

# PREDICCIÓN DE ACCIDENTES DE TRÁNSITO EN LA REGIÓN METROPOLITANA DE CHILE BASADO EN REDES NEURONALES

Cecilia Montt , Nibaldo Rodríguez, Alejandra Valencia, Álvaro Contreras  
Pontificia Universidad Católica de Valparaíso. Chile  
[cmontt@ucv.cl](mailto:cmontt@ucv.cl); [nibaldo.rodriguez@ucv.cl](mailto:nibaldo.rodriguez@ucv.cl); [alejandra.valencia@ucv.cl](mailto:alejandra.valencia@ucv.cl)

## RESUMEN

Los accidentes de tránsito están relacionados con pérdidas de vidas humanas y personas lesionadas. En este estudio se diseña un sistema basado en redes neuronales artificiales para pronosticar accidentes de tránsito con resultado grave y menos grave. La información recolectada contiene causas que explican el modelo, por lo que sería posible identificar las causas que han producido los accidentes en estudio. De la predicción se concluye que el modelo es efectivo, pues entregó un 93% de la varianza explicada con un MAPE (Porcentaje de Error Medio Absoluto) del 7% y el 92% de los datos con error menor al 10%.

*Palabras claves: Seguridad Vial, Redes Neuronales, Predicción*

## ABSTRACT

Traffic accidents are related to loss of human lives and injured. In this study a system is designed on artificial neural networks (ANN) to predict traffic accidents with serious and less serious outcome. The information collected contains the causes of the model, it would be possible to identify the causes of accidents have been studied. Prediction is concluded that the model is effective, because it gave a 93% variance explained along with ASM 7% and 92% of the data have an error less than 10%.

*Keywords: Road Safety, Neural Networks, Prediction*

## 1. INTRODUCCION

Los accidentes son eventos complejos y aleatorios en los que se involucra una variedad de factores ya sean éstos humanos, ambientales o propios de la mecánica de los vehículos involucrados. Por lo tanto, identificar los factores relevantes que influyen en los accidentes de tránsito y predecir la cantidad de éstos que ocurrirán durante una ventana de tiempo, resulta ser una herramienta de gran ayuda al momento de llevar a cabo planes de seguridad vial y evitar que este tipo de siniestros siga en aumento.

La Organización Mundial de la Salud (OMS) resalta que la mayoría de las colisiones en las que las personas se ven involucradas son predecibles y prevenibles. En la actualidad los accidentes representan la octava causa de muerte en el mundo y la tendencia indica que el 2030 ésta será la quinta causa de fallecimiento (OMS, 2009).

En Chile los accidentes de tránsito es un tema preocupante, ya que pese a las diversas campañas éstos siguen teniendo cifras alarmantes y complejas de abordar. Por ejemplo el año 2014 los accidentes de tránsito alcanzaron una cifra de 1630 muertes, siendo el número de víctimas más significativas de los últimos 6 años. El año 2013 los accidentes arrojaron un 2,7% de muertos y un porcentaje de 19,2 % de graves y menos graves, siendo los atropellos, colisiones y choques los mayores porcentajes de siniestros (CONASET, 2015). El problema es que estos eventos no solo tienen consecuencias para quienes los protagonizan sino que también en sus familias y tienen importantes repercusiones en el aumento del gasto público y privado asociado al despliegue logístico y asistencial en que incurrir.

La mayoría de los trabajos que han estudiado los accidentes de tránsito, han tratado de clasificarlos, en base a la cantidad de lesionados, heridos, ilesos y muertos en distintas condiciones. Algunas aproximaciones de estimación de accidentes de tránsito que los autores han venido desarrollando desde hace algunos años son a través de Redes Bayesianas (Montt et al, 2010) o el uso de redes neuronales artificiales del tipo Perceptrón Multicapa (MLP) y máquinas de soporte vectorial (SVM) con LS -SVM (Montt et al, 2011). Se ha observado que el uso de redes neuronales artificiales en estimación de predicción de accidentes de tránsito es reducido y su enfoque depende plenamente de cómo el investigador construya el modelo. Como antecedente previo se tiene que Rubio et al (2013), construyeron un modelo de predicción de la cantidad de personas lesionadas y fallecidas en accidentes de tránsito, utilizando las redes neuronales con algoritmos de optimización por enjambre de partículas (PSO) y Levenberg-Marquart (LM), para la V región.

