

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

Jose De la rosa
Camilo Rodríguez



Introducción

Las empresas de taxis de la ciudad de Bogotá - Colombia, han estado recibiendo en sus plataformas por parte de sus usuarios recurrentes una gran cantidad de inconformidades. Esto se viene presentando debido a la mala experiencia que tienen al momento de solicitar un servicio de taxi.

Problema

Realizando la revisión del comportamiento de los servicios en el periodo comprendido entre enero y mayo del 2022 de la plataforma de transporte Taxis Libres. Esta recibió mensualmente en promedio 1.6 millones de solicitudes de los cuales el 67.9% fueron abandonados o rechazados. Así mismo, se realizaron unas encuestas para determinar las principales causas de las inconformidades, dando como resultado: el vehículo se demoró en llegar, el conductor solicitó cancelar el viaje y el conductor nunca llegó.

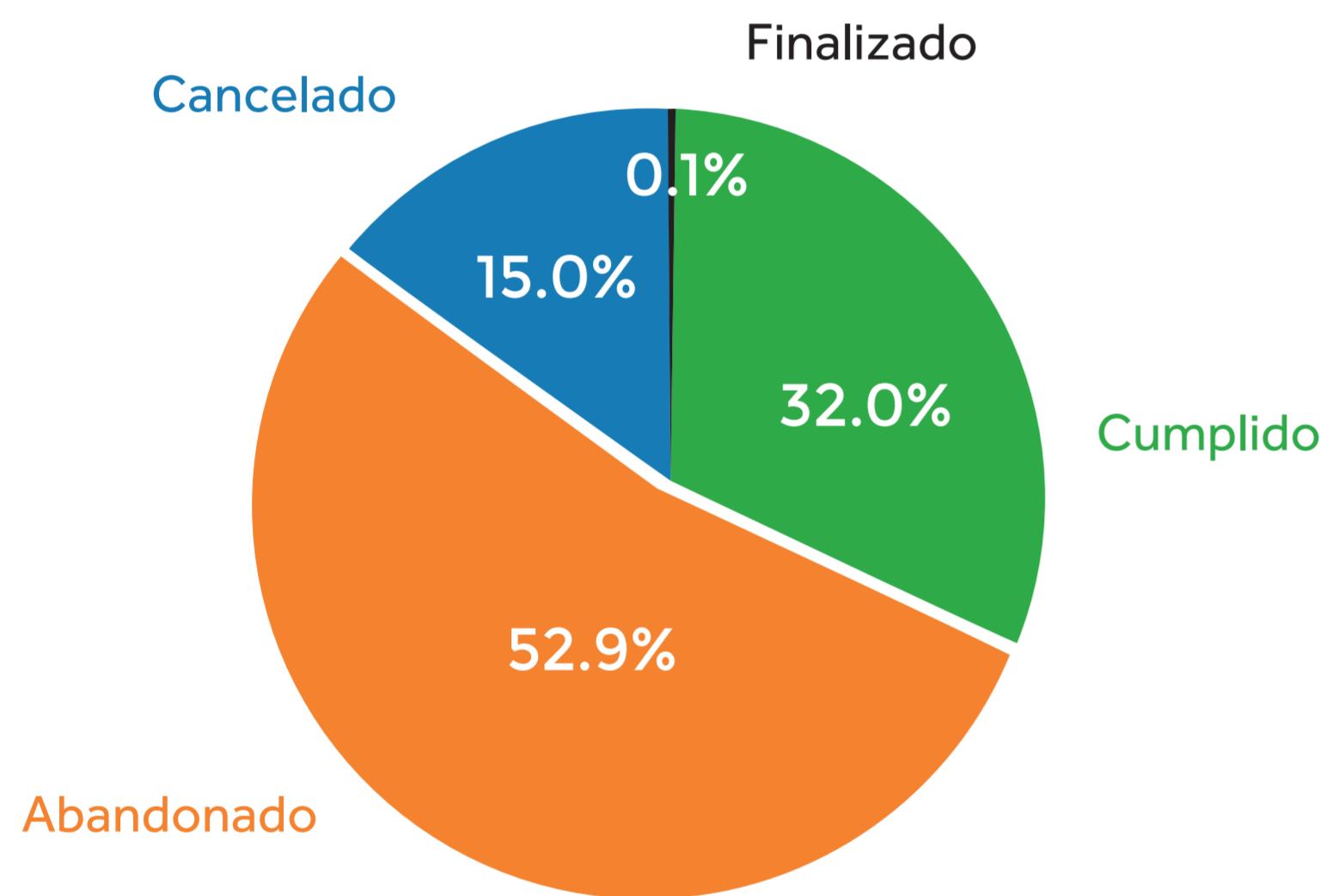


Figura 1: Servicios por estados

Con base en el problema dicho, la empresa se planteó la siguiente pregunta problema: **¿Cómo podemos reducir la cantidad de servicios cancelados y rechazados para mejorar experiencia del usuario?**

Hipótesis

Diseñar una solución tecnológica que se suponga que logre reducir un **15%** en la cantidad de solicitudes abandonados y rechazados que tienen actualmente las plataformas de taxis, esto con el propósito de mejorar la experiencia en la prestación del servicio de taxis en los usuarios recurrentes.

Objetivos

Diseñar un método que permita predecir con al menos 80% de probabilidad, las solicitudes de servicios de los usuarios recurrentes de empresas de taxis.

Objetivos específicos:

- Extraer y procesar los datos de la plataforma móvil desde el periodo de diciembre del 2021 hasta mayo del 2022 para evaluar los modelos.
- Realizar experimentos con los modelos (probabilístico, KNN y Random Forest) para validar sus métricas de desempeño.
- Probar los modelos con al menos 1.000 usuarios para determinar los mejores hiperparámetros a utilizar.
- Validar aquellos modelos que superen el 80% de la métrica de precisión para determinar cual representa mejor rendimiento.

