

# ANÁLISIS DE ACCIDENTES DE TRÁNSITO CON INTELIGENCIA COMPUTACIONAL

Cecilia Montt, Pontificia Universidad Católica de Valparaíso, [cmontt@ucv.cl](mailto:cmontt@ucv.cl)  
José Miguel Rubio, Pontificia Universidad Católica de Valparaíso, [jose.rubio.l@ucv.cl](mailto:jose.rubio.l@ucv.cl)  
Silvana Lanata, Pontificia Universidad Católica de Valparaíso, [silvanalanatad@gmail.com](mailto:silvanalanatad@gmail.com)

## RESUMEN

El objetivo de la investigación es el desarrollo y evaluación de modelos de predicción que estimen el número de personas lesionadas y fallecidas en accidentes de tránsito, relacionado con las causas que producen el accidente. Se utilizarán redes neuronales artificiales en combinación con algoritmos de técnicas de inteligencia artificial. Este artículo presenta la primera parte del trabajo, que corresponde al análisis de los datos para el diseño de los modelos de estimación. Como resultado de esta etapa se obtuvieron las causas más frecuentes de los últimos 8 años.

*Palabras claves: redes neuronales, inteligencia artificial, accidentes de tránsito*

## 1. INTRODUCCIÓN

Un informe de seguridad vial realizado por la organización mundial de la salud en el 2009, sostuvo como conclusión que la tasa más alta de letalidad por cien mil habitantes correspondían a países de ingresos bajos y medios.

Desde hace algunos años, se han venido desarrollando diversas investigaciones en accidentes de tránsito que utilizan técnicas como Redes Bayesianas (Montt, 2010), Perceptrón Multicapa (MLP), maquinas de soporte vectorial (LS-SVM) (Montt et al, 2011 y Kunt et al ,2011) comparó algoritmos genéticos y redes neuronales artificiales con patrones de búsqueda. Sin embargo, el uso de redes neuronales artificiales en estimación de accidentes de tránsito es reducido y su enfoque depende plenamente de cómo el investigador construya el modelo de predicción

Sin embargo, los resultados obtenidos no permiten decidir si estos modelos son realmente efectivos. Es por ello que este proyecto busca mejorar las técnicas de inteligencia computacional ya utilizadas con el fin de obtener mejores resultados y obtener predicciones, no solo clasificaciones, para poder contribuir con un conjunto de modelos que sean un real aporte al estudio de la seguridad vial.

Las Técnicas de Inteligencia Computacional (IC) permiten visualizar patrones y variables significativas que ayudan a identificar de mejor manera las condiciones en las que ocurren los accidentes, considerando que cada siniestro implica la conjugación de un gran número de variables, de las cuales se deben identificar sólo las más relevantes para determinar la cantidad de accidentes que se producen en la actualidad.

En este artículo se presenta un análisis preliminar de los datos, es decir, su pre- procesamiento, que corresponde a la primera parte de la metodología para el diseño de los modelos de estimación. En esta etapa se determinan de las causas más frecuentes que producen los accidentes de tránsito, en 8 años de estudio, desde 2004 hasta el 2011.

En los puntos siguientes primero se presenta el estado del arte, después el marco teórico que se utilizara para la estimación de los modelos de predicción, enseguida se presenta la metodología, finalmente los resultados de la primera parte del proyecto y las conclusiones.

## 2. ESTADO DEL ARTE

Un estudio realizado por (Abdelwahab y Abdel-Aty, 1997), sobre los datos de accidentes en el área de Florida, tenía por objetivo analizar los accidentes ocurridos en intersecciones señalizadas. Las lesiones se clasificaron en: no lesiones, posibles lesiones y lesiones incapacitantes. En el estudio se comparó el desempeño de Fuzzy ARTMAP y el Perceptron Multicapa, dando mejores resultados este último. Para MLP se utilizó Lavenberg-Marquardt como algoritmo de entrenamiento, obteniendo una precisión del 60.4% al probarlo. Para el caso de Fuzzy ARTMAP se logró una precisión del 56.1%.

