

Universidad de Lima
Facultad de Ingeniería y Arquitectura
Carrera de Ingeniería de Sistemas



SISTEMA DE ALERTA Y NOTIFICACIÓN DE VARIACIONES EN EL PATRÓN DE CONDUCCIÓN

Tesis para optar el grado académico de título en Ingeniería de Sistemas

Kevyn Pool Franco Caldas

Código 20120513

Asesor

Juan Manuel Gutiérrez Cárdenas

Lima – Perú
Marzo del 2021

SISTEMA DE ALERTA Y NOTIFICACIÓN DE VARIACIONES EN EL PATRÓN DE CONDUCCIÓN

Kevyn Pool Franco Caldas
20120513@aloe.ulima.edu.pe
Universidad de Lima

Resumen Cada año se incrementa en 3% la cantidad de fallecidos por accidentes en el área urbana (INEI, 2018), debido a una conducción imprudente y sistemas de seguridad poco eficientes para evitar accidentes automovilísticos. En Perú se hace notoria la necesidad de un sistema que permita alertar de manera auditiva a los conductores cuando su patrón de conducción difiera de lo habitual y evitar accidentes automovilísticos.

La presente investigación tiene como objetivo el desarrollo un sistema que permita generar alertas cuando se detecte una conducción que difiera de los patrones habituales, aplicando las metodologías Modelo de Red Neuronal de Kohonen y el Modelo oculto de Markov e implementarlas en un aplicativo móvil para Android propio. En la experimentación se determinaron dos estados de conducción: “no accidente” y “previo accidente”, donde los conductores entre 20 y 30 años, dentro de un área controlada, mostraban en promedio que el 86% de las veces aceleraban cuando se encontraban en el estado de “previo accidente”, pero también se observó que la probabilidad promedio de cambiar del estado “previo accidente” a “no accidente” fue del 91%, lo que indico una tendencia de conducción dentro de los patrones habituales. Los resultados obtenidos para la generación de patrones de conducción permiten alertar al conductor cuando su estilo de conducción difiere de los patrones habituales y evitar accidentes automovilísticos. Este hallazgo contribuirá a mejorar nuestra comprensión de los vínculos entre la variabilidad de patrones de conducción y la incidencia de accidentes automovilísticos.

Palabras Clave: Smartphone, OBDII, Patrones, Accidentes automovilísticos, Clasificación Hidden Markov Model, Kohonen Model.

Abstract: Each year, the number of deaths from accidents in urban areas increases by 3% (INEI, 2018), due to reckless driving and inefficient safety systems to avoid car accidents, therefore, in Peru, the Need for a system that allows drivers to be audibly alerted when their driving pattern differs from usual and to avoid car accidents.

The objective of this research is to develop a system that allows to generate alerts when driving that differs from the usual patterns is detected, applying the Kohonen Neural Network Model methodologies and the hidden Markov Model and implementing them in a mobile application for Android itself. . In the experimentation, two driving states were determined: "No accident" and "Previous Accident", where drivers between 20 and 30 years old, within a controlled area, showed on average that 86% of the time they accelerated when they were on the road. "Previous Accident" status, but it is also pointed out that the average probability of changing from "Previous Accident" to "No accident" status was 91%, which indicates a driving trend within the usual patterns. The results obtained for the generation of driving patterns allow the driver to be alerted when his driving style differs from the usual patterns and to avoid car accidents. This finding contributes to improving our understanding of the links between variability in driving patterns and the incidence of motor vehicle accidents.

Keywords: Smartphone, OBDII, Patterns, Car accidents, Hidden Markov Model Classification, Kohonen Model.

1. INTRODUCCIÓN

El desarrollo tecnológico automovilístico como indican Zaldivar, Calafate, Cano, y Manzoni (2011), en materia de seguridad humana ha mostrado un crecimiento acelerado respecto a años anteriores, impulsado por la implementación de sistemas de seguridad para la prevención de accidentes. Existen proyectos en Europa encargados de abordar la seguridad en las carreteras, como, por ejemplo: ASSET-ROAD (EUROPA - ASSET Advanced Safety and Driver Support in Essential Road Transport | TRIMIS - European Commission, 2011), Cyber-Cars2 (EUROPA - Close Communications for Cooperation between Cybercars | TRIMIS - European Commission, 2009) y SAFETRIP (Final Report Summary - SAFETRIP (Satellite application for emergency handling, traffic alerts, road safety and incident prevention) | Report Summary | SAFETRIP | FP7 | CORDIS | European Commission, 2013). Por lo tanto, debido al creciente número de accidentes registrados en el país que asciende al 3% anual (INEI, 2018) y tomando como referencia diferentes investigaciones que buscan salvaguardar la vida, se tiene como objetivo para esta investigación desarrollar un sistema clasificador de patrones de conducción que alerte al conductor si su estilo de conducción difiere de sus patrones habituales y evitar accidentes automovilísticos, para lograr este objetivo se aplicarán los modelos de Red Neuronal de Kohonen y el Modelo

Oculto de Markov con información obtenida del puerto OBDII y sensores del Smartphone, la integración de los dos modelos antes mencionados se logró utilizando la investigación de Morimoto H. (2016) que menciona la necesidad de usar una variable que permita desarrollar ambas metodologías, en este caso se optó por usar la variable de ACELERACIÓN que se encuentra relacionada a la generación de patrones de conducción con el Modelo de Kohonen y que a la vez permite determinar a través de conteo las matrices de transición, emisión y probabilidades iniciales para el Modelo Oculto de Markov. Entonces el proceso requirió: recolectar datos de los sensores en ECU (Engine Control Unit) Automotriz y los sensores del Smartphone, utilizar la metodología del Modelo de Kohonen para generar patrones de conducción, utilizar la metodología del Modelo Oculto de Markov para generar estados de conducción, y por último evaluar y alertar al conductor si su patrón de conducción difiere de lo habitual con notificaciones audibles.

2. ESTADO DEL ARTE

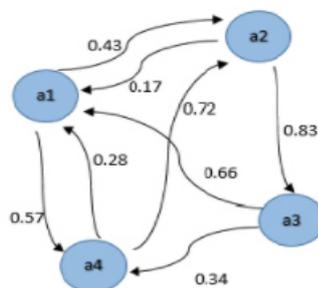
La investigación abordó los beneficios de las tecnologías On Board Display (OBD) y los sensores del Smartphone con aplicaciones en la seguridad automovilística. Meseguer, Calafate, Cano, y Manzoni (2013), muestran una gama de posibilidades que se presentan cuando se combinan los vehículos con los sensores del Smartphone, un ejemplo puede ser diagnosticar el estado del vehículo mediante el monitoreo constante de los sensores implementados en el motor, también es posible generar patrones de conducción a través de diferentes metodologías de clasificación para su posterior análisis.

