Data-driven Models for Predicting No-show Rates and Service Times in Outpatient Appointment Scheduling

Moustapha Fall
Computer Science and Digital Society
(LIST3N)
University of Technology of Troyes
12 Rue Marie Curie, Troyes, 1030.
moustapha.fall@utt.fr

Ilhem Slama
Laboratoire Génie Industriel
Centrale Supélec, Paris
3 Rue Joliot Curie, Gif-sur-Yvette, 9119.
ilhem.slama@centralesupelec.fr

Yassine Ouazene
Computer Science and Digital Society
(LIST3N)
University of Technology of Troyes
12 Rue Marie Curie, Troyes, 1030.
yassine.ouazene@utt.fr

Achraf Jabbeur Telmoudi
LISIER Lab
Higher National Engineering School of Tunis (ENSIT)
Tunis, Tunisia.
achraf.telmoudi@ieee.org

***Abstract*—Outpatient clinics are integral to healthcare, offering vital services without the need for hospital admission. However, appointment scheduling in these settings remains challenging due to uncertainties such as patient no-shows and variable service times. This study proposes a data-driven approach to minimize physician idle time and patient waiting time by analyzing an eight-year (2016–2023) dataset from primary and specialized care for American veterans. Adopting the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, four predictive models: Random Forest, Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Artificial Neural Networks (ANN), were developed for both classification (no-show) and regression (service time) tasks. The ANN model demonstrated superior predictive performance in both domains. The key predictors of no-shows included waiting time, type of care and care provider, while type of care, care provider, and veteran ZIP code were the most influential in forecasting service time. These findings highlight the potential of machine learning to improve appointment scheduling in outpatient clinics.**

***Index Terms*—Outpatient Appointment Scheduling, No-shows, Service time, Machine Learning Models.**

I. Introduction

Outpatient clinics play a vital role in modern healthcare systems [1], providing affordable, high-quality, and easily accessible medical services without the need for hospitalization, with stays typically lasting less than a day. This sector has seen significant growth, with the global outpatient care market estimated at €460 billion in 2024, projected to reach €500 billion by 2029, at an annual growth rate of 1.68% [2]. This expansion can be attributed to the ability of outpatient clinics to address major challenges such as geographic accessibility and an aging population, while maintaining quality care at affordable prices. To sustain this progress, it is crucial that outpatient clinics optimize resource use and enhance patient satisfaction through effective scheduling systems.

This work was partially funded by the Grand Est region in France as part of a collaborative project for Research, Development and Innovation under agreement number 23P07205.

However, efficient scheduling remains a significant challenge due to uncertainties [3]. For example, patients may arrive late or not at all for their appointment, or spend more or less time with the doctor than planned [4]. These uncertainties (or sources of variability) disrupt clinic workflows, creating inefficiencies, and negatively impacting patient and provider satisfaction.

This study takes advantage of the increasing availability of large datasets in clinical contexts to address two primary sources of variability in outpatient scheduling: patient no-show probabilities and variability in service times.

Machine learning algorithms (ML) provide powerful tools for analyzing patient behavior [5] and predicting the duration of service times [6]. While traditional approaches often relied on small, clinic-specific datasets, exploiting big data (e.g., multi-clinic, demographic diversity) now allows for more robust and generalizable predictions [7], addressing biases typically associated with earlier methods. By examining a large and representative dataset, this study aims to answer three key questions. Which ML algorithms perform best in predicting these sources of variability? What key factors influence these predictions? And what are the theoretical and practical implications that can enhance clinic operations?

The remainder of this paper is organized as follows. Section II reviews the relevant literature, III describes the solution approach, IV details the data mining steps, V discusses key findings and their implications, and finally, VI concludes the study.

II. Literature review

Research into appointment scheduling in ambulatory clinics draws on a rich history of operations research and quantitative systems analysis [3], beginning with Bailey's pioneering work in 1952 [8]. Since then, many advances and innovative approaches have been made in this field. The literature analysis can be broken down into two categories of approach: studies that tackle the problem using optimization models and others that reduce uncertainty using ML techniques.

A. Optimization methods

Classical optimization approaches structure decision making on three levels: strategic (long-term resource allocation), tactical (medium-term planning and resource adjustment), and operational (day-to-day appointment setting) [9]. Scheduling models aim to maximize resource utilization while reducing patient waiting times, often using techniques such as block scheduling, overbooking strategies, and dynamic slot management [10]–[12].

