporales. Se descubrió todas presentan estacionalidad, por ende, el modelo ARIMA se debe transformar a SARIMA.

**Estimación:** La estimación de los parámetros finalmente escogidos se hizo mediante la comparación de distintos modelos SARIMA, seleccionando el modelo que se ajusta de mejor manera a los datos bajo el criterio AIC (Criterio de información de Akaike).

**Examen diagnóstico:** Una vez seleccionado el mejor modelo, se realizaron las pruebas Jarque-Bera, ARCH y Ljung-Box a los residuos generados por este, con el objetivo de verificar que se ajusten a una distribución normal, al cumplirse cada una de las pruebas se puede dar paso al pronóstico de los datos. Cabe decir que, si las pruebas no se cumplen, se le realizan estos mismos test al siguiente mejor modelo hasta encontrar uno que satisfaga cada uno de los pasos.

**Pronóstico:** Al ser aceptadas cada una de las pruebas realizadas en el paso anterior se realizó la predicción de los delitos.

**Cálculo de los factores de comparación:** Se calculó los factores que se utilizarán en el análisis de los resultados y comprobación de la efectividad del modelo utilizado. En este caso se utilizó el error cuadrático medio, error porcentual absoluto medio y la desviación absoluta media.

Las Redes Neuronales Artificiales (RNA) son sistemas de procesamiento de la información cuya estructura

y funcionamiento están inspirados en las redes neuronales biológicas. Consisten en un conjunto de elementos simples de procesamiento llamados nodos o neuronas conectadas entre sí por conexiones que tienen un valor numérico modificable llamado peso. La actividad que una unidad de procesamiento o neurona artificial realiza en un sistema de este tipo es simple. Normalmente, consiste en sumar los valores de las entradas (inputs) que recibe de otras unidades conectadas a ella, comparar esta cantidad con el valor umbral y, si lo iguala o supera, enviar activación o salida (output) a las unidades a las que esté conectada. Tanto las entradas que la unidad recibe como las salidas que envía dependen del peso o fuerza de las conexiones por las cuales se realizan dichas operaciones [17].

La aplicación de Redes Neuronales Artificiales a la predicción de series de tiempo se realizó de acuerdo a las siguientes etapas utilizando el lenguaje de programación Python 3.0.

**Búsqueda de las variables de entrada:** Se determinó como variables de entrada en la red neuronal, los distritos de las comunas de Hualpén y Talcahuano.

**Preparación del conjunto de datos:** Para esto se escaló el conjunto de datos, lo que consiste en llevarlos a un intervalo [-1, 1]. Luego se utilizó la función “series\_to\_supervised()” de Python 3.0. Esta función toma una serie de tiempo univariante o multivariante y la enmarca como un conjunto de datos de aprendizaje supervisado. Convierte las filas del distrito en varias columnas y entrena la red neuronal con el algoritmo “backpropagation” [18]. La tabla de datos resultante para cada distrito es similar a la que muestra en la Figura 3 que ejemplifica el distrito uno de la comuna de Hualpen, HP1.

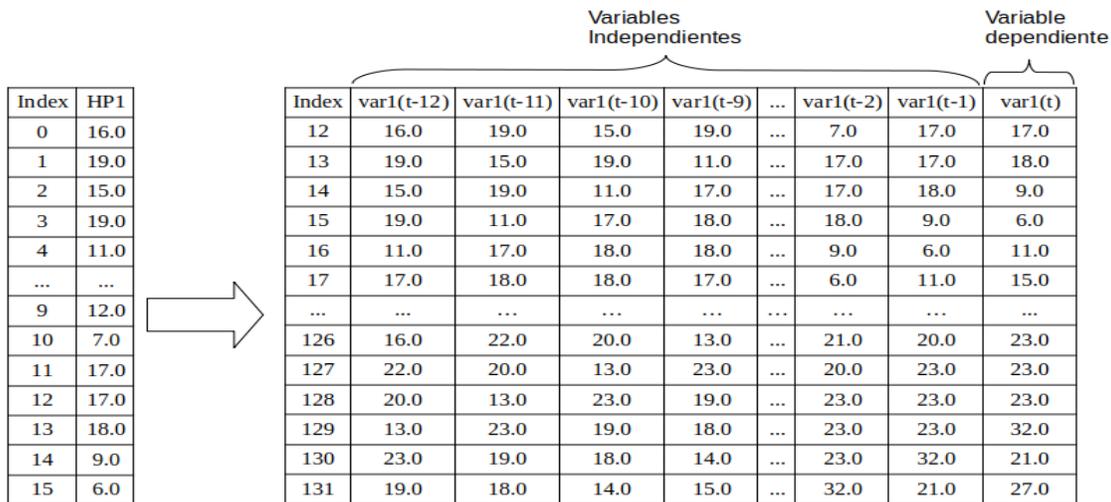


Fig. 3. Conversión de serie de tiempo a problema de tipo supervisado.