Uno de los temas pertinentes a la investigación es la prevención de accidentes automovilísticos tomando como punto base los patrones de conducción, con el objetivo de salvaguardar la vida humana. Desde esta perspectiva, podemos tomar como base la investigación de Morimoto H. (2016), realizada en la ciudad de Nagoya en Japón, utilizó como punto de partida la red de Kohonen y el Modelo Oculto de Markov con el objetivo de determinar la relación entre los patrones generados con datos climatológicos y los accidentes cerebrovasculares. Demostró que la red de Kohonen era un método efectivo para obtener patrones climáticos que servirían como "estados" de los modelos ocultos de Markov. Estos modelos se componen de una matriz de emisión, una matriz de transición y una matriz de probabilidades iniciales. La matriz de emisión se utiliza para mostrar la probabilidad de emitir algún estado basándose en la observación, la matriz de transición muestra la probabilidad de transitar de un estado anterior a un nuevo estado y la matriz de probabilidades iniciales muestra la probabilidad de iniciar en algún estado. Sus aplicaciones sobre Markov proporcionaron modelos efectivos para aclarar la relación existente entre los estados climáticos a las incidencias de accidentes cerebrovascular. La efectividad de los modelos ocultos de Markov se estimó mediante una prueba estocástica para los errores cuadráticos medios. Los Modelos Ocultos de Markov con estados generados por la red de Kohonen sirvieron como una descripción de la relación existente entre estados climáticos y las incidencias de accidentes cerebrovasculares. Esta perspectiva nos muestra la posibilidad de trabajar con modelos no relacionados aplicando información que permitan auto relacionarse y complementarse para lograr un objetivo específico, salvaguardar la vida humana. Esta investigación se utiliza como marco de referencia para lograr generar una interacción entre el Modelo de Kohonen y el Modelo Oculto de Markov.

Estas metodologías no son las únicas que se han utilizado para la prevención de accidentes automovilísticos tomando como punto base los patrones de conducción. Zaldivar, Calafate, Cano, y Manzoni (2011), hacen mención sobre la capacidad de los sistemas para percibir las variaciones de los patrones habituales, los cuales, generados a través de información recuperada de la ECU automotriz, y así poder tomar acciones que maximicen las posibilidades de una exitosa transmisión del mensaje de alerta de variación de patrones de conducción a diversos receptores. Desde esta perspectiva, la seguridad automovilística posee diversas maneras de ser abordada. Min Goo Lee, Yong Kuk Park y Kyung Kwon Jung (2011), en su investigación realizada en Corea de Sur, presentan la aplicación de un modelo polinómico lineal cuadrático para el cálculo de eficiencia de combustible en base a las variables revoluciones por minuto y posición del cuerpo de aceleración recuperadas de la ECU automotriz, este modelo permite generar una clasificación de conducción y mostrar si el estilo de conducción se torna agresivo para el consumo de combustible. Este modelo ha mostrado 71% de exactitud para predecir si el consumo de combustible se sale de los patrones habituales y verificar el manejo agresivo afecta el consumo de combustible a futuro. Adicionalmente, para mejorar el estilo de conducción, se utilizó la retroalimentación hacia los conductores, mostrándoles sus niveles de consumo actual y posteriormente reevaluando el modelo de consumo de combustible para mostrar avances positivos o negativos. A pesar del éxito obtenido con este modelo de clasificación lineal, se pueden encontrar diversos métodos de clasificación, Amarasinghe y otros (2015), en su investigación realizada en Sri Lanka, muestran que mediante el uso del modelo de Markov (Figura. 1), a través del cálculo de las variables de aceleración y desaceleración, que fueron recuperadas de la ECU automotriz, y aplicando un procesador de eventos complejos llamado Sindhi Query, se logra crear patrones de conducción analizando los datos observables y de esta manera alertar al conductor, de manera visual o auditiva sobre si su comportamiento frente al volante se

torna agresivo y puede generar riesgo para su vida o para la de los demás.

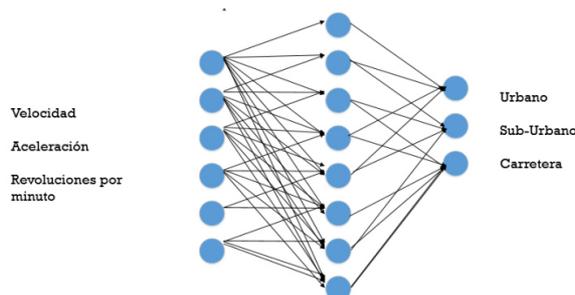
Figura 1: Diagrama de transición de aceleración



Fuente: Amarasinghe y otros. (2015).

Teniendo en cuenta que el modelo de Markov no es el único referente en tema de clasificación. Según Meseguer, Calafate, Cano y Manzoni (2013) realizada en España, apreciamos el uso del modelo de BackPropagation (Figura. 2) como clasificador para estilos de conducción, donde hacen uso de variables tales como velocidad, aceleración y revoluciones del motor que se recuperaron de la ECU automotriz, donde se utilizó información recopilada en un periodo de dos meses, dando como resultados conducción más segura al controlar aceleraciones bruscas y altas velocidades.

Figura 2: Red Neuronal BackPropagation



Fuente: Meseguer, Calafate, Cano y Manzoni. (2013).

De las metodologías de clasificación expuestas en la parte superior, se incluye el Modelo de Redes Neuronales de Kohonen, donde según la investigación de R.Dange y otros (2017), realizada en Italia, a través del ingreso de variables como aceleración y frenado, se logra obtener un mapa del patrón de conducción y de esta manera observar un patrón de conducción agresiva y moderada en un rango de tiempo.

No obstante, se debe tener en cuenta que las metodologías de clasificación son capaces de trabajar en paralelo bajo la misma estructura de datos, como menciona Le Bastard y otros (2011) en su investigación desarrollada en Estados Unidos, en la cual, a través de 3 metodologías de clasificación vehicular permiten diferenciar automóviles y camiones. En primer lugar, se utilizó el clasificador Bayesiano que aplica la búsqueda del mejor vector a partir de un conjunto de vectores, asumiendo N como vector de ruido, donde todos los elementos no están correlacionados, de igual manera, se aplica el método de lógica difusa, donde no se asume un vector de ruido y se ejecuta una función trapezoide, la cual genera el mejor resultado al convertirse en un triángulo, por último, se utiliza el método Support Vector Machine (SVM), donde se busca el punto más alejado del hiperplano para obtener la mejor clasificación.

3. ANTECEDENTES

De las diferentes investigaciones analizadas en el Estado del arte, se extraen diversas metodologías que son ideales para generar patrones de conducción. Se optó por aplicar en esta investigación la Red Neuronal de Kohonen y Modelo Oculto de Markov.