En (Mussonne, 1999) se realizó un estudio para analizar los accidentes vehiculares ocurridos en las intersecciones en Milán, Italia . Usó un modelo MLP feed-forward y se entrenó con BP. El modelo tenía diez nodos de entrada para ocho variables (día o noche, flujo de tránsito circulante en la intersección, número de puntos virtuales conflictivos, número de puntos real de real conflicto, tipos de intersección, tipos de accidentes, condiciones de la superficie del camino y condiciones del tiempo). Los nodos de salida fueron llamados índice de accidente, este fue calculado como un ratio entre el número de accidentes para una cierta intersección y el número de accidentes para la intersección más peligrosa. Los resultados mostraron que el mayor índice de accidentes ocurrió en intersecciones no señalizadas en la noche.

En (Sohn y Lee, 2002), se usaron varios algoritmos para mejorar la exactitud de clasificadores individuales (red neuronal y árbol de decisión) para dos categorías de gravedad en accidentes de tránsito (lesiones corporales y daño de propiedad). Se usaron tres enfoques distintos (classifier fusión basado en el algoritmo de Dempster-Shafer, procedimiento Bayesiano y modelo logístico); data ensemble fusión de datos basado en arcing y bagging; y clustering basado en el algoritmo k-means. Obteniendo los mejores resultados el algoritmo clustering basado en clasificación; un 73.94% y un 76.1%, para el método de clustering para red neuronal y árbol de decisión respectivamente. La red neuronal obtuvo un 70.86% de precisión.

En (Chong et al, 2004) se realizó una clasificación de accidentes según la severidad del accidente, clasificando este en cinco clases (no lesión, posible lesión, lesiones no incapacitantes, lesiones incapacitantes y lesiones fatales) y se compararon las técnicas de árbol de decisión, red neuronal (usando un entrenamiento que combinaba Back Propagation y Conjugate Gradient Descent), Hibrid DT-ANN y SVM, usando la función de base radial para el kernel. Como resultado se obtuvo que en todas las clases el árbol de decisión presento mejor precision

En (Akgüngör y Doganse, 2009) realizaron un estudio que comparó dos modelos analíticos (una modificación del modelo Smeed y el modelo de Andreassen adaptado) y una red neuronal artificial para estimar el número de fallecidos en accidentes de tránsito. Se realizaron pruebas utilizando los datos históricos de los accidentes de 20 años de tres grandes metrópolis de Turquía, los resultados mostraron que la red neuronal artificial predecía mejor los resultados al tener una mejor precisión que los modelos analíticos.

Por otra parte, se ha trabajado en el análisis de accidentes de tránsito de las principales regiones de Chile mediante técnicas estadísticas y algunas técnicas de minería de datos. Por ejemplo en (Montt et al , 2006), se analizaron los siniestros mediante estadísticas descriptivas y árboles de decisión, sin embargo, no se utilizó una metodología que ameritase un trabajo con gran cantidad de datos. Más tarde, Montt et al (2010) trabajaron con técnicas de minería de datos basadas principalmente en estructuras de reyes bayesianas. Procedieron a aplicar algoritmos de aprendizaje paramétrico y propagación de evidencia.

En (Montt, et al 2009), se utilizaron métodos de agrupamiento e índices de Calinski y Harabasz para encontrar agrupaciones que representaran mejor la información. Lograron

caracterizar los distintos tipos de accidentes, como choques con objeto, volcadura, caídas y los atropellos, encontrando características particulares en cada uno de los tipos de accidentes.

