Customer Appointment Analysis in Automobile Dealerships

Priya Saha, Joachim Hubele and Stephen McDonald
CDK Global, USA
{Priya.Saha, Joachim.Hubele, Stephen.McDonald}@cdk.com

Abstract

Appointment scheduling prior to in-person visit to vehicle service centers is a well-known activity in our daily lives; we often do that to save ourselves time during our visits. Prior knowledge of customer arrivals help the service store management navigate through their daily demand and estimate a revenue goal. However in real-world, several people end up missing their appointments. To avoid business losses, dealers often practise engaging with customers beyond their capacity, but it only leads to operational inefficiency. In our work, we extrapolate several vehicle dealer stores that have high as well as low missed appointment rates and show interesting customer visit patterns post-scheduling. Additionally, we propose a Machine Learning based solution to empower dealers with optimal allocation of appointment slots among their customers and generate maximum revenue from the arrangement. Our motivation in the paper is to enhance the daily service demand process in the dealerships with excellent customer care.

Keywords: Appointment; Automobile; Vehicle Service; Predictive Overbooking; Dealerships.

Introduction and Motivation

Vehicle servicing is one of the stable business segments of the automobile Original Equipment Manufacturers (OEMs) and is driven by their dealerships. Revenue that is generated from servicing, is an aggregated earning of the various fixed operations done on the customer vehicles during their visits. Therefore, the customer visits influence the dealership business to a great extent. Many vehicle owners prefer to walk into the dealer service centers, wait in line and meet an advisor to process their request. Several other customers schedule their appointments online or via business development centers and visit on the appointment day. Scheduling an appointment ahead of visit is a very common customer behavior, but missing those appointments lead to significant business loss. As a solution, many dealers practise blind appointment overbooking [1] and intend to maximize their daily customer acquisition (and eventually maximize revenue). However, a lack of proper estimate of the walk-in demand, in addition to the overbooked and the regular scheduled appointments result in an operational inefficiency. Issues such as lack of shop capacity-control, multiple customer-collisions, increased wait-time, technician unavailability, parts-shortage are some of the widely known consequences. In the longer term, customer dissatisfaction can also impact the vehicle sales business. Dealers who do not practise overbooking, may suffer from reduced revenue generation as the appointment slots of the absent customers remain under utilized. Although a more organized methodology would improve the daily dealer operations, what is more important is to understand the causes of the missed appointments and how we could develop a robust controlled booking system to minimize the revenue loss. Here, we try to investigate a few studies done in many areas to reason why customers do not show up during their scheduled appointments.

With the power of the Predictive Analytics, scientists have found that the root cause of missing appointments (or being absent) varies industry to industry. In organizations such as office workplaces, employees have been found to be habitually not-at-work or habitually being late (Al-Rasheed 2021). Factors such as age, education, day of the week, month, distance from work, transportation costs etc. (Araujo et al. 2019), (Ali Shah et al. 2020) have been found to be associated to such behavior. Patients missing appointments, is also a very prominent occurrence in clinics, and it economically affects the care unit as well as resource planning significantly (Kheirkhah et al. 2015). Industries including hotels and airlines often overbook because they are aware many customers will miss their reservations. Revenue generated from the customers who are acquired through the overbooking is used to balance the loss (Sveinsdóttir 2019) created by the ones who did not show up. Although blind overbooking is a popular business method to get by, there are very high

[1]https://digitaldealer.com/dealer-ops-leadership/the-wrong-way-to-increase-service-revenue/

chances of customer-collisions leading to poor customer experience. Only recently, studies have found predictive overbooking has brought substantial improvements in clinic efficiency as well as profitability than blind overbooking and no overbooking (Hargreaves and Lin 2020), (Huang and Hanauer 2014). While there is a plethora of studies available to understand the absenteeism, missing appointments and popular solutions in various industries, there's a scarcity of studies related to vehicle servicing. We can draw parallels between automobile dealerships and other industries, however, the vehicle servicing operation dynamics can be much different in a few aspects. In dealerships, the customers do not pay any penalty for losing their appointment, as opposed to clinics where sometimes customers are charged a fee for not showing up. In airline or hotel industries, there can be a non-refundable booking fee. While dealerships often provide services to delayed customers, in airlines, hotels or clinics, the customers may have to face severe wait-time or unavailability of services.