En la Figura 3 se muestra 12 variables independientes que corresponden a los delitos cometidos los 12 meses anteriores al mes a predecir, que es la variable dependiente. La fila 12 representa 12 meses de delitos, desde el mes 1 hasta el mes 12 para enseñar a la red a predecir el mes 13. La fila 13 representa 12 meses pero desde el mes 2 hasta el 13 para enseñar a la red a predecir el mes 14. Así hasta llegara a la fila 131 que representa también 12 meses desde el mes 120 al mes 131, para enseñar a la red a predecir el mes 132.

Determinación del conjunto de entrenamiento y conjunto de prueba: Se definieron 108 registros para el entrenamiento y 12 registros para probar el desempeño del modelo como se muestra en la Figura 4.

Los 108 registros de entrenamiento corresponden a los delitos cometidos entre los años 2009 y 20018. Los

12 registros de prueba corresponden a los delitos cometidos el año 2019.

Estructura de Red Neuronal: El modelo apropiado para serie de tiempo es definido por el comando “Secuencial()”. Se definió una arquitectura con 12 neuronas de entrada en forma secuencial, una capa oculta con 12 neuronas y función de activación tangente hiperbólica (dado los datos entre -1 y 1) . Esto se realizó con la función “Dense. Entre la capa oculta y la capa de salida se agregó la función “Flatten()”, la cual redujo la matriz de datos de las 12 neuronas en una fila con los datos por orden de entrada. Posterior a esto se agrega la capa de salida con una sola neurona y la misma función de activación. Como optimizador de la red se utilizó “Adam”.

	Index	var1(t-12)	var1(t-11)	var1(t-10)	...	var1(t-2)	var1(t-1)	var1(t)
Datos de Entrenamiento	12	-0.28571427	-0.07142854	-0.3571428	...	-0.92857146	-0.21428573	-0.21428573
	13	-0.07142854	-0.3571428	-0.07142854	...	-0.21428573	-0.21428573	-0.14285707
	14	-0.3571428	-0.07142854	-0.64285713	...	-0.21428573	-0.14285707	-0.78571427
	...	...	...	...	...	...	...	...
	117	0.38095248	0.07142866	-0.21428573	...	0.1428572	0.0	-0.5
	118	0.07142866	-0.21428573	0.2857144	...	0.0	-0.5	0.21428573
	119	-0.21428573	0.2857144	-0.28571427	...	-0.5	0.21428573	-0.07142854
Datos de Test	120	0.2857144	-0.28571427	0.35714293	...	0.21428573	-0.07142854	-0.14285707
	121	-0.28571427	0.35714293	0.21428573	...	-0.07142854	-0.14285707	-0.42857146
	...	...	...	...	...	...	...	...
	129	-0.5	0.21428573	-0.07142854	...	0.21428573	0.21428573	0.8571429
	130	0.21428573	-0.07142854	-0.14285707	...	0.21428573	0.8571429	0.07142866
	131	-0.07142854	-0.14285707	-0.42857146	...	0.8571429	0.07142866	0.5

Fig. 4. Subdivisión de la base de datos en entrenamiento y test.

Entrenamiento: en esta etapa se definen 22 épocas de entrenamiento. Se considera el algoritmo de entrenamiento supervisado, que ajusta los pesos sinápticos mediante la minimización del error, “backpropagation”.

Prueba: Se presenta a la red el conjunto de datos seleccionados para este fin y se obtienen los valores de la predicción del siguiente periodo para cada distrito.

Cálculo de los factores de comparación: Se utilizó el error cuadrático medio, error porcentual absoluto medio y la desviación absoluta media.

La red neuronal fue entrenada y testada con la base de datos sin aplicar transformación para volver estacionarias las series de tiempo dado a que se comprobó la

generación de mejores resultados.

**V.RESULTADOS**

A continuación, se presentarán los resultados de las métricas de error de pronóstico MAPE, MAD y MSE para ambos métodos. Para ejemplificar los resultados se seleccionó los tres distritos con mayor cantidad de delitos de las comunas de Hualpen y Talcahuano.

La Tabla 2, muestra la métrica de error de pronóstico MAPE de los 6 distritos censales seleccionados. Esta métrica nos indica la exactitud de la predicción como un porcentaje del error. En 4 de los 6 distritos, SARIMA presenta un menor error en sus pronósticos.

**TABLA 2. Comparación de MAPE del Modelo Estadístico y Red Neuronal.**

Modelo	HP1	HP2	HP6	TH1	TH9	TH14
Red Neuronal	21,55%	63,05%	19,77%	35,54%	36,09%	20,54%
SARIMA	25,5%	45,8%	39,83%	28,6%	35,8%	17,5%

La Tabla 3, muestra la métrica de error de pronóstico MAD de los 6 distritos censales seleccionados. Esta métrica indica exactitud en las mismas unidades que los datos, valores pequeños indican un mejor ajuste.

