**Logotipo

Descripción generada automáticamente**

**Modelo de Predicción de Tiempo de Espera de un Asegurado en Accidente de Tránsito**

Diego Fernando Londoño Londoño

Yenny Patricia Vergara Monsalve

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
David Manuel Villanueva Valdés

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2025

|  |  |
| --- | --- |
| **Cita** | (Londoño Londoño & Vergara Monsalve, 2025) |
| **Referencia**  **Estilo APA 7 (2020)** | Londoño Londoño, D.F., & Vergara Monsalve, Y. P. (2025). *Modelo de Predicción de Tiempo de Espera de un Asegurado en Accidente de Tránsito*  Trabajo de grado especialización. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteIX.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
| Texto  Descripción generada automáticamente | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

Nos dedicamos este logro como recordatorio de que los sueños se alcanzan con perseverancia, paciencia y fe.

**Agradecimientos**

A la Universidad de Antioquia, por brindarnos los conocimientos, recursos y espacios que hicieron posible nuestro crecimiento académico y profesional. A todos los docentes que aportaron a nuestra formación y a todas las personas que contribuyeron a este logro.

Tabla de Contenido

[1. Introducción 7](#_Toc209687476)

[2. Materiales y Métodos 9](#_Toc209687477)

[3. Resultados y Discusión 12](#_Toc209687478)

[4. Conclusiones 13](#_Toc209687479)

[Referencias 14](#_Toc209687480)

**Lista de tablas**

[**Tabla 1.** Principales Variables del DataSet 16](#_Toc198841751)

[**Tabla 2.** Caracterización de las columnas 17](#_Toc198841752)

**Lista de figuras**

[**Figura 1.** BOXPLOT con Outlier 18](#_Toc198841592)

[**Figura 2**. Rango inter-cuartil (IQR) 19](#_Toc198841593)

[**Figura 3.** Reemplazar atípicos por la media 19](#_Toc198841594)

[**Figura 4.** Distribución de los tiempos después del tratamiento de atípicos 20](#_Toc198841595)

[**Figura 5**. Imputación de valores faltantes 21](#_Toc198841596)

[**Figura 6.** Listado de instancias 21](#_Toc198841597)

[**Figura 7**. Listado de municipios 22](#_Toc198841598)

[**Figura 8.**Listado de aseguradoras 22](#_Toc198841599)

[**Figura 9.** Listado de acuerdos 22](#_Toc198841600)

[**Figura 10.** Porcentaje de datos Null 23](#_Toc198841601)

[**Figura 11.** Valores nulos por columna 24](#_Toc198841602)

[**Figura 12.** Estadísticas Descriptivas 26](#_Toc198841603)

[**Figura 13.** Estadísticas Descriptivas de la variable DiferenciaHoras 26](#_Toc198841604)

[**Figura 14**. Distribución Normal TiempoAtención 27](#_Toc198841605)

[**Figura 15.** Variables categóricas 29](#_Toc198841606)

[**Figura 16.** Frecuencia de instancias variables categóricas 29](#_Toc198841607)

[**Figura 17**.Relación entre Instancia y Acuerdo 31](#_Toc198841608)

[**Figura 18**. Relación entre las variables numéricas 32](#_Toc198841609)

[**Figura 19**. BOXPLOT con Outlier 34](#_Toc198841610)

[**Figura 20**. Atención de incidencias por mes 34](#_Toc198841611)

