Aplicación de redes bayesianas para el control de la frecuencia de los accidentes viarios

Application of Bayesian Networks to Control Road Accident the Frequency

Francisco Soler-Flores[1], Nicoleta González-Cancela[2], Beatriz Molina Serrano*[3]

Resumen

La teoría y la aplicación de eventos raros han sido muy importantes en los últimos años debido a su importancia práctica en campos muy diferentes, tales como seguros, finanzas, ingeniería o ciencias ambientales. Este artículo presenta una metodología para predecir eventos raros basado en redes bayesianas, que a su vez permite el estudio de escenarios alternativos para controlar la frecuencia de accidentes de tráfico. De esta manera, el modelo Naive-Poisson y ROCDM se presenta en este documento para su validación. El modelo desarrollado se usa para estimar y predecir los accidentes de tránsito como eventos raros y los resultados se han evaluado mediante el uso de la curva ROCDM. Un modelo Naive-Poisson y un modelo de validación basado en la curva ROC se usa para estudiar varias carreteras españolas y los resultados se muestran aquí.

Palabras clave: Accidentes de tráfico, eventos raros, inteligencia artificial, red bayesiana, Naive Bayes, Naive-Poisson.

Abstract

The theory and application of rare events have been very important in recent years due to their practical importance in very different fields such as insurance, finance, engineering or environmental sciences. This article presents a methodology for predicting rare events based on Bayesian networks, which in turn allows the study of alternative scenarios to control the frequency of traffic accidents. In this way, the Naive-Poisson and ROCDM model are presented in this document for validation. The developed model is used to estimate the prediction of traffic accidents as rare events and the results have been evaluated by using the ROCDM curve. Naive-Poisson model and a validation model based on the ROC curve are used to study several Spanish roads and the results are shown here.

Keywords: *Traffic accidents, rare events, artificial intelligence, Bayesian network, Naive Bayes, Naive-Poisson.*

1. INTRODUCCIÓN

En las últimas décadas, se han desarrollado muchas técnicas para el análisis y modelado de datos en diferentes áreas de estadísticas e Inteligencia Artificial, y varias se han aplicado al estudio de eventos raros (Rodriguez, 2015). Las técnicas de minería de datos, que incluyen aquellas que operan automáticamente con la mínima intervención humana, generalmente son eficientes para trabajar con las grandes cantidades de información disponible en las bases de datos de muchos problemas prácticos. Por lo tanto, los modelos de Inteligencia Artificial, tales como las Redes Bayesianas, se utilizan en este documento para estimar la probabilidad de que ocurra un evento raro.

La ingeniería de tráfico es una rama de la ingeniería civil que utiliza técnicas de ingeniería para lograr el movimiento seguro y eficiente de personas y bienes en las carreteras (Soler, Varela y González, 2015). Se centra principalmente en la investigación para un flujo de tráfico seguro y eficiente, como la geometría de la vía, las aceras y los cruces de peatones, las instalaciones de ciclo segregado, el marcado de carriles compartidos, las señales de tráfico, las marcas de la superficie de la carretera (marcas viales) y los semáforos. La ingeniería de tráfico se ocupa de la parte funcional del sistema de transporte, excepto la infraestructura provista.

El desarrollo de la infraestructura, carreteras y vías de alta capacidad en el mundo de hoy implica problemas sociales y económicos causados por los accidentes de tráfico. Los accidentes viales son un problema importante para los países desarrollados y son relevantes en la política. El estudio de la relación entre la frecuencia de accidentes de tráfico y las características de la vía, el tráfico, el medio ambiente y los usuarios es también una de las aplicaciones más importantes del análisis estadístico en el campo de la seguridad vial. Los modelos matemáticos utilizados para estimar la frecuencia de accidentes a lo largo de la historia han sido diferentes, desde la regresión lineal a los modelos de regresión multivariante, como la regresión logística o la regresión de Poisson o, más recientemente, el análisis de conglomerados y los árboles de clasificación. Los modelos de análisis de datos en el campo de la Inteligencia Artificial, como las redes neuronales o las redes bayesianas, comienzan a utilizarse

* Autora de contacto: bmolina@prointec.es

[1] Dr. en Tecnología y Sistemas de Información. Universidad Internacional de La Rioja, La Rioja (España).

[2] Dra. Ingeniera de Caminos, Canales y Puertos. Universidad Politécnica de Madrid, Madrid (España).

[3] Dra. Ingeniera de Caminos, Canales y Puertos. PROINTEC, S.A.U., Madrid (España).

en problemas relacionados con el estudio de la frecuencia de accidentes. Sin embargo, el Modelo Lineal Generalizado (Quishpe Tasiguano, 2015) es actualmente el más aceptado y, por lo tanto, el más utilizado. Por lo tanto, es ampliamente reconocido que la distribución de la frecuencia de accidentes sigue una distribución de Poisson, pero no existe un modelo estándar para el estudio de este problema en la comunidad científica. Además, no hay ningún software que permita trabajar con datos de accidentes para predecir estas frecuencias (Rodríguez et al., 2013).

Un evento Et (Weiss y Hirsh, 1998) es una observación que ocurre en un instante *t* y se describe mediante un conjunto de valores. Del mismo modo, una secuencia de eventos es una secuencia de eventos ordenados temporalmente, $S = \{Et_1, Et_2, \ldots Et_n\}$, que incluye todos los eventos dentro del intervalo de tiempo $t1 \leq t \leq tn$. Los eventos están asociados con un objeto de dominio D, que es la fuente o generador de eventos. El evento objetivo es el evento para predecir y especificar por un conjunto de variables.

En este siglo, han suscitado un gran interés la teoría y las aplicaciones de eventos raros y eventos extremos. Esto se debe a su relevancia práctica en diferentes campos, como los seguros, las finanzas, la ingeniería, las ciencias ambientales o la hidrología. El tratamiento de eventos raros, eventos que ocurren con una baja probabilidad, es un problema complejo e integral cuyo tratamiento cae dentro del campo de modelar la incertidumbre y la teoría de la decisión. La 'Ley de eventos raros', demostrada por Poisson (Delgado, 2017), se basa matemáticamente en el concepto de ocurrencia rara (Fernández, 2008). Esta ley que lleva su nombre también se llama ley de eventos raros de Poisson (1837) o también llamada 'ley de los pequeños números' (Ordóñez, 2014).

En las últimas décadas, se han desarrollado numerosas técnicas para el análisis y modelado de datos en diferentes áreas de las estadísticas (Tomz, King y Zeng, 2003), Flores, Mayora y Piña (Flores, Mayora y Piña, 2008) y la Inteligencia Artificial (Seiffert, et al., 2007). Estos se han aplicado al estudio de eventos raros. En este contexto, la minería de datos o data mining (MD) (Holmes, Tweedale y Jain, 2012) es un área interdisciplinaria moderna que abarca técnicas que operan automáticamente (requieren una intervención humana mínima) y también son eficientes para trabajar con las grandes cantidades de información disponible en las bases de datos de muchos problemas prácticos. Estas técnicas pueden extraer conocimiento útil (asociaciones entre variables, reglas, patrones, etc.) a partir de los datos brutos almacenados, lo que permite un mejor análisis y comprensión del problema. En algunos casos, este conocimiento también puede post-procesarse automáticamente (Ratner, 2017) y puede beneficiar la extracción de conclusiones e incluso tomar decisiones casi automáticamente en situaciones prácticas específicas (sistemas inteligentes). La aplicación práctica de estas disciplinas se extiende a muchos problemas de predicción comerciales o de investigación en campos de clasificación o diagnóstico.

2. INTELIGENCIA ARTIFICIAL: REDES BAYESINAS Y NAIVE BAYES

2.1. Redes Bayesianas

Las diversas técnicas disponibles en la minería de datos, las redes bayesianas (bayesian networks) o las redes probabilísticas permiten modelar toda la información relevante para un problema y sacar conclusiones basadas en el problema de la evidencia disponible mediante el uso de mecanismos de inferencia probabilísticos. Las redes bayesianas se han utilizado en el contexto de la estimación de la ocurrencia de eventos raros en algunos trabajos (Cheon, Kim, Lee y Lee (2009) y Ebrahimi y Daemi (2010)) sin señalar realmente un método de estimación general.

Las redes bayesianas (J. H. Kim y Pearl (1983), Pearl (1988)) son una representación compacta de una distribución de probabilidad multivariada. Formalmente, una red Bayesiana es un gráfico acíclico dirigido, donde cada nodo representa una variable aleatoria y las dependencias entre variables están codificadas en la estructura del gráfico (Figura 1) de acuerdo con el criterio de la separación (Castillo, Gutiérrez y Hadi (1997)). Asociado con cada nodo, en la red hay una distribución condicional de los padres de esa probabilidad de nodo, de modo que la distribución conjunta tiene en cuenta el producto de las distribuciones condicionales asociadas con los nodos de la red. Es decir, para una red con n variables, la ecuación es la siguiente (ecuación 1):

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | pa(x_i)) \qquad [1]$$

Las redes Bayesianas generalmente consideran variables discretas o nominales; entonces éstas deben ser discretizadas antes de construir el modelo. Aunque existen modelos de redes bayesianas con variables continuas, estas variables están limitadas a las relaciones gaussianas y lineales. Los métodos de discretización se dividen en dos tipos principales: supervisados y no supervisados (Dougherty, Kohavi y Sahami,1995).

El concepto de causalidad (Agueda, 2011) en una red bayesiana da como resultado un caso particular de estas redes, llamadas causales (Pearl, 2000). Las redes bayesianas pueden tener una interpretación causal y aunque a menudo se usan para representar relaciones causales, el modelo no tiene que representarlas de esta manera, la distribución de Naive-Bayes es un ejemplo de esto, las relaciones no son causales.

