



UNIVERSIDAD  
SERGIO ARBOLEDA

Predicción solicitudes de servicios  
de taxis basados en  
comportamientos de usuarios  
recurrentes en plataformas móviles

JOSE ALFREDO DE LA ROSA FONTALVO

UNIVERSIDAD SERGIO ARBOLEDA  
ESCUELA DE CIENCIAS EXACTAS E INGENIERÍA  
MAESTRÍA INTELIGENCIA ARTIFICIAL  
BOGOTÁ D.C.  
2023

# Predicción solicitudes de servicios de taxis basados en comportamientos de usuarios recurrentes en plataformas móviles

JOSE ALFREDO DE LA ROSA FONTALVO

DOCUEMNTTO FINAL DE MAESTRÍA

Asesor  
Camilo Rodríguez

UNIVERSIDAD SERGIO ARBOLEDA  
ESCUELA DE CIENCIAS EXACTAS E INGENIERÍA  
MAESTRÍA EN INTELIGENCIA ARTIFICIAL  
BOGOTÁ D.C.  
2023

---

Nota de aceptación:

---

---

---

---

---

---

---

Firma del presidente del jurado

---

Firma del jurado

---

Firma del jurado

---

## Dedicatoria

Nota mediante la cual el autor ofrece su trabajo, en forma especial, a personas o entidades [ICONTEC, 2008]. Su presentación es opcional y, aunque cada autor podrá determinar la distribución del texto en la página, se sugiere esta presentación. En ella el autor dedica su trabajo en forma especial a personas y/o entidades.

Por ejemplo:

A mis padres

o

A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.

Alan Turing

---

## Agradecimientos

Página de agradecimientos. En ella el (los) autor (es) expresa (n) el reconocimiento hacia las personas y entidades que asesoraron técnicamente, suministraron datos, financiaron total o parcialmente la investigación o contribuyeron significativamente al desarrollo del tema. Es opcional y debe contener, además de la nota correspondiente, los nombres de las personas con sus respectivos cargos y los nombres completos de las instituciones y su aporte al trabajo [[ICONTEC, 2008](#)].



UNIVERSIDAD  
SERGIO ARBOLEDA

---

## Resumen

Esta tesis pretende mostrar el estudio de diferentes técnicas de

El resumen es una presentación abreviada y precisa del contenido de un documento, sin agregar interpretación o crítica. Para documentos extensos como informes, tesis y trabajos de grado, no debe exceder de 500 palabras, y debe ser lo suficientemente breve para que no ocupe más de una página [ICONTEC, 2008]. Se recomienda que este resumen sea analítico, es decir, que sea completo, con información cuantitativa y cualitativa, generalmente incluyendo los siguientes aspectos: objetivos, diseño, lugar y circunstancias, pacientes (u objetivo del estudio), intervención, mediciones y principales resultados, y conclusiones. Al final del resumen se deben usar palabras claves tomadas del texto (mínimo 3 y máximo 10 palabras), las cuales permiten la recuperación de la información.

**Palabras clave:** (máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado).

**Palabras clave:** palabra 1, palabra 2, palabra 3, ...

---

## Abstract

**Keywords:** Keyword 1, Keyword 2, Keyword 3, ...



UNIVERSIDAD  
SERGIO ARBOLEDA

---



UNIVERSIDAD  
SERGIO ARBOLEDA



# Contenido

---

<b>Lista de figuras</b>	<b>11</b>
<b>Lista de tablas</b>	<b>13</b>
<b>1. Introducción</b>	<b>17</b>
1. Contexto del problema . . . . .	17
2. Hipótesis . . . . .	18
3. Justificación . . . . .	18
4. Retos y desafíos . . . . .	19
4.1. Movilidad urbana . . . . .	19
4.2. Industria del taxi . . . . .	21
5. Marco teórico . . . . .	22
5.1. Análisis general . . . . .	22
5.2. Investigaciones de autores . . . . .	23
5.3. Análisis comparativo . . . . .	25
6. Objetivos . . . . .	26
6.1. Objetivo general . . . . .	26
6.2. Objetivos específicos . . . . .	26
<b>2. Recolección y análisis del conjunto de datos</b>	<b>29</b>
<b>3. Modelos predictivos y características</b>	<b>31</b>
<b>4. Resultados</b>	<b>33</b>
<b>5. Conclusiones y recomendaciones</b>	<b>35</b>
1. Conclusiones . . . . .	35
2. Recomendaciones . . . . .	35
<b>6. Bibliografía</b>	<b>37</b>



# Lista de figuras

---

1.1. Modelo stacking propuesto por el autor . . . . .	24
1.2. Arquitectura propuesta por el autor . . . . .	24
4.1. Solución del problema . . . . .	34



# Lista de tablas

---

1.1. Tabla comparación de modelos utilizados por diferentes autores . . . .	26
---	----



# Glosario

---

En esta sesión, vamos a explicar un poco de los algoritmos o terminología utilizada a lo largo de este trabajo.

- **LSTM** - Long Short Term Memory, es un tipo de redes neuronales recurrentes, es comunmente utilizada para predicción de series de tiempo.
- **XGBoost** - eXtreme Gradient Boosting, es un algoritmo de tipo supervisado que utiliza un modelo más fuerte a partir de modelos débiles en particular para resolver problemas con problemas con datos heterogéneo[Mendoza, ].
- **CNN** - Convolutional Neural Network, es un algoritmo de aprendizaje profundo, empleado en visión por computación y reconocimiento de imágenes. Comúnmente en: reconocimiento de imágenes, detectando objetos en carros autónomos, reconocimiento de rostros en redes sociales y análisis de imágenes en medicina[Sharma, ].
- **RNN** - Recurrent Neural Network, s un tipo de red neuronal artificial, es comunmente usado con datos de series de tiempo o secuenciales para resolver problemas de regresión o clasificación como: traducción de texto, reconocimiento de voz, resumen de textos, análisis de call center, reconocimiento facial, entre otros[Education, ].
- **ARIMA** - Autoregressive Integrated Moving Average,es uno de los métodos de previsión más utilizados para la previsión de datos de series temporales univariantes[Brownlee, b].
- **SARIMA** - Seasonal Autoregressive Integrated Moving Average, es una extensión de ARIMA que permite modelar directamente el componente estacional de la serie[Brownlee, b].
- **SVM** - Support Vecttor Machine, son un conjunto de métodos de aprendizaje supervisado que se utilizan para la clasificación, la regresión y la detección de valores atípicos[McGregor, ].
- **LS-SVM** - Least-squares Support Vector Machine, es un método para analizar datos y reconoce patrones los cuales son muy comúnmente usados para análisis de regresión y clasificación[wikipedia, ].

- **Wavelets de Daubechies5**, son una familia de wavelets ortogonales que definen una transformada wavelet discreta y se caracterizan por un número máximo de momentos de fuga para un soporte determinado. Con cada tipo de wavelet de esta clase, existe una función de escalado (denominada wavelet padre) que genera un análisis ortogonal multirresolución[[wikipedia](#), ].
- **MLP** - multilayer perceptron, es una clase de red neuronal artificial (RNA), es comúnmente utilizado para predecir problemas de clasificación o regresiones[[Brownlee, c](#)].
- **NNA** - Neural Network Algorithms, es un algoritmo metaheurístico de optimización, inspirado en los sistemas nerviosos biológicos y en las redes neuronales artificiales (RNA), para resolver problemas complejos de optimización[[Adollah](#), ].
- **GAN** - Generative Adversarial Networks, es una aproximación al modelado generativo utilizando métodos de aprendizaje profundo, como las redes neuronales convolucionales. Además son una forma inteligente de entrenar un modelo generativo, enmarcando el problema como un problema de aprendizaje supervisado[[Brownlee, a](#)].



# Introducción

## 1. Contexto del problema

Las empresas operadoras de servicios de taxis de las grandes ciudades de Colombia se encuentran presentando un creciente desafío en mantener una constante disponibilidad de servicios de taxis en sus plataformas, esto ocurre principalmente en algunas zonas u horarios específicos.

La empresa operadora de transporte de Bogotá **Taxis Libres (TL)**, una de las más grandes operadoras de servicios de taxis del país, realizó una encuesta anónima a sus conductores para identificar los motivos que les permitiera entender el por qué no tomaban servicios en zonas u horarios de alta demanda por la plataforma, en el resultado lograron identificar lo siguiente por parte de los conductores:

- No cuentan con tiempo suficiente para llegar a cumplir con los servicios
- Elevado consumo de tiempo por la alta densidad del tráfico desde su ubicación hasta a la ubicación del usuario
- Alta demanda de servicios en las calles en los horarios picos (entre las 6 y 9 am y entre las 3 y 7 pm) y aún más cuando llueve
- El único medio de pago que se maneja en los servicios ofrecidos en las calles es el efectivo
- Falta de tarifas dinámicas por condiciones climáticas, eventos, horarios o zonas de altas demandas
- La tarifa del producido del día solo se entrega en efectivo a los dueños o administradores de los taxis

Como derivación a lo anterior mencionado, los conductores han optado por ofrecer más servicios en las calles y menos en las plataformas, evitando con esto mantener una constante disponibilidad de conductores que suplan la oferta de los usuarios. En consecuencia, la plataforma de la empresa **TL** ha empezado a recibir una gran cantidad de inconformidades por partes de los usuarios recurrentes de la plataforma. Solo en el mes de marzo del 2023 la plataforma recibió alrededor de 2.1 millones de solicitudes de servicios de taxis, de los cuales el 57 % de esas solicitudes fueron abandonadas o rechazadas, lo que los llevó a analizar qué estaba pasando con la prestación del servicio. En

## 2. Hipótesis

---

una revisión realizada durante el mismo periodo de tiempo se obtuvieron las principales causas reportadas en la plataforma y que se listarán a continuación:

- El usuario se fue
- El usuario no necesita el taxi
- El vehículo se demoró en llegar
- El usuario tomó otro taxi
- El conductor solicitó cancelar el viaje
- El usuario tiene otro servicio
- El conductor no acepta el método de pago

Al identificar las causas de los servicios abandonados o rechazados en relación con los motivos obtenidos por los conductores, la empresa planteó realizarse la siguiente pregunta al problema presentado.

¿Como reducir la alta cantidad de inconformidades de los usuarios recurrentes de la plataforma prediciendo sus comportamientos durante la semana y así programarles sus servicios de forma anticipada?

## 2. Hipótesis

Diseñar una solución tecnológica que logre reducir al menos un **10%** la cantidad de solicitudes abandonados o rechazados que presentan actualmente las plataformas de taxis, con el propósito de mejorar la experiencia en la prestación del servicio en los usuarios recurrentes.

## 3. Justificación

Los taxistas en las grandes ciudades tienen un gran desafío día a día prestando servicios en las calles. Ellos constantemente se exponen al peligro de recoger personas sin conocer quienes son, cual es su procedencia, a que destino quieren ir y lo más importante sus intenciones. Adicionalmente, pasan la mayor parte de su tiempo adivinando dónde se tienen que ubicar para así recoger en menor tiempo la mayor cantidad de pasajeros evitando estar dando vueltas por diferentes zonas de la ciudad, haciendo un uso ineficiente del tiempo y el combustible sin obtener un máximo provecho.

La intuición de ellos algunas veces le funciona y es hasta probable que muchas veces acierten por el conocimiento de la profesión en las calles durante años de experiencia,

pero la mayoría de las veces los comportamientos de los usuarios cambian por diferentes circunstancias como suelen ser: el tráfico, el clima, los accidentes, las manifestaciones, los conciertos u otro tipo de eventos que se presenten en la ciudad. Esto hace que los usuarios varíen su conducta con respecto a lo habitual.

Por otro lado, los pasajeros también necesitan obtener sus servicios de taxis los más pronto posible. Lograr tener disponibilidad de servicios de taxi de forma proactiva ayudaría tanto a los usuarios como a los conductores en su laburo diario.

El envío anticipado de servicios de reservas de taxis ayudaría a reducir el tiempo de espera de los clientes dándoles una mejor experiencia, así como también, hacer un uso eficiente de los recursos limitados de los conductores, para que con esto puedan ir programándose con tiempo los recorridos van a realizar, maximizando de esta forma el tiempo de trabajo durante el día.

Basado en lo anterior, este trabajo pretende definir el diseño de una solución que permita identificar el comportamiento de los usuarios recurrentes y predecir con anticipación cuando podrían tomar su próximo servicio. Todo esto sería posible solo si los usuarios mantienen hábitos o rutinas que permitan identificar sus comportamientos dentro de la plataforma. Además, esto le ayudará a que se puedan reservar con anticipación servicios, logrando disminuir la probabilidad de abandonos o rechazados en horarios donde habitualmente se necesiten, adicionalmente el conductor programará con anticipación la recogida del servicio logrando así más clientes y conductores felices.

## 4. Retos y desafíos

### 4.1. Movilidad urbana

La movilidad en las grandes ciudades se enfrenta a varios retos debido a factores como la densidad de población, la congestión del tráfico, la limitación de la infraestructura y los problemas de sostenibilidad.

Según Cabify en su blog de desafíos de movilidad urbana "La crisis climática es la espada de Damocles, es el mayor reto al que se enfrentan las ciudades en cuanto a movilidad: no hay transporte de futuro si no hay futuro. Los gobiernos del mundo trabajan ya en sus agendas verdes, poniendo muescas en el calendario con el cumplimiento de los (ODS) (Objetivos de Desarrollo Sostenible) siendo Europa el continente que pretende liderar el cambio hacia la neutralidad climática para el año 2050 a través de la nueva Ley Europea del Clima." [Cabify, 2021].

Aunque la crisis climática suena como uno de los temas más relevantes al hablar de movilidad urbana, existen muchos otros factores que contribuyen a agravarlo. Algunos

#### 4. Retos y desafíos

---

de los más apremiantes son:

- La congestión: es un problema importante en muchas grandes ciudades y puede tener un impacto significativo en el medio ambiente, la economía y la calidad de vida de los residentes. La congestión puede aumentar la contaminación atmosférica, acústica y los accidentes de tráfico. También puede dificultar los desplazamientos, con la consiguiente pérdida de productividad y aumento de los niveles de estrés.
- Transporte público ineficaz: este suele ser ineficiente en las grandes ciudades. Debido a varios factores como: infraestructuras anticuadas, falta de financiación y mala planificación. Un transporte público ineficaz puede dificultar los desplazamientos, lo que puede aumentar la dependencia del automóvil y los atascos.
- Falta de accesibilidad peatonal: muchas grandes ciudades son poco transitables. Esto puede dificultar los desplazamientos sin coche, lo que puede aumentar la congestión y reducir la actividad física.
- Elevado coste del transporte: el costo del transporte puede ser un obstáculo para muchas personas en las grandes ciudades. Esto puede deberse al costo de poseer un vehículo, al costo del transporte público y al costo de los estacionamientos. Los elevados costos del transporte pueden dificultar que la gente se permita vivir en las grandes ciudades.

Aunque existen varias alternativas que se pueden realizar para afrontar los retos de movilidad en las grandes ciudades. Muchas tienden a realizar estrategias de gobierno que les permitan planificar soluciones prometedoras que podrían disminuir los problemas, entre muchas soluciones las que se pueden considerar más importantes son:

- Invertir en transporte público: cuando se hace de manera correcta, puede lograr que la movilidad sea eficiente y asequible, lo que la vuelve en una opción más atractiva para las personas.
- Convertir las ciudades más transitables: estas pueden facilitar los desplazamientos sin coche de los peatones. Si embargo, esto no es posible sino se invierten en realizar más aceras, carriles de bicicleta o parques.
- Promover modos de transporte alternativos: los medios de transporte alternativos, como el carro compartido, el mono patín eléctrico o la bicicleta, pueden ayudar a reducir la congestión y mejorar la calidad del aire.
- Cambiar la actitud de la gente respecto al uso del transporte: cambiar la actitud de las personas puede ayudar a que estén más dispuestos a utilizar otros modos de transporte alternativos. Esto se puede realizar mediante campañas de educación y concienciación.

Aunque los retos de movilidad en las grandes ciudades son complejos, pueden hacer varias cosas para resolverlos. Invirtiendo en transporte público, haciendo las ciudades más transitables y promoviendo modos de transporte alternativos, podemos hacer que nuestras ciudades sean más habitables y sostenibles.

## 4.2. Industria del taxi

Desde 2016, el sector mundial del taxi ha experimentado un cambio en la forma en que los potenciales pasajeros interactúan a través de plataformas móviles. Con la llegada de nuevos competidores de servicios de transporte, las empresas buscan diferentes formas de predecir mejor la demanda y las eficiencias futuras.

La mayor revolución de la industria del taxi ha sido el auge de las plataformas de viajes habilitadas por aplicaciones móviles. Empresas como **Uber** y **Lyft** fueron pioneras en el uso de aplicaciones por teléfonos inteligentes para conectar a pasajeros con conductores de una forma más cómoda y eficiente, provocando cambios sustanciales en la dinámica del mercado.

La introducción de nuevas plataformas de viajes trajo consigo varios cambios significativos en el sector del taxi:

- **Competencia en el mercado:** las plataformas de transporte introdujeron una mayor competencia en el mercado del transporte. Ofrecían una alternativa más cómoda y, a menudo, más asequible que los taxis tradicionales, lo que atrajo a un número significativo de clientes.
- **Modelo entre iguales:** las plataformas de viajes crearon un modelo entre iguales en el que cualquiera con un vehículo que cumpliera los requisitos podía convertirse en conductor. Esto abrió oportunidades a particulares que querían obtener ingresos extra ofreciendo viajes en sus propios coches.
- **Avances tecnológicos:** las plataformas de transporte aprovecharon las aplicaciones móviles y la tecnología GPS para ofrecer experiencias de usuario sin brechas. Esta innovación tecnológica permitió funciones como el seguimiento en tiempo real, las transacciones sin efectivo y el envío automatizado, haciendo que la experiencia tradicional del taxi pareciera anticuada en comparación.
- **Comodidad y accesibilidad:** mediante el uso de aplicaciones para teléfonos inteligentes, los usuarios podían solicitar fácilmente un viaje con sólo pulsar un par de veces en su teléfono. La aplicación ofrece información en tiempo real sobre la ubicación del conductor y la hora estimada de llegada, lo que hace que el proceso sea más cómodo y transparente.

- Precios y transparencia: las plataformas de viajes introdujeron la tarificación dinámica, también conocida como tarificación por sobrecarga, que ajusta las tarifas en función de la oferta y la demanda. Este modelo de precios transparente eliminó el regateo y la incertidumbre asociados a menudo a los taxis tradicionales.
- Sistema de calificación de conductores: las plataformas de transporte introdujeron un sistema de puntuación que permitía a los pasajeros valorar a sus conductores y dar su opinión. Este bucle de retroalimentación ayudó a mantener un nivel de responsabilidad y mejoró la calidad del servicio.
- Funciones de seguridad mejoradas: las plataformas implementaron características de seguridad como la comprobación de los antecedentes del conductor, la información del vehículo y el seguimiento por **GPS**. Estas medidas ayudaron a mejorar la seguridad de los pasajeros y proporcionaron una sensación de seguridad.
- Mejora de la experiencia del cliente: las plataformas de transporte se centraron en mejorar la experiencia general del cliente. La posibilidad de valorar a los conductores, dar su opinión y acceder al servicio de atención al cliente a través de la aplicación contribuyó a mantener un nivel de servicio superior al de los taxis tradicionales.

A pesar que las nuevas plataformas de transporte han perturbado de forma significativa el sector tradicional del taxi. La aparición de empresas como **Uber** o **Lyft** han planteado nuevos retos y oportunidades al sector del taxi, provocando cambios sustanciales en la dinámica del mercado. Su llegada alteró radicalmente el sector tradicional y suscitó debates y discusiones sobre la normativa, las prácticas laborales y el futuro del transporte. También inspiró innovaciones en otras áreas del transporte, como los servicios de reparto de comida.

En general, el transporte de taxis fue perturbado al introducir en un sector tan tradicional tecnologías innovadoras, mejoras en la comodidad y la transparencia en poner en tela de juicio los modelos de negocio tradicionales. Esto obligó al sector del taxi a adaptarse y evolucionar para seguir siendo competitivo en el cambiante panorama del transporte.

## 5. Marco teórico

### 5.1. Análisis general

Partiendo del análisis de los textos literarios acerca de las investigaciones que se han venido realizado acerca de la predicción de la demanda de los servicios de transporte en diferentes ámbitos como taxis, renta de vehículos, transportes públicos, privados, entre otros. Se puede determinar que se ha dedicado mucho esfuerzo para entender como este tema es parte indispensables en el desarrollo de sistemas de transportes y ciudades

SERGIO ARBOLEDA

inteligentes.

En la última década, los modelos de predicción de servicios han venido presentando un tema candente debido a que tiene características temporales y dinámicas que lo hacen muy interesante. De hecho, un buen modelo de predicción nos podría ayudar pre-asignar recursos de conductores, mejorando la experiencia de los usuarios en las plataformas de servicios de transporte [Youfang, 2019]. Para realizar esta investigación, se analizaron más de 100 artículos publicados en diferentes fuentes de información, lo que permitió observar cómo se habían abordados los diferentes enfoques acerca de la predicción de la demanda de servicios como eje fundamental en los problemas de movilidad de las diferentes ciudades del mundo.

En esa revisión se evidenció algo muy interesante en las diferentes fuentes consultadas, cerca del 60 % de los artículos analizados se encuentran desarrollados para ciudades de chinas y esto ocurre dado a que menos del 10 % de los 1.400 millones de ciudadanos chinos posee un automóvil y la frecuencia con la que los ciudadanos chinos se desplazan en taxi, autobús y el metro es la más alta del mundo. Durante las últimas décadas, muchos investigadores han dedicado tiempo a desarrollar modelos de predicción de la demanda de transporte, esos métodos se podrían agrupar como lineales, no lineales, computacionales y combinados [Li, 2017].

## 5.2. Investigaciones de autores

Para entender la problemática que se está abordando es importante conocer en profundidad que la demanda de taxis y viajeros se encuentra variando temporalmente en diferentes zonas de la ciudad, lo que motiva a los conductores a estar en las áreas con un alto potencial de demanda de pasajeros y así mejorar la tasa de utilización de los taxis. Últimamente se ha investigado mucho la correlación entre la demanda de taxis y las dependencias relacionadas. Sin embargo, la previsión de la demanda de taxis sigue siendo un problema abierto y muy explorado. Con base a lo anterior, la mayoría de los artículos consultados proponen investigar cómo mover los taxis durante espacios de tiempo, ubicaciones de la ciudad e incluso con otras variables para tratar de predecir su comportamiento o las zonas donde se están moviendo los viajeros como los explican los siguientes autores.

El autor **Soumia Boumeddan** en su estudio [Boumeddane, 2021] propone utilizar un stack de modelos para predecir la demanda de transportes colectivos, entre los métodos propuestos se encuentran los modelos estadísticos (**ARIMA** y **SARIMA**) y de aprendizaje de máquina (**XGBoost**, **LSTM** y **MLP**).

En estudios anteriores, se han utilizados diferentes modelos de redes neuronales profundas para predecir la demanda de taxis. A pesar del esfuerzo estos ensayos no han funcionado adecuadamente debido a la información de contexto.

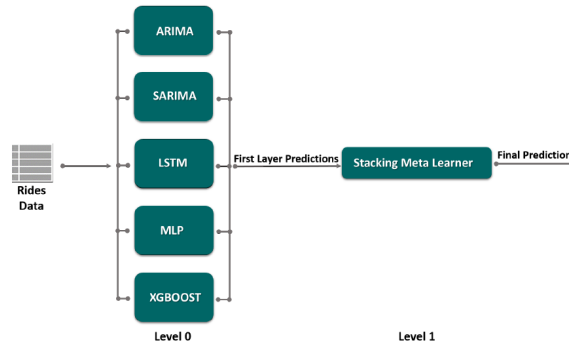


Figura 1.1: Modelo stacking propuesto por el autor

Sin embargo, en el trabajo de **Liu Tong**, utilizaron (**LSTM**, **CNN**, **RNN**) para extraer las características espacio-temporales de las demandas de taxi, que además tuvieron en cuenta la información contextual de las regiones logrando buenos resultados en la predicción de los servicios de taxi[Tong, 2020].

A su vez, **Yunxuan Li**, propuso predecir la demanda de trafico de corto plazo utilizando un modelo de **Wave SVM** en 4 etapas donde combinan análisis de wavelets de Daubechies5 y (**LS-SVM**).

Otro enfoque lo realizó **Sai Krishna**, propuso en su trabajo de investigación un modelo de redes profundas basado en **RNN** con la información de las solicitudes servicios actuales y la historia de varias zonas de la ciudad para predecir la hora de la demanda de los servicios de taxis, esto se explica a continuación como lo demuestra la siguiente arquitectura:

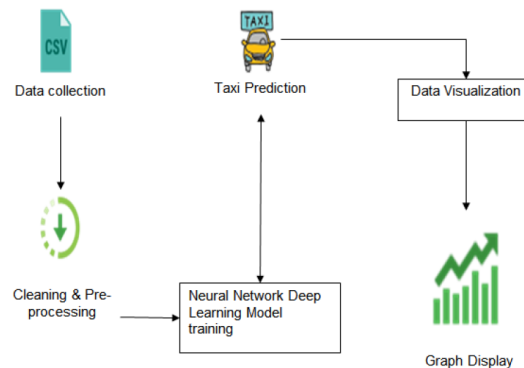


Figura 1.2: Arquitectura propuesta por el autor

Este diseño explica como proporcionar un modelo que pueda suplir la expectativa de



interés de los taxis para una premisa horaria y un tiempo específico en varios territorios de la ciudad.[A., 2021]

Por otra parte, el autor **Chung-Yi Lin** demuestra en su investigación los buenos resultados obtenidos de la predicción de taxis utilizando (**NNA**) que le permitió predecir un modelo utilizando las siguientes características:

- Día de la semana
- Temperatura
- Horario de lluvia
- Promedio de movimiento

Para lograr tener buenos rendimientos con sus resultados propuestos utilizó los modelos (**CNN**, **LSTM**, **RNN**, **GRU**)

Por último, el autor **Hasan A. H. Naji** decidió investigar la predicción de la demanda de taxis como lo hicieron muchos investigadores, pero en particular este autor decidió darle un giro distinto y aplicar el modelo de Generative Adversarial Networks (**GAN**) a múltiples fuentes de datos lo que le permitió aventurar un enfoque distinto a la mayoría de las propuestas. Sabiendo que estas aprenden automáticamente los patrones de los datos de tal manera que el modelo puede utilizarse para generar salida a nuevos ejemplos que podrían haberse extraído del conjunto de datos original, lo que si se aplica a diferentes fuentes de datos aprendería patrones de distintas fuentes lo que lo lleva a ser un modelo muy potente para esta propuesta[Naji, 2021].

Como se ha reafirmado en todos los trabajos estudiados, los taxis juegan un muy importante en los sistemas urbanos de transporte de cara a las ciudades inteligentes, logrando brindar un cómodo y confortable servicio a millones usuarios en todo el mundo.

### 5.3. Análisis comparativo

En esta sesión se realizó una cuadro comparativo de los autores investigados anteriormente donde se especifican los modelos que cada uno utiliza en sus tesis de investigación para resolver el problema de predicción de servicios de taxis. A continuación, se presenta la tabla con cada uno de los modelos propuestos:

## 6. Objetivos

Tabla de comparación de modelos								
Autor	LSTM	XGBoost	CNN	RNN	SVM	MLP	NNA	GAN
Soumia Bou-meddan	X	X				X		
Liu Tong	X		X	X				
Yunxuan Li					X			
Sai Krishna				X				
Chung-Yi Lin	X		X	X			X	
Hasan A. H. Naji								X

Cuadro 1.1: Tabla comparación de modelos utilizados por diferentes autores

## 6. Objetivos

### 6.1. Objetivo general

Diseñar un método que permita predecir con al menos un 70 % de probabilidad, las solicitudes de servicios de los usuarios recurrentes.

### 6.2. Objetivos específicos

A continuación se describen los objetivos específicos que detallan el objetivo general:

- Extraer y procesar los datos de una plataforma móvil de una empresa operadora de servicios de taxis de Bogotá - Colombia del periodo comprendido entre septiembre del 2022 hasta abril del 2023.
- Realizar un experimento con los siguientes modelos: probabilístico, Kneighbors, Gradient Boosting, Random Forest, Ada boost y Multilayer Perceptron. Para evaluar sus diferentes métricas de desempeño.
- Probar los modelos con los usuarios recurrentes que superen la validación de retención "del latín retentio, es la acción y efecto de retener" [Porto, 2010] (en el contexto del problema la retención se determina si el usuario toma un servicio al menos el 60 % de las veces el mismo día de la semana y a la misma hora) para determinar los mejores hiperparámetros a utilizar.

- Evaluar el resultado de los modelos determinando aquellos que presenten mejores métricas de: precision, F1, KAPPA, ROC, score test, recall y confusion matrix para determinar cual representa el mejor rendimiento.



## 6. Objetivos

---



UNIVERSIDAD  
SERGIO ARBOLEDA

# Recolección y análisis del conjunto de datos

---

Set de datos



UNIVERSIDAD  
SERGIO ARBOLEDA

---



UNIVERSIDAD  
SERGIO ARBOLEDA

# Modelos predictivos y características

---

Modelo predictivo



UNIVERSIDAD  
SERGIO ARBOLEDA

---



UNIVERSIDAD  
SERGIO ARBOLEDA



# Resultados

---

Se creó un modelo que permitiera predecir la hora y el día de la semana de un usuario que posiblemente tomará un servicio, esto con el fin de reservarle de manera anticipada un servicio por medio de una notificación en la plataforma móvil donde solo deberá confirmar si lo desea, y así, mejorar su experiencia.

A continuación, se explica con más detalle la solución propuesta:

Uno de los principales factores relevantes de un buen modelo predictivo es la data, para esto se analizó la información de **8.2 millones** de servicios que realizaron los usuarios en la plataforma de Taxis Libres entre los periodos de diciembre del 2021 a mayo del 2022.

Por último, se limpió los datos nulos y los outliers del set de datos; se tomó los atributos: teléfono del usuario, fecha, hora y día de la semana del servicio; se probó el modelo probabilístico, KNN y Random Forest.



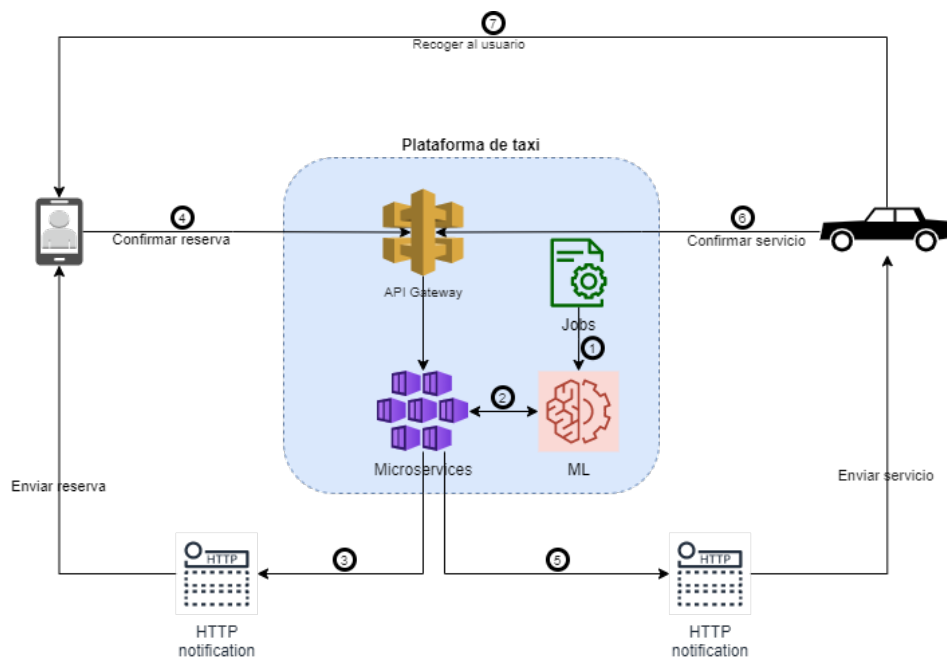


Figura 4.1: Solución del problema

# Conclusiones y recomendaciones

---

## 1. Conclusiones

Se seguirán realizando pruebas ajustando los hiperparámetros y las técnicas utilizadas, logrando incrementar la precisión en al menos un **80%** de las métricas de clasificación para que el modelo sea más confiable.

En esta investigación solo se tuvieron en cuenta las variables de tiempo para predecir el servicio, pero para una futura investigación se podría complementar con otras variables que permitan enriquecer el modelo como por ejemplo: días feriados, zona segura, clima o eventos en la ciudad.

## 2. Recomendaciones

Se presentan como una serie de aspectos que se podrían realizar en un futuro para emprender investigaciones similares o fortalecer la investigación realizada. Deben contemplar las perspectivas de la investigación, las cuales son sugerencias, proyecciones o alternativas que se presentan para modificar, cambiar o incidir sobre una situación específica o una problemática encontrada. Pueden presentarse como un texto con características argumentativas, resultado de una reflexión acerca de la tesis o trabajo de investigación.



## 2. Recomendaciones

---



UNIVERSIDAD  
SERGIO ARBOLEDA

# Bibliografía

- [A., 2021] A., S. K. G. V. (2021). Hourly demand prediction for taxi services using deep learning. *Annals of the Romanian Society for Cell Biology*, 25:63–70. 25
- [Adollah, ] Adollah, A. Neural network algorithm (nna). 16
- [Boumeddane, 2021] Boumeddane, S. (2021). A Model Stacking Approach for Ride-Hailing Demand Forecasting : a Case Study of Algiers. *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being*, 14(3):16–21. 23
- [Brownlee, a] Brownlee, J. A gentle introduction to generative adversarial networks (gans). 16
- [Brownlee, b] Brownlee, J. A gentle introduction to sarima for time series forecasting in python. 15
- [Brownlee, c] Brownlee, J. When to use mlp, cnn, and rnn neural networks. 16
- [Cabify, 2021] Cabify (2021). Desafíos de la movilidad urbana. 19
- [Education, ] Education, I. C. Recurrent neural networks. 15
- [ICONTEC, 2008] ICONTEC (2008). *Documentación. Presentación de Tesis, Trabajos de grado y otros trabajos de investigación. NTC1486*. Instituto Colombiano de Normas Técnicas y Certificación, Bogotá D.C., Colombia, sexta actualización edition. 4, 5, 6
- [Li, 2017] Li, Y. (2017). Taxi booking mobile app order demand prediction based on short-term traffic forecasting. *Transportation Research Record*. 23
- [McGregor, ] McGregor, M. Svm machine learning tutorial – what is the support vector machine algorithm, explained with code examples. 15
- [Mendoza, ] Mendoza, J. B. Tutorial: Xgboost en python. 15
- [Naji, 2021] Naji, H. A. H. (2021). Forecasting taxi demands using generative adversarial networks with multi-source data. *Applied Sciences*, 11:9675. 25
- [Porto, 2010] Porto, P. (2010). Retención - qué es, definición y concepto. 26

- 
- [Sharma, ] Sharma, P. Applications of convolutional neural networks(cnn). 15
- [Tong, 2020] Tong, L. (2020). Predicting Taxi Demands via an attention-based Convolutional Recurrent Neural Network. *School of Computer Engineering and Science, Shanghai University, Shanghai, China*, 206. 24
- [wikipedia, ] wikipedia. Daubechies wavelet. 15, 16
- [Youfang, 2019] Youfang, L. (2019). Spatio-temporal lstm for od passenger demand prediction. *Beijing Jiaotong Daxue Xuebao/Journal of Beijing Jiaotong University*, 43:114–121. 23