UNIVERSIDAD NACIONAL DEL ALTIPLANO

FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

ANÁLISIS DEL MODELO DE PREDICCIÓN EN LA FATALIDAD DE ACCIDENTES DE TRÁNSITO EN LA REGIÓN DE PUNO, 2022

TESIS

PRESENTADA POR:

MARIZOL LIZBETH SERRANO QUISPE

PARA OPTAR AL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2024

Reporte de similitud

NOMBRE DEL TRABAJO

ANÁLISIS DEL MODELO DE PREDICCIÓN EN LA FATALIDAD DE ACCIDENTES DE T RÁNSITO EN LA REGIÓN DE PUNO, 2022

AUTOR

MARIZOL LIZBETH SERRANO QUISPE

RECUENTO DE PALABRAS

22537 Words

RECUENTO DE CARACTERES

120111 Characters

RECUENTO DE PÁGINAS

110 Pages

TAMAÑO DEL ARCHIVO

1.9MB

FECHA DE ENTREGA

Jul 24, 2024 6:37 AM GMT-5

FECHA DEL INFORME

Jul 24, 2024 6:39 AM GMT-5

18% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 16% Base de datos de Internet
- 1% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 11% Base de datos de trabajos entregados

Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 10 palabras)
- Bloques de texto excluidos manualmente

D.Sc. Angel Javier Quispe Carita
ING. ESTADÍSTICO E INFORMÁTICO
CIP. N° 104299

Juan Carlos Juarez Vargas
Ing. Estadístico e Informático
CIP. 77859

Resumen

DEDICATORIA

Dedico esta tesis a mis padres, Jhon Alex Serrano Vargas y Sonia Quispe Canaviri, quienes son el pilar más fundamental en mi vida. Agradezco su amor, paciencia y esfuerzo, que han hecho posible que hoy pueda cumplir otro sueño y por enseñarme que, a pesar de las caídas, siempre puedo levantarme y seguir adelante.

Gracias también a mis hermanas Alejandra y Melisa por su constante apoyo, y a toda mi familia por estar siempre a mi lado. Sus oraciones, consejos y palabras de aliento me han hecho mejor persona y me han ayudado a lograr todos mis sueños y metas.

Marizol Lizbeth Serrano Quispe

AGRADECIMIENTOS

Agradezco a Dios por concederme sabiduría, salud y fortaleza, lo que hizo posible alcanzar este logro y me guío en el camino correcto para seguir adelante.

A la Universidad Nacional del Altiplano, por ofrecerme la oportunidad de formarme profesionalmente con sus conocimientos y enseñanzas.

A la Escuela Profesional de Ingeniería Estadística e Informática y a los docentes de la Facultad de Ingeniería Estadística e Informática por compartir sus conocimientos durante mi preparación profesional

A mis Padres, hermanas, amigos y compañeros por su apoyo constante.

También quiero expresar un agradecimiento especial a los jurados Dr. Bernabé Canqui Flores, M.Sc. Roberto Elvis Roque Claros, M.Sc. Alcides Ramos Calcina y en especial a mi director de tesis D.Sc. Angel Javier Quispe Carita, cuya orientación y respaldo fueron esenciales para terminar esta tesis.

Marizol Lizbeth Serrano Quispe

ÍNDICE GENERAL

Pág.

Área: Estadística

Tema: Regresión Logística

FECHA SUSTENTACIÓN: 1 de agosto 2024

ÍNDICE DE TABLAS

Pág.

ÍNDICE DE FIGURAS

Pág.

ÍNDICE DE ANEXOS

Pág.

ACRÓNIMOS

INEI:	Instituto Nacional de Estadística e Informática
OMS:	Organización Mundial de la Salud
PBI:	Producto Interno Bruto
MTC:	Ministerio de Transportes y Comunicaciones
OPS:	Organización Mundial de la Salud
SIDPOL:	Sistema de Denuncia Policiales
ONSV:	Observatorio Nacional de Seguridad Vial
PNP:	Policía Nacional del Perú
MRL:	Modelo de Regresión Logística

RESUMEN

Los modelos de predicción se han consolidado como herramientas valiosas en la gestión y evaluación de riesgos en el ámbito de la seguridad vial, abordando un desafío crucial: reducir el número de accidentes de tránsito, que son una de las principales causas de muerte en todo el mundo. El estudio tiene como objetivo determinar el modelo predictivo de regresión logística más óptimo para predecir la probabilidad de fatalidad en accidentes de tránsito en la región de Puno durante el año 2022. Se empleó un método de investigación hipotético deductivo, con un diseño no experimental. Para ello, se utilizó la recopilación retrospectiva de datos, procedentes de la base de datos de la X-MACREPOL-PUNO. El modelo de regresión logística se desarrolló utilizando el 70% de los datos para entrenamiento y el 30% para validación. De acuerdo a los resultados obtenidos el modelo clasificador es el siguiente: $Ln(odds) = -4.3398 - 0.9817x_1 - 1.0834x_2 + 0.8289x_3 + 1.6779x_4 - 0.836x_5 + 0.3876x_7 + 1.2369x_8 + 2.1454x_9$ El modelo presenta un rendimiento sobresaliente con un accuracy del 95%, para accidentes fatales, la precisión es del 86% y un recall del 89%, mientras que, para accidentes no fatales, la precisión alcanza 97% y el recall el 96%. La métrica F1 es de los 87% para accidentes fatales y del 97% para accidentes no fatales. El modelo se validó bajo los supuestos de linealidad, cumpliendo adecuadamente este supuesto, La prueba de independencia de Durbin-Watson dio como resultado un valor de 1.993, lo cual indica que no hay una correlación significativa entre los residuos. Además, el índice de multicolinealidad es 2.95, sugiere que no existen problemas de multicolinealidad.

Palabras clave: Accidentes fatales, Métricas, Modelo de Regresión Logística.

ABSTRACT

Prediction models have become valuable tools in risk management and evaluation in the field of road safety, addressing a crucial challenge: reducing the number of traffic accidents, which are one of the leading causes of death worldwide. The study aims to determine the most optimal logistic regression predictive model to predict the probability of fatalities in traffic accidents in the Puno region during the year 2022. A hypothetical-deductive research method with a non-experimental design was employed. Retrospective data collection was used, sourced from the X-MACREPOL-PUNO database. The logistic regression model was developed using 70% of the data for training and 30% for validation. According to the results obtained, the classifier model is as follows: $Ln(odds) = -4.3398 - 0.9817x_1 - 1.0834x_2 + 0.8289x_3 + 1.6779x_4 - 0.836x_5 + 0.3876x_7 + 1.2369x_8 + 2.1454x_9$ The model exhibits outstanding performance with an accuracy of 95%. For fatal accidents, the precision is 86% and the recall is 89%, while for non-fatal accidents, the precision reaches 97% and the recall is 96%. The F1 metric is 87% for fatal accidents and 97% for non-fatal accidents. The model was validated under the assumption of linearity, adequately fulfilling this assumption. The Durbin-Watson independence test resulted in a value of 1.993, indicating no significant correlation between residuals. Additionally, the multicollinearity index is 2.95, suggesting no multicollinearity issues.

Keywords: Fatal Accidents, Metrics, Logistic Regression Model.

CAPÍTULO I

INTRODUCCIÓN

Según la OMS (2018), los accidentes de tránsitos representan la principal causa de muerte y lesiones incapacitantes en niños y jóvenes adultos, particularmente en los grupos de edad de 5 a 29 años. A nivel global, constituyen la octava causa de mortalidad, generando alrededor de 1.35 millones de decesos anuales. Además, se destaca que los accidentes automovilísticos ocupan el segundo lugar como principal causa de muerte en la población mundial, siendo responsables de aproximadamente 1.15 millones de defunciones anuales.

En Perú, según el Observatorio Nacional de Seguridad Vial (ONSV, 2022) la situación es igualmente alarmante. En el año 2022 se registraron 83,897 accidentes de tránsito, lo que provocó la muerte de 3,328 personas y las lesiones de 53,552 otras. Más del 50% de estos accidentes se atribuye a la imprudencia del conductor y al exceso de velocidad.

La necesidad de entender y predecir estos eventos se ha vuelto crucial para implementar medidas preventivas efectivas. En este contexto, la regresión logística emerge como herramienta poderosa para analizar y predecir la probabilidad de eventos binarios, como la fatalidad en accidentes de tránsito. Esta técnica es capaz de manejar variables categóricas como continuas y ha demostrado ser eficaz en identificar variables significativas que influyen en la severidad de los accidentes, (Chen & Chen, 2020).

La presente investigación busca determinar el modelo predictivo de regresión logística más adecuado para explicar y predecir la fatalidad de accidentes de tránsito en

la región de puno durante el año 2022, proporcionando así una herramienta importante para la toma de decisiones y la ejecución de medidas preventivas.

La estructura del estudio se presenta de la siguiente manera: en el primer capítulo se aborda el planteamiento del problema de investigación, los objetivos planteados, formulación de hipótesis y la justificación; en el segundo capítulo se lleva a cabo la revisión de la literatura; el tercer capítulo explica la metodología del estudio, describiendo el tipo y diseño de la investigación, los métodos utilizados y el análisis de la información. El cuarto capítulo analiza los resultados obtenidos a través de las estimaciones del modelo, que cumplen con los objetivos propuestos. El quinto capítulo presenta los hallazgos de la investigación, el sexto capítulo hace sugerencias pertinentes y el séptimo capítulo detalla las referencias bibliográficas utilizadas.

1.1. PLANTEAMIENTO DEL PROBLEMA

Según la Organización mundial de la salud (2023), se estima que cerca de 1,19 millones de personas mueren por accidentes de tránsito cada año, lo que significa que una persona muere cada dos minutos. Además, hay alrededor de 20 a 50 millones de individuos que experimentan lesiones no mortales, algunas de las cuales pueden provocar discapacidades permanentes.

Según un informe de la Defensoría del Pueblo, a pesar de las restricciones e inmovilizaciones implementadas para controlar la propagación del COVID-19 en 2020 y 2021, las cifras de accidentes de tránsito en Perú han disminuido. Sin embargo, en 2022, la cantidad de accidentes de tránsito volvió a los niveles registrados en 2019. En 2020 se registró una de las cifras más bajas con 57,396 accidentes; luego hubo un pequeño aumento en 2021, con 74.624 accidentes. Sin embargo, durante los primeros seis meses de 2022, se registraron más de 41,000 accidentes de tránsito, casi la mitad de los casos

reportados en 2018 y 2019, cuando se superaron los 90,000 casos en ambos años. En el primer semestre de 2022, la cifra de personas fallecidas en accidentes de tránsito fue de 1,573, mientras que el número de personas heridas representó casi el 40% de los casos reportados en 2019.A estas estadísticas se debe sumar la información de los últimos 25 años, en los que se han registrado más de 2,08 millones de accidentes de tránsito con más de 1,2 millones de fallecidos (Defensoría del Pueblo, 2022).

Según el Ministerio de transportes y comunicaciones (2022), en su informe anual "Informe de siniestralidad de tránsito fatal – Región Puno", en el año 2021 los accidentes de tránsito representaron una preocupación significativa en Puno, con tasas de mortalidad superiores a la media nacional de Perú. Ese problema fue particularmente grave debido a las características demográficas, geográficas y de infraestructura de la región. Del análisis de vulnerabilidad de los usuarios, se observó que los conductores son los más afectados, representando el 42.7% de las muertes. En términos de edad, los adultos de 26 a 44 años son el grupo etario más impactado. Además, el 73.3% de las personas fallecidas son hombres. Geográficamente, la mayoría de los accidentes de tránsito fatales ocurren en zonas rurales (85%) y en carreteras (84%). En cuanto a las clases y causas de los siniestros, los choques son la clase de siniestro más común (39.7%) e imprudencia del conductor es la causa principal de los accidentes (68.7%), con el exceso de velocidad siendo el factor específico más frecuente (32.4%).

La importancia de crear estrategias efectivas para disminuir la frecuencia de accidentes de tránsito en Puno se destaca en estos datos. Como persona que viaja constantemente, enfrento la incertidumbre de transitar por carreteras peligrosas, con el constante temor de sufrir un accidente. En este contexto, la estadística puede ser una herramienta poderosa para abordar este problema, ofreciendo modelos de predicción que nos permitan anticipar futuros accidentes de tránsito. Al analizar datos históricos de

accidentes, es posible identificar patrones y factores de riesgo que contribuyen a la ocurrencia de estos eventos. La regresión logística, en particular, permitirá calcular la probabilidad de ocurrencia de un accidente en función de diversas variables explicativas, como el tipo de accidente, las causas, vehículos involucrados, el lugar y la hora del accidente, y las condiciones del conductor.

1.2. FORMULACIÓN DEL PROBLEMA

1.2.1. Problema general

¿Cuál es el modelo predictivo de regresión logística más óptimo para predecir la probabilidad de fatalidad en accidentes de tránsito en la región de Puno durante el año 2022?

1.3. OBJETIVOS

1.3.1. Objetivo General

Determinar el modelo predictivo de regresión logística más óptimo para predecir la probabilidad de fatalidad en accidentes de tránsito en la región de Puno durante el año 2022.

1.3.2. Objetivos Específicos

- Seleccionar las variables más significativas que explican la fatalidad en accidentes tránsitos en la región de Puno durante el año 2022.
- Estimar y validar el modelo predictivo de accidentes de tránsito en la región de Puno, para comprender la fatalidad de los accidentes.

1.4. HIPÓTESIS DE LA INVESTIGACIÓN

1.4.1. Hipótesis General

El modelo de regresión logística predice la probabilidad de fatalidad en accidentes de tránsito en la región de Puno durante el año 2022.

1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN

Los accidentes de tránsito, especialmente aquellos con víctimas mortales, ocurren con frecuencia en la región Puno y representan un gran desafío para las autoridades, comunidades y personas. Los accidentes de tránsito no sólo causan pérdidas de vidas, sino también altos costos médicos, pérdida de productividad y daños a la propiedad. Las familias afectadas enfrentan no sólo la pérdida emocional sino también el impacto financiero de no poder trabajar y los costos médicos. En este sentido, la regresión logística y otras técnicas de análisis estadístico se convierten en poderosas herramientas para identificar las variables más influyentes y predecir la probabilidad de accidentes fatales, y los resultados de este estudio ayudarán a mejorar la seguridad vial regional. Esto puede incluir mejorar la infraestructura vial, realizar campañas de educación y concientización sobre el tráfico vial e introducir tecnología avanzada de monitoreo y control del tráfico. Desde una perspectiva académica, este estudio contribuye al conjunto de conocimientos existentes sobre predicción de accidentes de tránsito. La aplicación del modelo de regresión logística a datos de accidentes específicos de la región de Puno brinda la oportunidad de validar y comparar el método en el contexto local. Además, proporciona un marco para futuras investigaciones y desarrollo en esta área. El modelo predictivo desarrollado en este estudio no sólo es aplicable a la región de Puno, sino que también puede adaptarse y utilizarse en otras regiones con características similares.

CAPÍTULO II

REVISIÓN LITERATURA

2.1. ANTECEDENTES DE INVESTIGACIÓN

2.1.1. Antecedentes Internacionales

Monroy y Díaz (2020), examinaron el modelo de pronóstico de la gravedad de los accidentes de tráfico en Bogotá empleando un modelo de regresión logística para identificar las variables fundamentales. Los resultados revelaron que alrededor del 52% de las personas fallecidas en estos accidentes son peatones, con un aproximado del 23% correspondiente a mujeres, y la mayoría de los accidentes mortales ocurren entre los 20 y 30 años. En la clasificación de incidentes, se destaca que cuatro de siete grupos contienen más del 50% de víctimas que son peatones, y se observa una prevalencia similar en la categoría de atropellos. Los grupos 4 y 5 están principalmente compuestos por hombres, en un 96% y 84% respectivamente, y en su mayoría son motociclistas (67% y 44%, respectivamente). Se identifica que la probabilidad de fallecer en un atropello es apenas un 4% mayor que en una colisión, pero esta cifra aumenta a 2.20 veces cuando se trata de peatones. Además, los hombres tienen una probabilidad 1.79 veces mayor de morir que las mujeres.

Cruz (2017) creo modelos de predicción destinados a estimar la probabilidad de que ocurran accidentes de tráfico en áreas específicas de Madrid, España; empleando redes neuronales en el lenguaje de programación Python. Los resultados revelaron que la probabilidad de que un accidente respondió a la combinación de factores del histórico de accidentes de tráfico y las variables

predictoras seleccionadas; por ejemplo, si en un día de la semana se le asigna un valor de 1, en una franja horaria de 0.25, en un trimestre de 0.5, no es un día festivo (valor 0), la velocidad media del viento es de 0.2, la temperatura media es 0.3, y las precipitaciones tienen un valor de 0.1, y en una de estas situaciones ocurre un accidente en la Calle M-30, carril 1, y en el kilómetro correspondiente a la zona 2, mientras que en los otros tres casos no se registra ningún accidente en esa área, la probabilidad de que ocurra un accidente en esa combinación específica de variables es del 25% (1 de cada 4 casos).

Por otra parte, Aguirre y Balarezo (2022) llevaron a cabo un análisis de los accidentes vehiculares utilizando técnicas de minería de datos para predecir la mortalidad de los accidentes de tránsito en el Distrito Metropolitano de Quito durante el período 2015-2019; para lo cual emplearon el Modelo Autorregresivo Integrado a las Medias Móviles (ARIMA), BoxJenkins y Vectores Autorregresivos (VAR); en el caso del modelo ARIMA, identificaron que en los años 2020-2021, hubo una disminución leve en la cantidad de accidentes; por otro lado, en el modelo VAR, se encontró que la relación entre los accidentes sin señalización y la influencia del alcohol en el mes siguiente mejoró, ya que a medida que aumentaban los accidentes en un mes, aumentaban los accidentes en el mes siguiente. Específicamente, observaron que cuando ocurrían accidentes en un mes, la cantidad de accidentes tendía a aumentar en el mes siguiente, especialmente aquellos relacionados con la falta de señalización en el área del accidente, así como en casos que involucraban sustancias alcohólicas, estupefacientes o psicotrópicas.

Asimismo, García (2021) empleo modelos basados en redes bayesianas y modelos multivariados basados en GLM para Bogotá, con énfasis en analizar y

evaluar la capacidad predictiva de diferentes modelos para estimar la cantidad de accidentes, incluyendo el sistema integrado de transporte público (SITP) y sus carriles prioritarios. Se comparan dos enfoques: el modelo lineal generalizado multivariado (GLM) y el modelo de red bayesiana probabilística (PBN), con el objetivo de determinar cuál ofrece un mejor rendimiento en la predicción de la cantidad de accidentes en la ciudad. Este análisis permite identificar los factores que tienen un impacto significativo en la aparición de accidentes de tráfico en la ciudad y proporciona una función de calibración precisa que puede ser empleada para estimar y anticipar el número de accidentes en la infraestructura vial de Bogotá.

2.1.2. Antecedentes Nacional

Pérez (2018) ajusto y valido un modelo de predicción de accidentes en el contexto específico de Perú, donde su precisión la justificó en aplicar la metodología del Manual de Seguridad Vial de Estados Unidos de 2010, y el uso de un indicador denominado "factor de calibración" para medir la cercanía de las predicciones a los valores reales observados en el terreno. Los resultados obtenidos son altamente satisfactorios, con un factor de calibración promedio de 1.06, lo que indica que el modelo realiza predicciones con errores pequeños a pesar de algunas limitaciones. En consecuencia, se concluye que es factible aplicar la metodología estadounidense en el contexto peruano, lo que ha llevado a una reducción significativa de accidentes, en torno al 40%.

Sin embargo, Estrada & Soto (2021) se enfocó en evaluar la seguridad vial en la Avenida Atahualpa, correspondiente a Cajamarca y Baños del Inca aplicando metodologías de inspección de seguridad vial del Ministerio de Transportes y

Comunicaciones (MTC) y la metodología HSM-2010. Los datos determinaron que la vía analizada presenta elementos de inseguridad que contribuyen en un 45% a la ocurrencia de accidentes de tránsito; al emplear la metodología HSM-2010, se determinó que la frecuencia promedio de accidentes en la vía corresponde a 17 accidentes al año. Luego, se implementaron propuestas de mejora que lograron reducir los accidentes de tránsito en un impresionante 78%. Estos resultados respaldan la hipótesis planteada inicialmente, que sugería deficiencias en la seguridad vial de la vía.