Las redes bayesianas automatizan el proceso de desarrollo probabilístico (Uusitalo, 2007) utilizando la expresividad del gráfico (Vázquez, et al., 2018). Los modelos

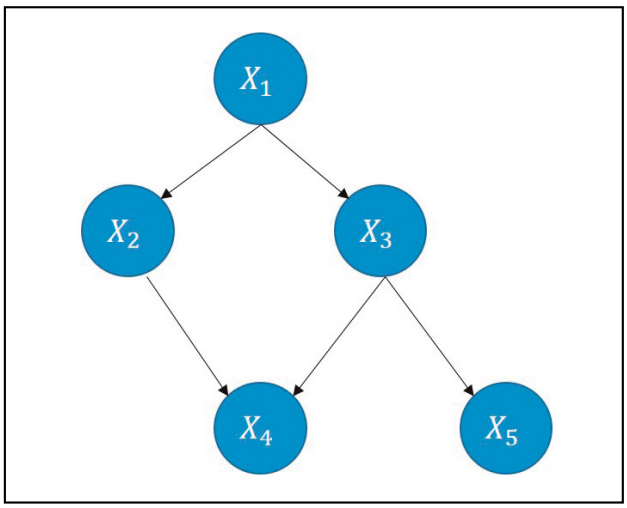

Figura 1. Ejemplo de red bayesiana.

resultantes combinan los resultados de la teoría de grafos (para representar las relaciones de dependencia e independencia de todas las variables) y la probabilidad (para cuantificar estas relaciones). Esta unión permite tanto la modelización eficiente de aprendizaje automático, a través del cálculo de parámetros (Sucar, 2006), que está modelado por una distribución Beta para el caso de variables binarias y variables multivaluadas, la distribución de Dirichlet (Tabla 1) y, por el otro lado, de la inferencia de la evidencia disponible. La base de conocimiento de tales sistemas es una estimación de la función de probabilidad conjunta de todas las variables en el modelo, mientras que el módulo de razonamiento es donde se realiza el cálculo de las probabilidades condicionales. El estudio de esta técnica proporciona una buena visión general del problema del aprendizaje estadístico y la minería de datos.

Tabla 1. Estimación de parámetros

Estimador	Expresión
Máxima verosimilitud (Maximum Likelihood). Multinomial	$\theta_k^* = \frac{N_k}{N}$
Estimación bayesianna (Bayesian estimation). Dirichlet	$\theta_k^* = \frac{N_K + a_k}{N + \sum_{i=1}^{r} a_i}$

2.2. Naive Bayes

Una de las formas más sencillas que se puede idear al considerar la estructura de una red bayesiana con objetivos que califican es la llamada Naive-Bayes (Duda, Hart y Stork, 2000). Su nombre proviene de las suposiciones ingenuas sobre las cuales se construye, qué es considerar que todas las variables predictoras son condicionalmente independientes dada la variable calificada (Figura 2).

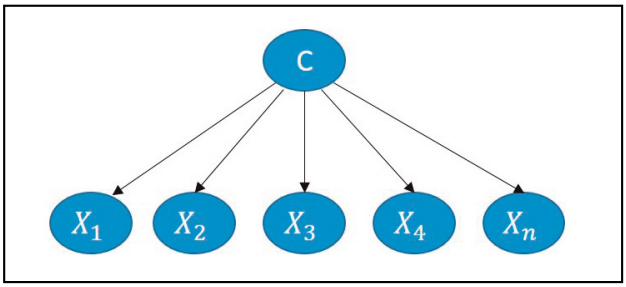

Figura 2. Naive Bayes.

El modelo Naive Bayes es muy utilizado porque (Holmes et al. (2012)):

- Es simple de construir y entender.
- El proceso de inducción es rápido
- Es muy robusto considerando atributos irrelevantes.
- Utiliza muchos atributos para hacer la predicción final.

Su poder predictivo es competitivo con otros clasificadores existentes, Naive-Bayes es uno de los clasificadores más efectivos (Duda et al., 2000). Este clasificador aprende la probabilidad condicional de cada atributo X_i dada la clase C de un conjunto de entrenamiento. El proceso de clasificación se obtiene aplicando la regla de Bayes, calculando la probabilidad de C, dadas las instancias de X_1, X_2, ..., X_n y tomando la probabilidad posterior más alta como clase predicha. Estos cálculos se basan en una fuerte suposición de independencia: todos los atributos X_i son condicionalmente independientes dado el valor de la clase C.

La probabilidad de que la *j-ésima* instancia perteneciente a la clase *i-ésima* de la variable C pueda aplicarse simplemente aplicando el teorema de Bayes, es como se indica a continuación (ecuación 2):

$$P(C = \theta_i | X_1 = x_{1j}, \dots, X_n = x_{nj}) \propto P(C = \theta_i) \cdot \cdot P(X_1 = x_{1j}, \dots, X_n = x_{nj} | C = \theta_i) \quad [2]$$

Dado que suponemos que las variables de predicción son condicionalmente independientes dada la variable C, obtenemos la ecuación 3:

$$P(C = \theta_i | X_1 = x_{1j}, \dots, X_n = x_{nj}) \propto P(C = \theta_i) \cdot \cdot \prod P(X_r = x_{rj} | C = \theta_i) \quad [3]$$

El modelo de Naive-Bayes combinado con la distribución de Poisson se utilizó para la clasificación de texto en el documento de S.-B. Kim, Seo y Rim (Kim, Seo y Rim, 2003) con buenos resultados. En este trabajo, se propone la adición de la minería de datos usando redes bayesianas y la aplicación de la distribución de probabilidad conocida para el estudio de eventos raros. De esta forma y con los valores conocidos de las variables utilizadas como predictores, se pueden estudiar diferentes escenarios y ver cuándo es más probable la ocurrencia de un evento raro.

3. METODOLOGÍA PROPUESTA

A continuación se va a explicar el modelo Naive-Poisson desarrollado de manera que esta metodología permita predecir eventos raros basándose en las redes bayesianas, que a su vez permite el estudio de escenarios alternativos para controlar la frecuencia de accidentes de tráfico.

3.1 Naive-Poisson para eventos raros

Las fases en las que se divide el modelo son las que se muestran en la Figura 3.

3.1.1. Preprocesamiento de datos

Las redes Bayesianas generalmente consideran variables discretas o nominales, por lo que el primer paso es discretizar y luego construir el modelo. Aunque existen modelos de redes bayesianas con variables continuas, estas variables están limitadas a relaciones gaussianas y lineales (Jimenez, et al., 2018). Los métodos de discretización se dividen en dos tipos principales: supervisados y no supervisados (Dougherty, 1995).

La discretización se aplica ampliamente en las aplicaciones de descubrimiento de conocimientos y aprendizaje automático, con el objetivo de:

1. reducir y simplificar los datos disponibles,
2. hacer que el aprendizaje de modelos sea más eficiente y

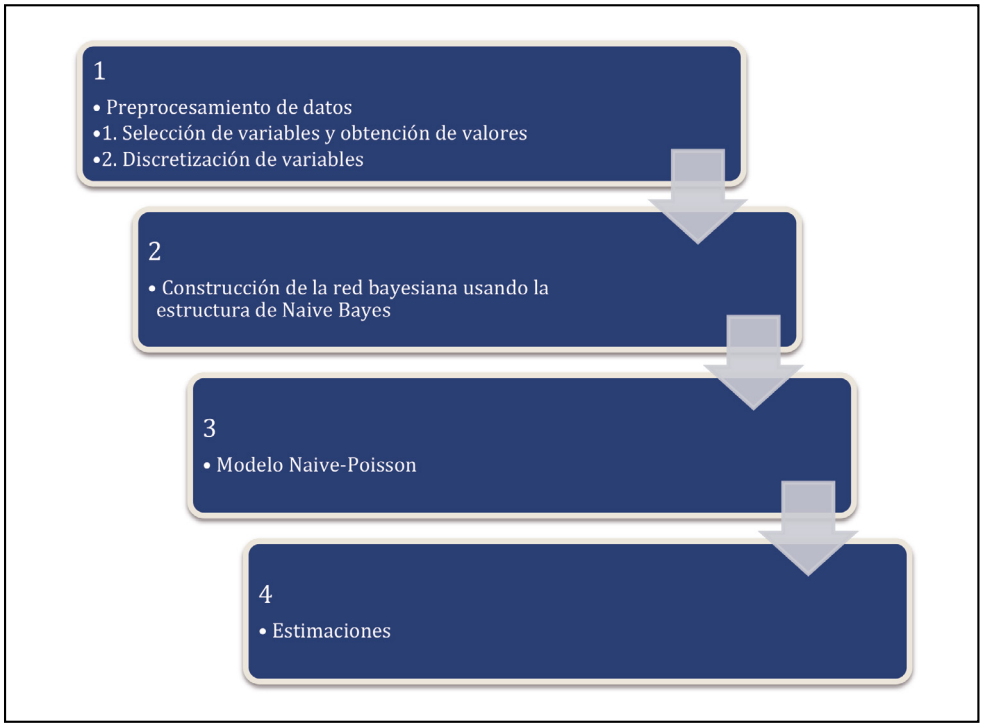

Figura 3. Fases de Naive Bayes para eventos raros.

3. obtener resultados más compactos y fácilmente interpretables (Liu et al., 2002).

A lo largo de los años, se han propuesto varios métodos diferentes de discretización, de los cuales solo unos pocos se utilizan ampliamente, mientras que otros pasan desapercibidos (García et al., 2013) (Yang et al., 2010) (Liu et al., 2002). Dado que la discretización de los datos generalmente da como resultado la pérdida de información (Li, 2007), (Uusitalo, 2007), el método de discretización empleado afectará la calidad predictiva de cualquier modelo aprendido a partir de los datos. Respecto a lo anterior, varios trabajos abordan la pregunta de qué método de discretización es más adecuado para la minería de datos en general (García et al., 2013) (Liu et al., 2002) o para el aprendizaje de redes bayesianas en particular (Lima et al., 2014), (Zhou et al., 2014), resultando que la mejor elección del método tiende a depender de la naturaleza y características de los datos disponibles.