En (Kunt et al, 2011), se desarrollaron modelos usando como parámetros de entrada la 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. Se construyeron los modelos basados en un total de 1000 choques que se produjeron durante el año 2007 en la autopista entre Teherán y Ghom. El algoritmo genético evalúa once ecuaciones para obtener la mejor. La red neuronal utilizó una arquitectura multicapa perceptron (MLP) y una red de tipo feedforward. El modelo de mejor ajuste se selecciona de acuerdo con el valor de la raíz media de los errores cuadrados (RMSE), la media de los errores absolutos (MAE) y la suma de errores cuadrados (SSE). El más alto valor de R que se obtuvo para las redes neuronales artificiales fue de alrededor de 0,87, lo que demuestra que estas proporcionan la mejor predicción.

En (Fierro et al, 2013) se utilizó técnicas de clasificación basadas en redes neuronales artificiales y los algoritmos de entrenamiento LDWPSO, QPSO y LDWQPSO, con diferentes topologías de red y tamaños de muestra para clasificar el grado de severidad de las lesiones con que resultan las personas en accidentes de tránsito en la región de Valparaíso, Chile.

En síntesis, las Redes Neuronales, Arboles de Decisión, Redes Bayesianas y Clustering son técnicas predominantes dentro de los estudios realizados referentes a los accidentes de tránsito. Sin embargo, solo clasifican los accidentes de tránsito, por lo cual con técnicas de inteligencia computacional (IC) ya utilizadas, en este trabajo se obtendrán predicciones, para poder contribuir con un conjunto de modelos que sean un real aporte al estudio de la seguridad vial.

### 3. MARCO TEORICO

La (IC) provee diversas técnicas que permiten el análisis de datos, dentro de las cuales se puede mencionar a las Redes Neuronales Artificiales (RNA), que corresponden a un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Estos mecanismos son utilizados para la clasificación y predicción de datos, y en combinación con algoritmos de computación evolutiva (optimización por enjambre de partículas (PSO) y optimización por colonias de hormiga (ACO), entre otros), se pueden establecer patrones que permiten estimar el número de personas lesionadas y fallecidas en accidentes de tránsito.

La neurona artificial es una emulación a menor escala de la neurona biológica. Posee un estado interno, el cual se conoce como nivel de activación, este estado puede cambiar según las señales que reciba, las que pueden provenir del medio o de otras neuronas artificiales.

El conjunto de estados que puede poseer la neurona se denomina S, el cual puede tener valores binarios, naturales o incluso valores en un intervalo continuo.

La neurona tiene un nivel de activación (ver figura 1), el cual está determinado por las entradas (parámetros correspondientes a los datos del problema) y los pesos sinápticos que esta posea, si bien las entradas son externas a ella, los pesos pueden cambiar para así adaptar el nivel de activación (se pretenden ajustar los pesos mediante PSO y ACO). Este cambio va a depender de las entradas que recibe la neurona mediante una función de activación, estas entradas pueden venir de otras neuronas o del exterior.

Como se desea calcular el nivel de activación, es preciso en primer lugar calcular la entrada total de la célula  $E_i$ , este valor se calcula mediante la fórmula 1, en donde se suma cada entrada, en este caso las causas de accidentes de tránsito, ( $X=[x_1, x_2, \dots, x_j, \dots, x_n]$ , representando las señales de la sinapsis en una neurona biológica) multiplicada por un peso sináptico ( $W=[w_{i1}, w_{i2}, \dots, w_{ij}, \dots, w_{in}]$ , representando la fuerza de una conexión sináptica) y se le resta un umbral  $\theta$ . La entrada total de la célula,  $E$ , es procesada por la función de activación  $f$ , produciendo una señal de salida  $S$ . Es así que dependiendo de la función  $f$  es que se tendrán diferentes modelos de autómatas.

$$E = (X^T * W) - \theta = x_1 * w_1 + \dots + x_n * w_n - \theta = \sum_{i=1}^n w_i * x_i - \theta \tag{1}$$