Chipana (2023), analizo y determino los factores que influyen en los accidentes de tránsito generados por el transporte público terrestre en Villa El Salvador, 2021. En los resultados encontró que la mayoría de los casos, el factor humano, que se refiere al conductor del vehículo, desempeñando el papel más significativo y es responsable de aproximadamente el 68% de los accidentes, debido a diversos motivos y circunstancias. El otro factor clave es el vehículo en sí, es decir, la unidad móvil. En este sentido, se han llevado a cabo investigaciones con diferentes categorías y subcategorías relacionadas con el tráfico vial. Este enfoque se ha implementado exitosamente en países como España, Colombia y Chile, y se recomienda a las autoridades peruanas seguir este modelo, ya que ha demostrado reducir el número de muertes y lesiones, especialmente entre los peatones.

Asimismo, Pérez (2020) determinaron la factibilidad de la aplicación de modelos probabilísticos para pronosticar los índices de accidentabilidad en la Compañía Minera Raura S.A.; mediante la distribución probabilística de variables aleatorias discretas de Poisson, resultando que: La probabilidad más alta, P(x) = 0.9758, corresponde a un índice de accidentabilidad (x) que va desde 0 hasta 0.9.

Esto indica que existe un 97.58% de posibilidades de que el índice de accidentabilidad para el próximo período se encuentre en el rango de 0 a 0.9. La probabilidad acumulativa más baja, P (< x1) = 1, se refiere a un índice de accidentabilidad (x) que oscila entre 2 y 2.8. Esto significa que hay un 100% de probabilidad de que el índice de accidentabilidad para el próximo período sea menor que 2.8. La probabilidad acumulativa más alta, P(>x) = 0.0242, se relaciona con un índice de severidad (x) que varía de 0 a 0.9. Esto sugiere que existe un 2.42% de probabilidad de que el índice de severidad para el próximo período sea mayor que 0.9, se puede afirmar que desarrollar modelos probabilísticos para prever los índices de accidentabilidad en la Compañía Minera Raura S.A. es viable con una probabilidad de ocurrencia superior al 80%.

Avalos et. al (2022), este trabajo propone un modelo para estimar el número de lesiones que pueden ocurrir en accidentes de tránsito en la provincia de Lima mediante la implementación de un modelo de regresión lineal múltiple basado en el registro de accidentes ocurridos entre 2016 y 2017. Del análisis realizado se puede concluir que la regresión lineal múltiple permite una predicción suficiente utilizando el modelo con variables explicativas. Combinaciones lineales de otras variables, el análisis de componentes principales permitió reducir el número de variables de 22 a 4. Esto se recomienda para trabajos futuros para simplificar el modelo sin perder demasiada variabilidad. En este caso, se utilizan varios métodos para reducir el número de variables, especialmente el tratamiento de variables categóricas, como el análisis de correspondencia. De esta forma, el objetivo es crear un modelo más sencillo, pero más preciso. Además, la información sobre el número de muertes por un accidente puede servir como

indicador de su gravedad, ya que agregar esta información al modelo de evaluación ayudará a tomar mejores decisiones sobre las prioridades de atención.

2.1.3. Antecedentes Locales

De acuerdo con Choque (2020), se estimó el costo económico indirecto causado por los accidentes de tránsito en el departamento de Puno durante el periodo 2013-2017. Se empleó un enfoque de estudio cuantitativo, analítico, longitudinal y retrospectivo. Los resultados revelaron las variables que influyen en los accidentes de tránsito, destacando que la colisión de vehículos fue el tipo de evento más común, con un 40% de los casos, y el exceso de velocidad fue la causa principal, con un 32%. El 79% de las víctimas fueron hombres, con el grupo de edad más afectado de 26 a 60 años, que agrupa el 51% del total de fallecimientos. Durante los cinco años, el costo económico de las víctimas fatales alcanzó los S/ 311, 135,443.10, mientras que el costo asociado a las víctimas no fatales fue de S/ 64, 949,913.54. En ambos casos, la variable que tuvo mayor influencia en la estimación del costo fue el ingreso promedio mensual.

Mamani & Ponce (2021), aplicaron el método predictivo del Highway Safety Manual 2010 para estimar los accidentes de tránsito y su impacto en la toma de decisiones sobre seguridad vial en la carretera Puno-Ilave. Utilizaron un enfoque cuantitativo con un diseño no experimental correlacional. Concluyeron que el diseño geométrico de la vía no influye en los accidentes de tránsito, mientras que el IMDA y los dispositivos de control de tránsito sí tienen un impacto significativo. El método predictivo de accidentes de tránsito del HSM se aplicó correctamente según las directrices del manual de seguridad vial (2017) para el territorio nacional, obteniendo un factor de calibración (Cr) de 0.839. Este valor,

comparado con estudios similares, sugiere que el procedimiento es prometedor para la evaluación de vías a nivel nacional.

Según, Quispe (2024), identifico y evaluó los puntos críticos de accidentes de tránsito en la vía Puno-Ilave entre los años 2021 y 2022, proponiendo medidas preventivas. Se empleó un enfoque cuantitativo con un diseño no experimental y un nivel descriptivo, tomando como población de estudio la vía Puno-Ilave, con una muestra de 20 puntos críticos. Los resultados, basados en los reportes policiales, indican que entre 2021 y 2022 se registraron 153 accidentes de tránsito en esta vía. Utilizando el método de control de calidad de la tasa de accidentes, se identificaron 5 puntos críticos de alta peligrosidad. En cuanto a los elementos geométricos en tramos curvos, el 30% no cumple con el radio ni con el peralte requerido, el 100% no cumple con el ancho mínimo de la calzada ni de la berma, y todas las curvas carecen de sobreancho. En los tramos rectos, el 73% no cumple con la tangente máxima y el 100% no cumple con el ancho mínimo de la calzada ni de la berma según la norma DG-2018.

2.2. MARCO TEÓRICO

2.2.1. Discriminación logística

Para, Peña (2002), el desafío en la discriminación logística reside en que, en su mayoría, los parámetros son desconocidos y necesitan ser estimados a partir de los datos. Además, los datos disponibles para la clasificación suelen no tener una distribución normal. Por lo tanto, en muchos problemas de clasificación, se recurre al uso de variables discretas.

La regresión logística es un método de análisis de regresión que se utiliza con frecuencia para predecir los resultados de una variable categórica basada en

variables independientes o predictoras. Su utilidad radica en la capacidad de modelar la probabilidad de que ocurra un evento en función de una variedad de factores. Se utiliza una función logística para modelar las probabilidades que describen los posibles resultados de un solo experimento, considerando las variables explicativas. Las dos formas principales de regresión logística son la regresión logística simple y la regresión logística múltiple, (Peña, 2002).

En 1958, David Cox introdujo la Regresión Logística Simple, un enfoque de regresión diseñado para calcular la probabilidad de una variable cualitativa binaria en relación con una variable cuantitativa. Una de las aplicaciones fundamentales de la regresión logística es la clasificación binaria, donde las observaciones se categorizan en uno de dos grupos según el valor de la variable utilizada como predictor.

La regresión logística utiliza modelos estadísticos para comprender las relaciones entre:

- Una variable dependiente cualitativa que puede ser dicotómica (regresión logística binaria o binomial) o tener más de una categoría (regresión logística multinomial).
- Las covariables son variables explicativas independientes que pueden ser cualitativas o cuantitativas.

Las variables explicativas cualitativas deben ser binarias, adoptando el valor 0 en ausencia y 1 en presencia. En caso de que la covariable tenga más de dos categorías, se requiere realizar una conversión mediante la creación de múltiples variables explicativas binarias ficticias (variables dummy). Al llevar a

cabo esta transformación, cada categoría de la variable se incorporaría al modelo de manera individual.

Los modelos de regresión logística cumplen con tres objetivos principales:

- Evaluar la magnitud de la relación entre cada covariable y la variable dependiente.
- Identificar la presencia de interacción y confusión entre las covariables en relación con la variable dependiente (por ejemplo, las odds ratio para cada covariable).
- Categorizar a los individuos en las distintas categorías (presente/ausente) de la variable dependiente.

2.2.2. Modelo Logit

Como resultado, la regresión logística, a diferencia de la regresión lineal, no tiene como objetivo predecir un valor específico de la variable y utilizando una o más variables predictoras (Xs). En cambio, su objetivo es calcular la probabilidad de que ocurra Y tomando en cuenta los valores de las variables Xs. La siguiente es la estructura general de la ecuación, (Bustamante, 2014):

$$P(y=1) = F(Z_i) = \frac{e^{Z_i}}{1+e^{Z_i}} = F(X_i\beta') = \frac{e^{X_i\beta'}}{1+e^{X_i\beta'}};\ Z_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \quad (1)$$

La ecuación de la función de distribución acumulada (FDA), utilizada es la función de distribución logística. El modelo Logit relaciona la variable dicotómica y_i con las variables $X_{2i} \ldots X_{Ki}$ a través de la ecuación:

$$Y_i = \frac{1}{1 + e^{-(\beta_1+\beta_2 X_{2i}+\cdots+\beta_K X_{Ki})}} + u_i \qquad (2)$$

De igual manera que el modelo de probabilidad lineal supone que $E(U_i) = 0$ y dado que la variable de respuesta es la dicotómica se puede denotar que:

$$P = (y_i = 1) = E(y_i/X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki})}} \qquad (3)$$

La importancia del modelo Logit son:

- F hacer referencia a la función de distribución logística.
- u_i es una variable aleatoria que distribuye normal $N(0,\ \sigma^2)$.
- Las variables o características X_i son fijas en el muestreo.
- La variable dependiente Y_i se puede tomar valores de cero y uno.

La interpretación del modelo Logit se puede dar del siguiente hecho: conocidos dados, los valores de la característica X_i, se les asigna una probabilidad, por ejemplo P_i de que la variable Y_i valga la unidad. Así que:

$$Prob(Y_i = 1/X_i) = P_i \qquad (4)$$

Pruebas de Significancia

Para el análisis de regresión logística se consideran las siguientes pruebas estadísticas que facilitan ver el nivel de significancia de las variables:

- Test de Wald: Es una evaluación estadística de los parámetros de la regresión logística, si los valores son significativamente diferentes de cero. El test de Wald matemáticamente se presenta de la siguiente manera:
- $wald = \frac{b^2}{s_b^2}$, que sigue una distribución normal estándar. Los parámetros son significativos si tienen probabilidad menor al 5%.
- Pseudo R^2: Es un estadístico que permite determinar la bondad de ajuste, esto indica cuánto mejora sin restricción a otro modelo con restricción, este

estadístico se expresa en términos de porcentaje, el estadístico Pseudo R^2 matemáticamente se presenta de la siguiente forma:

$$Pseudo\ R^2 = 1 - \frac{L_M^2}{L_0^2}$$

El valor del estadístico también llamado McFadden-R^2 está compuesto por L_M que es el log-likelihood para el modelo con ajuste y L_0 es el log-likelihood para el modelo nulo (solo el intercepto).

Se trata de medidas que evalúan el incremento de la verosimilitud del modelo: el cambio del estadístico -2log(L) o bien de L, la razón de verosimilitud que varía entre 0 y 1. Si $\Lambda = 2 - \log(L)$ entonces $L = \exp\left(-\frac{\Lambda}{2}\right)$ y los pseudo R^2 se calculan de la forma siguiente en el caso de Cox y Snell y del de Negelkerke:

Pseudo R^2 de Cox y Snell:

$$R^2 = 1 - \left(\frac{L_{constante}}{L_{modelo}}\right)^{\frac{2}{n}} = 1 - \exp\left(\frac{\Lambda_{modelo} - \Lambda_{constante}}{n}\right)$$

Estadísticos que varía entre 0 y 1,$0 \leq R^2 \leq 1$,y donde L(a) es el modelo de la constante mientras que L(a,b1,b2,...,bj) es el modelo completo considerado.

Pseudo R^2 de Negelkerke:

$$R_N^2 = \frac{1 - \left(\frac{L_{constante}}{L_{modelo}}\right)^{\frac{2}{n}}}{1 - (L_{constante})^{\frac{2}{n}}} = \frac{1 - \exp\left(\frac{\Lambda_{modelo} - \Lambda_{constante}}{n}\right)}{1 - \exp\left(\frac{-\Lambda_{constante}}{n}\right)}$$

Es un estadístico que varía también entre 0 y 1, con valores algo superiores en relación anterior. Tanto como el anterior son indicadores de la variabilidad explicada.

- Evaluación del estadístico de bondad de ajuste z^2

$$z^2 = \sum_{i=1}^{n} \frac{(p_i - \hat{p}_i)^2}{\hat{p}_i(1-\hat{p}_i)} = \sum_{i=1}^{n} \frac{R_i^2}{\hat{p}_i(1-\hat{p}_i)}$$

Donde R_i es el residuo entre la probabilidad observada y la probabilidad estimada del i-ésimo caso. z^2 Sigue una distribución chi-cuadrado. Para que el modelo sea significativo, la probabilidad asociada debe ser menor o igual a 0,05.

Supuestos del modelo de regresión logística

La regresión logística es un método estadístico que se utiliza para modelar una variable dependiente categórica, generalmente dicotómica (binaria), a partir de una o más variables independientes. A continuación, se presentan los supuestos fundamentales del modelo de regresión logística.

a) Linealidad

En la regresión logística, uno de los supuestos clave es que existe una relación lineal entre las variables independientes y el logit de la variable dependiente. El logit es el logaritmo de las probabilidades de que ocurra el evento de interés. Este supuesto difiere del de la regresión lineal, donde se asume una relación lineal directa entre las variables independientes y la variable dependiente, (Menard, 2002).

El supuesto de linealidad en la regresión logística establece que:

$$log\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Donde:

- $log\left(\frac{P(Y=1)}{P(Y=0)}\right)$ es el logit de la probabilidad de que el evento Y ocurra.
- β_0 es el intercepto del modelo.

- $\beta_1, \beta_2, \ldots \beta_n$ son los coeficientes de las variables independientes X_1, X_n.

En términos más simples, esto significa que se asume una relación lineal entre las variables independientes y el logit de la variable dependiente, no entre las variables independientes y la probabilidad directa de ocurrencia del evento, (Menard, 2002).

Verificación de la linealidad

Para verificar el supuesto de linealidad, se puede utilizar el enfoque de la "prueba del logit". Esto implica lo siguiente:

- Crear nuevas variables que sean los productos cruzados de las variables independientes y el logaritmo natural de sí mismas.
- Añadir estas nuevas variables al modelo logístico.
- Evaluar la significancia de estos términos adicionales. Si son significativos, sugiere que el supuesto de linealidad podría no ser válido.

b) Independencia

El supuesto de independencia es crucial en la regresión logística y en otros modelos estadísticos. Este supuesto establece que las observaciones deben ser independientes entre sí. En otras palabras, el valor de la variable dependiente para una observación no debe estar influenciado por el valor de la variable dependiente de otra observación, (Menard, 2002).

Importancia del supuesto de independencia

La violación del supuesto de independencia puede llevar a:

- Estimaciones sesgadas de los coeficientes del modelo.
- Inferencias estadísticas incorrectas.
- Sobreestimación o subestimación de los errores estándar, lo que afecta las pruebas de significancia.

Verificación del Supuesto de Independencia

Diseño Experimental: Asegurar un diseño experimental adecuado donde las observaciones sean recogidas de manera independiente. Por ejemplo, en estudios transversales, las observaciones se recogen en un único punto en el tiempo, reduciendo el riesgo de dependencia temporal, (Menard, 2002).

Análisis de Residuales: Examinar los residuales del modelo para detectar patrones que puedan indicar dependencia entre las observaciones. Los residuales deben parecer aleatorios y no mostrar estructuras o patrones sistemáticos, (Menard, 2002).

Pruebas Estadísticas: Utilizar pruebas estadísticas específicas como el estadístico Durbin-Watson para detectar autocorrelación en los residuales, aunque esta prueba es más común en la regresión lineal, (Menard, 2002).

c) Multicolinealidad

La multicolinealidad se refiere a una situación en la que dos o más variables independientes en un modelo de regresión están altamente correlacionadas. Esto puede dificultar la estimación precisa de los coeficientes de regresión y puede inflar los errores estándar de las estimaciones, (Menard, 2002).

Explicación del Supuesto de Ausencia de Multicolinealidad

En la regresión logística, se asume que las variables independientes no están altamente correlacionadas entre sí. La presencia de multicolinealidad puede causar varios problemas, tales como:

- Estimaciones Inestables: Los coeficientes de las variables independientes pueden cambiar drásticamente con pequeñas modificaciones en el modelo.
- Dificultad para Determinar el Impacto Individual: La redundancia de información puede hacer que sea difícil determinar el impacto individual de cada variable independiente en la variable dependiente.
- Inflación de los Errores Estándar: Los errores estándar de los coeficientes pueden ser inflados, lo que reduce la significancia estadística de las variables.

Diagnóstico de la Multicolinealidad

Existe la técnica del VIF para diagnosticar la presencia de multicolinealidad en un modelo de regresión logística:

Factor de Inflación de la Varianza (VIF)

Calcula la cantidad de inflación de los errores estándar de los coeficientes como resultado de la correlación entre las variables independientes. Un VIF superior a 10 suele considerarse indicativo de multicolinealidada problemática, (Menard, 2002).

2.2.3. Análisis de componentes principales (PCA)

Karl Pearson introdujo el Análisis de Componentes Principales (PCA) como método estadístico a finales del siglo XIX., especialmente en el contexto del análisis de factores. Sin embargo, el avance de esta técnica se vio limitado debido a la complejidad de los cálculos, hasta que la aparición de las computadoras permitió su desarrollo y aplicación efectiva, (Quiroga & Limon, 2011).

El propósito del Análisis de Componentes Principales (PCA) es agrupar las variables que están correlacionadas entre sí y separarlas de las que no lo están. En el PCA, no se analizan los factores de manera individual; en cambio, se interpreta el agrupamiento de las variables, (Cuadras, 2014).

El Análisis de Componentes Principales (PCA) posee dos características fundamentales que lo convierten en un método ampliamente utilizado para la reducción de la dimensionalidad.

- Las componentes principales capturan la máxima variabilidad o varianza de X de manera secuencial, minimizando la pérdida de información en términos de error de reconstrucción.
- Los componentes principales resultantes son ortogonales entre sí, lo que simplifica su procesamiento posterior al poder ser tratadas de manera independiente.

Los datos de entrada estarán organizados en una matriz $X \in R^{m*p}$, donde m es el número de variables de entrada y n es el número de observaciones. Los datos proyectados se disponen en una matriz $Y \in R^{m*p}$, con p variables de salida y n observaciones, siendo p menor que m (p<m). Cada vector $\boldsymbol{X_i} =$

$[X_1, X_2, \dots\ X_m]^{T_1}$ contiene todas las variables de entrada asociadas a una observación i, mientras que $\boldsymbol{Y_i} = [y_1, y_2, \dots\ y_m]^T$ incluirá las variables de salida.

El objetivo del PCA es identificar una matriz $U \in R^{m*p}$ que transforme linealmente el espacio de los datos de entrada X en una matriz Y con un número reducido de variables. Esto se logra mediante una combinación lineal de las variables originales, proyectándolas en las direcciones de mayor varianza de los datos para conservar la mayor cantidad de información posible. Estas direcciones de máxima varianza están definidas por los p vectores de proyección, u_k, que se encuentran en las columnas de la matriz U.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix} \xrightarrow[\boldsymbol{Y = U^T X}]{\text{PCA}} Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_p \end{bmatrix} \ (m > p)$$

En este proyecto, se denominará "componentes principales" a las variables de salida $\boldsymbol{y_k = u_k^T X}$, mientras que los vectores de proyección u_k serán considerados como los coeficientes o cargas de la k-ésima componente principal.

Para cada variable, la matriz U puede contener hasta p = m vectores de proyección. Sin embargo, en la mayoría de los casos se limita esta característica al mantener únicamente las proyecciones con mayor discrepancia, lo que conlleva inevitablemente a la pérdida de datos. Para minimizar esta pérdida, el algoritmo PCA concentra la mayor parte de la varianza en un número reducido de variables en el nuevo espacio. De este modo, se pueden eliminar aquellas variables del espacio de salida que tienen menor varianza, garantizando la menor pérdida posible de información. La premisa fundamental es conservar la mayor cantidad de información en el menor número de variables posible.

A continuación, se describirán dos enfoques para desarrollar el PCA y obtener la matriz de proyección del nuevo espacio. Estos enfoques tienen diferentes objetivos iniciales: El más común y conocido, busca maximizar la varianza en el nuevo espacio de datos. Aunque inicialmente parecen problemas distintos, se demostrará las formulaciones ofrecen soluciones equivalentes, puede utilizarse para calcular los vectores de proyección.

PCA como método de para maximizar la varianza

Dado que el objetivo del PCA es reducir el número de variables en los datos de entrada manteniendo la mayor cantidad de información posible, en esta sección se detallará el desarrollo del método, centrándose principalmente en la preservación de la varianza.