3.1. Red neuronal de Kohonen

La Red Neuronal de Kohonen tiene la característica principal de poseer una variable por cada entrada o nodo, por otro lado, en la capa de salida se ejecutan los procesos y se organizan los resultados de acuerdo con las características de cada entrada, además el flujo de información está conformado por las entradas que están conectadas a las salidas a través del peso sináptico W_{ij} . En una ejecución estándar los pesos permanecen estáticos, pero con el objetivo de mejorar la distribución de los datos, los pesos se inicializan de manera aleatoria con valor máximo igual a uno. El proceso inicia mediante la ejecución de cada neurona de la capa de procesamiento, se calcula la similitud entre el vector de entrada y vector de pesos y vence la que tiene mayor similitud al centroide, adicionalmente por función de vecindad las neuronas vecinas ajustan sus pesos después de generarse la ganadora.

La fórmula de aprendizaje está dada por:

$$D^2(W_{ij}, X) = \sum_k (W_{ij} - X_k)^2 \quad (1)$$

(Kohonen, 2001)

D = Distancia
 W = Peso asignado a la neurona
 i = Entrada en la j-ésima neurona
 j = Vector de Peso de j-ésima neurona
 X = Neurona de Entrada
 k = k-ésima posición de vector de entrada

Generada la neurona ganadora, se tienen que actualizar los pesos sinápticos propios y de los vecinos y se da por la siguiente fórmula:

$$W_{ij}(t+1) = W_{ijk}(t) + \alpha(t) * h(|i-g|, t) * (x_k(t) - W_{ijk}(t)) \quad (2)$$

(Kohonen, 2001)

t = Representa el tiempo
 g = Unidad de mejor correspondencia
 $\alpha(t)$ = Representa el ratio de aprendizaje.
 $h(|i-g|, t)$ = Representa la Función de vecindad.
 $(x_k(t) - W_{ijk}(t))$ = Representa la diferencia entre la entrada y el peso actual.

Adicionalmente la actualización del ratio de aprendizaje viene dada por la siguiente fórmula:

$$\alpha(t) = \alpha_0 + (\alpha_f - \alpha_0) * (t/t_\alpha) \quad (3)$$

(Kohonen, 2001)

Donde:
 α_0 = Ratio Inicial.
 α_f = Ratio Final.
 t/t_α = Numero de iteración sobre el máximo número de iteración.

3.2. Modelo de Markov Estándar

Este modelo se utiliza para el análisis del comportamiento de procesos estocásticos, es decir modelos que evolucionan a lo largo del tiempo dentro de un conjunto de estados donde la probabilidad del próximo estado se basa en función del estado anterior. La Tabla 1 muestra un ejemplo de una matriz de transición con las probabilidades de desencadenar un estado mientras se encuentra en un estado anterior, para su correcta aplicación se debe tener en cuenta las tres leyes de Markov:

- i. El modelo debe estar conformado por una matriz de transición cuadrada.
- ii. Los números dentro de las celdas debe estar entre 0 y 1.
- iii. La suma de probabilidades por fila debe ser igual a 1.

Tabla 1

Matriz de Transición, tabla ejemplo de estados con probabilidades de desencadenarse

Clima	Soleado	Nublado	Lluvioso
Soleado	0.5	0.3	0.2
Nublado	0.2	0.5	0.3
Lluvioso	0.15	0.45	0.4

Entonces se puede definir que el Modelo de Markov está definido por:

$$M = (Q, \pi, A) \tag{4}$$

(R. Durbin, S. Eddy, A. Krogh, 1998)

Donde Q representa el conjunto de estados que incluye un estado final F, π representa el valor de la probabilidad inicial y A representa el valor de la probabilidad de transición.

Donde la probabilidad de transición está dada por:

$$P(X_{t+1} = j | X_t = i) \tag{5}$$

(R. Durbin, S. Eddy, A. Krogh, 1998)

Donde P es la probabilidad de transicionar de un estado anterior a un nuevo estado, X es el estado, t es el tiempo del estado, j es el estado futuro e i es la probabilidad de que se dé el estado X en el tiempo t, que es entendida como la probabilidad de que X pase al estado j en el tiempo t+1 dado que X era i en el tiempo t.

3.3. Modelo Oculito de Markov

El Modelo Oculito de Markov a diferencia del modelo estándar está conformado por:

$$M = (Q, \epsilon, \pi, A, B) \tag{6}$$

(R. Durbin, S. Eddy, A. Krogh, 1998)

Donde Q representa el conjunto de estados que incluye un estado final F, ϵ representa el conjunto de símbolos o alfabetos finitos, π representa el valor de la probabilidad inicial, A representa el valor de la probabilidad de transitar a un nuevo estado y B representa la probabilidad de emitir un estado final.

Como se observa, a diferencia del modelo estándar, el modelo oculito de Markov nos agrega adicionalmente la probabilidad de emisión y un conjunto de símbolos alfabéticos, sin embargo, se mantiene la premisa del Modelo Estándar de Markov, la cual expresa que el estado en el tiempo $t+1$ solo depende del estado en el tiempo t y los estados visitados permanecen ocultos, de esta manera se tiene que al pasar del estado X_t al estado X_{t+1} sin conocer todos los estados anteriores y poder generar una cadena definida por ϵ , el cálculo de la probabilidad viene dada por:

$$P(X) = \sum q = q_1 + q_2 + \dots + q_F \tag{7}$$

(R. Durbin, S. Eddy, A. Krogh, 1998)

Que es definido como la sumatoria de toda secuencia de estados que genero la cadena donde q representa el estado en el tiempo t.

3.4. Heuristic Measure Hi-SOM

Esta ecuación propuesta por Benabdeslem y Lebbah (2007) permite calcular la importancia de las variables utilizadas en el Modelo de Kohonen y está conformado por:

$$H_{jc} = \gamma c * \frac{W_{cj}}{\sum_{k=1}^d W_{ck}} \quad (8)$$

(Benabdeslem y Lebbah, 2007)

$$\gamma c = \frac{Card(c)}{N} \quad (9)$$

(Benabdeslem y Lebbah, 2007)

$$H_j = \sum_{c=1}^N H_{jc} \quad (10)$$

(Benabdeslem y Lebbah, 2007)

Donde H_j representa la sumatoria del caculo de importancia de cada agrupación definida en H_{jc} , H_{jc} es la importancia de cada variable asociada a cada neurona, c representa la neurona, W_{cj} representa es el peso de una neurona, $\sum_{k=1}^d W_{ck}$ representa la sumatoria de todos los pesos de las neuronas, $Card(c)$ representa el numero elementos asignados a cada neurona, N representa el total de registros y γc representan la densidad de cada neurona.

Además, para verificar que la selección de características sea la correcta se utiliza una verificación del error topológico que presenta el porcentaje de vectores de datos, para los cuales la primera y la segunda mejor unidad de coincidencia, no son unidades adyacentes y se representa de la siguiente forma:

$$C_p = argmin TE(p) \quad (11)$$

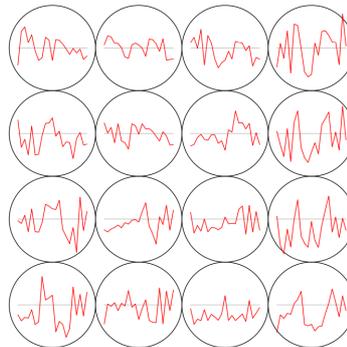
(Benabdeslem y Lebbah, 2007)

Donde C_p representa el mínimo error topológico calculado para todas las variables y $TE(p)$ representa el error topológico estimado.

