Taxi Cancellations Predictive Modeling Using Neural Networks

**Business Problem**

Yourcabs.com is currently facing a critical business challenge where drivers frequently cancel scheduled rides, leaving customers stranded and dissatisfied. To address this issue, I utilized historical booking data to develop a predictive model to classify rides as “cancelled” or “not cancelled.”

Below, I present my findings and insights derived from the analysis.

A screenshot of a graph

AI-generated content may be incorrect.

A graph of a number of columns

AI-generated content may be incorrect.

It is clear when analyzing the cancellation and non-cancellation rates that most rides are completed without any problems, with cancellations accounting for a very small percentage of the dataset. On the other hand, cancellations exhibit a very consistent trend throughout the Iek, with a few minor exceptions on specific days. With a somewhat greater cancellation rate than other days, Fridays and weekends stand out.

A graph with green and purple lines

AI-generated content may be incorrect.

Moreover, bookings are typically at their highest in the evening. Even though the pattern of trip hours is similar, a few notable variations may indicate delays between the time of booking and the duration of the trip.

A graph with blue squares and green text

AI-generated content may be incorrect.

A graph with a bar and text

AI-generated content may be incorrect.

Most bookings are made offline rather than online, while online bookings typically have feIr cancellations. However, a slight but considerable percentage of cancellations occur even with bookings made online. Hourly rentals and point-to-point trips account for a higher volume of bookings and cancellations‬

A comparison of a graph

AI-generated content may be incorrect.

‬

A blue square with numbers and a blue square with black text

AI-generated content may be incorrect.

A graph showing a curve

AI-generated content may be incorrect.

With a precision of 94% and a recall of 95%, my model demonstrates great dependability in forecasting non-cancellations. These measurements demonstrate how Ill it detects rides that are not cancelled. However, with a f1-score of 22% and a precision of 25%, the model's capacity to forecast cancellations is considerably loq. This suggests that the algorithm has trouble correctly predicting situations in which rides are probably going to be cancelled. The model's moderate ability to differentiate between the two classes (cancelled and non-cancelled) of rides is indicated by its AUC score of 0.72.

The model does a good job of forecasting non-cancellations, but it must do a better job of precisely predicting cancellations, especially when it comes to reducing false negatives.

**Feature Importance Analysis**

This is a crucial step in understanding which factors contribute the most to the target variable, i.e., car cancellations. I used a Random Forest model to identify the factors.

These are the key findings(ranked):

1. lead\_time(0.0938): Represents the time difference between booking and travel start. A longer lead-time often correlates with the likelihood of cancellation.

2. to\_date(0.0927): Indicates the planned end date of the trip. Certain patterns in ride durations may influence cancellations.

3. user\_id(0.0874): Highlights user-specific behavior. Some users have frequent cancelling records.

4. distance(0.0796): Shows how longer trips are prone to cancellations.

5. from\_long(0.0675): Refers to Pickup location longitude. Indicates spatial patterns in cancellations.

Followed by, from\_lat, from\_area\_id, booking\_hmy, to\_area\_id, to\_long.

A graph of a graph showing the amount of a number of objects

AI-generated content may be incorrect.

**Recommendations**

1. **Focus on crucial predictors:** Prioritize characteristics such as travel type (e.g., hourly rentals and point-to-point trips) and wait time, since greater delays between booking and travel are associated with higher cancellation rates. Improve operations in locations with greater cancellation rates, particularly during peak evening booking times and on Fridays and weekends.
2. **Encourage online booking:** Online bookings usually have low cancellation rates than offline methods. Promoting the use of online platforms with tailored incentives or discounts may encourage more customers to book online, thus lowering cancellations.
3. **Improve resource allocation:** Increase driver availability in high-demand locations and during peak hours to better meet customer needs and reduce ride cancellations.
4. **Keep track of outstation trips:** Outstation trips have a significantly greater cancellation rate despite being small. These journeys need extra attention to identify and deal with the underlying causes of cancellations.
5. **Identify and manage major delays:** Reduce lead times and improve consistency between booking and travel times. This can increase customer satisfaction while considerably lowering the risk of cancellations.