En términos generales, el funcionamiento del PCA se centra en encontrar, a partir de un conjunto de datos de entrada X con m variables, un vector de pesos u_1 que proyecte este conjunto de datos en la dirección de máxima varianza de X, donde u_k es un vector compuesto por m variables. Una vez obtenido u1, el siguiente paso es buscar otro vector u_2 ortogonal a u1 que capture la máxima varianza posible. Este proceso continúa hasta obtener el p-ésimo vector u_p, que maximiza la varianza y es ortogonal a u$u_1,\ u_2, \dots, u_{p-1}$1. La proyección de los datos de entrada en estos vectores produce las componentes principales: $\boldsymbol{y}_1 = \boldsymbol{u}_1^T X,\ y_2 = \boldsymbol{u}_2^T X, \dots, y_k = \boldsymbol{u}_k^T X$. Aunque es posible calcular hasta la m-ésima componente principal, normalmente solo se necesita calcular las primeras ppp componentes, donde se proyectará la mayor parte de la varianza. Las variables restantes se descartan, lo que provoca una pérdida de información, pero esta será mínima en términos de varianza.

Este enfoque puede explicarse de manera más sencilla utilizando la formulación matricial. Si consideramos U como la matriz compuesta por los p vectores de proyección, y X como la matriz de los datos de entrada, entonces la matriz Y, formada por las p componentes principales, se obtiene de la siguiente manera:

$$Y = U^T X \dots (1)$$

El objetivo del PCA es mantener la mayor cantidad de información en las variables de Y, asegurando que estas sean ortogonales. Una manera de alcanzar este objetivo es maximizando la covarianza de las componentes principales, lo cual se expresa de la siguiente manera:

$$\max_{U} = C_{YY} \dots (2)$$

$$sujeto\ a\ U^T U = \text{I} \dots (3)$$

Para concentrar la información en el menor número posible de componentes sin repetir dicha información, la transformación debe garantizar que las variables resultantes en el nuevo espacio sean ortogonales. Por esta razón, se impone la restricción $U^T U$=I. De este modo, al eliminar los atributos menos significativos, la pérdida de información será mínima. Para encontrar la matriz U que resuelve el problema mencionado en (3), primero se debe desarrollar su expresión de C_{YY} de la siguiente manera:

$$C_{YY} = \frac{1}{n} YY^T = \frac{1}{n}(U^T X)(U^T X)^T = \frac{1}{n} U^T X X^T U = U^T \left(\frac{1}{n} X X^T\right) U = U^T C_{XX} U \dots (4)$$

A partir de esta formulación, podemos redefinir el objetivo del PCA de la siguiente manera:

$$\max_{U} tr\{U^T C_{XX} U\}$$

$$sujeto\ a\ U^T U = I \ldots (5)$$

Nos encontramos ahora con un problema de optimización con restricciones. Para resolver esta expresión, se utilizarán multiplicadores de Lagrange. Primero, se define lo siguiente:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_p \end{bmatrix} \ldots (6)$$

Dado que estamos utilizando la matriz de multiplicadores de Lagrange, tenemos que:

$$L_p = Tr\{U^T C_{XX} U\} - (I - U^T U)\,\Lambda$$

$$C_{XX} = U\Lambda \ldots (7)$$

Se reduce a un problema de valores propios y vectores propios sobre la matriz de covarianza de los datos.

El PCA se puede entender como un problema relacionado con vectores propios y valores propios. Los vectores propios de C_{XX} proporcionan las direcciones de máxima varianza de X. Por lo tanto, si se elige u_1 como el k-ésimo vector propio de C_{XX}, entonces $\boldsymbol{u}_k^T X$ será la k-ésima componente principal. Los vectores propios se ordenan de mayor a menor varianza, y el valor propio asociado a cada vector propio indica la varianza de la componente principal derivada de él.

Usualmente, el cálculo de los vectores propios y valores propios se realiza mediante la técnica de descomposición en valores singulares (SVD). La solución SVD de X es:

$$X = U\Sigma V^T \ldots (8)$$

En esta descomposición, se sabe que, U y V son matrices ortonormales, donde $U = u_1,\ u_2, \ldots, u_m$ tiene en sus columnas los vectores propios de XX^T. Σ es una matriz diagonal que contiene los valores singulares de XX^T. Cada valor singular se define como la raíz cuadrada del valor propio correspondiente $(\sigma_i \equiv \sqrt{\lambda_i})$.

De esta forma se obtiene mediante la SVD de X que:

$$XX^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma^2 U^T$$

$$XX^T U = U\Sigma^2 \ldots (9)$$

Esta última expresión es comparable a (7), aunque no utiliza C_{XX}. Con la SVD de X, se calculan los vectores propios y valores propios de XX^T en lugar de C_{XX}. Para igualar ambas expresiones y obtener finalmente la matriz UUU con las proyecciones ortogonales de máxima varianza de XXX, es necesario aplicar la SVD a 1/√n X.

Después de aplicar el Análisis de Componentes Principales (PCA) a un conjunto de variables y observaciones, se genera un espacio reducido que facilita significativamente el análisis y la interpretación de los datos. Este nuevo espacio es el resultado de combinaciones lineales de las variables originales.

En el análisis de componentes principales, se crea un nuevo sistema de coordenadas basado en los datos originales. En este sistema, el componente con la mayor varianza se identifica como el primer componente principal, seguido por el segundo componente principal, que tiene la segunda mayor varianza, y así sucesivamente, (Pla, 1986). El primer componente principal proporciona la mayor

cantidad de información. A medida que se avanza a los siguientes componentes, la cantidad de información disminuye, es decir, el segundo componente contiene menos información que el primero, y así sucesivamente, (Black, Babin, & Tatham, 2006).

Fundamentos de PCA

- **Transformación Lineal:** El PCA transforma los datos originales X en un nuevo conjunto de variables Z (componentes principales) mediante una transformación lineal. La relación se puede expresar como:

$$Z=XW$$

Dónde: W es la matriz de pesos (auto vectores) y Z son los componentes principales.

- **Varianza y Covarianza**: Los componentes principales son direcciones en el espacio de las variables originales que maximizan la varianza. La primera componente principal es la dirección en la que los datos varían más, la segunda es la dirección ortogonal a la primera con la siguiente mayor variabilidad, y así sucesivamente.
- **Descomposición en Autovalores y Autovectores**: La matriz de covarianza de los datos originales se descompone mediante PCA en autovalores y autovectores. Mientras que los autovalores indican la cantidad de varianza explicada por cada componente, los vectores propios definen las direcciones de los componentes principales.

Ventajas y Limitaciones del PCA

Ventajas:

- **Reducción de Dimensionalidad:** Permite reducir el número de variables mientras se retiene la mayor parte de la variabilidad de los datos.
- **Descorrelación:** Las componentes principales están ortogonalmente no correlacionadas, lo que puede ser útil en modelos estadísticos.
- **Eliminación de Ruido:** Al eliminar componentes con baja varianza, se puede reducir el ruido en los datos.

Limitaciones:

- **Interpretabilidad:** Las componentes principales pueden no tener un significado fácil de interpretar en el contexto original de las variables.
- **Sensibilidad a la Escala:** Requiere que los datos estén correctamente estandarizados, ya que es sensible a las escalas de las variables.
- **Linealidad:** PCA es una técnica lineal y puede no capturar relaciones no lineales entre las variables.

2.2.4. Eficiencia de la predicción

La matriz de confusión y la curva ROC se utilizan durante el proceso de validación para evaluar los modelos de clasificación de regresión logística.

a) Matriz de confusión

Según Fawcett (2005), la matriz de confusión es una herramienta crucial para evaluar el rendimiento de los modelos de clasificación. Se muestran los números de verdaderos positivos, falsos positivos, verdaderos

negativos y falsos negativos para proporcionar un resumen detallado del desempeño del modelo. El cálculo de otras métricas importantes, como la precisión, la sensibilidad y la especificidad, se facilita con esta matriz.

- **Accuracy**

  La precisión Accuracy es una métrica crucial en la evaluación de modelos de clasificación, incluyendo la regresión logística. Esta métrica indica la proporción de observaciones correctamente clasificadas sobre el total de observaciones, (Fawcett, 2005).

  La precisión se define como la capacidad del modelo para clasificar correctamente tanto los positivos como los negativos. Se calcula utilizando la siguiente fórmula:

  $$(Accuracy) = \frac{TP+TN}{TP+TN+FP+FN}$$

  Donde:

  - TP (True Positives): Verdaderos positivos.
  - TN (True Negatives): Verdaderos negativos.
  - FP (False Positives): Falsos positivos.
  - FN (False Negatives): Falsos negativos.

  Importancia de la Precisión

  La precisión es una métrica importante porque proporciona una visión general del rendimiento del modelo, indicando qué tan bien está clasificando las observaciones en general. Sin embargo, es importante considerar que, en problemas de clasificación desbalanceada, la precisión por sí sola puede ser engañosa y debe

complementarse con otras métricas como la sensibilidad (recall) y la especificidad, (Fawcett, 2005).

- **Precisión**

Es la métrica que indica el porcentaje total de valores correctamente clasificados, abarcando tanto los positivos como los negativos.

$$(Precisión) = \frac{TP}{TP+FP}$$

Donde:

- TP (True Positives): Verdaderos positivos.
- FP (False Positives): Falsos positivos.

- **Especificidad (Specificity)**

La especificidad es una medida utilizada para evaluar el rendimiento de los modelos de clasificación, como la regresión logística. Esta métrica mide la capacidad del modelo para identificar correctamente las observaciones negativas, es decir, los verdaderos negativos, (Fawcett, 2005).

La proporción de observaciones verdaderamente negativas entre el total de observaciones es conocida como especificidad. Se calcula utilizando la fórmula a continuación:

Especificidad $(Specificity) = \frac{TN}{TN+FP}$

Donde:

- TN (True Negatives): Verdaderos negativos.
- FP (False Positives): Falsos positivos.

Importancia de la Especificidad

La especificidad es importante porque muestra la capacidad del modelo para evitar falsos positivos. En contextos donde es crucial minimizar los errores de clasificación de los casos negativos (por ejemplo, en diagnósticos médicos donde un falso positivo podría llevar a tratamientos innecesarios), la especificidad es una métrica clave. Complementa a la sensibilidad (recall), proporcionando una imagen más completa del rendimiento del modelo, (Fawcett, 2005).

- **Sensibilidad (Sensitivity)**

Una métrica crucial para evaluar el rendimiento de los modelos de clasificación, como la regresión logística, es la sensibilidad, también conocida como recuperación o tasa de verdaderos positivos. Esta métrica mide la capacidad del modelo para detectar observaciones positivas, es decir, verdaderos positivos, (Fawcett, 2005).

La sensibilidad es la proporción de observaciones positivas verdaderas entre el total. Se calcula utilizando la fórmula a continuación:

$$\text{Sensibilidad } (Sensitivity) = \frac{TP}{TP+FN}$$

Donde:

- TP (True Positives): Verdaderos positivos.
- FN (False Negatives): Falsos negativos.

Importancia de la Sensibilidad

La sensibilidad es importante porque muestra la capacidad del modelo para identificar correctamente los casos positivos. Es especialmente crucial en contextos donde es fundamental minimizar los errores de clasificación de los casos positivos (por ejemplo, en diagnósticos médicos donde un falso negativo podría llevar a la falta de tratamiento necesario). La sensibilidad complementa a la especificidad, proporcionando una imagen más completa del rendimiento del modelo, (Fawcett, 2005).

- **F1 Score en la Evaluación de Modelos de Clasificación**

El puntaje F1 es una medida que se utiliza para evaluar el rendimiento de los modelos de clasificación al combinar la sensibilidad y la precisión en una sola medida. Es especialmente útil cuando se necesita un equilibrio entre precisión y sensibilidad, y cuando hay una distribución desbalanceada entre las clases, (Fawcett, 2005).

El F1 Score se define como la media armónica de la precisión y la sensibilidad. La fórmula para calcular el F1 Score es:

$$\text{F1 Score} = 2 * \frac{Precisión * Sensibilidad}{Precisión + Sensibilidad}$$

Donde:

- $\text{Precisión} = \frac{TP}{TP+FP}$
- $\text{Sensibilidad} = \frac{TP}{TP+FN}$
- TP (True Positives): Verdaderos positivos.
- FP (False Positives): Falsos positivos.
- FN (False Negatives): Falsos negativos.

Importancia del F1 Score

El F1 Score es crucial porque proporciona una medida única que considera tanto los falsos positivos como los falsos negativos, equilibrando la sensibilidad y la precisión. Esto es crucial en escenarios donde ambas métricas son importantes y no se puede priorizar una sobre la otra. Por ejemplo, en diagnósticos médicos, se necesita tanto una alta precisión para minimizar los falsos positivos como una alta sensibilidad para minimizar los falsos negativos, (Fawcett, 2005).

b) Curva de ROC

La capacidad discriminativa de los modelos de clasificación se evalúa principalmente utilizando la curva ROC y su área bajo la curva (AUC). La tasa de verdaderos positivos y la tasa de falsos positivos se muestran en la curva ROC a diferentes umbrales de clasificación. El AUC es una métrica integral que resume la capacidad del modelo para distinguir entre clases, donde un valor más cercano a 1 indica un mejor rendimiento, (Fawcett, 2005).

2.3. MARCO CONCEPTUAL

2.3.1. Accidente

Es un incidente imprevisto, carente de intencionalidad, se desencadenó de manera súbita, resultando en una lesión y causando daños sustanciales a personas u objetos. La rapidez del suceso provocó consecuencias contundentes, poniendo en riesgo la salud y seguridad de quienes estaban presentes, así como exponiendo al medio ambiente a daños en su entorno, (OPS, 2018). La Organización Mundial de la Salud lo describe como un "evento imprevisto, no intencionado, sin voluntad

alguna, originado por una fuerza externa que actúa de manera rápida y se manifiesta a través de la presencia de lesiones físicas o trastornos mentales, (Valdés, Ferrer, & Ferrer, 1996). Un incidente puede manifestarse como una intoxicación o envenenamiento debido al contacto con sustancias sólidas o líquidas, una caída, ahogamiento, incendio, quemadura, torcedura, heridas, entre otros. Estos eventos, que no pueden controlarse ni preverse, no son evitables incluso con precauciones meticulosas, ya que están determinados por el azar, (OPS, 2018). En un eventual accidente de tránsito, se trata de un comportamiento no intencional que escapa al control, aunque esta descripción no resulta la más apropiada. Las desgracias en las vías suelen deberse a acciones irresponsables que podrían evitarse, (EDUVIA, 2007).

2.3.2. Accidente de tránsito

Para, Tabasco (2015), se refiere al daño causado a una persona o propiedad durante un desplazamiento específico, resultado principalmente de la acción arriesgada, negligente o irresponsable de un conductor, pasajero o peatón. No obstante, en muchas ocasiones, también puede atribuirse a fallos mecánicos inesperados, errores en el transporte de carga, condiciones ambientales adversas, cruce de animales en el tráfico e incluso a deficiencias en la infraestructura vial, como fallas en la señalización y en el diseño de caminos y carreteras.

2.3.3. Accidentes de tránsitos fatales

La mortalidad derivada de eventos de tráfico a nivel global ha experimentado un incremento durante el último siglo, como consecuencia del uso excesivo e imprudente de vehículos motorizados y no motorizados. Estos vehículos, aunque proporcionan beneficios para la sociedad y mejoran la calidad

de vida, también conllevan un panorama perjudicial para la salud de quienes transitan por las vías, (OMS, 2018). La pérdida de vidas debido a accidentes de tránsito representa un significativo desafío en términos de salud pública. A diario, más de 3,800 personas fallecen a nivel mundial a causa de estos incidentes, sumando aproximadamente 1.25 millones de víctimas al año. Además, entre 20 y 50 millones de personas resultan afectadas por traumatismos no mortales, (OMS, 2018).

2.3.4. Accidentes de tránsitos no fatales

Los accidentes de tránsito no fatales son aquellos en los que las personas involucradas sufren lesiones, pero no fallecen. Aunque no resultan en muertes, estos accidentes pueden tener consecuencias graves, como lesiones severas que pueden llevar a discapacidades a largo plazo y afectaciones significativas en la calidad de vida de las víctimas. Se estima que por cada persona que muere en un accidente de tránsito, hay muchas más que sufren lesiones no fatales (Murray & Lopez, 1996). La atención a los accidentes no fatales es crucial, ya que representan una carga considerable para los sistemas de salud y para la economía, debido a los costos de atención médica y la pérdida de productividad.

2.3.5. Lugar de ocurrencia de accidentes de tránsito

a. Autopista

Las autopistas están diseñadas para el tránsito rápido y eficiente de vehículos a alta velocidad. No obstante, estas velocidades elevadas provocan que los accidentes en autopistas tiendan a ser graves y frecuentemente fatales. La ausencia de intersecciones a nivel y la presencia de múltiples carriles pueden incrementar la severidad de los accidentes.

b. Calles

Las calles urbanas tienen velocidades más bajas en comparación con las autopistas y están diseñadas para facilitar tanto el tránsito vehicular como el de peatones. Los accidentes en estas vías suelen implicar interacciones con peatones y ciclistas, y frecuentemente se deben a la congestión y a las numerosas maniobras que se deben realizar en estos entornos.

c. Jirones

Los jirones, característicos de algunas ciudades latinoamericanas, son calles angostas con alta densidad de peatones y tráfico local. Los accidentes en estos lugares suelen ser causados por la falta de visibilidad y el espacio limitado para maniobras.

d. Avenida

Las avenidas son arterias principales en las ciudades que permiten un flujo vehicular más eficiente que las calles secundarias. No obstante, debido a la alta densidad de tráfico y la presencia de intersecciones, las avenidas son frecuentemente escenarios de accidentes de tránsito.

e. Carreteras

Las carreteras, particularmente en zonas rurales, conllevan un riesgo considerable debido a las altas velocidades y a una infraestructura de seguridad menos desarrollada. Los accidentes en estas vías suelen resultar en colisiones frontales y salidas de la carretera.

2.3.6. Tipos accidentes de tránsito

Se refieren a incidentes en los que está involucrado un único vehículo en movimiento a lo largo de una vía, y tienen una conexión directa o indirecta con el factor humano. A su vez, los accidentes de tránsito se dividen en categorías específicas:

- **Choque**

Se refiere a la colisión de un vehículo en movimiento contra cualquier objeto, ya sea permanente o temporalmente fijo, o contra otro vehículo estacionado. Los choques simples pueden clasificarse según la forma de colisión, siendo estas frontal, angular, lateral o posterior.

- **Volcadura**

Consiste en el vuelco de un vehículo en movimiento, pudiendo ocurrir sobre sus lados, hacia adelante o hacia atrás. En casos específicos, la volcadura puede adquirir denominaciones particulares, como la volcadura en tonel (izquierdo o derecho), donde el número en la parte superior indica el número de vueltas de giro y el número en la parte inferior indica el lado sobre el que queda apoyado.

- **Despiste**

Se refiere a la pérdida de contacto de las llantas de un vehículo con la superficie normalmente transitable de la vía. El despiste puede clasificarse en dos tipos.

- **Incendio**

Como un tipo de accidente de tránsito, ocurre cuando el vehículo está en movimiento y suele ser causado por fallas mecánicas, como roturas en la tubería de alimentación, problemas en el tiempo de explosión que resultan en la expulsión de gasolina no quemada, o la inflamación del combustible por diversas circunstancias casuales. En algunos casos, puede presentarse un incendio debido a un cortocircuito en el sistema eléctrico.

- **Atropello**

Se trata de la interacción entre un vehículo en movimiento y un peatón, y la clasificación de los atropellos se realiza en función de la forma de la colisión:

Proyección: Se manifiesta cuando el vehículo colisiona con el peatón y lo desplaza, ya sea hacia adelante (en la dirección de la circulación del vehículo) o lateralmente.

Volteo: curre cuando el vehículo impacta al peatón y, debido a la forma de la carrocería y a la acción del peatón, este es elevado y cae sobre el vehículo, pudiendo rodar hacia atrás o lateralmente.

Aplastamiento: Este tipo de atropello se presenta cuando cualquiera de las ruedas del vehículo pasa sobre cualquier parte del cuerpo del peatón.

Compresión: Similar al aplastamiento, esta categoría se configura cuando una o varias ruedas del vehículo pasan sobre alguna parte del cuerpo del peatón.

Arrastre: Se produce cuando un vehículo arrastra a un peatón al engancharse con alguna parte, ya sea del cuerpo o vestimenta, con alguna parte del vehículo.

Encontronazo: Sucede cuando el peatón impacta directamente contra el vehículo.

2.3.7. Tipos de vehículos en los accidentes de tránsito

- **Automóviles**

Los automóviles son los vehículos que más frecuentemente se ven involucrados en accidentes de tránsito, debido a su alta presencia en las carreteras. Estos accidentes pueden variar desde colisiones leves hasta choques mortales, dependiendo de factores como la velocidad y las condiciones del camino.