3.5 Implementación del Modelo de Kohonen y Modelo Oculto de Markov

Según Hiroshi M. (2016), tomó como punto de partida utilizar información climática e información de transportes de emergencia en la ciudad de Nagoya – Japón entre los años 2002 hasta 2004 para determinar si existía una relación entre los accidentes cerebrovasculares y el clima. El primer paso fue clasificar la información climática, para esto el autor utiliza el Modelo de Kohonen y genera un Mapa Auto Organizado (SOM) tomando una grilla de 16 neuronas que generó lo siguiente:

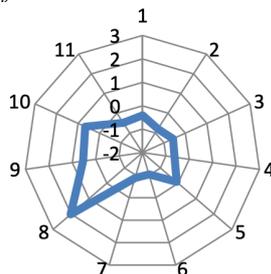
Figura 3: Mapa Auto Organizado de Estados Climáticos



Fuente: Hiroshi M. (2016).

El gráfico en cada unidad mostraba el valor de vector que incluye valores como: (nivel del mar-presión, temperatura máxima, temperatura mínima, precipitación, etc) de los días pertenecientes a la unidad correspondiente, al examinar el detalle de cada grafica se obtuvo:

Figura 4: Grafica Cobweb de Estado Climatico Abril



Fuente: Hiroshi M. (2016).

El gráfico de Cobweb por ejemplo para abril de 2002 hasta 2004 Figura 4 mostró que una alta velocidad del viento (número 8) era una característica de la unidad. El número en el gráfico significaba parámetros, como 1 presión al nivel del mar, 2 temperatura media, 3 máxima temperatura, 4 mínima de temperatura, 5 rango diario de temperatura, 6 presión de vapor de equilibrio, 7 relativa humedad, 8 velocidad del viento, 9 horas de sol, 10 cantidad de radiación solar global, 11 precipitación. Todos los parámetros se escalaron para que la media se convirtiera en cero y la desviación en 1. Los datos diarios de los elementos meteorológicos se clasificaron en 16 grupos de unidades por SOM. Cada grupo podría considerarse como una especie de "patrón meteorológico" o "estado meteorológico". Entoces esto significó que se generaron 16 grupos o estados que se utilizarán en el Modelo Oculito de Markov, pero se muestra la necesidad de reducir la cantidad de estados debido a la complejidad de diferenciar las características de cada estado, de esta manera se redujeron a los siguientes estados: "Presión alta caliente", "Presión alta fría", "Presión baja lluviosa" y "Presión baja fría".

Por otro lado, para realizar la carga de los parámetros del Modelo Oculito de Markov se consideró como observación el número de personas que fueron llevadas a los hospitales en ambulancia en el tiempo donde el estado fue generado y tuvieron un accidente cerebrovascular. Entonces para generar las matrices de transición, matriz de emisión y matriz de probabilidades iniciales, se realizó de la siguiente forma: para el caso de la matriz de transición se contabiliza la cantidad de veces que se cambió de estado a otro, es decir se cambió de un clima a otro, con el resultado obtenido se divide sobre la cantidad de ocurrencias totales, para generar la matriz de probabilidades iniciales se contabiliza la cantidad de veces que ocurrió la observación de tener un paciente con un accidente cerebrovascular y el resultado se divide entre el total de registros, por ultimo para realizar el cálculo de la matriz de probabilidades de emisión se contabiliza la cantidad de veces que ocurra un estado cuando se desencadene la observación de tener un paciente con un accidente cerebrovascular y se divide entre el total de registros. Con la información recolectada al desarrollar la serie de tiempo se logró identificar si existe relación entre el clima y los accidentes cerebrovascular, además que al realizar múltiples combinaciones de los 4 estados el cálculo del error medio cuadrático con el "p-value" mostró resultados muy alentadores.

Tabla 2
Análisis de probabilidad HMM

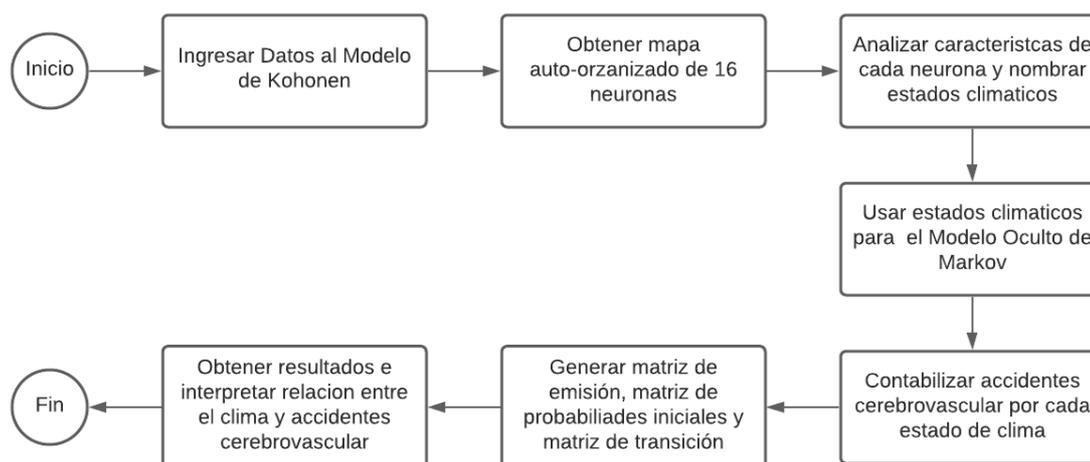
	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
HMM	2.75	2.93	2.73	2.53	3.28	2.21	2.68	3.22	2.68	3.74	2.47	2.43
p-values	0.42	0	0.79	0	0.6	0	0.06	0.02	0	0.08	0	0.39

4. METODOLOGÍA

Para lograr una integración entre el Modelo Oculito de Markov y las redes de Kohonen se optó por utilizar la investigación de Morimoto H. (2016), su investigación busco determinar si los patrones climatológicos tenían alguna relación al desencadenar accidentes cerebrovasculares. La investigación toma como punto de inicio la generación de "estados" o patrones a partir de las redes de Kohonen utilizando información climática que genero resultados como por ejemplo: "Alta presión cálida"; "Alta presión fría"; "Baja presión de lluvia"; "Baja presión fría", por otro lado, para iniciar la ejecución del Modelo Oculito de Markov se utilizaron los patrones generados en las redes de Kohonen, es decir, tomando como base la información aplicada para generar los patrones en la red de Kohonen se comienza a

cargar los parámetros necesarios para ejecutar este modelo, dicho de otra manera, cargar información en la matriz de transición, matriz de probabilidades iniciales y matriz de emisión. Los parámetros requeridos por el Modelo Oculto de Markov se extrajeron de los datos diarios de transporte de emergencia relacionados a los accidentes cerebrovasculares de la ciudad de Nagoya en Japón, las cuales fueron contabilizadas y utilizadas para determinar las matrices de transición, emisión y probabilidades iniciales. El flujo de integración entre el Modelo de Kohonen y el Modelo Oculto de Markov se muestra en la Figura 5.