**TABLA 3. Comparación de MAD del Modelo Estadístico y Red Neuronal.**

Modelo	HP1	HP2	HP6	TH1	TH9	TH14
Red Neuronal	4,08	6,75	3,91	1,66	3,83	9,41
SARIMA	5,72	5,15	6,27	1,9	3,7	8,4

La Tabla 4, muestra la métrica de error de pronóstico MSE de los 6 distritos censales seleccionados. Esta métrica mide la exactitud de los valores ajustados, sirve para cuantificar la precisión de una técnica de predicción. La red neuronal obtuvo en general una mejor exactitud de los valores ajustados. De la comparación, 4 de los 6 distritos observados con menor MSE corresponden a la red neuronal.

**TABLA 4. Comparación de MSE del Modelo Estadístico y Red Neuronal.**

Modelo	HP1	HP2	HP6	TH1	TH9	TH14
Red Neuronal	27,9	56,08	40,08	4,5	19,16	106,08
SARIMA	55,55	34,85	58,89	5,9	19,6	113,9

Con el fin de encontrar diferencias estadísticamente significativas en las métricas resultantes y tener mayor certeza sobre que modelo presenta mejor desempeño, se realizó un aprueba de medias. Para elegir la mejor prueba se comprobó la normalidad de los datos mediante la prueba de Shapiro Wilk y la homogeneidad de varianzas mediante la prueba de Barlett. Estos resultados se muestran en la Tabla 5.

**TABLA 5. Resultados test estadísticos.**

Prueba	Valor-p MAPE	Valor-p MSE	Valor-p MAD
Shapiro - Wilk Red Neuronal	0,008187	5,707E-05	0,003065
Shapiro - Wilk SARIMA	0,5357	0,0001968	0,02684
Barlet	0,02545	0,9911	0,8041
Kruskal - Wallis	0,511	0,9626	0,7692

En Tabla 5, los valores -p para la prueba de Shapiro-Wilk muestra que los errores no se comportan normal, excepto MAPE para SARIMA. La prueba de Bartlett muestra que solo para MSE y MAD las varianzas son homogéneas. Dado a que no se presenta normalidad y homogeneidad

neidad de la varianza en forma conjunta, para probar si existen diferencias significativas según cada medida de error de la Red Neuronal y SARIMA se aplica la prueba no paramétrica de Kruskal - Wallis [19]. El valor-p mayor a 0,05 para cada tipo de error muestra que no existe evidencia significativa para señalar que los resultados entregados por ambos modelos son diferentes.

Si observamos los promedios de las medidas de medición del error de pronóstico MAPE, MAD y MSE de los dos modelos empleados en esta investigación, ver tabla 6. Podemos observar que el modelo SARIMA es

ligeramente más preciso que la red neuronal, sin embargo, ambos modelos presentan una varianza considerable en sus medidas de error de pronóstico. De acuerdo a los resultados de los test estadísticos, no existe evidencia estadísticamente significativa para señalar que un modelo presenta mejores rendimientos en sus predicciones. Ambos modelos pueden ser aplicados por Fiscalía de la Región del BioBio para predecir el número de delitos contra la propiedad en los sectores geográficos estudiados.

**Tabla 6. Promedios métricas totales para ambos modelos**

Modelo	MAPE	MAD	MSE
SARIMA	41.87% +- 14.43	3.49 +-1.98	25.45 +-27.83
Red Neuronal	43.33% +- 18.29	3.63 +-2.09	25.35 +-27.76

## VI.CONCLUSIONES

Los modelos creados en esta investigación entregan resultados variables en términos de las medidas de medición del error de pronóstico MAPE, MAD y MSE. Las variaciones y falta de precisión pueden deberse a que la variable temporal es solo unos de los elementos que explica el crimen. Hay otros como la oportunidad en el hecho criminal, variables macroeconómicas y a algunos factores de impacto social como el estallido social ocurrido en Chile durante el año 2019.

Sin embargo, se determinó con los tests estadísticos aplicados, que no es posible asegurar diferencia significativa en que un modelo tiene mejor desempeño que el otro. Es decir, el fenómeno es posible modelarlo con estadística clásica o algoritmos de machine learning.

Se recomienda el uso de Redes Neuronales dada la simpleza en su implementación. Esta simpleza radica en que no tienen supuestos que validar como en el caso de los modelos ARIMA.

Este modelo puede ser utilizada como una herramienta de apoyo a la toma de decisiones de Unidad de Análisis Criminal de la Fiscalía Regional del Biobío predecir delitos y establecer planes de acción para cada distrito censal sobre todo en los meses donde los pronósticos indican un mayor aumento de los delitos.

Se recomienda sobre todos la creación de estos planes en los sectores en donde los errores fueron bajos.

## RECONOCIMIENTO

Se agradece el apoyo financiero otorgado por la Universidad del Bío-Bío, Chile, mediante el Proyecto de Iniciación en Investigación código 2060204IF/I. Se agradece también el apoyo técnico y los datos facilita-

dos por la Unidad de Análisis Criminal de la Fiscalía Regional del Biobío, Chile mediante un acuerdo de colaboración.

## REFERENCIAS

- [1]ENUSC, « Síntesis Delictual ENUSC 2020,» PAÍS, 2020.
- [2]AMUCH, «Delitos de mayor connotación social en las comunas de Chile,» Santiago, 2018.
- [3]Fiscalía de Chile, «Fiscalía de Chile,» Abril 2018. [En línea]. Available: [http://www.fiscaliadechile.cl/Fiscalia/fiscalias\\_lagos/noticias\\_det.do?id=14462](http://www.fiscaliadechile.cl/Fiscalia/fiscalias_lagos/noticias_det.do?id=14462). [Último acceso: Diciembre 2021].
- [4]C. Á. Vargas, «Sistemas Judiciales,» Mayo 2020. [En línea]. Available: [https://sistemasjudiciales.org/wp-content/uploads/2020/05/Analisis\\_Criminal\\_Chile.pdf](https://sistemasjudiciales.org/wp-content/uploads/2020/05/Analisis_Criminal_Chile.pdf). [Último acceso: Diciembre 2021].
- [5]P. Chen, Y. Hongyong y S. Xueming, «Forecasting Crime Using the ARIMA Model,» 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 627-630, 2008.
- [6]E. A. Anderson y J. Díaz, «Using Process Control Chart Techniques to Analyse Crime Rates in Houston, Texas,» The Journal of the Operational Research Society, vol. 47, nº 7, pp. 871-881, 1996.
- [7]F. H. T. Espinosa, «Prediction of Recidivism in Thefts and Burglaries Using Machine Learning,» Indian Journal of Science and Technology, vol. 13, p. 696 – 711, 2020.
- [8]P. P. Ascencio, «Desarrollo de un modelo predictivo de robo a casas basado en redes neuronales,» Santiago, 2020.
- [9]R. F. Reier Forradellas y S. L. Nãñez Alonso,

«Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction,» Social Sciences, 2021.

[10]F. H. Troncoso Espinoza y N. E. Fernández Rozas, «Limpieza, corrección y geocodificación de grandes bases de datos de direcciones mediante minería de texto,» *uct*, vol. 25, pp. 80-87, junio 2021.

[11]U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, *Advances in Knowledge Discovery and Data Mining*, American Association for Artificial Intelligence, 1996.

[12]S. Zhang, «Nearest neighbor selection for iteratively kNN imputation,» *Journal of Systems and Software*, vol. 85, n° 11, pp. 2541-2552, 2012.

[13]J. A. Mauricio, *Introducción al Análisis de Series de Tiempo*, Madrid: Universidad de Madrid, 2007.

[14]V. Yakovyna y O. Bachkai, «The Comparison of Holt - Winters and Box - Jenkins Methods for Software Failures Prediction,» *COLINS*, 2018.

[15]D. Gujarati y D. Porter, *Econometría*, México D.F.:

McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V, 2010.

[16]R. krispin, *Hands-On Time Series Analysis with R: Perform time series analysis and forecasting using R*, Birmingham: Packt, 2019.

[17]D. Mercado y L. Pedraza, «Comparación de Redes Neuronales aplicadas a,» *PROSPECTIVA*, vol. 13, n° 2, pp. 88-95, 2015.

[18]J. Brownlee, «Machine Learning Mastery,» 8 Mayo 2017. [En línea]. Available: <https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/>. [Último acceso: Diciembre 2021].

[19]E. Ramalle-Gómara y J. Andres de Llano, «Utilización de métodos robustos en la estadística inferencial,» *Atención Primaria*, vol. 32, n° 3, pp. 177-181, 2003.

[20]C. Chatfield, «The Holt-Winters Forecasting Procedure,» *Applied Statistics*, vol. 27, n° 3, pp. 264-279, 1978.

## RESUMEN CURRICULAR



**Fredy Troncoso Espinosa**, Doctor en Sistemas de Ingeniería, Universidad de Chile, Ingeniero Civil Industrial Universidad del Bío-Bío, Chile. Académico e Investigador Departamento de Ingeniería Industrial, Universidad del Bío-Bío. Concepción, Chile



**Danilo Gómez Correa**, Magister en Estadística Aplicada, Universidad de Concepción, Profesor de Matemática Universidad de Concepción, Chile. Académico tiempo parcial Universidad del Bío-Bío. Concepción, Chile y consultor de empresas.