A decorative background featuring a network diagram with nodes and connecting lines. The nodes are represented by small circles, some of which are highlighted with blue outlines or filled with blue. The lines are thin and gray, forming a complex web-like structure. The overall aesthetic is clean and modern, with a focus on connectivity and data flow.

Detección de anomalías de conducción

Univ. Evelyn Cusi Lopez
Tutor: Ph.D.. Eduardo Di Santi



1.

Detección de anomalías

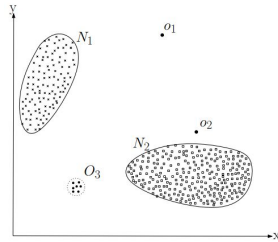
Definición, tipos y enfoques

“

La *detección de anomalías* se trata de encontrar patrones de interés (valores atípicos, excepciones, peculiaridades, etc.) que se desvían del comportamiento esperado dentro de los conjuntos de datos.



Tipos de anomalías

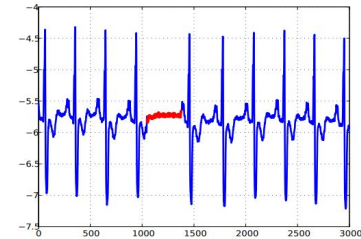
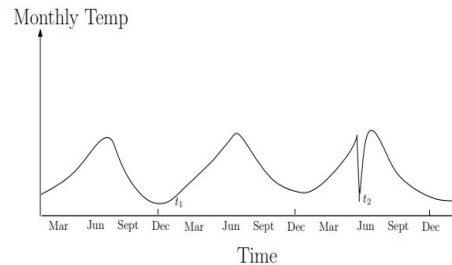


Anomalías de Punto

Las anomalías de punto son simplemente instancias únicas y anómalas dentro de un conjunto de datos más grande, es decir, éstas se encuentran separadas del resto de los datos.

Anomalías Contextuales

Estos son puntos que sólo se consideran anómalos en un contexto específico.



Anomalías Colectivas

Si una colección de instancias de datos relacionadas es anómala con respecto a todo el conjunto de datos, se denomina anomalía colectiva.

Enfoques de la detección de anomalías

Supervisada

Se basa en construir un modelo predictivo para clases normales vs. anomalías, posteriormente tomar cualquier instancia de datos no visto, comparar con el modelo y determinar a qué clase pertenece.

Semi-supervisada

Se basa en construir un modelo para la clase correspondiente al comportamiento normal, y usar el modelo para identificar anomalías en los datos de prueba

No supervisada

Estas técnicas suponen que las instancias normales son mucho más frecuentes que las anomalías en los datos de prueba, en caso de que esta suposición no sea cierta, tales técnicas sufren de una alta tasa de falsas alarmas.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

2.

Planteamiento del problema

Accidentes de tránsito

Cada año se pierden aproximadamente 1,35 millones de vidas como consecuencia de los accidentes de tránsito.



Entre 20 millones y 50 millones

de personas sufren traumatismos no mortales,
y muchos de esos traumatismos provocan una
discapacidad.

Objetivo General

El objetivo general del presente trabajo es desarrollar un mecanismo de detección de anomalías de conducción, mediante el uso de un dispositivo móvil y algoritmos de Aprendizaje Automático, con el fin de alertar de forma oportuna el hallazgo de un patrón anómalo en el manejo, tal como cansancio, ebriedad, o problemas de salud, ej, epilepsia.



Objetivos Especificos



1

Capturar los parámetros de manejo de un conductor mediante el uso de sensores de un dispositivo móvil.



2

Escalar los parámetros de manejo mediante técnicas de pre-procesamiento de datos.



3

Generar un modelo de aprendizaje automático que se ajuste a un comportamiento normal de manejo



4

Definir un método de detección de anomalías para generar una alerta de conducción anormal.



5

Evaluar el método de detección de anomalías con nuevas muestras.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels or types of nodes. The lines are thin and gray, connecting the nodes in a non-linear fashion.

3. **Metodología**

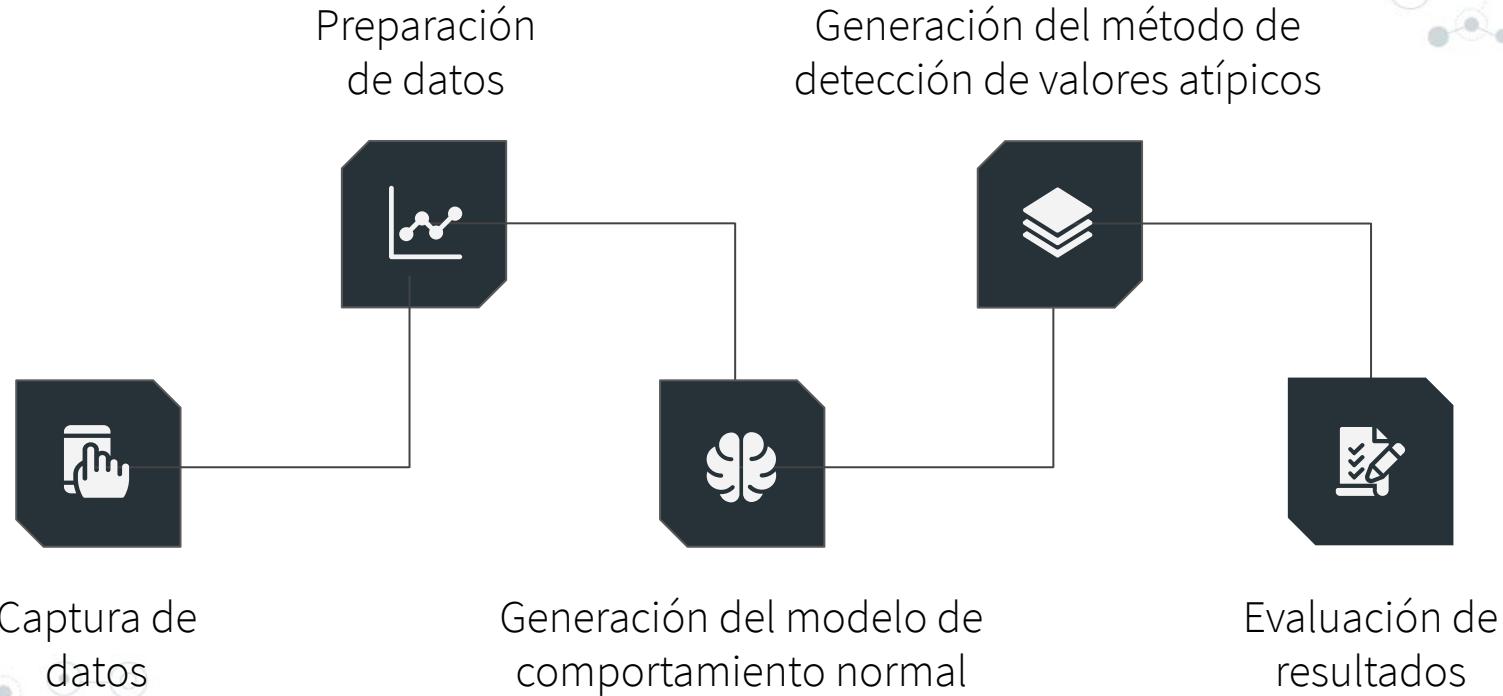


Metodología de investigación

El presente estudio se realizó con un enfoque **experimental**, teniendo como **hipótesis** la siguiente:

¿Es posible detectar anomalías de conducción mediante el uso de un dispositivo móvil y algoritmos de Aprendizaje Automático?