- **Motocicletas**

Las motocicletas implican un alto riesgo en accidentes de tránsito debido a su menor estabilidad y la falta de protección para el conductor. Los motociclistas tienen una mayor probabilidad de sufrir lesiones graves o mortales en comparación con los ocupantes de vehículos cerrados.

- **Camiones**

Los camiones, debido a su gran tamaño y peso, pueden provocar accidentes extremadamente graves. Las colisiones que involucran camiones a menudo resultan en daños significativos y tasas de mortalidad más altas, especialmente para los ocupantes de vehículos más pequeños.

- **Autobuses**

Los accidentes que involucran autobuses pueden tener consecuencias severas debido al gran número de pasajeros. Aunque son menos comunes que los

accidentes de automóviles y motocicletas, pueden resultar en múltiples lesiones debido a la cantidad de personas a bordo.

- **Bicicletas**

Los ciclistas son especialmente vulnerables en accidentes de tránsito debido a su falta de protección y baja visibilidad. Los accidentes que involucran bicicletas a menudo provocan lesiones graves para los ciclistas, particularmente en áreas urbanas con alto tráfico.

- **Vehículos de carga ligera**

Los vehículos de carga ligera, como furgonetas y camionetas, están involucrados en una cantidad considerable de accidentes de tránsito. Aunque son más grandes que los automóviles, no ofrecen la misma protección que los camiones, lo que puede llevar a lesiones graves para los ocupantes.

2.3.8. Causas de accidentes de tránsito

- **Exceso de Velocidad**

El exceso de velocidad es una de las causas más importantes de accidentes de tránsito, siendo responsable de casi un tercio de los accidentes mortales. Conducir a altas velocidades disminuye el tiempo de reacción del conductor y aumenta la gravedad de las colisiones.

- **Conducción Bajo los Efectos del Alcohol**

El consumo de alcohol por parte de los conductores contribuye a aproximadamente un tercio de los accidentes de tránsito mortales. El alcohol

perjudica la coordinación, el juicio y el tiempo de reacción, lo que incrementa significativamente el riesgo de accidentes.

- **Distracciones al Volante**

Las distracciones al volante, como el uso de teléfonos móviles, comer o interactuar con los pasajeros, son una causa importante de accidentes de tránsito. Estas distracciones pueden desviar la atención del conductor y aumentar el riesgo de colisiones.

- **Condiciones de las Carreteras**

Las malas condiciones de las carreteras, como los baches, la falta de señalización y la iluminación inadecuada, contribuyen de manera significativa a los accidentes de tránsito. Una infraestructura deficiente puede dificultar la maniobrabilidad y aumentar el riesgo de colisiones.

2.3.9. Incidencia diaria de accidentes de tránsito

La frecuencia diaria de accidentes de tránsito se refiere al número de incidentes que ocurren en un día determinado. Esta métrica es crucial para identificar y analizar los patrones diarios de accidentes, lo que permite desarrollar estrategias específicas y efectivas para prevenirlos en momentos críticos del día, como las horas pico. Al comprender cuándo y por qué ocurren estos accidentes, las autoridades pueden implementar medidas de seguridad, campañas de concienciación y cambios en la infraestructura vial para reducir la incidencia de estos eventos y mejorar la seguridad en las carreteras.

2.3.10. Mortalidad en peatones por accidentes de tránsito

Cuando abordamos la mortalidad de peatones causada por accidentes de tránsito, resulta crucial examinar este fenómeno en el contexto de un proceso continuo de transformación en los sistemas de movilidad, influenciado por cambios en la infraestructura, vehículos y entorno. Estos factores inciden en la mayoría de los casos creando un ambiente que no favorece la seguridad vial, introduciendo frecuentemente nuevos riesgos en las vías transitadas. De este modo, los errores cometidos por los usuarios de las vías públicas están vinculados a limitaciones naturales, como la visibilidad nocturna, la visión periférica, la estimación de velocidad y distancias, el procesamiento de información por el cerebro, así como otros factores psicológicos relacionados con la edad y el género, que afectan el riesgo de estar involucrado en accidentes.

Los peatones, considerados una población altamente vulnerable, forman parte de los "usuarios vulnerables de la vía pública", según la clasificación de la Organización Mundial de la Salud (OMS), que incluye a peatones, ciclistas y motociclistas. La variabilidad en edad, género y situación socioeconómica de los peatones que sufren lesiones o fallecen revela la diversidad de este grupo. Estas características difieren considerablemente entre países y regiones, destacando la importancia de recopilar y analizar datos a nivel local para obtener un conocimiento completo del problema en dicha escala, (OMS, 2018).

2.3.11. Factores de los accidentes de tránsito

El Instituto de Tráfico y Seguridad Vial de España (INTRAS: 2007: 13) señala que varios estudios han resaltado las notables dificultades para disminuir la cantidad de accidentes. Para lograr este objetivo, es esencial adquirir un mejor

entendimiento de las características, causas y consecuencias de los accidentes. En este sentido, se propone la implementación de programas de investigación dedicados a la accidentalidad, que permitan el acceso a las estadísticas y el análisis de los datos, (Vellacorta, 2015).

Se refieren a elementos que ayudan a comprender cómo contribuyeron al desarrollo de un accidente de tránsito. Estos factores revelan la conducta u operación de los usuarios de la vía, ya sea peatones, conductores u ocupantes (pasajeros). La causa basal del incidente, es decir, la causa principal que determinó las circunstancias del accidente de tránsito, se considera un factor determinante. Por otro lado, los factores contributivos son aquellos que guardan una relación indirecta con los eventos que ocasionaron el accidente de tránsito y que también forman parte integral del suceso en cuestión, (Arapa, 2019)

2.3.12. Factores determinantes de los accidentes de tránsito

De acuerdo a Arapa (2019), el factor determinante se define como la causa principal y fundamental del accidente, destacando su importancia y prevalencia en los hechos. Este factor es responsable de desencadenar el accidente de tránsito en su totalidad. Establecer estos factores de manera precisa resulta crucial para la imposición de sanciones administrativas al conductor u operador del vehículo, así como a los usuarios de la vía.

2.3.13. Factores contributivos de los accidentes de tránsito

Se consideran factores contributivos aquellos que surgen como resultado o condiciones del incidente, pero que tienen una importancia reducida al compararlos con la causa principal, según el Manual de normas y procedimientos para intervención e investigación de accidentes de tránsito (2013). Estos factores,

de alguna manera, están relacionados con las responsabilidades de un conductor al circular en la vía y, como su nombre indica, contribuyen al desenlace final del accidente de tránsito. Principio del formulario.

CAPÍTULO III

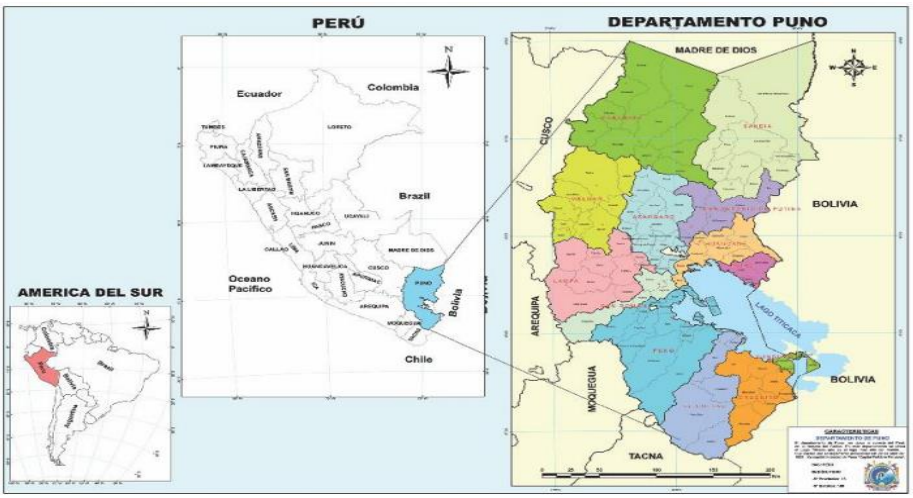
MATERIALES Y MÉTODOS

3.1. LOCALIZACIÓN GEOGRÁFICA DEL ESTUDIO

Puno constituye uno de los veinticuatro departamentos que, junto con la Provincia Constitucional del Callao, forman la República del Perú. La región Puno se encuentra en la sierra sudeste del país, en la meseta del Collao, situada entre los 13°0066'00" y 17° 17′ 30″ de latitud sur, así como los 71° 6′ 57″ y 68° 48′ 46″ de longitud oeste respecto al meridiano de Greenwich. Sus límites son el sur con la región Tacna, al este con el Estado Plurinacional de Bolivia, y al oeste con las regiones de Cusco, Arequipa y Moquegua. La capital es la ciudad homónima de Puno, y Juliaca es la localidad más poblada. Situado en el sureste de la nación, limita al norte con el departamento de Madre de Dios, al este con Bolivia (Departamento de La Paz), al suroeste con Tacna y Moquegua, y al oeste con Arequipa y Cuzco. Con 66,997 km2, es el quinto departamento más grande después de Loreto, Ucayali, Madre de Dios y Cuzco. Fue fundado el 26 de abril de 1822**.**

Figura 1

Ubicación de la región de Puno



Nota: (Canales, 2012).

3.2. MÉTODO DE ESTUDIO

El método hipotético-deductivo consta de varios pasos fundamentales: la observación del fenómeno a estudiar, la formulación de una hipótesis para explicar dicho fenómeno, la deducción de consecuencias o proposiciones más simples derivadas de la hipótesis, y la verificación de la verdad de estos enunciados mediante la comparación con la experiencia, (Fernández & Baptista, 2014).

3.3. DISEÑO DE ESTUDIO

El diseño de investigación adoptado es no experimental, se elige esta modalidad debido a que no se manipulan variables ni se realizan intervenciones; en cambio, se enfoca en la observación y análisis de la relación entre las variables, Asimismo, se lleva a cabo mediante la aplicación de métodos estadísticos que siguen un proceso secuencial y se respaldan en evidencia, (Fernández & Baptista, 2014).

3.4. POBLACIÓN Y MUESTRA

3.4.1. Población

La población objetivo está determinada por todos los accidentes de tránsitos ocurridos en la región de Puno en el año 2022, que fueron registrados por el por la X-MACREPOL-PUNO, por la División De Estadística De La Policía Nacional Del Perú.

3.4.2. Muestra

La muestra está conformada por los 1823 accidentes de tránsito fatales y no fatales en carreteras (vías de la región de Puno) reportados por la X-MACREPOL-PUNO durante el año 2022. La muestra es no probabilística,

incluyen una técnica de selección basada en las particularidades del estudio, frente a un criterio estadístico de generalización, (Fernández & Baptista, 2014).

3.5. TÉCNICA DE RECOLECCIÓN Y PROCESAMIENTOS DE DATOS

3.5.1. Plan de procesamiento y análisis de datos

Se realizo las coordinaciones, con el responsable de la División De Estadística De La Policía Nacional Del Perú, obteniendo la autorización para la revisión de los datos presentado en tablas de frecuencia correspondientes a los accidentes de tránsito durante el año 2022.

Una vez solicitadas los datos se sometieron a un proceso de codificación y transferidos a una base de datos mediante el programa Excel 2016 y el lenguaje de programación Python; Posteriormente a los datos se les realizo un análisis descriptivo y luego realizó el análisis de regresión logística para finalmente ser interpretados. Los resultados se organizaron en tablas y figuras estadísticas, de acuerdo a los objetivos de la investigación.

Selección de Variables a través del Análisis de Componentes Principales

Debido a la cantidad de variables junto a sus categorías, se optó por utilizar la técnica de análisis de componentes principales con el propósito de agrupar variables y no perder información. El Análisis de Componentes Principales, permite transformar un conjunto de variables posiblemente correlacionadas en un conjunto de valores de variables linealmente no correlacionadas, denominadas componentes principales. Este método es especialmente útil en contextos donde se maneja un gran número de variables, ya que ayuda a reducir la dimensionalidad

del conjunto de datos, manteniendo la mayor parte de la varianza presente en los datos originales. En este estudio, se aplicó PCA para condensar 79 características en 9 variables principales, facilitando así un análisis más eficiente y manejable sin sacrificar información crucial sobre los accidentes de tránsito.

Análisis estadístico

La base de datos empleada en este estudio fue gentilmente proporcionada por la X-MACREPOL-PUNO, la cual contiene información detallada sobre accidentes de tránsito fatales y no fatales ocurridos durante el año 2022. El conjunto de datos, en formato Excel, incluía una tabla de frecuencias que abarcaba todas las variables relevantes para el análisis. Todas las variables fueron categorizadas como cualitativas categóricas, lo que requirió un enfoque meticuloso para su codificación y análisis.

Para transformar y organizar estos datos, se utilizaron herramientas de Python, un programa ampliamente empleado en el análisis de datos. En particular, se utilizó la técnica de codificación one-hot, una técnica crucial para convertir variables categóricas en un formato que los modelos de aprendizaje automático puedan comprender. La codificación one-hot genera una nueva variable binaria, conocida como dummy (0 = No o 1 = Sí), para cada categoría de una variable, garantizando que cada categoría se represente de forma única sin establecer un orden arbitrario. Este método es esencial para evitar relaciones espurias entre categorías y permite a los algoritmos procesar los datos con eficiencia y precisión.

Codificación One-Hot: Esta técnica se utilizó para convertir las variables categóricas en columnas binarias. La codificación one-hot asegura que cada categoría se represente de manera única y sin imponer un orden arbitrario,

evitando así relaciones espurias entre categorías y permitiendo una interpretación precisa por parte de los modelos de aprendizaje automático, (Hastie, Tibshirani, & Friedman, 2009)

Estandarización de variables: Para el cálculo de los modelos, las variables se estandarizaron para tener una media de 0 y una desviación estándar de 1. De lo contrario, las variables con mayor varianza dominarían al resto, (Quiroga & Limon, 2011).

Análisis de Componentes Principales (PCA): Se utilizó PCA para reducir la dimensionalidad del conjunto de datos. Esta técnica es fundamental para simplificar el modelo sin perder datos importantes, (Quiroga & Limon, 2011).

Modelo Logit: Finalmente, se aplicó un modelo logit para el análisis principal. Este modelo se benefició de las etapas previas de codificación y reducción de dimensionalidad, permitiendo una interpretación y predicción más robusta y precisa de los datos, (Peña, 2002).

De lo que se trata es de plantear un modelo de regresión logística que estime la probabilidad de fatalidad en accidentes de tránsito en la región de Puno,2022.

3.6. OPERACIONALIZACIÓN DE VARIABLES

Tabla 1

Operacionalización de variables

Variables		Tipo de variable	Codificación
Dependiente	Fatalidad de accidentes de tránsito	Cualitativa	1= Accidente fatal 0= Accidente no fatal BINARIO
Independientes	Lugar de ocurrencia	Cualitativa	Autopista, Calle, Jirón, Pasaje, Avenida, Curva, Cruce de avenidas, Cruce de calles, Carretera, Otros, BINARIO, 0=NO, 1=SI
	Tipo de accidentes de tránsito	Cualitativa	Choque, Atropello, Choque y atropello, Caída, Volcadura, Incendio de vehículo, Choque y fuga, Atropello y fuga, Despiste y volcadura, Colisión, Despiste, Colisión y fuga, Otros, BINARIO, 0=NO, 1=SI
	Tipos de vehículos involucrados en el accidente de tránsito	Cualitativa	Automóvil, Station wagon, Camioneta pick up, Camioneta rural, Camioneta panel, Ómnibus, Camión, Remolcador, Remolque y semirremolque, Vehículo no identificado, Moto lineal, Motocar, Triciclo, Furgoneta, Bicicleta, Vehículo no identificado, Otros, BINARIO,0=NO,1=SI
	Causas de los accidentes de tránsitos	Cualitativa	Exceso de velocidad, Imprudencia del conductor, Ebriedad del conductor, Imprudencia del peatón, Imprudencia del pasajero, Exceso de carga, Desacato señal de tránsito por parte del conductor, Desacato señal de tránsito por parte del peatón, Falla mecánica, Falta de luces, Vía en mal estado, Señalización defectuosa, Invasión de carril / maniobra no permitida, Vehículo mal estacionado, Factor ambiental, Estado ebriedad del peatón, No identifica la causa, No tiene la certeza de determinar la causa, Otros ,BINARIO,0=NO,1=SI

Incidencia Diaria	Cuantitativa	Lunes, martes, miércoles, jueves, viernes, sábado, domingo, BINARIO, 0=NO, 1=SI
Tiempo de ocurrencia del accidente de tránsito (Hora)	Cuantitativa	Hora 00:01 - 02:00 hrs. Hora 02:01 - 04:00 hrs. Hora 04:01 - 06:00 hrs. Hora 06:01 - 08:00 hrs. Hora 08:01 - 10:00 hrs. Hora 10:01 - 12:00 hrs. Hora 12:01 - 14:00 hrs. Hora 14:01 - 16:00 hrs. Hora 16:01 - 18:00 hrs. Hora 18:01 - 20:00 hrs. Hora 20:01 - 22:00 hrs. Hora 22:01 - 24:00 hrs. BINARIO,0=NO,1=SI
Heridos	Cuantitativa	Heridos Femenino Heridos Masculino BINARIO,0=NO,1=SI
Victimas Muertos	Cuantitativa	Muertos Femenino Muertos Masculino BINARIO,0=NO,1=SI
Genero del conductor	Cualitativa	Masculino Femenino BINARIO,0=NO,1=SI

Nota: Elaboración propia.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. IDENTIFICACIÓN DE FACTORES

4.1.1. Análisis Univariado de la Fatalidad en Accidentes de Tránsito

La tabla 2, muestra las consecuencias de los accidentes de tránsito en la región de Puno en 2022. Se registraron 1823 accidentes de tránsito. De estos incidentes, la gran mayoría fueron no fatales, sumando un total de 1449 casos, lo que representa el 79.5% del total. Esto indica que, aunque los accidentes de tránsito son frecuentes, la mayoría no resultan en fatalidades. Por otro lado, los accidentes fatales, aunque menos comunes, constituyeron una proporción significativa del total. Hubo 374 accidentes fatales, representando el 20.5% de todos los accidentes registrados.

Tabla 2

Accidentes de tránsito registrados en la región de Puno en el 2022

Consecuencia	**Accidentes de tránsitos**	
	Frecuencia	**Porcentaje (%)**
No fatal	1449	79.5%
Fatal	374	20.5%
Total	**1823**	**100%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la siguiente tabla 4, presenta los accidentes de tránsito registrados durante el año 2022 en diferentes lugares de la región de Puno. Las carreteras tuvieron una alta incidencia, constituyendo el 24.0% del total de accidentes y destacándose con un 36.6% de accidentes fatales, lo que indica un mayor riesgo en estas vías. Las avenidas y calles fueron escenarios comunes de accidentes no

fatales, con un 14.4% y un 15.0% respectivamente, pero presentaron menores proporciones de accidentes fatales, con un 7.5% y un 3.7% respectivamente. Los jirones representaron el 11.9% del total de accidentes, con un 13.7% de no fatales y un 5.1% de fatales. Los menos frecuentes fueron los accidentes en cruces de avenidas y calles, curvas, y pasajes, que juntos sumaron una pequeña fracción del total. Un 2.7% de los accidentes ocurrieron en lugares categorizados como "otros", lo que indica la presencia de incidentes en ubicaciones variadas y menos definidas.