Dado lo anterior el objetivo de este trabajo es predecir el número de personas graves y menos graves de accidentes de tránsito en la Región Metropolitana, para el período Enero/2003 a diciembre/2012 (CONASET, 2013), a través de redes neuronales utilizando el algoritmo de LM y MLP. Se espera con esta herramienta analizar la posibilidad de predecir el resultado de los accidentes en las personas. Esta herramienta será de ayuda tanto al sector público como privado, en el primer caso a Carabineros de Chile para implementar planes acotados a determinados tipos

de conducta y situaciones riesgosas y a los privados en temas de seguros y ahorros en cuantificación de accidentes.

## 2. ESTADO DEL ARTE

Numerosos trabajos se han abocado al estudio de los accidentes de tránsito, muchos de ellos tratando de clasificarlos, en base a la cantidad de lesionados, heridos, ilesos y muertos en distintas condiciones. A modo de ejemplo, encontramos el modelo logarítmico (Shou-Ren et al, 2010) para estudiar los factores claves para determinar la gravedad de los participantes en accidentes producidos en cruces ferroviarios o el de Hovden et al, (2011) quienes abordaron los factores o indicadores que potencialmente apoyan o contribuyen al aprendizaje multinivel de procesos para la identificación de factores que producen los accidentes tanto aéreos, marítimos como ferroviarios, con lo que nos evidencian que la preocupación por este tipo de siniestros es a nivel global.

Para predicción con redes neuronales artificiales Kunt et al (2011), Dia y Rose (1997) y Bedard et al (2002) son los primeros que desarrollaron modelos con mayor información del accidente como: edad y género del conductor; el uso del cinturón de seguridad; el tipo y la seguridad del vehículo; las condiciones meteorológicas; la superficie de la carretera; la relación de velocidad; tiempo de caída; el tipo de accidente; tipo de colisión y el flujo de tráfico. Dia y Rose (1997) por su parte utilizó datos del mundo real para comparar la técnica red neuronal MLP, con un modelo heurístico de detección de accidentes en la autopista de Melbourne. Los resultados mostraron que el modelo de la red neuronal era más fiable, ya que podía detectar más rápido los accidentes sobre el modelo que estaba en operación de las autopistas. El último de ellos Bedard et al (2002) aplicaron una regresión logística para determinar la independencia entre la contribución del conductor, choque y características de los vehículos. Los autores encontraron que la prevención de fatalidades radicaba en el mayor uso del cinturón de seguridad, en la reducción de velocidad y la reducción de impactos del lado del conductor.

Wang et al (2011), utilizó un modelo de predicción de accidentes de tránsito en base a la lógica difusa. El resultado de la prueba muestra que hay una buena relación entre los números observados y los números estimados por el modelo, lo que permite según el autor, una herramienta para la mejora de la gestión de la seguridad del tráfico.

Deublein et al (2013) presenta una metodología para la predicción de la ocurrencia de los accidentes de tráfico, esta utiliza una combinación de tres métodos estadísticos: gamma análisis de regresión de Poisson y logarítmica normal multivariante jerárquica. Se demuestra que con la metodología propuesta es posible el desarrollar modelos para estimar la incidencia de los accidentes de tráfico de cualquier red carretera, siempre que se disponga de los datos necesarios.

Otra aplicación de inteligencia artificial es en Barba et al (2014), donde se predice accidentes de tránsito con regresión funcional, para la V región de Chile.

### 3. METODOLOGIA

De la base de datos entregada por CONASET (2013), se estudia primero la información por semana las cuales corresponden a 681, correspondientes a las semanas en que ocurrieron los accidentes, donde la información contiene el número de personas lesionadas con resultado de lesiones graves y menos graves en accidentes ocurridos en la región Metropolitana de Chile entre los años 2003 y 2012.

La información recolectada contiene las causas que explicaran el modelo para predecir los accidentes. Cada tipo de accidente estudiado, contiene información referente primero al accidente que comprende el identificador del este, fecha, hora de accidente, comuna, causas, tipo de accidente, estado atmosférico, condición calzada, tipo de calzada estado calzada, urbano/rural, además si la persona resulto con lesiones graves y menos graves, también contiene el sexo, calidad (peatón, conductor, pasajero), edad, y el tipo de vehículo, si es particular o de servicio. En cada accidente puede haber una o más personas involucradas y uno o más vehículos involucrados.