3.1.2. Construcción de la red bayesiana

Desde la fase anterior, la construcción de la red consiste en lo siguiente:

- Parte cualitativa (estructura): identificar relaciones causales; analizar las variables en términos de dependencias e independencia
- Parte cuantitativa (evaluación de probabilidad): cuantificar relaciones e interacciones

En este caso, se propone que un modelo específico, el Naive-Bayes, adecuado para el proceso de clasificación y la variable para la que desea estimar los eventos raros, sea la variable denominada red principal (Castillo et al., 1997).

Se opta por el desarrollo de la metodología para el modelo Naive Bayes. El modelo de Naive Bayes es muy utilizado porque tiene, entre otras, ciertas ventajas:

- Es simple de construir y entender.
- Las inducciones son extremadamente rápidas, y solo requieren un paso para hacerlo.
- Es muy robusto considerando atributos irrelevantes.

Una vez que se da la estructura de la red, se calculan las tablas de probabilidad que permiten la descomposición de la distribución de probabilidad y luego la inferencia se basa en evidencias y se calcula la distribución de probabilidad asociada con la red.

Esta información proporciona la red completa construida utilizando el modelo Naive-Poisson que se describe a continuación.

3.1.3. Modelo Naive-Poisson

El modelo o procedimiento mediante el cual, a partir de la red bayesiana construida por Naive-Bayes, se obtiene la distribución de probabilidad asociada con la estimación de la probabilidad de ocurrencia de un evento raro es el llamado Naive-Poisson).

El proceso de asignación de la distribución de probabilidad consta de las siguientes secciones (El-Gheriani, Khan, Zuo, Ming, 2017):

- Asunción de Poisson
- Construcción de la distribución de probabilidad

Se acepta en la comunidad científica que la distribución de la frecuencia de eventos raros es compatible con una distribución de Poisson (Poisson, 1837) y (Scaillet, Treccani & Trevisan, 2017). Por lo tanto, se supone para la construcción del modelo que la frecuencia de eventos raros

sigue una distribución de Poisson. Tras obtener los valores de la distribución real, los datos fueron ajustados para este tipo de distribución (Melchers & Beck, 2018).

La distribución de probabilidad proporciona una red bayesiana construida para estimar la probabilidad de diferentes valores de cada uno de los valores de las variables que proporciona la discretización. A partir de los resultados obtenidos, la red se ajusta con una distribución de Poisson, que como sabemos está determinada por su media λ (ecuación 4).

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad [4]$$

Para el cálculo del parámetro que determina la distribución de Poisson, se toma la media λ, su estimador de máxima verosimilitud, que viene dada por la ecuación 5, donde xi son los valores discretos de los accidentes y *p(xi)* la probabilidad que proporciona la red construida a través del algoritmo de Naive-Bayes, siendo *C* la variable discreta para clasificar, que es la variable cuyos valores se consideran eventos raros que queremos estudiar.

$$\lambda = \sum_{i=1}^{n} x_i \cdot p(x_i) \quad [5]$$

Para cada uno de los valores de los estratos a_1, a_2, ….*an* que proporciona la discretización de la variable para estudiar los eventos raros, se toman los valores reales de la frecuencia del evento raro que se va a estimar, excepto para a_n, en el cual se asigna un valor dado promedio de los valores más altos que x_i en el estrato superior a_{n-1} (ecuación 6).

$$a_n = \bar{x}_i \; con \; x_i > a_{n-1} \quad [6]$$

De esta forma se obtiene la distribución de Poisson asociada a cualquiera de las situaciones (j = 1, Idots, m) para estudiar (ecuación 7) con cualquier conjunto de valores de las diferentes variables seleccionadas, lo que permite determinar la situación donde hay una mayor probabilidad de ocurrencia de eventos de baja probabilidad una vez que se han detectado los valores de la variable estudiada, que dada su distribución, resulta que los eventos raros son de baja probabilidad.

$$P_j(X = x) = e^{-\lambda j}\frac{\lambda_j^x}{x!} \quad [7]$$

Como resultado del proceso se obtiene un modelo probabilístico basado en redes bayesianas y una distribución de probabilidad que determina la probabilidad de ocurrencia de los diferentes valores de accidentalidad en diferentes puntos de la carretera, a partir de variables geométricas y de tráfico de la misma. A partir de la red construida y al realizar inferencia que, como es sabido, consiste en "dadas ciertas variables conocidas (evidencia), calcular la probabilidad posterior de las demás variables (desconocidas)", a partir de las variables físicas de la carretera y variables de tráfico conocidas, se determina la frecuencia de accidentes mediante el cálculo de la probabilidad a posteriori y se compara con diferentes opciones.

3.2. Validación. Curva ROCDM

Estimar la exactitud (accuracy) de un clasificador inducido por un algoritmo de aprendizaje automático, es decir, validar un clasificador, es importante no solo para predecir su futuro comportamiento, sino también para poder escoger un clasificador (selección de modelo) (Schaffer, 1993) dentro de un conjunto de posibilidades, o para combinar clasificadores. Para estimar la exactitud final de un clasificador, lo deseable es tener un método con poca varianza.

Así, la exactitud de un clasificador es la probabilidad con la que dicho clasificador clasifica correctamente una instancia seleccionada al azar. Algunos investigadores, sobre todo en la comunidad estadística, usan ratios de error (uno menos la exactitud) en lugar de la exactitud.

Para los conjuntos de datos reales, el valor de la exactitud más alta no es conocido, pero probablemente no sea 100% en la mayor parte de los dominios. El poder computacional ha crecido hasta un punto en que los métodos de computación intensivos para estimar la exactitud son utilizados más a menudo y en conjuntos de datos más grandes. Entre otros se puede hablar de:

- Matrices de confusión. Permiten ver mediante una tabla la distribución de los errores cometidos.
- Métodos estándares de validación. Permiten obtener la exactitud de los clasificadores.
- Curva ROC (Receiver Operation Characteristic). Procedimiento que permite evaluar la calidad de los clasificadores.

Este artículo presenta un modelo y una metodología que, a partir de los datos brutos de los que se dispone habitualmente a la hora de estudiar cualquier problema, en el marco del análisis de datos, sistematiza el estudio de sucesos raros y es capaz de estudiar diferentes situaciones que modifican la probabilidad de ocurrencia de estos. Para su evaluación se propone la utilización de una modificación de la curva ROC, la curva ROCDM.

Para evaluar si un clasificador supervisado es mejor que otro, una posible comparación que se puede realizar es que el clasificador que mejor tanto por ciento de bien clasificados tenga sea el mejor clasificador. Sin embargo, hay un enfoque más formal, basado en el cálculo del área bajo la curva ROC del clasificador. Cuanto mayor sea el área bajo la curva ROC, mejor será el clasificador.

La definición dada se centra en los casos en los que solo hay dos valores posibles de clase a clasificar. Sin embargo, las curvas ROC son extensibles para cualquier número de clases. El análisis ROC Receiver Operating Characteristic es una metodología desarrollada en el seno de la Teoría de la Decisión en los años 50 (Green and Swets, 1966), y cuya primera aplicación fue motivada por problemas prácticos en la detección de señales por radar. Con el tiempo se comenzó a utilizar en el área de la biomedicina, inicialmente en radiología.

Un primer grupo de métodos para construir la curva ROC lo constituyen los llamados métodos no paramétricos. Se caracterizan por no hacer ninguna suposición sobre la distribución de los resultados del clasificador. El más simple de estos métodos es el que suele conocerse como empírico, que consiste simplemente en representar todos los pares (1 - especificidad, sensibilidad) para todos los posibles valores de corte que se puedan considerar con la muestra particular de que se disponga. Desde un punto de vista técnico, este método sustituye las funciones de densidad teóricas por una estimación no paramétrica de ellas, es

decir, la función de densidad empírica construida a partir de los datos.

La mayor exactitud diagnostica de una prueba se traduce en un desplazamiento hacia arriba y a la izquierda de la curva ROC. Esto sugiere que el área bajo la curva ROC se puede emplear como un índice conveniente de la exactitud global de la prueba: la exactitud máxima correspondería a un valor del área bajo la curva de 1 y la mínima a un valor de 0.5. Cuando la curva ROC se genera por el método empírico, independientemente de que haya empates o no, el área puede aproximarse mediante la regla trapezoidal, es decir, como la suma de las áreas de todos los rectángulos y trapecios (correspondientes a los empates) que se forman bajo la curva.

La exactitud clasificatoria se expresa como sensibilidad y especificidad (Greiner, Matthias & Gardner, 2000). Cuando los datos de la prueba son dicotómicos (si-no, sano-enfermo,…) se pueden expresar los resultados mediante tablas de contingencia y a partir de ellas obtener los valores de sensibilidad y especificidad de la prueba (Tabla 2).

Tabla 2. Valores de sensibilidad y especificidad de la prueba

		Real	
		1	0
Prueba	1	Verdadero positivo	Falso positivo
	0	Falso negativo	Verdadero negativo

Siendo

- Sensibilidad = probabilidad de clasificar correctamente a un sujeto enfermo.
- Especificidad = probabilidad de clasificar correctamente a un sujeto sano.
- V PP = probabilidad de que un sujeto enfermo dé positivo en la prueba (valores predictivos positivos).
- V PN = probabilidad de que un sujeto sano dé negativo en la prueba (valores predictivos negativos).
- Exactitud = probabilidad de resultados correctos de la prueba.

El análisis de las curvas ROC surgió a principios de los años cincuenta (Zweig and Campbell, 1993) para el análisis de la detección de las señales de radar. En medicina el análisis ROC se ha utilizado de forma muy extensa en epidemiología e investigación médica, de tal modo que se encuentra muy relacionado con la medicina basada en la evidencia. El análisis ROC es la técnica de preferencia para evaluar nuevas técnicas de diagnóstico por imagen.