Figura 5: Flujo de la implementación del Modelo Kohonen con el Modelo Oculto de Markov



El paso previo para el inicio de esta investigación fue determinar que variables que eran las más relevantes para la red de Kohonen. Morimoto H. (2016) generó una clasificación tomando una grilla de 16 neuronas con el Modelo de Kohonen y cada neurona según sus características pasó a ser considerado un estado para el Modelo Oculto de Markov. Cabe señalar que la información recuperada de los participantes de esta investigación no tenía una clasificación previa y tampoco se conocía si las variables seleccionadas eran representativas para los modelos que se aplicarían, entonces para poder continuar con la investigación se optó por aplicar un método de selección de variables propuesto por Benabdeslem y Lebbah (2007) llamado Heuristic Measure Hi-SOM. Este método se aplicó sobre las variables escogidas que son: revoluciones por minuto, aceleración, velocidad y desplazamiento vertical y horizontal tales, ya que son las más eficientes y utilizadas en los estudios mencionados previamente. Según los estudios de Amarasinghe y otros (2015) se menciona que estas variables generan grandes beneficios en el análisis de datos vehiculares, además según Meseguer y otros (2013), a pesar de que estas variables son las más importantes, debe evaluarse su importancia en el modelo trabajado con el objetivo de evaluar la relevancia de cada variable. Este último argumento es validado por Min Goo Lee, Yong Kuk Park y Kyung Kwon Jung (2011) que inició su experimento con múltiples variables, pero al realizar una validación para determinar si existía una buena relación entre sus variables seleccionadas, se llegó a la conclusión de solo utilizar las variables de revoluciones por minuto y el valor de posición del cuerpo de aceleración debido a su alta correlación. El proceso de ejecución del Modelo de Kohonen consiste en:

1. Seleccionar todas las variables relevantes para el Modelo de Kohonen
2. Calcular la importancia de la variable para el modelo
3. Eliminar la variable con el menor valor calculado del paso anterior
4. Evaluar el error topológico y error cuadrático
5. Repetir desde el punto número dos hasta que el error topológico y error cuadrático sea cercano a cero

Cuando se realizó la selección de las variables, se determinó que las más significativas fueron: aceleración, velocidad y desplazamiento horizontal y vertical. Con la selección de variables finalizada, el siguiente paso fue validar la cantidad total de grupos adecuados para generar los patrones de conducción, para esto se aplicó una validación con el método de separación de homogeneidad, donde la homogeneidad se refiere a la similitud entre nodos dentro de un mismo clúster y la separación se refiere a la disimilitud entre patrones de diferentes clúster, basándose en los

centroides existentes en cada clúster, de esta forma al aplicar este método se obtuvieron los siguientes resultados :

Tabla 3
Separación de Homogeneidad

	K=2	K=3
Homogeneidad	5,2855	4,0531
Separación	9,1864	8,8659

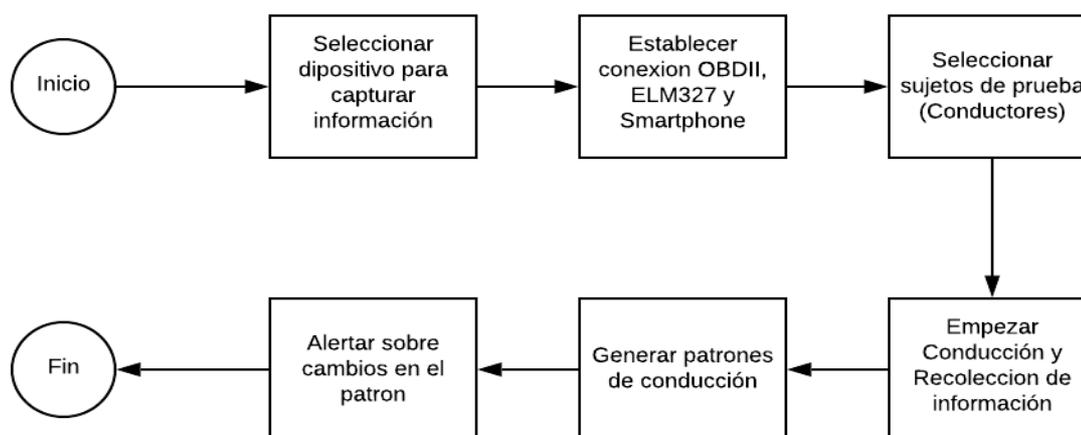
Como se aprecia en la tabla 3, al aplicar este método de separación de homogeneidad se observa que la mejor elección para agrupar la información de los participantes fue de 3 clúster, porque la variable de homogeneidad cumple con el concepto de minimizar las distancias dentro del grupo, incluso mucho más que al contar solo con dos grupos, para la selección de la cantidad de grupos nos basamos en el total de grupos presentados por las investigaciones previamente mostradas y de esta manera representar las pruebas más significativas. Por otro lado, la variable de separación logró maximizar las distancias de los grupos existentes. Por último, se seleccionó la variable de aceleración como dato de observación del Modelo Oculto de Markov, para esta elección se utilizó el mismo acercamiento de Morimoto H. (2016), es decir seleccionar una variable que se encuentra relacionada a la generación de patrones y que permite determinar a través de conteo las matrices de transición, emisión y probabilidades iniciales, con esta perspectiva se puede aplicar el Modelo de Kohonen como un visualizador y clasificador de patrones de conducción y utilizar los patrones generados con el Modelo Oculto de Markov para evaluar si mantener los patrones de conducción puede generar un posible accidente automovilístico.

5. EXPERIMENTACIÓN

Esta investigación fue experimental debido a que se utilizaron sujetos de prueba para evaluar si el uso de modelos matemáticos probabilísticos y clasificatorios lograban mejorar los estilos de conducción. Se utilizó un conjunto de vehículos con conector OBDII (On Board Display Two) y se definió un área de conducción de 15 km altamente transitada entre las 07:00 am y 08:00 am, para el inicio de este experimento se seleccionó un dispositivo de comunicación OBDII que cumpliera con los estándares actuales de los automóviles, por otro lado, se utilizaron Smartphone con el sistema operativo Android, que integra los sensores de acelerómetro y giroscopio que recolectaran información en un periodo de 27 días, donde al terminar este plazo se aplicó un proceso de aprendizaje sobre los modelos seleccionados, con el objetivo de delimitar los patrones de conducción de cada conductor y alertar cuando estos difieran de lo habitual.