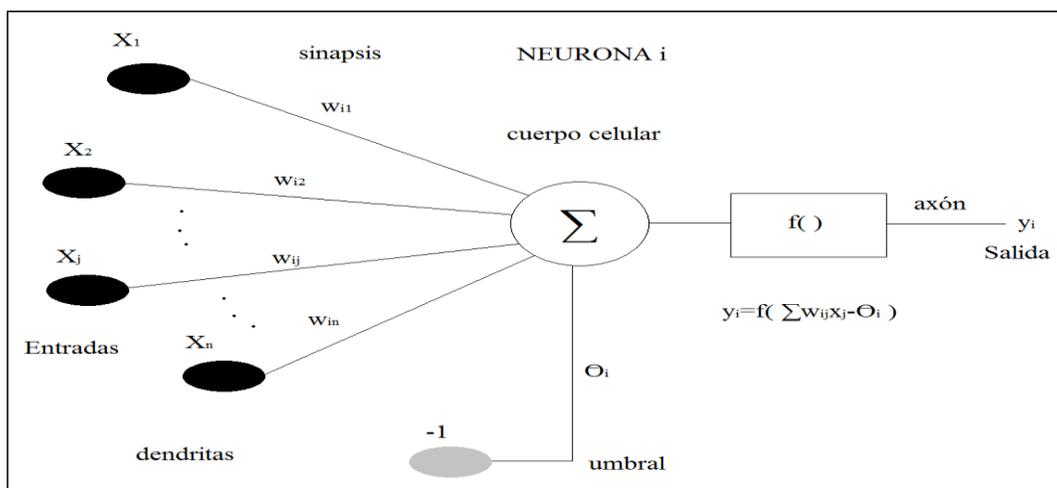


Figura 1. Esquema general de una red neuronal artificial

El aprendizaje en las redes neuronales artificiales es visto como el proceso de ajuste de pesos. Se parte de un conjunto de pesos sinápticos aleatorios, con el fin de buscar un conjunto de pesos que permitan a la red desarrollar correctamente una tarea en particular. El proceso de aprendizaje es un proceso iterativo, en el cual se va refinando la solución hasta alcanzar un nivel de operación que de conformidad. Una gran cantidad de métodos de entrenamiento utilizados en las redes neuronales con conexión hacia delante consisten en proponer una función de error que mida el rendimiento de la red en función de los pesos sinápticos. El objetivo que cumple el entrenar la red es obtener un conjunto de pesos sinápticos que minimizan (o maximizan) la función. El método de optimización proporciona una regla de actualización de los pesos que en función de los patrones de entrada modifica iterativamente los pesos hasta alcanzar el punto óptimo de la red neuronal. Una forma de

obtener mejores resultado en la obtención de estos pesos, es mediante la utilización de algoritmos evolutivos, como por ejemplo PSO y ACO (Fierro et al, 2012).

El backpropagation (retro-propagación) es el método de entrenamiento más utilizado en redes con conexión hacia delante. Este es un método de aprendizaje supervisado, en el que se aplica un patrón de entrada, el cual se propaga por las distintas capas que componen la red hasta producir la salida de la misma. La salida se compara con la salida deseada y se calcula el error cometido por cada neurona de salida. Estos errores se transmiten hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de las capas intermedias. Cada neurona recibe un error que es proporcional a su contribución sobre el error total de la red. Basándose en el error recibido, se ajustan los errores de los pesos sinápticos de cada neurona. Por lo general una forma de entrenar redes neuronales es mediante el error cuadrático medio, este es el método utilizado mayormente para fines predictivos y esta dado por la ecuación 2:

$$E = \frac{1}{N} * \sum_{i=1}^N (d_i(n) - y_i(n))^2 \quad (2)$$

Donde N es la cantidad de registros de dato para entrenamiento de la red,  $d_i$  es el valor deseado considerando las causas del accidente, es decir, el valor real para el i-ésimo registro de entrenamiento e  $y_i$  es el valor estimado entregado por la red para el i-ésimo registro de entrenamiento.