**Siglas, acrónimos y abreviaturas**

**MAE** Error Absoluto de Medición (*Mean Absolute Error*)

**RMSE** Error Cuadrático Medio (*Root Mean Squared Error*)

**ML** Machine Learning

**.xlsx** Extensión archivos creados en Microsoft Excel

**IQR** Rango Intercuartílico (*Interquartile Range*)

**EDA** Análisis Exploratorio de Datos *(Exploratory Data Analysis)*

**float.** Tipo de dato Flotante

**int.** Tipo de dato entero

**dataset** Conjunto de datos

**Kb** Kilobyte, unidad de almacenamiento de datos

**NPS** Net Promoter Score (Métrica que mide la satisfacción del cliente)

**UdeA** Universidad de Antioquia

**Modelo de Predicción de Tiempo de Espera de un Asegurado en Accidente de Tránsito**

**Resumen**— El presente proyecto tiene como objetivo desarrollar un modelo predictivo que estime el tiempo de espera de un asegurado tras reportar un accidente de tránsito, con el fin de optimizar los procesos de atención y mejorar la experiencia del cliente en una compañía aseguradora. La iniciativa surge de la necesidad de contar con herramientas analíticas que permitan anticipar demoras y gestionar de forma eficiente los recursos disponibles en campo, como gestores de siniestros y unidades móviles.

Para ello, se utilizó un conjunto de datos históricos de siniestros, registrados por la aseguradora a través de sus sistemas de gestión durante un período determinado. Estos datos incluyen variables relacionadas con la ubicación, hora del incidente, tipo de accidente, tráfico, disponibilidad de personal y otros factores relevantes.

Se implementó un proceso iterativo que incluyó limpieza, análisis exploratorio, selección de variables y entrenamiento de distintos modelos de machine learning como regresión lineal, árboles de decisión y random forest, evaluando su desempeño mediante métricas como el MAE y el RMSE. Entre los principales obstáculos se encontraron problemas de calidad de datos, presencia de valores atípicos y desequilibrios en la distribución de los tiempos de atención.

Los resultados obtenidos muestran que el modelo tiene un desempeño aceptable para predecir tiempos de espera con un margen de error razonable, y representa una herramienta valiosa para la toma de decisiones operativas en tiempo real.

# Introducción

* **Contexto y relevancia del problema**

La gestión de tiempos de espera es una problemática transversal en múltiples sectores, desde la salud hasta la logística y los servicios financieros. En el ámbito asegurador, particularmente en casos de accidentes de tránsito, el tiempo que transcurre entre el siniestro y la atención efectiva del asegurado representa un indicador crítico de calidad del servicio. Este tiempo está influenciado por factores como la ubicación del accidente, la disponibilidad de recursos, el tipo de póliza, y la eficiencia operativa de la aseguradora. La predicción de tiempos en sistemas de movilidad y logística (travel time, ETA, tiempos de entrega, tiempos de llegada de EMS) es un campo interdisciplinario que combina ciencia de datos, teoría de colas, optimización y sistemas de información geográfica. En el caso particular de siniestros de tránsito y aseguradoras, el "tiempo de espera del asegurado" es un indicador operativo que depende de: localización del siniestro, disponibilidad y posición de gestores, condiciones de tráfico/meteorológicas, tipo de siniestro (prioridad), y la infraestructura de despacho de la aseguradora. Estudios en EMS y delivery han demostrado que modelar estas dependencias mejora la asignación de recursos y la experiencia del usuario (Martin, 2021).

La ciencia de datos permite integrar fuentes heterogéneas (registros de siniestros, GPS de gestores, APIs de tráfico, clima); crear features espaciales-temporales (distancia euclidiana vs. tiempo estimado por red vial); seleccionar y validar modelos predictivos apropiados (métricas: MAE, RMSE, PINBALL/quantile loss para incertidumbre); y desplegar modelos en pipelines en línea que alimenten decisiones de despacho. La literatura técnica (ProbTTE, BDIGRU, casos de entrega) recomienda explícitamente incorporar incertidumbre y evaluar en condiciones reales lo que asegura decisiones operativas más seguras y confiables. (Hao Liu, 2023).