In our paper, we discussed the data processing steps, followed by various interesting observations in customer visits and a proposed Machine Learning solution framework. We trained and tested our model with eleven months and one month of data respectively. Using our methodology, we tried to understand why the customers miss their appointments in dealerships. Given that the customers often exhibit this behavior, we also prescribed a few ways to leverage our solution and enhance the dealership operational planning. We finished the paper with our future work, acknowledgments and references.

Data Generation and Curation

We collected our data from CDK Service, a popular product of CDK Global [2] and competing against other well-known service providers such as Xtime, DealerFX, AutoLoop, etc. The product generates high business revenue and is being consumed across several OEM dealerships nationwide. CDK Appointment is one of the main components of the product which is configured in the dealership websites to book appointments. For an online scheduling experience, customers can visit dealer websites and book their service appointments. When the customer arrives in the dealership post-scheduling, the advisor in the dealer lane diagnoses the vehicle and generates a list of operations (often referred as repair-order) to be performed on the vehicle. In our work, we considered the data generated by the CDK Appointment instrumented in the dealerships of a top-tier US OEM with data consent agreement. Due to business security reasons, we were required to keep the OEM identification classified.

We studied two different datasets. First is the Appointment data, which is a repository of daily scheduling of the web appointments by the customers. Second is the Repair-Order data (service data), which is a collection of daily vehicle repair work at the service centers. The two datasets were processed to a refined form that helped us reach our findings. Our data collection is limited to 2021. We did not consider studying older data due to the bias that might have been created by the pandemic in 2020. Table 1 gives an idea of the refined data statistics.

Dataset 1: Appointment Data Our Appointment data spanned over 47 states and as many as 559 cities. Appointments can have 5 different status groups - “complete”, “no-show”, “cancelled”, “working” and “paused”. The “complete”, “working” and “paused” status indicate that the appointments have been serviced, or, are in the process of servicing. No visit by the customer on the day of the appointment without informing the dealership is referred as a “no-show”. A scheduled visit being cancelled by the customer is referred as “ cancelled”. In our paper, the “no-show” and “cancelled” combined is referred as the “missed” appointments.

Dataset 2: Repair-Order Data The availability of a customer Repair-Order (RO) record of a vehicle at the dealership confirms that the owner had visited the dealership to get his car fixed. Similarly, an unavailability of a RO record indicates that the customer did not show up. ROs are generated for walk-in as well as web appointment scheduled customers. In our data curation step, we discarded the walk-in ROs.

Data Curation Although, we had a status available to understand the state of the appointment in the Appointment data, due to process issues in the dealerships, we added a pre-processing step. Our RO records show a significant time-lag with appointment time due to the customer wait time involved in the dealership during visit. Therefore, our extraction step only involved the date of the appointment and the RO. For every month, we aggregated and validated all the appointments and their respective ROs. At first, for every web appointment, we recovered all the ROs based on the VIN (Vehicle Identification Number) and the appointment-RO date; this step accounts for all the vehicles that were serviced on the same day as the scheduled appointment at any dealership. Second, for every web appointment, we recovered all the ROs based on VIN and store id; here we made sure we captured the RO service records that were delayed more than a day as well as serviced at the same store as the appointment. Third, we extracted the ROs based on the customer information (email, phone number) for appointments that do not have VINs (such as guest customers). If we could not find a valid RO for an appointment in the same month, we marked the appointment “missed”. We combined all the months to analyze the dataset. The curated data was used for building our solution framework.

[2] https://www.cdkglobal.com/fixed-ops/service/cdk-service

Table 1: Descriptive statistics about the dataset for the year 2021.

Statistic	Measurement
Start Date	Jan, 01
End Date	Dec, 31
Number of Appointment Records	334,434
Number of Dealerships	590
Number of Cities	559
Number of Vehicles	265,112
Min Vehicle Year Serviced	1978
Max Vehicle Year Serviced	2022

Analysis and Observations

At first, we studied our original Appointments data. We aggregated all the "completed" and the "missed" appointments by stores. Note that only "completed" status means a revenue was generated without any time restrictions. We referred to the stores that have higher "missed" appointments than the "completed" appointments as *Under-Performers* and the stores that have higher "completed" appointments than the "missed" ones as *Over-Achievers*. There were 204 *Under-Performers* and 386 *Over-Achievers* in our data. Almost 77% appointments came from the *Over-Achievers* and 23% came from the *Under-Performers*. We tested the normality of the monthly appointment frequency distribution of the cohorts and found that none of them follow a normal curve. The KS statistics of the *Over-Achievers* and the *Under-Performers* were 0.96 and 0.95 (pvalue = 0.00) respectively. Due to uncertainty of the existence of the normal distribution, we chose to examine our analysis with Mann-Whitney test as discussed below. Furthermore, in Table 2, we show that the distributions are vastly different due to their high and statistically significant U value (Test Num 1).

