

Hsiao-Wei Liu

Liangzhu Chen

Michael Liu

Yanmin Luo

Yanke Sun

Xinyu Wang

Jason Xu

Jiixin Xu

Aixin Zhou

Professor Josue Martinez

APAN 5900

5 December 2023

Results and Recommendations of Customer No-Show Predictive Model

Executive Summary

The project goal is to create a predictive solution for reducing customer no-show rates in maintenance appointments for Ryder System Inc., an American transportation and logistics firm renowned for its fleet management, supply chain, and transportation services. Our approach involves training a Random Forest model on 80% of 100,000 appointment records obtained through stratified sampling. Deliverables include an integration-ready model for the Operations Platform and data visualizations providing insights. From these insights, we offer recommendations to optimize resource allocation and enhance customer engagement.

Project Background

In this project, we are addressing Ryder's challenge of reducing customer no-shows for maintenance appointments. Our proposed solution is to develop a predictive model that accurately forecasts the likelihood of customers attending their scheduled appointments across 21 different time slots throughout the week, distributed over three daily shifts - morning, afternoon, and night. This model aims to enhance Ryder's operational efficiency by identifying the most reliable scheduling times, thereby minimizing the need for frequent rescheduling calls and improving customer satisfaction. Moreover, this predictive tool holds potential for integration with Ryder's automated scheduling system, offering optimal maintenance time slots for each customer or vehicle. This integration is expected to significantly reduce administrative costs associated with the current scheduling process.

Data Summary

In our analysis of three large datasets that encompass 1.4 million observations across 115 attributes. These datasets span four critical areas: Customers, Appointments, Fleet, and Geography. We categorized the data into four types: date, categorical, numeric, and binary. Our data showed that about 80% of the records are drive-up visits, which are not relevant to our analysis. We used the remaining 250,000 observations of appointment data as the master data to our predictive modeling.

Analytics Method

We first combined datasets from the years 2018 to 2020, creating a comprehensive pool for analysis. We employed correlation matrix, VIF test, and other techniques to select features. In

addressing missing values, specifically in `unit_checkin_date`, `appointment_priority_id`, and `pm_type`, which constituted 4% of the data, we opted to drop these values. Additionally, we enriched the dataset by introducing new `DateTime` features that capture details about weekdays and different time periods (morning, night, and afternoon). To ensure the data was machine-learning ready, we applied one-hot encoding, transforming all categorical data attributes into dummy variables.

An integral part of our approach was the ground truth labeling. We created labels based on whether the appointment and check-in dates, weekdays, and time periods aligned. If all these factors matched, the `Y` variable was labeled as 1, indicating the customer was not late; otherwise, it was labeled as 0. To achieve a balanced dataset, we conducted stratified sampling on all appointment data, resulting in 50,000 observations each for “show” and “no-show” categories (100,000 in total). However, our ability to perform stratified sampling was limited on weekends and night time due to low numbers of original data during those times.

For the final input data, we used dummy variables of customer information and appointment datetime data as the independent variables. The dependent variable was represented as the “`SHOW_UP`” column, which was structured in a binary format to indicate whether a customer showed up (1) or did not show up (0) for their appointment.

We developed a random forest model to achieve a balance of interpretability and robustness., which was trained on 80% of the data and tested against the remaining 20%. We also performed a randomized grid search for hyperparameter tuning to optimize our model. The model achieved an 88% AUC on test data and 85% AUC on train data, which indicated a strong consistency and no significant risk of overfitting. For both the train and test data, precision and recall rates for both show-up and non-show-up appointments were balanced.

In our feature importance analysis, we discovered that Appointment Priority IDs 5 and 6 are the most crucial, accounting for about half of the model's total feature importance.

Additionally, we found that night time appointments impact the model's outcomes more than morning or afternoon appointments. Beyond the top 10 most important features, the remaining 64 features collectively make up about a quarter of the total feature importance, indicating that while certain features are dominant, a wider range of factors also plays a key role in the model's performance.