Tabla 3

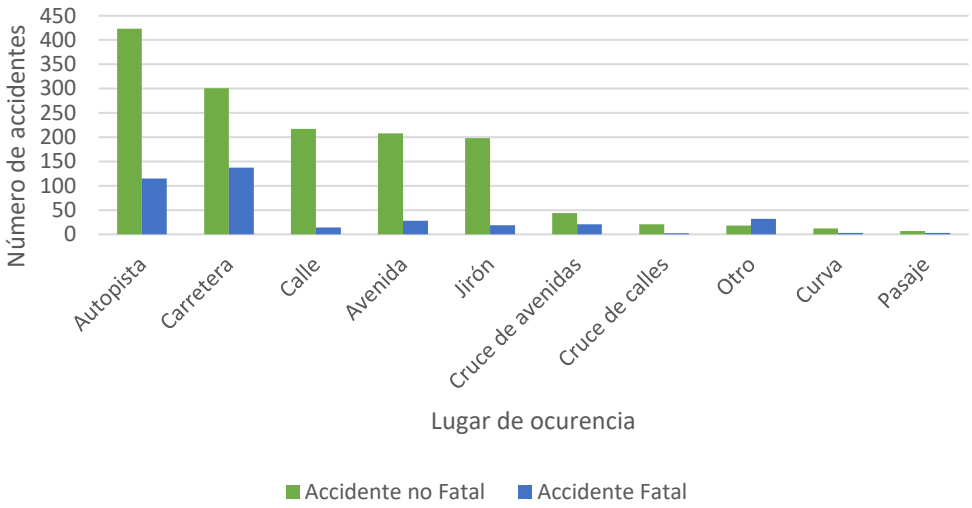
Lugares de ocurrencia de accidentes de tránsito registrados en el 2022

Lugar de ocurrencia del accidente de Tránsito	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Autopista	423	29.2%	115	30.7%	538	29.5%
Avenida	208	14.4%	28	7.5%	236	12.9%
Calle	217	15.0%	14	3.7%	231	12.7%
Carretera	301	20.8%	137	36.6%	438	24.0%
Cruce de avenidas	44	3.0%	21	5.6%	65	3.6%
Cruce de calles	21	1.4%	2	0.5%	23	1.3%
Curva	12	0.8%	3	0.8%	15	0.8%
Jirón	198	13.7%	19	5.1%	217	11.9%
Pasaje	7	0.5%	3	0.8%	10	0.5%
Otro	18	1.2%	32	8.6%	50	2.7%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 2

Distribución de accidentes de tránsito por lugar de ocurrencia



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 4, se observan diversos tipos de accidentes registrados en la región de Puno durante el año 2022. Los choques encabezaron la lista de incidentes, constituyendo el 34.1% del total, con una notable proporción de 37.6% no fatales y 20.6% fatales. Los atropellos también fueron prevalentes, representando el 25.8% de los accidentes, con una distribución similar entre no fatales (25.8%) y fatales (25.9%). Los despistes, aunque menos comunes, mostraron una mayor tendencia a la fatalidad, con un 12.6% del total de accidentes, de los cuales el 15.5% fueron fatales y el 11.9% no fatales. Otros tipos de accidentes, como caídas, colisiones, incendios de vehículos y volcaduras, sumaron una fracción menor del total, pero no menos significativa en términos de gravedad. Los accidentes de "choque y fuga" y "atropello y fuga" destacaron por su peligrosidad, constituyendo el 3.0% y el 2.7% del total, respectivamente. Finalmente, la categoría "otros" representó el 12.6% de los incidentes, evidenciando la variedad de situaciones peligrosas en las vías de Puno.

Tabla 4

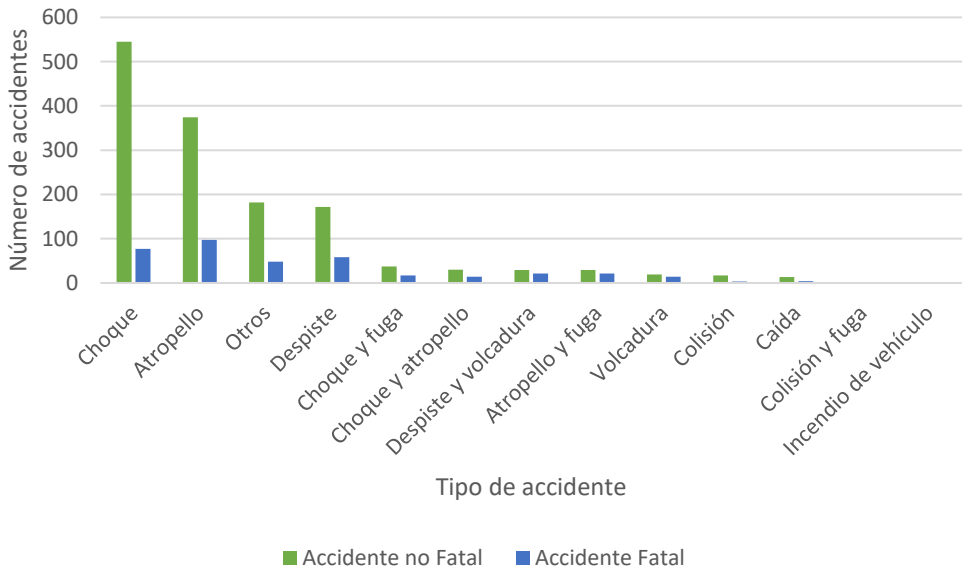
Tipos de accidentes de tránsito registrados en el 2022

Tipo de accidente de Tránsito	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Atropello	374	25.8%	97	25.9%	471	25.8%
Caída	13	0.9%	4	1.1%	17	0.9%
Choque	545	37.6%	77	20.6%	622	34.1%
Choque y atropello	30	2.1%	14	3.7%	44	2.4%
Choque y fuga	37	2.6%	17	4.5%	54	3.0%
Colisión	17	1.2%	3	0.8%	20	1.1%
Colisión y fuga	1	0.1%	0	0.0%	1	0.1%
Despiste	172	11.9%	58	15.5%	230	12.6%
Despiste y volcadura	29	2.0%	21	5.6%	50	2.7%
Incendio de vehículo	1	0.1%	0	0.0%	1	0.1%
Volcadura	19	1.3%	14	3.7%	33	1.8%
Atropello y fuga	29	2.0%	21	5.6%	50	2.7%
Otros	182	12.6%	48	12.8%	230	12.6%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 3

Distribución de accidentes de tránsito por clase



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 5, se identifican los tipos de vehículos implicados en los accidentes de tránsito ocurridos en la región de Puno durante el año 2022. Los automóviles fueron los más comúnmente involucrados, representando el 38.6% del total de accidentes, con un 42.5% de los no fatales y un 23.3% de los fatales. Las motos lineales también fueron significativas, constituyendo el 15.2% del total, con una proporción de 14.0% en accidentes no fatales y 19.8% en fatales. Las camionetas rurales y pick up también presentaron una alta incidencia, sumando el 9.6% y el 8.8% respectivamente, siendo la rural responsable del 12.6% de los accidentes fatales. Otros vehículos como camiones (5.2%), camionetas panel (4.2%), y motocar (3.8%) también mostraron una participación notable en los accidentes. Aunque menos frecuentes, vehículos como bicicletas, triciclos, y ómnibus estuvieron involucrados en una pequeña fracción de los accidentes. Es importante destacar que un 6.6% de los accidentes involucraron otros tipos de vehículos no especificados.

Tabla 5

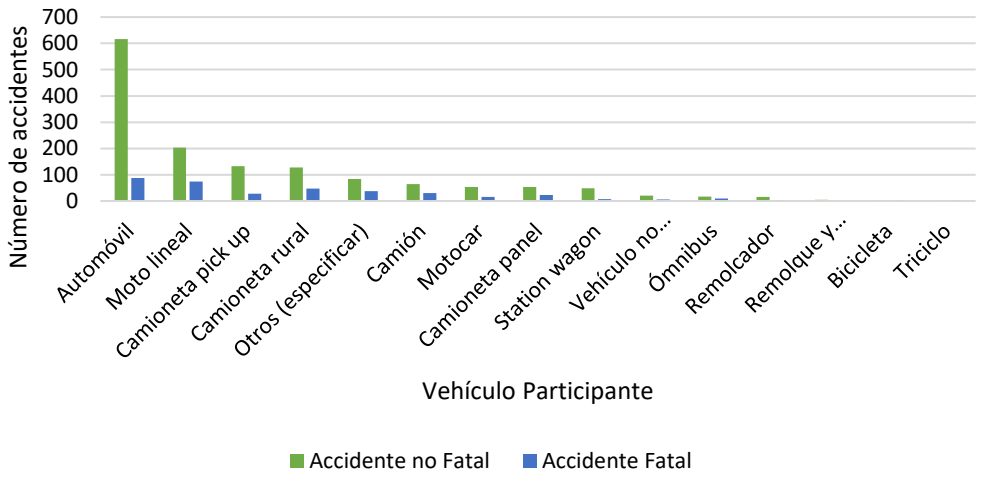
Tipos de vehículos en los accidentes de tránsito involucrados en el 2022

Tipo de vehículo involucrado en el accidente de Tránsito	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Automóvil	616	42.5%	87	23.3%	703	38.6%
Bicicleta	4	0.3%	3	0.8%	7	0.4%
Camión	65	4.5%	30	8.0%	95	5.2%
Camioneta panel	53	3.7%	23	6.1%	76	4.2%
Camioneta pick up	133	9.2%	28	7.5%	161	8.8%
Camioneta rural	128	8.8%	47	12.6%	175	9.6%
Moto lineal	203	14.0%	74	19.8%	277	15.2%
Motocar	54	3.7%	16	4.3%	70	3.8%
Ómnibus	17	1.2%	9	2.4%	26	1.4%
Remolcador	16	1.1%	4	1.1%	20	1.1%
Remolque y semirremolque	5	0.3%	0	0.0%	5	0.3%
Station wagon	49	3.4%	7	1.9%	56	3.1%
Triciclo	2	0.1%	3	0.8%	5	0.3%
Vehículo no identificado	20	1.4%	6	1.6%	26	1.4%
Otros (especificar)	84	5.8%	37	9.9%	121	6.6%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 4

Distribución de accidentes de tránsito por vehículo participante



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

La Tabla 6, muestra las causas de los accidentes de tránsito registrados en la región de Puno en 2022. Analizando las causas de estos accidentes, se puede ver que el exceso de velocidad fue la causa principal, representando el 27,2 % del total de accidentes, con un 26,3 % de los accidentes no fatales y un 30,5 % de los accidentes fatales. La imprudencia del conductor y el desacato a las señales de tránsito también fueron causas significativas, constituyendo el 16.8% y el 18.5% del total de accidentes, respectivamente. Aunque la ebriedad del conductor fue responsable del 18.4% de los accidentes, se observó una menor proporción de accidentes fatales (10.4%) en comparación con los no fatales (20.5%). Otros factores, como la imprudencia del peatón y el pasajero, la falla mecánica y el factor ambiental, también contribuyeron a la ocurrencia de accidentes, aunque en menor medida. Es notable que el 4.9% de los accidentes no identificaron una causa específica, y en el 2.9% no se pudo determinar con certeza la causa.

Tabla 6

Causa de accidentes de tránsito registrados en el 2022

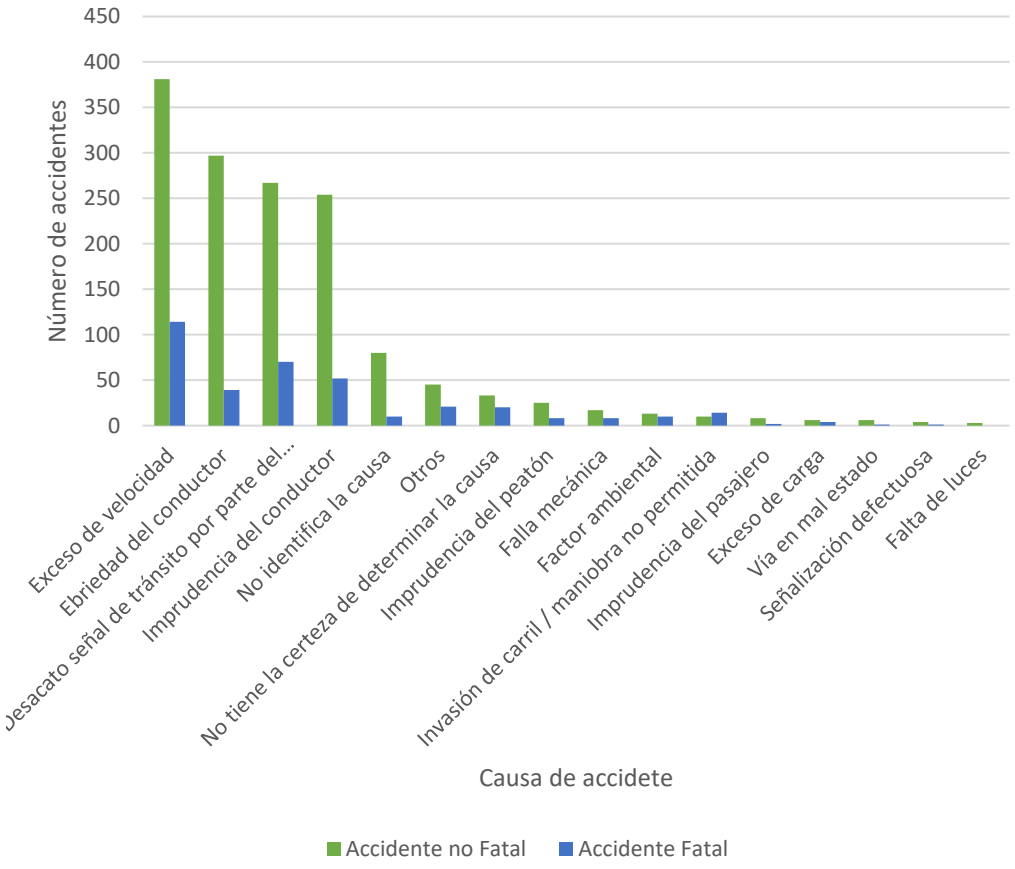
Causas de los accidentes de Tránsito	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Imprudencia del peatón	25	1.7%	8	2.1%	33	1.8%
Desacato señal de tránsito por parte del conductor	267	18.4%	70	18.7%	337	18.5%
Ebriedad del conductor	297	20.5%	39	10.4%	336	18.4%
Exceso de carga	6	0.4%	4	1.1%	10	0.5%
Exceso de velocidad	381	26.3%	114	30.5%	495	27.2%
Factor ambiental	13	0.9%	10	2.7%	23	1.3%
Falla mecánica	17	1.2%	8	2.1%	25	1.4%
Falta de luces	3	0.2%	0	0.0%	3	0.2%
Imprudencia del conductor	254	17.5%	52	13.9%	306	16.8%
Imprudencia del pasajero	8	0.6%	2	0.5%	10	0.5%

Causas de los accidentes de Tránsito	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Invasión de carril / maniobra no permitida	10	0.7%	14	3.7%	24	1.3%
Señalización defectuosa	4	0.3%	1	0.3%	5	0.3%
Vía en mal estado	6	0.4%	1	0.3%	7	0.4%
No identifica la causa	80	5.5%	10	2.7%	90	4.9%
No tiene la certeza de determinar la causa	33	2.3%	20	5.3%	53	2.9%
Otros	45	3.1%	21	5.6%	66	3.6%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 5

Distribución de accidentes de tránsito por causas



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 7, se observan el horario de los accidentes de tránsito registrados en la región de Puno durante el año 2022. La mayor incidencia de accidentes se registró entre las 00:01 y las 02:00 horas, representando el 21.0% del total, con un 21.7% de accidentes no fatales y un 18.4% de accidentes fatales. Las horas pico de la tarde y la noche también mostraron una alta incidencia, particularmente entre las 18:01 y las 20:00 horas, que representaron el 13.5% del total de accidentes, con un 14.1% de no fatales y un 11.0% de fatales. Los accidentes en el intervalo de 16:01 a 18:00 horas representaron el 8.9% del total, con un 8.4% de no fatales y un 11.0% de fatales. Otras franjas horarias con menor pero significativa incidencia incluye las primeras horas de la mañana (06:01 - 08:00 horas) y las últimas horas de la noche (22:01 - 24:00 horas), ambas con alrededor del 6.1% del total de accidentes. Es notable que los accidentes durante la madrugada (02:01 - 04:00 horas) y las primeras horas del día (04:01 - 06:00 horas) representaron un menor porcentaje, pero con una proporción relativamente alta de accidentes fatales, destacándose especialmente las horas de 04:01 a 06:00 con un 8.3% de accidentes fatales. Estos datos indican que hay un mayor riesgo de accidentes graves por la noche y por la mañana, subrayando la necesidad de reforzar las medidas de seguridad vial durante estos períodos, como la implementación de controles de velocidad, mayor iluminación y patrullaje, y campañas de concienciación sobre los peligros de conducir en horarios de mayor riesgo.

Tabla 7

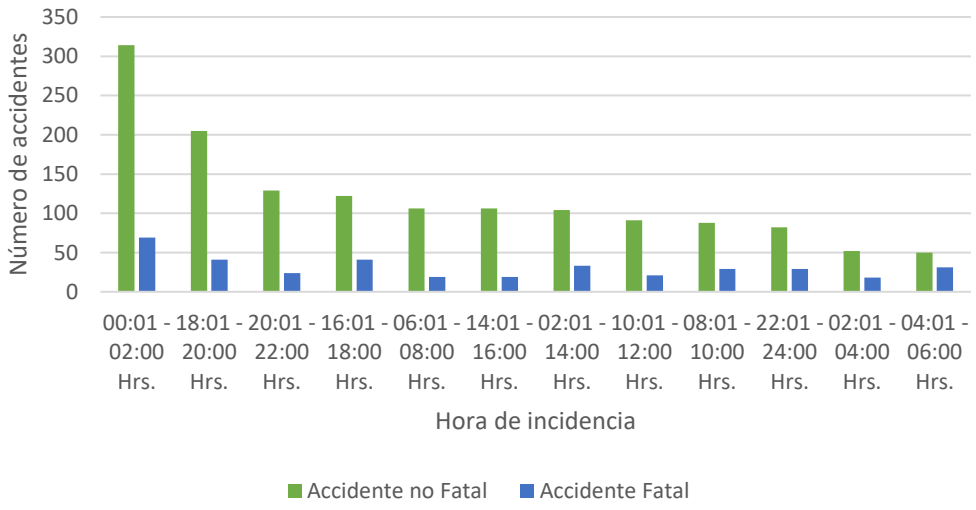
Horario de accidentes de tránsito registrados en el 2022

Incidencia Horaria	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
00:01 - 02:00 Hrs.	314	21.7%	69	18.4%	383	21.0%
02:01 - 04:00 Hrs.	52	3.6%	18	4.8%	70	3.8%
04:01 - 06:00 Hrs.	50	3.5%	31	8.3%	81	4.4%
06:01 - 08:00 Hrs.	106	7.3%	19	5.1%	125	6.9%
08:01 - 10:00 Hrs.	88	6.1%	29	7.8%	117	6.4%
10:01 - 12:00 Hrs.	91	6.3%	21	5.6%	112	6.1%
02:01 - 14:00 Hrs.	104	7.2%	33	8.8%	137	7.5%
14:01 - 16:00 Hrs.	106	7.3%	19	5.1%	125	6.9%
16:01 - 18:00 Hrs.	122	8.4%	41	11.0%	163	8.9%
18:01 - 20:00 Hrs.	205	14.1%	41	11.0%	246	13.5%
20:01 - 22:00 Hrs.	129	8.9%	24	6.4%	153	8.4%
22:01 - 24:00 Hrs.	82	5.7%	29	7.8%	111	6.1%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 6

Distribución de accidentes de tránsito por hora de incidencia



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 8, se observan los accidentes de tránsito registrados en la semana en la región de Puno durante el año 2022. Los domingos registraron la

mayor incidencia de accidentes, representando el 30.7% del total, con un 31.7% de accidentes no fatales y un 26.7% de fatales. Los sábados también presentaron una alta incidencia, sumando el 17.0% del total de accidentes, con un 17.9% de no fatales y un 13.6% de fatales. Los viernes siguieron en frecuencia, representando el 17.0% del total de accidentes, con un 15.6% de no fatales y un 22.2% de fatales, indicando un alto riesgo en este día. Los días entre semana mostraron una distribución más uniforme de los accidentes, con los jueves y miércoles registrando el 10.1% y el 9.5% del total, respectivamente. Es notable que los jueves presentaron una mayor proporción de accidentes fatales (11.8%) en comparación con otros días laborales. Los martes y lunes tuvieron la menor incidencia, representando el 8.2% y el 7.5% del total de accidentes, respectivamente, con una distribución similar entre accidentes no fatales y fatales. Estos datos indican que los fines de semana, especialmente los domingos, tienen una mayor incidencia de accidentes, posiblemente debido a un aumento en la movilidad y la relajación de las medidas de seguridad vial.