Mann-Whitney U test The test compares the difference between two independent groups that are not normally distributed. The U value determines the difference. In this test, we have - **Ho:** The two represented groups have the same distribution of scores, and, **Ha:** The groups are different in their shape or spread. We can reject the null hypothesis by supporting the test with statistically significant pvalue (0.05 or less) of the U statistic.

Days of the Week We extracted the days on which the appointments were scheduled in both the cohorts and found that the number of appointments were high on the week days and low on the weekends. The Mann-Whitney test as reported in Table 2 (Test Num 2) concluded that the weekly appointment distributions were different in the cohorts.

Geographic Region We mapped the dealership stores according to the states. Although majority of stores were from some of the very populous states, such as TX, CA, FL, PA and OH (these states also have many cities), they were disproportionately distributed.

Table 2: Mann-Whitney U test Resuts

Test Num	Test Name	U-value	pvalue
1	Appointment Frequency	9499779995.00	pvalue < 0.01
2	Days of the Week	9925342771.50	pvalue < 0.05
3	Delayed Visits	3266618.00	pvalue < 0.01
4	Days Lagged	7771307.00	pvalue < 0.01
5	Days-Out	8917610592.00	pvalue = 0.00

In Table 3, we described the total number of dealer stores as well as the number of *Under-Performers* and *Over-Achievers*. Out of 42 stores in PA, 30 were *Over-Achievers*, and 30-out-of-37 stores in OH were *Over-Achievers*. In contrast, TX and FL had almost equal division. Also, we found that states such as NJ and MA have 11-out-of-17 and 18-out-of-23 *Over-Achievers* respectively. Therefore, we concluded that stores in a few states in general have less missed appointments.

Delayed Visits The lack of penalty in dealerships not only results in missed appointments, but also in visits that are not standardized by the appointment time on the same day. There are many customers who schedule appointments and visit dealerships at a much different time on the same day. However, this creates a confusion on the dealer's side; such a behavior may not be anticipated ahead of time, thus resulting in operational inefficiency. In the *Over-Achiever* and the *Under-Performer* cohorts, we found about 1.07% and 3.69% of the appointments respectively were visited by the customers with a significant time-difference. Figure 1 shows the percentage cohort population according to the time-difference. We observed that the *Under-Performers* demonstrated a tendency of visiting the dealerships before their appointment time, and it is more than the *Over-Achievers*. One reason to explain such behavior might be due to the high wait-time in the *Under-Performers*; the customers prefer to drop-off their vehicles earlier than scheduled. Additionally, we noticed that the peak time difference between appointment time and the visit time is between 0 to 4 hours. The Mann-Whitney test as reported in Table 2 (Test Num 3) confirmed that the time-difference distributions of the cohorts were different from each other.

We also discovered that there's a tendency of customers to not show-up on appointment day at all, but visit the dealership with a time-lag of one-to-several days. There could be one or more visits post the scheduled appointment. For simplicity, we have computed the most recent visit after the scheduled appointment and measured the days-lag based on that. This behavior was observed in 33.41% and 55.58% of the appointments in the *Over-Achiever* and the *Under-Performer* cohorts respectively. We confirmed with a Mann-Whitney test (Test Num 4 in Table 2) that the days-lagged delay distributions of the cohorts were different from each other.

It is to keep in mind that the above characteristic is observed in only a sample of the population. It is hard to know ahead of time if a customer would be delayed, as a not-showing-up on the scheduled time does not always

Table 3: Comparative analysis of the *Under-Performers* and *Over-Achievers* according to their locations.