La validación de los modelos propuestos se realizará mediante la técnica de Validación Cruzada. Este método consiste en dividir los datos en dos partes, una para pruebas y otra para entrenamiento de la red neuronal, un ejemplo es usar el 80% de los datos para entrenamiento y un 20% para validación. Además se guardarán los valores de los pesos que den mejores resultados en el entrenamiento.

#### 4. METODOLOGÍA

Para el estudio se utilizarán los datos suministrados por CONASET, estos datos se encuentran almacenados en planillas Excel. En esta primera etapa del trabajo se analizaron solo tres años de accidentes de tránsito en la V Región, el 2004 al 2011.

Los datos se pueden separar en tres tipos de entidades, cada una con sus respectivos atributos para fines de registro:

- Accidentes: identificador del accidente, fecha, hora de accidente, comuna, causas, tipo de accidente, estado atmosférico, condición calzada, tipo de calzada, estado calzada, urbano/rural, y la información que indica la cantidad de ilesos leves, graves, menos graves y muertos.
- Persona: identificador del accidente, sexo, calidad (peatón, conductor, pasajero), edad, y resultado (muerto, grave, menos grave, leve, ileso).
- Vehículo: identificador del accidente, servicio, y el tipo de vehículo.

En cada accidente puede haber una o muchas personas involucradas y uno o muchos vehículos involucrados.

#### 4.1. Diseño de Modelos

Para llegar a la obtención de los modelos de predicción se ha de seguir una serie de pasos, estos se detallan a continuación:

1. Tratamiento de datos: en este paso se realiza un pre-procesamiento de los datos, en el cual se eliminan redundancias e inconsistencias para luego normalizar los datos para facilitar las tareas posteriores.
2. Elección de la red neuronal artificial y algoritmo de entrenamiento: en esta etapa se seleccionan el tipo de red usada para cada modelo.
  - a. Selección de algoritmo de entrenamiento (PSO, ACO o alguna de sus variantes).
  - b. Elección de topologías: se define el número de nodos de entrada de la red, el número de nodos de la capa oculta y de la capa de salida.
3. Entrenamiento de la red: la red se entrena mediante PSO y ACO, de manera de obtener los pesos de la red. Como se indicó anteriormente, el entrenamiento se realiza utilizando un 80% de los datos.
4. Testeo de la red: se obtienen los resultados de la estimación (número de accidentes y número de personas fallecidas y lesionadas), para ello se utilizará el 20% de los datos restantes que no se utilizaron para entrenamiento.
5. Comparación de resultados: en la etapa final se compararán los resultados obtenidos por los diversos modelos.

Selección y codificación de datos de entrada para los modelos, donde los atributos de entrada para los modelos de estimación que se obtendrán en una segunda parte de este trabajo, son el estado (fallecido, lesionado o ileso) de las personas involucradas en accidentes de tránsito, las causas (99 causas las cuales se agruparon en fallas mecánicas, adelantamientos, conducción, pasajero, señalización, velocidad, carga y / o descarga y otras infracciones) y tipos de accidentes (atropello, colisiones, choques y otros). Con esta información se procederá a diseñar los modelos de estimación para predecir la gravedad de la lesión de las personas.

La formulación esperada de los modelos de predicción sería por ejemplo:

$$Y = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3)$$

Donde Y por ejemplo sería la persona conductor en estado grave, donde los  $x_i$  serían las causas descritas anteriormente,  $a_i$  los parámetros asociados a cada causa.

## 5. RESULTADOS PRIMERA ETAPA DEL PROYECTO

Del estudio de 8 años de accidentes de tránsito, se identifican las causas que tienen la mayor frecuencia, para los conductores, pasajeros y peatones involucrados en los accidentes. El total

de involucrados en accidentes de tránsito durante esos 8 años son en promedio 1,5 a 2 involucrados por accidentes.