El modelo utilizado en este trabajo es un modelo autorregresivo univariado, es decir depende solo de **m-muestras previas** de accidentes con resultados graves y menos graves. Cada uno de esos accidentes está compuesto por la información descrita en el párrafo anterior, es decir, sería posible estimar el número de accidentes de tránsito. Sin embargo, este estudio no considera un análisis de las variables exógenas más correlacionados con los accidentes, pero es posible realizar un análisis de frecuencias para dichas causas. Por ejemplo en un accidente se identifican por ejemplo la hora; causa que puede ser “no mantener la distancia prudente y razonable”; edad de la persona; estado en que quedo en este caso grave, para realizar este modelo de dichos dato solo se consideró el estado en que quedo la persona, es decir, grave o menos grave.

#### 3.1 Análisis estadístico de las causas de accidentes.

En la siguiente figura se muestra el porcentaje promedio de las causas más frecuentes para los accidentes graves y menos graves durante los años del estudio.

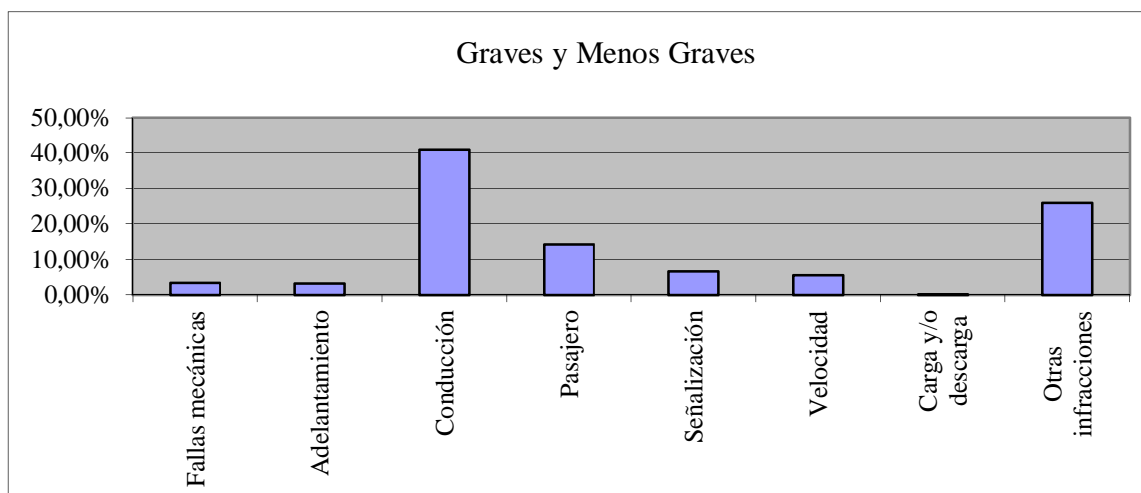


Figura 1 Causas más frecuentes de los accidentes graves y menos graves

Como se visualiza en la figura las causas con mayor frecuencia son la conducción y pasajero.

Las componentes de la causa conducción y su porcentaje promedio en los años de estudio son:

Conducción	Porcentaje
Conducir no atento condiciones transito momento	26,00%
Conducir sin mantener distancia razonable ni prudente	22,44%
No respetar derecho. Preferente paso peatón	13,17%
Conducir en estado de ebriedad	11,89%
No respetar derecho. Preferente paso vehículo	8,29%
Cambiar sorpresivamente pista circulación	7,51%
Conducir bajo la influencia del alcohol	5,51%
Conducir contra sentido del transito	2,60%
Conducir condiciones físicas deficientes (cansancio, sueño)	2,01%
Conducir por el eje izquierdo de la calzada	0,51%
Conducir bajo influencia de drogas o estupefacientes	0,06%
No respetar derecho preferente de paso	0,00%

Tabla 1: Porcentaje promedio de las causas conducción

El comportamiento de las causas más frecuentes de la conducción durante los años de estudio es:

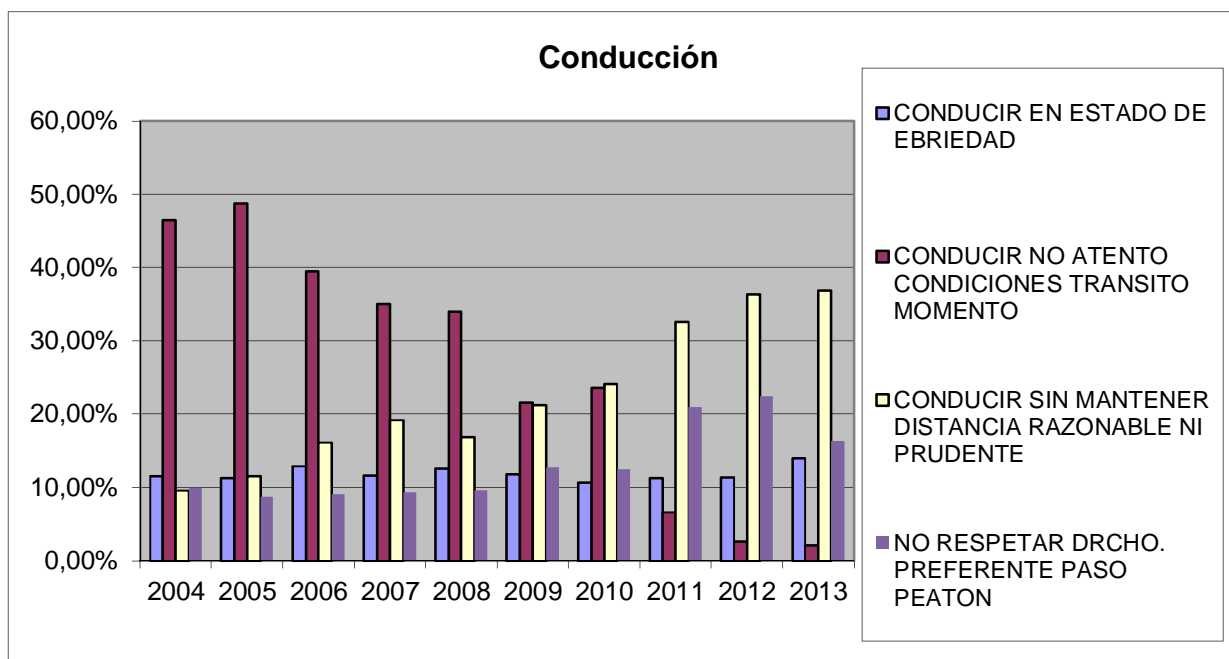


Figura 2: Causas anuales más frecuentes de la conducción

Ahora, las causas más recurrentes de accidente, tanto para graves y menos graves, para la causa pasajero son los siguientes:

Pasajero	Porcentaje
Peatón cruza en forma sorpresiva o descuidada	44,55%
Imprudencia, del peatón	22,01%
Peatón cruza calzada fuera paso peatones	10,21%
Ebriedad del peatón	8,16%
Peatón permanece en la calzada	4,73%
Pasajero sube o baja de vehículo en movimiento	4,57%
Imprudencia del, pasajero	2,34%
Pasajero viaja pisadera vehículo	1,52%
Peatón cruza camino o carretera sin adoptar precauciones	1,38%
Ebriedad del pasajero	0,52%
Otras peatón	0,00%

Tabla 2: Porcentaje promedio de las causas pasajero

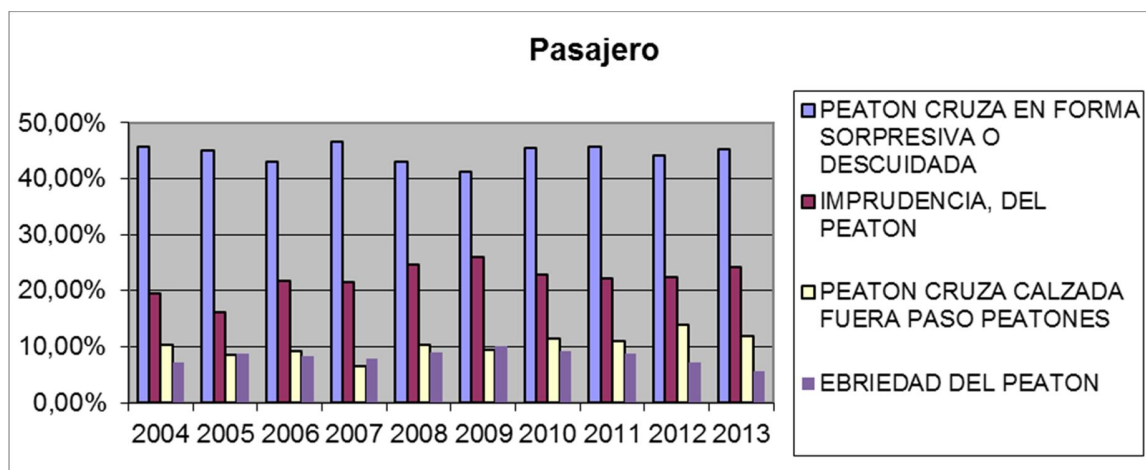


Figura 3: Causas anuales más frecuentes del pasajero

Se destaca que la variabilidad de las frecuencias mostrada en los gráficos anteriores son las posibles causas subyacentes que generan los accidentes en estudio, donde la causas con mayor frecuencia para el caso conducción son “Conducir no atento condiciones transito momento” y “Conducir sin mantener distancia razonable ni prudente”, para el caso del pasajero son “Peatón cruza en forma sorpresiva o descuidada” y “Imprudencia, del peatón”

### 3.2 Predicción.-

La metodología utilizada se basa en las Redes Neuronales Artificiales (RNA), que son un método inspirado en la neurofisiología para las tareas de clasificación, identificación, diagnóstico predicción. Las RNA tienen la capacidad de adaptarse dinámicamente modificando los pesos sinápticos (parámetros) de sus interconexiones.