Dentro de las ventajas principales se puede decir:

- Es una representación fácilmente comprensible de su capacidad de discriminación en todo el rango de valores.
- No requieren un nivel de decisión particular porque comprende todo el rango posible de valores.
- Es independiente de la prevalencia.

Y las desventajas:

- No se muestran los puntos de corte, sólo se muestran su sensibilidad y especificidad asociadas.
- Tampoco se muestra el número de sujetos.
- Al disminuir el tamaño de la muestra, la curva tiende a hacerse más escalonada y desigual.

El índice de precisión global de la prueba de diagnóstico viene dado por el valor del área bajo la curva, este valor está comprendido entre 0,5 (azar) y 1 (perfecta discriminación).

Swets (Swets, 1988) clasifica la exactitud de la prueba del siguiente modo: si el valor del área está comprendido entre 0,5 y 0,7 entonces la exactitud es baja, si está comprendido entre 0,7 y 0,9 la exactitud es regular-alta (dependiendo de lo que se esté estudiando) y si es superior a 0,9 la exactitud de la prueba es alta.

Por lo tanto, el valor del área bajo la curva resume la curva ROC en su conjunto, la utilización de este valor permite hacer comparaciones de puntos de dos curvas que tengan igual sensibilidad o especificidad y proporciona un enfoque global de comparación de la exactitud de las pruebas comparando sus respectivas áreas bajo la curva. Gráficamente tendrá mayor precisión aquella curva que esté situada más arriba y la izquierda.

Una curva ROC (acrónimo de característica de funcionamiento del receptor) es un gráfico de la sensibilidad frente a (1-especificidad) para un sistema clasificador binario como umbral de discriminación (Hanley & McNeil, 1982).

Los clasificadores discretos como los árboles de decisión (Decision Trees) y los sistemas de reglas (Rule Systems) arrojan resultados numéricos dados como valores de etiquetas binarias. Se proporciona un único punto en el espacio ROC cuando estos clasificadores se usan con un conjunto específico de instancias para clasificar o predecir el rendimiento del clasificador (Ferri, Flach & Hernández-Orallo, 2002).

Para otros clasificadores, como el clasificador Naive Bayes, una red neuronal artificial bayesiana o red, los valores de salida probablemente representen la medida en que uno pertenece a una de dos clases, por ejemplo. De esta forma, se necesita un nuevo indicador.

Este modelo generaliza el caso binario discreto. La idea es considerar la predicción del modelo para cada uno de los valores de la variable y calcular la curva ROC para cada caso individual y tratarla como binaria.

La propuesta se convierte en una variable con n valores discretos (*A = a1, a2, ldots, an*) que llamamos una variable binaria multivaluada, utilizando la función llamada *v (x)* para que pueda construirse para cada uno de los valores discretos asociados a la curva ROC, obteniendo al final n curvas ROC que describan la precisión global de predicción de la calificación (Tabla 3).

Tabla 3. Curva de ROCDM

Curva	Valor	Área bajo la curva
a_0	$v(x) = \begin{Bmatrix} 1 & x = a_1 \\ 0 & x \in \overline{A - a_1} \end{Bmatrix}$	s_1
a_2	$v(x) = \begin{Bmatrix} 1 & x = a_2 \\ 0 & x \in \overline{A - a_2} \end{Bmatrix}$	s_2
…	…	…
…	…	…
a_n	$v(x) = \begin{Bmatrix} 1 & x = a_n \\ 0 & x \in \overline{A - a_n} \end{Bmatrix}$	s_n

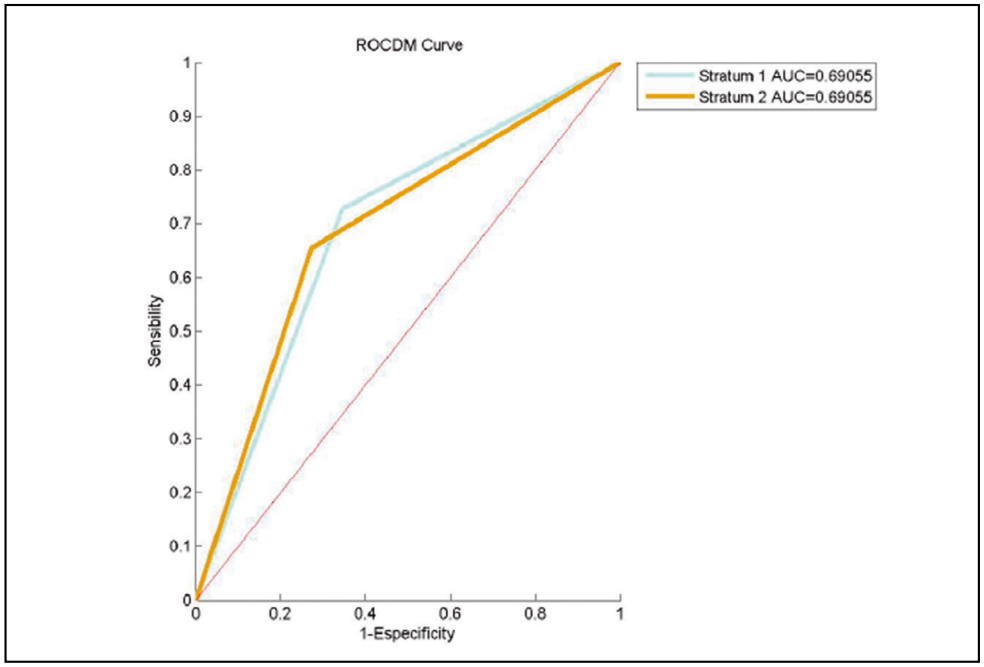

Figura 4. Curva de ROCDM.

De este modo, se obtiene una curva ROC para cada uno de los *n* valores de la variable y cuando se combinan todas las curvas ROC, resulta el gráfico de ROCDM (Figura 4).

La interpretación es la misma que el modelo binario, pero incluye todos los valores variables en el mismo gráfico. El área debajo de cada curva (definida por la diagonal del primer cuadrante) es el valor de la curva ROC para cada valor de la variable, por lo que el valor del área de la curva ROCDM como el promedio de *n* es áreas definidas *atextrmROCDM = verlines1, s2, ldots, sn,* que toma valores entre 1 (prueba perfecta) y 0.5 (prueba inútil).

4. RESULTADOS DE LA IMPLEMENTACIÓN

Las variables a seleccionar para el estudio propuesto se agrupan en las siguientes categorías:

- Factores relacionados con el tráfico
- Factores relacionados con la infraestructura
- Entorno de la carretera

Las variables y sus valores asociados se tomaron de resultados de investigaciones anteriores. Así, no es objeto de la investigación la obtención de variables y sus valores.

El primer paso para el desarrollo del estudio consta de la definición de las variables que permitan reflejar en la mayor medida posible los efectos sobre la seguridad de las interrelaciones entre los valores de las distintas características y los de su variación a lo largo de la carretera partiendo de los datos disponibles. Algunas de ellas son calculadas a partir de la base de datos original. Posteriormente se analiza la correlación de estas variables con los índices de accidentalidad para seleccionar las que efectivamente tienen una mayor influencia en los niveles de seguridad.

Entre las variables consideradas se incluyeron algunas que reflejan la variación de características de cada tramo respecto de los situados en sus inmediaciones, como son el índice cuadrático de la inclinación y la disminución de la velocidad específica respecto de los tramos contiguos. De igual forma se incluyeron variables que dependen de la interacción de otras, como son los límites de velocidad y la visibilidad, que depende de la combinación de los elementos del trazado en planta y alzado, de la sección transversal y de las restricciones al campo de visión del conductor impuestas por la configuración del entorno de la carretera.

Para analizar el grado de asociación de las variables características de la carretera con las que miden la accidentalidad se aplicaron dos procedimientos:

1. Determinación de los índices de correlación entre las variables consideradas para determinar el grado de asociación entre las mismas.
2. Ajuste de curvas de regresión entre las variables con índices de correlación significativos.

El índice de correlación refleja el grado de asociación entre variables. Puede adoptar valores entre -1 y 1. Un coeficiente de correlación de 0 indica que no hay ninguna asociación entre los valores de las dos variables, siendo sus variaciones independientes, mientras que un valor absoluto del coeficiente de correlación igual a 1 indica una asociación perfecta en la que una variable es totalmente dependiente de la otra, existiendo una relación biunívoca entre los valores de ambas. Cuanto más se acerca a 1 el valor absoluto del coeficiente de correlación, mayor es el grado de dependencia entre las variables. El signo del coeficiente de correlación positivo significa que cuando el valor de una variable aumenta el de la otra tiende también a aumentar, mientras que el signo negativo corresponde a casos en los que cuando el valor de una de las variables crece, el de la otra disminuye. Del análisis de valoración y posible correlación entre la variable dependiente (ACV) y las posibles variables explicativas se obtienen los resultados para la selección de las variables finales. Para cada par de variables se obtiene el coeficiente de correlación de Pearson y el "p-valor" de un contraste de significación bilateral. Las correlaciones con un p-valor inferior a 0,005 resultan significativas con un nivel de confianza del 99%, mientras

que si el p-valor es inferior al 0,05 el nivel de significación es del 90%.

De la misma forma se seleccionaron las variables con mayor coeficiente de correlación con el índice de peligrosidad, intrínsecamente relacionado con la frecuencia de accidentes.

Se va a optar para el desarrollo de la metodología por el modelo Naive Bayes. Dicho modelo es muy utilizado debido a que presenta, entre otras, ciertas ventajas:

- Generalmente, es sencillo de construir y de entender.
- Las inducciones de son extremadamente rápidas, requiriendo solo un paso para hacerlo.
- Es muy robusto considerando atributos irrelevantes.
- Toma evidencia de muchos atributos para realizar la predicción final.