Despite their robustness, traditional models often rely on simplifying assumptions, such as average no-show rates or service times, that ignore patient heterogeneity and stochastic variability [3]. These deterministic approaches, though analytically tractable, frequently underperform in real world settings, resulting in suboptimal patient throughput, increased waiting times, or underuse by clinicians [6].

B. Prediction methods

To address these limitations, recent research increasingly integrates ML-based prediction techniques, leveraging clinical data sets to improve scheduling accuracy [13]. ML approaches enable flexible modeling of complex, non-linear relationships and help identify the most relevant features driving scheduling variability [14]. With the growing availability of healthcare data and advancements in predictive analytics, clinics now have new opportunities to optimize appointment systems [15].

These methods primarily target two key sources of uncertainty: patient no-show probabilities and consultation service time variability.

No-show prediction models typically incorporate variables such as patient demographics, previous attendance patterns, lead time, and provider-related factors. Advanced algorithms, including ensemble methods such as Random Forest and Gradient Boosting, as well as neural networks, have demonstrated superior predictive performance compared to traditional statistical techniques [16], [17]. Their precision is commonly assessed using metrics such as the Area Under the Curve (AUC), precision, recall, and the F1 score [18].

Service time prediction is generally treated as a regression task, using patient characteristics, physician attributes, and appointment details to estimate consultation durations [6]. These predictions support more precise, individualized slot allocations, better aligning capacity with expected demand. Evaluation metrics typically include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) [18].

Other sources of variability, such as patient arrival time or in-clinic waiting durations, have also been modeled using time series forecasting techniques (e.g., SARIMA, Simple Exponential Smoothing) and supervised ML approaches [14], [19]. However, many of these studies are limited to single-institution datasets or address only one dimension of variability, reducing the generalizability of their findings in diverse healthcare settings.

C. Towards Generalizable, Data-Driven Scheduling

Classical optimization methods, while foundational, struggle to address real-world uncertainties due to their reliance on homogenized assumptions. Solely relying on these methods risks inefficiencies such as prolonged wait times, clinician idle time, and staff frustration from rigid, suboptimal schedules. However, ML offers a paradigm shift by leveraging clinic-specific data to predict and mitigate these uncertainties. ML models, such as ensemble methods and neural networks, outperform traditional approximations in predicting no-shows and service times, enabling dynamic, patient-centric scheduling.

Despite their promise, existing studies remain constrained by narrow datasets (often single-clinic) and a fragmented focus on isolated variability sources. This limits the generalizability of the proposed solutions in various care settings.

This study develops a ML framework to predict patient no-show probabilities and service time variability, leveraging multi-clinic data to enhance generalizability across ambulatory care settings. By replacing rigid optimization assumptions with dynamic, data-driven predictions, the methodology enables clinics to tailor appointment allocations to patient-specific behaviors, improving scheduling efficiency without relying on deterministic heuristics.

The following section details this methodology, emphasizing its adaptability to diverse ambulatory care contexts.

III. Solution approaches

Fig. 1 provides a detailed view of our approach. It incorporates the various phases that make up the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. Five of the six phases were used to predict these sources of variability. The exploration process took us from business understanding to evaluation, through data understanding, data preparation, and modeling. In the latter, four data mining models were used for both classification (no-show probability prediction) and regression (service time prediction) tasks: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Artificial Neural Networks (ANN). Each classification model is evaluated using the following metrics: Precision, recall, area under curve (AUC), F1 score, and accuracy. Similarly, each of the regression models is evaluated with the following metrics: Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), taking care to exclude no-shows case. The importance of the predictors is then assessed for each of the eight optimized models. The two models with the best trade-offs between the performance and reliability best score importance variable were selected to provide guidelines for outpatient clinics.