Las etapas que contempló este experimento fueron las siguientes:

Figura 6: *Flujo de la implementación*



- A. Elementos para la recolección de información: Se conectó un dispositivo OBDII - ELM327 al automóvil y se estableció la comunicación entre el dispositivo ELM327 y el Smartphone a través de una conexión Bluetooth.
- B. Selección de sujetos de prueba: Para seleccionar a los sujetos de prueba las investigaciones de Crizzle y otros

Tabla 4
Matriz de Emisión

Estados/Variables	Aceleración	Desaceleración
No Accidente	0,29	0,71
Previo Accidente	0,83	0,17

Tabla 5
Matriz de Transición

Variables	No Accidente	Previo Accidente
No Accidente	0,97	0,03
Previo Accidente	0,91	0,09

Tabla 6
Matriz de Probabilidades iniciales

	Probabilidad
No Accidente	0,8
Previo Accidente	0,2

Las tablas en la parte superior muestran los estados de “no accidente” y “previo accidente”, los cuales se definieron por un cambio de aceleración negativa entre 3.4 m/s² – 4.5 m/s² propuesto por The American Association of State Highway and Transportation Officials, además se muestra la variable de “Aceleración” que fue seleccionada como variable de observación, para lograr poblar estas tablas en la parte superior se utilizó la información aplicada sobre el Modelo de Kohonen, es decir, utilizar información con las clasificaciones de “previo accidente” y “no accidente” con el objetivo de predecir con una serie de tiempo si el conductor desencadenaría el estado de “previo accidente” utilizando el Modelo Oculto de Markov, el proceso comienza: Utilizando registros previamente clasificados con el estado de “no accidente” o “previo accidente”, es decir los registros que presentan un cambio de aceleración negativa entre 3.4 m/s² – 4.5 m/s² tomando como punto de inicio la información registro anterior. Para cargar la matriz de emisión: se contabiliza la cantidad de veces que se muestran el estado de “no accidente” mientras se está acelerando y desacelerando sobre cada resultado se aplica una división sobre el total de registros, se aplica la misma lógica para el estado de “previo accidente”, en el caso de la matriz de transición para cargar información se contabiliza los cambios de estado, es decir, contabilizar la cantidad de veces que un registros pasa de “previo accidente” al estado de “no accidente” o viceversa y aplicar una división sobre el total de registros, por último, para cargar la matriz de probabilidades iniciales se contabiliza la cantidad de registros que se tiene con el estado de “previo accidente” y “no accidente” luego se aplica una división sobre el total de registros. Entonces la información mostrada en las matrices de transición, emisión y probabilidades iniciales se analiza de la siguiente forma:

- Tabla 4: Representa la probabilidad de que se emita un estado, es el resultado de un conteo de variaciones de la aceleración. Apreciamos que sobre el total de información sin procesar, mientras el estado de “no accidente” estuviera activo el 71% de las veces, el grupo mantenía una desaceleración, por el contrario, mientras el estado de “previo accidente” estuviera activo la desaceleración era del 17%, lo que se puede traducir en que existe una mayor aceleración mientras se encuentre en el estado “previo accidente”, esta información nos indica que existe una relación inversa para la variable de aceleración en los datos procesados para este grupo de conductores.
- Tabla 5: Es resultado de contabilizar la cantidad de veces que el estado de “no accidente” y “previo accidente” se desencadenaba según el cambio de aceleración propuesto por The American Association of State Highway and Transportation Officials, por lo tanto el 97% de las veces este grupo de conductores se mantenía en el estado de “no accidente” si el estado anterior era de “no accidente”, de igual forma existe una probabilidad del 91% de que si el estado actual es “previo accidente” pase a un estado de “no accidente”, esto indica que este grupo de conductores tiene una tendencia mucho mayor a estar en estado de “no accidente”.
- Tabla 6: Las probabilidades iniciales de este grupo de conductores indican que pueden comenzar una conducción en estado de “no accidente” el 80% de las veces, lo que reafirma lo declarado con anterioridad.

Finalmente, la salida de información generada por la serie de tiempo de 10 segundos con un conjunto de observaciones generó lo siguiente:

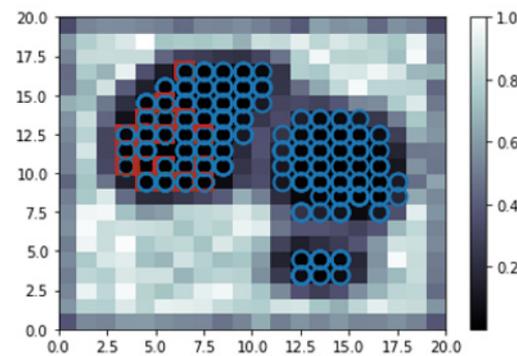
Tabla 7
Resultados en serie de 10 segundos

Tiempo/Grupos	Grupo 1	Grupo 2	Grupo 3
1s	No accidente	No accidente	No accidente
2s	No accidente	No accidente	No accidente
3s	No accidente	No accidente	Previo Accidente
4s	No accidente	No accidente	Previo Accidente
5s	No accidente	No accidente	Previo Accidente
6s	No accidente	No accidente	No accidente
7s	No accidente	No accidente	No accidente
8s	No accidente	Previo Accidente	No accidente
9s	No accidente	No accidente	No accidente
10s	No accidente	No accidente	No accidente

Los resultados de la Tabla 6 indican que los conductores del grupo 1 no ingresa en el estado de “previo accidente”, por lo que no se generó una alerta que indique una variación en el Patrón habitual de conducción que podría desencadenar un accidente automovilístico.

- El segundo grupo de 4 participantes con un total de 388,800 registros, como en el caso anterior contempla conductores con un estado en mayor medida de “no accidente” como se muestra en el siguiente mapa de Kohonen.

Figura 8: *Mapa de Kohonen Grupo 2*



A pesar de que existe una visualización mayor de recuadros rojos superpuestos que en el caso anterior, aun no es suficiente para poder clasificar este grupo de conductores como propensos a mantenerse en un estado de “previo accidente”, por lo tanto, analizaremos las siguientes tablas:

Tabla 8
Matriz de Emisión

Estados/Variabes	Aceleración	Desaceleración
No Accidente	0,32	0,68
Previo Accidente	0,81	0,19

Tabla 9
Matriz de Transición

Variabes	No Accidente	Previo Accidente
No Accidente	0,96	0,04
Previo Accidente	0,93	0,07

Tabla 10
Matriz de Probabilidades iniciales

	Probabilidad
No Accidente	0,8
Previo Accidente	0,2

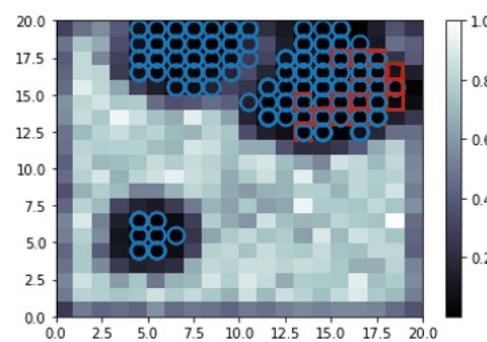
A diferencia del caso anterior, la información mostrada en las matrices de transición, emisión y probabilidades iniciales se analiza de la siguiente forma:

- Tabla 8: Representa la probabilidad de que se emita un estado, es el resultado del conteo de variaciones de la aceleración. Apreciamos que sobre el total de información sin procesar, mientras el estado de “no accidente” estaba activo el 68% de las veces el grupo se encontraba desacelerando, por el contrario, mientras el estado de “previo accidente” estaba activo, el 19% del tiempo se encontraban desacelerando lo que se puede traducir en que existe una mayor aceleración mientras se encuentre en el estado “previo accidente”, esta información nos indica que existe una relación inversa entre la aceleración y desaceleración con los datos procesados para este grupo de conductores.
- Tabla 9: Aquí se muestra que el 96% de las veces este grupo de conductores se mantenía en el estado de “no accidente” si el estado anterior es de “no accidente”, de igual forma existe una probabilidad del 93% de que si el estado actual es “previo accidente” pase a un estado de “no accidente”, esto indica que este grupo de conductores tienen una tendencia mucho mayor a estar en estado de “no accidente”.
- Tabla 10: Esta tabla indica que este grupo de conductores pueden comenzar una conducción en estado de “no accidente” el 80% de las veces, lo que reafirma lo declarado en el punto anterior.

Desde este punto de vista, la salida de información generada por la serie de tiempo de 10 segundos del grupo 2 en la tabla 6, a diferencia del caso anterior si desencadenaron la alerta de variación de Patrón de conducción debido a que se registró un estado de “Previo Accidente, por lo tanto, se puede decir que este grupo de conductores pueden tener tendencia a tener un posible accidente automovilístico.

3. El tercer grupo de 3 participantes con 291600 registros, contempla el grupo de conductores con mayor tendencia a estar en el estado de previo accidente, como se muestra en el siguiente mapa de Kohonen.

Figura 9: Mapa de Kohonen Grupo2



La cantidad de los recuadros rojos superpuestos es mucho mayor que en las muestras anteriores, lo que indica que los conductores de este grupo tienen una tendencia mayor a estar en el estado de “previo accidente”. Las siguientes tablas nos permiten una visualización de esta afirmación:

Tabla 11
Matriz de Emisión

Estados/Variables	Aceleración	Desaceleración
No Accidente	0,29	0,71
Previo Accidente	0,94	0,06

Tabla 12
Matriz de Transición

Variablen	No Accidente	Previo Accidente
No Accidente	0,97	0,03
Previo Accidente	0,91	0,09

Tabla 13
Matriz de Probabilidades iniciales

	Probabilidad
No Accidente	0,7
Previo Accidente	0,3

Según el mapa de Kohonen y la información mostrada en las tablas, estos conductores pueden mantenerse fuera de un accidente automovilístico, pero la probabilidad de mantenerse en el estado de “previo accidente” es mayor que los casos anteriores, además según sus probabilidades iniciales el conductor tiene tendencia a iniciar en un estado de “no accidente” y esto se ve reflejado también directamente en la información sin procesar donde los umbrales de “previo accidente” fueron alcanzados con mas frecuencia.

- Tabla 11: Representa la probabilidad de que se emita un estado, es el resultado de un conteo de variaciones de la aceleración. Apreciamos que sobre el total de información sin procesar, mientras el estado de “no accidente” estuviera activo el 71% de las veces el grupo mantenía una desaceleración, por el contrario, mientras el estado de “previo accidente” estuviera activo, el 6% del tiempo se encontraban desacelerando, lo que se puede traducir en que existe una mayor aceleración mientras se encuentre en el estado “previo accidente”, esto es mayor que en los casos anteriores, esta información nos indica que existe una relación inversa para la variable de aceleración en los datos procesados para este grupo de conductores.
- Tabla 12: Muestra que el 97% de las veces este grupo de conductores se mantendría en el estado de “no accidente” si el estado anterior es de “no accidente”, de igual forma existe una probabilidad del 91% de que si el estado actual es “previo accidente” pase a un estado de “no accidente”, esto indica que en gran medida este grupo de conductores tiene una tendencia a transitar al estado de “no accidente”, pero a diferencia de los grupos anteriores, se ve impactado en mayor medida por las probabilidades de la matriz de emisión que indican una tendencia mayor a transitar al estado de “previo accidente”.

Tabla 13: la matriz de probabilidades iniciales de este grupo de conductores indica que pueden comenzar una conducción en estado de “no accidente” del 70% de las veces, lo que es un indicador más claro de que el estado de “previo accidente” puede desencadenarse ya que en los otros casos la probabilidad de iniciar en el estado de “no accidente” era mayor.

De esta forma, la salida de información generada por la serie de tiempo con un conjunto de observaciones generó los resultados mostrados en la Tabla 6, lo que indica que este grupo de conductores tiene una tendencia aun mayor que el grupo anterior a entrar al estado de “previo accidente” lo cual tiene mayor probabilidad de desencadenar un accidente automovilístico.

Finalmente, con los datos recolectados y los resultados analizados de los tres grupos evaluados durante esta sección del documento, entendemos que al conocer los patrones de conducción de cada grupo de conductores podemos predecir si desencadenará un posible accidente automovilístico. Entonces el aplicativo desarrollado accionará la alerta auditiva para indicar variaciones negativas en el patrón de conducción y evitar los accidentes automovilísticos.

Desde otra perspectiva, las investigaciones de Min Goo Lee, Yong Kuk Park y Kyung Kwon Jung (2011) y Meseguer, Calafate, Cano y Manzoni (2013), toma como referencia la clasificación de patrones de conducción con el objetivo de mejorar consumos de combustible y clasificar la ruta de conducción respectivamente, mientras que esta investigación no solo genera patrones de conducción para conocer los estilos de cada usuario, sino que busca predecir si el estilo de conducción de cada usuario puede desencadenar un posible accidente automovilístico, este objetivo es similar a la investigación de Amarasinghe y otros (2015) que busca mediante su Modelo de Markov predecir si un conductor se involucraría en un accidente automovilístico, pero con la diferencia de que no cuentan con patrones de conducción clasificados que expliquen de manera visual la información con la que se realizó la evaluación del Modelo

de Markov, sino que utiliza un cálculo promedio para determinar estados y calcular las matrices necesarias para el Modelo de Markov y con esta información procesada alertar al conductor cuando sobrepase el umbral establecido por The American Association of State Highway and Transportation Officials.

7. CONCLUSIONES

Mediante la presente investigación y con la información recopilada a través de pruebas, validaciones y correcciones de las metodologías antes mostradas, es decir, el modelo de Red Neuronal de Kohonen y el Modelo Oculto de Markov, se desarrolló un sistema que permitió generar alertas cuando el Patrón de conducción detectaba en la serie de tiempo un estado de “previo accidente” con el objetivo de mejorar los estilos de conducción, utilizando tres variables: aceleración, velocidad y valores de desplazamiento horizontal y vertical.