Predecir es el conocimiento anticipado de lo que ocurrirá en el futuro, basándose en el conocimiento del pasado. Por ejemplo, lo que ocurre a diario en el ámbito de la salud cuando un equipo médico en base a ciertos síntomas podrá saber el desarrollo previsible de una enfermedad. La predicción tiene un fuerte impacto en la toma de decisiones en negocios, predicción de tiempo, fallas, en la bolsa de acciones, desgaste de piezas, etc. Para este estudio las redes neuronales se utilizan para predecir, lo que significa que dado un conjunto de  $n$  datos, el objetivo es pronosticar el número de accidentes para el período siguiente (en este caso una semana), por ejemplo en el tiempo.

Las redes neuronales para predecir los accidentes de tráfico, se procede en primer lugar al análisis y a la normalización de los datos. Posteriormente se inicia el proceso para la estimación, para lo cual primero se calibran y se testean los datos y luego éstos últimos se pasan por el método de suavizado. Una vez realizada dichas etapas se procede a evaluar con el modelo propuesto para utilizar la red neuronal, que en este caso es el de Levenberg Marquardt que es un algoritmo que tiene como propósito entrenar los datos mediante un gradiente de segundo orden sin tener que computar la matriz Hessiana (Mejía, 2004). El algoritmo de Levenberg Marquardt es una

modificación del método de Newton, el que fue diseñado para minimizar funciones que sean la suma de los cuadrados de otras funciones no lineales; es por ello, que tiene un excelente desempeño en el entrenamiento de redes neuronales.

Para trabajar los datos se utiliza una variación del enfoque tradicional, donde después de probados los datos se pasan directamente a la etapa de entrenamiento de la red con el algoritmo Levenberg-Marquart (LM) y Multilayer Perceptron (MPL) (ver Figura 4), como se ilustra en la figura los datos de entrada  $x_i$  pasan al entrenamiento de la red y la comparación de resultados  $y_{(n)}$ . Sin embargo en este trabajo se introduce una técnica de suavizado (smooth) a la data de entrada, previo a la aplicación de los algoritmos de LM y MPL, lo que permitiría una generalización mayor a la red neuronal, sin embargo se compara el resultado con la información en bruto obtenida. Este método se esquematiza en la Figura 5.

En este caso, X corresponde al accidente con resulta de lesiones graves o ilesos, y el resultado Y es la estimación de estos.