Tabla 8

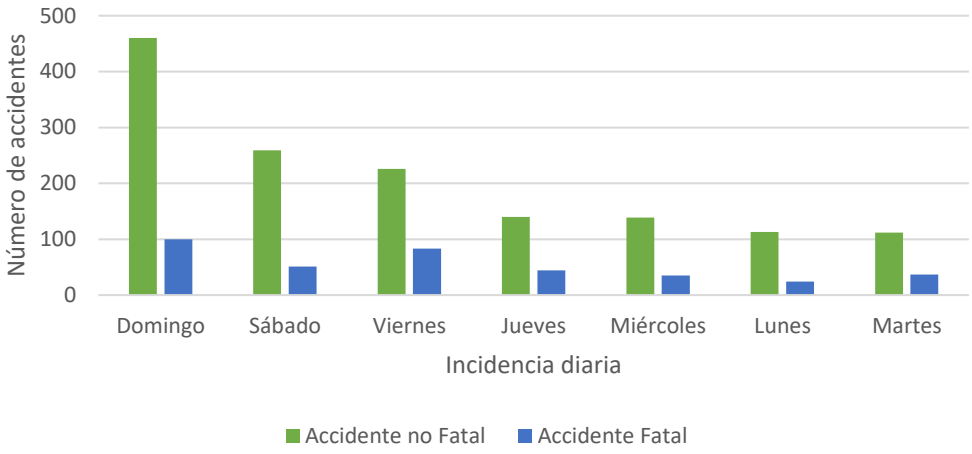
Accidentes de tránsitos registrado durante la semana en el 2022

Incidencia Diaria	Fatalidad de accidente de transito				Total	
	Accidente no Fatal		Accidente Fatal			
	N	%	N	%	N	%
Lunes	113	7.8%	24	6.4%	137	7.5%
Martes	112	7.7%	37	9.9%	149	8.2%
Miércoles	139	9.6%	35	9.4%	174	9.5%
Jueves	140	9.7%	44	11.8%	184	10.1%
Viernes	226	15.6%	83	22.2%	309	17.0%
Sábado	259	17.9%	51	13.6%	310	17.0%
Domingo	460	31.7%	100	26.7%	560	30.7%
Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 7

Distribución de accidentes de tránsito por día de incidencia



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 9, se observan los accidentes de tránsito registrados victimas por género en la región de Puno durante el año 2022. En los accidentes de tránsito fatales, los hombres constituyen la mayoría de las víctimas mortales con un 66.3% (248 muertes), en comparación con las mujeres que representan el 33.7% (126 muertes). Esto indica que los hombres tienen más del doble de probabilidad de morir en accidentes fatales que las mujeres, lo que sugiere la necesidad de políticas de seguridad vial y campañas de concienciación específicas para abordar los factores que contribuyen a esta mayor mortalidad entre los conductores masculinos.

Tabla 9

Accidentes de tránsitos registrados por género muertos en el 2022

Victimas Muertos	**Fatalidad de accidente de transito**	
	Accidente Fatal	
	N	**%**
Muertos Femeninos	126	33.7%
Muertos Masculinos	248	66.3%
Total	**374**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 8

Cantidad de muertos por género en los accidentes de tránsito

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 10, se observan los heridos en los accidentes de tránsito registrados en la región de Puno durante el año 2022. En los accidentes de tránsito no fatales, las mujeres representan el 85.1% de los heridos (1233), mientras que los hombres constituyen el 14.9% (216), mostrando que las mujeres son significativamente más propensas a resultar heridas en accidentes no fatales en comparación con los hombres.

Tabla 10

Género de heridos por los accidentes de tránsito en el 2022

Heridos	**Fatalidad de accidente de transito**	
	Accidente no Fatal	
	N	**%**
Heridos Femeninos	1233	85.1%
Heridos Masculinos	216	14.9%
Total	**1449**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 9

Cantidad de heridos en los accidentes de tránsito

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 11, en los accidentes de tránsito, los conductores masculinos son los más incidentes, representando el 76.5% del total de conductores involucrados (1395 de 1823), con un 94.5% (1369) de participación en accidentes fatales y un 7.0% (26) en accidentes no fatales. Por otro lado, las conductoras femeninas constituyen el 23.4% del total de conductores involucrados (426 de 1823), con una participación del 5.5% (80) en accidentes fatales y del 92.5% (346) en accidentes no fatales. Además, en un 0.1% de los casos (2), no se conoce el género del conductor debido a la fuga. Esto evidencia que los conductores masculinos no solo están más involucrados en accidentes en general, sino que también tienen una mayor incidencia en accidentes fatales en comparación con las conductoras femeninas.

Tabla 11

Género de los conductores en los accidentes de tránsito en el 2022

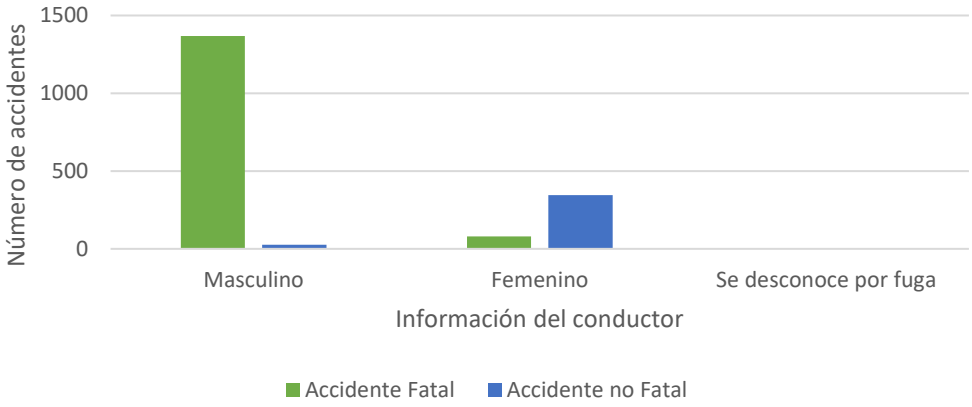
Conductor	Fatalidad de accidente de transito				Total	
	Accidente Fatal		Accidente no Fatal			
	N	%	N	%	N	%
Masculino	1369	94.5%	26	7.0%	1395	76.5%
Femenino	80	5.5%	346	92.5%	426	23.4%
Se desconoce por fuga	0	0.0%	2	0.5%	2	0.1%

Total	**1449**	**100.0%**	**374**	**100.0%**	**1823**	**100.0%**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Figura 10

Distribución de accidentes de tránsito por información del conductor



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

4.2. ESTIMACIÓN DEL MODELO DE PREDICCIÓN DEL ESTUDIO

4.2.1. Selección de variables

El conjunto de datos incluye 1,823 registros y 9 variables. De estos, 374 corresponden a accidentes no fatales (valor = 0) y 1,449 a accidentes fatales (valor = 1). Para el proceso de entrenamiento del modelo, se utilizó el 70% de los registros (1,276), mientras que el 30% restante (547) se reservó para la fase de prueba.

Análisis de Componentes Principales (PCA)

El análisis de componentes principales (PCA) se realizó para reducir la dimensionalidad de las variables transformadas. La siguiente tabla muestra los resultados del PCA, incluyendo la desviación estándar, la proporción de varianza explicada por cada componente y la proporción acumulada de varianza explicada:

Tabla 12

Resultados del análisis de componentes principales

Componente Principal	Desviación Estándar	Proporción de Varianza	Proporción Acumulada
PC1	1.835	0.374	0.374
PC2	1.2743	0.1804	0.5545
PC3	1.085	0.1308	0.6853
PC4	0.9868	0.1082	0.7935
PC5	0.9301	0.0961	0.8896
PC6	0.7032	0.0549	0.9445
PC7	0.4725	0.0248	0.9693
PC8	0.4366	0.0212	0.9905
PC9	0.2923	0.0095	1

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Estos resultados indican que los primeros componentes principales explican la mayor parte de la variabilidad en los datos. En particular, los primeros cuatro componentes principales explican aproximadamente el 79.35% de la varianza total. Esta reducción dimensional es crucial para simplificar el análisis manteniendo la mayoría de la información relevante.

En la tabla 12, observa las variables agrupadas significativas para el modelo:

PC1 X1: Accidentes por clase: Choque, Atropello, Choque y atropello, Caída, Volcadura, Incendio de vehículo, Choque y fuga, Atropello y fuga, Despiste y volcadura, Colisión, Despiste, Colisión y fuga, Otros.

PC2 X2: Causas de los accidentes: Exceso de velocidad, Imprudencia del conductor, Ebriedad del conductor, Imprudencia del peatón, Imprudencia del pasajero, Exceso de carga, Desacato señal de tránsito por parte del conductor, Desacato señal de tránsito por parte del peatón, Falla mecánica, Falta de luces, Vía en mal estado, Señalización defectuosa, Invasión de carril / maniobra no

permitida, Vehículo mal estacionado, Factor ambiental, Estado ebriedad del peatón, No identifica la causa, No tiene la certeza de determinar la causa, Otros.

PC3 X3: Vehículo Participante: Automóvil, Station wagon, Camioneta pick up, Camioneta rural, Camioneta panel, Ómnibus, Camión, Remolcador, Remolque y semirremolque, Vehículo no identificado, Moto lineal, Motocar, Triciclo, Furgoneta, Bicicleta, Vehículo no identificado, Otros.

PC4 X4: Lugar de ocurrencia: Autopista, Calle, Jirón, Pasaje, Avenida, Curva, Cruce de avenidas, Cruce de calles, Carretera, Otros.

PC5 X5: Incidencia Diaria: lunes, martes, miércoles, jueves, viernes, sábado, domingo.

PC7 X7: Victimas Muertos: Muertos femeninos y muertos masculinos

PC8 X8: Heridos: Heridos masculinos y heridos femeninos.

PC9 X9: Conductor: género masculino, género femenino y se desconoce por fuga.

En la tabla 13, se muestra la significancia de las variables X1, X2, X3, X4, X5, X7, X8, X9 en el modelo; por lo tanto, influyen en el modelo. Cabe mencionar que se eliminó una variable no significativa la cual es X6 (Incidencia Horaria) porque no era significativo para el modelo.

Tabla 13

Variables que ingresan al modelo

	Coef	Std err	z	P>|z|	[0.025	0.975]
Const	-4.339	0.469	-9.25	0.0000	-5.259	-3.42
X1: Accidentes por clase	-0.981	0.355	-2.769	0.0060	-1.677	-0.287
X2: Causas de los accidentes	-1.083	0.318	-3.409	0.0010	-1.706	-0.461
X3: Vehículos Participantes	0.8289	0.267	3.105	0.0020	0.306	1.352
X4: Lugar de ocurrencia	1.6779	0.303	5.536	0.0000	1.084	2.272
X5: Incidencia Diaria	-0.836	0.22	-3.805	0.0000	-1.267	-0.405
X7: Victimas Muertos	0.3876	0.118	3.275	0.0010	0.156	0.619
X8: Heridos	1.2369	0.301	4.105	0.0000	0.646	1.827
X9: Conductor	2.1454	0.158	13.565	0.0000	1.835	2.455

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 14 muestra que el modelo de regresión logística tiene un buen poder explicativo. El R cuadrado de Cox y Snell, con un valor de 0.724, indica que las variables independientes explican el 72.4% de la variabilidad en la clasificación de accidentes. A su vez, el R cuadrado de Nagelkerke, que es de 0.806, sugiere que el modelo explica aproximadamente el 80.6% de la variabilidad de los datos, lo que demuestra un ajuste robusto y efectivo del modelo para predecir la fatalidad de los accidentes de tránsito.

Tabla 14

Prueba del coeficiente de determinación

Resumen del modelo		
Logaritmo de la verosimilitud	**R cuadrado de Cox y Snell**	**R cuadrado de Negelkerke**
-180.2787	0,724	0,806

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Verosimilitud del modelo:

En la tabla 15, se evidencia la significancia estadística mediante la prueba de Wald, donde se destaca que la constante en el modelo es estadísticamente significativa. Por lo tanto, es necesario incluir la constante en el modelo, ya que esta provoca un impacto en la variable de fatalidad de accidentes (ya sean fatales o no fatales).

Tabla 15

Prueba del coeficiente de intercepción

Variables en la ecuación						
	B	Error estándar	Wald	gl	Sig.	Exp(B)
Constante	-1,330	,064	437,566	1	,000	,265

Nota: Elaboración propia en base a los datos de la PNP

Al examinar la hipótesis nula (βi = 0), junto con la significancia estadística asociada y el valor de la Razón de Probabilidades (OR) exp (B), se determina que las variables que contribuyen al modelo, como accidentes por (X1), causas de los accidentes (X2), vehículos involucrados(X3), lugar de ocurrencia(X4), incidencia diaria(X5) victimas muertos(X7), heridos (X8) y genero del conductor (X9) son estadísticamente significativas a un nivel de significación del 5%. Por consiguiente, el modelo estimado se presenta de la siguiente manera:

El modelo logístico puede escribirse como:

$$P(x) = \frac{1}{1 + e^{[-4.3398 - 0.9817x_1 - 1.0834x_2 + 0.8289x_3 + 1.6779x_4 - 0.836x_5 + 0.3876x_7 + 1.2369x_8 + 2.1454x_9]}}$$

$$Ln(odds) = -4.3398 - 0.9817x_1 - 1.0834x_2 + 0.8289x_3 + 1.6779x_4 - 0.836x_5 + 0.3876x_7 + 1.2369x_8 + 2.1454x_9$$

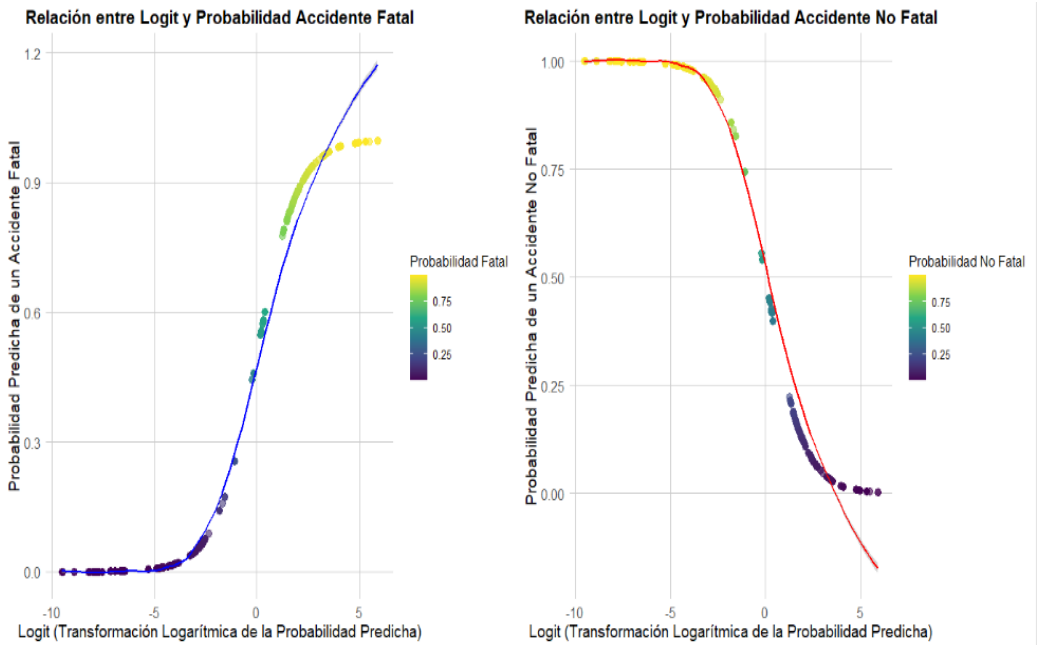
4.2.2. Supuestos de la regresión logística

- **Linealidad**

Para verificar el cumplimiento del supuesto de linealidad del modelo de regresión logística, se observó la relación entre el logit de la probabilidad de un accidente fatal y las variables predictoras. En la figura de dispersión entre el logit y la probabilidad predicha de accidentes fatales y no fatales indica que el modelo de regresión logística se ajusta adecuadamente a los datos. Ambas curvas de tendencia muestran la esperada forma sigmoide, sugiriendo que la relación entre las variables independientes y el logit de la probabilidad de ocurrencia de accidentes sigue una tendencia aproximadamente lineal. Esto confirma que el supuesto de linealidad en la regresión logística se cumple, validando la idoneidad del modelo para predecir las probabilidades de accidentes fatales y no fatales de manera precisa y consistente.

Figura 11

Relación del modelo logit y las variables predictoras



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

- **Independencia**

Para analizar la presencia de autocorrelación en los residuos del modelo de regresión, se usó la prueba de Durbin-Watson. Este estadístico, con un valor calculado de 1.993 y un p-valor de 0.902, indica que no hay una correlación significativa entre los residuos. En términos prácticos, esto significa que las observaciones en el estudio parecen ser independientes unas de otras, lo cual es esencial para garantizar la solidez de los hallazgos. La falta de autocorrelación sugiere que cualquier patrón en los datos no se debe a la dependencia temporal o estructural entre las observaciones, validando así la robustez del análisis y la interpretación de los resultados de la investigación.

Tabla 16

Prueba del Durbin-Watson

Prueba de Durbin -Watson			
Lag	Autocorrelation	D-W statistic	p-value
1	0.003566119	1.99258	0.902

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

- **Multicolinealidad**

Los factores de inflación de la varianza (VIF) calculados indican el grado de multicolinealidad entre las variables predictoras del modelo. Con un VIF promedio de 2.95, el nivel de multicolinealidad es moderado y, dado que todos los valores de VIF son menores a 10, se puede concluir que no existe una multicolinealidad severa entre las variables predictoras. Esto sugiere que las variables del modelo no están altamente correlacionadas entre sí, cumpliendo así uno de los supuestos fundamentales de la regresión logística. Por lo tanto, se considera que el modelo es robusto y adecuado para el análisis.

Tabla 17

Prueba de Multicolinealidad

Variables	Factor de Inflación de la Varianza
Accidentes por clase	7.107992
Causas de los accidentes	2.199613
Vehículos Participantes	4.26528
Lugar de ocurrencia	2.872857
Incidencia Diaria	2.568538
Victimas Muertos	1.365791
Heridos	2.512513
Conductor	1.296147
VIF promedio	**2.95**

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

Eficiencia de la predicción

La tabla 18, muestra los resultados de las métricas de regresión logística utilizando un conjunto de 9 variables principales obtenidas mediante PCA para predecir la fatalidad en accidentes de tránsito. Las métricas de evaluación para el modelo de clasificación de la fatalidad de los accidentes indican que tiene una accuracy del 95%, lo que significa que clasifica correctamente el 95% de todos los accidentes como fatales o no fatales. Los accidentes fatales tienen una precisión del 86%, lo que significa que el modelo predice como fatales los que realmente lo son. El recall del 89% indica que el modelo identifica correctamente los accidentes fatales reales. La puntuación F1 del 87% muestra un buen equilibrio entre precisión y recall en la identificación de accidentes fatales.

Para los accidentes no fatales, la precisión es del 97%, indicando que el 97% de los accidentes predichos como no fatales realmente lo son. El recall del 96% muestra que el modelo identifica correctamente los accidentes no fatales reales. El F1 score del 97% refleja un excelente equilibrio entre precisión y recall para esta categoría.

Tabla 18

Métricas de la regresión logística con variables PCA

Fatalidad de accidentes de transito	Métricas			
	Accuracy	**Precisión**	**Recall**	**F1**
Accidente Fatal	95%	86%	89%	87%
Accidente no Fatal		97%	96%	97%

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la tabla 19, se presenta la matriz de confusión del modelo de regresión logística con el 30% de datos de validación (547 datos). Los resultados son los siguientes:

- **Verdaderos Positivos (VP): 424** - Los casos en los que el modelo predijo "Accidente no fatal" y realmente fueron "Accidente no fatal".
- **Verdaderos Falsos (VF): 16** -Los casos en los que el modelo predijo "Accidente no fatal" pero realmente fueron "Accidente fatal".
- **Falsos Positivos (FP): 12**-Los casos en los que el modelo predijo "Accidente fatal" pero realmente fueron "Accidente no fatal".
- **Falsos Negativos (FN): 95** - Los casos en los que el modelo predijo "Accidente fatal" y realmente fueron "Accidente fatal".

Tabla 19

Matriz de confusión regresión

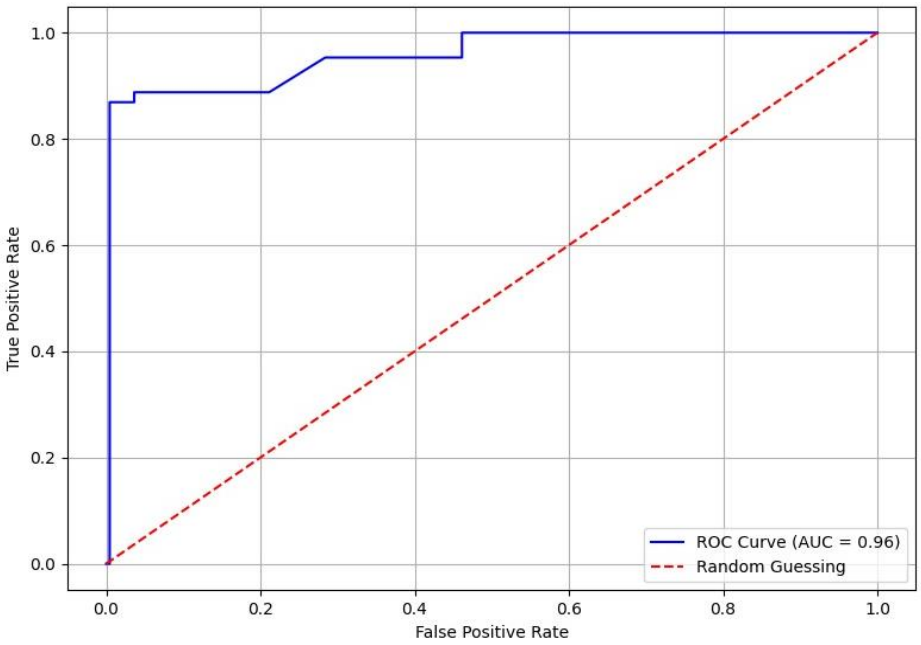
		Predichas	
		0: Accidente no fatal	**1: Accidente Fatal**
Reales	0: Accidente no Fatal	VP 424	VF 16
	1: Accidente Fatal	FP 12	FN 95

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

En la figura 12, se observa la curva ROC de la regresión logística, donde presenta un AUC de 0.96 indica que el modelo tiene un rendimiento excelente en la clasificación de accidentes fatales y no fatales. Cuanto más cercano esté el AUC a 1, mejor es el modelo para distinguir entre las clases.