La primera conclusión fue que nuestro modelo se ve beneficiado al contar solo con estas tres variables, porque al realizar la validación de características se aprecia que estas impactan significativamente al modelo de manera positiva al tener alta relación entre ellas, además el error calculado decrece de manera considerable.

La segunda conclusión está relacionada directamente al Modelo Oculto de Markov, ya que existe una relación inversa entre el estado de “no accidente” y el “previo accidente”, donde el primero se relaciona principalmente con mayores desaceleraciones que aceleraciones, caso contrario ocurre en el segundo estado donde se tiene un índice mayor de aceleraciones que desaceleraciones, por ejemplo el primer grupo mostró que mientras se encuentre en estado de “previo accidente” se encontraban 83% de tiempo acelerando y solo 17% desacelerando, por el contrario cuando estaban en el estado de “no accidente” solo aceleraban 29% del tiempo y desaceleraban el 71% del tiempo.

Por último, la tercera conclusión fue que el estado de “previo accidente” se desató con más frecuencia en los conductores con mayor tendencia a acelerar donde los datos de las matrices de transición, emisión y probabilidades iniciales indicaban una tendencia menor de pasar desde un estado de “previo accidente” a “no accidente”.

8. TRABAJOS FUTUROS Y RECOMENDACIONES

De acuerdo a Amarasinghe (2015); Crizzle, Myers, Almeida, y Vrkljan (2011); Meseguer (2013); Morimoto, (2016); R.Dange (2017), para obtener datos la mejor alternativa es tener un espacio controlado donde las condiciones de recolección de información sean muy similares entre los conductores. Por otro lado, con la metodología actual es posible orientar el trabajo a campos no solo de conducción urbana sino también al sector empresarial, como por ejemplo la conducción de camiones de transporte de carga o pasajeros, con el objetivo de evitar que se generen accidentes durante las jornadas laborales.

Respecto a las metodologías de la red neuronal de Kohonen y El Modelo Oculto de Markov, se recomienda tener información que permita una interacción directa entre estos dos modelos, como se mostró en esta investigación a través de la variable de aceleración que sirvió como base para construir patrones de conducción en la red de Kohonen y determinar los estados de Markov con las observaciones necesarias para generar los datos requeridos para el modelo.

REFERENCIAS

- Alvear, O., Calafate, C. T., Cano, J.-C., y Manzoni, P. (2015). Validation of a vehicle emulation platform supporting OBD-II communications. *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, 880–885. <https://doi.org/10.1109/CCNC.2015.7158092>
- Amarasinghe, M., Kottegoda, S., Arachchi, A. L., Muramudalige, S., Bandara, H. M. N. D., y Azeez, A. (2015). Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics. *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, 243–249. <https://doi.org/10.1109/ICTER.2015.7377695>
- Benabdeslem, K., y Lebbah, M. (2007). Feature Selection for Self-Organizing Map. *2007 29th International Conference on Information Technology Interfaces*, 45–50. <https://doi.org/10.1109/ITI.2007.4283742>
- Crizzle, A. M., Myers, A. M., Almeida, Q. J., y Vrkljan, B. H. (2011). *Using in-vehicle devices to examine exposure and patterns in drivers with Parkinson's disease compared to an age-matched control group* (Proc 6th i).
- EUROPA - ASSET Advanced Safety and Driver Support in Essential Road Transport | TRIMIS - European Commission. (s/f). Recuperado el 16 de enero de 2021, de <https://trimis.ec.europa.eu/project/asset-advanced-safety-and-driver-support-essential-road-transport>
- EUROPA - Close Communications for Cooperation between Cybercars | TRIMIS - European Commission. (s/f). Recuperado el 16 de enero de 2021, de <https://trimis.ec.europa.eu/project/close-communications-cooperation-between-cybercars>
- Final Report Summary - SAFETRIP (Satellite application for emergency handling, traffic alerts, road safety and

incident prevention) | Report Summary | SAFETRIP | FP7 | CORDIS | European Commission. (s/f). Recuperado el 16 de enero de 2021, de <https://cordis.europa.eu/project/id/233976/reporting/pl>

- INEI. (2018). *PERU Instituto Nacional de Estadística e Informática INEI*.
<https://www.inei.gob.pe/estadisticas/indice-tematico/traffic-accidents/>
- Kohonen, T. (2001). *Self-Organizing Maps (Vol. 30)*. Berlin, Heidelberg: Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-642-56927-2>
- Le Bastard, C., Guilbert, D., Delepouille, A., Boubezoul, A., Ieng, S.-S., y Wang, Y. (2011). Vehicule identification from inductive loops application : Travel time estimation for a mixed population of cars and trucks. *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 507–512.
<https://doi.org/10.1109/ITSC.2011.6082802>
- Meseguer, J. E., Calafate, C. T., Cano, J. C., y Manzoni, P. (2013). DrivingStyles: A smartphone application to assess driver behavior. *2013 IEEE Symposium on Computers and Communications (ISCC)*, 000535–000540.
<https://doi.org/10.1109/ISCC.2013.6755001>
- Min Goo Lee, Yong Kuk Park, Kyung Kwon Jung, J. J. Y. (2011). Estimation of Fuel Consumption using In-Vehicle Parameters. *Computers & Industrial Engineering*, 59(1), 157–165.
<https://www.semanticscholar.org/paper/Estimation-of-Fuel-Consumption-using-In-Vehicle-Pa-Lee-Park/452a214ed37c7ff2a942b8922f3fcfe505eaa933>
- Morimoto, H. (2016). *Hidden Markov Models and Self-Organizing Maps Applied to Stroke Incidence*. March, 158–168.
- R. Durbin, S. Eddy, A. Krogh, G. M. (1998). *Biological sequence analysis probabilistic models of proteins and nucleic acids* (C. U. PRESS (ed.)). CAMBRIDGE UNIVERSITY PRESS.
- Dange, G. R., Paranthaman, P. K., Bellotti, F., Samaritani, M., Berta, R., & De gloria, A. (2017). Assessment of driver behavior based on machine learning approaches in a social gaming scenario. *Lecture Notes in Electrical Engineering*, 409, 205–218. https://doi.org/10.1007/978-3-319-47913-2_24
- Wonang Jang, Daeseong Jong, y Dohoon Lee. (2016). Methodology to improve driving habits by optimizing the in-vehicle data extracted from OBDII using genetic algorithm. *2016 International Conference on Big Data and Smart Computing (BigComp)*, 313–316. <https://doi.org/10.1109/BIGCOMP.2016.7425936>
- Zaldivar, J., Calafate, C. T., Cano, J. C., y Manzoni, P. (2011). Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. *2011 IEEE 36th Conference on Local Computer Networks*, 813–819. <https://doi.org/10.1109/LCN.2011.6115556>