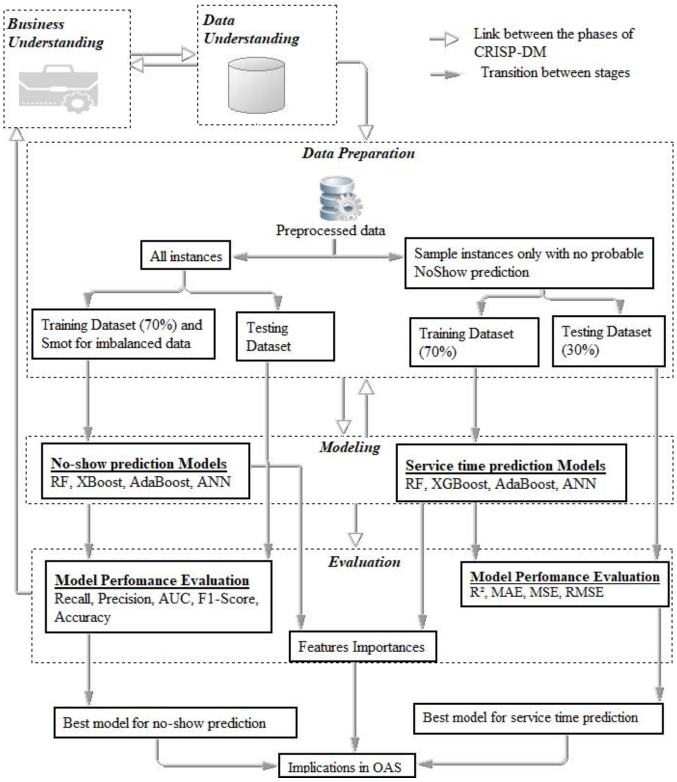


Fig. 1: Detailed Research Framework With CRISP-DM Steps

IV. Data mining

The adoption of a rigorous methodological approach is essential in data mining to guarantee the reliability of processes, ensure the reproducibility of analyses, and make the results accessible to stakeholders, even at a low technical level. CRISP-DM, one of the most widely used methodologies, offers a structured and iterative framework that facilitates project management while ensuring the robustness of the analysis.

A. Business Understanding

Ambulatory clinics face a critical challenge in balancing resource optimization and patient satisfaction, as shown in Fig. 2. Key sources of variability, patient no-shows, unpunctuality, and inconsistent service times disrupt operations by increasing physician idle time or patient waiting times. For instance, no-shows and late arrivals lead to underutilized resources, while prolonged service durations delay subsequent appointments. In contrast, shorter-than-expected consultations waste capacity and walk-ins or service interruptions exacerbate delays. These inefficiencies reduce patient satisfaction (due to long waits) [20], [21] and resource optimization (due to idle time) [22], creating a cyclical detriment to clinic performance. To address this, the study focuses on predicting no-shows and service-time variability, aiming to mitigate uncertainty and enhance both resource optimization and patient experience. This approach is consistent with previous research that emphasizes the need to tackle these dual challenges in outpatient scheduling [3], [22].

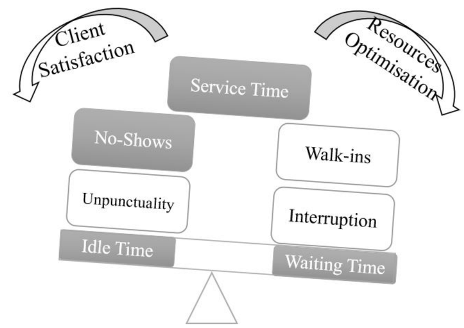


Fig. 2: Trade-off Between Resource Utilization and Patient Satisfaction, Given Uncertainty

B. Data Understanding

A description and context of the data used was provided [23]. These data come from the Veterans Health Administration (VHA) and Community Medical Center (CC) Facility Wait Time Study. The raw anonymized data include more than 51 million consultations over 10 years, from January 2014 to August 2023. They provide a framework for analyzing this set of data on a national scale. The scandal surrounding excessive waiting times at VHAs in 2014 led to the creation of the Veterans Choice Program (VCP). It allows veterans whose estimated travel time to the nearest VHA facility is more than 60 minutes to receive care through VHA-contracted Community Centers.

In a more recent study based on earlier findings, care was classified into three types: primary care clinics, mental health services clinics, and clinics specializing in other fields [24]. Their work identified two of the three typical environments encountered in the literature [3]: groups of clinics providing primary care and those providing specialist care.

We started by eliminating redundant columns, especially those that aggregate other variables (*dtot*, which represents the total waiting time derived from two other waiting times), as well as superfluous columns such as *year*, already included in the *activitydatetime* column.

The VCP program was introduced in 2014, making 2014 and 2015 pivotal years. As a result, we have excluded these years and focused on an eight-year period from 2016 to 2023.

The variable *non_va* indicates the local health care provider in charge of the patient. It is a binary variable that takes the value 1 if it is a CC or 0 if it is a VHA (see Table I for the description of the variables after pre-processing). A comparison between the number of absences for these two categories of care providers, illustrated by Fig. 3, reveals a higher absence rate in CCs than in VHAs. In fact, around 6. 1% of consultations in CCs result in absence, compared to only 1. 7% in VHAs. This rate is calculated by dividing the number of absences by the total number of consultations for each category, taking into account the fact that consultations in VHAs are twice as frequent as in CCs.