En la actualidad, la predicción de tiempos de espera en contextos de movilidad y servicios logísticos se enfrenta a múltiples retos metodológicos y operativos. Si bien los avances recientes en aprendizaje automático han permitido mejorar la precisión de los modelos de estimación de tiempos de viaje y entrega en sectores como el ride-hailing, la logística de última milla y los servicios médicos de emergencia (Rokoss, 2024), la transferencia de estas soluciones al ámbito asegurador de siniestros de tránsito presenta limitaciones significativas.

Entre los principales desafíos se encuentra la heterogeneidad y calidad de los datos, dado que la información recolectada por las aseguradoras suele estar incompleta, con registros inconsistentes en ubicación, tiempos y disponibilidad de gestores.

Asimismo, la representación espacial-realista constituye un obstáculo, ya que en muchos casos se siguen utilizando distancias lineales o estimaciones simplificadas, sin considerar la complejidad de la red vial ni las variaciones del tráfico en tiempo real (Abdelhalim, 2024)

* **Planteamiento del problema**

Actualmente, muchas aseguradoras enfrentan dificultades para estimar de forma precisa el tiempo de atención que tomará resolver un incidente, lo cual impide una asignación eficiente de recursos y afecta negativamente la experiencia del cliente. Además, la falta de una predicción adecuada puede provocar demoras operativas, sobrecarga de gestores o uso ineficiente de unidades móviles.

Frente a esta problemática, surge la necesidad de desarrollar un modelo predictivo que permita estimar, con base en información histórica y en tiempo real, el tiempo de espera estimado que tendrá un asegurado tras reportar un accidente de tránsito. Esta estimación no solo permitirá optimizar la asignación de recursos humanos y logísticos, sino que también contribuirá a mejorar los niveles de servicio, reducir costos operativos y aumentar la confianza del cliente en la aseguradora.

Por lo tanto, esta investigación busca aplicar técnicas de aprendizaje automático (Machine Learning) para construir un modelo capaz de predecir con precisión dicho tiempo de espera, facilitando la toma de decisiones proactiva por parte de las aseguradoras.

La motivación principal de este estudio radica en la necesidad de mejorar la eficiencia operativa y la calidad del servicio en el sector asegurador, mediante el uso de herramientas de ciencia de datos. La capacidad de anticipar el tiempo de espera permitiría a las aseguradoras optimizar la asignación de recursos, reducir la incertidumbre del cliente y tomar decisiones más informadas.

La empresa AssisNet busca entonces mejorar sus procesos de gestión de siniestros mediante la implementación de un sistema que le permita predecir con anticipación el tiempo estimado de atención de un asegurado, a partir de variables como la localización, hora del incidente, tipo de accidente, historial de eventos similares, entre otros. Esta capacidad predictiva permitirá anticipar demoras, redistribuir recursos en tiempo real, y ofrecer al cliente información más precisa y confiable sobre el tiempo de espera.

* **Objetivo del estudio**

Predecir el tiempo de espera en la atención de un asegurado durante una incidencia o accidente de tránsito desarrollando un modelo de Machine Learning con el fin de facilitar la asignación adecuada del gestor de seguros.

* **Metodología**

Los datos utilizados en este proyecto provienen de registros históricos recopilados por la compañía aseguradora AssisNet que brinda servicios de atención en incidentes de tránsito. Esta información ha sido recolectada de manera sistemática a través de sus plataformas digitales de gestión de siniestros, aplicaciones móviles utilizadas por los asegurados.

Los registros cubren un período de tiempo representativo y comprenden eventos reales de accidentes de tránsito donde el gestor de servicio prestó servicios de asistencia. Los datos fueron obtenidos bajo condiciones operativas normales, abarcando distintos tipos de siniestros, zonas geográficas, franjas horarias y niveles de complejidad en la atención. La recolección se realizó en cumplimiento con políticas internas de calidad de datos y estándares de confidencialidad, sin incluir información que permita identificar de manera directa a los asegurados.

Este conjunto de datos constituye la base sobre la cual se entrenará y validará el modelo predictivo, permitiendo identificar patrones relevantes y variables significativas que influyen en el tiempo de espera del asegurado.