Las causas que mas se repiten son, para:

Conductor:

1. Conducción sin mantener distancia razonable ni prudente. (Causa 28)
2. Perdida control vehículo. (Causa 95)
3. Conducción no atentos a las condiciones de transito(causa 27)
4. Señalización, desobedecer señal pare (causa 65)
5. Conducción, cambiar sorpresivamente pista circulación (causa 29)

Las dos primeras causas se repiten en todos los años.

Pasajero:

1. Conducción sin mantener distancia razonable ni prudente. (Causa 28)
2. Perdida control vehículo. (Causa 95)
3. Conducción en estado de ebriedad (causa 24)
4. Conducción no atentos a las condiciones de transito(causa 27)
5. Velocidad no razonable ni prudente (causa 72)

Las tres primeras causas se repiten en todos los años.

Peatón:

1. No respetar derecho preferente de paso peatón. (Causa 31)
2. Peatón, cruza calzada forma sorpresiva o descuidada. (Causa 52)
3. Imprudencia del Peatón. (Causa 53)
4. Ebriedad del peatón. (causa 24)
5. Causas no determinadas.(causa 97)

Las tres primeras causas se repiten en todos los años.

A continuación, en la figuras 2, 3 y 4 se muestra la tendencia de la causa para cada tipo de involucrado, indicando en el eje de las "Y" el numero de involucrados:

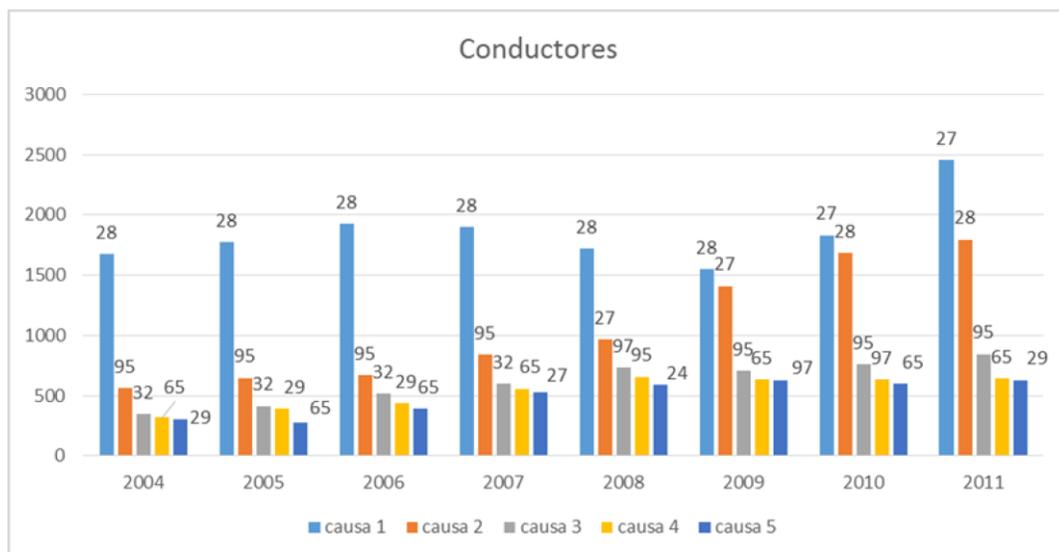


Figura 2.- Conductores involucrados por causas

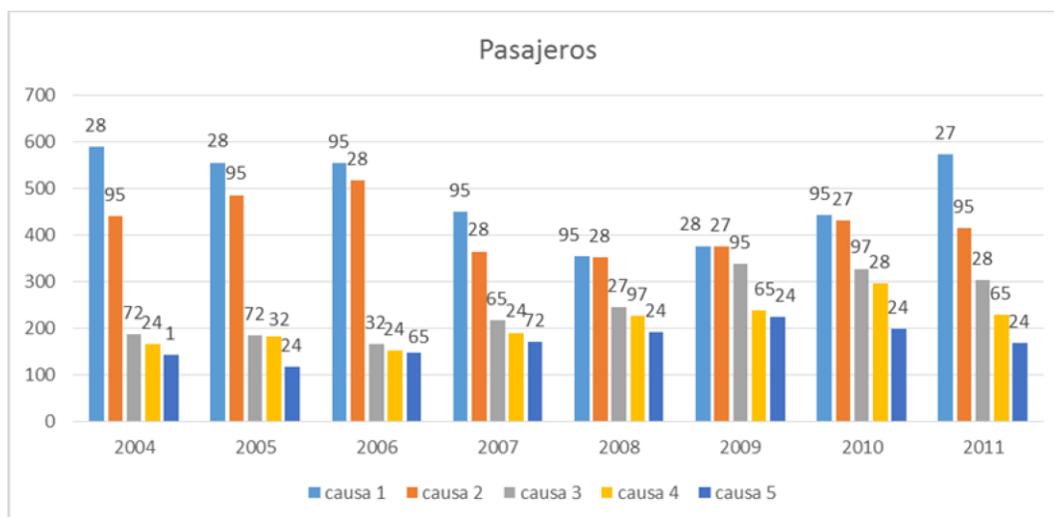


Figura 3.- Pasajeros involucrados por causas

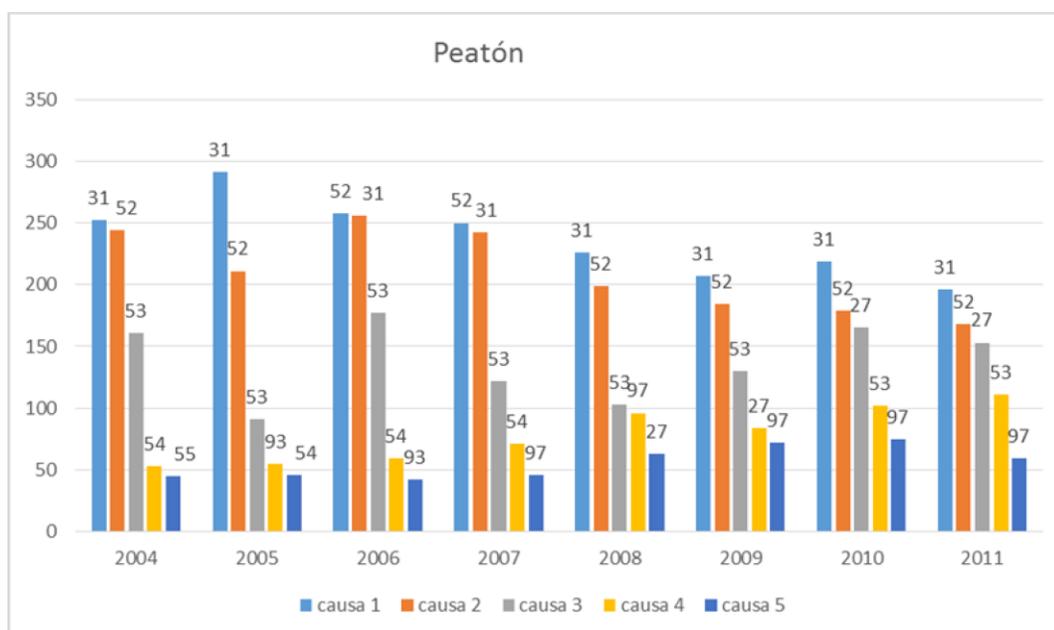


Figura 4.- Peatones involucrados por causas

Como si vizualiza en los graficos para los conductores y pasajeros la causa no atentos a las condiciones de transito ha aumentado en realacion a la causa no mantener la distancia prudente y razonable. Para los patones las causas que mas se repiten es no respetar el derecho preferente a paso y peaton cruza sorpresivamente la calzada.

Con la informacion encontrada, se podra ahora estimar modelos para predecir el estado que queda la persona involucrada en un accidente de transito, dado las causas mas frecuentes encontradas en esta primera etapa del trabajo.