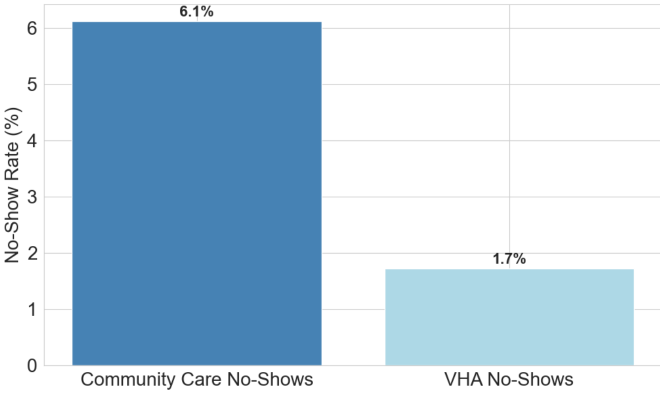


Fig. 3: Comparison of No-show rates in VAHs and CCs

C. Data Preparation

Data preparation addressed critical quality issues in the raw dataset. The missing values were managed by removing the *AdministrativeFlag* column (53% missing) and applying the removal of the list to minor gaps in *zip* (postal code) and *state* (residence). Outliers such as negative service times, resulting from scheduling errors, were discarded, while extreme values reflecting clinic diversity (e.g., primary versus specialized care) were retained.

The target variables were derived as follows: **Service time** was calculated as the difference between the date of consultation completion and the scheduled appointment date. **No-show** was obtained by decomposing the categorical *disp* variable (original categories: *Completed*, *Discontinued*, *Cancelled*) into two binary variables using one-hot encoding. The *no-show* variable was assigned 1 for all absences (including cancellations) and 0 for attended appointments, while *discontinued* flagged appointment interruptions. The *Completed* status was inferred when both variables were 0.

Feature engineering included historical metrics (*prior_no-show_rate*, *prior_service_time_avg*) and clinic-type categorization (*type_care*). Redundant variables such as *dtc* (total consultation time) were removed due to overlap with *dts* (scheduling delay; correlation coefficient: 0.7).

The data contained 3.1% no-shows, indicating a class imbalance. Unbalanced data are those in which the distribution of data classes shows very different proportions [25]. This imbalance can be managed in two ways: sampling methods such as oversampling the minority class, undersampling the majority class, or penalized learning. At this stage of the analysis, we used the synthetic minority oversampling technique (SMOTE). As there is no universal method that works a priori [26], we systematically evaluated various resampling techniques combinations (e.g., RandomUnderSampler and Cluster Centroid) with different sampling strategy configurations. In the end, integrating SMOTE with class weighting at the modeling level yielded the most satisfactory results after hyperparameter optimization. The final transformations and variable definitions are cataloged in Table I, with workflows illustrated in Fig. 4.

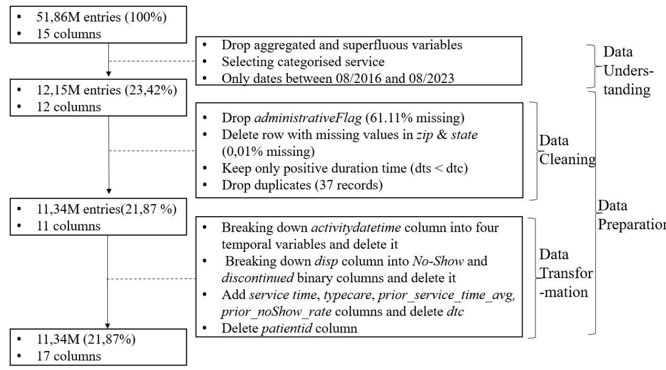


Fig. 4: Various Transformations Undergone by Data During the CRISP-DM data Understanding and Preparation Phases

D. Modelling

After data preparation, the data set is divided into a training set (70%) and a test set (30%). Two types of predictive models are developed: classification models to predict patient no-shows and regression models to estimate service time. Four machine learning algorithms are used for both tasks: Random Forest (RF), XGBoost, AdaBoost, and Artificial Neural Networks (ANNs).

The study employs ensemble methods (bagging and boosting) and ANNs due to their demonstrated effectiveness in predicting these sources of variability, as evidenced in the literature, and their capability to handle both discrete and continuous results. Ensemble models leverage the "Wisdom of the Crowd" principle, combining weak learners to improve accuracy. RF relies on bootstrapping to ensure diversity among learners, while XGBoost and AdaBoost improve performance through gradient-based optimization and sample weighting, respectively. Hyperparameters are optimized via grid search and validated using cross-validation to mitigate overfitting. ANNs, specifically Multi-Layer Perceptron (MLPs), are also utilized for their capacity to model complex nonlinear relationships, an approach increasingly adopted in healthcare research.

The eight resulting ML models are categorized as follows.

- Classification: RF_C, $XGBoost_C$, $AdaBoost_C$ and ANN_C.
- Regression: RF_R, $XGBoost_R$, $AdaBoost_R$ and ANN_R.

E. Evaluation

1) Classification Model: To assess the performance of classification models predicting patient no-shows in an imbalanced dataset (3.11% no-shows), we employed standard evaluation metrics: AUC, Precision, Recall, F1-score, and Accuracy. The formulas for these metrics, as well as those used for regression, are presented in Fig. 5. Among the classification metrics, AUC is the most reliable for imbalanced data, as it evaluates a

TABLE I: VARIABLE DESCRIPTION AFTER DATA PRE-PROCESSING

Name	Datatype	Source	Description
Predictors			
patientsid	Numerical	Raw	Unique identifier for the patient in the health system.
sta3n	Categorical	Raw	VHA facility Identifier, used to differentiate between different sites or installations.
stopcode	Categorical	Raw	Type of primary or speciality care provided during the consultation.
dta	Continuous	Raw	Time between the consultation request and its approval.
dts	Numerical	Raw	Time between the consultation approval and its scheduling.
non_va	Binary	Raw	1 if community care, 0 if VHA care.
zip	Categorical	Raw	Patient's ZIP Code of residence (First three digits).
state	Categorical	Raw	State code of the patient's residence.
discontinued	Binary	Raw	1 if there are disruptions, 0 otherwise.
hour	Numerical	Derived	Hour of the day.
month	Numerical	Derived	Month of the year.
days_of_week	Numerical	Derived	Day of the week.
days_of_month	Numerical	Derived	Day of the month.
prior_service_time_avg	Binary	Derived	Mean of all previous services before the current consultation.
prior_no-show_rate	Continuous	Derived	Percentage of previous not shown-up consultations before the current consultation.
type care	Binary	Derived	1 if primary care, 0 if specialized care.
Dependent Variables			
no-show	Binary	Derived	1 if the patient has not shown up, 0 otherwise.
service time	Continuous	Derived	Time spent by the patient with the doctor during a consultation.

model's ability to distinguish between classes regardless of their distribution. Table II shows the evaluation results of our different models. In this study, the highest AUC was achieved by $XGBoost_C$ (0.86), followed by RF_C (0.83), indicating a strong discriminatory power. Although $AdaBoost_C$ recorded the highest precision (79.02%), this came with the lowest recall (50.46%), which is more critical if we then want to set up overbooking strategy. In contrast, RF_C obtained the best recall (70.79%), demonstrating its ability to identify most non-shows effectively. $XGBoost_C$ delivered a good balance between precision and recall, as reflected in its F1-score (0.69), though it showed signs of overfitting. All models maintained an acceptable generalization capability, despite the application of SMOT sampling solely on training data. Although ANN_C did not outperform in any one metric, it provided a well-rounded and reliable performance, with moderate overfitting and more cautious confidence levels in incorrect predictions.

Classification metrics

$$\text{AUC} = \int_0^1 TPR(FPR)\, d(FPR)$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Notation:
TP: True Positives
FP: False Positives
TN: True Negatives
FN: False Negatives
TPR: True Positive Rate

Regression Metrics

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Notation:
y_i: real values
$\hat{y}_i$: predicted values
$\bar{y}$: mean of real values

Fig. 5: Classification and Regression Metric Formulas

TABLE II: PREDICTION RESULTS OF CLASSIFICATION MODELS ON TRAINING AND TEST DATA

	RF$_C$		**XGBoost$_C$**		**AdaBoost$_C$**		**ANN$_C$**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
AUC	0.97	0.83	1.00	0.86	0.81	0.81	0.85	0.80
Precision (%)	86.71	56.86	99.92	72.77	67.91	67.67	81.29	61.66
Recall (%)	89.95	70.79	99.99	66.91	52.78	52.75	66.58	64.09
F1-score	0.88	0.59	0.99	0.69	0.54	0.54	0.69	0.63
Accuracy (%)	90.73	89.88	99.99	96.78	96.79	96.78	81.96	95.04

2) Regression Model: For service time prediction, regression models were evaluated using R^2, MAE, MSE, and RMSE.

The wide range and high variability of service times (0–365 minutes, with a standard deviation of 25.22) made accurate prediction challenging, particularly in diverse clinical settings. Table III presents the results of our regression models. RF_R showed the strongest explanatory power (R^2 of 20. 88% in the test data), but ANN_R delivered the most accurate predictions overall, with the lowest MAE (12.22 min), MSE (231.41) and RMSE (15.21). These results indicate that ANN_R was the best at limiting large prediction errors and aligning closely with the actual duration of the service.

TABLE III: PREDICTION RESULTS OF REGRESSION MODELS ON TRAINING AND TEST DATA

	RF$_R$		XGBoost$_R$		AdaBoost$_R$		ANN$_R$	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
R^2(%)	34.89	20.88	20.14	18.16	4.04	4.03	4.74	4.74
MAE	13.81	14.91	15.12	15.27	16.22	16.89	12.21	12.22
MSE	414.47	502.72	508.30	609.99	609.77	231.41	231.22	231.41
RMSE	20.36	22.42	22.55	22.80	24.21	24.69	15.21	15.21

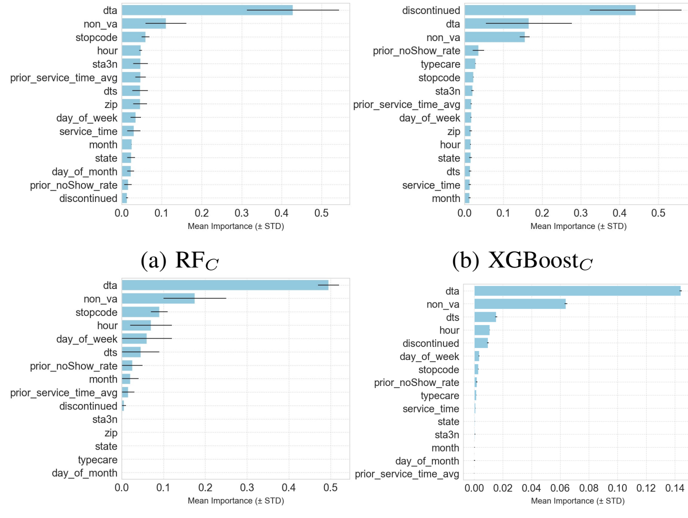
F. Feature Importance Analysis

Figs. 6a–6d and Figs. 7a–7d present the ranked importance of predictors across classification and regression models, respectively. The horizontal axis indicates the average importance score, while the vertical axis lists the predictor names. For ensemble models, importance was calculated based on impurity reduction (entropy for classification, variance for regression), while for neural networks, it was derived from internal weighting mechanisms. To enhance robustness, bootstrapping was applied, and the standard deviation of importance estimates is shown via black bars in the figures. That estimates the variability of feature importance to assess their sensitivity to sampling variation.

In the classification models (RF_C, $XGBoost_C$, $AdaBoost_C$, ANN_C), indirect waiting time variable, *dta* (latency before approval), emerged as the most important features. This finding underscore the persistent impact of waiting times on veterans' no-show behavior, even under the VCP initiative to reduce delays. Other significant predictors included *non_va* (care provider), *stopcode* (type of care) and *hour*. The history of past no-show behavior (*prior_noShow_rate*), obtained in the feature engineering phase, is also important.

In the regression models (RF_R, $XGBoost_R$, $AdaBoost_R$, ANN_R), *stopcode* and *non_va* consistently had the highest importance, highlighting the role of consultation type in determining service time. Additional key features included *zip*, *dts* and *stat3n*. The history of past service times (*prior_service_time_avg*), obtained in the feature engineering phase, proved to be also important for predicting service times.

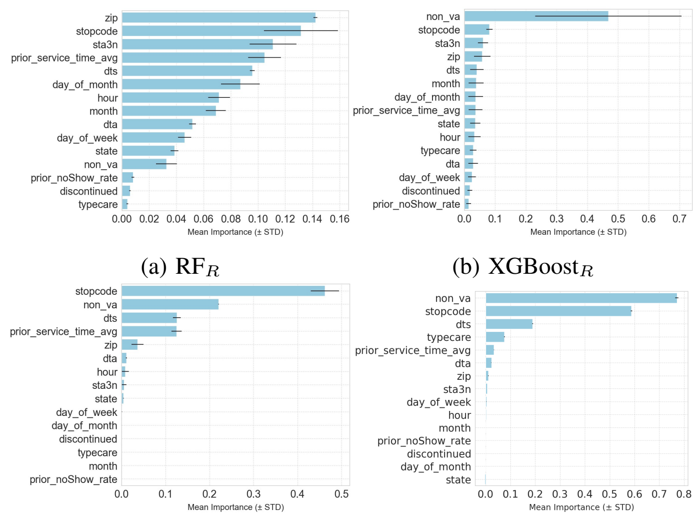
Temporal variables (e.g., *day_of_week*, *month*) consistently ranked low in both tasks, suggesting a limited seasonal effect on no-shows and service time. These findings align with prior studies, which also downplay the role of temporal factors in similar contexts [6], [16], [18], [27], [28].



(a) RF$_C$ (b) XGBoost$_C$

(c) AdaBoost$_C$ (d) ANN$_C$

Fig. 6: Importance of Variables in Classification Models



(a) RF$_R$ (b) XGBoost$_R$

(c) AdaBoost$_R$ (d) ANN$_R$

Fig. 7: Importance of Variables in Regression Models

V. DISCUSSION

This study demonstrates that machine learning provides a practical and effective solution to the persistent challenges of outpatient scheduling challenges that traditional methods often fail to address adequately. On one hand, decision-making without an analysis tool can be time-consuming and demotivating for staff; on the other, solely using optimization methods falls short in effectively addressing uncertainty. By contrast, our ML-based approach enables flexible exploration of clinical data and highlights the most influential features for predicting patient no-shows and service time variability.

A major challenge in applying data mining techniques in healthcare lies in the contextual dependence of the models, which are often constrained by the specificity of the datasets used. To overcome this limitation, we exploited heterogeneous data drawn from multiple clinics, thereby enhancing the robustness and generalisability of our models across diverse outpatient environments.

From a practical point of view, the study supports the implementation of data-driven interventions. These include deploying real-time decision support systems powered by ANN models, using predicted no-show probabilities to trigger preventive actions such as targeted reminders or strategic overbooking and dynamically adjusting appointment slot durations based on expected service times.

Future integration of these predictive tools with overbooking strategy could offer further improvements by optimizing the trade-off between physician idle time and patient waiting time [29]. This combination would harness the uncertainty-handling strengths of machine learning and the efficiency-maximizing capabilities of operational research.

Moreover, clinics can capitalize on the key predictors identified. For example, patient waiting time was identified as a significant predictor of no-show behavior. If patients face extended delays between requesting and receiving an appointment, this waiting time can negatively influence their likelihood of attending. Clinics can proactively address this issue by sending a reminder.

VI. Conclusion

This research offers a novel and scalable ML framework to predict appointment-related uncertainties in outpatient care. Unlike many previous studies focused on single providers, single institutions, or single sources of variability, this work leverages a large, representative dataset to improve the generalizability of predictions. It shows that even in low no-show rate contexts, no-show behavior can be predicted effectively and that incorporating data-driven insights into scheduling processes has significant potential to improve operational efficiency and patient satisfaction.

Future research should focus on deploying this framework in real-world clinical settings and integrating it with optimization approaches. This integration could provide a robust foundation for next-generation appointment management systems.

References

[1] H. Feng, Y. Jia, T. Huang, S. Zhou, and H. Chen, "An adaptive decision support system for outpatient appointment scheduling with heterogeneous service times," *Scientific Reports*, vol. 14, no. 1, p. 27731, 2024.

[2] statista, "Outpatient Care - Worldwide | Statista Market Forecast," 2024.

[3] D. Gupta and B. Denton, "Appointment scheduling in health care: Challenges and opportunities," *IIE Transactions*, vol. 40, no. 9, pp. 800–819, 2008.

[4] A. Kuiper, J. de Mast, and M. Mandjes, "The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice," *Omega*, vol. 98, p. 102122, 2021.

[5] L. H. A. Salazar, W. D. Parreira, A. M. d. R. Fernandes, and V. R. Q. Leithardt, "No-Show in Medical Appointments with Machine Learning Techniques: A Systematic Literature Review," *Information*, vol. 13, no. 11, p. 507, 2022.

[6] D. Golmohammadi, L. Zhao, and D. Dreyfus, "Using machine learning techniques to reduce uncertainty for outpatient appointment scheduling practices in outpatient clinics," *Omega*, vol. 120, p. 102907, 2023.

[7] T. Daghistani, H. AlGhamdi, R. Alshammari, and R. H. AlHazme, "Predictors of outpatients' no-show: big data analytics using apache spark," *Journal of Big Data*, vol. 7, no. 1, p. 108, 2020.

[8] J. D. Welch and N. J. Bailey, "Appointment systems in hospital outpatient departments," *The Lancet*, vol. 259, no. 6718, pp. 1105–1108, 1952.

[9] A. Ahmadi-Javid, Z. Jalali, and K. J. Klassen, "Outpatient appointment systems in healthcare: A review of optimization studies," *European Journal of Operational Research*, vol. 258, no. 1, pp. 3–34, 2017.

[10] T. Cayirli and E. Veral, "Outpatient scheduling in health care: a review of literature," 2003.

[11] T. Cayirli and K. K. Yang, "A Universal Appointment Rule with Patient Classification for Service Times, No-Shows, and Walk-Ins," *Service Science*, vol. 6, no. 4, pp. 274–295, 2014.

[12] L. LaGanga and S. Lawrence, "Appointment Overbooking in Health Care Clinics to Improve Patient Service and Clinic Performance," *Production and Operations Management*, vol. 21, 2012.

[13] A. Ala and F. Chen, "Appointment Scheduling Problem in Complexity Systems of the Healthcare Services: A Comprehensive Review," *Journal of Healthcare Engineering*, vol. 2022, p. 5819813, 2022.

[14] A. Pak, B. Gannon, and A. Staib, "Predicting waiting time to treatment for emergency department patients," *International Journal of Medical Informatics*, vol. 145, p. 104303, 2021.

[15] K. Topuz, T. L. Urban, R. A. Russell, and M. B. Yildirim, "Decision support system for appointment scheduling and overbooking under patient no-show behavior," *Ann Oper Res*, 2024.

[16] G. Fan, Z. Deng, Q. Ye, and B. Wang, "Machine learning-based prediction models for patients no-show in online outpatient appointments," *Data Science and Management*, vol. 2, pp. 45–52, 2021.

[17] R. M. Goffman, S. L. Harris, J. H. May, A. S. Milicevic, R. J. Monte, L. Myaskovsky, K. L. Rodriguez, Y. C. Tjader, and D. L. Vargas, "Modeling Patient No-Show History and Predicting Future Outpatient Appointment Behavior in the Veterans Health Administration," *Military Medicine*, vol. 182, no. 5-6, pp. e1708–e1714, 2017.

[18] S. Srinivas and H. Salah, "Consultation length and no-show prediction for improving appointment scheduling efficiency at a cardiology clinic: A data analytics approach," *International Journal of Medical Informatics*, vol. 145, p. 104290, 2021.

[19] L. Luo, L. Luo, X. Zhang, and X. He, "Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models," *BMC Health Serv Res*, vol. 17, no. 1, p. 469, 2017.

[20] J. Sun, Q. Lin, P. Zhao, Q. Zhang, K. Xu, H. Chen, C. J. Hu, M. Stuntz, H. Li, and Y. Liu, "Reducing waiting time and raising outpatient satisfaction in a Chinese public tertiary general hospital-an interrupted time series study," *BMC Public Health*, vol. 17, no. 1, p. 668, 2017.

[21] W.-J. Chang and Y.-H. Chang, "Design of a Patient-Centered Appointment Scheduling with Artificial Neural Network and Discrete Event Simulation," *Journal of Service Science and Management*, vol. 11, no. 1, pp. 71–82, 2018.

[22] H. Zhu, Y. F. Chen, E. Leung, and X. Liu, "Outpatient appointment scheduling with unpunctual patients," *International Journal of Production Research*, vol. 56, no. 5, 2018.

[23] Y. Feyman, A. Legler, and K. N. Griffith, "Appointment wait time data for primary & specialty care in veterans health administration facilities vs. community medical centers," *Data in Brief*, vol. 36, p. 107134, 2021.

[24] Y. Feyman, D. A. Asfaw, and K. N. Griffith, "Geographic Variation in Appointment Wait Times for US Military Veterans," *JAMA Network Open*, vol. 5, no. 8, p. e2228783, 2022.

[25] F. Sağlam and M. A. Cengiz, "A novel SMOTE-based resampling technique trough noise detection and the boosting procedure," *Expert Systems with Applications*, vol. 200, p. 117023, 2022.

[26] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, p. 27, 2019.

[27] D. Golmohammadi, "A Decision-Making tool based on historical data for service time prediction in outpatient scheduling," *International Journal of Medical Informatics*, vol. 156, p. 104591, 2021.

[28] Y. Han, M. E. Johnson, X. Shan, and M. Khasawneh, "A multi-appointment patient scheduling system with machine learning and optimization," *Decision Analytics Journal*, vol. 10, p. 100392, 2024.

[29] M. Fall, I. Slama, and Y. Ouazene, "Managing no-shows and resource utilisation for outpatient appointment scheduling," *11th IFAC Conference on Manufacturing Modelling, Management and Control (IFAC MIM2025), Trondheim, Norway.*, 2025.