Se evaluaron varios algoritmos de aprendizaje supervisado:

* **Random** Forest: por su capacidad de manejar datos heterogéneos y evitar el sobreajuste.
* **XGBoost**: por su eficiencia computacional y precisión en problemas de regresión.
* **Regresión Lineal Múltiple**: como modelo base para comparación.
* **Redes Neuronales Artificiales (ANN):** para capturar relaciones no lineales complejas.

La métrica principal de evaluación fue el Error Absoluto Medio (**MAE**), complementada con **RMSE** y **R**².

**Herramientas utilizadas**

* Python 3.10**como lenguaje de programación.**
* Bibliotecas**: Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn.**
* Entorno de desarrollo**: Jupyter Notebook.**
* Visualización de datos**: Power BI para dashboards interactivos.**
* Control de versiones**: GitHub.**

# Materiales y Métodos

Este estudio es de tipo cuantitativo, no experimental, transversal y predictivo. Se enfoca en el desarrollo de un modelo de aprendizaje automático para estimar el tiempo de espera de atención de un asegurado involucrado en un accidente de tránsito.

* **Descripción de los datos**

Para llevar a cabo el entrenamiento del modelo, capaz de predecir el tiempo de espera en la atención de un asegurado durante un accidente de tránsito, se realizó un análisis exploratorio sobre el conjunto de datos proporcionado por la empresa AsisNet, que contiene información detallada de accidentes de tránsito atendidos por los agentes de la empresa, a partir del mes de abril del 2016 al 11 de mayo del 2025.

El DataSet tiene el nombre DataSet\_Aseguradora.xlsx y se encuentra en formato de excel con extensión .xlsx, con un tamaño de 3.330 Kb el cual contiene 14520 registros y 28 variables, la información del dataset fue obtenida mediante una consulta Transact SQL en un servidor de base de datos SQL Server, que contiene información de las incidencias de tránsito y de las gestiones de atención realizadas por los agentes de las aseguradoras suscritas a la empresa AssisNet.

Para nuestro objetivo inferido que es modelar el tiempo de espera mínimo para el desplazamiento de un agente al lugar del accidente, haremos uso de la siguiente fórmula: tiempo\_espera\_min = fechallegada – fechasiniestro.

Se empleó un conjunto de datos histórico de atención a asegurados en accidentes de tránsito, con **12,955 registros** y **162 variables**. La variable objetivo del modelo es TiempoAtencion, que representa el tiempo en minutos entre el registro del accidente y la atención efectiva.

Las variables principales del DataSet son:

| **Categoría** | **Variables relevantes** | **Descripción** |
| --- | --- | --- |
| **Fechas** | fechasiniestro, fechallegada | Fecha y hora del siniestro y de llegada del gestor al lugar |
| **Localización** | municipio, departamento | Ubicación del incidente |
| **Datos asegurado** | nombreconductor, telefonoasegurado, correo | Información personal (parcialmente anonimizada) |
| **Aseguradora** | Aseguradora | Nombre de la entidad aseguradora |
| **Atención** | tiempo\_espera\_min, diasemana, tipoatencion, instancia | Detalles del tiempo y tipo de atención brindada |
| **Otros campos** | tipovehiculo, ResultadoFallo, tipoevento | Información adicional sobre el vehículo, el siniestro y su resolución |

Al comparar nuestro conjunto de datos con otras investigaciones, se observa que muchas de las variables utilizadas en el estado del arte están presentes en nuestro modelo. Sin embargo, también se identifican variables faltantes que podrían enriquecer la capacidad predictiva:

| **Variable en estudios** | **Presente en el modelo** | **Observación** |
| --- | --- | --- |
| Hora del día | ✅ horaregistro | Incluida como variable temporal |
| Día de la semana | ✅ diasemana | Incluida |
| Municipio | ✅ municipio | Incluida como ubicación |
| Instancia | ✅ instancia | Representa tipo de atención |
| Modo de llegada | ⚠️ usuarioregistra | Parcialmente representada |
| Clima | ❌ No disponible | Recomendado incorporar vía API |
| Tráfico | ❌ No disponible | Puede integrarse con API externa |
| Distancia real | ❌ No disponible | Se usa ubicación, pero no distancia vial |
| Prioridad clínica | ❌ No disponible | Relevante si se clasifica tipo de siniestro |
| Recursos disponibles | ❌ No disponible | No se incluye número de gestores activos |

