



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA
INFORMACIÓN



ANÁLISIS DE DATOS DE SENSORES PARA CLASIFICAR LA CONDUCCIÓN VEHICULAR PRUDENTE / IMPRUDENTE

Trabajo recepcional bajo la modalidad de Tesis
Que para obtener el grado de
Licenciado en Sistemas Computacionales

Presenta:

Daniel Rodríguez López

Directores:

Dr. Pablo Pancardo García

Dr. Gerardo Arceo Moheno

Cuerpos Académicos:

Sistemas Distribuidos
Innovación en Organismos Sociales

Líneas de Generación y Aplicación del Conocimiento:

Sistemas Ubicuos
Las tecnologías de información (TI) en la evaluación de
aspectos estratégicos de las organizaciones para formulación
de propuestas innovadoras



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA
INFORMACIÓN



ANÁLISIS DE DATOS DE SENSORES PARA
CLASIFICAR LA CONDUCCIÓN VEHICULAR PRUDENTE
/ IMPRUDENTE

Trabajo recepcional bajo la modalidad de Tesis

Que para obtener el grado de

Licenciado en Sistemas Computacionales

Presenta:

Daniel Rodríguez López

Directores:

Dr. Pablo Pancardo García

Dr. Gerardo Arceo Moheno

Jurado Revisor:

Dr. José Adán Hernández Nolasco

M.S.I. María Evilia Magaña

Dra. María Alejandrina Almeida Aguilar

Dra. Martha Patricia Silva Payró

Dr. Ricardo Gómez Crespo

Cuerpos Académicos:

Sistemas Distribuidos

Innovación en Organismos Sociales



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



Oficio No.029/2021/DACYTI/D
5 de enero 2021

Dr. Pablo Pancardo García
Profesor investigador
Presente

De acuerdo al artículo 72 del Reglamento de Titulación, de la Universidad Juárez Autónoma de Tabasco, me permito designarle Director del Trabajo Recepcional titulado **"Clasificación de conducción vehicular prudente/imprudente"**, con folio 202001-SD-02-2 a realizarse por los CC. **Daniel Rodríguez López y José Adrián Ramos García**, para obtener el título de la Licenciatura en Sistemas Computacionales, bajo la modalidad de Tesis.

No omito manifestarle que con fundamento en el artículo 78 del reglamento antes mencionado, a partir de la fecha expedida tiene como plazo máximo un año para concluir la tesis de lo contrario, el estudiante deberá optar por otra modalidad de titulación.

Sin otro particular, agradezco de antemano su colaboración y apoyo, enviándole un cordial saludo.

Atentamente

MTE. Oscar Alberto González González

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

Director

C.c.p. Coordinador de Estudios Terminales
Tesis
Archivo
Consecutivo

MTE. *OAGG/MA* JAGR



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



Oficio No.030/2021/DACYTI/D
5 de enero 2021

Dr. Gerardo Arceo Moheno
Profesor investigador
Presente

De acuerdo al artículo 72 del Reglamento de Titulación, de la Universidad Juárez Autónoma de Tabasco, me permito designarle Director del Trabajo Recepcional titulado **"Clasificación de conducción vehicular prudente/imprudente"**, con folio 202001-SD-02-2 a realizarse por los CC. **Daniel Rodríguez López y José Adrián Ramos García**, para obtener el título de la Licenciatura en Sistemas Computacionales, bajo la modalidad de Tesis.

No omito manifestarle que con fundamento en el artículo 78 del reglamento antes mencionado, a partir de la fecha expedida tiene como plazo máximo un año para concluir la tesis de lo contrario, el estudiante deberá optar por otra modalidad de titulación.

Sin otro particular, agradezco de antemano su colaboración y apoyo, enviándole un cordial saludo.

Atentamente

MTE. Oscar Alberto González González

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

Director

C.c.p. Coordinador de Estudios Terminales
Tesis
Archivo
Consecutivo

MTE. *OAGG/MA*JAGR

F6: Liberación de Director de Tesis

2021

Cunduacán, Tabasco., a 22 de junio de 2021.

Asunto: Liberación de dirección de tesis.

M.T.E. Oscar Alberto González González
Director de la DACYTI
Presente

Por medio de la presente nos permitimos comunicarle que después de haber realizado las asesorías correspondientes al Trabajo Recepcional: "Clasificación de conducción vehicular prudente / imprudente", con folio: 202001-SD-02-2; elaborado por el C. Daniel Rodríguez López, de la Licenciatura en Sistemas Computacionales, consideramos que lo ha concluido satisfactoriamente, por lo que puede continuar con los trámites de titulación.

Sin otro particular, aprovechamos la ocasión para enviarle un cordial saludo.

Atentamente



Dr. Pablo Pancardo García



Dr. Gerardo Arceo Moheno

c.c.p. M.A. José Alberto Gutiérrez Rosique. - Coordinador de Estudios Terminales
Integrantes de la Comisión Revisora
Estudiantes



Cunduacán, Tabasco, a 23 de junio de 2021.

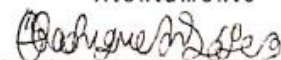
Asunto: Solicitud de Jurado

M.T.E. Oscar Alberto González González
Director de la DACYTI
Presente

Por este medio me permito informarle que el trabajo recepcional: "Clasificación de conducción vehicular prudente / imprudente", *bajo la modalidad de tesis*, ha sido liberado por mis directores: Dr. Pablo Pancardo García y Dr. Gerardo Arceo Moheno, por lo que en atención a ello me dirijo a usted con la finalidad de solicitarle tenga a bien nombrar al jurado para que evalúe el citado trabajo.

Sin otro particular, aprovecho la ocasión para enviarle un cordial saludo.

Atentamente



Daniel Rodríguez López

Licenciatura:	Licenciado en Sistemas computacionales
Matrícula:	162H8002
Domicilio:	Fraccionamiento Los Naranjos
Localidad:	Nacajuca, Tabasco
Teléfono:	993 146 6164
E-mail:	danro198@outlook.com

c.c.p. M.A. José Alberto Gutiérrez Rosique, - Coordinador de Estudios Terminales.
Estudiantes.





UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO
"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



"2022, Año de Ricardo Flores Magón"

Oficio No.598/2022/DACYTI/D
27 de mayo 2022

Asunto: Autorización cambio de nombre a trabajo Recepcional

C. Daniel Rodríguez López
Presente

En virtud a su solicitud de cambio de nombre del trabajo Recepcional denominado "Clasificación de conducción vehicular prudente/imprudente", con folio 202001-SD-02-2 y al visto bueno de sus directores, así como, el común acuerdo de su Comisión Revisora, informo a usted que se **Autoriza el cambio de nombre del trabajo Recepcional**, por el nombre de: **"Análisis de datos de sensores para clasificar la conducción vehicular prudente/imprudente"** bajo la modalidad de titulación Tesis.

Sin otro particular aprovecho la oportunidad para saludarle.

Atentamente

MTE. Oscar Alberto González González
Director

C.c.p. MA José Alberto Gutiérrez Rosique. -Coordinador de Estudios Terminales
Dr. Pablo Pancardo García. -Director Trabajo Recepcional
Dr. Gerardo Arceo Moheno. -Director Trabajo Recepcional
Integrantes de la Comisión Revisora
Archivo
Consecutivo

MTE. *OAGG/MA*JAGR

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.
Cunduacán, Tabasco, México.
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx

Cunduacán, Tabasco., a 14 de noviembre de 2022.

Asunto: Respuesta de jurado

M.T.E. Oscar Alberto González González
Director de la DACYTI
Presente

En atención a los oficios girados por usted, en los que se nos designa como parte del jurado para efectuar la revisión del trabajo recepcional titulado "*Análisis de datos de sensores para clasificar la conducción vehicular prudente / imprudente*", bajo la modalidad de tesis, realizado por el C. Daniel Rodríguez López, alumno de la Licenciatura en Sistemas computacionales, nos permitimos informarle que en virtud de que el estudiante ha atendido las observaciones realizadas, otorgamos nuestra aprobación para que continúe los trámites correspondientes a su titulación.


Sin otro particular, aprovechamos la ocasión para enviarle un cordial saludo.

Atentamente integrantes del Jurado



Dr. José Adán Hernández Nolasco

Dr. Ricardo Gómez Crespo

Dra. Martha Patricia Silva Payró

M.S.L. María Evilia Magaña

Dra. María Alejandrina Almeida Aguilar

C.c.p. M.A. José Alberto Gutiérrez Rosique. - Coordinador de Estudios Terminales
Estudiantes



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



"2022, Año de Ricardo Flores Magón"

Oficio No. 1473/2022/DACYTI/D
14 de noviembre de 2022

Asunto: Autorización Impresión trabajo recepcional

C. Daniel Rodríguez López

En virtud de cumple satisfactoriamente los requisitos establecidos en el Reglamento de Titulación vigente en la Universidad Juárez Autónoma de Tabasco, informo a usted que se autoriza la Impresión del trabajo Recepcional denominado **"Análisis de datos de sensores para clasificar la conducción vehicular prudente / imprudente"**, para presentar Examen Profesional y obtener el título de **Licenciado en Sistemas Computacionales**, bajo la modalidad de **Tesis**.

Sin otro particular aprovecho la oportunidad para saludarle.

Atentamente

MTE. Óscar Alberto González González
Director



C.c.p. Asesor del Trabajo Recepcional: Dr. Pablo Pancardo García
Asesor del Trabajo Recepcional: Dr. Gerardo Arceo Moheno
Archivo
Consecutivo

MTE. *OAGG/MA*JAGR

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.
Cunduacán, Tabasco, México.
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx

Carta de Autorización

A quien corresponda:

El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada, **“Análisis de datos de sensores para clasificar la conducción vehicular prudente / imprudente”**, de la cual soy autor y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa más no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en éste documento.

Se firma la presente autorización en la ciudad de Villahermosa, Tabasco a los 23 días del mes de noviembre del año 2022.

Autorizo



Daniel Rodríguez López

Agradecimientos

Este trabajo fue posible gracias a mis directores de tesis el Dr. Pablo Pancardo García y el Dr. Gerardo Arceo Moheno, por haberme guiado durante el proceso de investigación del presente proyecto.

De igual manera agradezco a los docentes de la División Académica de Ciencias y Tecnologías de la Información de la Universidad Juárez Autónoma de Tabasco, por haber compartido sus conocimientos a lo largo de mi formación profesional.

Agradezco también a mi tutor el L.I.A. Ericsson Saldívar Correa Robles por orientarme en la planeación de mi trayectoria académica.

Finalmente agradezco a mi compañero José Adrián Ramos García por su valiosa colaboración en esta investigación.

Dedicatorias

La presente tesis está dedicada principalmente a mis padres, por su amor y apoyo incondicional, siempre dispuestos a escucharme y estar conmigo en momentos difíciles e inspirarme a dar lo mejor de mí.

A mis hermanas, por todos los momentos llenos de alegría compartidos a lo largo de nuestras vidas que nos fueron uniendo cada vez más con el paso de los años.

A mis amigos, quienes tuve la dicha de conocer y convivir junto a ellos gratas experiencias en diferentes etapas de mi vida.

Resumen

El propósito de la presente investigación fue determinar cuando la conducción de un vehículo está siendo prudente o imprudente. En primer lugar, se analizó un dataset creado previamente en otra investigación, en la cual se empleó un vehículo de juguete para realizar las pruebas y que considera tres tipos diferentes de caminos. El dataset se dividió en dos partes, un 70% de los datos para realizar el entrenamiento y el porcentaje restante se empleó para las pruebas y calcular algunas funciones estadísticas a partir de los datos.

Las funciones estadísticas que se calcularon son la media, moda, mediana, desviación estándar, varianza, valor cuadrático medio y energía. A partir de lo anterior, se realizaron una serie de gráficas para comparar visualmente todos los resultados estadísticos y se eligió aquellas características estadísticas que presentaban mayor variabilidad y, por tanto, permitirían clasificar cada tipo de camino. De las estadísticas iniciales se seleccionaron tres de ellas. Estas fueron las que aportaban mayores elementos para diferenciar entre los distintos tipos de caminos y velocidades.

Posteriormente, se realizó una comparación entre los valores estadísticos obtenidos en la fase de prueba respecto a los obtenidos durante el entrenamiento. Se pudo confirmar que, efectivamente, los valores son muy cercanos, por lo que se comprueba que el análisis mediante valores estadísticos de los datos es eficiente para poder obtener la clasificación de los tipos de conducción.

Finalmente, es importante mencionar que se consideró conducción prudente cuando un conductor al circular sobre un camino con baches y topes, lo hace de forma lenta o moderada. Mientras que una conducción imprudente fue calificada así cuando el conductor circula sobre un camino con baches y topes a una velocidad rápida. En lo que respecta a una superficie plana y sin obstáculos, se consideró prudente siempre y cuando cumpla con las normas de conducción con seguridad.

Introducción

Este trabajo provee un método de clasificación de conducción vehicular que pretende proporcionar una utilidad a la gran cantidad de información que se obtiene de los vehículos en su conjunto. Se busca analizar un dataset que contiene datos de la aceleración de un vehículo de juguete en diferentes superficies y velocidades. Dicho dataset fue construido empleando mediciones de un acelerómetro.

La clasificación de la conducción vehicular está sujeta al hábito personal de comportamiento, ya que al manejar nos encontramos con distintos obstáculos o problemas. Por otro lado, se emplearon funciones estadísticas para conseguir una clasificación a los datos generados por los vehículos.

El fin fue analizar el comportamiento del conductor, de esta manera se logra crear un modelo de análisis de datos para que las empresas relacionadas con los vehículos, por ejemplo, empresas de transporte público, transporte de mercancías, etcétera; puedan informarse acerca del uso que los conductores hacen de los vehículos. Asimismo, las empresas o particulares podrían tener un control y por consiguiente proporcionar mayor seguridad a los pasajeros, así como ahorrar costos derivados de una conducción deficiente, por ejemplo, el incremento en el consumo de diésel y/o gasolina. Consecuentemente, el desafío consistió en generar una clasificación adecuada de los datos y precisar si la conducción que se está llevando a cabo es prudente o imprudente. La investigación realizada se encuentra estructurada en cinco capítulos que se describen a continuación:

Capítulo I, Generalidades: en este capítulo se muestran los antecedentes referentes al tema, se plantea y define la problemática de investigación, además del objetivo general y los específicos. Es aquí mismo donde se justifica el hecho de realizar el trabajo de investigación y se determinan los alcances y limitaciones que lo conforman.

Capítulo II, Marco teórico: en el marco referencial se hallan las síntesis de trabajos previos que se han hecho sobre el tema de la investigación o de temas semejantes, los cuales sirven de apoyo para este trabajo recepcional; mientras que el marco conceptual se

conforma por las definiciones de los conceptos más relevantes con el propósito de fortalecer la comprensión del tema.

Capítulo III, Aplicación de la metodología y desarrollo: se determina la muestra y se presenta la explicación de los instrumentos empleados para recabar datos durante el desarrollo. Se muestran todas las funciones estadísticas utilizadas y cuáles de ellas se determinaron como más relevantes.

Capítulo IV, Pruebas y resultados: contiene el producto de la aplicación de los instrumentos presentados en el capítulo anterior mediante tablas y gráficas, además de estar complementado con el análisis e interpretación de estas. Se explica de qué manera las funciones estadísticas fueron empleadas para ayudar al análisis de la clasificación de conducción vehicular.

Capítulo V, Conclusiones y trabajos futuros: se plasman las conclusiones que resultan de la investigación y se sugieren los trabajos que pueden surgir con relación al tema.