Las variables y sus valores asociados se tomarán de resultados de investigaciones anteriores. Así, no ha sido objeto de la investigación la obtención de variables y sus valores, empleándose datos oficiales de la DGT y datos públicos. De forma resumida se puede concluir que los factores que influyen en la frecuencia de accidentes se recogen en "Estimación de sucesos poco probables mediante redes bayesianas" (Soler Flores, 2014).

Las administraciones de carreteras cuentan con inventarios de los datos geométricos de sus carreteras, así como sistemas de gestión de la información que proporcionan una fuente precisa de información para poder ser tratada. Así mismo, a partir de registros de tráfico se dispone de las bases de datos de accidentes de tráfico. Estas fuentes de información son utilizadas generalmente para la tarea de mantenimiento de carreteras e información estadística de la accidentalidad y sus causas. Originalmente, los inventarios contienen registros de las características de la carretera en tramos de 10 metros de longitud, que son tratados para obtener la información correspondiente a las longitudes de tramos consideradas en la calibración de los modelos. Para la obtención de algunas variables puede ser necesario el tratamiento estadístico y computacional de algunas de las existentes o ser necesario efectuar tomas de datos adicionales para, por ejemplo, la localización de las travesías y de los accesos e intersecciones o para establecer la ramificación de la red.

El clasificador Naive Bayes constituye la red bayesiana más sencilla que se puede construir orientada a clasificación. Este clasificador es muy eficiente en ciertos dominios, debido a que es muy robusto ante atributos irrelevantes. Con el objetivo de la simplificación del modelo y comprobando los mejores resultados se trabaja con las siguientes variables descritas en la Tabla 4.

Basado en datos públicos de carreteras españolas durante un período de cinco años, con una selección de tramos de 500 metros y la selección de variables dadas y descritas en la Tabla 4, aplicando el modelo Naive-Poisson, se obtienen las distribuciones de probabilidad de la frecuencia de accidentes para cada tramo y cada variable, y entonces estas variables se discretizan en respuesta a los estratos que figuran en la Tabla 5.

Tabla 4. Descripción de la variable

Variable	Description-unidad
IMD	Intensidad media diaria media en el período de calibración. Tráfico (vehículos/año)
ACC	Número total de accidentes con víctimas durante el período de calibración. Frecuencia de los accidentes. (acv)
DACIN	Densidad de intersecciones y accesos (accesos/km)
INME	Valor medio de la inclinación (rampa o pendiente) en el tramo (%)
PPAD	Proporción de kilómetros en los que está señalizada la prohibición de adelantamiento (%)
RV85M	Disminución de la velocidad específica relativa a los tramos adyacentes de 1 km (km / h)

$$[a_0, a_1, \dots, a_n] = \begin{Bmatrix} 1 & if & x \le a_0 \\ 2 & if & a_0 < x \le a_1 \\ \dots & \dots & \dots \\ n & if & a_{n-1} < x \le a_n \\ n+1 & if & x > a_n \end{Bmatrix}$$

En la presente propuesta el método de discretización empleado es el método de "*modificado manual de los intervalos de las clases*", de manera que el experto del dominio decidirá el número de intervalos, rechazar o corregir instancias con outliers (Soler-Flores, Mayora, y Piña, 2008). Para la discretización de las variables se implementaron diferentes scripts y aplicaciones en Matlab que permitieron realizar esta tarea de manera automática.

Tabla 5. Discretización

Variable	Discretización
IMD	[524.2, 1055.2,2104.4]
ACC	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100]
DACIN	[0, 1, 2, 3, 4, 5, 6, 7, 8]
INME	[10, 15, 20, 25, 30, 35, 40, 45]
PPAD	[0, 0.25, 0.75]
RV85M	[5, 15, 20]

Una vez creada la red Bayesiana es necesaria su evaluación y verificación de su utilidad; por ejemplo, mediante el análisis de sensibilidad para comprobar cómo la variación de valores introducidos como evidencias en ciertas variables afectan a los resultados en el resto de variables. No solo modelan de forma cualitativa el conocimiento, sino que además expresan de forma numérica la fuerza de las relaciones entre las variables. Esta parte cuantitativa del modelo suele especificarse mediante distribuciones de probabilidad como una medida de la creencia que tenemos sobre las relaciones entre las variables de modelo.

En este apartado se recogen los resultados de las medidas de bondad de ajuste seleccionadas para los resultados del procedimiento llevado a cabo.

Es necesario ahora determinar una discretización de los accidentes. Para ello se realizaron hasta 10 análisis diferentes obteniéndose las distintas discretizaciones que se recogen en la Tabla 6. Para su estudio se realiza el ajuste de la Red Bayesiana construida, utilizando el criterio de bondad de ajuste definido.

Tabla 6. Discretización de ACC

Análisis	Discretización
1	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100]
2	[100]
3	[10]
4	[10, 100]
5	[50, 100]
6	[50]
7	[20]
8	[20, 40]
9	[30]

Teniendo en cuenta los valores predictivos y el ajuste de la curva de probabilidad, se selecciona la discretización número 3. Observar eventos raros representa que se dé el caso de más de 10 accidentes en un tramo en el período estudiado, ya que los accidentes de tráfico en la ingeniería de tráfico representan un evento raro, es decir, un evento que ocurre con una probabilidad muy baja. Para seleccionar la mejor predicción, los valores de eventos raros predichos se analizan estudiando los diferentes valores de probabilidad (Figura 5) y probando sus probabilidades (Figura 6).

En esta discretización se han tomado los siguientes cortes: estrato 1, intervalo 20 representados en la Figura 7. De esta forma, la discretización considerada se muestra en la

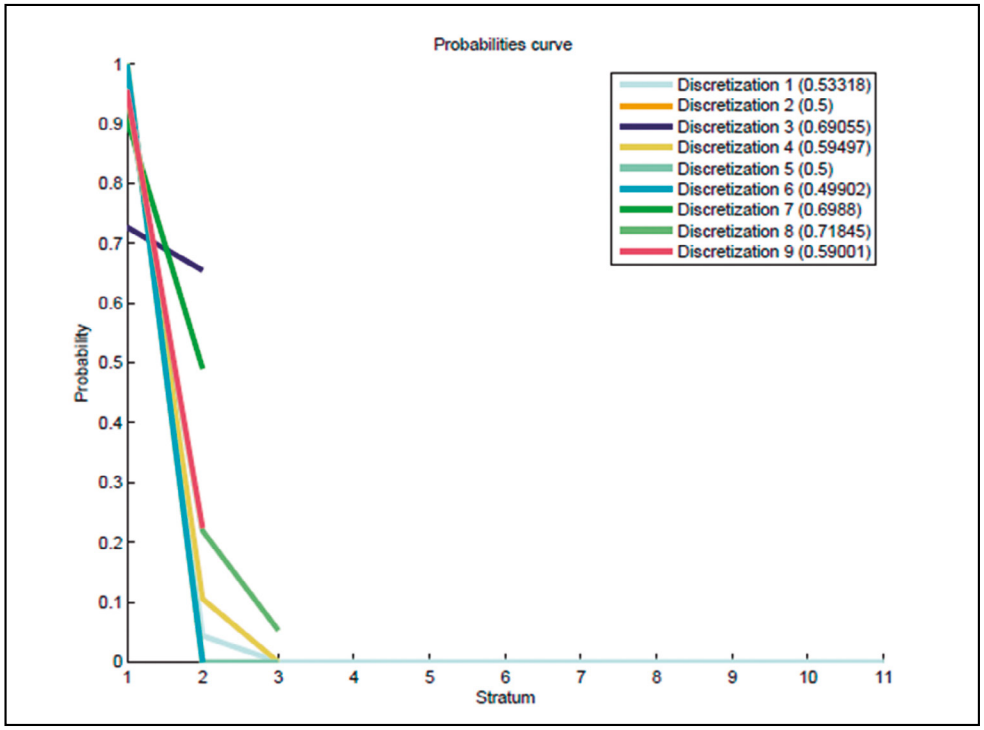

Figura 5. Comparación de probabilidades.

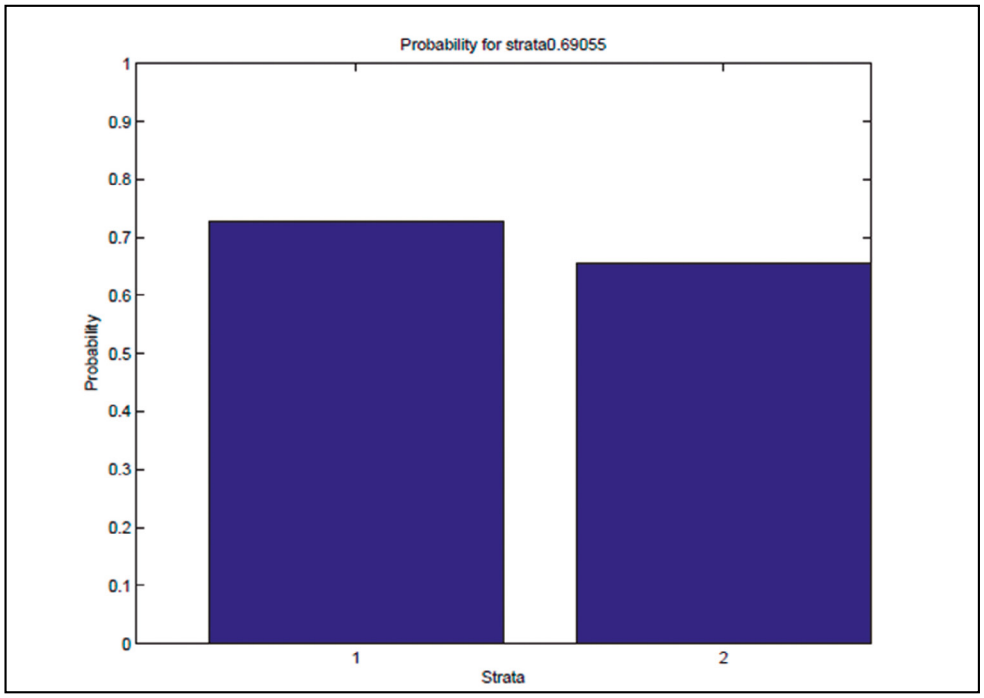

Figura 6. Probabilidades.