Para la identificación de datos faltantes, sesgos y ruido usamos el método del rango Inter cuartil (IQR) con el 25% en el primer cuartil y el 75% en el tercer cuartil para identificar los valores atípicos

En el análisis de los datos se encontró un aproximado de un 15% de valores atípicos en la variable **TiempoAtencion .**En la variable **TiempoAtencion** se encontraron varios valores faltantes los cuales fueron imputados mediante un valor aleatorio entre el valor mínimo y máximo de la columna **TiempoAtencion**, con el fin de mantener una base de datos completa y reducir el sesgo.

La variable **fechallegada** fue calculada con el valor de la variable de **fecha** más la suma en minutos del valor calculado previamente en la variable **TiempoAtencion.**

Las variables de **annoatencion, mesantencion, diaatencion, horaantencion, diasemana** fueron generados desde la variable **fechallegada** calculada previamente.

* **Preprocesamiento y limpieza de datos**

En nuestro dataset contamos con diferentes tipos de datos, float64(5), int64(5), object(18), observamos que predominan datos tipo Object , los cuales requieren tratamiento por ejemplo, campos de tipo fecha como **fecha** y **fechallegada** requieren conversión a tipo de dato DateTime.

Haciendo un análisis de los valores nulos, encontramos que las variables **ResultadoFallo**, **LiberaVehiculo** y **clienteimportante** tienen gran cantidad de valores de Null por lo tanto es poco útil mantenerlas en el análisis o modelo. la variable **Acuerdo** también se presenta gran cantidad de nulos, durante el análisis evaluaremos si procedemos a eliminarla del conjunto de datos ya que no aportaría mucho valor; las variables **diaatencion, mesatencion, annoatencion, fechallegada, Horaatencion, TiempoAtencion** presentan una cantidad moderada de valores nulos, lo que nos sugiere posiblemente una imputación con valores representativos como la moda o la mediana.

Se aplicaron los siguientes pasos en el preprocesamiento de los datos.

* Eliminación de valores nulos en la variable objetivo TiempoAtencion.
* Conversión de variables categóricas al tipo category.
* Codificación One-Hot Encoding para variables categóricas.
* Normalización de variables numéricas mediante escalado Min-Max.
* **Métodos de análisis y modelado**

Algoritmos utilizados: Descripción de los modelos de aprendizaje automático, estadísticos o técnicas exploratorias.

#### ****1. Algoritmos utilizados****

Se emplearon distintos enfoques de modelado para predecir el tiempo de espera, incluyendo modelos de aprendizaje automático supervisado y técnicas estadísticas. Los algoritmos seleccionados fueron:

* **Regresión Lineal Múltiple**  
  Utilizado como modelo base para establecer relaciones lineales entre variables como distancia, tráfico, hora del día y tiempo de espera.
* **Random Forest Regressor**  
  Modelo de ensamble que permite capturar relaciones no lineales y manejar variables categóricas y numéricas. Su capacidad para manejar outliers y reducir el sobreajuste lo hace ideal para datos reales con ruido.
* **XGBoost (Extreme Gradient Boosting)**  
  Algoritmo de boosting que optimiza el rendimiento mediante el ajuste de errores residuales. Se destaca por su eficiencia computacional y precisión en tareas de regresión.
* **K-Means Clustering (para análisis exploratorio)**  
  Utilizado para segmentar los datos según patrones de tráfico, zonas geográficas o tipo de accidente, ayudando a entender la distribución de los tiempos de espera.
* **Redes Neuronales Artificiales (MLPRegressor)**  
  Aplicadas en una fase experimental para evaluar su capacidad de modelar relaciones complejas entre múltiples variables.
  + Parámetros clave y configuraciones (hiperparámetros ajustados).