Solución

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.

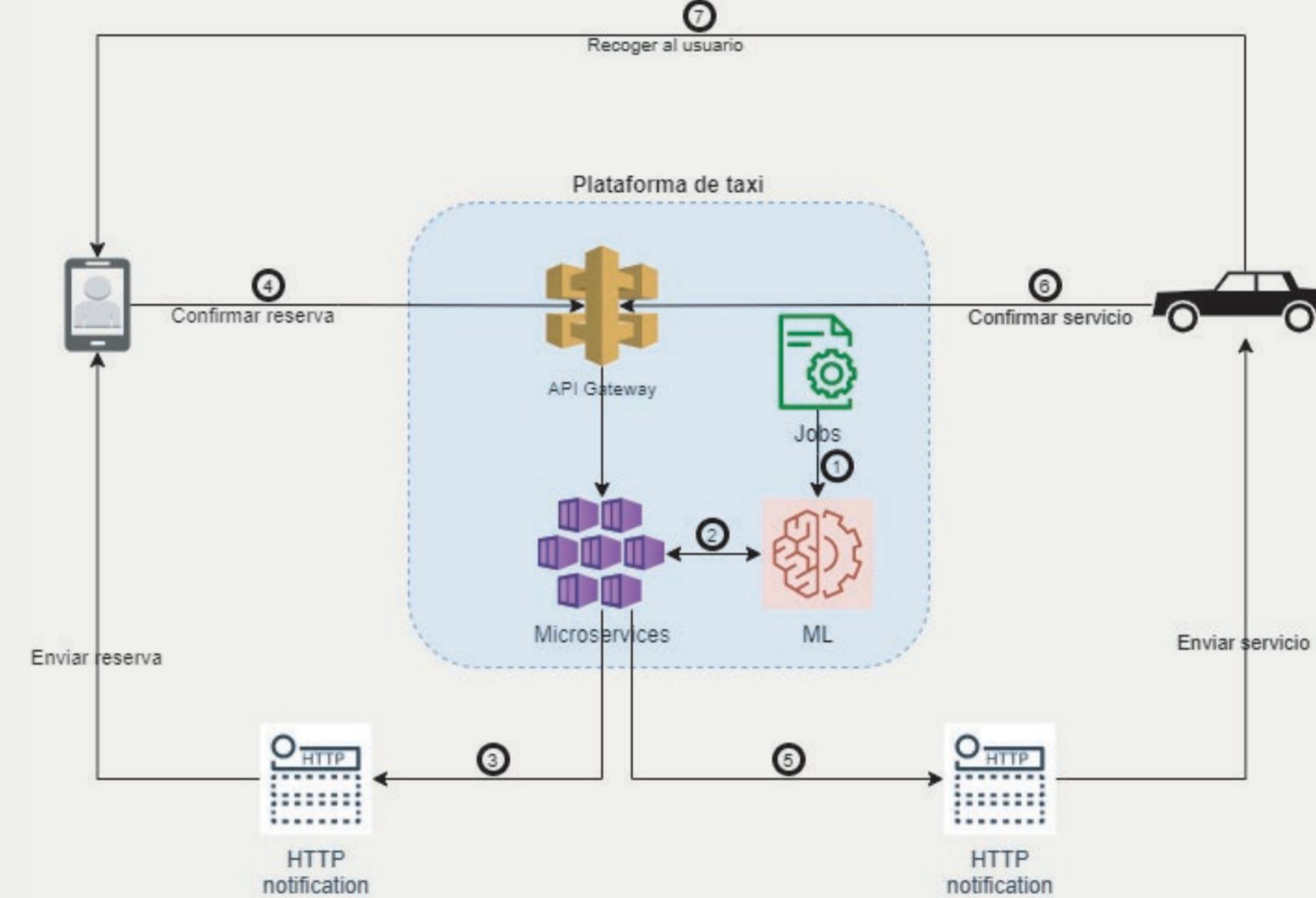


Figura 2: Gráfico de la solución

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 períodos 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.

Resultados

Como parte de las validaciones, se realizaron varias pruebas utilizando la información de 1.000 usuarios aleatorios durante una ventana de tiempo de 7 días a partir del 2022-04-23. Estas validaciones se realizaron con los modelos (probabilístico, KNN y Random Forest) dando como resultado una precisión muy baja lo cual no era lo que se esperaba de los modelos. A continuación, se encuentra el resultado de las métricas obtenidas de las pruebas realizadas:

Modelo	Precisión	Recall	F1	TN	TP	FP	FN
Probabilistic	0.4045	0.0927	0.1508	131838	301	443	2946
KNN	0.3304	0.0930	0.1451	1316	302	612	2945
Randomforest	0.4175	0.0748	0.1269	131942	243	339	3004

Tabla 1: Tabla de métricas de los modelos

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.

Referencias

- Yunxuan Li. "Taxi Booking Mobile App Order Demand Prediction Based on Short-Term Traffic Forecasting". In: Transportation Research Record (2017). doi: <http://dx.doi.org/10.3141/2634-10>.
- Lin Youfang. "Spatio-temporal LSTM for OD passenger demand prediction". In: Beijing Jiaotong Daxue Xuebao/Journal of Beijing Jiao-tong University 43 (2019), pp. 114-121.
- Hasan A. H. Naji. "Forecasting Taxi Demands Using Generative Adversarial Networks with Multi-Source Data". In: Applied Sciences 11 (2021), p. 9675. doi: <https://doi.org/10.3390/app11209675>