Índice general

Capítulo 1. Generalidades	1
1.1 Antecedentes	1
1.2 Planteamiento del problema	2
1.2.1 Definición del problema	2
1.2.2 Delimitación de la investigación	3
1.2.3 Preguntas de investigación	3
1.3 Objetivos	3
1.3.1 Objetivo general	3
1.3.2 Objetivos específicos	3
1.4 Justificación	4
1.5 Metodología utilizada	5
1.5.1 Enfoque de la investigación	5
1.5.2 Fuentes de investigación	5
1.5.3 Técnicas de recolección de datos	6
1.5.4 Selección de herramientas	6
Capítulo 2. Marco teórico	8
2.1 Marco referencial	8
2.2 Marco conceptual	10
2.2.1 Conceptos y teorías relacionados con la investigación	10
2.2.2 Conceptos y teorías relacionados con el marco tecnológico	11
2.2.3 Conceptos y teorías relacionados con el método a utilizar	12
2.3 Marco tecnológico	16
2.3.1 Herramientas de Software	16
2.3.2 Herramientas de Hardware	16
2.4 Marco legal	17
Capítulo 3. Aplicación de la metodología y desarrollo	18
3.1 Valores estadísticos para los ejes del acelerómetro	18
3.2 Determinación y tabulación de características de los datos	28

3.3	Graficación de los resultados estadísticos.....	32
3.4	Características más relevantes.....	34
Capítulo 4.	Pruebas y resultados.....	36
4.1	Interpretación de los resultados.....	36
4.2	Resultados de clasificación de conducción prudente.	37
4.3	Resultados de clasificación de conducción imprudente.....	40
Capítulo 5.	Conclusiones y trabajos futuros	43
Bibliografía	45
Glosario	49
Anexos.....		51
Anexo A.	Dataset utilizado para el análisis de la clasificación de conducción vehicular.....	52
Anexo B.	Código fuente utilizado para el análisis de la clasificación de conducción vehicular. ...	61

Índice de ilustraciones

Ilustración 1: Estadística descriptiva (Requena, 2014a)	12
Ilustración 2: Graficación de la media	32
Ilustración 3: Graficación de la desviación estándar	33
Ilustración 4: Graficación de la varianza	33
Ilustración 5: Camino BVB	37
Ilustración 6: Camino RVB	38
Ilustración 7: Camino RVM	38
Ilustración 8: Camino RVR	39
Ilustración 9: Camino TVB	39
Ilustración 10: Camino BVM	40
Ilustración 11: Camino BVR	41
Ilustración 12: Camino TVM	41
Ilustración 13: Camino TVR	42
Ilustración 14: Código fuente. Clase Data	61
Ilustración 15: Código fuente. Funciones. Mostrar tabla	61
Ilustración 16: Código fuente. Funciones. Formatear texto y contenido	62
Ilustración 17: Código fuente. Importaciones	62
Ilustración 18: Código fuente. Datos de entrenamiento	63
Ilustración 19: Código fuente. Listado de bases de datos	63
Ilustración 20: Código fuente. Selección de base de datos para pruebas	64
Ilustración 21: Código fuente. Creación de datos de prueba	64
Ilustración 22: Código fuente. Creación de datos de prueba (continuación)	65
Ilustración 23: Código fuente. Clasificación. Comparación de valores	65
Ilustración 24: Código fuente. Clasificación. Comparación de valores (continuación)	66
Ilustración 25: Código fuente. Clasificación. Determinación de resultados	66
Ilustración 26: Código fuente. Clasificación. Muestra de resultados	67

Índice de ecuaciones

Ecuación 1: Suma de vectores.....	13
Ecuación 2: Media.....	13
Ecuación 3: Media (x).....	13
Ecuación 4: Mediana para n impar.....	13
Ecuación 5: Mediana para n par.....	14
Ecuación 6: Moda.....	14
Ecuación 7: Desviación estándar.....	14
Ecuación 8: Valor cuadrático medio	15
Ecuación 9: Valor absoluto de un número	15
Ecuación 10: Energía	16

Índice de tablas

Tabla 1: Primeros 20 registros de la tabla de camino con baches a velocidad baja.....	18
Tabla 2: Primeros 20 registros de la tabla de camino con baches a velocidad moderada.....	19
Tabla 3: Primeros 20 registros de la tabla de camino con baches a velocidad rápida	20
Tabla 4: Primeros 20 registros de la tabla de camino recto a velocidad baja	20
Tabla 5: Primeros 20 registros de la tabla de camino recto a velocidad moderada	21
Tabla 6: Primeros 20 registros de la tabla de camino recto a velocidad rápida	22
Tabla 7: Primeros 20 registros de la tabla de camino con topes a velocidad baja	22
Tabla 8: Primeros 20 registros de la tabla de camino con topes a velocidad moderada	23
Tabla 9: Primeros 20 registros de la tabla de camino con topes a velocidad rápida.....	24
Tabla 10: Media, valores máximos y mínimos de la tabla Bache velocidad baja.....	24
Tabla 11: Media, valores máximos y mínimos de la tabla Bache velocidad moderada	25
Tabla 12: Media, valores máximos y mínimos de la tabla Bache velocidad rápida	25
Tabla 13: Media, valores máximos y mínimos de la tabla Recta velocidad baja	26
Tabla 14: Media, valores máximos y mínimos de la tabla Recta velocidad moderada	26
Tabla 15: Media, valores máximos y mínimos de la tabla Recta velocidad rápida	27
Tabla 16: Media, valores máximos y mínimos de la tabla Tope velocidad baja	27
Tabla 17: Media, valores máximos y mínimos de la tabla Tope velocidad moderada	27
Tabla 18: Media, valores máximos y mínimos de la tabla Tope velocidad rápida	28
Tabla 19: Valores estadísticos para los datos de la tabla Bache velocidad baja	29
Tabla 20: Valores estadísticos para los datos de la tabla Bache velocidad moderada	29
Tabla 21: Valores estadísticos para los datos de la tabla Bache velocidad rápida.....	29
Tabla 22: Valores estadísticos para los datos de la tabla Recta velocidad baja	30
Tabla 23: Valores estadísticos para los datos de la tabla Recta velocidad moderada	30
Tabla 24: Valores estadísticos para los datos de la tabla Recta velocidad rápida.....	30
Tabla 25: Valores estadísticos para los datos de la tabla Tope velocidad baja	31
Tabla 26: Valores estadísticos para los datos de la tabla Tope velocidad moderada.....	31
Tabla 27: Valores estadísticos para los datos de la tabla Tope velocidad rápida.....	31
Tabla 28: Datos de entrenamiento.....	36

Tabla 29: Datos de prueba.....	37
Tabla 30: Muestra del 0.003% del dataset de la tabla Bache velocidad baja.....	52
Tabla 31: Muestra del 0.003% del dataset de la tabla Bache velocidad moderada.....	53
Tabla 32: Muestra del 0.003% del dataset de la tabla Bache velocidad rápida	54
Tabla 33: Muestra del 0.003% del dataset de la tabla Recta velocidad baja.....	54
Tabla 34: Muestra del 0.003% del dataset de la tabla Recta velocidad moderada.....	55
Tabla 35: Muestra del 0.003% del dataset de la tabla Recta velocidad rápida	56
Tabla 36: Muestra del 0.003% del dataset de la tabla Tope velocidad baja	57
Tabla 37: Muestra del 0.003% del dataset de la tabla Tope velocidad moderada	58
Tabla 38: Muestra del 0.003% del dataset de la tabla Tope velocidad rápida	59

Capítulo 1. Generalidades

1.1 Antecedentes

Actualmente, manejar un vehículo es una actividad muy usual para las personas. Los accidentes de tránsito son la segunda causa de muerte en México y según cifras oficiales, en 2013 fallecieron más de 36 mil personas a causa de algún tipo de accidente. Aunque el riesgo de morir aumenta en las carreteras del país por las altas velocidades que en ellas se alcanzan, la mayor parte de los accidentes se registran en las zonas urbanas y suburbanas; aproximadamente el 80 por ciento de los accidentes de tránsito se deben a una conducta equivocada de los conductores (Secretaría de Salud, 2015).

Los accidentes de tránsito son un problema que afecta a la población y son una de las principales causas de muerte y discapacidad en el mundo. Cada año mueren cerca de 1.3 millones de personas en las carreteras del mundo entero, y entre 20 y 50 millones padecen traumatismos no mortales. Los accidentes de tránsito son una de las principales causas de muerte en todos los grupos etarios, y la primera entre personas de entre 15 y 29 años (Organización Mundial de la Salud, 2017).

Esos accidentes provocados, además de poner un alto costo emocional y económico al familiar afectado, repercuten en la familia y en la comunidad pues ocasionan fracturas y/o lesiones en la persona, y en casos críticos incluso puede causar la muerte. México ocupa el séptimo lugar mundial en siniestros viales. De acuerdo con datos del Instituto Nacional de Salud Pública (INSP), nuestro país ocupa el séptimo lugar a nivel mundial y el tercero en la región de Latinoamérica en muertes por siniestros viales, con 22 decesos de jóvenes de entre 15 y 29 años al día y 24 mil decesos en promedio al año. Los siniestros viales constituyen la primera causa de muerte en jóvenes entre 15 y 29 años de edad y la quinta entre la población general (Instituto Nacional De Salud Pública, 2020).

1.2 Planteamiento del problema

1.2.1 Definición del problema

Existen diversos factores que pueden afectar la forma de conducir de las personas, ocasionando inconvenientes en el momento en que una persona se disponga a conducir un vehículo. Estos factores pueden ser ocasionados por el estado emocional y psicológico en que se encuentre el conductor, el tipo de vehículo que disponga, las condiciones en que se encuentre la vía, las señales y el tráfico vehicular, el tiempo de las personas tras el volante, la educación vial y el clima, entre otros factores.

Como conductor podemos hallar diversas afectaciones al estar detrás del volante que pueden ocasionar grandes pérdidas. Los seres humanos, de manera natural, reaccionan ante un cierto evento dependiendo fuertemente del estado psicoemocional en el que se encuentran. Un problema serio es cuando las emociones adquieren un efecto negativo en el comportamiento y afectan el proceso de conducción, tales como manejar sobre el límite máximo de velocidad, no respetar los señalamientos, etc., y esto consecuentemente, provoca cierta propensión a accidentes.

Dado que no es factible estar vigilando a cada conductor para determinar si está realizando una conducción prudente o imprudente, es necesario que se desarrollen soluciones tecnológicas que lo puedan realizar de forma automatizada. Sin embargo, el problema es clasificar la conducción empleando datos obtenidos con sensores durante el proceso de manejo del automóvil.

De acuerdo con lo expresado anteriormente, se planteó la siguiente pregunta de investigación:

¿Cómo obtener una clasificación de conducción vehicular en prudente o imprudente empleando datos de sensores de movimientos?

1.2.2 Delimitación de la investigación

1.2.2.1 Alcances

Construcción de un modelo que permita conocer la manera de manejar de un conductor clasificándolo en prudente o imprudente dependiendo de patrones de comportamiento extraídos a partir de un dataset dado.

1.2.2.2 Limitaciones

La investigación se limitó a la información proporcionada por sensores que puedan medir el comportamiento de movimientos del vehículo, el cual puede verse afectado por diferentes situaciones que pueden presentarse cuando la persona se encuentra tras el volante.

El dataset de conducción empleado ya se encuentra construido en un trabajo previo.

1.2.3 Preguntas de investigación

¿Cómo obtener una clasificación de conducción vehicular en prudente o imprudente empleando datos de sensores de movimientos?

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar un modelo basado en datos de sensores de movimientos y métodos de estadística descriptiva que sea capaz de clasificar la conducción en prudente o imprudente para la toma de decisiones hacia un sistema de transportación inteligente.

1.3.2 Objetivos específicos

- Obtener los valores estadísticos descriptivos de los datos contenidos en el dataset de conducción vehicular.

- Analizar los datos empleando técnicas de aprendizaje automático para la clasificación de la conducción en prudente e imprudente.
- Tabular y graficar los resultados estadísticos para obtener las características más relevantes.
- Realizar pruebas con las características seleccionadas para construir y afinar un modelo de clasificación.
- Ofrecer el modelo adquirido para la toma de decisiones.

1.4 Justificación

Un conductor es responsable de las vidas de las personas que se encuentren en el vehículo. Al medir los riesgos y movimientos que el vehículo pueda presentar mientras se encuentre en circulación, a través de los resultados capturados por los sensores, fue posible calificar los movimientos que pueden tener un mal o un buen conductor.

El impacto de esta investigación se vio reflejado en un incremento del conocimiento en las Ciencias de la Computación derivado del procedimiento aplicado al dataset de conducción vehicular (obtenido mediante el uso de tecnología de sensores y técnicas de estadística descriptiva), lo que sirvió para poder categorizar el estilo de conducción en prudente o imprudente. A partir de esta información se logró determinar las debidas precauciones para evitar cualquier irregularidad en el trayecto del vehículo y, por consiguiente, reducir el número de accidentes automovilísticos.

Los beneficios fueron:

- Para los conductores, minimizar los accidentes de tránsito.

- A los alumnos y personas en general, prevenir los riesgos a los que se exponen cuando ellos toman el transporte público o privado para trasladarse a la escuela, el trabajo y de vuelta a su casa.
- Asimismo, los resultados de la investigación se pueden ofrecer a las autoridades de gobierno u organismos encargados de la regulación o gestión del servicio de transporte de pasajeros o ciudadanos, por ejemplo, autoridades de la Secretaría de Movilidad.

1.5 Metodología utilizada

1.5.1 Enfoque de la investigación

De acuerdo con Hernández et al. (2014) el enfoque cuantitativo tiene la finalidad de establecer pautas de comportamiento utilizando la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico. La presente investigación tuvo un enfoque cuantitativo debido a que se utilizaron los datos obtenidos de la investigación de Hernández (2020) para aplicar un algoritmo que emplea fórmulas estadísticas y proporciona valores para determinar la clasificación de conducción vehicular en prudente e imprudente.

1.5.2 Fuentes de investigación

De acuerdo con la definición de Cabeza y Carvajal (s.f.), se consideran fuentes de información los materiales originales o elaborados que aportan datos a través de los cuales se accede al conocimiento; las fuentes de información son el conjunto de datos que permite conocer los elementos necesarios para cubrir la demanda de información o conocimiento de un usuario.

La información utilizada para el desarrollo de esta investigación pertenece a una fuente de información primaria ya que la información fue extraída de una tesis referente a la elaboración de un dataset con datos del desplazamiento de un vehículo en una ruta definida que fueron recolectados a través de sensores de movimientos.

1.5.3 Técnicas de recolección de datos

Se utilizó el conjunto de datos creado por Hernández (2020) en su investigación titulada “Construcción de Dataset de conducción vehicular”. Este conjunto de datos contiene valores en ejes x, y, z de los movimientos hechos por un vehículo de juguete al acelerar sobre una ruta, el cual experimentó distintas condiciones en la vía, tales como un camino plano, un camino con topes y uno con huecos o baches. Los resultados conseguidos de dicha investigación fueron analizados para aplicar fórmulas de estadística descriptiva con la finalidad de categorizar el estilo de conducción con base en los valores obtenidos por los sensores.

1.5.4 Selección de herramientas

1.5.4.1 XAMPP.

XAMPP es el entorno más popular de desarrollo con PHP. XAMPP es una distribución de Apache completamente gratuita y fácil de instalar que contiene MariaDB, PHP y Perl. El paquete de instalación de XAMPP ha sido diseñado para ser increíblemente fácil de instalar y usar. (Apache Friends, 2022)

Cada letra de XAMPP se refiere a uno de sus componentes, donde X es multiplataforma (Cross-Platform en inglés), A significa Apache (el servidor web), M es MySQL o MariaDB (los gestores de bases de datos) y las Ps son los lenguajes de programación PHP y Perl. De igual manera incluye otros componentes que no aparecen en las siglas, tal como lo es phpMyAdmin, la cual es una herramienta que permite administrar bases de datos de forma sencilla.

1.5.4.2 Python

Python es un lenguaje de programación interpretado, interactivo y orientado a objetos. Incorpora módulos, excepciones, tipado dinámico, tipos de datos dinámicos de muy alto nivel y clases. Soporta múltiples paradigmas de programación más allá de la programación orientada a objetos, como la programación procedimental y funcional. (Python Software Foundation, 2022)

Python ha sido uno de los lenguajes de programación más populares en los últimos años y esto se debe a su gran versatilidad. Cuenta con módulos para distintos tipos de desarrollos, tales como: desarrollo web, cálculo numérico y álgebra lineal, preprocesamiento y modificación de datos, entre muchos otros. Por lo que es una buena opción emplearlo como una herramienta para el análisis de datos.