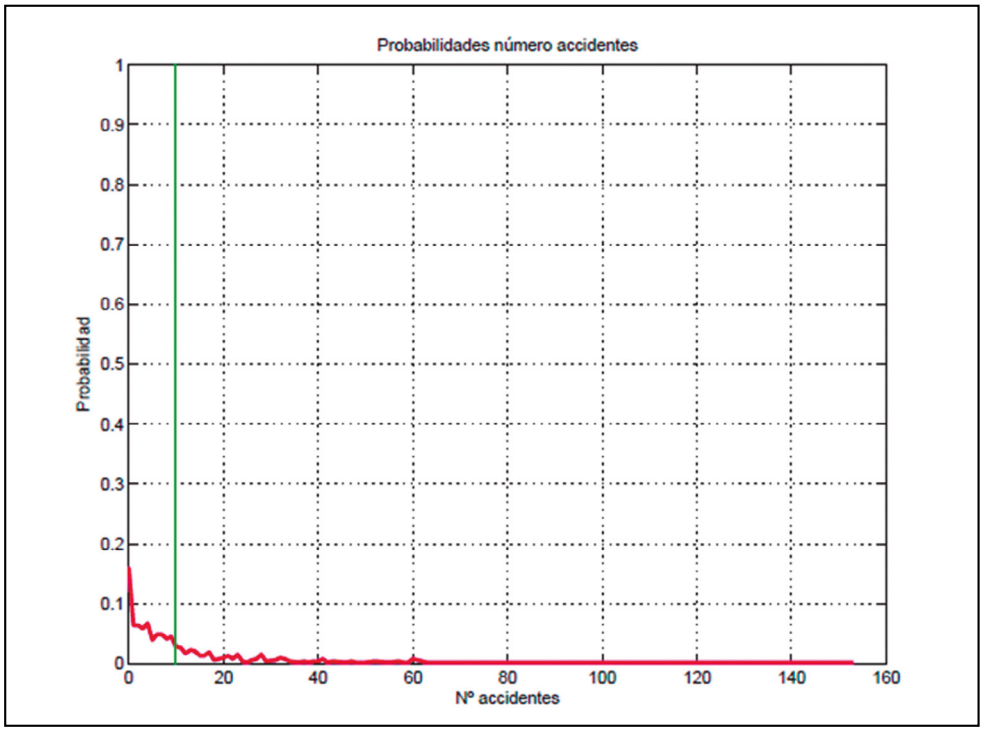

Figura 7. Estratificación de los accidentes.

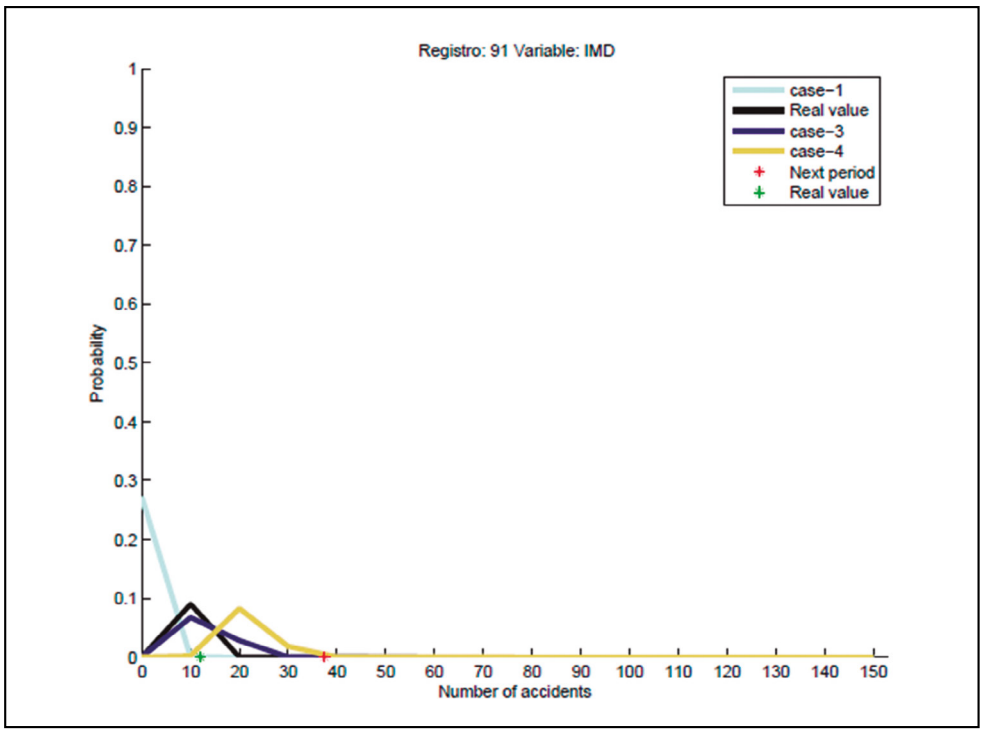

Figura 8. Análisis IMD.

Figura 7 y su curva ROCDM (Figura 4), para un ejemplo de sección seleccionada y el estado de cada una de las variables descritas. En este ejemplo, en cada una de las figuras, se observa la distribución de accidentes para cada uno de los estratos de cada variable y en negro en el estudio de situación.

Mediante el procedimiento de estimación resultante se dispone de las distribuciones de probabilidad asociadas a cada posible caso. Como ejemplos del procedimiento, para algunos casos se tienen las siguientes figuras que comparan diferentes distribuciones para diferentes estratos de las variables indicadas.

Análisis IMD

La Figura 8 muestra que la distribución de accidentes para el caso 3 hace que sea más probable que ocurran más de 10 accidentes, pero las diferencias entre los otros casos no son significativas, es decir, sólo la IMD tiene una alta probabilidad con respecto a los otros para ser eventos raros. Por tanto, lo que representa es la distribución de accidentes en base a los diferentes estratos de la IMD, de forma que, en la Figura 8 se muestra, por ejemplo, la distribución de accidentes de tráfico para el estrato 3 de IMD (IMD superior a 2104) en color azul.

Análisis de PPAD (proporción de kilómetros con prohibición de adelantamiento)

En este caso Figura 9 se muestra que la modificación del estrato de densidad de acceso al tramo 1 (0% de kilómetros con prohibición de adelantamiento) sería una probabilidad menor de eventos raros. Por tanto, lo que representa es la distribución de accidentes en base a los diferentes estratos de la PPAD, por ejemplo en la Figura 9 en azul se define la distribución de accidentes de tráfico para el estrato 3 de PPAD (PPAD superior a 0,75).

Análisis INME (valor medio de la inclinación)

La variable INME en el estrato, en el que se ubica (Figura 10) para el ejemplo de tramo, podría modificarse para mejorar la situación con eventos raros. Así, lo que representa es la distribución de accidentes en base a los diferentes estratos de la IMD, por ejemplo en la Figura 10 en amarillo se tiene la distribución de accidentes de tráfico para el estrato 4 de INME (INME entre 20 y 25).

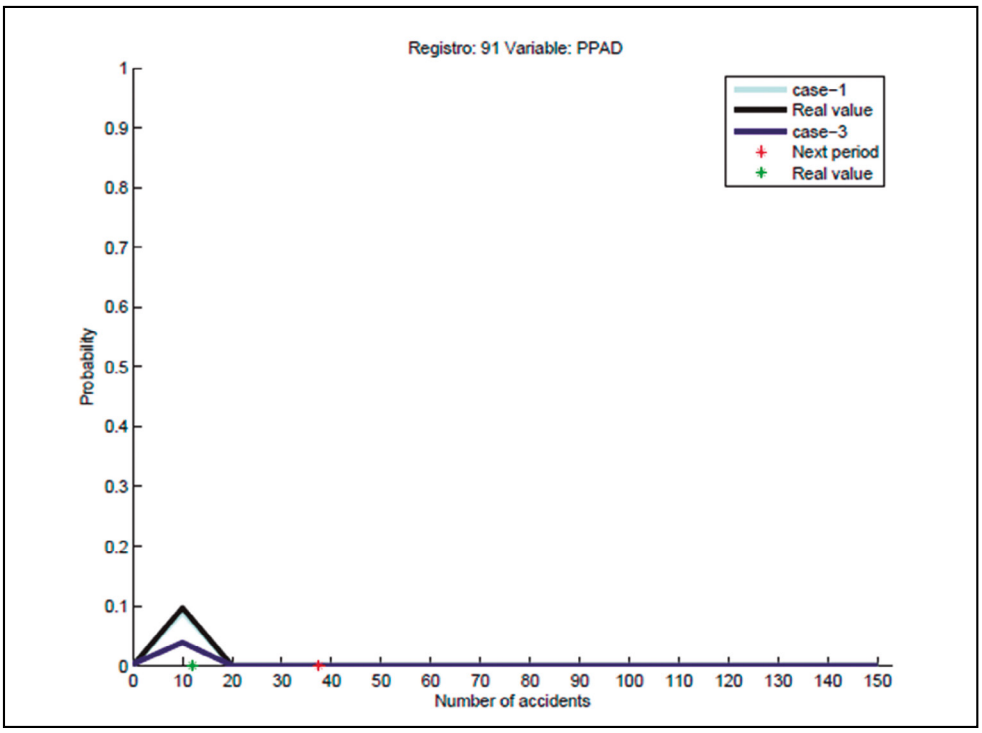

Figura 9. PPAD.