Figura 12

Curva ROC regresión logística con variables PCA



Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

4.3. PREDICCION DE LA FATALIDAD EN ACCIDENTES

En la tabla 16 se observa la probabilidad de que ocurran accidentes de tránsitos en la región de Puno para el año 2022. Donde el caso Y_{20} se presenta de un accidente de tránsito fatal (0.738088), el accidente es por choque y atropello causado por exceso de carga, el vehículo participante es un automóvil, el lugar de ocurrencia es una autopista, es un día lunes, la víctima es un hombre, el herido es un hombre y el conductor del automóvil es un hombre.

Tabla 20

Probabilidad en los accidentes de tránsitos en la región de Puno 2022

Variables	Caso	
Y: Fatalidad de accidentes de transito	Y_{20} **Atributo**	
X_1:Accidentes por clase	Choque y atropello	
X_2:Causas de los accidentes	Exceso de carga	
X_3:Vehículo Participante	Automóvil	
X_4:Lugar de ocurrencia	Autopista	
X_5:Incidencia Diaria	Lunes	
X_7:Victimas Muertos	Muertos Masculinos	
X_8:Heridos	Heridos Masculinos	
X_9:Conductor	Masculino	
Probabilidad	**Accidente Fatal**	0.738088
	Accidente no Fatal	0.261912

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

El odds ratio (OR) a partir de las probabilidades proporcionadas:

$$Odds = \frac{Probabilidad\ de\ accidente\ fatal}{Probabilidad\ de\ accidente\ no\ fatal}$$

Los Odds para un accidente fatal y no fatal son:

- **Accidente fatal:** $Odds = \frac{0.738088}{1-0.738088} \approx 2.82$

- **Accidente no fatal:** $Odds = \frac{0.261912}{1-0.261912} \approx 0.355$

Sin embargo, el odds ratio en este caso es simplemente el odds de un accidente fatal, ya que estamos comparando las probabilidades de un solo evento (accidente fatal) frente a no tenerlo:

$$Odds = \frac{Odds\ de\ Accidente\ Fatal}{Odds\ de\ accidente\ no\ Fatal} = \frac{2.82}{0.355} \approx 7.94$$

El odds ratio es aproximadamente 7.94, lo que indica que la probabilidad de tener un accidente fatal es casi 8 veces mayor que la de no tenerlo, dado el mismo conjunto de condiciones.

En la tabla 21, se evidencia resumen de varios casos en la probabilidad de accidentes de tránsitos en la región de Puno en el año 2022.

Tabla 21

Resumen de casos en la probabilidad de casos de accidentes de tránsitos

Accidentes	Actual	Predicción	Accidente no fatal	Accidente fatal	Odds
555	0	0	0.998829	0.001171	0.00117237
1742	0	0	0.916941	0.083059	0.09058271
297	1	1	0.278544	0.721456	2.59009708
733	0	0	0.998828	0.001172	0.00117338
785	0	0	0.999668	0.000332	0.00033211
239	1	1	0.018664	0.981336	52.5790827
1106	0	0	0.998643	0.001357	0.00135884
1708	0	1	0.4051	0.5949	1.46852629
212	1	1	0.059429	0.940571	15.8268017

Nota: Elaboración propia en base a los datos de la X-MACREPOL-PUNO

4.4. DISCUSIÓN

En primer lugar, los hallazgos de Choque (2020) en la región de Puno son consistentes con los resultados obtenidos en 2022. Al igual que en el periodo 2013-2017, los choques siguen siendo el tipo de accidente más frecuente en 2022, representando el 34.1% del total, con una significativa proporción de incidentes no fatales (37.6%) y fatales (20.6%). El exceso de velocidad fue identificado como la principal causa de accidentes en ambos estudios, con un 32% según Choque (2020) y una incidencia igualmente preocupante en 2022, representando el 27.2% del total de accidentes, con un 26.3% de los no fatales y un 30.5% de los fatales. En cuanto a las víctimas, Choque (2020) reporta que el 79% de las muertes correspondieron a hombres, un patrón que se mantiene

en 2022, donde la mayoría de los accidentes de tránsito involucraron a conductores masculinos. Estos puntos de concordancia subrayan la persistencia de ciertos factores de riesgo en la región y la necesidad de estrategias continuas y enfocadas para mitigar estos problemas.

En segundo lugar, los modelos de regresión con análisis de componentes principales (PCA) muestran una capacidad razonable de R cuadrado de 0.72, lo que indica que estos componentes pueden explicar una proporción significativa de la variación en los datos. Esta metodología permite reducir la complejidad del modelo al disminuir el número de variables de 79 a 9 componentes principales, manteniendo la capacidad predictiva del modelo. La importancia de los componentes principales se refuerza en el estudio de Avalos et al. (2022) en Lima, donde aplicaron PCA en sus modelos de regresión y lograron reducir el número de variables de 22 a 4 componentes principales, obteniendo un R cuadrado aproximado de 0.74. Este resultado destaca la eficacia del PCA para simplificar modelos complejos sin sacrificar precisión en las predicciones, confirmando su utilidad en el contexto del análisis de accidentes de tránsito.

En tercer lugar, los principales factores que con un 5% de nivel de significancia se encontraron como influyentes en la fatalidad de accidentes tránsitos en la región de Puno durante el 2022 son: accidente de tránsito por clase, causas de los accidentes, vehículos participantes en los accidentes de tránsitos, lugar de ocurrencia del accidente de tránsito, incidencia diaria de accidentes, victimas por los accidentes de tránsitos, heridos de accidentes de tránsitos, genero del conductor; panorama similar es el encontrado por Aguirre y Balarezo (2022) que afirma por medio de su estudio en Quito la existencia de un efecto positivo en la relación entre los accidentes sin señalización y la influencia del alcohol; opinión que también se refuerza por Chipana (2023) en su estudio en Villa El Salvador, con el cual señala que de los casos referentes a accidentes de

tránsito, el factor humano desempeña el papel más significativo y es responsable de aproximadamente el 68% de los accidentes; además de tener otro factor importante como lo es la unidad móvil.

En cuarto lugar, el modelo predictivo de la fatalidad de accidentes de tránsitos en la región de Puno, que permite comprender la ocurrencia de los accidentes en la presente investigación el modelo de regresión logística, bajo la validación de las métricas: accuracy, recall, F1 y el área debajo de las curvas ROC, tal como lo hizo (Silva & Roman, 2021).

V. CONCLUSIONES

PRIMERA: La presente investigación identifico los factores correlacionados con los accidentes de tránsito fatales en la región de Puno en el año 2022, destacando que en términos generales la incidencia de los accidentes de tránsito fatales se sitúa en un 20.9%. Con base en los resultados del ajuste del modelo, se logró estimar la contribución de los factores asociados a la fatalidad de accidentes. El modelo clasificador, aplicado a través de una muestra representativa, fue sometido a la prueba de determinación de Nagelkerke del 80.6%. Además, la eficiencia de la predicción demostró un acurracy impresionante del 95%. Específicamente, en la predicción de accidentes fatales, el modelo alcanzo una precisión del 86% y una sensibilidad (recall) del 89%, indicando su capacidad para identificar correctamente la mayoría de casos fatales. En el caso de accidentes no fatales, la precisión es aún mayor, con un 97%, y un recall del 96%. La métrica F1, que equilibra la precisión y el recall, es del 87% para accidentes fatales y del 97% para accidentes no fatal. Además, la curva ROC con un AUC de 0.96 indica que el modelo tiene un rendimiento excelente en la clasificación de accidentes fatales y no fatales.

SEGUNDA: Mediante la aplicación de la técnica de análisis de componentes principales, reduciendo la dimensionalidad de las variables se logró identificar los elementos que aumentan el riesgo de fatalidad en accidentes de tránsito en la región de Puno. Esta metodología posibilitó la simplificación y clasificación de las variables pertinentes, entre las que se destacan los accidentes por clase, la causa de los accidentes, los vehículos

involucrados, el lugar de ocurrencia de los accidentes de tránsito, el día, las victimas muertas, los heridos y el género del conductor.

TERCERA: Los factores que inciden en la fatalidad de los accidentes de tránsito en la región de Puno son: X1: Clase de accidente, X2: Causas de los accidentes de tránsito, X3: Vehículos participantes, X4: Lugar de ocurrencia del accidente, X5: Incidencia diaria y X7: Victimas muertos, X8: Heridos y X9: Género del conductor. Se logró desarrollar el siguiente modelo:

$$Ln(odds) = -4.3398 - 0.9817x_1 - 1.0834x_2 + 0.8289x_3 + 1.6779x_4 - 0.836x_5 + 0.3876x_7 + 1.2369x_8 + 2.1454x_9$$

El modelo de regresión logística se estimó y se validó considerando los supuestos de linealidad, independencia y multicolinealidad. La evaluación de la linealidad mostró que el modelo presenta una pendiente pronunciada, indicando una relación lineal clara entre las variables predictoras y el logit de la probabilidad de un accidente fatal, confirmando así que el supuesto de linealidad se cumple adecuadamente. La prueba de independencia de Durbin-Watson arrojó un valor calculado de 1.993 y un p-valor de 0.902, indicando que no existe una correlación significativa entre los residuos, cumpliéndose el supuesto de independencia. En cuanto a la multicolinealidad, el promedio del Factor de Inflación de la Varianza (VIF) es de 2.95, sugiriendo una correlación moderada entre las variables predictoras, sin problemas significativos de multicolinealidad en el modelo. Estos resultados validan la idoneidad del modelo de regresión logística para predecir la probabilidad de accidentes fatales de manera precisa y consistente.

VI. RECOMENDACIONES

PRIMERA: A partir de los factores determinantes identificados, se podría enfocar en capacitar y concientizar a toda la población de la Región Puno. Es especialmente importante trabajar a través de las instituciones educativas para fortalecer la cultura vial desde temprana edad. Además, en los distritos con mayor incidencia de accidentes de tránsito, se podrían implementar campañas de seguridad vial (proyectos de inversión), con el objetivo de que estas iniciativas tengan un efecto multiplicador. Asimismo, sería útil modificar las normas vigentes para sancionar de manera más estricta las conductas relacionadas directamente con los factores determinantes hallados en la presente tesis.

SEGUNDA: Se recomienda a futuras investigaciones referentes a la fatalidad de accidentes de tránsitos construir modelos logísticos, pero con enlaces asimétricos, como cloglog, probit scobit, etc., las cuales muchas veces presentan mejores indicadores de los modelos de regresión logística (logit) tradicionales.

TERCERA: Se recomienda construir modelo por subpoblaciones, con el objetivo de obtener resultados más finos que permitan tomar decisiones más acertadas.

CUARTA: Se recomienda encarecidamente trabajar con bases de datos originales. Durante el desarrollo de este estudio, se proporcionó una tabla de frecuencias, lo cual requirió un proceso adicional de desglosamiento para obtener una base de datos adecuada para el análisis. Este proceso no solo fue laborioso, sino que también aumentó el riesgo de introducir errores o sesgos en los datos.

VII. REFERENCIAS BIBLIOGRÁFICAS

Aguirre, S., & Balarezo, A. (2022). *Análisis predictivo de la accidentalidad vehicular en el Distrito Metropolitano de Quito en el periodo 2015-2019.* Universidad Central del Ecuador, Quito. Obtenido de http://www.dspace.uce.edu.ec/bitstream/25000/27180/1/FCE-CEST-AGUIRRE%20CARLOS-BALAREZO%20ANAIZ.pdf

Arapa, K. (2019). *Identificación de los factores determinantes de la los accidentes de tránsitos fatales en las provincias de Arequipa, Caylloma e Islay 2013-2018.* Universidad Católica Santa María, Arequipa.

Avalos, A., Cárdenas, J., Caballero, R., & Ayma, V. (2022). Estiamción de la cantidad de herido en accidentes de tránsito dentre de la provincia de Lima utilizando modelo de regresión lineal múltiple y PCA. *Universidad Pacifico del Perú*.

Ayala, H., Valenzuela, G., & Espeinoza, A. (2020). Obtención de un modelo de minería de datos aplicado a la deserción universitaria del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca. *Revista Ontare*, 135-150. doi:https://journal.universidadean.edu.co/index.php/Revistao/article/view/2676/2087

Berlanga, S., Rubio, H., & Vila, B. (2013). Cómo aplicar árboles de decisión en SPSS. *Revista de Innovación e Investigación en Educación*, 79. doi:https://doi.org/10.1344/reire2013.6.1615

Berlanga, V., Rubio, M., & Vila, R. (20213). Cómo aplicar árboles de decisión en SPSS. *REIRE*, 15. Obtenido de https://doi.org/10.1344/reire2013.6.1615

Black, H., Babin, A., & Tatham. (2006). Multivariate Data Analysis. *Scientific Research*.

Bustamante. (2014). Modelos de variable dependiente discreta: El Modelo Logit y Probit. *UNMSM*, 31.

Canales, A. (2012). *Diagnostico ambiental regional (DAR) Puno.* Obtenido de https://sinia.minam.gob.pe/sites/default/files/siar-puno/archivos/public/docs/1307.pdf

Chen, M.-M., & Chen, M.-C. (2020). Modeling Road Accident Severity with Comparisons of Logistic Regression, Decision Tree and Random Forest. *MDPI*, 23.

Chipana, J. (2023). *Factores que influyen en los accidentes de tránsito ocacionados por el transporte público terrestre en Villa el Salvador, 2021*. Universidad Autónoma del Perú, Lima. Obtenido de https://repositorio.autonoma.edu.pe/bitstream/handle/20.500.13067/2273/Chipana%20Miranda%2C%20Jorge.pdf?sequence=1&isAllowed=y

Choque, J. (2020). *Estimación de los costos económicos indirectos de los accidentes de tránsito en el departemento de Puno periodo 2013-2017*. Universidad Nacional del Altiplano, Puno.

Choqueguanca, V., Cárdenas, F., & Mendoza, W. (2010). Perfil epidemiológico de los accidentes de tránsitos en el Perú, 2005-2009. *Med Exp Salud Publica*, 8. Obtenido de https://www.scielosp.org/pdf/rpmesp/v27n2/a02v27n2.pdf

Cruz, L. (2017). *Modelos predictivos de accidentes de tráfico en Madrid.* Universidad Internacional de la Rioja, Madrid. Obtenido de file:///C:/Users/Yonas/Downloads/CRUZ%20BELLAS,%20LUIS%20(1).pdf

Cuadras, C. (2014). *Nuevos métodos de multivariante.* Barcelona: CMC Editions.

Defensoría del Pueblo. (2022). *Defensoría del Pueblo: Cifra de accidentes de tránsitos en 2022 alcanza niveles registrados antes de la pandemia.* Lima.

EDUVIA. (2007). *Accidente, siniestro o incidente vial.* Obtenido de http://www.eduvia.com.ar/2010/02/08/accidente-siniestro-o-incidente-vial-¿cual-es-la-definicioncorrecta

Estrada, L., & Soto, S. (2021). *Análisis de la seguridad vial en la AV. Atahualpa, que une los distritos de Cajamarca y Baños del Inca, aplicando la metodología de inspección de seguridad vial y el método predictivo de manual HSM 2010, para la reducción de accidentes de tránsito 2021*. Universidad Privada del Norte, Cajamarca. Obtenido de file:///C:/Users/Yonas/Downloads/Estrada%20S%C3%A1nchez,%20Luz%20Marina%20-%20Soto%20D%C3%ADaz,%20Saira%20Patricia.pdf

Fawcett, T. (2005). https://doi.org/10.1016/j.patrec.2005.10.010. *ELSEVIER*, 861-874.

Fernández, C. (2023). *Reporte Defensorial de accidentes de tránsito.* Lima: Defensoria del Pueblo. Obtenido de https://www.defensoria.gob.pe/wp-content/uploads/2023/04/Reporte-Defensorial-de-accidentes-de-tr%C3%A1nsito-N01-Abril-2023.pdf

Fernández, C., & Baptista, P. (2014). *Metodología de la investigación.* Santa Fe: S. A. De C. V.

García, J. (2021). *Modelo de predicción de siniestros viales basados en redes bayesianas para corredores de la red vial arterial de la ciudad de Bogotá.* Universidad nacional de Colombia, Bogotá. Obtenido de https://repositorio.unal.edu.co/bitstream/handle/unal/81200/1.015.424.410.2021.pdf?sequence=3&isAllowed=y

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning Data Mining, Inference, and Prediction.* Springer.

INEI. (2015). *Perú III censo nacional de comisarias 2014.* Lima: Instituto Nacional de Estadística e Informatica. Obtenido de https://bvs.minsa.gob.pe/local/MINSA/3596.pdf

INEI. (2021). *Reporte de aciidentes de tránsito.* Lima. Obtenido de https://www.defensoria.gob.pe/wp-content/uploads/2021/12/Reporte-de-Adjunt%C3%ADa-de-seguridad-vial-j.pdf

Mamani, A., & Ponce, F. (2021). *Predicción de accidentes de tránsito aplicando el manual HSM 2010 en la carretera Puno-Ilave.* Universidad Nacional del Altiplano, Puno.

Menard, S. (2002). *Applied Logistic Regression Analysis* . London: Sage university paper.

Monroy, S., & Díaz, H. (2020). Modelo de predicción de gravedad de accidentes de tránsito: un análisis de los siniestros en Bogotá, Colombia. *Universidad Nacional de Colombia*, 20.

MTC. (2022). *Informe de víctimas fatales en siniestros de tránsito, nivel nacional, 2021.* Lima: Observatorio Nacional. Obtenido de

https://www.onsv.gob.pe/post/informe-de-victimas-fatales-de-siniestros-de-transito-nivel-nacional-2021/

Ninahuanca, Y. (2021). *Accidentabilidad y la seguridad vial en el Jr. Santa Isabel, El Tambo, Provincia Huancayo.* Universidad Peruana los Andes, Huancayo. Obtenido de https://repositorio.upla.edu.pe/bitstream/handle/20.500.12848/3088/T037_21288235_M.pdf?sequence=1&isAllowed=y

OMS. (2018). *los progresos han sido insuficientes en abordar la falta de seguridad en las vías de tránsito del mundo.* Washington: World Health Organization.

OMS. (13 de Diciembre de 2023). *Organización Mundial de la Salud*. Obtenido de https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries

ONSV. (2022). *Observatorio Nacional de Seguridad Vial*. Obtenido de https://www.onsv.gob.pe/

OPS. (2018). *Prevención de accidentes y lesiones.*

Organización Mundial de la Salud. (13 de Diciembre de 2023). Obtenido de https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries

Peña, D. (2002). *Análisis de datos multivariantes.* Madrid: University Carlos III.

Pérez, J. (2018). Accidentabilidad y rediseño de la carretera Poroy-Urubamba, aplicando el modelo de predicción de accidentes en vías rurales del manual norteamericano highway safety manual 2010. *Rev Yachay*, 8. Obtenido de file:///C:/Users/Yonas/Downloads/fmiranda,+P%C3%A9rez,+J.+(2018).+339-346%20(1).pdf

Perez, W. (2020). *Planteamiento de un modelo probabilístico para pronosticar riesgos de accidentes en la compañía minera Raura S. A.* Universidad Nacional Del Centro del Perú, Huancayo. Obtenido de https://repositorio.uncp.edu.pe/bitstream/handle/20.500.12894/6219/Tesis%20Wilfried%20Bryan%20PEREZ%20PARRAGUEZ.pdf?sequence=1&isAllowed=y

Pla, L. (1986). *Análisis multivariado : método de componentes principales* . Washington: Secretaria General de la Organizacion.

Quiroga, & Limon. (2011). *Estudio de la correlación entre las diferentes bolsas financieras en el mundo, usando el análisis multivariado (PCA y LDA).* Universidad de Guadalajara, Jalisco.

Quispe, C. (2024). *Localización y evaluación de los puntos críticos de accidentes de tránsito e la vía Puno-Ilave entre los años 2021-2022 y propuesta de medidas preventivas.* Universidad Nacional del Altiplano, Puno.