Capítulo 2. Marco teórico

2.1 Marco referencial

Chachumba et al. (2020) desarrollaron una investigación consistente en realizar un entorno gráfico en realidad virtual juntamente con un vehículo dentro del cual se implementa un asistente de conducción autónomo, enfocado a la evasión de obstáculos presentes en la trayectoria del vehículo. Mediante un controlador; el algoritmo inteligente toma la decisión de maniobrar para evadir el obstáculo de acuerdo con los datos de posición de este; el proceso sistemático de control del automóvil es realizar la maniobra evasiva, evitar el obstáculo, volver el vehículo hacia su trayectoria original y devolver el control del vehículo al usuario para que este siga su camino. Una vez terminado el asistente de conducción, el aporte de su implementación será significativo en el aumento de seguridad activa de los vehículos al ser capaz de prevenir colisiones contra objetos que se encuentren obstaculizando el paso de este.

Arias (2017) abordó el diseño de un sistema de diagnóstico de fallas para el sistema de dirección asistida de un vehículo de categoría M1 (categorizado por el Ministerio de Transportes) con el objetivo de diagnosticar las fallas más relevantes de este sistema. El desarrollo se inició con un estudio del modelamiento matemático del sistema de dirección asistido eléctrico (EPS por sus siglas en inglés). Posteriormente, se diseñó el sistema de diagnóstico basado en dos etapas. La primera fue la detección de fallas, la cual está basada en la generación de Relaciones de Redundancia Analítica. Como segunda etapa, se diseñó el sistema de diagnóstico de fallas utilizando Redes Neuronales Artificiales a fin de poder reconocer los tipos de fallas de manera más robusta ante las perturbaciones. Las pruebas de validación del sistema de diagnóstico se realizaron utilizando los softwares de ingeniería Matlab y Carsim. Con estas pruebas se validó el adecuado funcionamiento del sistema de diagnóstico de fallas propuesto en un vehículo de

categoría M1. Finalmente se propuso un sistema para implementación en un vehículo real utilizando la plataforma Arduino.

Anaya et al. (2017) desarrollaron un soporte de comunicaciones basadas en topología física de los nodos o GeoNetworking. Este concepto significa que la red ad-hoc vehicular (VANET) basa la transmisión de los paquetes de datos en la localización topográfica de los diferentes nodos (vehículos) de la red, organizando el flujo de datos de una manera óptima que cubra los alrededores de cada vehículo. El núcleo de este sistema GeoNetworking es el algoritmo GeoRouting, que soporta el encaminamiento óptimo de los paquetes de datos y reorganiza la estructura de la red en función de las posiciones de los nodos. Este algoritmo se ha diseñado, implementado y probado en vehículos reales en carreteras reales con flujo de tráfico libre. Los resultados sugieren que las características del algoritmo de enrutamiento diseñado se pueden aplicar como núcleo de cualquier tipo de red vehicular para soportar aplicaciones de asistencia a la conducción basadas en sistemas de seguridad cooperativos.

Sarabia et al. (2016) diseñaron e implementaron un simulador de conducción vehicular utilizando un motor de videojuegos. El proceso de elaborar videojuegos requirió diversos conocimientos como diseño gráfico (elaboración de texturas y modelos 3D) y programación. Los motores de videojuegos son herramientas de programación utilizadas para la elaboración de proyectos de videojuegos. El propósito de este simulador es usarlo como base para trabajos de tutoría. El simulador de conducción vehicular cuenta con señales de tránsito, las cuales son validadas y penaliza al usuario en caso de cometer infracciones según las leyes ecuatorianas.

Márquez (2011) habla del surgimiento de nuevas tecnologías para un automóvil. Una de ellas es el BMW Foresight Transmission Control, es decir, una transmisión automática que tiene la capacidad de predecir curvas. Esta tecnología pretende, a través del navegador GPS, utilizar la información que éste provee para la anticipación de curvas, mostrar y predecir cualquier desperfecto en una carretera, entre otras, para que el

sistema tome la decisión de escoger las mejores rutas alternas. En caso de no encontrar una ruta alterna, el sistema reduce la velocidad del automóvil para evitar riesgos de frenado de motor en caso de encontrarse una curva o algún obstáculo en la trayectoria tomada para el destino del automóvil. Esta tecnología se puede ampliar más con las diferentes cámaras y sensores que se le pueden proporcionar al automóvil.

2.2 Marco conceptual

2.2.1 Conceptos y teorías relacionados con la investigación

Seguridad vial: La seguridad es un instrumento de formación, el cual propone una mejor solución para prevenir los accidentes automovilísticos, sin embargo, la utilización de estas leyes depende de cómo el conductor del automóvil las emplee. Se pretende analizar las consideraciones que puede tener la persona que maneje el automóvil y determinar si tuvo las debidas precauciones (Ucha, 2015).

Sociedad: La sociedad es la que puede exponerse mayormente a las regularidades de tomar un automóvil para el traslado de un lugar a otro, al realizar esto ellos están en constante riesgo, buscar la mejor solución y prevención puede verse constantemente afectada, ya sea la elección de transporte que pueda elegir ya sea privado o público, puede verse afectada por las condiciones en que pueda encontrarse el conductor o la carretera vial que él tome (Imaginario, 2019).

Conductor de vehículo: Es la persona que tiene la responsabilidad de tomar las máximas precauciones para evitar accidentes. Esta persona debe ser consciente de los riesgos a los que se puede exponer él y las personas que se encuentren pasajeras en un automóvil (Ramos, 2020).

Factores: Un factor es un elemento que actúa en conjunto con otros. La idea de riesgo, por su parte, alude a la cercanía o la inminencia de un daño. Esto quiere decir que a medida que un sujeto se vea afectado por más factores de riesgo, es más probable que experimente un cierto trastorno: cuenta con un riesgo sanitario superior al que tienen las personas que se encuentran alejadas de esos factores (Pérez y Gardey, 2018).

Prudente: Es un término que se emplea como sinónimo de sensatez, mesura, templanza, cautela o moderación. Se trata de la virtud que lleva a alguien a desenvolverse de modo justo y adecuado. (Pérez y Gardey, 2012).

Imprudente: Una imprudencia es un acto irreflexivo o torpe que implica un cierto riesgo para el que lo comete o para otros. A la persona que actúa de esta manera se le llama imprudente. En español abundan los sinónimos de este adjetivo, como atrevido, irresponsable, insensato, negligente o descuidado (Definición MX, 2017).

2.2.2 Conceptos y teorías relacionados con el marco tecnológico

Vehículo: Es un medio de locomoción que permite el traslado de un lugar a otro de personas o cosas. Cuando se traslada animales u objetos es llamado vehículo de transporte, como por ejemplo el tren, el automóvil, el camión, el carro, el barco, el avión, la bicicleta y la motocicleta, entre otros (Pérez y Merino, 2016).

Inteligencia Artificial: La inteligencia está vinculada a saber elegir las mejores opciones para resolver algún tipo de problema. Existen diversos tipos de inteligencia según sus atributos y procesos, como la inteligencia operativa, la inteligencia biológica o la inteligencia psicológica. Artificial, por otra parte, es un adjetivo que señala aquello hecho por mano, arte o ingenio del hombre. Lo artificial también permite nombrar a lo no natural o falso (Pérez y Gardey, 2014).

Sensores: El analizar los modos de conducir un automóvil, puede medir las frecuencias con la cual las personas tienen la capacidad de manejar el automóvil y cómo ellos pueden evitar los obstáculos que se encuentran en la carreta o ruta que tome el automóvil. Los resultados que pueden obtenerse con los registros de los sensores servirán para la identificación de mejores soluciones (Pérez y Gardey, 2010).

Dataset: Un conjunto de datos o dataset corresponde a los contenidos de una única tabla de base de datos o una única matriz de datos de estadística, donde cada columna de la tabla representa una variable en particular, y cada fila representa a un miembro determinado del conjunto de datos que estamos tratando. En un conjunto de datos o

dataset se tienen todos los valores que puede tener cada una de las variables, como por ejemplo la altura y el peso de un objeto, que corresponden a cada miembro del conjunto de datos. Cada uno de estos valores se conoce con el nombre de dato (Balagueró, 2018).

2.2.3 Conceptos y teorías relacionados con el método a utilizar

Estadística descriptiva: La estadística descriptiva es la rama de la estadística que recolecta, analiza y caracteriza un conjunto de datos (peso de la población, beneficios diarios de una empresa, temperatura mensual) con el objetivo de describir las características y comportamientos de este conjunto mediante medidas de resumen, tablas o gráficos (Requena, 2014a). En la ilustración 1 se muestra un ejemplo de estadística descriptiva.

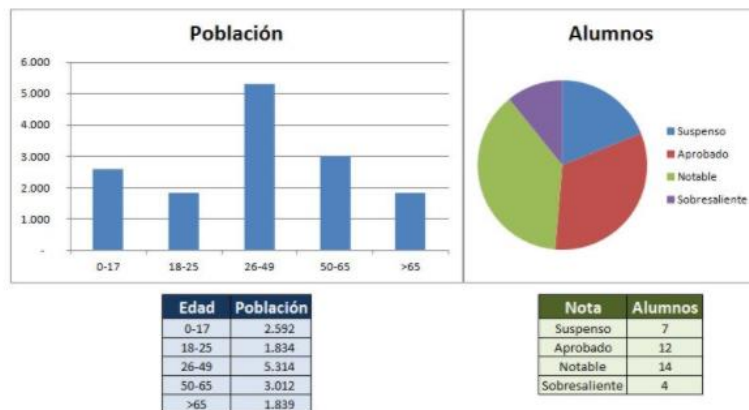


Ilustración 1: Estadística descriptiva
(Requena, 2014a)

Suma vectorial: Conocido también como suma del vector de magnitudes de aceleración (VAM en inglés). Si se suman dos magnitudes escalares, basta con sumar sus valores numéricos. Por ejemplo 10 w más 20 w son 30 w de potencia. Por el contrario, para la suma de vectores el proceso es más complejo, pues debemos de tener en cuenta dirección y sentido. Conociendo las componentes cartesianas de los vectores a sumar,

el vector resultante tendrá como componentes cartesianos la suma, eje a eje, de cada vector (Requena, 2017). En la ecuación 1 se muestra la fórmula de la suma vectorial.

$$VAM = \sum_{t_0}^{t_0+T} (|a_x| + |a_y| + |a_z|)dt$$

Ecuación 1: Suma de vectores

Media: También llamada promedio o media aritmética de un conjunto de datos (X_1, X_2, \dots, X_N) al valor característico de una serie de datos resultado de la suma de todas las observaciones dividido por el número total de datos (Requena, 2014c). En las ecuaciones 2 y 3 se muestran dos maneras de representar la fórmula de la media.

$$Media(X) = \bar{x} = \frac{\sum_{i=1}^N X_i}{N}$$

Ecuación 2: Media

Siendo (X_1, X_2, \dots, X_N) el conjunto de observaciones

$$Media(X) = \bar{x} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

Ecuación 3: Media (x)

Mediana: La mediana es un estadístico de posición central que parte la distribución en dos, es decir, deja la misma cantidad de valores a un lado que a otro. Para calcular la mediana es importante que los datos estén ordenados de mayor a menor, o al contrario de menor a mayor. Esto es, que tengan un orden (Marco, 2020). En la ecuación 4 se muestra la fórmula de la mediana cuando N es impar y en la ecuación 5 la fórmula de la mediana cuando N es par.

Si N es impar, la mediana es el valor que está al medio, es decir:

$$Mediana(X) = X_{\frac{N+1}{2}}$$

Ecuación 4: Mediana para n impar

Si N es par, la mediana es la media de los valores del centro, $N/2$ y $N/2+1$:

$$Mediana(X) = Media\left(X_{\frac{N}{2}}, X_{\frac{N}{2}+1}\right) = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2}$$

Ecuación 5: Mediana para n par

Moda: La moda es una medida de tendencia central que indica el valor que más se repite en un grupo de números. En un mismo estudio puede haber más de una moda, esto ocurre cuando dos (bimodal) o más números (multimodal) se repiten la misma cantidad de veces siendo este es el máximo número de veces del conjunto (Requena, 2014d). En la ecuación 6 se muestra la fórmula de la media.

$$M = L_i + \left(\frac{D_1}{D_1 + D_2}\right) A_i$$

Ecuación 6: Moda

Desviación estándar: La desviación estándar o desviación típica es una medida que ofrece información sobre la dispersión media de una variable. La desviación estándar es siempre mayor o igual que cero. Para entender este concepto necesitamos analizar 2 conceptos fundamentales.

Esperanza matemática, valor esperado o media: Es la media de nuestra serie de datos.

Desviación: La desviación es la separación que existe entre un valor cualquiera de la serie y la media (López, 2020). En la ecuación 7 se muestra la fórmula de la desviación estándar.

$$\sigma = \sqrt{\frac{\sum_1^N (x_i - \bar{X})^2}{N}}$$

Ecuación 7: Desviación estándar

$X \rightarrow$ Variable

$x_i \rightarrow$ Observación número i de la variable x

$N \rightarrow$ Número de observaciones

$\bar{X} \rightarrow$ Es la media de la variable X

Valor cuadrático medio (RMS en inglés): La media cuadrática o RMS (Root Mean Square) de un conjunto de valores (X_1, X_2, \dots, X_N) es una medida de posición central. Esta se define como la raíz cuadrada del promedio de los elementos al cuadrado (Requena, 2014b). En la ecuación 8 se muestra la fórmula del valor cuadrático medio.

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

Ecuación 8: Valor cuadrático medio

Valor absoluto: El valor absoluto lo conocemos como los no negativos y significa quitar el signo negativo y como resultado es que todos los números son positivos, se representan con dos líneas verticales y paralelas, dentro de estas dos líneas paralelas se colocan los números: $|x|$ (Muñoz, s. f.). En la ecuación 9 se muestran los diferentes resultados que puede tener el valor absoluto dependiendo el valor de x.

$$|x| = \begin{cases} x, si \ x > 0 \\ -x, si \ x < 0 \\ 0, si \ x = 0 \end{cases}$$

Ecuación 9: Valor absoluto de un número

Energía: Está relacionado con la capacidad de generar movimiento o lograr la transformación de algo. En el ámbito económico y tecnológico, la energía hace referencia a un recurso natural y los elementos asociados que permiten hacer un uso industrial del mismo (Pérez y Merino, 2008). En la ecuación 10 se muestra la fórmula de la energía.

$$E = \frac{1}{n} \sum_{i=1}^n |FFT_i|^2$$

Ecuación 10: Energía

2.3 Marco tecnológico

A continuación, se describirá la tecnología utilizada, las herramientas para el análisis del dataset y el desarrollo del método automático para la clasificación de conducción vehicular prudente o imprudente.

2.3.1 Herramientas de Software

XAMPP: Este entorno de desarrollo fue utilizado para crear un servidor local y administrar desde un navegador de internet una base de datos para almacenar los registros del dataset de conducción vehicular.

Python: La utilización de este lenguaje de programación fue para el análisis de los datos del dataset por medio de una computadora que fue capaz de medir y calcular resultados de funciones estadísticas, en las cuales los resultados obtenidos se utilizaron para evaluar la conducción en prudente e imprudente.

2.3.2 Herramientas de Hardware

Computadora: se requirió la utilización de una computadora para la visualización de los datos donde fue puesto a supervisión de un algoritmo, el cual tiene la capacidad de evaluar si los datos registrados pertenecen a una conducción prudente o imprudente. Para ello se calculó valores como la media, mediana, moda, desviación estándar, etcétera.

2.4 Marco legal

En la realización de este proyecto se consideró la ley de derecho de autor, el cual considera la utilización de cualquier investigación tomada para la realización del proyecto. La cual propone: La presente Ley, reglamentaria del artículo 28 constitucional, tiene por objeto la salvaguarda y promoción del acervo cultural de la Nación; protección de los derechos de los autores, de los artistas intérpretes o ejecutantes, así como de los editores, de los productores y de los organismos de radiodifusión, en relación con sus obras literarias o artísticas en todas sus manifestaciones, sus interpretaciones o ejecuciones, sus ediciones, sus fonogramas o videogramas, sus emisiones, así como de los otros derechos de propiedad intelectual.

Artículo 5.- La protección que otorga esta Ley se concede a las obras desde el momento en que hayan sido fijadas en un soporte material, independientemente del soporte, destino o modo de expresión. El reconocimiento de los derechos de autor y de los derechos conexos no requiere registro ni documento de ninguna especie ni que sea subordinado al cumplimiento de formalidad alguna.