## 6. CONCLUSIONES

En esta primera etapa del trabajo se analizan y preprocesan los datos con los cuales se obtendrán modelos de predicción, en esta etapa se obtienen las causas más frecuentes para los peatones, conductores y pasajeros, y tendencias al respecto, para los accidentes ocurridos en la V Región.

Los resultados de esta primera etapa, indican que las causas más frecuentes de accidentes de tránsito en los últimos 8 años son “conducción sin mantener distancia razonable ni prudente” y “perder el control del vehículo”, estas causas son para pasajeros y conductores, en cambio para peatones las causas más frecuentes son “no respetar derecho preferente de paso peatón” y “peatón, cruza calzada forma sorpresiva o descuidada”.

Entonces, con los datos ya analizados en esta etapa, se diseñarán los modelos de estimación para predecir la gravedad de la lesión de las personas después del accidente, es decir, lesionada o fallecida, según sus causas.

Con lo anterior se podrán hacer campañas de prevención con 8 años de accidentes de tránsito, y validar los modelos con los datos del año 2012, además se construirá un prototipo de Sistema de Estimación de Personas Fallecidas y Lesionadas en Accidentes de Tránsito.

## Referencias

Akgüngör A. P. y Dogan E.( 2009) An application of modified Smeed, adapted Andreassen and artificial neural network accident models to three metropolitan cities of Turkey, pp. 906-913.

Abdelwahab H. T. y Abdel-Aty M. A.(1997), Development of Artificial Neural Network Models to Predict Driver injury Severity in Traffic Accidents at Signalized Intersections. Transportation Research Record 1746: pp. 6-13.

CONASET. Estadísticas generales, Enero 2013.

Chong M. M., Abraham A.y Paprzycki M. (2004), Traffic Accident Analysis Using Decision Trees and Neural Network. Oklahoma State University.

Fierro, Jose, Montt, Cecilia; Rodriguez, Nibaldo(2012) “Clasificación de accidentes de tránsito utilizando redes neuronales con algoritmo PSO” Actas de Congreso Latino Americano de Investigación de Operaciones, Septiembre 24-28, 2012 – Rio de Janeiro Brasil

Kunt, M.M., Aghayan, I.a , Noii, N.b (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 December 2011, Pages 353-366

Mussone L., Ferrari A. y Oneta M.(1999) An analysis of urban collisions using an artificial intelligence model. Accident Analysis and Prevention, Milán, Montt, Cecilia, Castro, Felix, y Rodriguez Nibaldo “Análisis de accidentes de tránsito con máquinas de

soporte vectorial LS-SVM”, Valparaíso, Actas del Congreso Chileno de Ingeniería de Transporte, (Pendrive) Octubre 2011, Santiago, Chile, paper 54.

Montt C. (2006) “Definición de variables que afectan al individuo en su comportamiento respecto a la seguridad vial”, Proyecto de Investigación Interno PUCV, DGPI N°288.733/2006

Montt, Cecilia, Musso, Reynaldo y Chacón, Max. (2009) “Análisis de Accidentes de Tránsito con Métodos de Agrupamiento” Actas del VIII Congreso Chileno de Investigación de operaciones, OPTIMA 2009- , Universidad del Bio- Bio, Concepción, Chillan.

Montt, Cecilia, Zúñiga, Alejandro y Chacón, Max (2010) “Identificación de factores determinantes en accidentes de tránsito que afecten a las personas mediante redes bayesianas, Actas de XVI PANAM, July 15-18, 2010 – Lisboa, Portugal.

OMS. Prevención de la violencia y los traumatismos y discapacidad, Agosto 2011

Sohn S. Y. y Lee S. H. (2002) Data fusion, Ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea.

World Health Organization. Informe Sobre la Situación Mundial de la Seguridad Vial: Es Hora de Pasar a la Acción. World Health Organization, 2009.