Enfoque sobre el problema





4.

Preparación de datos

Preparación de datos

PRE-PROCESAMIENTO



1

Eliminación de
ruido de los
datos



2

División del
conjunto de
datos



3

Normalización
de datos



4

Reducción de
datos



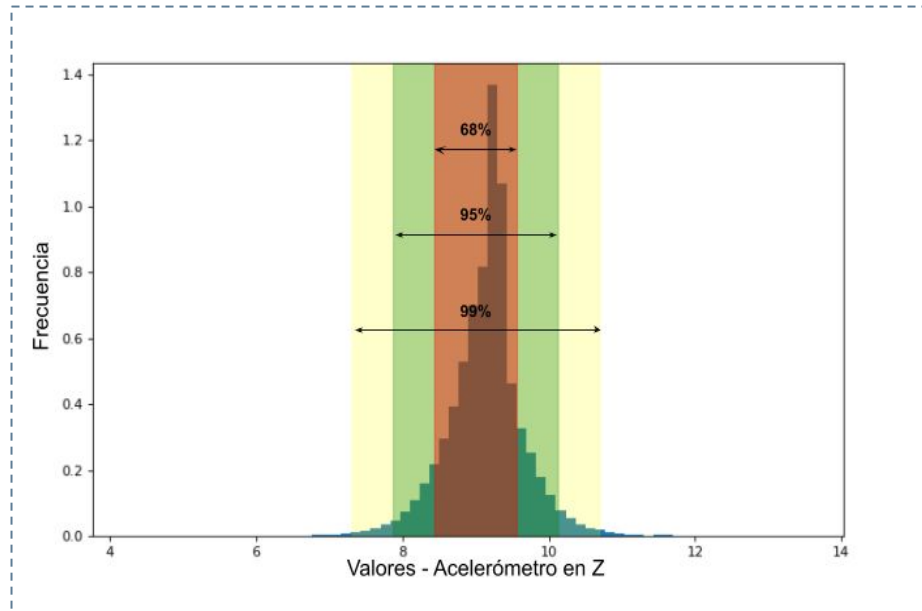
5

Transformación
de datos

Preparación de datos

ELIMINACIÓN DE RUIDO

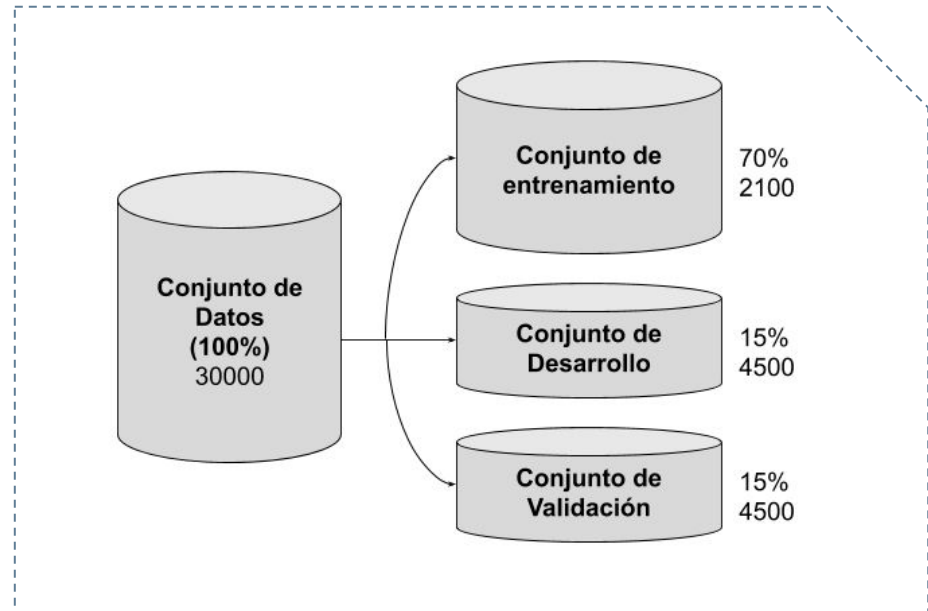
El ruido presente en el conjunto de datos puede afectar en gran medida el resultado de los algoritmos de Aprendizaje Automático, por lo tanto aquellos datos que contengan ruido deben ser eliminados en lo posible.

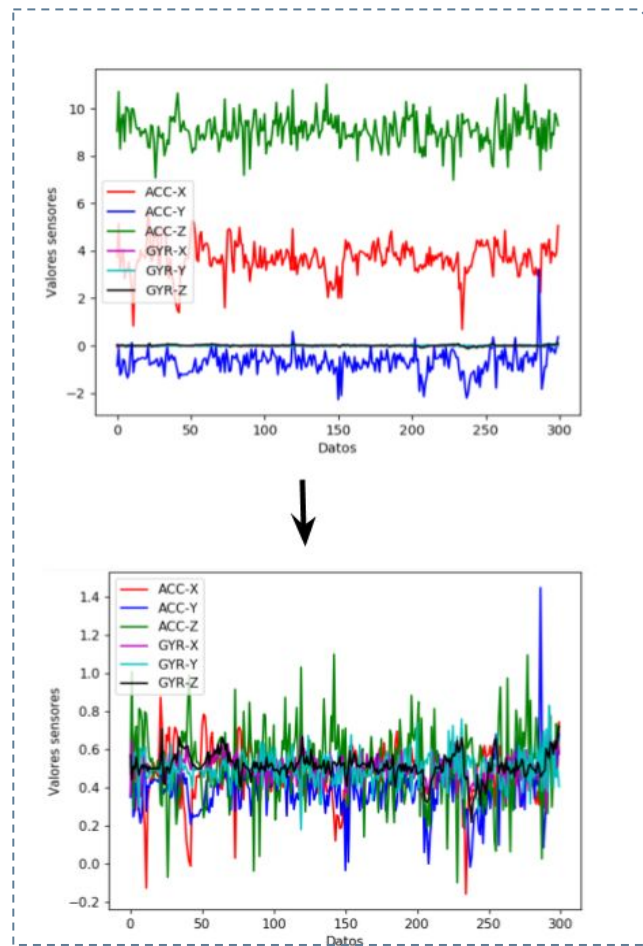


Preparación de datos

DIVISIÓN DEL CONJUNTO DE DATOS

La división del conjunto de datos tiene un gran impacto en la productividad, por lo cual es importante que al elegir los subconjuntos, estos tengan la misma distribución y deben ser elegidos aleatoriamente del conjunto de datos.





Preparación de datos

NORMALIZACIÓN DE DATOS

La normalización es el proceso de reescalar datos sin cambiar su comportamiento o naturaleza.

En la presente investigación se normalizo los datos usando el normalizado Min-Max, en un rango entre 0 y 1.