Dealer State	Counts		
	Total	*Under-Performers*	*Over-Achievers*
TX	55	29	26
FL	33	16	17
PA	42	12	30
CA	33	12	21
MI	28	11	17
MA	23	5	18
NJ	17	6	11
OH	37	7	30

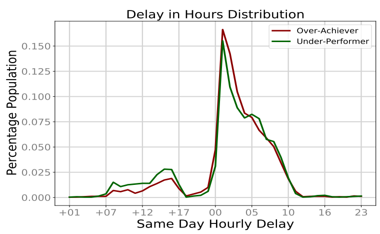


Figure 1: Distribution of the difference (in hours) in the appointment and the visit time in both cohorts. A '+' refers to cases where the visit was made prior to the appointment, and no '+' refers to cases where the visit was made later than the appointment. For example, '+01' means the visit was made one hour prior to the appointment and '01' means the visit was made one hour later to the appointment.

guarantee a delayed visit. There is also a significant part of the population who missed their appointment; they never visited the dealership post the appointment date.

Days-Out Days-out is the number of days between the day of the appointment scheduling and the day of the visit. It is often assumed, that a high days-out may lead to customer forgetfulness and eventually result in a missed appointment. We studied the 4 days distribution for the missed appointments in each of the cohorts with respect to the days-out (see Table 4). The *Under-Performers* show higher missed appointments in the first 4 days than the *Over-Achievers*. Thus, the first 4 days seem to be very critical for scheduled appointments in *Under-Performer*. The Mann-Whitney test in Table 2 (Test Num 5) confirmed that the distributions of the two cohorts were different from each other.

Determining Missed Appointments and Delay

Irrespective of whether a store is an *Under-Performer* or an *Over-Achiever*, a missed appointment is a missed revenue opportunity. The biggest pain-point for dealers is not being able to identify the appointments that may not be visited. In addition, we already have seen that many customers do not even follow their appointment time strictly. Therefore, we decided to build a predictive solution that can give dealers an edge. We built a state-of-the-art Machine Learning framework to identify the appointments that are likely to be missed or delayed, and eventually create opportunities for dealers to collect more revenue. We proposed a chained ensemble architecture as depicted in Figure 2, to address two problems - (i) whether a web appointment would be visited by a customer at all in the month, and (ii) if visit is ensured, would the visit be aligned to the scheduled appointment time. The 2-fold structure was built using LGBM Classifier algorithm and was trained with data from January through November, while being tested on data from December. For the model development, we considered all the appointment records that were validated through our curation step.

Table 4: Days-Out comparisons of missed appointments among *Under-Performers* and *Over-Achievers*.

Days Out	Percentage Population	
	Under-Performers	*Over-Achievers*
0	8.77%	5.50%
1	16.05%	12.09%
2	11.19%	9.87%
3	8.79%	8.21%

LightGBM The LightGBM (LGBM) is a gradient-boosted ensemble algorithm where the trees grow leaf-wise. Hence, it is very efficient in speed, performance and memory usage (Ke et al. 2017). The LGBM Classifier can be evaluated using the metrics described below, where TP= True Positive, TN = True Negative, FP = False Positive and FN = False Negative outcomes of the model.

- Accuracy (TP+TN/TP+TN+FP+FN): A fraction of correct predictions.
- Precision (TP/TP+FP): Measures proportion of correct positive identifications.
- Recall (TP/TP+FN): Measures proportion of actual positives that were identified correctly.
- F1-Score (2*Recall* Precision/Recall+Precision): A weighted average of Precision and Recall.

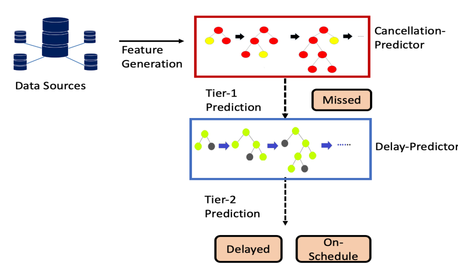


Figure 2: Proposed Model Framework Solution.

Proposed Solution Architecture First, we built a binary classification model (named, *Cancellation-Predictor*) using LGBM algorithm to predict whether an appointment would be visited post-scheduling. We used features that best describes an appointment such as, Email-Opted (whether customer record has an email id), Vehicle-Year (manufacturing year of the vehicle), Sms-Opted (whether customer record has a phone number to receive text), State (state location of the dealer), Mileage (mileage of the vehicle when appointment was made), Days-Out (number of days between the day of the appointment scheduling and the day of the visit), Mon-Appt (month of the scheduled appointment), Mon-Appt-Loop (month when appointment was created) and Day-Of-Week (day of the week of the scheduled appointment). For the categorical features, we used label encoding. Due to the high class imbalance, we up-sampled our minority class to have proportionate class distribution. The optimal hyper-parameters of the *Cancellation-Predictor* were retrieved from the parameter search space as discussed in Table 5. In Table 6, we reported the evaluation metrics collected from the best iteration.