Artículo 6.- Fijación es la incorporación de letras, números, signos, sonidos, imágenes y demás elementos en que se haya expresado la obra o de las representaciones digitales de aquellos que, en cualquier forma o soporte material, incluyendo los electrónicos, permita su percepción, reproducción u otra forma de comunicación.

Artículo 15.- Las obras literarias y artísticas publicadas en periódicos o revistas o transmitidas por radio, televisión u otros medios de difusión no pierden por ese hecho la protección legal.

Capítulo 3. Aplicación de la metodología y desarrollo

3.1 Valores estadísticos para los ejes del acelerómetro.

Se analizó la información del dataset por medio de secuencias de comandos en Python donde se calculó la media, es decir el promedio de cada una de las tablas que forman el dataset, mostrando una pequeña parte de los registros de cada tabla. Se contó con un registro de 9,000 datos capturados por cada una de las tablas. Esta información se muestra de las tablas 1 a la 9 donde x, y, z indican los ejes de movimiento del vehículo, el tiempo (medido en milisegundos), el tiempo que transcurre entre cada cambio de movimiento y la suma vectorial que se refiere a la suma de los valores de los ejes para cada movimiento.

x	y	z	Tiempo (ms)	Suma vectorial
-0.297	-0.624	9.634	0	0.015417469
0.095	-0.533	9.277	19	0.052723297
-0.013	-0.676	9.487	20	0.03047271
-0.166	-0.756	9.569	20	0.021380963
-0.178	-0.387	9.562	19	0.024313605
-0.195	-0.555	9.095	20	0.07094757
-0.528	-0.721	9.714	20	0.005604455
-0.422	-0.823	9.504	20	0.02661603
-0.3	-0.748	9.274	19	0.05107527
-0.022	-0.769	9.548	20	0.023553211
0.538	-0.796	9.49	20	0.027674891

Tabla 1: Primeros 20 registros de la tabla de camino con baches a velocidad baja.

x	y	z	Tiempo (ms)	Suma vectorial
0.149	-0.915	9.254	20	0.051955203
-0.177	-0.959	9.538	19	0.022658091
-0.038	-0.907	9.643	20	0.012677282
0.243	-0.856	9.356	20	0.041975634
0.111	-1.163	9.566	20	0.017627255
-0.08	-0.98	9.534	19	0.022979751
0.055	-0.652	9.299	20	0.049746
0.153	-0.677	9.397	20	0.039490553
-0.029	-0.951	9.573	19	0.019351191

Tabla 1: Primeros 20 registros de la tabla de camino con baches a velocidad baja (continuación).

x	y	z	Tiempo (ms)	Suma vectorial
0.517	-1.143	9.464	0	0.026831667
0.043	-0.621	9.429	19	0.036745623
-0.055	-0.708	9.538	20	0.025035754
-0.224	-0.501	9.538	20	0.026118734
0.061	-0.173	9.651	20	0.016030256
-0.055	-0.293	9.257	19	0.055881843
-0.044	-0.32	9.616	20	0.019222875
-0.091	-0.421	9.542	20	0.026328604
0.069	-0.512	9.354	19	0.045029967
0.136	-0.458	9.766	20	0.00329465
0.119	-0.442	9.445	20	0.036076927
0.001	-0.399	9.263	20	0.054883848
0.172	-0.495	9.659	19	0.013944471
0.264	-0.466	9.713	20	0.008383773
0.18	-0.252	9.288	20	0.052684877
-0.083	-0.155	9.661	19	0.015025504
-0.224	-0.166	9.314	20	0.05013538
-0.081	-0.372	9.337	20	0.047425215
0.166	-0.196	9.992	20	0.01888896
-0.122	-0.466	9.044	19	0.076776829

Tabla 2: Primeros 20 registros de la tabla de camino con baches a velocidad moderada

x	y	z	Tiempo (ms)	Suma vectorial
-0.098	-0.157	9.179	0	0.064131942
-0.093	-0.388	9.876	19	0.007549062
-0.481	0.018	9.602	20	0.019973818
0.118	-0.063	9.117	20	0.070542177
0.055	0.02	9.808	19	0.000186075
0.343	-0.078	9.341	20	0.047133459
0.044	-0.079	9.484	20	0.033187453
0.023	0.003	9.565	20	0.024971649
0.279	0.053	9.45	19	0.036262362
-0.081	-0.231	9.562	20	0.024960976
0.231	-0.14	9.506	20	0.03059767
-0.331	-0.048	9.563	20	0.024582359
0.195	-0.127	9.365	19	0.045067192
-0.263	0.13	9.338	20	0.047644506
-0.424	0.019	9.9	20	0.01010129
-0.164	-0.2	9.42	20	0.039393468
-0.209	-0.205	9.24	19	0.057631335
0.19	-0.388	10.096	20	0.030095743
1.285	-0.524	9.205	20	0.051068447
-0.819	0.51	9.183	19	0.058761996

Tabla 3: Primeros 20 registros de la tabla de camino con baches a velocidad rápida

x	y	z	Tiempo (ms)	Suma vectorial
-0.624	-1.428	8.905	0	0.078457592
0.776	-1.458	9.415	19	0.02560915
-0.347	-1.381	9.698	20	0.00081773
0.546	-1.688	9.41	20	0.023875651
-0.325	-1.489	9.326	19	0.036726794
0.105	-1.589	9.531	20	0.014972378
-0.212	-1.402	9.317	20	0.039319152
-0.025	-1.509	9.567	20	0.012710683
-0.275	-1.331	9.254	19	0.046557364
-0.138	-1.326	9.52	20	0.020092413

Tabla 4: Primeros 20 registros de la tabla de camino recto a velocidad baja

x	y	z	Tiempo (ms)	Suma vectorial
-0.244	-1.142	9.338	20	0.040699738
-0.056	-1.199	9.481	20	0.025822815
0.034	-1.459	9.494	19	0.020844787
0.444	-1.27	9.395	20	0.032534057
-0.029	-1.245	9.495	20	0.023820664
-0.119	-1.112	9.505	20	0.024407155
-0.397	-1.127	9.286	19	0.045610578
-0.131	-1.141	9.623	20	0.012100554
0.268	-1.209	9.316	20	0.042003618
0.155	-1.202	9.553	20	0.018392353

Tabla 4: Primeros 20 registros de la tabla de camino recto a velocidad baja (continuación).

x	y	z	Tiempo (ms)	Suma vectorial
-0.258	-1.601	9.928	0	0.025440369
-0.219	-1.442	8.986	19	0.072008222
0.146	-1.287	9.793	20	0.006960855
0.015	-1.422	9.235	20	0.047517833
-0.316	-1.586	9.562	20	0.011438512
-0.02	-1.578	9.342	19	0.034214333
-0.203	-1.429	9.501	20	0.020386559
-0.171	-1.308	9.328	20	0.039672632
0.12	-1.457	9.593	20	0.010830081
-0.031	-1.339	9.325	19	0.039684437
-0.164	-1.35	9.427	20	0.029094262
-0.033	-1.505	9.469	20	0.022638842
-0.004	-1.412	9.438	20	0.027213093
-0.099	-1.149	9.514	19	0.023074189
-0.128	-0.985	9.38	20	0.038486801
-0.225	-1.179	9.471	20	0.026834481
-0.051	-1.339	9.469	20	0.025143672
-0.098	-1.32	9.559	19	0.016288841
-0.087	-1.143	9.447	20	0.029939581
-0.311	-0.797	9.369	20	0.040980614

Tabla 5: Primeros 20 registros de la tabla de camino recto a velocidad moderada

x	y	z	Tiempo (ms)	Suma vectorial
-0.509	-1.203	8.906	0	0.082437851
0.095	-1.185	9.56	19	0.017978471
-0.015	-1.386	9.669	20	0.004297219
-0.295	-1.459	9.205	20	0.049482495
-0.335	-1.3	9.36	19	0.036107786
-0.083	-1.205	9.49	20	0.024815794
-0.018	-1.324	9.399	20	0.032435034
-0.268	-1.41	9.409	20	0.029782248
-0.265	-1.325	9.366	19	0.035375102
0.029	-1.908	9.076	20	0.054594131
-0.621	-4.921	9.311	20	0.075404742
0.518	-3.887	9.58	20	0.055198485
-0.99	-2.32	9.208	19	0.026785061
-0.675	-0.985	8.961	20	0.078470074
-0.866	-0.888	9.723	20	0.000836164
0.325	-1.906	9.608	20	0.00095631
0.897	-1.177	9.269	19	0.043181547
-0.51	-0.948	9.078	20	0.068134348
-0.951	-0.379	9.905	20	0.015062605
0.056	-0.604	9.143	20	0.065942919

Tabla 6: Primeros 20 registros de la tabla de camino recto a velocidad rápida

x	y	z	Tiempo (ms)	Suma vectorial
-1.155	-0.579	9.119	0	0.061154736
-1.056	-0.523	8.993	19	0.075445516
-1.118	-0.43	9.342	20	0.039910174
-1.073	-0.484	9.586	20	0.015494304
-0.976	-0.502	9.404	20	0.034879763
-0.971	-0.5	9.195	19	0.056102305
-1.008	-0.56	9.371	20	0.037545449
-0.877	-0.44	9.561	20	0.020263527
-0.632	-0.236	9.464	20	0.032822188
-0.491	-0.218	9.368	19	0.043487143

Tabla 7: Primeros 20 registros de la tabla de camino con topes a velocidad baja

x	y	z	Tiempo (ms)	Suma vectorial
-0.499	-0.296	9.537	20	0.026031432
-0.487	-0.301	9.601	20	0.019566317
-0.394	-0.157	9.535	20	0.027071537
0.085	2.116	10.09	19	0.050951992
-1.589	3.451	10.383	20	0.127040335
-1.442	0.063	9.344	20	0.036205655
0.24	0.69	7.556	19	0.226173888
-0.565	-0.324	8.869	20	0.093488006
-0.505	-0.599	10.128	20	0.035500328
-0.723	-0.303	9.344	20	0.044156327

Tabla 7: Primeros 20 registros de la tabla de camino con topes a velocidad baja (continuación).

x	y	z	Tiempo (ms)	Suma vectorial
0.24	0.164	9.063	0	0.075671723
0.161	0.024	9.059	19	0.076405469
0.096	0.153	9.938	20	0.013215218
0.6	0.163	9.349	20	0.044887712
0.601	0.157	9.331	19	0.046722442
0.55	0.321	9.483	20	0.031156119
0.065	0.449	9.69	20	0.011150385
0.031	0.524	9.311	20	0.04935937
-0.167	0.845	9.372	19	0.040622
0.057	0.638	9.421	20	0.037436255
0.185	0.571	9.655	20	0.013900207
0.131	0.431	9.483	20	0.0322433
0.25	0.629	9.669	19	0.011961032
-0.098	0.765	9.538	20	0.024553389
0.067	0.644	9.042	20	0.075927375
-0.241	0.079	9.18	20	0.063863129
-0.208	0.021	9.596	19	0.021582367
-0.064	-0.461	9.449	20	0.035631449
0.209	-0.622	9.568	20	0.022377791
0.291	-0.031	9.596	20	0.0213597

Tabla 8: Primeros 20 registros de la tabla de camino con topes a velocidad moderada

x	y	z	Tiempo (ms)	Suma vectorial
0.927	0.105	8.962	0	0.081505894
0.786	0.156	9.289	19	0.049592249
0.297	0.135	9.71	20	0.009635159
0.297	0.099	9.433	20	0.037900751
0.356	0.02	9.328	19	0.048439117
0.755	0.03	9.225	20	0.056483923
0.983	0.026	9.384	20	0.038187427
0.971	-0.1	9.575	20	0.018896218
0.818	-0.23	9.534	19	0.024282279
0.558	-0.22	9.399	20	0.039947097
0.401	-0.119	9.319	20	0.04909453
0.748	-0.125	9.492	20	0.029332593
0.514	0.171	9.276	19	0.052823289
0.616	1.364	9.435	20	0.026201129
-0.59	1.487	9.849	20	0.017133484
0.992	0.03	8.484	20	0.129271041
0.917	-0.739	9.069	19	0.067772667
1.053	-1.59	10.008	20	0.03854026
0.453	-0.72	10.263	20	0.049764827
1.186	-0.305	8.598	20	0.114702374

Tabla 9: Primeros 20 registros de la tabla de camino con topes a velocidad rápida

En las tablas 10 a la 18 se muestran las medias (promedios) de los datos capturados en cada una de las tablas anteriores, así como los valores de pico de cada eje (valores máximos y mínimos) y de la suma vectorial (determinación de pico general).

Media	0.066319213
Pico máximo	
Eje x	9.705
Eje y	5.121
Eje z	21.336

Tabla 10: Media, valores máximos y mínimos de la tabla Bache velocidad baja

Pico mínimo	
Eje x	-9.425
Eje y	-9.292
Eje z	1
Determinación de pico general	
Pico máximo	1.345807685
Pico mínimo	3.16929E-06

Tabla 10: Media, valores máximos y mínimos de la tabla Bache velocidad baja (continuación).

Media	0.108736358
Pico máximo	
Eje x	15.892
Eje y	6.779
Eje z	29.557
Pico mínimo	
Eje x	-9.996
Eje y	-12.702
Eje z	-3.999
Determinación de pico general	
Pico máximo	2.21927073
Pico mínimo	1.37682E-06

Tabla 11: Media, valores máximos y mínimos de la tabla Bache velocidad moderada

Media	0.177153827
Pico máximo	
Eje x	18.786
Eje y	9.836
Eje z	33.53
Pico mínimo	
Eje x	-14.548
Eje y	-14.698
Eje z	-10.506

Tabla 12: Media, valores máximos y mínimos de la tabla Bache velocidad rápida

Determinación de pico general	
Pico máximo	2.677567641
Pico mínimo	1.04222E-05

Tabla 12: Media, valores máximos y mínimos de la tabla Bache velocidad rápida (continuación).

Media	0.048159709
Pico máximo	
Eje x	2.65
Eje y	1.576
Eje z	12.975
Pico mínimo	
Eje x	-3.227
Eje y	-3.47
Eje z	5.1
Determinación de pico general	
Pico máximo	0.478533528
Pico mínimo	6.57755E-06

Tabla 13: Media, valores máximos y mínimos de la tabla Recta velocidad baja

Media	0.058150251
Pico máximo	
Eje x	4.4
Eje y	3.228
Eje z	13.845
Pico mínimo	
Eje x	-2.67
Eje y	-5.043
Eje z	5.102
Determinación de pico general	
Pico máximo	0.429369626
Pico mínimo	4.04735E-06

Tabla 14: Media, valores máximos y mínimos de la tabla Recta velocidad moderada

Media	0.077976536
Pico máximo	
Eje x	4.214
Eje y	2.261
Eje z	17.496
Pico mínimo	
Eje x	-4.157
Eje y	-6.591
Eje z	4.929
Determinación de pico general	
Pico máximo	0.791527086
Pico mínimo	1.12224E-06

Tabla 15: Media, valores máximos y mínimos de la tabla Recta velocidad rápida

Media	0.100071391
Pico máximo	
Eje x	6.619
Eje y	9.208
Eje z	28.624
Pico mínimo	
Eje x	-8.533
Eje y	-18.472
Eje z	-9.898
Determinación de pico general	
Pico máximo	2.040261948
Pico mínimo	1.07028E-06

Tabla 16: Media, valores máximos y mínimos de la tabla Tope velocidad baja

Media	0.163836078
Pico máximo	
Eje x	9.508
Eje y	14.728
Eje z	41.223

Tabla 17: Media, valores máximos y mínimos de la tabla Tope velocidad moderada

Pico mínimo	
Eje x	-11.861
Eje y	-27.238
Eje z	-14.439
Determinación de pico general	
Pico máximo	3.57641634
Pico mínimo	1.18406E-05

Tabla17: Media, valores máximos y mínimos de la tabla Tope velocidad moderada
(continuación)

Media	0.295648429
Pico máximo	
Eje x	26.405
Eje y	10.795
Eje z	52.113
Pico mínimo	
Eje x	-15.943
Eje y	-28.008
Eje z	-16.262
Determinación de pico general	
Pico máximo	4.796759424
Pico mínimo	7.57515E-06

Tabla 18: Media, valores máximos y mínimos de la tabla Tope velocidad rápida

3.2 Determinación y tabulación de características de los datos.

En las tablas 19 a la 27 se muestran los resultados de las determinaciones de la mediana, la moda, la raíz de la media cuadrática, la desviación estándar, la varianza y la energía de cada tabla que conforma el dataset.