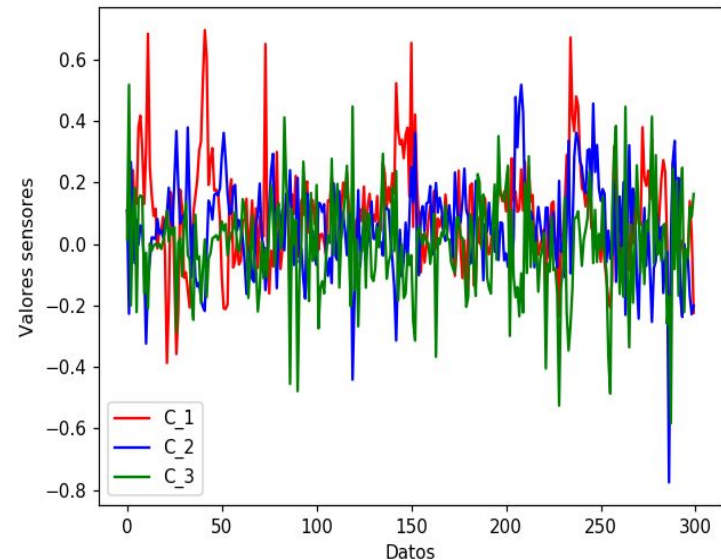
$$x_{normalizado} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Preparación de datos

REDUCCIÓN DE DATOS

Este proceso se basa en la adopción de algunas estrategias, tal que, el análisis de datos reducidos produce la misma información producida por los datos originales.

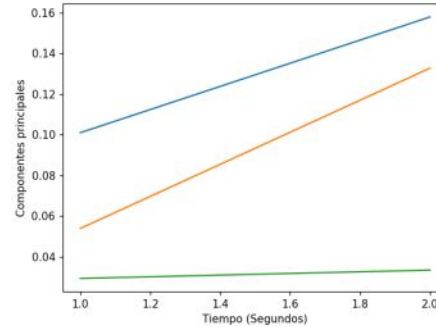
En este proyecto se usó PCA, la cual es una técnica estadística no supervisada y no paramétrica, que se utiliza para la reducción de dimensionalidad.



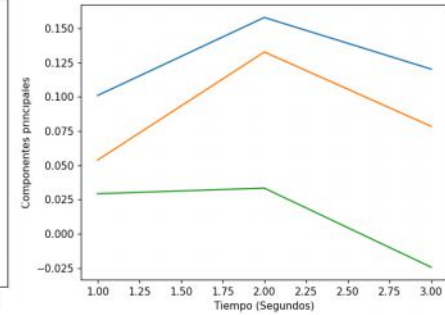
Preparación de datos

INGENIERÍA DE CARACTERÍSTICAS

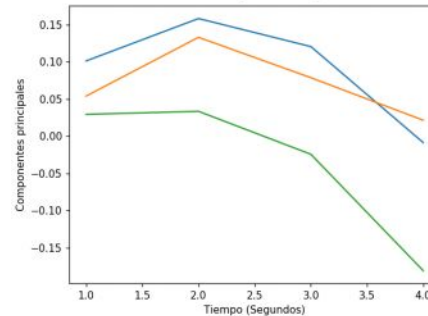
Esta técnica es útil para generar información adicional a partir de aquellos datos que no son lo suficientemente representativos por sí mismos.



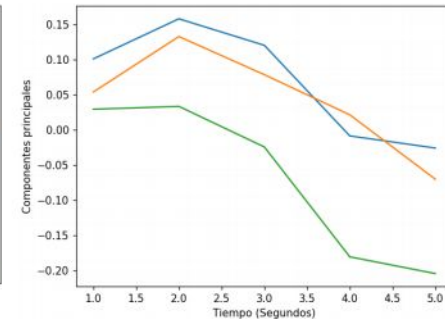
A: Serie de tiempo de 2 pasos.



B: Serie de tiempo de 3 pasos.



C: Serie de tiempo de 4 pasos.



D: Serie de tiempo de 5 pasos.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue and others in grey.

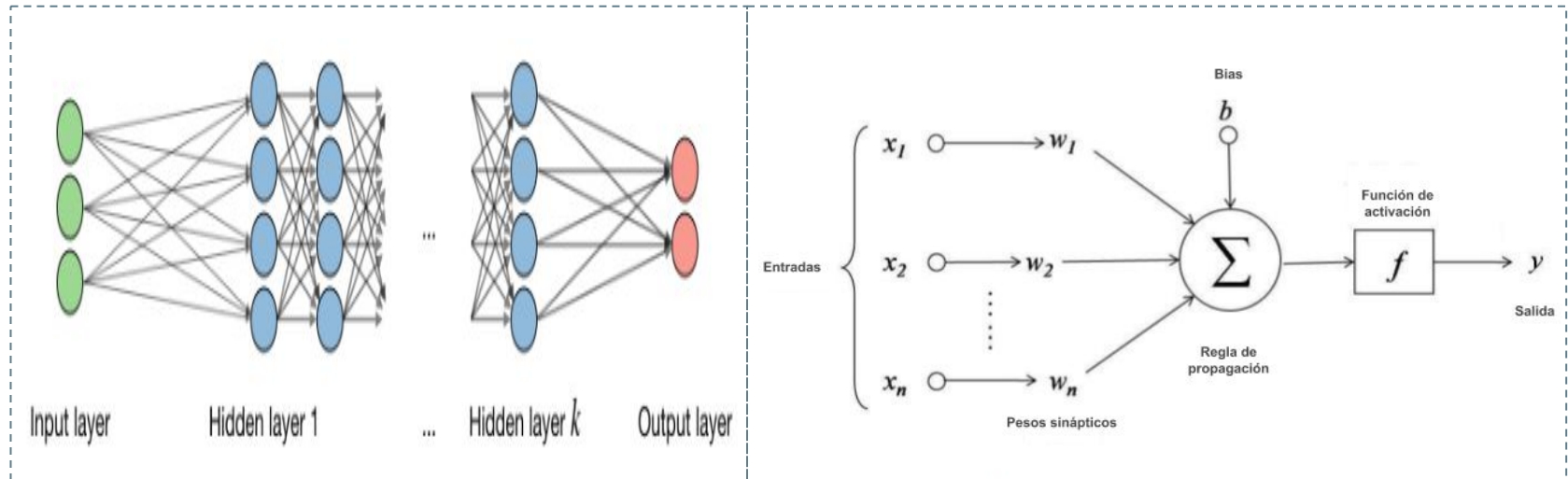
5.

Modelo de detección de anomalías

Fundamentos teóricos, generación del modelo de comportamiento normal y generación del método de detección de valores atípicos

Redes neuronales

La Red Neuronal Artificial (ANN), es un modelo computacional basado en la estructura y funciones de las redes neuronales biológicas. Es como un sistema nervioso humano artificial para recibir, procesar y transmitir información. ANN es un conjunto de neuronas conectadas organizadas en capas: **entrada**, **oculta** y de **salida**.