Table 5: Model Parameter Tuning Search Space.

Model Parameter	Model	
	Cancellation-Predictor	*Delay-predictor*
Learning Rate	[0.03,0.10]	[0.02,0.07]
Min Data in Leaf	[200,400]	[100,300]
Number Leaves	[200,400]	[200,300]
Objective	['binary']	['binary']
Max Depth	[5,7]	[3,5]
Number of Boost Rounds	[1000,2000]	[1000,2000]
Early stopping	[50,100]	[20,50]

Table 6: Representing Model Accuracy Metrics.

Model Test Statistic	RF Model		LGBM Model	
	Cancellation-Predictor	*Delay-predictor*	*Cancellation-Predictor*	*Delay-predictor*
Accuracy	0.82	0.70	0.87	0.74
Precision	0.84	0.58	0.85	0.61
Recall	0.77	0.73	0.88	0.85
F1-Score	0.81	0.65	0.87	0.71
Precision-Recall AUC Score	0.86	0.71	0.87	0.67

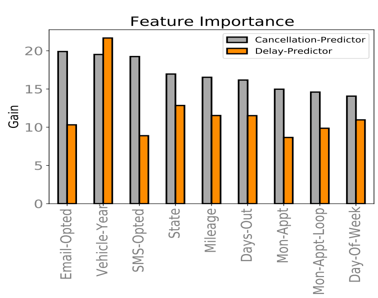


Figure 3: Feature Importance of Predictor models by gain.

Our next step was to predict whether the appointments that had high probability of visiting the dealership were punctual (within 4 hours of appointment time) or delayed (delay refers to a visit that is beyond 4 hours of the appointment time, we accounted the 4 hours as the wait-time). We built another binary classification model (named, *Delay-Predictor*) using LGBM. In addition to the baseline features, we trained the *Delay-Predictor* with the predicted outcome of the *Cancellation-Predictor*. The class imbalance was tackled using up-sampling of the minority class. The parameter search space for the optimal performance of the model is reported in Table 5. Using the optimal hyper-parameters, the performance metrics of the test set is reported in Table 6. Although we found that the performance of the *Delay-Predictor* is slightly lower than the *Cancellation-Predictor*, the former is doing an excellent task in classifying the Negative Class (Delayed). We show the confusion matrix in Figure 5 and demonstrate that out of 48% of the total predicted delayed population, 42% was identified accurately using our model. This gave us a good confidence that our model can be robust in predicting class of the appointments. We also evaluated our competency with a Random Forest stack (model metrics in Table 6). Finally, we chose to build the architecture with the LGBM models due to less training time and superior performance in key metrics.

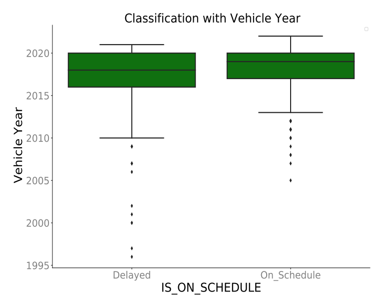
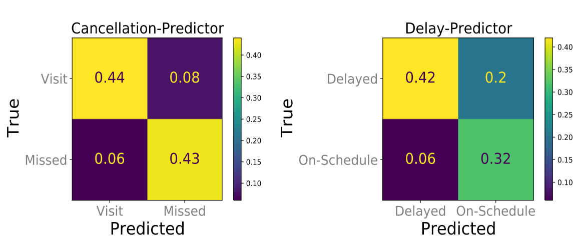


Figure 4: Vehicle year distribution in the Delay-Predictor segments.

Using our 2-tiered architecture, we extracted segments of population. From the Tier-1, we derived a segment that was predicted by the *Cancellation-Predictor* as *Missed*. The *Missed* segment is a set of appointments that were at high-risk of not getting visited by the customers within the month. We observed that when 51% population was predicted *Missed*, we correctly identified 43% (Figure 5a). The part of the population that was predicted as not *Missed* was passed through the Tier-2. We extracted two more segments from the process: *Delayed* and *On-Schedule*. The *On-Schedule* segment consists of the appointments which were expected to be visited within 4 hours of their appointment time. The *Delayed* segment represents the appointments which were expected to be visited beyond 4 hours to any day within the month. Our model predicted 48% and 52% of the population as *Delayed* and *On-Schedule* respectively. Out of 52% predicted *On-Schedule* segment population, 32% was correct. Additionally, 42% of the *Delayed* segment was correctly found when the model predicted that 48% (Figure 5b).