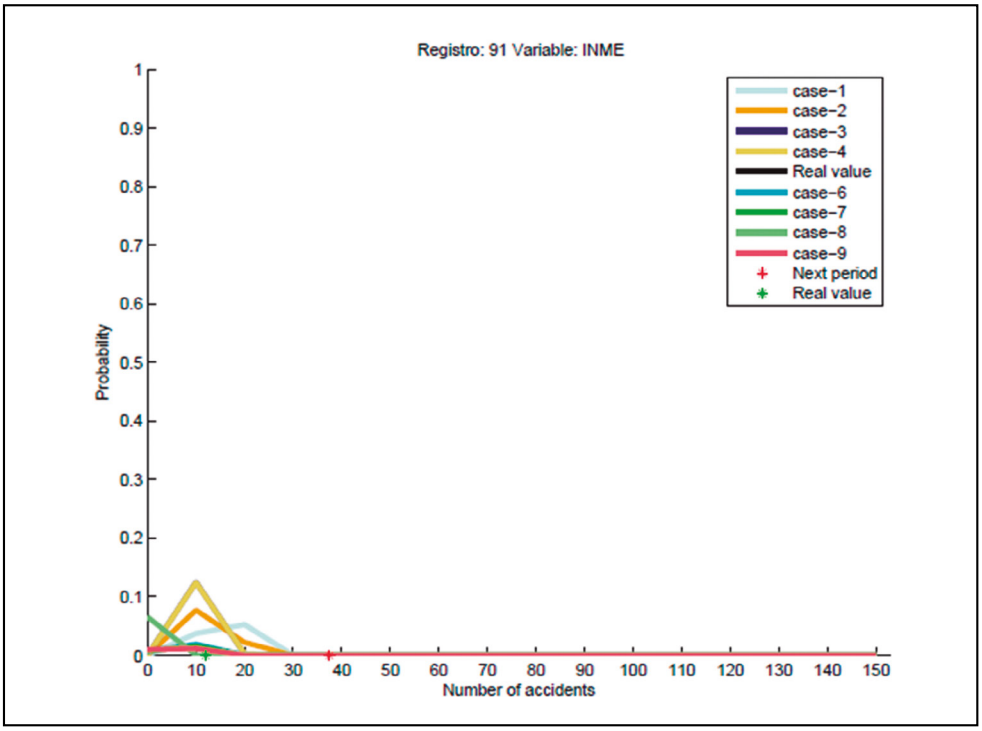

Figura 10. INME.

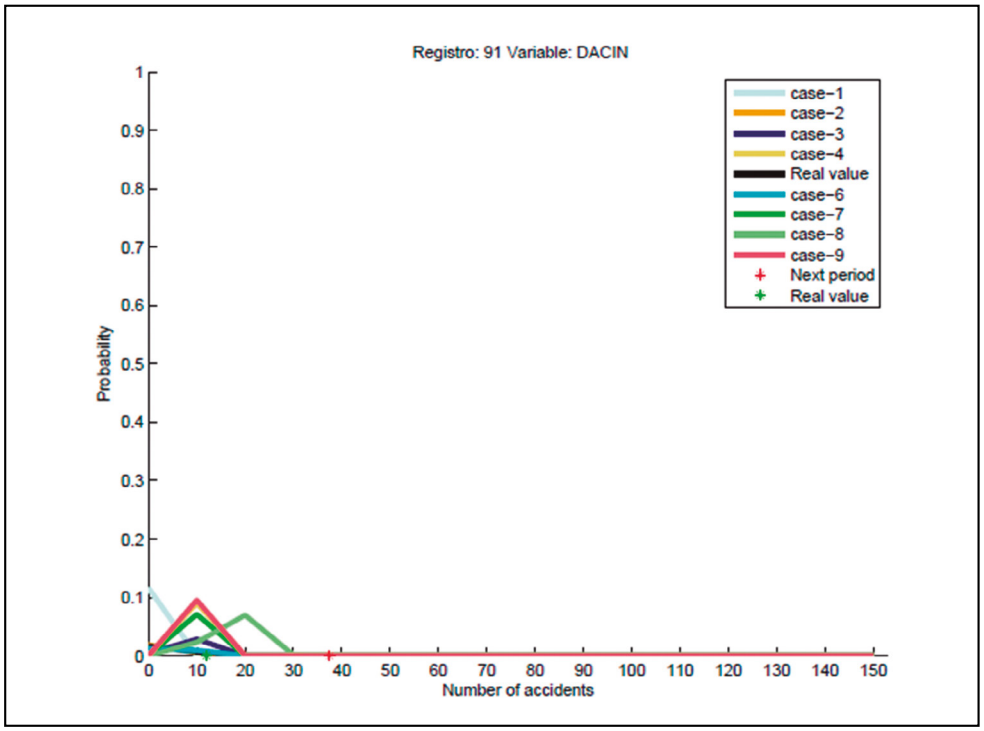

Figura 11. DACIN.

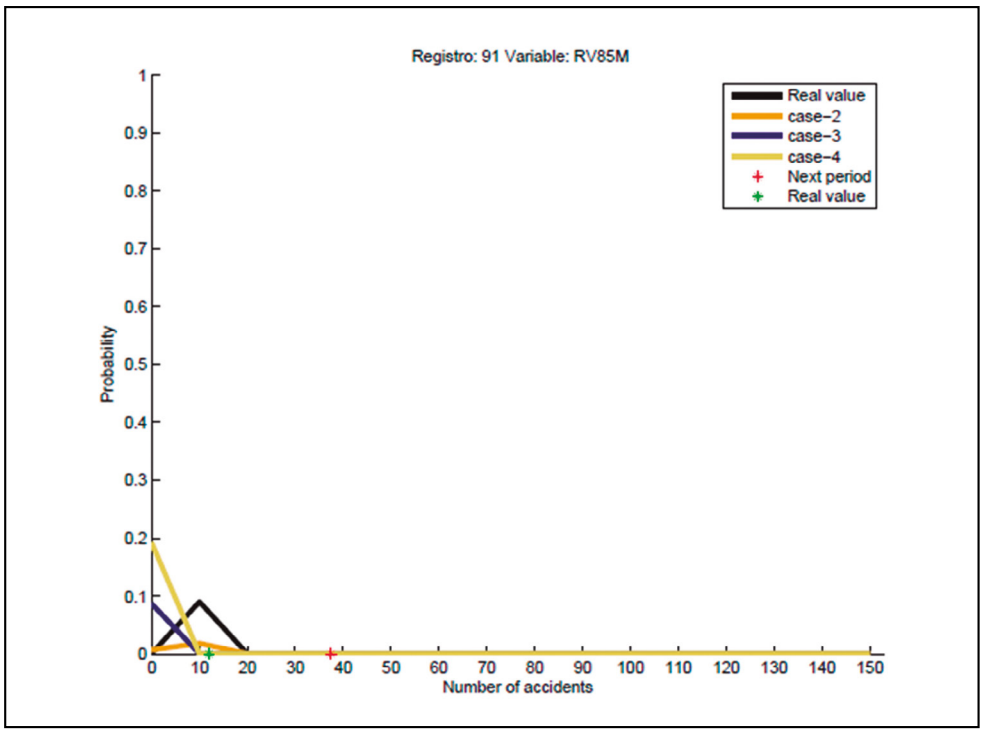

Figura 12. RV85M.

Análisis DACIN (densidad de accesos e intersecciones)

En este tramo particular, la variable DCIN (Figura 11) se puede mejorar, de forma que lo que representa es la distribución de accidentes en base a los diferentes estratos de la variable DACIN.

Análisis RV85M (disminución de la velocidad específica)

La capa proporciona 6 para la variable RV85M (Figura 12) baja probabilidad para la ocurrencia de eventos raros. Así, lo que representa es la distribución de accidentes en base a los diferentes estratos de la variable RV85M, por ejemplo, en la Figura 12 en naranja se tiene la distribución de accidentes de tráfico para el estrato 2 de la variable RV85M (RV85M entre 5 y 15)

5. CONCLUSIONES

Con el desarrollo de la investigación se ha cumplido el objetivo de definir e implementar una metodología que permita estimar a partir de datos a priori y mediante redes bayesianas y la distribución de Poisson la ocurrencia de sucesos de probabilidad de ocurrencia baja.

Se ha desarrollado un modelo basado en las Redes Bayesianas que permite estimar sucesos raros. Se han estudiado los modelos utilizados para estimar sucesos raros como base de conocimiento, analizando las carencias actuales que hacen necesario el desarrollo de este trabajo, de manera que se ha estudiado en detalle la distribución de Poisson asociada a este problema para optimizar el modelo. El modelo desarrollado fundamenta su precisión mediante criterios de bondad de ajuste, además del desarrollo de un criterio que se adapta al caso real propuesto de la frecuencia de los accidentes de tráfico.

Las Redes bayesianas permiten establecer una nueva metodología para estimar sucesos raros y, como caso particular, estimar la frecuencia de accidentes de tráfico.

Además, permite inferir diferentes situaciones para modificar la ocurrencia de los

diferentes sucesos. El modelo propuesto en este trabajo:

- Propone un modelo basado en redes bayesianas para el estudio de sucesos raros. Modelo Naive-Poisson.
- Propone una extensión de la representación gráfica, curva ROC para variables no binarias. Curva ROCDM.
- Presenta el desarrollo computacional de las propuestas.

Este artículo presenta una metodología y una aplicación de dicha metodología para determinar la frecuencia de accidentes de tránsito y controlarla. Las principales contribuciones de este documento se resumen de la siguiente manera:

1. El tratamiento de eventos raros es esencial en problemas reales y es por eso por lo que el desarrollo de un modelo y un método para el tratamiento de estos permitiendo el uso de manera sistemática se hace necesario.
2. A partir de los datos brutos y definiendo la variable a partir de la cual se desea estimar sus eventos de baja probabilidad, el modelo desarrollado define la distribución de probabilidad de los mismos y compara diferentes alternativas, tomando así las decisiones apropiadas en función de los resultados. El modelo Naive-Poisson combina el potencial de calificación del modelo Naive-Bayes para Redes Bayesianas tomado con el ajuste clásico de la distribución de frecuencias de los llamados "eventos raros", la distribución de Poisson.
3. La curva llamada ROC permite dibujar la probabilidad de clasificar correctamente una variable binaria. El modelo proporcionado en este documento, la curva ROCDM, permite extender el original que se puede considerar como variables no binarias.
4. Este documento se desarrolla a partir de su aplicación a casos reales, y los modelos de uso enumerados se presentan para estimar eventos de baja probabilidad y estudiar diferentes alternativas. El modelo desarrollado en este documento es válido para su aplicación a cualquier problema de estimación de ocurrencia de eventos de baja probabilidad.