Mediana	0.040930947
Moda	No hay
Raíz de la media cuadrática	0.110515093
Desviación estándar	0.088404456
Varianza	0.007815348
Energía	
Eje x	74.28389205
Eje y	82.62944572
Eje z	122.8971274

Tabla 19: Valores estadísticos para los datos de la tabla Bache velocidad baja

Mediana	0.05693756
Moda	No hay
Raíz de la media cuadrática	0.191095332
Desviación estándar	0.157142707
Varianza	0.02469383
Energía	
Eje x	98.49655849
Eje y	92.47062
Eje z	129.6459922

Tabla 20: Valores estadísticos para los datos de la tabla Bache velocidad moderada

Mediana	0.094854383
Moda	0.015047184
Raíz de la media cuadrática	0.295624963
Desviación estándar	0.236665672
Varianza	0.056010641
Energía	
Eje x	139.3168897
Eje y	129.4139642
Eje z	143.2850713

Tabla 21: Valores estadísticos para los datos de la tabla Bache velocidad rápida

Mediana	0.039378967
Moda	No hay
Raíz de la media cuadrática	0.061290086
Desviación estándar	0.037909327
Varianza	0.001437117
Energía	
Eje x	10.21333993
Eje y	35.04129688
Eje z	97.73070246

Tabla 22: Valores estadísticos para los datos de la tabla Recta velocidad baja

Mediana	0.046845845
Moda	0.014284194
Raíz de la media cuadrática	0.074448474
Desviación estándar	0.046487887
Varianza	0.002161124
Energía	
Eje x	14.45278467
Eje y	54.94046042
Eje z	102.0827921

Tabla 23: Valores estadísticos para los datos de la tabla Recta velocidad moderada

Mediana	0.061983129
Moda	0.066811066
Raíz de la media cuadrática	0.101655027
Desviación estándar	0.065218129
Varianza	0.004253404
Energía	
Eje x	33.35111972
Eje y	94.80140488
Eje z	108.6515974

Tabla 24: Valores estadísticos para los datos de la tabla Recta velocidad rápida

Mediana	0.049368287
Moda	0.002807639
Raíz de la media cuadrática	0.196281822
Desviación estándar	0.168855768
Varianza	0.02851227
Energía	
Eje x	78.86042333
Eje y	144.3747688
Eje z	165.9578274

Tabla 25: Valores estadísticos para los datos de la tabla Tope velocidad baja

Mediana	0.071189932
Moda	No hay
Raíz de la media cuadrática	0.338949752
Desviación estándar	0.296723228
Varianza	0.088044674
Energía	
Eje x	117.2570412
Eje y	162.9643725
Eje z	200.9280069

Tabla 26: Valores estadísticos para los datos de la tabla Tope velocidad moderada

Mediana	0.139326826
Moda	No hay
Raíz de la media cuadrática	0.586847039
Desviación estándar	0.506933383
Varianza	0.256981454
Energía	
Eje x	232.2675479
Eje y	203.2479381
Eje z	284.6224663

Tabla 27: Valores estadísticos para los datos de la tabla Tope velocidad rápida

3.3 Graficación de los resultados estadísticos.

Tomando en cuenta la tabulación de los datos determinados, se grafican los datos que se consideran tomar para el análisis y toma de decisión. De esta manera, en las ilustraciones 2 a la 4 se muestran la graficación de la media, desviación estándar y varianza, donde se puede visualizar la gran diferencia entre las formas de las velocidades, como también en qué tipo de superficie se encuentra conduciendo.

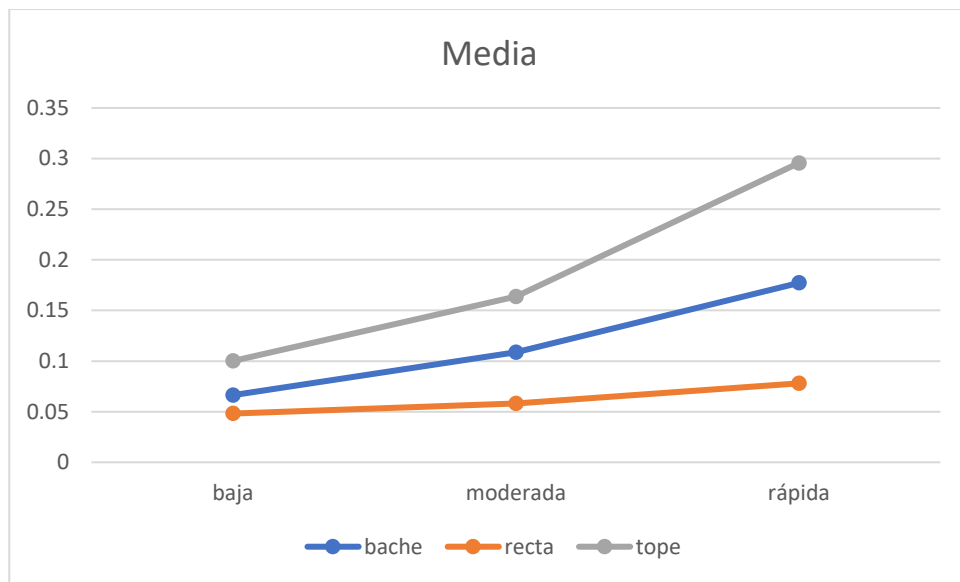


Ilustración 2: Graficación de la media

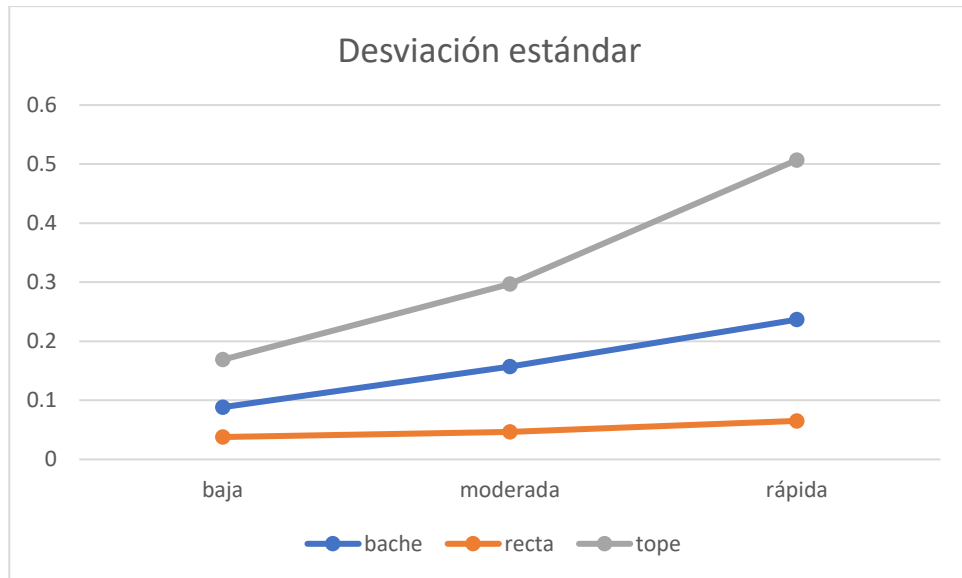


Ilustración 3: Graficación de la desviación estándar

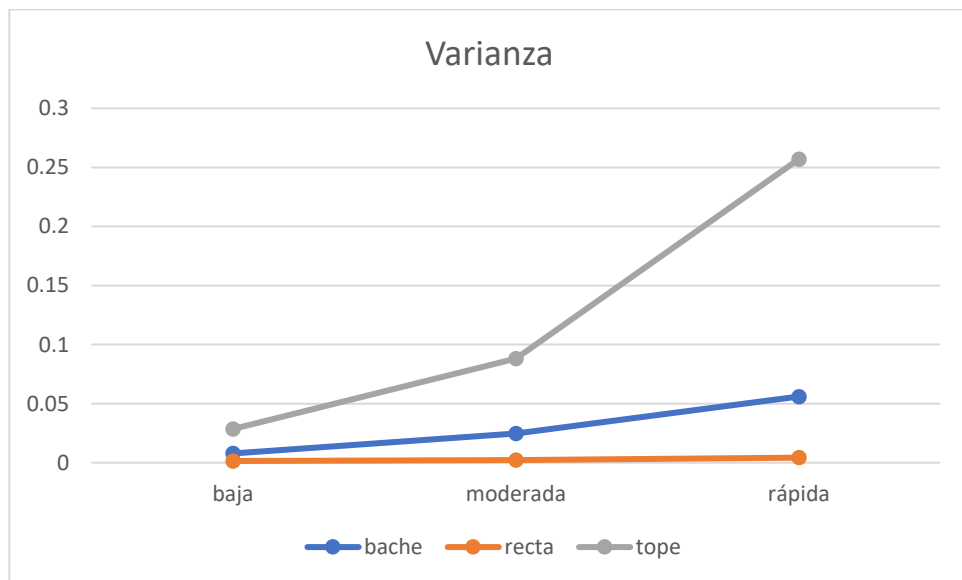


Ilustración 4: Graficación de la varianza

3.4 Características más relevantes.

Cada una de las fórmulas empleadas brinda información acerca de la relación que existe entre los movimientos de un vehículo y las distintas condiciones que pueden presentarse en un trayecto, dada el estado del camino. Al momento de ver sus valores graficados, hay ciertos valores estadísticos que ayudan a identificar con mayor facilidad la velocidad y un tipo de camino, en particular por la variación entre los valores. Y que resultan ser las características más relevantes para diseñar una clasificación de los estilos de conducción.

Un claro ejemplo de estas fórmulas estadísticas es la media, cuyos valores en las simulaciones del movimiento del vehículo hechas sobre un camino plano (camino recto) presentaron valores por debajo de 0.1, en las tres variaciones de velocidades. Por otra parte, en las simulaciones donde se encontraban topes en la vía, la media fue mayor a 0.1, independientemente de la velocidad empleada para cruzar por dicho camino. Otro ejemplo de fórmulas convenientes es la desviación estándar. Los valores obtenidos en las tres variaciones de velocidades para un camino recto son menores al valor más bajo de un camino con baches (0.088404456), que es la desviación estándar para un recorrido con baches en velocidad baja.

Por último, la otra fórmula estadística que también fue de mucha utilidad es la varianza. Para el caso de la varianza se logra apreciar que hay mucha similitud entre los valores de camino recto, camino con baches y camino con topes cuando estos son cruzados a baja velocidad, todos ellos presentan un valor por debajo de 0.05. Sin embargo, a medida que la velocidad incrementa para los tres tipos de caminos, la diferencia de valores entre cada tipo de camino es cada vez más evidente.

En las gráficas de las tres fórmulas descritas anteriormente se observa que el camino con topes es el que presentó variaciones más notorias en sus valores, con base en el aumento de la velocidad; en cambio el camino recto es el que menos alteraciones tiene en las gráficas, sobre todo en la varianza donde el cambio apenas es perceptible. Y el

camino con baches resultó ser un punto intermedio, en donde hay cambios más grandes entre la variación de las velocidades que las que se presentan en un camino recto, aunque no son incrementos tan altos como en el camino con topes. Estas fórmulas al ser las que muestran una diferenciación de valores más clara entre las diferentes alteraciones en el camino y cambios de velocidades servirán para crear funciones de clasificación de conducción que tomarán únicamente los valores necesarios para distinguir entre un estilo de conducción y otro. Consecuentemente las demás fórmulas no serán consideradas ya que la información que brindan en algunos casos no es de mucha utilidad para hacer esta diferenciación entre los estilos de conducción lo más notable posible.

Un caso incuestionable de esta falta de precisión para distinguir los estilos de conducción es la moda. Debido a que se presentaron muchas variaciones entre los datos captados por los sensores de movimiento, hay una enorme cantidad de datos con valores únicos, por lo que en algunos casos no fue posible encontrar una moda en los conjuntos de datos, y en aquellos donde se pudo establecer una moda, solo unos cuantos datos se repetían, como en el caso de bache velocidad rápida, recta velocidad moderada, recta velocidad rápida y tope velocidad baja, donde los valores repetidos no eran más de dos. Por esta razón la moda es una de las fórmulas que no fueron utilizadas para la graficación de los resultados estadísticos puesto que no presenta información relevante para diferenciar los estilos de conducción.

Capítulo 4. Pruebas y resultados

4.1 Interpretación de los resultados

En las tablas 28 y 29 se muestran las superficies y velocidades de conducción vehicular, donde se clasifica como bache velocidad baja (BVB), bache velocidad moderada (BVM), bache velocidad rápida (BVR), recta velocidad baja (RVB), recta velocidad moderada (RVM), recta velocidad rápida (RVR), tope velocidad baja (TVB), tope velocidad moderada (TVM) y tope velocidad rápida (TVR). Se determinó como condiciones la media, desviación estándar y varianza general; dichas funciones estadísticas se aplicaron a una muestra del 70% del dataset para la creación de los datos de entrenamiento y estas mismas funciones se aplicaron a una muestra del 30% restante para crear los datos de prueba, ya que contando con estos principales datos como base se determinó la clasificación de conducción vehicular.

Superficie y velocidad	Media	Desviación estándar	Varianza
BVB	0.068315951	0.092828463	0.008617124
BVM	0.110289016	0.159314255	0.025381032
BVR	0.173289517	0.238015991	0.056651612
RVB	0.046910711	0.037077582	0.001374747
RVM	0.055613521	0.044964751	0.002021829
RVR	0.080458747	0.067831764	0.004601148
TVB	0.099195935	0.166684144	0.027783604
TVM	0.166772723	0.295441171	0.087285485
TVR	0.294388886	0.50498005	0.255004851

Tabla 28: Datos de entrenamiento

Prueba	Media	Desviación estándar	Varianza
BVB	0.063890068	0.080849412	0.006536627
BVM	0.106884765	0.154569148	0.023891621
BVR	0.18727238	0.236574024	0.055967269
RVB	0.049938047	0.038087541	0.001450661
RVM	0.063463627	0.04913051	0.002413807
RVR	0.07312312	0.059574519	0.003549123
TVB	0.102610742	0.174486534	0.03044555
TVM	0.160161603	0.30422243	0.092551287
TVR	0.308702379	0.521043493	0.271486322

Tabla 29: Datos de prueba

4.2 Resultados de clasificación de conducción prudente.

Para la prueba y clasificación de conducción prudente, en las ilustraciones 5 a la 9 se muestra la determinación como acción prudente, el camino de las superficies y velocidades: BVB, RVB, RVM, RVR y TVB.

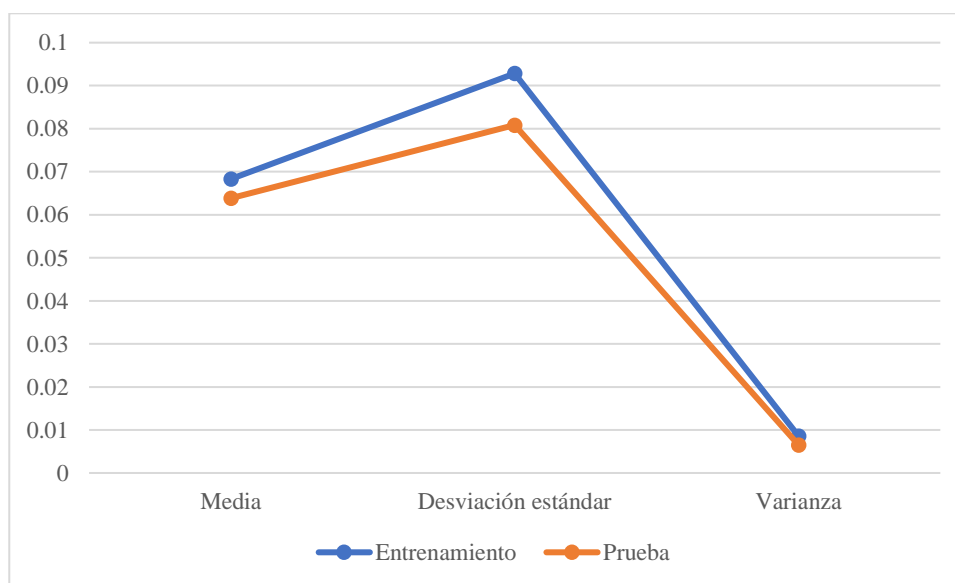


Ilustración 5: Camino BVB

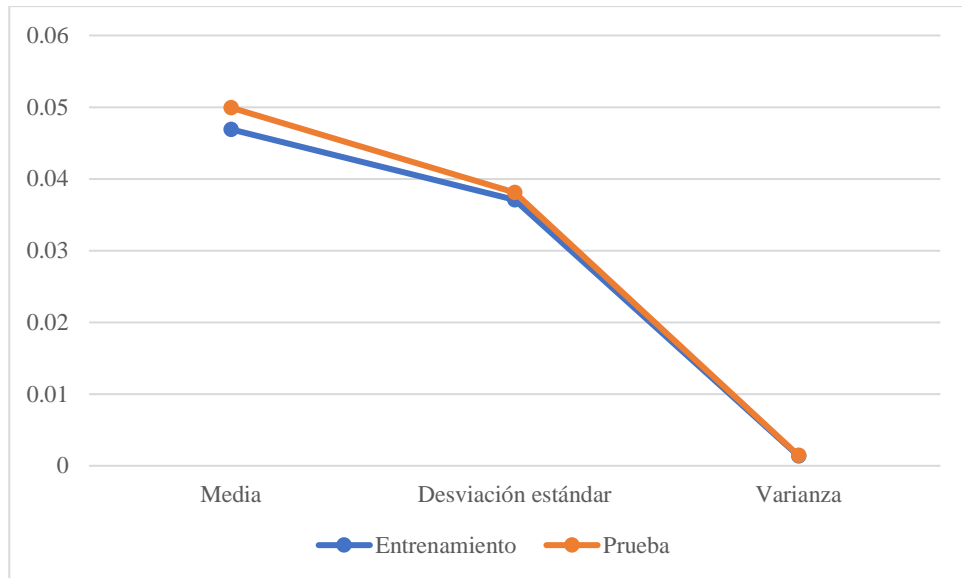


Ilustración 6: Camino RVB

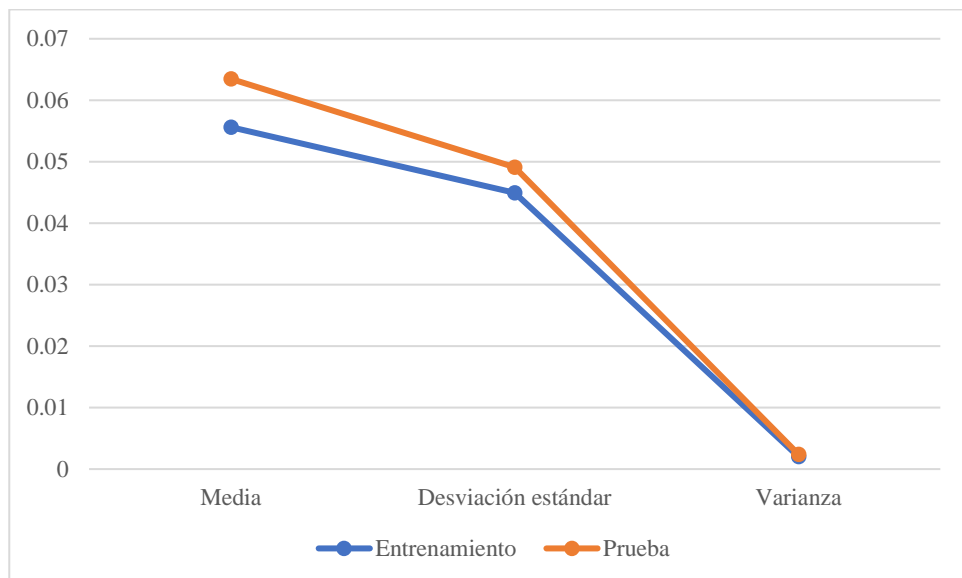


Ilustración 7: Camino RVM

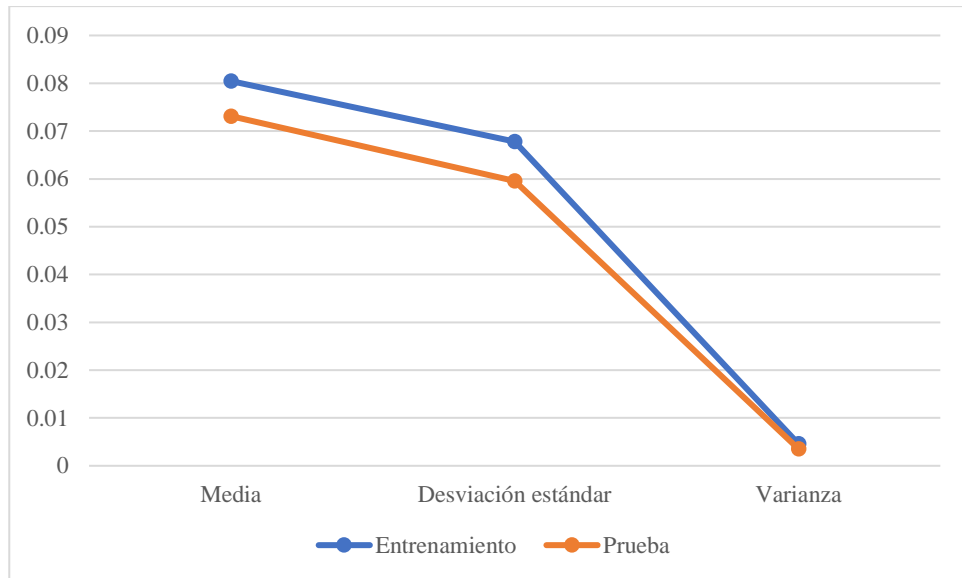


Ilustración 8: Camino RVR

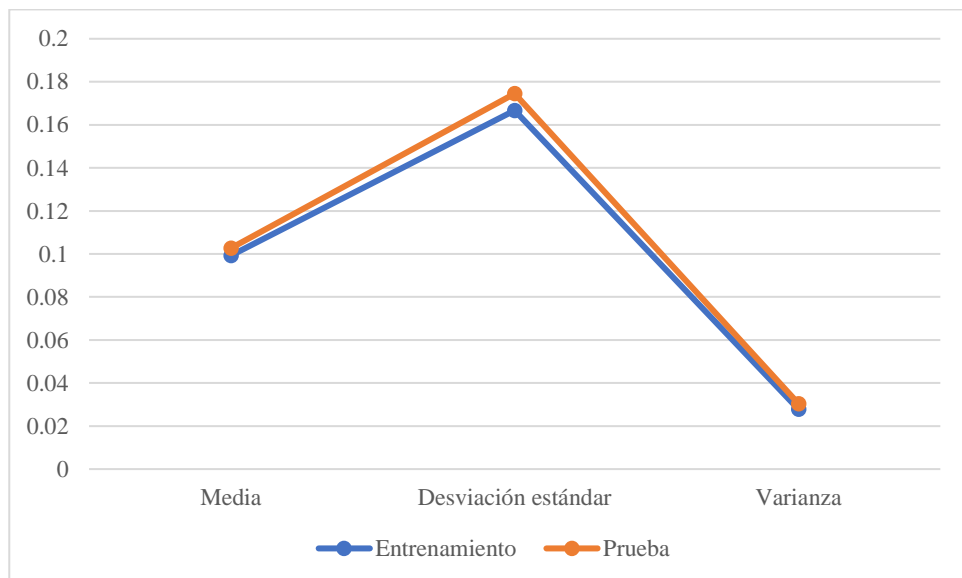


Ilustración 9: Camino TVB

Como se puede apreciar en las gráficas de resultados de clasificación de conducción prudente, los valores estadísticos media, desviación estándar y varianza de los datos de prueba muestran un comportamiento similar a los datos de entrenamiento. Por lo tanto,

es correcta la clasificación de los datos de prueba ya que se logró identificar a qué tipo de superficie y velocidad pertenecen los datos que fueron extraídos para las pruebas utilizando únicamente estos tres valores como características más relevantes.

4.3 Resultados de clasificación de conducción imprudente.

Para la prueba y clasificación de conducción imprudente, en las ilustraciones 10 a la 13 se muestra la determinación como acción imprudente, el camino de las superficies y velocidades: BVM, BVR, TVM y TVR.

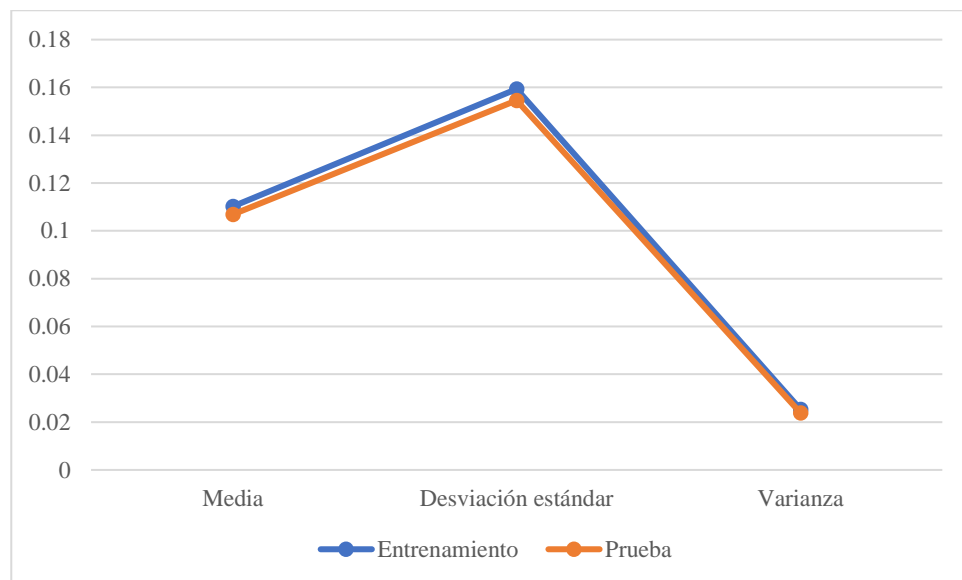


Ilustración 10: Camino BVM

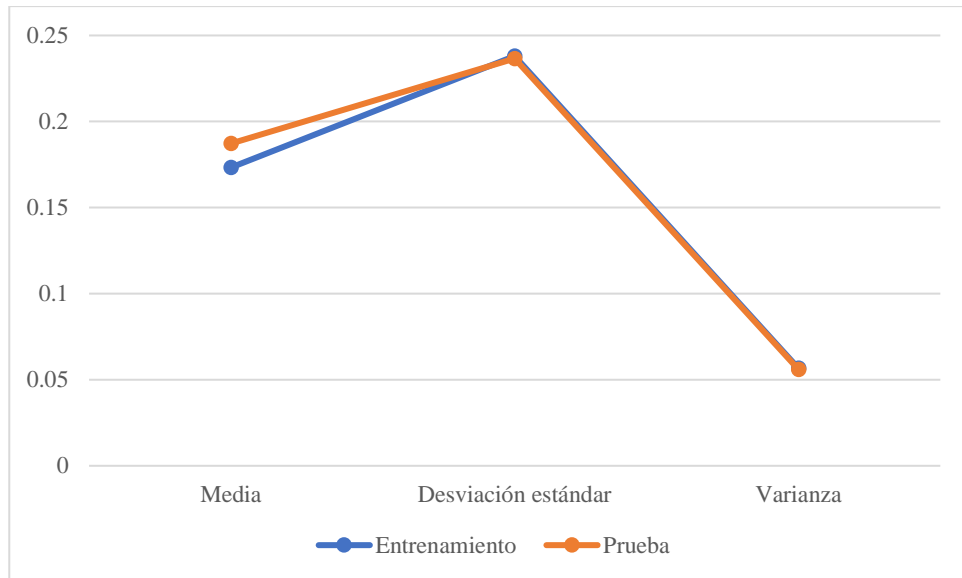


Ilustración 11: Camino BVR

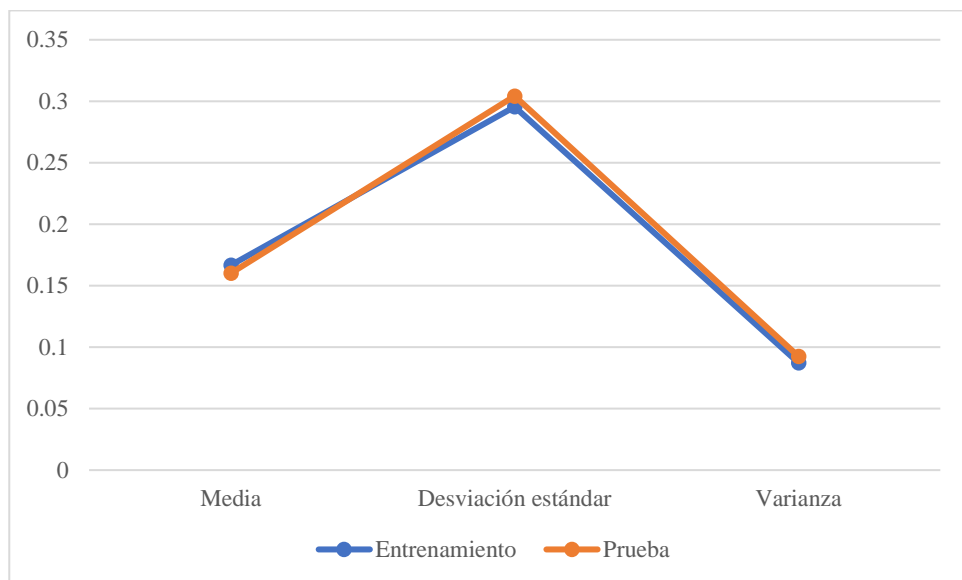


Ilustración 12: Camino TVM

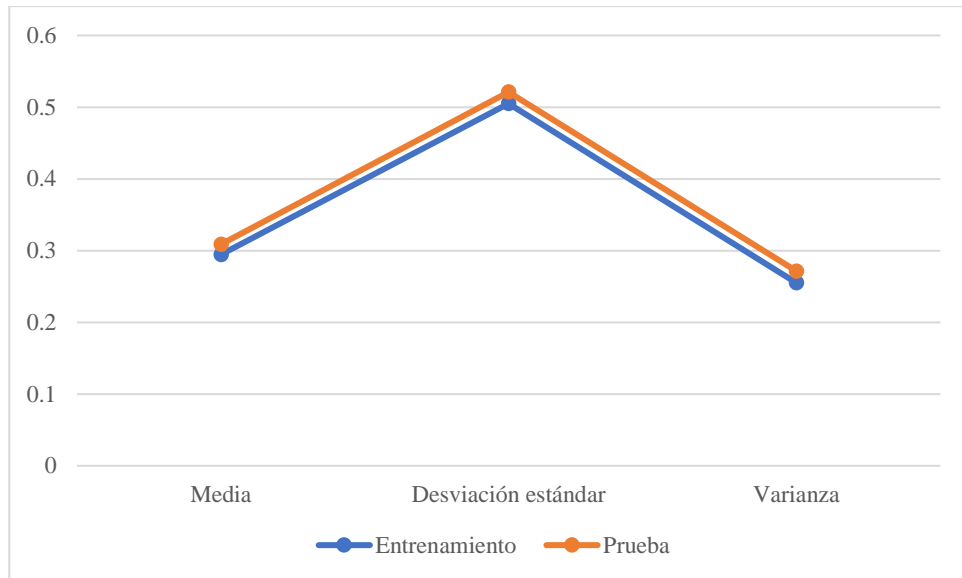


Ilustración 13: Camino TVR

Como se puede apreciar en las gráficas de resultados de clasificación de conducción imprudente, los valores estadísticos media, desviación estándar y varianza de los datos de prueba muestran un comportamiento similar a los datos de entrenamiento. Por lo tanto, es correcta la clasificación de los datos de prueba ya que se logró identificar a qué tipo de superficie y velocidad pertenecen los datos que fueron extraídos para las pruebas utilizando únicamente estos tres valores como características más relevantes.

Capítulo 5. Conclusiones y trabajos futuros

Este proyecto tuvo como objetivo clasificar la conducción prudente/imprudente de un vehículo, y para ello se empleó un dataset construido a partir de las lecturas del acelerómetro contenido en un smartphone colocado en un vehículo de prueba y recorriendo tres tipos de camino (camino con baches, camino con topes y camino plano), a distintas velocidades. A partir del análisis del dataset se pudo concluir que el uso de un acelerómetro permite capturar de manera eficiente la aceleración de un vehículo y por consiguiente tener la posibilidad de determinar el estilo de conducción.

Para el análisis de los datos capturados con el acelerómetro, se aplicaron las fórmulas estadísticas de la media, mediana, moda, energía, valor cuadrático medio, desviación estándar y varianza. Una vez hecho lo anterior, se analizaron y seleccionaron aquellas que eran útiles para diferenciar entre un estilo de conducción y otro. Lo anterior permitió la clasificación eficiente, por lo que podemos concluir que las características seleccionadas son suficientes para clasificar la conducción prudente o imprudente.

Es importante mencionar que existieron pequeñas diferencias entre los valores estadísticos obtenidos con los datos de entrenamiento y los valores del conjunto de prueba, por lo que la clasificación podría no ser 100% efectiva. Lo cual es común en procesos de clasificación empleando datos de sensores. Sin embargo, las diferencias no son tan relevantes como para considerar ineficiente la clasificación, esto es, las diferencias fueron pequeñas y no afectaron la calidad de la clasificación.

El uso de datasets con información de conducción para clasificar los estilos de conducción beneficia a la sociedad en general, porque sin ser necesario que una persona de modo personal vigile la forma de conducir, se puede conocer de modo automatizado la manera en que el chofer se comporta durante un trayecto.

El conocimiento adquirido durante esta investigación puede ser útil para quienes contratan choferes (por ejemplo, empresas con flotillas de transporte), ya que pueden realizar pruebas de conducción con los candidatos, sin tener necesidad de vigilarlos personalmente, y luego con los datos del sistema de clasificación que aquí se propone, conocer cuál es su estilo de conducción. También puede ser útil para las autoridades encargadas de gestionar el transporte público de pasajeros, ya que se podría identificar a los choferes que conducen en forma imprudente.

Cabe hacer mención que esta tesis de licenciatura forma parte de los productos comprometidos en el Proyecto con apoyo institucional denominado: “Sistemas inteligentes para el bienestar, la salud y la seguridad en ciudades inteligentes”. Por consiguiente, forma parte de los resultados hacia el logro de sistemas de transportación inteligentes.

Entre los trabajos futuros que se han considerado se encuentra utilizar otras características estadísticas para determinar si puede haber una mejor clasificación. Otro trabajo futuro puede ser emplear un algoritmo perteneciente al área de ciencia de datos para procesar los datos del acelerómetro y conocer los resultados que se obtendrían aplicando inteligencia artificial. Otra idea más es utilizar datasets que contengan información de otro tipo de sensores, como el giroscopio, para identificar comportamientos durante la conducción, por ejemplo, los giros bruscos.

Bibliografía

- Anaya, J. J.; Talavera, E.; Naranjo, J. E.; Jiménez, F. y Zato, J. G. (2017). Vehicular GeoNetworking based communications to support cooperative safety systems. *Securitas Vialis*. Vol. 9. Pp. 69-75. Recuperado de: <https://link.springer.com/article/10.1007/s12615-016-9096-8>
- Apache Friends (2022). ¿Qué es XAMPP? Recuperado de: <https://www.apachefriends.org/es/index.html>
- Arias, L. Á. (2017). Desarrollo de un sistema de diagnóstico de fallas en la dirección asistida eléctrica de automóviles. Tesis de Maestría en Ingeniería de Control y Automatización. Escuela de Posgrado. Pontificia Universidad Católica del Perú. Recuperado de: <https://search.proquest.com/docview/2425583510?pq-origsite=summon>
- Balagueró, T. (2018). ¿Qué son los datasets y los dataframes en el Big Data? Recuperado de: <https://www.deustoformacion.com/blog/programacion-diseno-web/que-son-datasets-dataframes-big-data>
- Cabeza, A. y Carvajal, A. (s. f.). Introducción a las Fuentes de Información en CC. Matemáticas. Recuperado de: <https://www.rua.unam.mx/portal/recursos/ficha/7952/concepto-y-tipos-de-fuentes-de-informacion-comunicacion-cientifica-el-proceso-de-busqueda-bibliografica>
- Cachumba, S. J.; Briceño, P. A.; Andaluz, V. H. y Erazo, G. (2020). Asistente Inteligente de Conducción Autónoma para Evitar Obstáculos y Colisiones. *Revista Ibérica de Sistemas e Tecnologías de Informação*. Núm. 28. Pp. 532-544. Recuperado de: <https://search.proquest.com/openview/fcf5293f3478d9a6fc22445b8c14a103/1?pq-origsite=gscholar&cbl=1006393>
- Definición MX (2017). Imprudencia. Recuperado de: <https://definicion.mx/imprudencia/>
- Dirección General de Tráfico. (2017). Psicología aplicada a la conducción. Recuperado de: <http://www.dgt.es/Galerias/seguridad-vial/formacion-vial/cursos-para-profesores->

y-directores-de-autoescuelas/XX-Curso-de-Profesores/Psicologia-aplicada-a-la-conduccion.pdf

Hernández, J. A. (2020). Construcción de Dataset de conducción vehicular. Tesis de Licenciatura en Sistemas Computacionales. División Académica de Ciencias y Tecnologías de la Información. Universidad Juárez Autónoma de Tabasco.

Hernández, R.; Fernández, C. y Baptista, P. (2014). Metodología de la investigación. 6ª ed. México: Ed. McGraw-Hill de México.

Imaginario, A. (2019). Sociedad. Recuperado de: <https://www.significados.com/sociedad/>

Instituto Nacional De Salud Pública. (2020). México, séptimo lugar mundial en siniestros viales. Recuperado de: <https://www.insp.mx/avisos/4761-seguridad-vial-accidentes-transito.html>

Jiménez, A. C. (2013). Estadística, representación tabular y grafica de datos. Recuperado de: <http://estadisticacrisanto.blogspot.com/2013/10/representacion-tabular-y-grafica-de.html>

López, J. F. (2020). Desviación estándar o típica. Economipedia. Recuperado de: <https://economipedia.com/definiciones/desviacion-tipica.html>

Marco, F. J. (2020). Mediana. Economipedia. Recuperado de: <https://economipedia.com/definiciones/mediana.html>

Márquez, R. (2011). BMW crea una tecnología capaz de predecir las curvas y prevenir accidentes. Recuperado de: <https://www.motorpasion.com/coches-hibridos-alternativos/bmw-crea-una-tecnologia-capaz-de-predecir-las-curvas-y-prevenir-accidentes>

Muñoz, Ó. A. (s. f.). Valor absoluto. Recuperado de: <https://www.uaeh.edu.mx/scige/boletin/prepa4/n2/m2.html>

Organización Mundial de la Salud. (2017). 10 datos sobre la seguridad vial en el mundo. Recuperado de: <https://www.who.int/features/factfiles/roadsafety/es/>

Pérez, J. y Gardey, A. (2010). Definición de sensor. Recuperado de: <https://definicion.de/sensor/>

- Pérez, J. y Gardey, A. (2012). Definición de prudencia. Recuperado de: <https://definicion.de/prudencia/>
- Pérez, J. y Gardey, A. (2014). Definición de inteligencia artificial. Recuperado de: <https://definicion.de/inteligencia-artificial/>
- Pérez, J. y Gardey, A. (2018). Definición de factores de riesgo. Recuperado de: <https://definicion.de/factores-de-riesgo/>
- Pérez, J. y Merino M. (2008). Energía. Definición. Recuperado de: <https://definicion.de/energia/>
- Pérez, J. y Merino, M. (2016). Definición de vehículo. Recuperado de: <https://definicion.de/vehiculo/>
- Python Software Foundation (2022). What is Python? Recuperado de: <https://docs.python.org/3/faq/general.html>
- Requena, B. (2014a). Estadística descriptiva. Universo formulas. Recuperado de: <https://www.universoformulas.com/estadistica/descriptiva/>
- Requena, B. (2014b). Media cuadrática. Universo formulas. Recuperado de: <https://www.universoformulas.com/estadistica/descriptiva/media-cuadratica/>
- Requena, B. (2014c). Media. Universo formulas. Recuperado de: <https://www.universoformulas.com/estadistica/descriptiva/media/>
- Requena, B. (2014d). Moda. Universo formulas. Recuperado de: <https://www.universoformulas.com/estadistica/descriptiva/moda/>
- Requena, B. (2017). Suma vectorial. Universo formulas. Recuperado de: <https://www.universoformulas.com/fisica/vectores/suma-vectores/>
- Sarabia, L.; Ginnio, A.; Guananga, B.; Jairo, A. (2016). Diseño e implementación de un simulador de conducción vehicular utilizando un motor de videojuegos. Tesis de Titulación de la Carrera de Ingeniería en Sistemas Computacionales. Universidad Católica de Santiago de Guayaquil. Recuperado de: <http://repositorio.ucsg.edu.ec/handle/3317/6749>

Secretaría de Salud. (2015). Accidentes viales, segunda causa de muerte en México. Recuperado de: <https://www.gob.mx/salud/articulos/accidentes-viales-segunda-causa-de-muerte-en-mexico>

Ucha, F. (2015). Definición de Seguridad vial. Recuperado de: <https://www.definicionabc.com/general/seguridad-vial.php>

Glosario

B

BVB: Bache velocidad baja.

BVM: Bache velocidad moderada.

BVR: Bache velocidad rápida.

E

EPS: Electronic Positioning System (Sistema de Dirección Asistido Eléctrico).

G

GPS: Global Positioning System (Sistema de Posicionamiento Global).

I

INSP: Instituto Nacional de Salud Pública.

R

RVB: Recta velocidad baja.

RVM: Recta velocidad moderada.

RVR: Recta velocidad rápida.

T

TVB: Tope velocidad baja.

TVM: Tope velocidad moderada.

TVR: Tope velocidad rápida.

V

VANET: Vehicular Ad-Hoc Network (Red Ad-Hoc Vehicular).

A N E X O S

Anexo A. Dataset utilizado para el análisis de la clasificación de conducción vehicular.

En las tablas 30 a la 38 se muestra el 0.003% del dataset obtenido por sensores para el análisis de una clasificación de conducción vehicular prudente/imprudente, dando como datos el eje x, y, z, tiempo en milisegundos y la suma vectorial, tomando en cuenta las distintas superficies y velocidades.

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.297	-0.624	9.634	0	0.015417469
0.095	-0.533	9.277	19	0.052723297
-0.013	-0.676	9.487	20	0.03047271
-0.166	-0.756	9.569	20	0.021380963
-0.178	-0.387	9.562	19	0.024313605
-0.195	-0.555	9.095	20	0.07094757
-0.528	-0.721	9.714	20	0.005604455
-0.422	-0.823	9.504	20	0.02661603
-0.3	-0.748	9.274	19	0.05107527
-0.022	-0.769	9.548	20	0.023553211
0.538	-0.796	9.49	20	0.027674891
0.149	-0.915	9.254	20	0.051955203
-0.177	-0.959	9.538	19	0.022658091
-0.038	-0.907	9.643	20	0.012677282
0.243	-0.856	9.356	20	0.041975634
0.111	-1.163	9.566	20	0.017627255
-0.08	-0.98	9.534	19	0.022979751
0.055	-0.652	9.299	20	0.049746
0.153	-0.677	9.397	20	0.039490553
-0.029	-0.951	9.573	19	0.019351191
-0.202	-0.713	9.548	20	0.02378029

Tabla 30: Muestra del 0.003% del dataset de la tabla Bache velocidad baja

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.122	-0.575	9.498	20	0.029951974
-0.085	-0.688	9.417	20	0.037463653
0.01	-0.691	9.365	19	0.042766201
-0.079	-0.581	9.577	20	0.021923286
-0.128	-0.506	9.595	20	0.020470393
-0.107	-0.757	9.328	20	0.045945173
0.189	-0.82	9.62	19	0.015623396

Tabla 31: Muestra del 0.003% del dataset de la tabla Bache velocidad baja (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.517	-1.143	9.464	0	0.026831667
0.043	-0.621	9.429	19	0.036745623
-0.055	-0.708	9.538	20	0.025035754
-0.224	-0.501	9.538	20	0.026118734
0.061	-0.173	9.651	20	0.016030256
-0.055	-0.293	9.257	19	0.055881843
-0.044	-0.32	9.616	20	0.019222875
-0.091	-0.421	9.542	20	0.026328604
0.069	-0.512	9.354	19	0.045029967
0.136	-0.458	9.766	20	0.00329465
0.119	-0.442	9.445	20	0.036076927
0.001	-0.399	9.263	20	0.054883848
0.172	-0.495	9.659	19	0.013944471
0.264	-0.466	9.713	20	0.008383773
0.18	-0.252	9.288	20	0.052684877
-0.083	-0.155	9.661	19	0.015025504
-0.224	-0.166	9.314	20	0.05013538
-0.081	-0.372	9.337	20	0.047425215
0.166	-0.196	9.992	20	0.01888896
-0.122	-0.466	9.044	19	0.076776829
-0.256	-0.281	9.729	20	0.007500177
0.1	-0.084	9.675	20	0.01367162
0.113	-0.373	8.972	20	0.084560538
-0.271	-0.217	9.808	19	0.000422278
-0.374	-0.22	9.677	20	0.01256645
-0.196	-0.446	8.965	20	0.084788295

Tabla 32: Muestra del 0.003% del dataset de la tabla Bache velocidad moderada

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.098	-0.157	9.179	0	0.064131942
-0.093	-0.388	9.876	19	0.007549062
-0.481	0.018	9.602	20	0.019973818
0.118	-0.063	9.117	20	0.070542177
0.055	0.02	9.808	19	0.000186075
0.343	-0.078	9.341	20	0.047133459
0.044	-0.079	9.484	20	0.033187453
0.023	0.003	9.565	20	0.024971649
0.279	0.053	9.45	19	0.036262362
-0.081	-0.231	9.562	20	0.024960976
0.231	-0.14	9.506	20	0.03059767
-0.331	-0.048	9.563	20	0.024582359
0.195	-0.127	9.365	19	0.045067192
-0.263	0.13	9.338	20	0.047644506
-0.424	0.019	9.9	20	0.01010129
-0.164	-0.2	9.42	20	0.039393468
-0.209	-0.205	9.24	19	0.057631335
0.19	-0.388	10.096	20	0.030095743
1.285	-0.524	9.205	20	0.051068447
-0.819	0.51	9.183	19	0.058761996
-0.948	-0.645	9.446	20	0.030036967
-0.153	0.079	9.816	20	0.000765563
0.104	0.189	9.174	20	0.064573292
-0.449	0.213	9.332	19	0.047377874
-0.669	-0.133	9.742	20	0.004500573
-0.47	0.423	9.438	20	0.035763695
0.096	0.114	8.936	20	0.08896608
-0.129	-0.385	9.825	19	0.00238395

Tabla 33: Muestra del 0.003% del dataset de la tabla Bache velocidad rápida

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.624	-1.428	8.905	0	0.078457592
0.776	-1.458	9.415	19	0.02560915
-0.347	-1.381	9.698	20	0.00081773
0.546	-1.688	9.41	20	0.023875651

Tabla 34: Muestra del 0.003% del dataset de la tabla Recta velocidad baja

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.325	-1.489	9.326	19	0.036726794
0.105	-1.589	9.531	20	0.014972378
-0.212	-1.402	9.317	20	0.039319152
-0.025	-1.509	9.567	20	0.012710683
-0.275	-1.331	9.254	19	0.046557364
-0.138	-1.326	9.52	20	0.020092413
-0.244	-1.142	9.338	20	0.040699738
-0.056	-1.199	9.481	20	0.025822815
0.034	-1.459	9.494	19	0.020844787
0.444	-1.27	9.395	20	0.032534057
-0.029	-1.245	9.495	20	0.023820664
-0.119	-1.112	9.505	20	0.024407155
-0.397	-1.127	9.286	19	0.045610578
-0.131	-1.141	9.623	20	0.012100554
0.268	-1.209	9.316	20	0.042003618
0.155	-1.202	9.553	20	0.018392353
-0.173	-1.226	9.449	19	0.028565269
-0.235	-1.115	9.462	20	0.028504899
0.205	-1.041	9.355	20	0.04026771
0.214	-1.318	9.403	19	0.031872301
-0.013	-1.403	9.535	20	0.017566134
-0.191	-1.25	9.445	20	0.028616615
-0.124	-0.678	9.3	20	0.049387779
0.111	-1.317	9.529	19	0.01934547

Tabla 35: Muestra del 0.003% del dataset de la tabla Recta velocidad baja (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.258	-1.601	9.928	0	0.025440369
-0.219	-1.442	8.986	19	0.072008222
0.146	-1.287	9.793	20	0.006960855
0.015	-1.422	9.235	20	0.047517833
-0.316	-1.586	9.562	20	0.011438512
-0.02	-1.578	9.342	19	0.034214333
-0.203	-1.429	9.501	20	0.020386559
-0.171	-1.308	9.328	20	0.039672632

Tabla 36: Muestra del 0.003% del dataset de la tabla Recta velocidad moderada

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.12	-1.457	9.593	20	0.010830081
-0.031	-1.339	9.325	19	0.039684437
-0.164	-1.35	9.427	20	0.029094262
-0.033	-1.505	9.469	20	0.022638842
-0.004	-1.412	9.438	20	0.027213093
-0.099	-1.149	9.514	19	0.023074189
-0.128	-0.985	9.38	20	0.038486801
-0.225	-1.179	9.471	20	0.026834481
-0.051	-1.339	9.469	20	0.025143672
-0.098	-1.32	9.559	19	0.016288841
-0.087	-1.143	9.447	20	0.029939581
-0.311	-0.797	9.369	20	0.040980614
0	-0.758	9.335	19	0.045288057
0.009	-1.011	9.478	20	0.028361626
0.322	-1.083	9.766	20	0.002155023
0.075	-0.9	9.425	20	0.034845019
-0.224	-0.578	9.449	19	0.034728691
-0.444	-1.015	9.525	20	0.022506426
0.28	-0.82	9.184	20	0.059654957
0.108	-1.311	10.015	20	0.029665672

Tabla 37: Muestra del 0.003% del dataset de la tabla Recta velocidad moderada (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.509	-1.203	8.906	0	0.082437851
0.095	-1.185	9.56	19	0.017978471
-0.015	-1.386	9.669	20	0.004297219
-0.295	-1.459	9.205	20	0.049482495
-0.335	-1.3	9.36	19	0.036107786
-0.083	-1.205	9.49	20	0.024815794
-0.018	-1.324	9.399	20	0.032435034
-0.268	-1.41	9.409	20	0.029782248
-0.265	-1.325	9.366	19	0.035375102
0.029	-1.908	9.076	20	0.054594131
-0.621	-4.921	9.311	20	0.075404742
0.518	-3.887	9.58	20	0.055198485

Tabla 38: Muestra del 0.003% del dataset de la tabla Recta velocidad rápida

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-0.99	-2.32	9.208	19	0.026785061
-0.675	-0.985	8.961	20	0.078470074
-0.866	-0.888	9.723	20	0.000836164
0.325	-1.906	9.608	20	0.00095631
0.897	-1.177	9.269	19	0.043181547
-0.51	-0.948	9.078	20	0.068134348
-0.951	-0.379	9.905	20	0.015062605
0.056	-0.604	9.143	20	0.065942919
0.639	-1.192	9.217	19	0.050387319
-0.513	-1.671	10.031	20	0.037936743
-0.212	-0.859	8.953	20	0.082914132
-0.506	-0.702	9.548	19	0.022718217
0.475	-0.9	9.417	20	0.034472234
0.257	-0.871	9.84	20	0.007320706
-2.699	-3.861	8.427	20	0.01586764
0.217	-1.082	10.542	19	0.08048957

Tabla 39: Muestra del 0.003% del dataset de la tabla Recta velocidad rápida (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
-1.155	-0.579	9.119	0	0.061154736
-1.056	-0.523	8.993	19	0.075445516
-1.118	-0.43	9.342	20	0.039910174
-1.073	-0.484	9.586	20	0.015494304
-0.976	-0.502	9.404	20	0.034879763
-0.971	-0.5	9.195	19	0.056102305
-1.008	-0.56	9.371	20	0.037545449
-0.877	-0.44	9.561	20	0.020263527
-0.632	-0.236	9.464	20	0.032822188
-0.491	-0.218	9.368	19	0.043487143
-0.499	-0.296	9.537	20	0.026031432
-0.487	-0.301	9.601	20	0.019566317
-0.394	-0.157	9.535	20	0.027071537
0.085	2.116	10.09	19	0.050951992
-1.589	3.451	10.383	20	0.127040335
-1.442	0.063	9.344	20	0.036205655

Tabla 40: Muestra del 0.003% del dataset de la tabla Tope velocidad baja

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.24	0.69	7.556	19	0.226173888
-0.565	-0.324	8.869	20	0.093488006
-0.505	-0.599	10.128	20	0.035500328
-0.723	-0.303	9.344	20	0.044156327
-1.206	0.006	9.415	19	0.032423242
-1.284	-0.229	9.119	20	0.0609786
0.378	-0.44	9.476	20	0.032238742
0.095	-0.036	9.655	20	0.015745721
-0.19	0.723	10.164	19	0.038884152
-0.086	0.398	9.55	20	0.025619094
-0.164	-0.261	9.373	20	0.044029838
-0.62	-0.355	9.21	20	0.05834161

Tabla 41: Muestra del 0.003% del dataset de la tabla Tope velocidad baja (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.24	0.164	9.063	0	0.075671723
0.161	0.024	9.059	19	0.076405469
0.096	0.153	9.938	20	0.013215218
0.6	0.163	9.349	20	0.044887712
0.601	0.157	9.331	19	0.046722442
0.55	0.321	9.483	20	0.031156119
0.065	0.449	9.69	20	0.011150385
0.031	0.524	9.311	20	0.04935937
-0.167	0.845	9.372	19	0.040622
0.057	0.638	9.421	20	0.037436255
0.185	0.571	9.655	20	0.013900207
0.131	0.431	9.483	20	0.0322433
0.25	0.629	9.669	19	0.011961032
-0.098	0.765	9.538	20	0.024553389
0.067	0.644	9.042	20	0.075927375
-0.241	0.079	9.18	20	0.063863129
-0.208	0.021	9.596	19	0.021582367
-0.064	-0.461	9.449	20	0.035631449
0.209	-0.622	9.568	20	0.022377791
0.291	-0.031	9.596	20	0.0213597

Tabla 42: Muestra del 0.003% del dataset de la tabla Tope velocidad moderada

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.372	-0.227	8.961	19	0.085464793
0.269	-0.61	9.664	20	0.012541446
0.423	-0.263	9.979	20	0.018493708
0.1	-0.096	9.59	19	0.022323973
0.354	-0.054	9.082	20	0.073490628
0.449	-0.172	9.253	20	0.055506217
0.326	-0.107	9.845	20	0.004177075
0.21	0.104	9.209	19	0.060960126

Tabla 43: Muestra del 0.003% del dataset de la tabla Tope velocidad moderada (continuación)

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.927	0.105	8.962	0	0.081505894
0.786	0.156	9.289	19	0.049592249
0.297	0.135	9.71	20	0.009635159
0.297	0.099	9.433	20	0.037900751
0.356	0.02	9.328	19	0.048439117
0.755	0.03	9.225	20	0.056483923
0.983	0.026	9.384	20	0.038187427
0.971	-0.1	9.575	20	0.018896218
0.818	-0.23	9.534	19	0.024282279
0.558	-0.22	9.399	20	0.039947097
0.401	-0.119	9.319	20	0.04909453
0.748	-0.125	9.492	20	0.029332593
0.514	0.171	9.276	19	0.052823289
0.616	1.364	9.435	20	0.026201129
-0.59	1.487	9.849	20	0.017133484
0.992	0.03	8.484	20	0.129271041
0.917	-0.739	9.069	19	0.067772667
1.053	-1.59	10.008	20	0.03854026
0.453	-0.72	10.263	20	0.049764827
1.186	-0.305	8.598	20	0.114702374
0.844	-0.415	8.769	19	0.100989541
0.525	0.108	9.804	20	0.000880804
0.94	0.165	10.114	20	0.035568618
1.045	0.364	9.306	20	0.044693061

Tabla 44: Muestra del 0.003% del dataset de la tabla Tope velocidad rápida

Eje x	Eje y	Eje z	Tiempo (milisegundos)	Suma vectorial
0.881	0.61	8.257	19	0.151249516
0.143	2.044	9.477	20	0.011623434
-1.207	1.042	10.073	20	0.039595149
0.558	-0.503	8.002	19	0.180714869

Tabla 45: Muestra del 0.003% del dataset de la tabla Tope velocidad rápida (continuación)

Anexo B. Código fuente utilizado para el análisis de la clasificación de conducción vehicular.

En las ilustraciones 14 a la 26 se muestra el código fuente escrito en Python donde se realiza una conexión a una base de datos, en la cual se encuentran almacenados los registros del dataset y posteriormente se emplean dichos valores para crear los datos de entrenamiento y los datos de prueba que utiliza el algoritmo para hacer la clasificación del tipo de conducción vehicular.

```
# Clase que se utilizará para crear los datos de entrenamiento y los datos de prueba.
class Data:
    def __init__(self, driving_type, mean, std_deviation, variance):
        self.driving_type = driving_type
        self.mean = mean
        self.std_deviation = std_deviation
        self.variance = variance
```

Ilustración 14: Código fuente. Clase Data

```
# Funciones para añadir un formato a los resultados mostrados.
def print_table(list):
    print("+-----+")
    print("|Conducción|      Media      |Desv. estándar|  Varianza  |")
    print("+-----+")
    for element in list:
        driving_type = element.driving_type
        if driving_type.find("_") > -1:
            driving_type = driving_type.split("_")
            acronym = ""
            for word in driving_type:
                acronym = acronym + word[0:1]
            driving_type = " " + acronym.upper() + " "
        value = "{:.10f}"
        print("| " + driving_type + " ", end="|")
        print(" " + value.format(element.mean) + " ", end="|")
        print(" " + value.format(element.std_deviation) + " ", end="|")
        print(" " + value.format(element.variance) + " |")
    print("+-----+\n")
```

Ilustración 15: Código fuente. Funciones. Mostrar tabla

```

def format_content(content, reference):
    while len(content) < len(reference):
        if len(content) == len(reference) -1:
            content = content + "|"
            break
        content = content + " "
    return content

def format_text(text):
    text = text.capitalize()
    text = text.replace("_", " ")
    text = text.replace("rapida", "rápida")
    return text

```

Ilustración 16: Código fuente. Funciones. Formatear texto y contenido

```

# Importación de módulos para crear una conexión a una base de datos,
# trabajar con arreglos y utilizar funciones estadísticas como:
# media "numpy.mean()", desviación estándar "numpy.std()" y varianza "numpy.var()".
import mysql.connector
import numpy

db = mysql.connector.connect(
    host="localhost",
    user="root",
    password="",
    database="dataset_conduccion_vehicular"
)
cursor = db.cursor()

```

Ilustración 17: Código fuente. Importaciones


```

# Creación de los datos de entrenamiento.
cursor.execute("SHOW TABLES")
table_names = []
training_data = []

for table in cursor:
    table_name = str(table)
    table_names.append(table_name.strip("()", ","))

for table_name in table_names:
    cursor.execute("SELECT `suma_vectorial` FROM `" + table_name + "` WHERE id BETWEEN 1 AND 6000")
    values = numpy.array(cursor.fetchall())
    training_data.append(
        Data(
            table_name,
            numpy.mean(values),
            numpy.std(values),
            numpy.var(values)
        )
    )

print("\nDatos de entrenamiento:")
print_table(training_data)

```

Ilustración 18: Código fuente. Datos de entrenamiento

```

# Mostrar las bases de datos que existen.
database_list = []
print("Listado de bases de datos:")
cursor.execute("SHOW DATABASES")

for database in cursor:
    print(database)
    database_name = str(database)
    database_list.append(database_name.strip("()", ","))

print()

```

Ilustración 19: Código fuente. Listado de bases de datos

```

# El usuario seleccionará la base de datos de la cual se extraerán las pruebas.
x = True

while x:
    selected_database = input("Escriba el nombre de la base de datos para las pruebas: ")
    selected_database = selected_database.strip()
    selected_database = selected_database.strip("()'",")

    for existing_database in database_list:
        if selected_database == existing_database:
            x = False
            db.database = selected_database
            break

print("\nBase de datos seleccionada:", selected_database)

```

Ilustración 20: Código fuente. Selección de base de datos para pruebas

```

# Creación de los datos de prueba
cursor.execute("SHOW TABLES")
table_names = []
testing_data = []
test_number = 1

for table in cursor:
    table_name = str(table)
    table_names.append(table_name.strip("()'","))

```

Ilustración 21: Código fuente. Creación de datos de prueba

```

for table_name in table_names:
    query = "SELECT `suma_vectorial` FROM `" + table_name + "`"

    if selected_database == "dataset_conduccion_vehicular":
        query = query + " WHERE id BETWEEN 6001 AND 9000"

    cursor.execute(query)
    values = numpy.array(cursor.fetchall())
    testing_data.append(
        Data(
            "Prueba " + str(test_number),
            numpy.mean(values),
            numpy.std(values),
            numpy.var(values)
        )
    )
    test_number += 1

print("\nDatos de prueba:")
print_table(testing_data)

```

Ilustración 22: Código fuente. Creación de datos de prueba (continuación)

```

# Clasificación de las pruebas en base a los datos de entrenamiento.
test_number = 1
for test in testing_data:
    test.driving_type = []

    # Comparar la diferencia entre los valores de la prueba y los valores de cada uno
    # de los datos de entrenamiento.
    diff_between_means = []
    diff_between_std_deviations = []
    diff_between_variances = []

    for element in training_data:
        diff_between_means.append(abs(element.mean - test.mean))
        diff_between_std_deviations.append(abs(element.std_deviation - test.std_deviation))
        diff_between_variances.append(abs(element.variance - test.variance))

```

Ilustración 23: Código fuente. Clasificación. Comparación de valores

```

# La diferencia más pequeña de valores (media, desviación estándar y varianza)
# representa el tipo de conducción al que se aproxima más la prueba.
min_diff_between_means = min(diff_between_means)
min_diff_between_std_deviations = min(diff_between_std_deviations)
min_diff_between_variances = min(diff_between_variances)

for i in range(len(training_data)):
    if min_diff_between_means == diff_between_means[i]:
        test.driving_type.append(training_data[i].driving_type)
        break

for i in range(len(training_data)):
    if min_diff_between_std_deviations == diff_between_std_deviations[i]:
        test.driving_type.append(training_data[i].driving_type)
        break

for i in range(len(training_data)):
    if min_diff_between_variances == diff_between_variances[i]:
        test.driving_type.append(training_data[i].driving_type)
        break

```

Ilustración 24: Código fuente. Clasificación. Comparación de valores (continuación)

```

# Revisión de las respuestas del tipo de conducción de la prueba.
driving_types = test.driving_type
results = {}
for driving_type in driving_types:
    total_matches = driving_types.count(driving_type)
    results[driving_type] = total_matches

# El tipo de conducción repetido más veces es el resultado.
maximum = max(results.values())
for key, value in results.items():
    if value == maximum:
        test.driving_type = key
        break

```

Ilustración 25: Código fuente. Clasificación. Determinación de resultados

```

# Mostrar resultados del tipo de conducción de la prueba.
dividing_line = "+-----+"
print(dividing_line)

print("|   Determinación del tipo de conducción de la prueba", test_number, "   |")
print(dividing_line)

content = "|En base a la media: " + format_text(driving_types[0])
print(format_content(content, dividing_line))

content = "|En base a la desviación estándar: " + format_text(driving_types[1])
print(format_content(content, dividing_line))

content = "|En base a la varianza: " + format_text(driving_types[2])
print(format_content(content, dividing_line))

content = "|Resultado final: " + format_text(test.driving_type)
print(format_content(content, dividing_line))
print(dividing_line + "\n")

test_number += 1

```

Ilustración 26: Código fuente. Clasificación. Muestra de resultados