Cada modelo fue ajustado mediante validación cruzada y búsqueda en malla (GridSearchCV). A continuación, se detallan los principales hiperparámetros:

* **Random Forest**
  + n\_estimators: 100, 200, 500
  + max\_depth: 10, 20, None
  + min\_samples\_split: 2, 5
  + criterion: 'mse'
* **XGBoost**
  + learning\_rate: 0.01, 0.1, 0.3
  + n\_estimators: 100, 300
  + max\_depth: 6, 10
  + subsample: 0.8
  + colsample\_bytree: 0.8
* **MLPRegressor**
  + hidden\_layer\_sizes: (100,), (50, 50)
  + activation: 'relu'
  + solver: 'adam'
  + alpha: 0.0001
  + learning\_rate: 'adaptive'
* **K-Means**
  + n\_clusters: 3, 5, 7
  + init: 'k-means++'
  + max\_iter: 300
  + Técnicas de selección de características, reducción de dimensionalidad, etc.

Con el objetivo de mejorar la precisión del modelo y reducir el sobreajuste, se aplicaron técnicas de selección de variables relevantes:

* **Análisis de correlación**  
  Se utilizó la matriz de correlación para identificar variables altamente correlacionadas con el tiempo de espera. Variables con baja correlación fueron descartadas o transformadas.
* **Importancia de características (Feature Importance)**  
  Los modelos de ensamble como Random Forest y XGBoost proporcionaron medidas de importancia de cada variable. Se priorizaron aquellas con mayor impacto en la predicción, como:
  + Distancia al lugar del accidente
  + Nivel de tráfico en tiempo real
  + Hora del día
  + Tipo de accidente
  + Ubicación geográfica (codificada)
* **Selección automática (Recursive Feature Elimination - RFE)**  
  Se aplicó RFE con regresión lineal y Random Forest para seleccionar el subconjunto óptimo de variables, eliminando iterativamente las menos relevantes.
* **Codificación de variables categóricas**  
  Se utilizó One-Hot Encoding para variables como tipo de vehículo, zona del accidente y día de la semana, permitiendo su inclusión en modelos numéricos.
* **Evaluación del modelo**
  + Métricas de evaluación utilizadas (precisión, recall, F1-score, RMSE, etc.).
  + Métodos de validación (validación cruzada, bootstraping).
  + Comparación entre modelos (si se probaron varios enfoques).

# Resultados y Discusión

En este apartado se deben presentar los resultados obtenidos y su interpretación en relación con el problema planteado. Debe ser clara, objetiva y respaldada por métricas, gráficos y comparaciones con otros estudios o enfoques.

**Elementos clave de Resultados y Discusión**

* **Presentación de resultados**
* **Resumen de los hallazgos clave**: Indicar los resultados más relevantes de los modelos o análisis.
* **Desempeño de los modelos**: Comparación con métricas como precisión, recall, F1-score, RMSE, AUC-ROC, entre otras.
* **Visualización de datos**: Gráficos, tablas y figuras que respalden los hallazgos.
* **Análisis e interpretación**
* **Explicación de los resultados**: ¿Qué significan en el contexto del problema?
* **Comparación con estudios previos** (si aplica): ¿Cómo se comparan los resultados con enfoques anteriores?
* **Factores que influyeron en el desempeño**: Calidad de los datos, hiperparámetros, técnicas de preprocesamiento.
* **Evaluación crítica**
* **Limitaciones del estudio**: Problemas encontrados, sesgos en los datos, restricciones del modelo.
* **Posibles mejoras**: Métodos alternativos, necesidad de más datos, ajustes en la metodología.