Ya hemos demostrado que es posible modelar los eventos de baja probabilidad de ocurrencia mediante el uso de Redes Bayesianas y su posible aplicación a ciertos problemas. Tiene grandes impactos e implicaciones prácticas en una amplia gama de aplicaciones. Dado que el marco propuesto es robusto para grandes variaciones dentro de la clase, también se puede utilizar en la industria, en el análisis de fallos. Aunque el trabajo marco propuesto ha superado los métodos existentes, hay mucho más que mejorar, los datos que usamos para evaluar el sistema de propuesta son relativamente simples.

Los aspectos a considerar en futuras investigaciones para continuar el trabajo desarrollado en este documento, se pueden delimitar en las siguientes líneas:

1. Extensión a otros algoritmos de aprendizaje como K2 o DVNSST.
2. Generalización de la metodología a problemas dentro del paradigma de Big Data. Algoritmos Map-Reduce.
3. Desarrollo de aplicaciones implementando la metodología.
4. Aplicación del modelo a otros estudios de casos.

A partir de la distribución resultante obtenida a partir de la Red bayesiana, el ajustar los datos a una distribución de Poisson permite estimar la probabilidad de los diferentes sucesos y comparar varias situaciones a la vez.

En el caso práctico, el modelo Naive-Poisson cuantifica las diferentes probabilidades de las diferentes frecuencias de accidentes y determina su función de densidad de probabilidad. Así, a partir de los modelos construidos se ha estimado una distribución de Poisson para cada caso, esa distribución permite estimar la accidentalidad de una manera no determinística, lo que permite comparar diferentes opciones según los valores de las variables.

Respecto al caso real estudiado, a partir de los resultados se comprueba que las redes bayesianas solventan algunos de los inconvenientes del modelo lineal generalizado y mejora el ajuste en los modelos de predicción de frecuencias de accidentes de tráfico utilizando minería de datos.

6. BIBLIOGRAFÍA

Agueda, C. P. (2011). Causality in sciencie. *Pensamiento Matemático* (1) 12.

Castillo, E., Gutiérrez, J.M. y Hadi, A.S. (1997). Expert Systems and Probabilistic Network Models. Springer Verlag.

Cheon, S.-P., Kim, S., Lee, S.-Y. & Lee, C.-B. (2009) Bayesian networks based rare event prediction with sensor data. *Knowledge-Based Systems* 22 (5) 336–343.

Delgado de la Iglesia, E. (2017). Estudio y simulación de eventos raros mediante el método de aceleración RESTART. Tesis Doctoral. ETSI_Informatica, UPM.

Dougherty, M. (1995). A review of neural networks applied to transport, *Transportation Research Part C: EmeTingTechnologies*, vol. 3, no. 4, pp. 247-260.

Dougherty, J., Kohavi, R. & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features, ICML, 194–202.

Duda, R. O., Hart, P. E. & Stork, D. G. (2000). Pattern classification, NY WileyID: 129

Ebrahimi, A. & Daemi, T. (2010). Considering the rare events in construction of the bayesian network associated with power sys-

tems. Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on, IEEE, 659–663.

El-Gheriani, M., Khan, F., Zuo, M.J. (2017). Rare Event Analysis Considering Data and Model Uncertainty. ASCE-ASME Journal of Risk and Uncertainty. *Engineering Systems, Part B: Mechanical Engineering*, vol. 3, no 2, p. 021008.

Fernández, B. (2008). La Ley de los Eventos Raros, Legado de Siméon Denis Poisson. Memorias Escuela Regional de Probabilidad y Estadística, Villahermosa, UJAT.

Ferri, C., Flach, P., & Hernández-Orallo, J. (2002, July). Learning decision trees using the area under the ROC curve. *ICML*, Vol. 2, pp. 139-146.

Flores, F. S.; Mayora, J. P.; Piña, R. J. (2008). Tratamiento de outliers en los modelos de predicción de accidentes de tráfico. VIII Congreso de Ingeniería del Transporte, España.

Flores, F.S.; Varela, J.A.O.; González, M.D.L. (2015). Sucesos raros en Ingeniería de Tráfico. *Pensamiento Matemático*, vol. 5, no 1, p. 63-74.

García, S., Luengo, J., Saez, A., Lopez, V., Herrera, F. (2013). A survey of discretization techniques: taxonomy and empirical analysis in supervised learning IEEE Transactions on Knowledge and Data Engineering, pp. 734-750

Green, D.M. and Swets, J.A. (1966) Signal Detection Theory and Psychophysics. Wiley, New York.

Greiner, Matthias & Gardner, Ian. (2000). Epidemiologic Issues in the validation of veterinary diagnostic tests. Preventive veterinary medicine. 45. 3-22. 10.1016/S0167-5877(00)00114-8.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

Holmes, D. E., Tweedale, J., Jain, L. C. (2012). Data mining techniques in clustering, association and classification. *Data Mining: Foundations and Intelligent Paradigms* p. 1–6.

Jimenez, O.L.; Garrido, G.C.B.; Rodríguez, C.R.B. (2018). Comparación de clasificadores sobre multiples datasets con pruebas estadísticas no paramétricas. *Universidad&Ciencia*, vol. 7, no 2, p. 64-82.

Kim, J. H. & Pearl, J. (1983) A computational model for causal and diagnostic reasoning in inference systems, Proceedings of the 8th International Joint Conference on Artificial Intelligence, Citeseer, 190–193.

Kim, S.-B., Seo, H.-C., Rim, H.-C. (2003). Poisson naive bayes for text classification with feature weighting, Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11, Association for Computational Linguistics, 33–40.

Li, Y. (2007). Control of spatial discretisation in coastal oil spill modelling I*nt. J. Appl. Earth Observ.*, 9, pp. 392-402

Lima, M.D., Nassar, S.M., Rodriges, P.I.R., Freitas-Filho, P., Jacinto, C.M.C. (2014). Heuristic discretization method for Bayesian networks *J. Comput. Sci.*, 10 (5), pp. 869-878

Liu, H., Hussain, F., Lim, C., Dash, M. (2002). Discretization: an enabling technique *Data Mining Knowl. Discov*., 6, pp. 393-423

Melchers, R. E., & Beck, A. T. (2018). Structural reliability analysis and prediction. John Wiley & Sons.

Ordoñez, H. O. La complejidad del problema de la inseguridad vial. CARRETERAS, 2014, vol. 13, no 1, p. 105.

Pearl, J. (2000). Causality: models, reasoning and inference, Vol. 29, Cambridge Univ Press.

Quishpe Tasiguano, I.D.(2015). Factores de riesgo de siniestralidad y cálculo de primas de los vehículos asegurados en el Ecuador mediante modelos lineales generalizados. Tesis de Licenciatura. Quito: EPN, 2015.

Rodríguez García, T. (2015). Predicción de tráfico de contenedores a corto plazo mediante técnicas de minería de datos: redes neuronales artificiales y redes bayesianas. Tesis Doctoral. ETSI Caminos, Canales y Puertos, UPM.

Ratner, B. (2017). Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data. Chapman and Hall/CRC.

Rodríguez, J. M., Medina, M. H., Campuzano, J. C., Bangdiwala, S. I., & Villaveces, A. (2013). Methodological proposal for implementing an intervention to prevent pedestrian injuries, a multidisciplinary approach: the case of Cuernavaca, Morelos, Mexico. *Injury prevention*, injuryprev-2013.

Scaillet, O., Treccani, A., & Trevisan, C. (2017). High-frequency jump analysis of the bitcoin market.

Schaffer, C. (1993). Mach Learn. 13: 135. https://doi.org/10.1007/BF00993106

Seiffert, Chris, et al. (2007). Mining data with rare events: a case study. *Tools with Artificial Intelligence*. ICTAI 2007. 19th IEEE International Conference on. IEEE, p. 132-139.

Soler Flores, F. (2014). Estimación de sucesos poco probables mediante redes bayesianas.

Soler-Flores, F., Mayora, J. M. P., y Piña, R. J. (2008). Tratamiento de outliers en los modelos de predicción de accidentes de tráfico. VIII CIT, 02/07/2008-04/07/2008, La Coruña, España.

Sucar, L. E. (2006). Redes bayesianas

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. Science, 240(4857):1285–1293.

Tomz, M.; King, G.; Zeng, L. (2003). ReLogit&58; Rare Events Logistic Regression. *Journal of statistical software*, vol. 8, no 1, p. 1-27.

Uusitalo, L. (2007). Advantages and challenges of bayesian networks in environmental modelling. *Ecological Modelling* 203 (3), 312–318.

Vázquez, M.Y.L., et al. (2018). Obtención de modelos causales como ayuda a la comprensión de sistemas complejos. *Revista de la Facultad de Ciencias Médicas,* vol. 18, no 2.

Weiss, G. M., & Hirsh, H. (1998). Learning to Predict Rare Events in Event Sequences. In *KDD*, 359-363.

Yang, Y., Webb, G., Wu, W. (2010). Discretization methods Data Mining and Knowledge Discovery Handbook, Springer, pp. 101-116

Zhou, Y., Fenton, N., Neil, M. (2014). Bayesian network approach to multinomial parameter learning using data and expert judgments. *J. Approx. Reason*, 55, pp. 1252-1268

Zweig, M. H. and Campbell, G. (1993). Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4):561–577