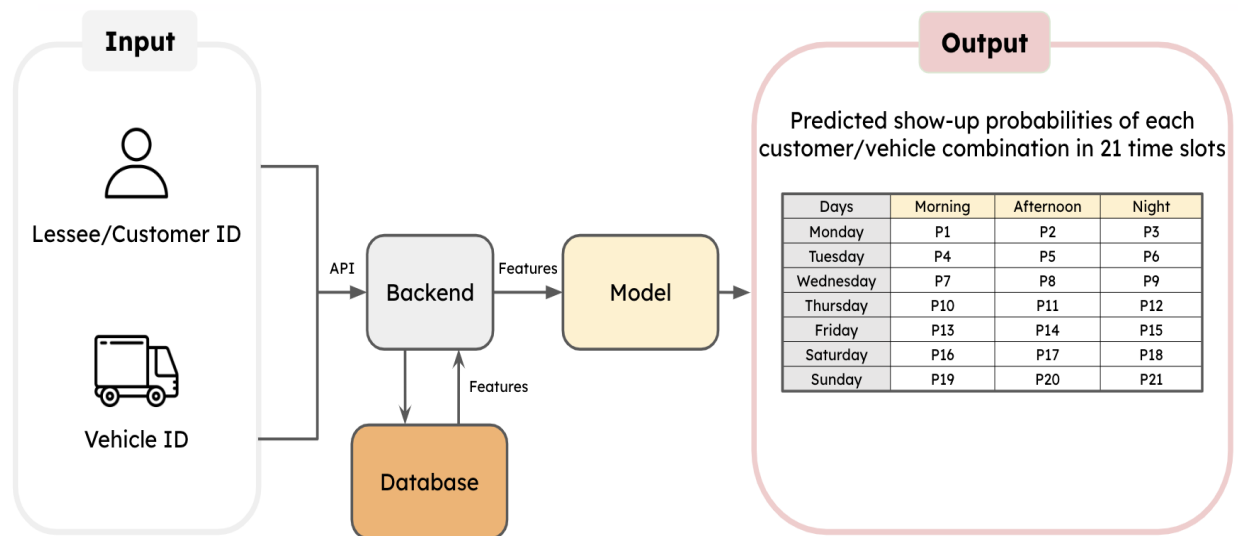
Business Use Case

Our model can be used in two ways: First, predict show-up rate by customer segments, and second, predict the show-up rate of an individual customer and vehicle ID combination. In order to gain business insights for a specific customer segment based on historical data, for example region or industry, we first use `predict.proba()` to get the show-up probability for all appointments data, then filter the predicted results based on selected segment criteria and calculate the average show-up rate for each time slot. We could also consider predicting binary outcomes using `predict()` for each row of data and compute the proportion of show-ups versus all data in each time slot to obtain the show-up rate; however, because the data contained low number of nighttime and weekend appointments, using this methodology led to inaccurate results, hence we pursued the “average method” instead.

To predict the show-up rates for an individual customer and vehicle ID combination, we filter all appointment data by the ID combination first. Because of the limited number of data available for one customer, we remove appointment date time features of all records and

duplicate the remaining personal information features on 21 artificial time slots, and finally use the same prediction mechanism to estimate the show-up rate across all time slots.

Ryder may integrate this use case with the auto scheduling platform. The diagram below shows the workflow of the proposed system:

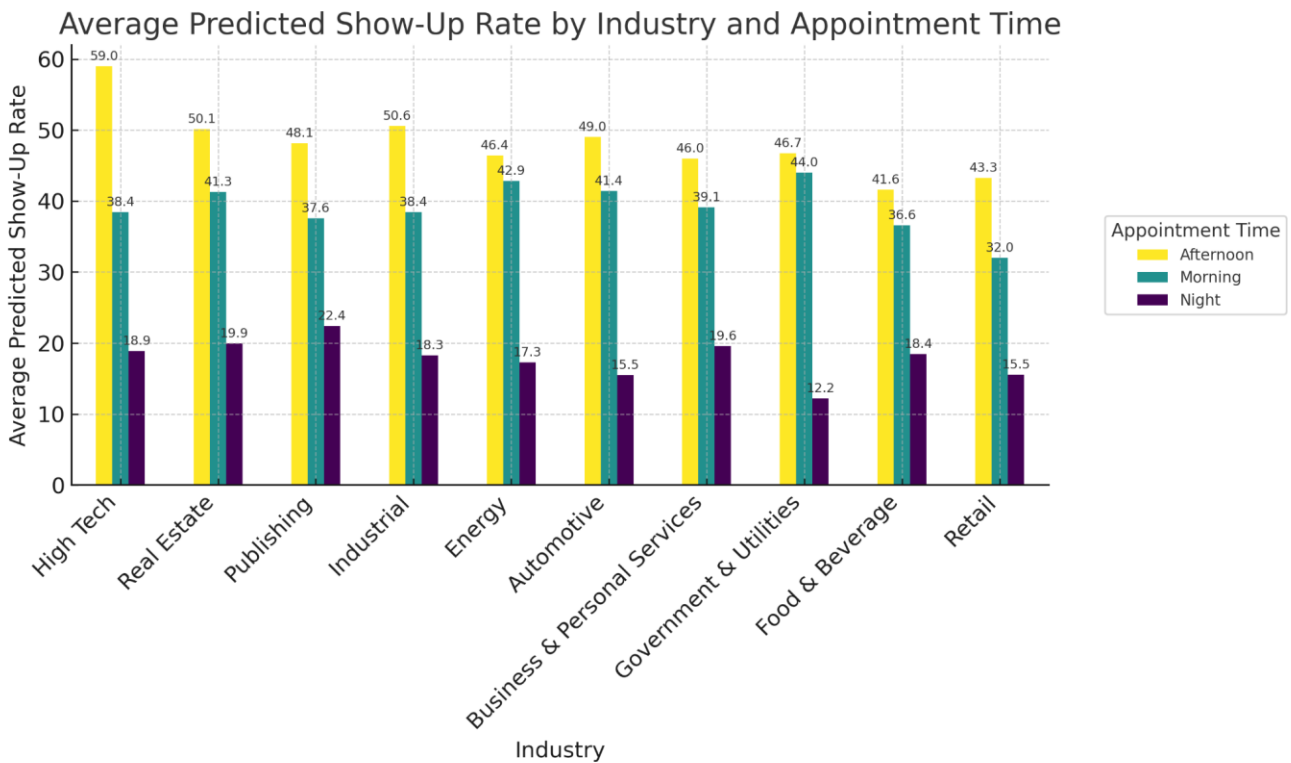


A customer would enter their personal and vehicle ID on a webpage, i.e., frontend, as inputs, and then the backend looks up the corresponding features in the database. With the attributes gathered, the model takes in the data and produces the output. Ryder may set a threshold for a show-up rate that triggers the system to automatically schedule an appointment for a customer in the optimal time slot, or, if the threshold is not met by the show-up rate in any of the 21 time slots, pick the time slot with the maximum show-up rate instead. If implemented successfully, this enables Ryder to cut human labor costs significantly through eliminating the manual appointment scheduling process.

Business Insights

Industries

In our analysis of various industries, we observed that there's no significant differences across industries. Notably, show-up rates vary considerably depending on the time of day. On weekdays, the show-up rate was moderate in the morning and peaked in the afternoon, then dropped significantly in the evening. However, during weekends, the pattern shifts, with mornings emerging as the time with the highest show-up rates, especially on Saturdays.