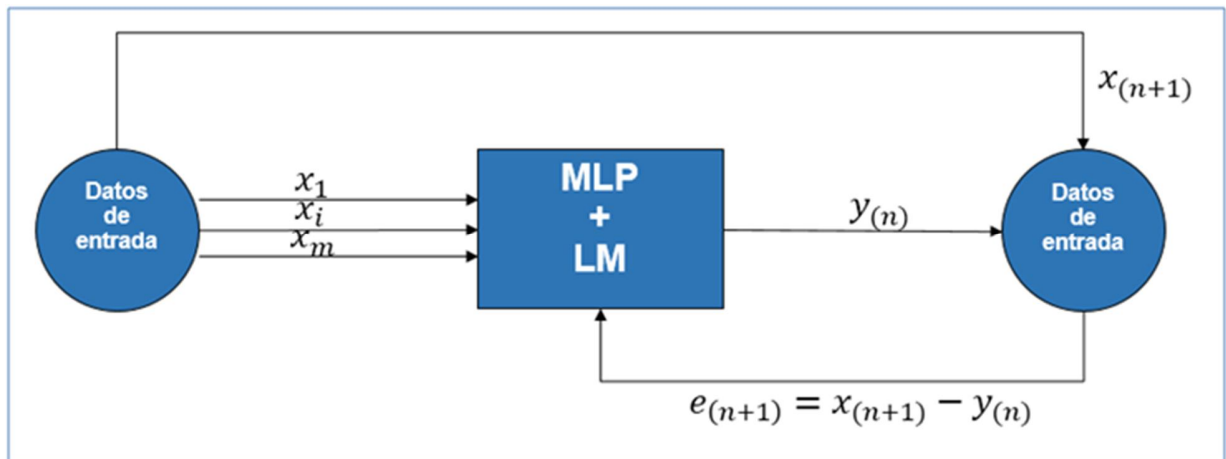


Figura 4: Método de aprendizaje tradicional

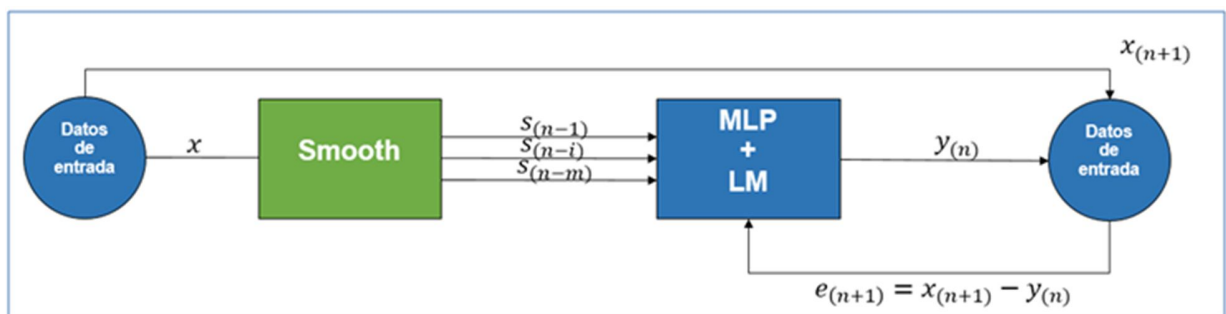


Figura 5: Método de aprendizaje propuesto

### 3.3 Métricas de Evaluación.

Para la selección de la topología y, por consiguiente de los modelos que presentan mejores resultados, en cuanto al porcentaje de datos de entrenamiento, se utilizaron un conjunto de



métricas de exactitud calculadas entre los datos observados (valores reales) y los datos pronosticados (obtenidos por medio de la red). Estas métricas se nombran a continuación:

- Raíz del Error Cuadrático Medio (RMSE). Consiste en la suma de las diferencias entre los datos observados y los datos proyectados por el modelo.

$$RSME = \sqrt{\frac{\sum_{i=1}^n (d_i - y_i)^2}{N}}$$

- Coeficiente de Determinación. Mide la dependencia entre los datos reales y los pronosticados. El 0 indica dependencia y el 1 lo contrario.

$$R^2 = 1 - \frac{\sum_{i=1}^n (d_i - y_i)^2}{\sum_{i=1}^n (d_i - \bar{y})^2}$$

- Porcentaje de Error Medio Absoluto (MAPE). Proporciona una indicación de que tan grandes son los errores de pronóstico comparados con los valores reales de la serie.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{d_i - y_i}{d_i} \right|}{N} \times 100 \quad d_i \neq 0$$

- Error Relativo. Nos presenta una indicación de cómo se comporta el porcentaje de error en los datos obtenidos

$$RR = \frac{d_i - y_i}{\frac{y_i + \bar{y}}{2}} \times 100$$

Se define para las formulas presentadas a  $d_i$  al valor observada en la semana  $i$ ,  $y_i$  al valor pronosticado en la semana  $i$ ,  $\bar{y}_i$  como la media de la data observada y  $N$  como el número total de semanas computadas.

#### 4. RESULTADOS

Para trabajar los datos estos se agruparon por semanas y se consideraron sólo aquellos en los cuales hubo como resultado personas con lesiones Graves y Menos Graves. Por lo que se decide trabajar con un porcentaje del total de datos para entrenamiento de la red 70%, 75% y 80%, es decir con estos datos se predice los datos siguientes es decir el, 30%, 25% y 20%. , donde los resultados son:

Luego del análisis de los tres modelos propuestos, se puede concluir que el modelo con suavizado de la data es efectivo, pues nos entregó en dos de los tres casos sobre el 93% de la varianza explicada con un mape bordeando el 7% además podemos afirmar que el 92% de los datos tendrá un error menor al 10% aplicable a todos los modelos.

Para la etapa de entrenamiento en este resumen se muestran los resultados con un 70% los datos, que entrego los mejores resultados para predecir. Ver Tabla 3.