Proin sodales mattis dignissim. Pellentesque egestas quam semper erat rhoncus posuere. Maecenas ipsum nisi, bibendum at (ver **Tabla 1**).

Tabla 1. Reporte de clasificación para la red U-Net.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support\* |
| Background  (positive class) | 0.95 | 0.99 | 0.97 | 5419983 |
| Foreground  (negative class)) | 0.95 | 0.75 | 0.83 | 1133617 |
|  |  |  |  |  |
| accuracy |  |  | 0.95 | 6553600 |
| macro avg | 0.95 | 0.87 | 0.90 | 6553600 |
| weighted avg | 0.95 | 0.95 | 0.95 | 6553600 |

\**Support: número de muestras (o instancias) que pertenecen a cada clase en el conjunto de datos utilizado para la evaluación (número de pixels).*

# Conclusiones

El modelo propuesto se alinea con enfoques utilizados en estudios de predicción de tiempos en hospitales y logística, como Random Forest y XGBoost. Las métricas obtenidas (MAE, RMSE, R²) muestran un desempeño competitivo, aunque se reconoce que la incorporación de variables externas podría mejorar significativamente la precisión.

Para enriquecer el modelo nos vimos en la necesidad de integrar APIs para incorporar datos de tráfico y clima en tiempo real mediante servicios como Google Maps.

El presente estudio desarrolló y evaluó un modelo de predicción del tiempo de espera de un asegurado en accidentes de tránsito, integrando técnicas de analítica de datos, aprendizaje automático y análisis geoespacial. Los resultados obtenidos demuestran que es posible mejorar significativamente la estimación de tiempos de atención mediante el uso de modelos predictivos avanzados.

Entre los modelos evaluados, **XGBoost** y **Random Forest** se destacaron por su precisión y capacidad de adaptación a variables dinámicas como tráfico, ubicación y hora del día. Estas técnicas superan ampliamente los enfoques tradicionales basados en promedios o reglas estáticas, lo que sugiere una oportunidad clara para modernizar los sistemas de atención en emergencias viales.

La selección adecuada de características, junto con la reducción de dimensionalidad y el análisis exploratorio, permitió construir un modelo eficiente, interpretable y aplicable en contextos reales. Además, la comparación con modelos logísticos tipo UBER/domicilios refuerza la viabilidad de aplicar estos enfoques en el sector asegurador y de atención médica.

El uso de modelos de predicción basados en datos puede transformar la experiencia del asegurado, optimizar recursos operativos y contribuir a una atención más rápida y eficiente en situaciones críticas. Este trabajo sienta las bases para futuras investigaciones que integren datos en tiempo real, APIs de tráfico y sistemas de geolocalización para mejorar aún más la precisión y utilidad del modelo.

# Referencias

* Breiman, L. (2001). *Random Forests*. Machine Lea
* Breiman, L. (2001). *Random Forests*. Machine Learning.
* Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. ACM SIGKDD.
* Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
* Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. JMLR.
* Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters.

# Bibliografía

Abdelhalim, A. &. (2024). *Computer vision for transit travel time prediction: an end-to-end framework using roadside urban imagery. Public Transport, 17, 221-246. .* Obtenido de https://link.springer.com/article/10.1007/s12469-023-00346-3

Ahmad, O. F., Soares, A. S., Mazomenos, E., Brandao, P., Vega, R., Seward, E., . . . Lovat, L. B. (2019). Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *The lancet. Gastroenterology & hepatology, 4*(1), 71-80. doi:10.1016/s2468-1253(18)30282-6

American Cancer Society. (3 de 10 de 2023). *Colonoscopy*. Obtenido de Cancer.org: https://www.cancer.org/cancer/diagnosis-staging/tests/endoscopy/colonoscopy.html