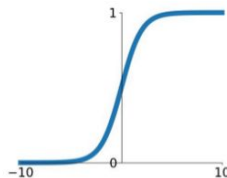


Tipos de funciones de activación

Activation Functions

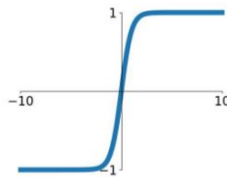
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



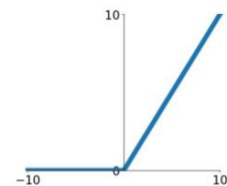
tanh

$$\tanh(x)$$



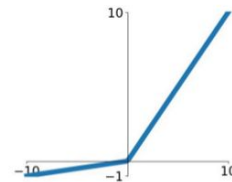
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

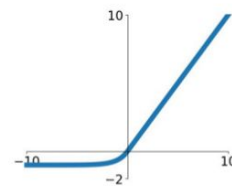


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

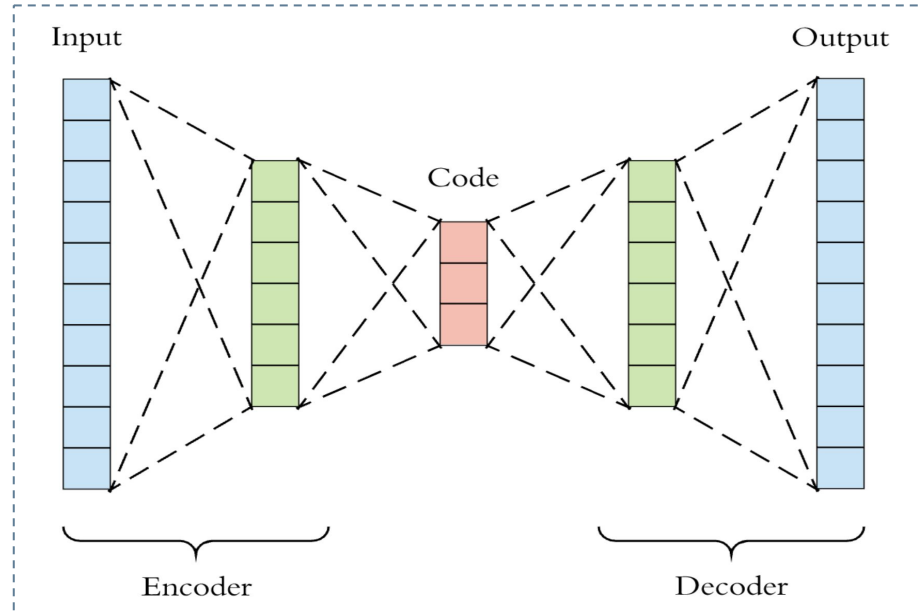
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Autoencoder

Se utiliza para reducir la cantidad de variables aleatorias consideradas, de modo que el sistema puede aprender una representación para un conjunto de datos y, por lo tanto, procesar modelos de datos generativos.



Isolation Forest

Es un algoritmo de detección de valores atípicos no supervisado que aprovecha el hecho de que los valores atípicos son "**pocos y diferentes**", lo que significa que son menos numerosos y tienen valores de características inusuales en comparación con la clase normal.

Isolation Forest

Es un algoritmo de detección de valores atípicos no supervisado que aprovecha el hecho de que los valores atípicos son "pocos y diferentes", lo que significa que son menos numerosos y tienen valores de características inusuales en comparación con la clase normal.



Aislando
una clase
normal

Isolation Forest

Es un algoritmo de detección de valores atípicos no supervisado que aprovecha el hecho de que los valores atípicos son "pocos y diferentes", lo que significa que son menos numerosos y tienen valores de características inusuales en comparación con la clase normal.



Aislando
una clase
anormal

Métricas de evaluación

MATRIZ DE CONFUSION

		Predicción	
		Clase Positiva	Clase Negativa
Reales	Clase Positiva	Verdadero Positivo (VP)	Falso Negativo (FN)
	Clase Negativa	Falso Positivo (FP)	Verdadero Negativo (VN)

CUADRO 3.1: Matriz de confusión, para una clasificación binaria (Elaboración propia).

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

$$\text{Precision} = \frac{VP}{VP + FP}$$

$$\text{Sensibilidad} = \frac{VP}{FN + VP}$$

$$\text{Especificidad} = \frac{VN}{FP + VN}$$

Modelo de comportamiento normal

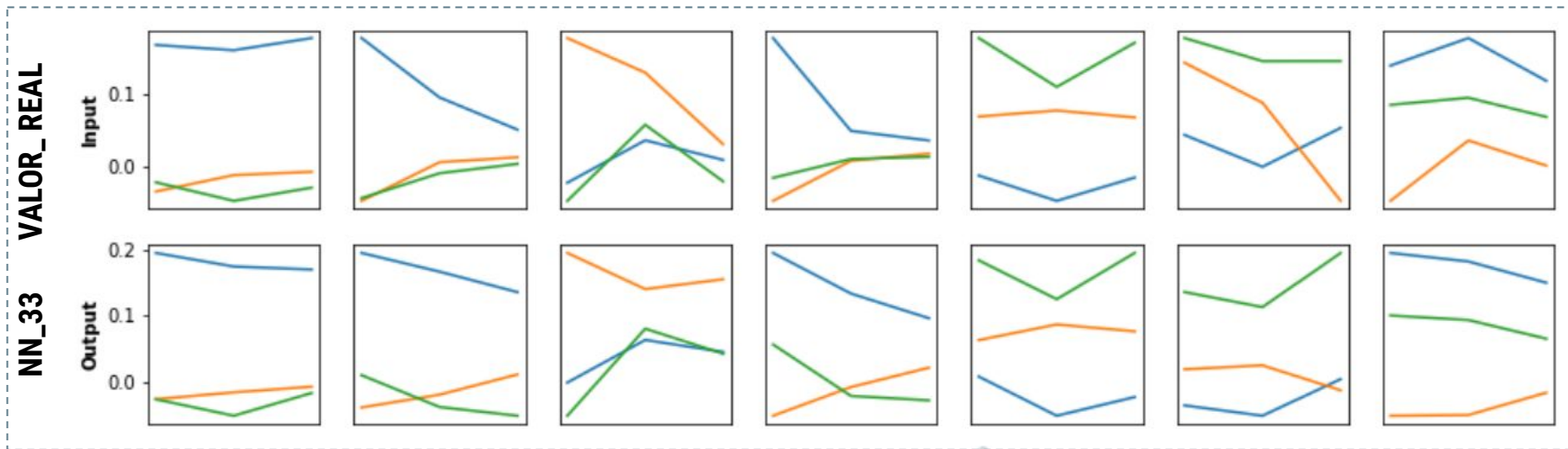
Esta etapa es una de las partes más importantes de éste trabajo, debido a que el rendimiento del modelo de detección de anomalías depende en gran parte de la precisión de esta etapa.

Arquitectura
del modelo

Arquitectura Densa			
NN_33			
Tipo	Salida	Activación	# Parámetros
Input	(3,3)		0
Flatten	9		0
Dense	8	elu	80
Dense	5	elu	45
Dense	8	elu	88
Dense	9	tanh	81
Reshape	(3,3)		0

Resultado del modelo de comportamiento normal

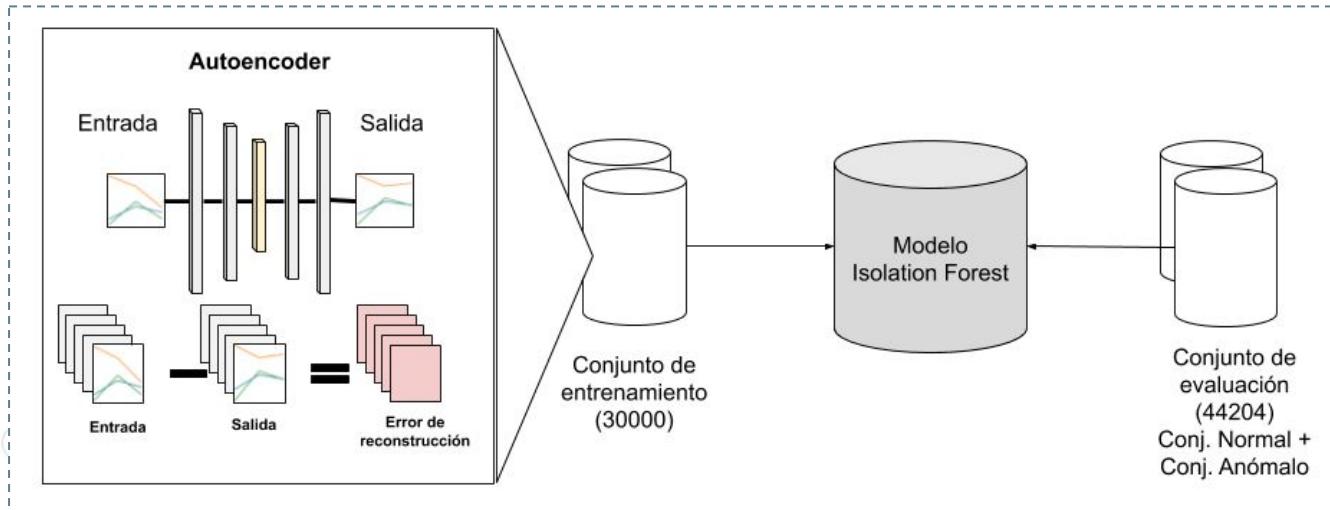
Precision	Tiempo de ejecución
0.90007407	26us/step



Método de detección de anomalías

Tipo de anomalía	No. anomalías	No. datos
Giros en Zig Zag	5	105
Giros a la derecha e izquierda a alta velocidad	7	39
Frenos en seco	6	24

CUADRO 5.1: Tabla del conjunto de anomalías (Elaboración propia).

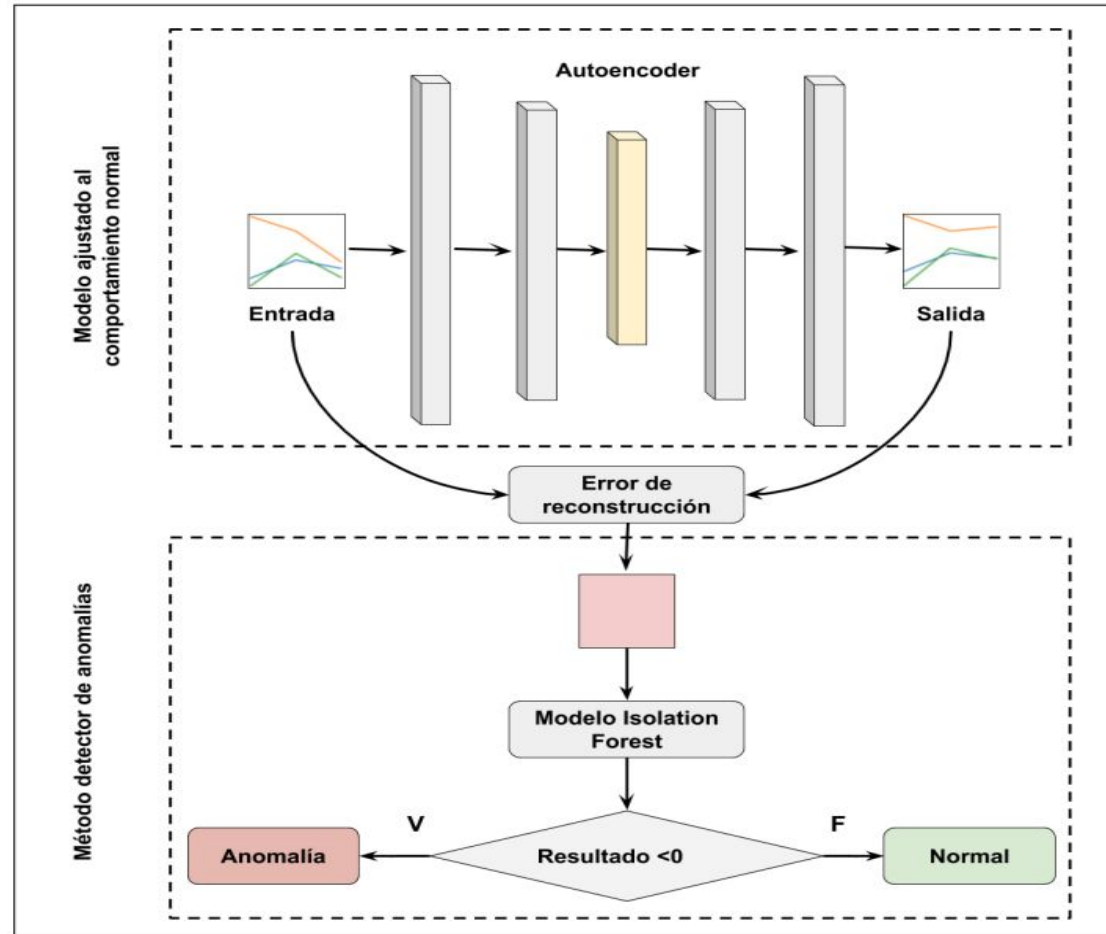


Resultado del metodo de deteccion de anomalias

Contaminación (C)	VP	VN	FN	FP	Sensibilidad	Especificidad
0.0025	103	43898	61	142	0.6280	0.9968
0.0050	108	43792	56	248	0.6585	0.9944
0.0075	111	43688	53	352	0.6768	0.9920

CUADRO 5.9: Evaluación de la detección de anomalías usando Isolation forest para errores de reconstrucción (Elaboración propia).

Mecanismo de detección de anomalías



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

6.

Resultados y evaluación

Resultado general

		Predicción	
		Anomalía	Clase Normal
Reales	Anomalía	111	53
	Clase Normal	352	43688

CUADRO 6.1: Matriz de confusión, para el mecanismo de detección de anomalías (Elaboración propia).

Detección de anomalías del tipo zig zag

Giros en Zig Zag		
VP	FN	Total
69	36	105

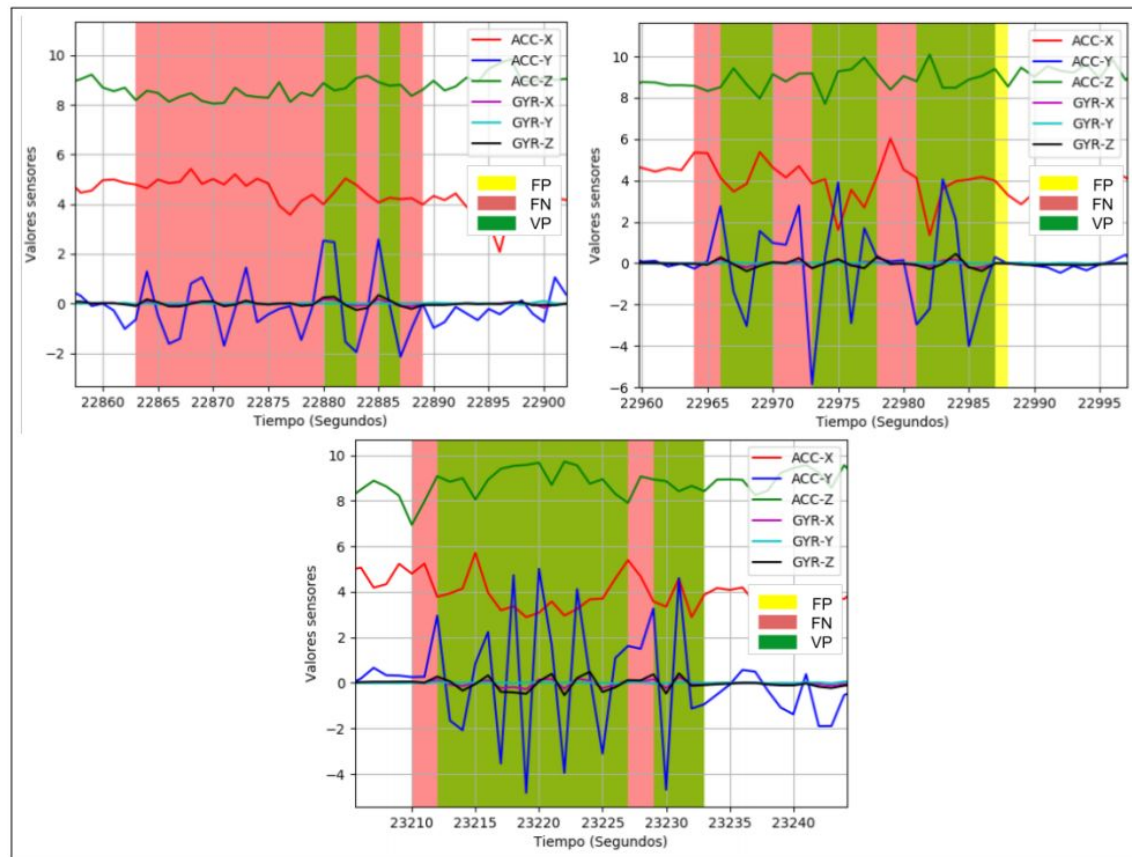


FIGURA 6.1: Resultados de la detección de anomalías del tipo zig zag (Elaboración propia).

Detección de anomalías del tipo giros a alta velocidad

Giros a alta velocidad		
VP	FN	Total
23	12	35

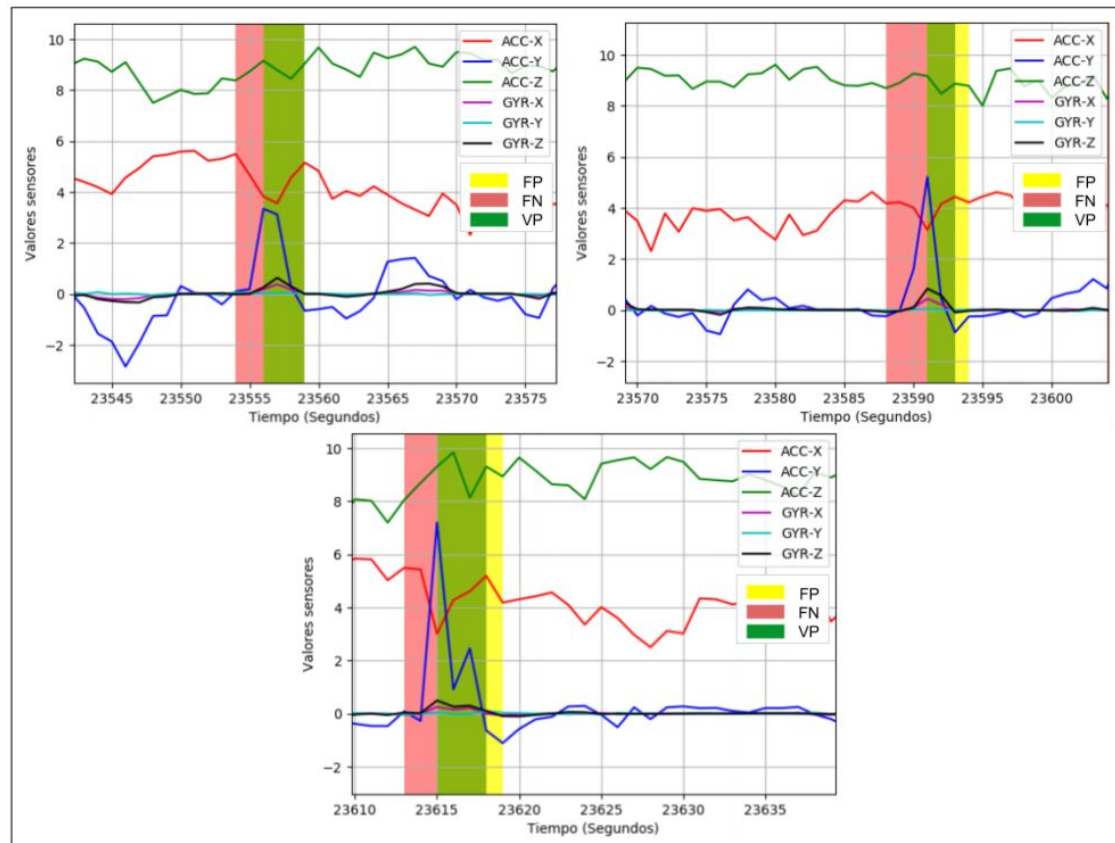


FIGURA 6.2: Resultados de la detección de anomalías del tipo giros a alta velocidad (Elaboración propia).

Detección de anomalías del tipo frenos en seco

Frenos en seco		
VP	FN	Total
19	5	24

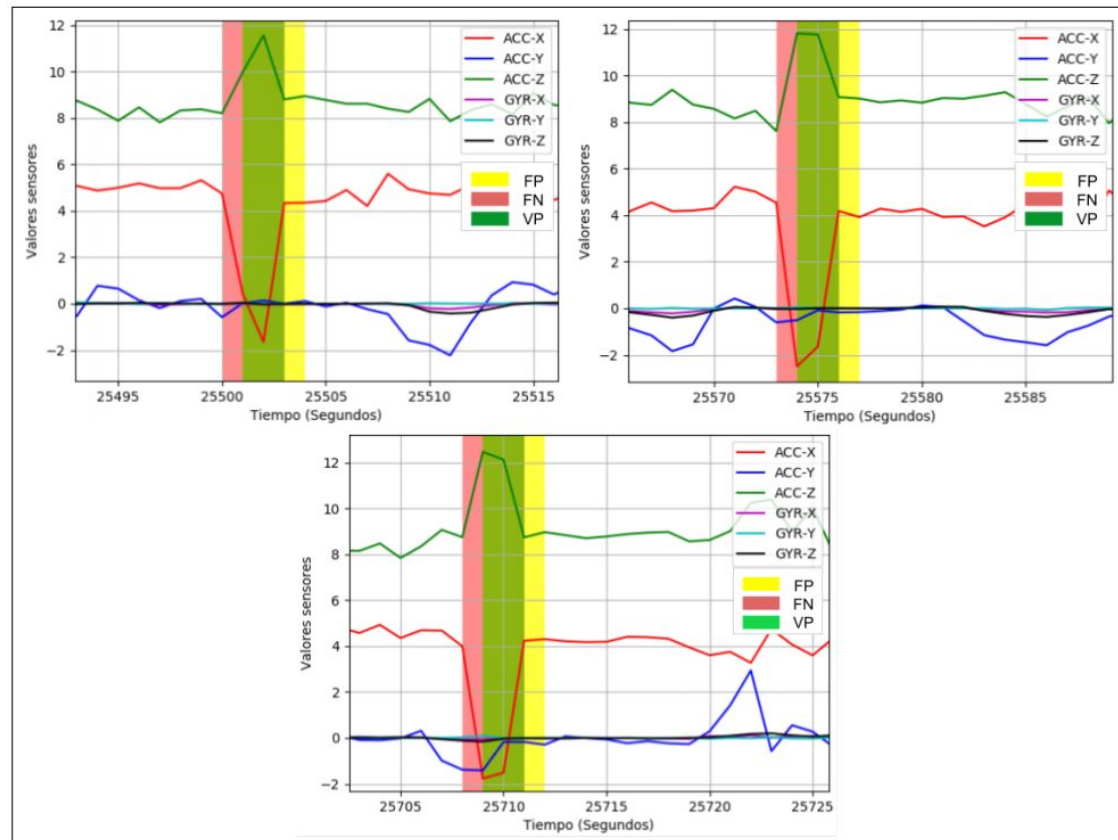


FIGURA 6.3: Resultados de la detección de anomalías del tipo frenos en seco (Elaboración propia).

Detección de falsos positivos

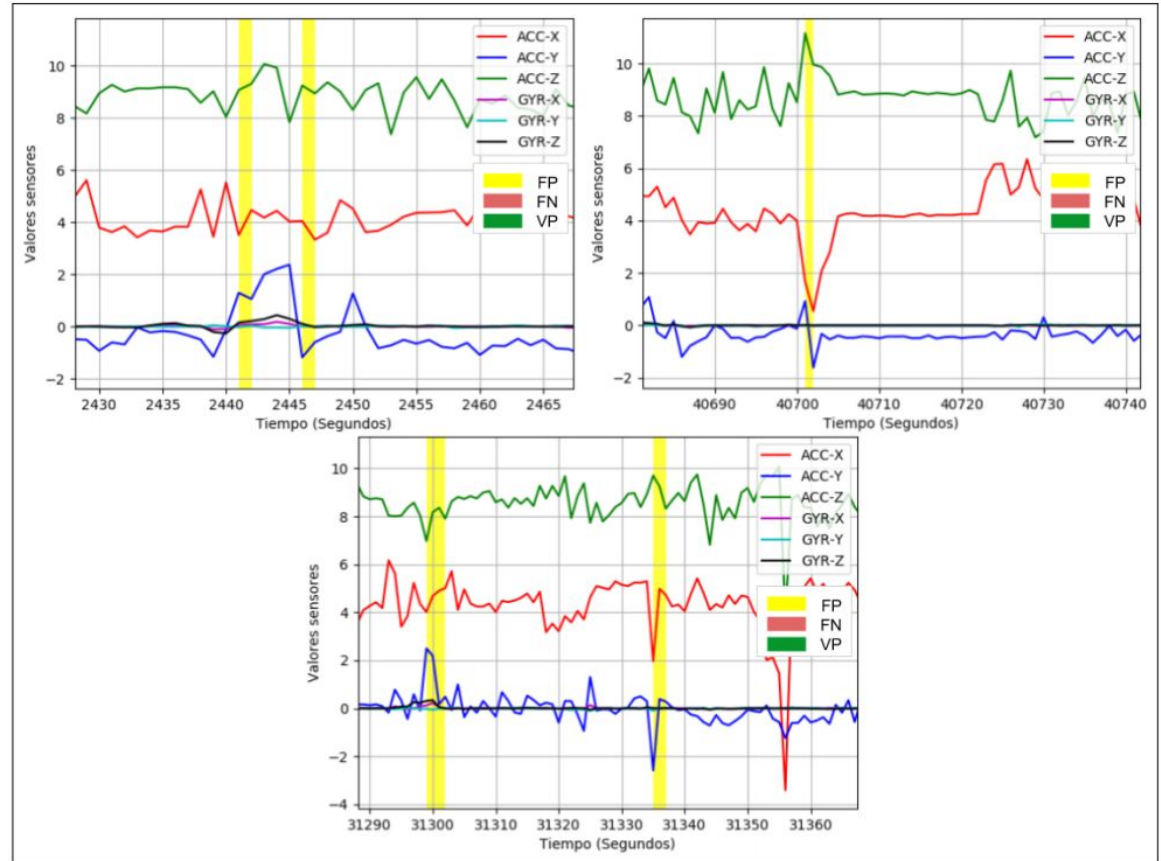


FIGURA 6.4: Resultados de la detección de falsos positivos (Elaboración propia).

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

7.

Conclusiones y trabajos futuros



Conclusiones

- © Se comprueba, la capacidad con la que cuentan los sensores inerciales de un dispositivo móvil para representar correctamente el movimiento de un automóvil y de esa manera ser capaz de alimentar, con un previo preprocesamiento, un mecanismo de detección de anomalías.
- © La red más simple logró los mejores resultados tanto en precisión como en tiempo empleado durante el proceso de predicción; para generar el modelo del comportamiento normal.
- © Los bosques de aislamiento fueron la mejor técnica para definir un método de detección de anomalías, debido a su simplicidad el entrenamiento y a su robusto resultado.
- © Integrando el modelo del comportamiento normal y el método de detección de anomalías, los cuales sólo fueron entrenados con el conjunto de datos "normal", se logra la creación de un mecanismo capaz de identificar valores atípicos de la conducción de cada agente.
- © Los resultados del conjunto de evaluación el cual presenta muestras anómalas, dio como resultado la correcta detección del 67.68 % de las muestras, que presentan anomalías en el conjunto de evaluación, así como también presenta una tasa de tan sólo 0.80 % de muestras normales detectadas como anomalías. Siendo un gran avance en el ámbito de la detección de anomalías con un enfoque semi-supervisado.