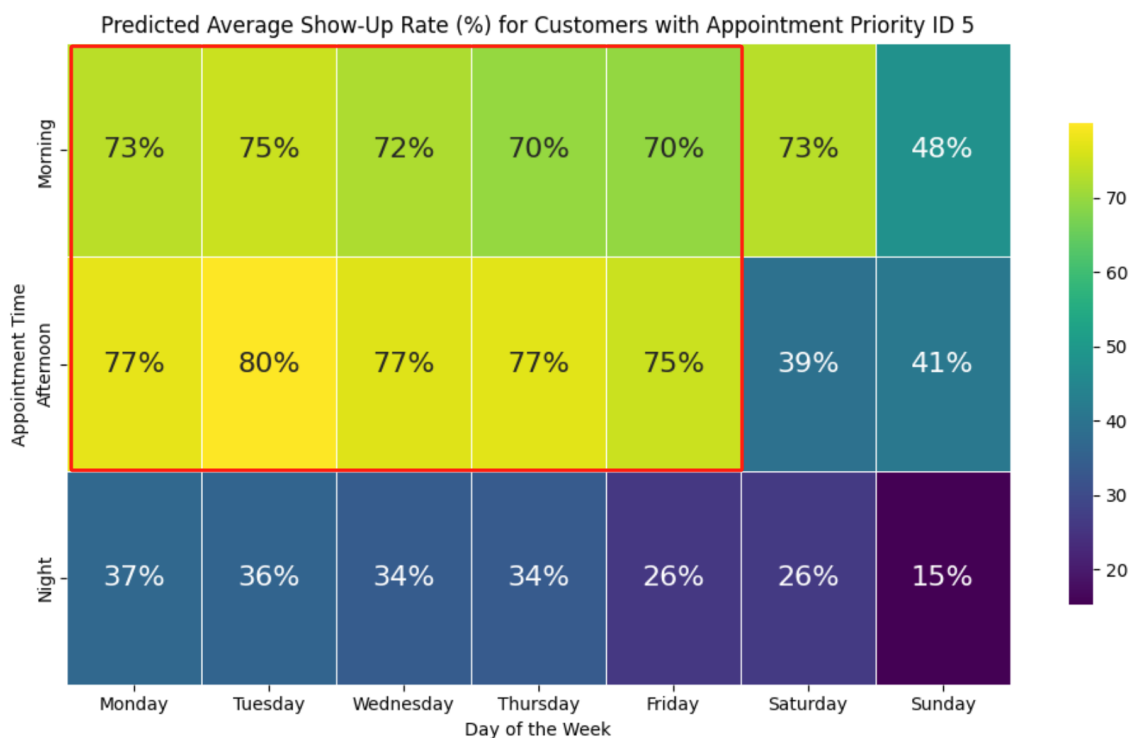
Appointment Priority IDs

In the Appointment Priority IDs, we observed that Appointment Priority ID 5 (hard schedule appointment) has the highest show-up rate, particularly on weekdays and around noon, showing a significantly high attendance rate. However, Appointment Priority ID 6 (Driver

Leaving with no Substitution Vehicle) has a relatively lower show-up rate, and customers with this type of appointment tend to concentrate on Monday to Wednesday evenings.

Vehicle Types

Based on provided data, It shows a distinct pattern in vehicle arrivals for appointments. Specifically, tractors, trailers, and trucks are more likely to show up on Saturday mornings than



at other times in the morning. This finding is consistent with a broader trend observed across all vehicle types. Generally, there is a higher likelihood of vehicles attending appointments during weekday afternoons, while the likelihood significantly drops during nighttime.

Regions

Our analysis across six regions, categorized into morning, afternoon, and evening slots, reveals a trend: customers generally prefer afternoon appointments for truck maintenance, with lower show-up rates at night. A notable exception is in the Central and Southeastern regions, where Saturday morning appointments see significantly higher show-up rates of 67% and 68%, respectively, exceeding the average afternoon rates across all regions. This insight is crucial for optimizing scheduling and minimizing no-shows.

Recommendations

Based on our insights, our first recommendation is to re-evaluate resource allocation according to customers' high probability show-up rates. The second recommendation is to enhance customer engagement through multifaceted initiatives. This includes implementing feedback surveys to delve into the preferences and patterns influencing appointment attendance, particularly during periods of low turnout. Additionally, we recommend deploying targeted communication strategies, such as email and SMS reminders, focusing on customers with reservations in historically low show-up slots. Complementing this, offering incentives to customers who book appointments during these less popular times could significantly bolster attendance rates.

These recommendations will help the company reduce the no-show rate, saving resources while gaining insights into customer behavior. This, in turn, will enable Ryder's team to provide more efficient and satisfactory services to customers, enhancing Ryder's competitiveness in the industry.