In Figure 3, we show the importance of the baseline features. We used the "gain" parameter as importance type, which computes the average gain of the feature when it is used in trees. Although the ranks of the features are different in each of the models, we found the manufacturing year of the vehicle as a very critical one. Figure 4 shows that the vehicle year distribution in the *Delayed* segment has a much longer tail than the *On-Schedule*. Customers who own relatively newer vehicles tend to be more punctual of their appointment. The median vehicle year of the *On-Schedule* segment is also more recent than the *Delayed*. Thus, it is fair to say that with the age of the vehicles, customers start falling into the *Delayed* segment. They may also tend to service in independent shops as the vehicles get old. Other critical common features are the dealer state, mileage, days-out and day-of-week. As we have already seen earlier, dealer location is very pronounced in the *Over-Achievers* and the *Under-Performers*; it is not surprising that it highly contributed towards the missing appointment decision-making. It is also interesting to note that probably consumers with lower mileage are prone to more servicing on time than the other vehicles with higher mileage, as mileage gets accumulated with the age of the vehicle.

Prescriptive Solution in the Dealership Business Model Using our 2-tiered architecture in the dealership business, we prescribed a segmented solution for booking appointments: (i) *Missed* segment - once this segment of customers is identified, we recommend the service center to connect to them and provide them incentives that can entice them to the dealership; this can also help the dealer build a relation with the customer and eventually win trust. Additionally, the appointment slots from this segment can be used to acquire more customers; (ii) *Delayed* segment - the dealership can open up the slots that were scheduled by this segment and accept appointments from other customers; (iii) *On-Schedule* segment - this is the most trusted segment of the dealership business, hence we recommend giving them elevated customer experience. The segments can be leveraged easily by integrating our framework within the dealer appointment system.



(a) (b)

Figure 5: Confusion Matrix with population annotated in a scale of 0 to 1.

Conclusion and Future Work

In this paper, we have explored the visit characteristics of the customers who schedule web appointments with the dealerships in a prominent OEM in USA. Not only we unveiled the various customer visit patterns, we also proposed a 2-tiered machine learning based solution that can be used ahead of time to optimally allocate the appointments among customers. With a better estimate of the appointments that have higher likelihood to get missed or delayed, dealers can open up the booked slots to other customers for a controlled booking and servicing. Such arrangement can help the dealers adjust their walk-in demand and resources. This also helps dealers generate more revenue and win long-term customer loyalty. Ultimately, dealers get a much more organized system, better revenue and a higher customer satisfaction. Our work can also be extended to understand other OEMs.

Acknowledgments

We thank Brad Bowers (Director, Sales), Guillermo Leon (Director, Customer Success) and Tammy Horner (Sr Dir, Manufacturer Solutions) from the CDK Global team to support our work.

References

Al-Rasheed, A. 2021. Identification of important features and data mining classification techniques in predicting employee absenteeism at work. *International Journal of Electrical & Computer Engineering (2088-8708)* 11(5).

Ali Shah, S. A.; Uddin, I.; Aziz, F.; Ahmad, S.; Al-Khasawneh, M. A.; and Sharaf, M. 2020. An enhanced deep neural network for predicting workplace absenteeism. *Complexity* 2020.

Araujo, V. S.; Rezende, T. S.; Guimarães, A. J.; Araujo, V. J. S.; and de Campos Souza, P. V. 2019. A hybrid approach of intelligent systems to help predict absenteeism at work in companies. *SN Applied Sciences* 1(6):1–13.

Hargreaves, C., and Lin, T. S. 2020. A machine learning overbooking algorithm for enhancing clinic efficiency.

Huang, Y., and Hanauer, D. A. 2014. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Applied clinical informatics* 5(03):836–860.

Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; and Liu, T.-Y. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30:3146–3154.

Kheirkhah, P.; Feng, Q.; Travis, L. M.; Tavakoli-Tabasi, S.; and Sharafkhaneh, A. 2015. Prevalence, predictors and economic consequences of no-shows. *BMC health services research* 16(1):1–6.

Sveinsdóttir, A. G. 2019. *Passenger overbooking, using no-show probability.* Ph.D. Dissertation.