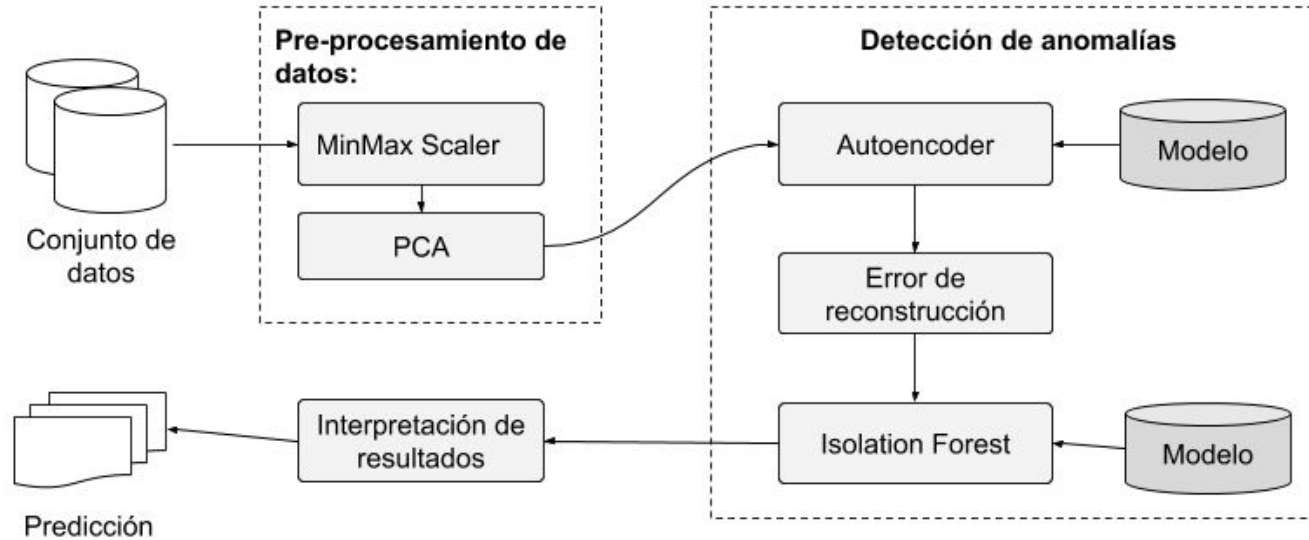
Trabajos futuros

- © Agregar la velocidad del vehículo como un nuevo parámetro del conjunto de datos, debido a que esto podría brindar un mejor entendimiento del comportamiento normal de conducción, así como también de las anomalías.
- © En lugar de trabajar con los datos en crudo, usar la diferencia entre un dato capturado en el tiempo t y un dato capturado en $t - 1$ ($\text{diff} = \text{datot} - \text{datot}-1$), la aplicación de éste pre-procesamiento de datos podría maximizar la detección de aquellas anomalías que presentan elevadas diferencias entre los datos consecutivos.
- © Validar el modelo con nuevos tipos de anomalías como por ejemplo: derrapes, choques, giros en U a alta velocidad, entre otros. Esto debido a que el proyecto se limitó al reconocimiento de sólo tres tipos de anomalías por la dificultad y peligro que conlleva su captura.
- © Extender el modelo para que sea capaz de determinar no sólo una anomalía sino también el tipo al que dicha anomalía pertenece.
- © Migrar el modelo del comportamiento normal de keras a Tensorflow 2.0.
- © Implementar un sistema de información para monitorear las anomalías mediante el mecanismo de detección propuesto en el presente trabajo.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels or types of nodes. The lines are thin and gray, connecting the nodes in a non-linear fashion.

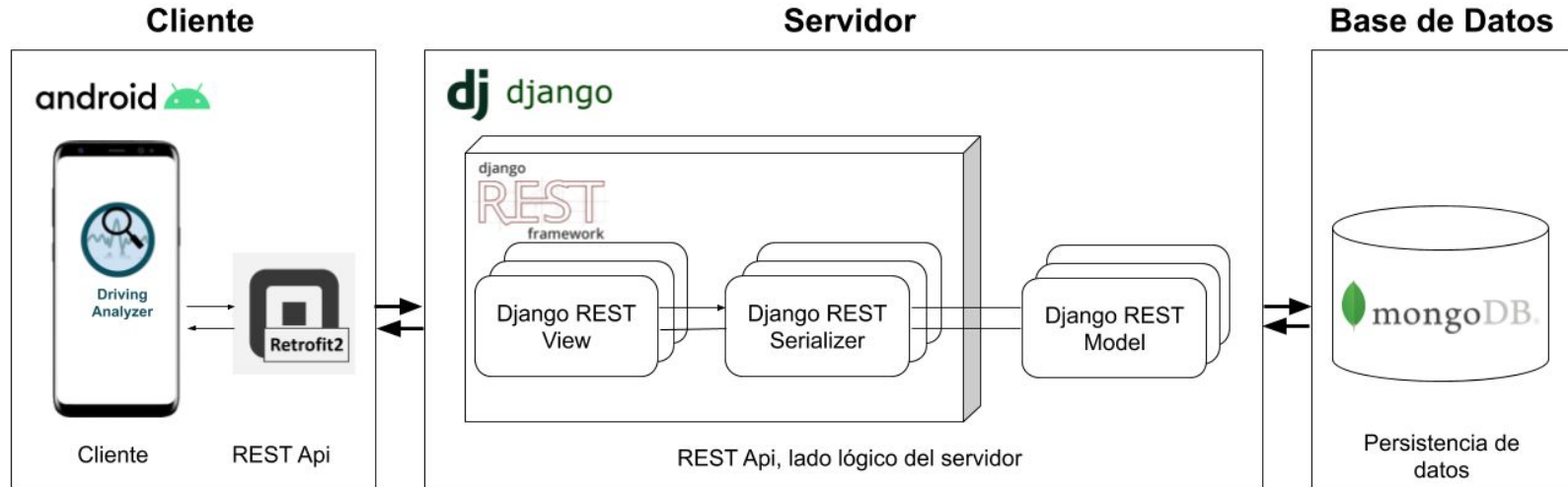
7. **Demostracion**

Arquitectura Logica



La arquitectura lógica está conformada por dos principales partes del mecanismo de detección propuesto: el pre-procesamiento y la detección de anomalías, como se puede observar en la Figura.

Arquitectura Física



La arquitectura física del prototipo cuenta con tres partes principales: el cliente, el servidor y la base de datos, como se puede observar en la gráfica.


Want big
impact?

Use big image.



The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic-looking structure that fills the entire background.

Gracias...!



https://es.slideshare.net/josem.alvarez/master-thesis-slides?qid=8fba12a6-36e9-4240-a13f-7e16f74680c2&v=&b=&from_search=13

slides

https://es.slideshare.net/llformiga/llus-formiga-phd-thesis?qid=8fba12a6-36e9-4240-a13f-7e16f74680c2&v=&b=&from_search=54

https://es.slideshare.net/rubIQ/vxc-computer-vision-presentation?qid=8fba12a6-36e9-4240-a13f-7e16f74680c2&v=&b=&from_search=69

https://es.slideshare.net/azubiaga/master-thesis-presentation-8444195?qid=8fba12a6-36e9-4240-a13f-7e16f74680c2&v=&b=&from_search=5

Slides tesis Ciencias de la comp

http://www.maia.ub.es/~sergio/soluciones2_006.htm



Para editar fotos

<https://pixlr.com/e>

Isolation Forest