Rivas, M., Suárez, A., & Serebrisky, T. (2019). Hechos estilizados de transporte urbano en América Latina y el Caribe. *BID*, 14. Obtenido de file:///C:/Users/Yonas/Downloads/Hechos_estilizados_de_transporte_urbano_en_Am%C3%A9rica_Latina_y_el_Caribe_es_es.pdf

Ruiz, M. (2018). *Análisis de sensibilidad mediante Random Forest.* Politécnica, Madrid.

Silva, J., & Roman, N. (2021). Predicting Dropout in Higher Education: a Systematic Review. *Sociedad Brasileira de Computacao*.

SUTRAN. (2022). *Accidentes de tránsito ocurridos en carrteras* . Lima.

Tabasco, C. (2015). Paradigmas, teorías y modelos de la seguridad y la inseguridad vial. 74. Obtenido de http://94.23.80.242/~aec/ivia/tabasso_124.pdf

Valdés, P., Ferrer, N., & Ferrer, A. (1996). Accidentes en los niños, un problema de salud actual. *SiElo*, 12. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0864-21251996000300012

Velasquez, A. (2016). La carga de enfermedad y lesiones en el Perú y las prioridades del plan esencial de aseguramiento universal. *Scielo*, 222-231. Obtenido de http://www.scielo.org.pe/scielo.php?pid=S1726-46342009000200015&script=sci_abstract

Vellacorta, M. (2015). *Limitaciones en la recopilación y uso de la información de accidentes de tránsito en la Policía Nacional del Perú.* PUCP, Lima. Obtenido de http://hdl.handle.net/20.500.12404/6689

Zhang, Z. (2020). A Bayesian Network Incremental Algorithm for Public Safety Data Analysis. *University of Newcastle*, 4. Obtenido de https://sci-hub.se/10.1109/ICMCCE51767.2020.00410

ANEXOS

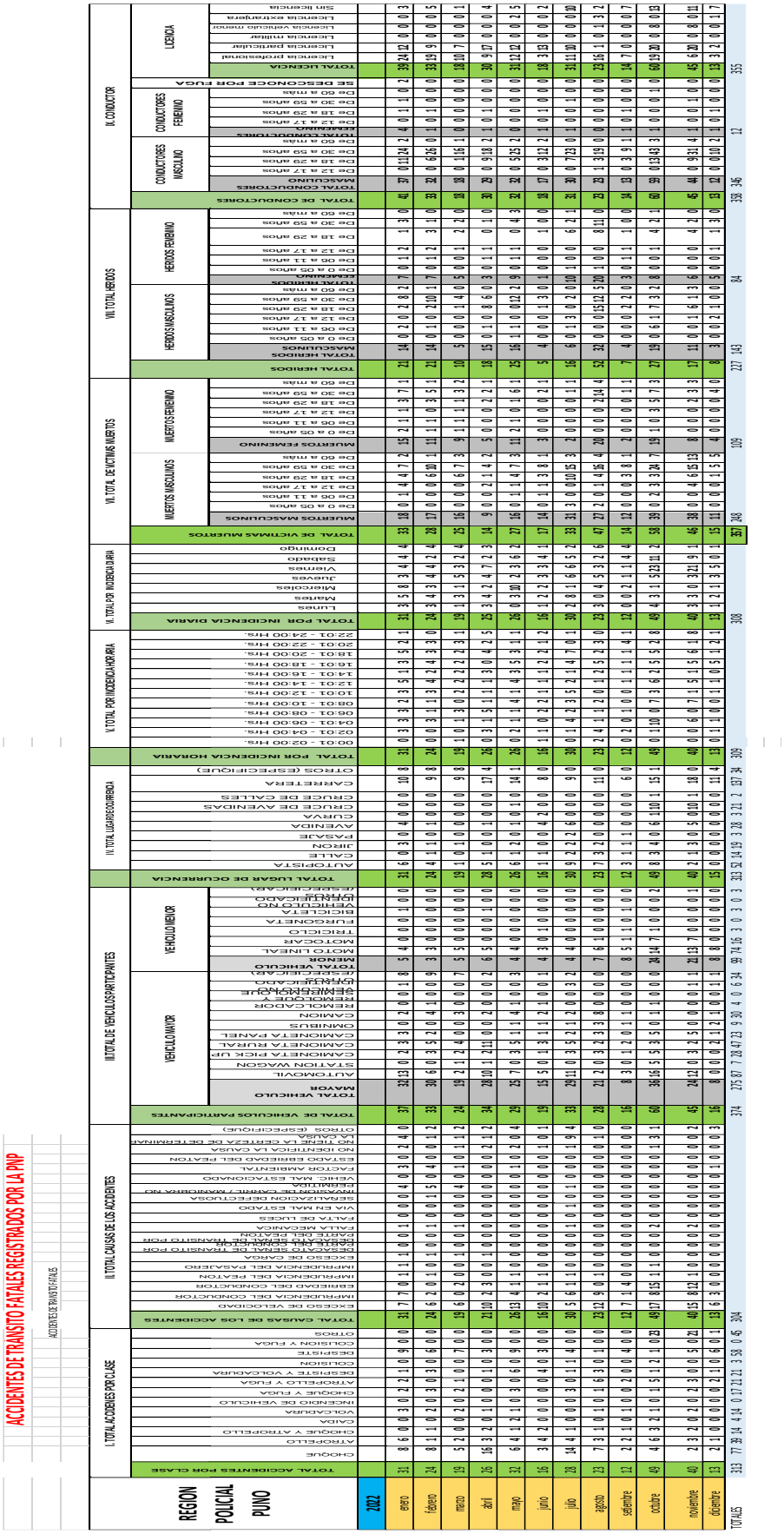
ANEXO 1. Código de la estimación del modelo de regresión logística

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split, cross_val_score, StratifiedKFold
from sklearn.metrics import classification_report, roc_auc_score, roc_curve, confusion_matrix
import matplotlib.pyplot as plt
from statsmodels.stats.outliers_influence import variance_inflation_factor
import numpy as np
import statsmodels.api as sm

# Ruta al archivo CSV
file_path = r'C:\Users\NEO\Documents\TESIS\PROBAR2.csv'

# Cargar los datos desde CSV
datos = pd.read_csv(file_path, sep=';')  # Ajusta el separador si es necesario

# Función para aplicar PCA y agregar las nuevas variables al DataFrame
def apply_pca(data, n_components=1):
    scaler = StandardScaler()
    data_scaled = scaler.fit_transform(data)
    pca = PCA(n_components=n_components)
    pca_result = pca.fit_transform(data_scaled)
    return pca, pca_result

# Aplicar PCA a cada grupo de variables y añadir al DataFrame
pca_accidentes_por_clase, datos['accidentes_por_clase_pca'] = apply_pca(datos.iloc[:, 0:13])
pca_causas_de_los_accidentes, datos['causas_de_los_accidentes_pca'] = apply_pca(datos.iloc[:, 13:29])
pca_vehiculos_participantes, datos['vehiculos_participantes_pca'] = apply_pca(datos.iloc[:, 29:44])
pca_lugar_de_ocurrencia, datos['lugar_de_ocurrencia_pca'] = apply_pca(datos.iloc[:, 44:54])
pca_incidencia_horaria, datos['incidencia_horaria_pca'] = apply_pca(datos.iloc[:, 54:66])
pca_incidencia_diaria, datos['incidencia_diaria_pca'] = apply_pca(datos.iloc[:, 66:73])
pca_victimas_muertos, datos['victimas_muertos_pca'] = apply_pca(datos.iloc[:, 73:75])
pca_heridos, datos['heridos_pca'] = apply_pca(datos.iloc[:, 75:77])
pca_conductor, datos['conductor_pca'] = apply_pca(datos.iloc[:, 77:80])

# Obtener nombres de las columnas PCA
columnas_pca = ['accidentes_por_clase_pca','causas_de_los_accidentes_pca','vehiculos_participantes_pca',
                'lugar_de_ocurrencia_pca', 'incidencia_diaria_pca',
                'victimas_muertos_pca', 'heridos_pca', 'conductor_pca']

# Crear un nuevo DataFrame solo con las variables PCA
variables_pca = datos[columnas_pca]

# Calcular el VIF para un conjunto de componentes principales
def calculate_vif(X):
    vif_data = pd.DataFrame()
    vif_data["Variable"] = X.columns
    vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in range(len(X.columns))]
    return vif_data
```

```
# Mostrar el VIF de todas las variables
print("\nVIF de todas las variables:")
todos_vif = calculate_vif(variables_pca)
print(todos_vif)

# Calcular y mostrar el promedio de los VIF
promedio_vif = np.mean(todos_vif["VIF"])
print(f"\nPromedio de los VIF de todas las variables: {promedio_vif}")

# Dividir los datos en entrenamiento y prueba
acctes = datos['ACCTES']
X_train, X_test, y_train, y_test = train_test_split(variables_pca, acctes, test_size=0.3, random_state=42)

# Inicializar y entrenar el modelo de regresión logística con todas las variables PCA
model_lr = LogisticRegression()
model_lr.fit(X_train, y_train)

# Resumen del modelo de regresión logística con todas las variables PCA
# Añadir constante al conjunto de entrenamiento para el modelo de statsmodels
X_train_const = sm.add_constant(X_train)

# Inicializar y ajustar el modelo de regresión logística con statsmodels
model_sm = sm.Logit(y_train, X_train_const).fit()

# Mostrar el summary del modelo de regresión logística con todas las variables PCA
print("\nSummary del modelo de regresión logística con todas las variables PCA:")
print(model_sm.summary())

# Evaluar el modelo de regresión logística con todas las variables PCA
accuracy_lr = model_lr.score(X_test, y_test)
print(f'\nAccuracy del modelo de regresión logística con todas las variables PCA: {accuracy_lr}')

# Predicciones y métricas de regresión logística con todas las variables PCA
y_pred_lr = model_lr.predict(X_test)
print("\nClassification Report del modelo de regresión logística con todas las variables PCA:")
print(classification_report(y_test, y_pred_lr))

# Calcular la matriz de confusión del modelo de regresión logística con todas las variables PCA
conf_matrix_lr = confusion_matrix(y_test, y_pred_lr)
print("\nMatriz de confusión del modelo de regresión logística con todas las variables PCA:")
print(conf_matrix_lr)

# Calcular probabilidades para la curva ROC del modelo de regresión logística con todas las variables PCA
y_probs_lr = model_lr.predict_proba(X_test)[:, 1]

# Calcular el área bajo la curva ROC (AUC) del modelo de regresión logística con todas las variables PCA
auc_lr = roc_auc_score(y_test, y_probs_lr)
print(f"\nAUC del modelo de regresión logística con todas las variables PCA: {auc_lr}")

# Calcular la curva ROC del modelo de regresión logística con todas las variables PCA
fpr_lr, tpr_lr, thresholds_lr = roc_curve(y_test, y_probs_lr)
```

```
# Graficar la curva ROC del modelo de regresión logística con todas las variables PCA
plt.figure(figsize=(8, 6))
plt.plot(fpr_lr, tpr_lr, color='blue', label=f'ROC Curve - Logistic Regression (AUC = {auc_lr:.2f})')
plt.plot([0, 1], [0, 1], color='red', linestyle='--', label='Random Guessing')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic (ROC) Curve - Logistic Regression with PCA Variables')
plt.legend()
plt.grid(True)
plt.show()

# Validación cruzada con k=5 y k=10
kfold_5 = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
kfold_10 = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)

cv_scores_5 = cross_val_score(LogisticRegression(max_iter=1000, solver='liblinear', penalty='l2', C=1.0),
                              variables_pca, acctes, cv=kfold_5, scoring='roc_auc')
cv_scores_10 = cross_val_score(LogisticRegression(max_iter=1000, solver='liblinear', penalty='l2', C=1.0),
                               variables_pca, acctes, cv=kfold_10, scoring='roc_auc')

print(f"\nAUC-ROC (Cross-Validation, k=5): {cv_scores_5.mean()} (std: {cv_scores_5.std()})")
print(f"AUC-ROC (Cross-Validation, k=10): {cv_scores_10.mean()} (std: {cv_scores_10.std()})")

# Calcular e imprimir el logaritmo de la verosimilitud
log_likelihood = model_sm.llf
print(f"\nLogaritmo de la verosimilitud: {log_likelihood}")

# Calcular e imprimir el R cuadrado de Cox y Snell
cox_snell_r2 = model_sm.prsquared
print(f"R cuadrado de Cox y Snell: {cox_snell_r2}")

# Calcular e imprimir el R cuadrado de Nagelkerke
# Se calcula usando la fórmula: R2_N = R2_C&S / (1 - exp(-L(0) / n))
null_log_likelihood = model_sm.llnull
n = len(y_train)
nagelkerke_r2 = cox_snell_r2 / (1 - np.exp(-null_log_likelihood / n))
print(f"R cuadrado de Nagelkerke: {nagelkerke_r2}")

# Prueba de Wald
print("\nPrueba de Wald para cada variable:")
print(model_sm.wald_test_terms())

# Prueba de bondad de ajuste de Hosmer-Lemeshow
from scipy.stats import chi2

# Crear la tabla de contingencia de Hosmer-Lemeshow
def hosmer_lemeshow_test(model, y, X, g=10):
    data = pd.DataFrame({'prob': model.predict(X), 'actual': y})
    data['group'] = pd.qcut(data['prob'], g, duplicates='drop')
    observed = data.groupby('group')['actual'].agg(['sum', 'count'])
    observed
```

ANEXO 2. Base de datos



ANEXO 3. Probabilidad de accidentes de tránsito para el caso 20

Caso 20					
Probabilidad		**Accidente fatal**	**0.738088**		
		Accidentes no fatal	**0.261912**		
Accidentes por clase			Vehículos Participantes		
Características	Dummy	Valores PCA	Características	Dummy	Valores PCA
Atropello	0		Automóvil	1	
Caída	0		Bicicleta	0	
Choque	0		Camión	0	
Choque y atropello	1		Camioneta panel	0	
Choque y fuga	0		Camioneta pick up	0	
Colisión	0		Camioneta rural	0	
Colisión y fuga	0	1.548	Moto lineal	0	
Despiste	0		Moto car	0	1.496
Despiste y volcadura	0		Ómnibus	0	
Incendio de vehículo	0		Remolcador	0	
Volcadura	0		Remolque y semirremolque	0	
Atropello y fuga	0		Station wagon	0	
Otros	0		Triciclo	0	
Causas de los accidentes			Vehículo no identificado	0	
Características	Dummy	Valores PCA	Lugar de ocurrencia		
Imprudencia del peatón	0		Características	Dummy	Valores PCA
Desacato señal de tránsito por parte del conductor	0		Autopista	1	
Ebriedad del conductor	0	1.768	Avenida	0	
Exceso de carga	1		Calle	0	-0.773
Exceso de velocidad	0		Carretera	0	
Factor ambiental	0		Cruce de avenidas	0	

Falla mecánica	0	
Falta de luces	0	
Imprudencia del conductor	0	
Imprudencia del pasajero	0	
Invasión de carril / maniobra no permitida	0	
Señalización defectuosa	0	
Vía en mal estado	0	
No identifica la causa	0	
No tiene la certeza de determinar la causa	0	
Otros	0	

Victimas Muertos		
Características	Dummy	Valores PCA
Muertos Femeninos	0	-1.928
Muertos Masculinos	1	

Heridos		
Características	Dummy	Valores PCA
Heridos Femeninos	0	1.953
Heridos Masculinos	0	

Conductor		
Características	Dummy	Valores PCA
Femenino	0	2.305
Masculino	1	
Se desconoce por fuga	0	

Cruce de calles	0	
Curva	0	
Jiron	0	
Pasaje	0	
Otro	0	

Incidencia Diaria		
Características	Dummy	Valores PCA
Lunes	1	-0.006
Martes	0	
Miércoles	0	
Jueves	0	
Viernes	0	
Sábado	0	
Domingo	0	

ANEXO 4. Carta policial para la obtención de datos

CARTA POLICIAL

SEÑOR(A) : Marizol Lizbeth SERRANO QUISPE.

REF. : Solicitud presentada por Marizol Lizbeth SERRANO QUISPE de fecha 18SET23

1.Mediante la presente, se hace de su conocimiento que habiéndose recibido su solicitud de fecha 18 de setiembre del 2023, donde solicita información sobre accidentes de tránsito (2018-2023), de la ciudad de Puno.

2.Se tiene que el TUO de la ley de transparencia y acceso a la información pública de la ley N° 27806, aprobado por D.S N° 021-2019-JUS, en su artículo 10° regula que "Las entidades de la administración pública tiene la obligación de proveer la información requerida si se refiere a la contenida en documentos escritos, fotografías, grabaciones, soporte magnético o digital, o en cualquier otro formato, siempre que haya sido creada u obtenida por ella o que se encuentre en posesión o bajo su control. Asimismo, para los efectos de esta Ley, se considera como información pública cualquier tipo de documentación financiada por el presupuesto público que sirva de base a una decisión de naturaleza administrativa, así como las actas de reuniones oficiales."; El artículo 13° De la acotada norma, establece que : "La solicitud de información no implica la obligación de las entidades de la Administración Publica de crear a producir información con la que no cuente o no tenga obligación de contar al momento de efectuarse el pedido. En este caso, la entidad de la Administración Publica deberá comunicar por escrito que la denegatoria de la solicitud se debe a la inexistencia de datos en su poder respecto de la información solicitada"; y que: "Esta ley tampoco permite que los solicitantes exijan a las entidades que efectúen evaluaciones o análisis de la información que poseen".(...), por otro lado, es propia la norma que regula excepciones establecidos en los artículos 15°, 16° y 17°, cuando la información se considera secreta, reservada o confidencial.

3. Que conforme a los numerales 4 y 6 del artículo 87 del Decreto Supremo N° 026-2017-IN, que aprueba el reglamento de la ley de la PNP, se tiene que entre otras funciones de la División de Estadística que integra la Dirección de Tecnología dela información y Comunicaciones, se encuentre respectivamente el "Administrar el contenido y el tratamiento de la información que se requiera para el sistema estadístico policial y otros relacionados a la función policial, en el ámbito de su competencia; " Y el "formular y proponer informes, boletines y anuarios de la información estadística de la Policía Nacional del Perú;". Por su parte el numeral 15 del artículo 208 del mismo cuerpo legal, establece como una de las funciones de las jefaturas de las Macro Regiones Policiales el "Dirigir y supervisar el proceso de registro, recopilación y análisis de la información estadística que produzca la Macro Región Policial a su cargo, para una adecuada toma de decisiones de conformidad con los lineamientos que dicte la División de Estadística de la Policía Nacional del Perú".

4.Conforme a lo anterior, se tiene que la información a la que pretende acceder no estaría incursa dentro de las excepciones reguladas por la Ley de Transparencia y acceso a la información pública; por lo que su entrega resulta **PROCEDENTE.**

5.En ese sentido se hace de su conocimiento que la información estadística requerida se remitió a través del correo electrónico mlserramoq28@gmail.com, de acuerdo a la solicitud de la referencia.

Aprovecho la oportunidad para expresarle los sentimientos de mi especial consideración y estima personal.

Puno 28 de setiembre del 2023

FHAM/app.

POLICIA NACIONAL DEL PERU

PNP 297632
Freddy Henry ARELLANO MENDOZA
CMDTE. PNP
JEFE UNIPLEDU X MACREPOL PUNO

Universidad Nacional del Altiplano Puno | Vicerrectorado de Investigación | Repositorio Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Marizol Lizbeth Serrano Quispe,
identificado con DNI 75374500 en mi condición de egresado de:

☑ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ingenieria Estadística e Informática,

informo que he elaborado el/la ☑ **Tesis** o ☐ **Trabajo de Investigación** denominada:

"ANÁLISIS DEL MODELO DE PREDICCIÓN EN LA FATALIDAD DE ACCIDENTES DE TRÁNSITO EN LA REGIÓN DE PUNO, 2022"

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 23 de Julio del 2024

FIRMA (obligatoria)

Huella

Universidad Nacional del Altiplano Puno | Vicerrectorado de Investigación | Repositorio Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Marizol Lizbeth Serrano Quispe,
identificado con DNI 75374500 en mi condición de egresado de:

☑ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ingenieria Estadística e Informática,

informo que he elaborado el/la ☑ **Tesis** o ☐ **Trabajo de Investigación** denominada:

" ANÁLISIS DEL MODELO DE PREDICCIÓN EN LA FATALIDAD DE ACCIDENTES DE TRÁNSITO EN LA REGIÓN DE PUNO, 2022 "

para la obtención de ☐**Grado,** ☑ **Título Profesional** o ☐ **Segunda Especialidad.**

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: https://creativecommons.org/licenses/by-nc-sa/4.0/

En señal de conformidad, suscribo el presente documento.

Puno 23 de Julio del 2024

FIRMA (obligatoria)

Huella