Métricas de evaluación	70% de los datos para la fase entrenamiento	
	Mejor	$\bar{X}$
RMSE	0,0181	0,01862
MAE	0,0128	0,01334
MAPE(%)	7,3041	7,81721
R <sup>2</sup> (%)	93,9594	93,8229

Tabla 3: Métrica de evolución de las predicciones

Cada uno de los modelos fue ejecutado 30 veces para obtener los resultados deseados. Analizando la Tabla 3, podemos extraer que la técnica de suavizado en los datos es efectiva, disminuyendo el ruido, por cuanto el esfuerzo en la generalización por parte del modelo es bajo y permite resultados muy cercanos a los reales. El modelo presenta un desempeño aceptable, pues se obtienen valores de correlación por sobre el 90%, hay pequeñas variaciones que nos permiten explicar el por qué es necesario analizar los datos que se dispone tanto para la etapa de entrenamiento y en consecuencia para el testing. Es así como en el caso donde se encontró un mejor desempeño es para el modelo mostrado en este resumen, con un 70% de la data para la etapa de entrenamiento, esto sin duda se debe a que se permite una buena generalización como un universo más grande para la evaluación de la etapa de prueba, lo que transforman el modelo en uno mucho más estable.

Evaluando el modelo propuesto, al predecir con datos reales (valor deseado), se tiene que línea histórica de los datos pronosticados se acerca a la línea de los datos reales, ver Figura 6 y 7, esto quiere decir que los datos pronosticados son ideales, esto debido a que el valor del porcentaje de error medio absoluto es bajo, muy cercano al 7%, como se muestra en la siguiente figura.

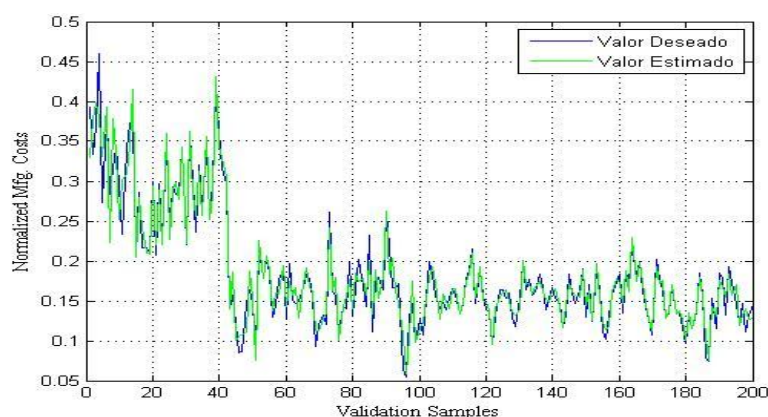


Figura 6: Comparación de los datos con el modelo propuesto

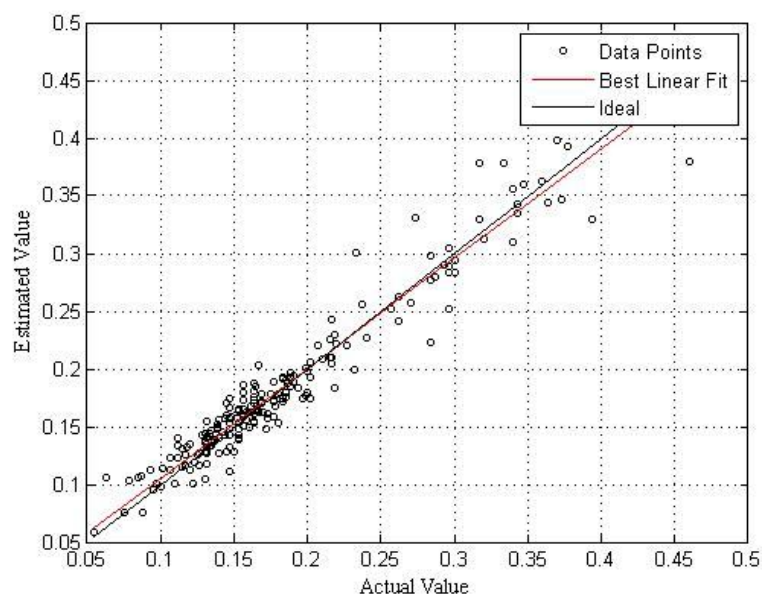


Figura 7: Correlación de los datos con el modelo propuesto

## 5. CONCLUSIONES

En este proyecto se presentó e implementó tres modelos para predecir la cantidad de accidentes de tránsito con resultado de graves y menos graves en la Región Metropolitana de Chile con redes neuronales utilizando algoritmo de Levenberg-Marquart.

Debido a que en la ocurrencia de este tipo de accidentes está en riesgo el bien más preciado para el hombre, como lo es la vida, se hace imprescindible la realización de modelos que permitan predecir y entreguen información sobre la cual tomar medidas a fin de evitar estos sucesos.

Como se muestra en el análisis estadístico de los accidentes graves y menos graves, se prevé que son causados principalmente por la causa conducción que contempla, es decir “Conducir no atento condiciones tránsito momento” y “Conducir sin mantener distancia razonable ni prudente” donde la frecuencia de ambas causas son mayor a un 46% de las causas por lo que habría evitar dichas formas de conducir por otra para los pasajeros que principalmente se incluyen los peatones, las causas más recurrentes son “Peatón cruza en forma sorpresiva o descuidada” e “Imprudencia, del peatón” ambas son un 60% de las causas en los últimos 10 años.

Dado lo anterior se estudiará en el futuro la predicción de las causas de accidentes con un modelo autorregresivo multivariado, considerando las causas descritas en el párrafo anterior. Por ejemplo además de predecir los accidentes con resultado de graves, se podrá predecir la causa, que podría no mantener la distancia prudente y razonable.

## 6. REFERENCIAS

- Barba, L, Rodríguez, N, Montt, C Hindawi (2014) Smoothing Strategies Combined with ARIMA and Neural Networks to Improve the Forecasting of Traffic Accidents , Publishing Corporation Scientific World Journal , Volume 2014, Article ID 152375, 12 pages.
- Bedard, M., Guyatt, G. H., Stones, M. J., y Hireds, J. P. (2002). The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. s.l. Accident analysis and Prevention. págs. 717-727.
- Comisión Nacional de Seguridad de Tránsito (2013) Estadísticas Generales, Cantidad de siniestros de tránsito y de víctimas. Disponible vía web [http://www.conaset.cl/conaset\\_web/contenido.php?id=73](http://www.conaset.cl/conaset_web/contenido.php?id=73), 2000-2012.
- Contreras, A, (2014) Red neuronal basada en el algoritmo Levenberg-Marquardt para pronosticar la cantidad de accidentes de tránsito en la región metropolitana de Chile. Proyecto final para optar al título de Ingeniero Civil Informático, Pontificia Universidad Católica de Valparaíso.
- Deublein, M., Schubert, M., Adey, B., Köhler, J. y Faber, M. (2013). Prediction of road accidents, Accident Analysis and Prevention, Volume 51.
- Dia, H y Rose, G. (1997). Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. s.l.: Transportation Research C, págs. 313-331.
- Hovden J, F. Størseth y R Kviseth (2011) multilevel learning from accidents – Case studies in transport. ELSEVIER Safety Science 49 (2011) 98–105.
- Kunt, M., Aghayan, I. y Noi, N. (2011). Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods Transport Volume 26, Issue 4, 1, Pages 353-366.
- Mejía, J.A. (2004) Sistema de detección de intrusos en redes de comunicaciones utilizando redes neuronales. Universidad de las Américas Puebla, México.
- Montt, C., Castro, F. y Rodríguez, N. (2011) Análisis de accidentes de tránsito con maquinas de soporte vectorial LS-SVM, Valparaíso, Actas del XV Congreso Chileno de Ingeniería de Transporte, Santiago, Chile, paper 54.
- Montt, C., Zúñiga, A. y Chacón, M. (2010) Identificación de factores determinantes en accidentes de tránsito que afecten a las personas mediante redes bayesianas, Actas de XVI Congreso Panamericano de Ingeniería de Tránsito, Transporte y Logística. Lisboa, Portugal.
- Organización Mundial de la Salud (2009) Los accidentes de tránsito, el suicidio y las afecciones maternas figuran entre las principales causas de muerte de los jóvenes. Disponible vía web en [http://www.who.int/mediacentre/news/releases/2009/adolescent\\_mortality\\_20090911/es/](http://www.who.int/mediacentre/news/releases/2009/adolescent_mortality_20090911/es/), 2009.

Rubio J, Montt, C. y Lanata S. (2013) Análisis de datos para el entrenamiento de redes neuronales para estimación de accidentes de tránsito, Actas del Congreso Chilena de Investigación de Operaciones, OPTIMA, paper 59, Concepción, Chile.

Shou-Ren Hua, Chin-Shang Li b, Chi-Kang Lee (2010). Investigation of key factors for accident severity at railroad grade crossings by using a logit model. ELSEVIER Safety Science 48 (2010) 186–194.

Wang, H., Zheng, L. y Meng, X. (2011). Traffic accidents prediction model based on fuzzy logic. Communications in Computer and Information Science. Volume 201 CCIS, Issue PART 1.