Obtenido de https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21772

Conteduca, V., Sansonno, D., Russi, S., & Dammacco, F. (2013). Precancerous colorectal lesions. *International journal of oncology, 43*(4), 973-984. doi:10.3892/ijo.2013.2041

Doubeni, C. A., Corley, D. A., Quinn, V. P., Jensen, C. D., Zauber, A. G., Goodman, M., . . . Fle. (2018). Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut, 67*(2), 291-298. doi:10.1136/gutjnl-2016-312712

Hao Liu, W. J. (4 de agosto de 2023). *Uncertainty-Aware Probabilistic Travel Time Prediction for On-Demand Ride-Hailing at DiDi.* Obtenido de Uncertainty-Aware Probabilistic Travel Time Prediction for On-Demand Ride-Hailing at DiDi: https://dl.acm.org/doi/10.1145/3580305.3599925?utm\_source=chatgpt.com

Hsu, C.-M., Hsu, C.-C., Hsu, Z.-M., Shih, F.-Y., Chang, M.-L., & Chen, T.-H. (2021). Colorectal polyp image detection and classification through grayscale images and deep learning. *Sensors (Basel, Switzerland), 21*(1/), 5995. doi:10.3390/s21185995

Martin, R. J. (2021). *Predicting emergency medical service call demand.* Obtenido de Predicting emergency medical service call demand: https://www.sciencedirect.com/science/article/pii/S2211692321000011?utm\_source=chatgpt.com

Nishihara, R., Wu, K., Lochhead, P., Morikawa, T., Liao, X., Qian, Z. R., . . . Ogino, S. (2013). Long-term colorectal-cancer incidence and mortality after lower endoscopy. *The New England journal of medicine, 369*(12), 1095-1105. doi:10.1056/nejmoa1301969

Park, S. B., & Cha, J. M. (2022). Quality indicators in colonoscopy: the chasm between ideal and reality. *Clinical endoscopy, 55*(3), 332-338. doi:10.5946/ce.2022.037

Rokoss, A. (2024). *Case study on delivery time determination using a machine learning approach. Springer / tesis. .* Obtenido de https://link.springer.com/article/10.1007/s10845-023-02290-2?utm\_source=chatgpt.com

Sánchez-Montes, C., Bernal, J., García-Rodríguez, A., Córdova, H., & Fernández-Esparrach, G. (2020). Revisión de métodos computacionales de detección y clasificación de pólipos en imagen de colonoscopia. *Gastroenterologia y hepatologia, 43*(4), 222-232. doi:10.1016/j.gastrohep.2019.11.004

Sánchez-Montes, C., García-Rodríguez, A., Córdova, H., Pellisé, M., & Fernández-Esparrach, G. (2020). Tecnologías de endoscopia avanzada para mejorar la detección y caracterización de los pólipos colorrectales. *Gastroenterologia y hepatologia, 43*(1), 46-56. doi:10.1016/j.gastrohep.2019.09.008

Shine, R., Bui, A., & Burgess, A. (2020). Quality indicators in colonoscopy: an evolving paradigm: Quality indicators in colonoscopy. *ANZ journal of surgery, 90*(3), 215-221. doi:10.1111/ans.15775

Tomar, N. K. (11 de 01 de 2021). *Automatic polyp segmentation using fully convolutional neural network.* Obtenido de arXiv [eess.IV]: http://arxiv.org/abs/2101.04001https://arxiv.org/abs/2101.04001

Williams, J. G., Pullan, R. D., Hill, J., Horgan, P. G., Salmo, E., Buchanan, G. N., . . . Haboubi, N. (2013). Management of the malignant colorectal polyp: ACPGBI position statement. *Colorectal disease: the official journal of the Association of Coloproctology of Great Britain and Ireland, 15*(s2), 1-38. doi:10.1111/codi.12262

World Health Organization. (11 de 07 de 2023). *Colorectal cancer*. Obtenido de Who.int: https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer

* rning.
* Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. ACM SIGKDD.
* Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
* Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. JMLR.
* Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters.