TUFTS UNIVERSITY SCHOOL OF MEDICINE
DEPARTMENT OF PUBLIC HEALTH AND COMMUNITY MEDICINE

PREDICTING PATIENTS' COMMITMENT TO MEDICAL APPOINTMENTS DURING THE PANDEMIC

ROULA KRAYEM

TABLE OF CONTENTS

1. ABSTRACT

This study aimed to identify factors influencing missed medical appointments using machine learning algorithms and logistic regression analysis. Data from the Ipsos Knowledge Panel Survey, collected during the COVID-19 pandemic, was used to train, and evaluate three machine learning models—Decision Tree, Random Forest, and XGBoost. After addressing data leakage, the models identified several important features related to general health, COVID-19-related concerns, and social determinants of health. Logistic regression analysis revealed five statistically significant variables affecting missed appointments, including physical activity, having a medical condition, psychological well-being, the likelihood of getting a COVID-19 vaccine for someone under the participant's care, and being a caregiver. The study provides valuable insights for healthcare providers and policymakers to design targeted interventions to reduce missed appointments and improve healthcare access. Recommendations include leveraging electronic medical records to build and train models, creating proactive plans for patients at risk of missing appointments, updating workflow and resource allocation, and continuously evaluating and adjusting the model. However, limitations such as different healthcare settings, non-representative data, and potential inaccuracy of predictive models should be considered. Future research should focus on validating the identified factors in larger and more diverse samples and examining the long-term effects of these factors on appointment adherence and overall healthcare outcomes.

2. INTRODUCTION

The COVID-19 pandemic has profoundly affected healthcare systems worldwide, leading to an unprecedented strain on resources and significant changes in patient behavior (Filip et al., 2022). One crucial aspect affected by the pandemic is patients' commitment to attending medical appointments. An increase in no-show rates has been observed during this period (Ayele et al., 2022), resulting in a myriad of issues, such as delayed or missed diagnoses, wasted resources, increased healthcare costs, and overbooking and schedule disruptions. Addressing this problem is critical for ensuring timely access to medical care and optimizing resource utilization in healthcare facilities (Huang & Hanauer, 2014).

This study aims to tackle the issue of patient no-shows during the pandemic by developing a predictive machine learning algorithm that can forecast a patient's likelihood of attending their medical appointment. By leveraging advanced machine learning techniques, healthcare providers can identify at-risk patients and implement targeted interventions to improve attendance rates. The objectives of this study are: (1) develop a predictive machine learning algorithm that estimates patients' commitment to their medical appointments during the COVID-19 pandemic, (2) evaluate the performance of the predictive models using real-world data, and (3) analyze the factors that influence patients' commitment to their medical appointments.

To achieve the objectives, we will use a dataset provided by the Viswanath lab at Dana Farber Cancer Institute, which the Ipsos Knowledge Panel collected between July and August 2020. The target population for the study includes non-institutionalized adults aged 18 and older residing in the United States, with over-samples of African Americans, Hispanics, and adults in low-income households. We will employ decision trees, random forest, and XGBoost machine learning algorithms to build our predictive models and assess their performance using various evaluation metrics.

In the following sections, we will review the relevant literature, discuss the data and methods used in the study, present the results of our predictive models, and analyze the factors that influence patients' commitment to medical appointments during the pandemic. By understanding these factors and harnessing the power of machine learning, healthcare providers can develop strategies to mitigate the impact of no-shows and enhance the overall efficiency of healthcare delivery during these challenging times (Car et al., 2018).

3. Literature Review

In this section, we review the existing literature related to patient no-shows, the impact of the COVID-19 pandemic on healthcare systems, and the application of machine learning techniques for predicting patient commitment to medical appointments.

3.1. Patient No-Show and their impact

No-shows, defined as patients who fail to attend their scheduled medical appointments without prior notification, are a significant concern for healthcare systems (Marbouh et al, 2020). No-shows result in various negative consequences for both healthcare institutions and patients.

For healthcare institutions, no-shows lead to wasted resources, as staff and equipment are reserved for appointments that do not take place (Kheirkhah et al., 2015). This wasted time and effort could be better utilized by attending to other patients, improving patient outcomes, and increasing efficiency. In addition, no-shows contribute to increased healthcare costs, as healthcare providers often overbook appointments to compensate for anticipated no-shows (Huang & Hanauer, 2014). Overbooking can lead to longer patient waiting times, diminished patient satisfaction, and increased stress on healthcare staff. Furthermore, missed appointments hinder the ability of healthcare providers to optimize their schedules, resulting in potential revenue loss and a reduction in the overall quality of care provided.

For patients, no-shows can lead to delayed or missed diagnoses, as healthcare providers cannot evaluate and treat medical conditions on time (Kheirkhah et al., 2015). This issue can result in worsening patients' health conditions and the need for more intensive and costly treatments in the future. Moreover, missed appointments can disrupt the continuity of care, which is crucial for managing chronic conditions and ensuring optimal health outcomes.

3.2. Impact of COVID-19 pandemic on healthcare systems

Healthcare systems have experienced unprecedented strain on resources, as the high demand for COVID-19 care led to a depletion of personal protective equipment (PPE) and medical supplies for healthcare staff (Ranney et al., 2020). Healthcare facilities had to adapt their infrastructure, creating additional wards, and reorganizing existing spaces to accommodate the surge in COVID-19 patients. This caused a diversion of resources away from routine care and elective procedures, potentially compromising the quality of care for patients with non-COVID health concerns (Peiffer-Smadja et al., 2020).

The pandemic also led to changes in patient behavior, with many individuals avoiding healthcare facilities due to concerns about contracting the virus or overburdening the healthcare system. The fear of infection led to an increase in patient no-shows and canceled appointments, exacerbating existing problems related to access to care and patient commitment to medical appointments (Ayele et al., 2022). Furthermore, the pandemic prompted an increase in the use of telemedicine as an alternative to in-person consultations, which has both benefits and limitations in terms of patient care (Mubaraki et al, 2021).

3.3. Machine-Learning techniques in healthcare

Machine learning techniques have been increasingly employed in various healthcare applications, such as predicting patient outcomes, disease diagnosis, personalized treatment, and patient commitment to medical appointments (Rajkomar et al., 2019; Kheirkhah et al., 2020). Popular machine learning algorithms used in healthcare include decision trees, random forests, and XGBoost. Decision trees are simple, interpretable models that can capture complex interactions between features, but they are prone to overfitting, especially when dealing with noisy data or data with many features. In contrast, random forests address the overfitting issue by constructing an ensemble of decision trees, introducing randomness to improve generalization performance, but at the cost of reduced interpretability (Breiman et al., 1984).

XGBoost, an advanced gradient boosting algorithm, builds an ensemble of decision trees using a gradient boosting framework and incorporates regularization techniques to prevent overfitting, often achieving state-of-the-art performance in various healthcare tasks. However, like random forests, the interpretability of XGBoost models can be limited due to the complexity of the ensemble (Chen & Guestrin, 2016). Despite these limitations, machine learning techniques have demonstrated significant potential in improving healthcare outcomes, optimizing resource allocation, and enhancing patient care.

4. Methods

4.1. Data collection

The dataset for this study was obtained from Dana Farber Cancer Institute's Viswanath Lab. Ipsos Knowledge Panel collected data through a 63-question survey between July and August 2020. The survey included questions regarding general health, appointment-related topics, COVID-19-related issues, mental health, social life, and social determinants of health. The targeted population consisted of non-institutionalized adults aged 18 and older residing in the United States, with over-samples of African Americans, Hispanics, and adults in low-income households. The raw data contained 1,012 observations and 335 variables.

The outcome variable of interest was Q12, which asked participants if they missed any healthcare services due to the COVID-19 outbreak. The respondents' responses included "Yes, I missed" or "No, I did not miss," "I did not need to receive healthcare services," and those who refused to answer. Since the project focuses on participants with scheduled appointments, the dataset was filtered to include only those who answered "Yes, I missed" or "No, I did not miss." The process resulted in a reduced dataset of 575 observations, with a No/Yes ratio of 66:34 (Figure 1).

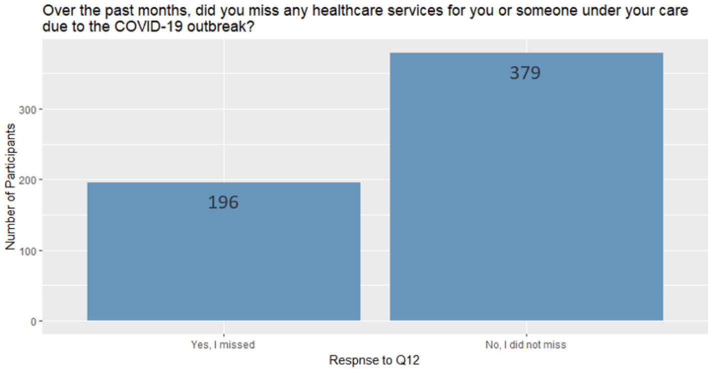


Figure 1: the graph shows the dataset after removing the observations of the participants who refused to answer.or did not have a scheduled appointment.

4.2. Data cleaning

The data cleaning and preprocessing were performed using R and RStudio, with the following packages: haven, labelled, dplyr, ggplot2, VIM, mice, boot, table1, gmodels, and stats. The process involved several steps:

a. Handling missing data: For variables with more than 10% NAs, missing values were replaced with -1 or 0, depending on the values in each variable. For variables with less than 10% NAs, the mice package was used to impute the missing values using a random forest algorithm. Lastly, text variables were transformed into categorical variables, with NAs assigned a value of -1.

b. Transforming Q13: The text variable "What are the services you missed?" was transformed into a categorical variable (Q13cat) with values ranging from 1 to 6, based on the type of missed appointments, with NAs assigned a value of -1. The categories were general visit, specialty visit, scan, surgery/procedure, cancer-related visit, virtual visit, and NA/empty.

c. Merging variables: A score was created by taking the mean of the statements for 18 questions related to well-being, mindfulness, neighborhood social cohesion, support, resilience, loneliness, fear of COVID, COVID misinformation, mask usage in the community, COVID symptoms knowledge, susceptibility, severity, COVID safety guidelines, trust in government and healthcare workers, behavioral compliance, health mavenism, and health mavenism on social media.
d. Checking for duplicate, constant, and quasi-constant (99%) variables: Variables that were constant or nearly constant across all observations were removed, as they do not contribute valuable information for model training. Duplicate variables were also identified and removed to avoid redundancy.
e. Consistency in Yes and No questions: The values for yes and no across the dataset were made consistent by changing some questions to carry the value 0 for No and 1 for Yes.
f. Creating negative and positive datasets: The negative dataset records those who refused to answer or had NAs as -1, while the positive dataset records those who refused to answer or had NAs as 0. The negative dataset retains information about those who rejected to answer, while the positive dataset does not differentiate between those who answered no and those who refused to answer.
g. One-hot encoding: Categorical variables were one-hot encoded to create binary features for each category. One-hot encoding is a common preprocessing step for machine learning algorithms, allowing them to treat categorical variables as separate, independent features. This step prevents the algorithms from misinterpreting the ordinal relationships between the categories, which could lead to incorrect predictions.
h. Splitting the dataset: The dataset was split into training and testing sets using a 70:30 ratio. This step ensures that the models are trained on a large portion of the data while still having a separate, unseen data for evaluation. The split helps to gauge the performance of the models and avoid overfitting, ensuring that the models can generalize well to new, unseen data.

After completing these preprocessing steps, the dataset was ready for use in training and evaluating the machine learning models.

4.3. Machine Learning Models: Details and Implementation

To develop the machine learning models, Python and Visual Studio Code were used as the programming language and integrated development environment, respectively. The primary libraries used in the process included pandas, numpy, matplotlib, seaborn, and sklearn.

Three machine learning algorithms were employed to build the predictive models: decision trees, random forest, and XGBoost. Each algorithm was trained and evaluated using both the negative and positive versions of the preprocessed dataset to find the best model to predict patient commitment to medical appointments during the COVID-19 pandemic.

A function was implemented for the decision tree algorithm to test multiple alpha hyperparameter values (also known as the cost complexity parameter). This parameter controls the trade-off between tree complexity and performance, with the optimal value chosen based on the model's performance on the validation dataset.

Random forest and XGBoost algorithms required a more comprehensive hyperparameter tuning process. Cross-validation and GridSearch were utilized to search for the best combination of hyperparameters for each model. GridSearch systematically tests multiple combinations of hyperparameters, while cross-validation divides the training data into several folds to reduce the risk of overfitting and to obtain a more reliable performance estimate. The optimal hyperparameters were selected based on the models' performance on the validation dataset during cross-validation.

After hyperparameter tuning, the models were trained on the entire training dataset using the optimal hyperparameters. The final performance evaluation was conducted on the test dataset to assess the models' ability to generalize to new, unseen data and to compare the performance of the models trained on the negative and positive datasets.

4.4. Model Evaluation

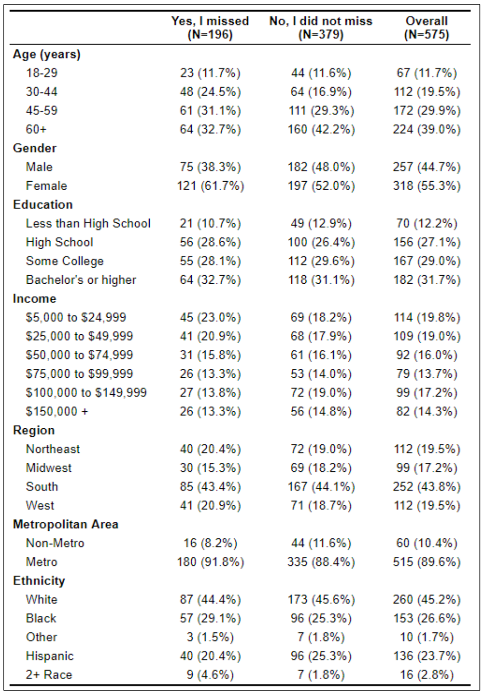
For the model evaluation, the dataset was split into training and testing sets using stratified sampling to ensure that the distribution of the target variable was preserved in both subsets. The evaluation metrics used to assess the performance of the decision tree, random forest, and XGBoost models were precision, recall, F1 score, and accuracy. Additionally, confusion matrices were created to visualize the performance of each model in terms of true positives, true negatives, false positives, and false negatives.

Feature importance was also extracted from each of the three models, providing insights into which variables contributed most to the predictions. A subset of the data containing only the important features, identified by the decision tree, random forest, and XGBoost models, was created to analyze further the relationship between these features and the outcome variable. A logistic regression model was fitted to the subset of important features and the p-values of the interactions between the independent variables and the outcome variable were assessed to determine the strength of the relationships.

5. Results

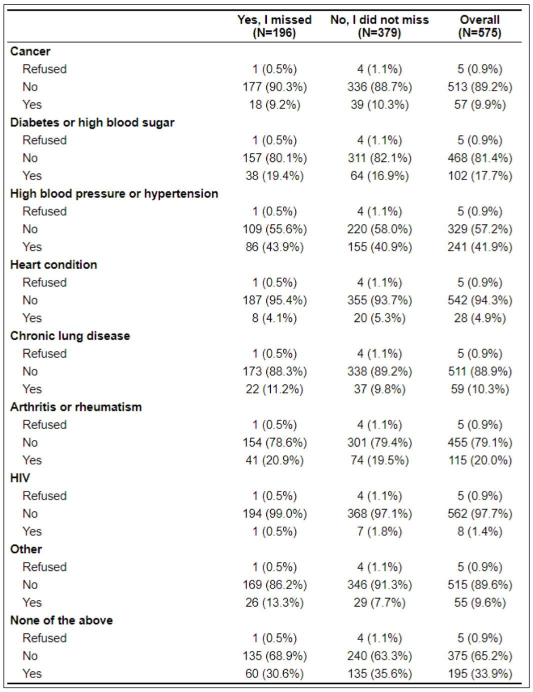
5.1. Participants Demographics

The participant demographics showed that 39% of the participants were 60 years old or older, while 30% were between 45 and 59 years old. A majority of 55% were females. In terms of education, 31% had a bachelor's degree or higher, and 29% had some college education. For income distribution, 19.8% had an annual income of $5,000 to $24,999, 19% had an income of $25,000 to $49,999, and the remaining participants had an annual income of over $50,000. Geographically, most participants (43.8%) were from the southern region of the United States, and 89.6% were from metropolitan areas. Lastly, the racial distribution showed that 42% were white, 26% were black, and 23% were Hispanic (Table 1(a)). Regarding medical conditions, 17% of the participants had diabetes, 45% had hypertension, 10% had chronic lung disease, and 20% had arthritis (Table 1(b)).



	Yes, I missed (N=196)	No, I did not miss (N=379)	Overall (N=575)
Age (years)			
18-29	23 (11.7%)	44 (11.6%)	67 (11.7%)
30-44	48 (24.5%)	64 (16.9%)	112 (19.5%)
45-59	61 (31.1%)	111 (29.3%)	172 (29.9%)
60+	64 (32.7%)	160 (42.2%)	224 (39.0%)
Gender			
Male	75 (38.3%)	182 (48.0%)	257 (44.7%)
Female	121 (61.7%)	197 (52.0%)	318 (55.3%)
Education			
Less than High School	21 (10.7%)	49 (12.9%)	70 (12.2%)
High School	56 (28.6%)	100 (26.4%)	156 (27.1%)
Some College	55 (28.1%)	112 (29.6%)	167 (29.0%)
Bachelor's or higher	64 (32.7%)	118 (31.1%)	182 (31.7%)
Income			
$5,000 to $24,999	45 (23.0%)	69 (18.2%)	114 (19.8%)
$25,000 to $49,999	41 (20.9%)	68 (17.9%)	109 (19.0%)
$50,000 to $74,999	31 (15.8%)	61 (16.1%)	92 (16.0%)
$75,000 to $99,999	26 (13.3%)	53 (14.0%)	79 (13.7%)
$100,000 to $149,999	27 (13.8%)	72 (19.0%)	99 (17.2%)
$150,000 +	26 (13.3%)	56 (14.8%)	82 (14.3%)
Region			
Northeast	40 (20.4%)	72 (19.0%)	112 (19.5%)
Midwest	30 (15.3%)	69 (18.2%)	99 (17.2%)
South	85 (43.4%)	167 (44.1%)	252 (43.8%)
West	41 (20.9%)	71 (18.7%)	112 (19.5%)
Metropolitan Area			
Non-Metro	16 (8.2%)	44 (11.6%)	60 (10.4%)
Metro	180 (91.8%)	335 (88.4%)	515 (89.6%)
Ethnicity			
White	87 (44.4%)	173 (45.6%)	260 (45.2%)
Black	57 (29.1%)	96 (25.3%)	153 (26.6%)
Other	3 (1.5%)	7 (1.8%)	10 (1.7%)
Hispanic	40 (20.4%)	96 (25.3%)	136 (23.7%)
2+ Race	9 (4.6%)	7 (1.8%)	16 (2.8%)

(a)



	Yes, I missed (N=196)	No, I did not miss (N=379)	Overall (N=575)
Cancer			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	177 (90.3%)	336 (88.7%)	513 (89.2%)
Yes	18 (9.2%)	39 (10.3%)	57 (9.9%)
Diabetes or high blood sugar			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	157 (80.1%)	311 (82.1%)	468 (81.4%)
Yes	38 (19.4%)	64 (16.9%)	102 (17.7%)
High blood pressure or hypertension			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	109 (55.6%)	220 (58.0%)	329 (57.2%)
Yes	86 (43.9%)	155 (40.9%)	241 (41.9%)
Heart condition			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	187 (95.4%)	355 (93.7%)	542 (94.3%)
Yes	8 (4.1%)	20 (5.3%)	28 (4.9%)
Chronic lung disease			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	173 (88.3%)	338 (89.2%)	511 (88.9%)
Yes	22 (11.2%)	37 (9.8%)	59 (10.3%)
Arthritis or rheumatism			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	154 (78.6%)	301 (79.4%)	455 (79.1%)
Yes	41 (20.9%)	74 (19.5%)	115 (20.0%)
HIV			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	194 (99.0%)	368 (97.1%)	562 (97.7%)
Yes	1 (0.5%)	7 (1.8%)	8 (1.4%)
Other			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	169 (86.2%)	346 (91.3%)	515 (89.6%)
Yes	26 (13.3%)	29 (7.7%)	55 (9.6%)
None of the above			
Refused	1 (0.5%)	4 (1.1%)	5 (0.9%)
No	135 (68.9%)	240 (63.3%)	375 (65.2%)
Yes	60 (30.6%)	135 (35.6%)	195 (33.9%)

(b)

Table 1: (a) shows the participants demographics, (b) shows participants’ medical conditions.

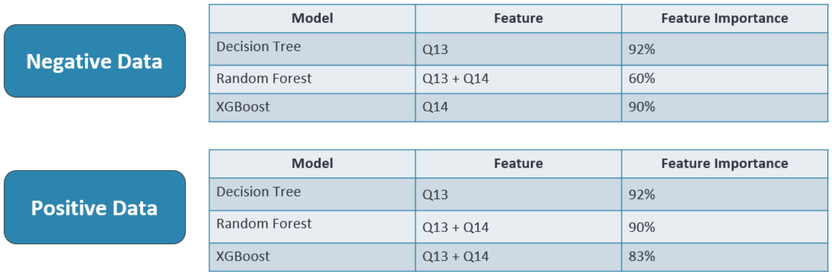
5.2. Initial model performance

In this study, we trained and evaluated three machine learning models—Decision Tree, Random Forest, and XGBoost—on negative and positive datasets to predict whether a patient would miss their medical appointment. The performance of the models was assessed using precision, recall, and accuracy.

Initially, the models demonstrated high accuracy. The Decision Tree model achieved an accuracy of 0.98 for both the negative and positive datasets, with a precision of 0.98 and recall of 0.98 for the negative dataset and a precision of 0.98 and recall of 0.97 for the positive dataset. The Random Forest model had an accuracy of 0.96 for the negative dataset and 0.98 for the positive dataset, with corresponding precision and recall values of 0.95 and 0.97 for the negative data and 0.98 and 0.97 for the positive data. The XGBoost model exhibited an accuracy of 0.96 for the negative dataset and 0.99 for the positive dataset, with precision and recall values of 0.95 and 0.97 for the negative data and 0.99 and 0.98 for the positive data (Appendix 1, 2, and 3).

5.3. Feature Importance and Data leakage

Upon examining the feature importance of the initial models, it became apparent that all six models heavily relied on Q13 and Q14 for their predictions. Further analysis revealed that these features were highly correlated with the outcome variable due to the structure of the survey questions. This correlation introduced data leakage into the models, causing the high performance observed in the initial results (Table 2).



Negative Data

Model	Feature	Feature Importance
Decision Tree	Q13	92%
Random Forest	Q13 + Q14	60%
XGBoost	Q14	90%

Positive Data

Model	Feature	Feature Importance
Decision Tree	Q13	92%
Random Forest	Q13 + Q14	90%
XGBoost	Q13 + Q14	83%

Table 2: the importance of Q13 and Q14 for the initial models.

5.4. Model performance after addressing data leakage

To address the data leakage issue, Q13, and Q14 were removed from both the negative and positive datasets, and the models were retrained. The performance of the decision tree, random forest, and XGBoost models decreased substantially for both datasets. For the negative dataset, the decision tree had a precision of 0.6, a recall of 0.6, and an accuracy of 0.63. The positive dataset showed a precision of 0.53, recall of 0.54, and accuracy of 0.57. The performance of random forest and XGBoost models also decreased for both datasets, emphasizing the importance of carefully examining the data and features used for training machine learning models (Appendix 4, 5, and 6).

5.5. Important features and logistic regression

Despite the decline in performance, the retrained models identified several important features related to general health (Q2, Q3, Q9_9), COVID-19-related concerns (Q26_7, Q32b, Q36e, Q38j, Q40d, Q42, Q44_2, Q45i), and social determinants of health (Q58, Q63, ppeduc, ppmarit, ppage, pphispan). A logistic regression analysis, constructed using these important features, revealed five statistically significant variables, suggesting that these factors may have a genuine impact on missed appointment outcomes:

a) Q2 (Physical Activity) with a coefficient of 0.05422 and a p-value of 0.02.
b) Q9_9 (Having a Medical Condition) with a coefficient of -0.39771 and a p-value of 0.047.
c) Q19 (Psychological Well-being) with a coefficient of -0.33588 and a p-value of 0.0002.

d) Q32b (Getting COVID-19 Vaccine for Someone Under Your Care) with a coefficient of 1.48659 and a p-value of 0.025.
e) Q59 (Being a Caregiver) with a coefficient of 0.91543 and a p-value of 1.08e-05.

5.6. Factors Affecting Missed Appointments

Missing Appointments with Physical Activity: A boxplot was created to compare the number of physical activities per week for participants who missed and did not miss their appointments. The boxplot for participants who missed their appointments was skewed to the right with a higher mean than those who did not miss their appointments. The results suggest that there might be a relationship between physical activity levels and missed appointments (Figure 2 (a)).

Missing Appointments and Psychological Well-being: A boxplot was created to compare the psychological well-being scores of participants who missed and did not miss their appointments. Both boxplots were skewed to the left; however, the mean and median psychological well-being scores were higher for participants who attended their appointments than those who did not attend them. This finding implies that there might be a relationship between psychological well-being and appointment attendance (Figure 2 (b)).

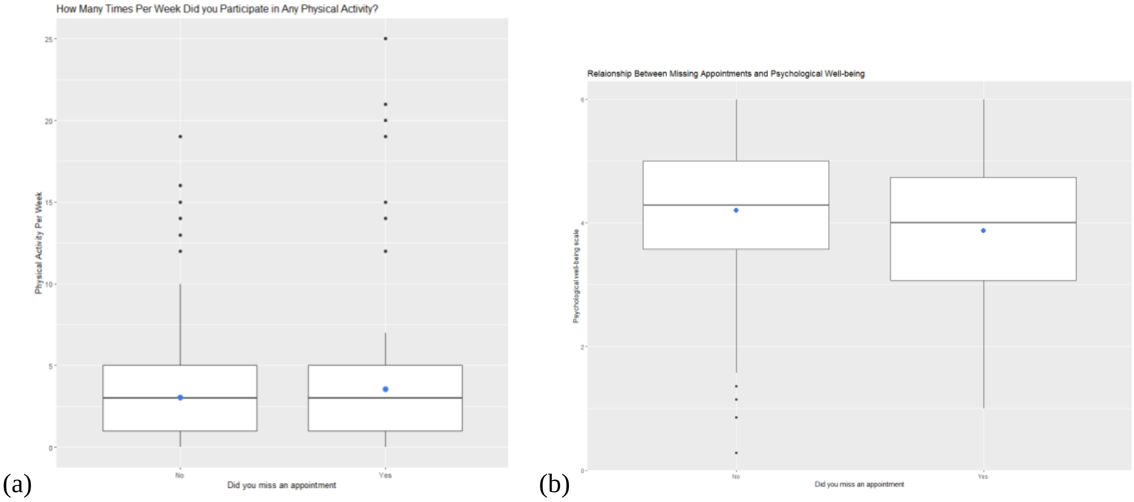
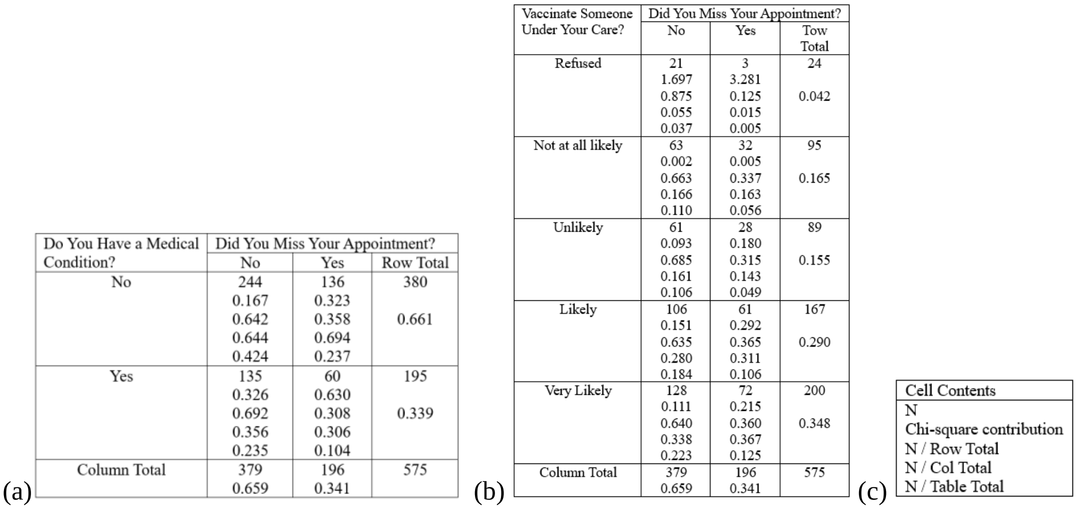


Figure 2: (a) Plotting missing appointments with physical activity. Participants who missed their appointments have a higher mean than those who did not miss their appointments. (b) Plotting missing appointments with psychological well-being. The mean and median psychological well-being scores were higher for participants who attended their appointments than those who did not attend them.

Missing Appointments and Having a Medical Condition: A cross-table was generated to explore the relationship between having a medical condition and missing appointments. The results showed that 34% of the survey participants had medical conditions. 70% of participants with a medical condition did not miss their visits, while 30% missed their appointments. On the other hand, 64% of those who do not have a medical condition did not miss their visit, while 36% missed their visit. This analysis indicates there might be a slight association between these factors. Participants with medical conditions tended to have a higher attendance rate (70%) compared to those without medical conditions (64%) (Table 2 (a)).

Missing Appointments and Getting COVID-19 Vaccine for Someone Under Your Care: A cross-table was generated to examine the relationship between participants who missed and did not miss their appointments and their likelihood of vaccinating someone under their care. The results showed that 63% of the participants were likely to vaccinate those under their care, with 64% of them having missed their appointments. This finding suggests a potential association between vaccine decisions for dependents and appointment attendance (Table 3 (b)).



(a)

Do You Have a Medical Condition?	Did You Miss Your Appointment?		
	No	Yes	Row Total
No	244 0.167 0.642 0.644 0.424	136 0.323 0.358 0.694 0.237	380 0.661
Yes	135 0.326 0.692 0.356 0.235	60 0.630 0.308 0.306 0.104	195 0.339
Column Total	379 0.659	196 0.341	575

(b)

Vaccinate Someone Under Your Care?	Did You Miss Your Appointment?		
	No	Yes	Tow Total
Refused	21 1.697 0.875 0.055 0.037	3 3.281 0.125 0.015 0.005	24 0.042
Not at all likely	63 0.002 0.663 0.166 0.110	32 0.005 0.337 0.163 0.056	95 0.165
Unlikely	61 0.093 0.685 0.161 0.106	28 0.180 0.315 0.143 0.049	89 0.155
Likely	106 0.151 0.635 0.280 0.184	61 0.292 0.365 0.311 0.106	167 0.290
Very Likely	128 0.111 0.640 0.338 0.223	72 0.215 0.360 0.367 0.125	200 0.348
Column Total	379 0.659	196 0.341	575

(c)

Cell Contents
N Chi-square contribution N / Row Total N / Col Total N / Table Total

Table 3: (a) a cross-table for missing appointments and having a medical condition. (b) a cross table for missing appointments and the likelihood of the participants vaccinating someone under their care. (c) Cell components.

Missing Appointments and Being a Caregiver: A bar plot was created to compare the caregiving status of participants who missed and did not miss their appointments. The plot revealed that almost half of the participants who missed their appointments were caregivers for children or older adults. On the other hand, most participants who attended their visits were not caregivers. This analysis indicates that being a caregiver may be an influential factor in missing visits (Figure 3).

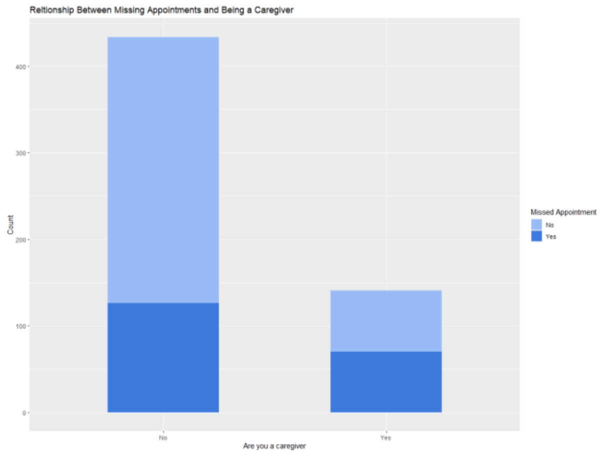


Figure 3: Plotting missing appointments and being a caregiver. Almost half of the participants who missed their appointments were caregivers for children or older adults, and most participants who attended their visits were not caregivers.

6. Discussion, Recommendations, and Limitations

This study aimed to identify factors that influence missed medical appointments using machine learning algorithms and logistic regression. Although the models' performance decreased after addressing data leakage, the study demonstrates the potential of machine learning and logistic regression in identifying factors that affect missed medical appointments. The findings provide valuable insights for healthcare providers and policymakers in designing interventions to reduce missed appointments and improve healthcare access.

One recommendation based on the findings is for healthcare settings to leverage their electronic medical records (EMR) to build and train their models. Rich data available in EMRs can be used to identify patients at higher risk of missing appointments, allowing for more targeted interventions. Healthcare providers can also create a proactive plan for patients classified as not likely to show up for their appointments, such as sending additional reminders or offering alternative appointment options to increase attendance. This can help reduce the number of missed appointments and ensure that patients receive the necessary care.

Another recommendation is to update the settings' workflow and resource allocation based on the predictive models' findings. For example, for clinics that double book providers, the clinic can double book the provider with patients who are less likely to show up to their appointment. This can help optimize resource utilization and ensure that providers have a predictable schedule. Additionally, healthcare providers should continuously evaluate and adjust the model to ensure its accuracy and effectiveness in predicting missed appointments. By adapting to changes in patient populations and healthcare practices, these models can remain a useful tool for reducing missed appointments and improving healthcare access.

Despite these recommendations, it is important to recognize the limitations of the study. Different healthcare settings or departments will require different models, as patient populations and factors influencing appointment attendance may vary. Moreover, models produced using data during the pandemic might not be applicable to non-pandemic times, as healthcare-seeking behaviors and appointment attendance may change in response to external factors. Predictive models may also be biased if the data used to train them is not representative of the population, leading to inaccurate predictions or perpetuation of existing disparities in healthcare access. Lastly, predictive models can never be 100% accurate, and there may be factors that contribute to missed appointments that are not captured in the data. Thus, healthcare providers should use these models as supplementary tools in conjunction with clinical judgment and patient context.

Future research should consider a larger and more diverse sample to further validate the identified factors and their impact on appointment attendance. Additionally, longitudinal studies can help in understanding the long-term effects of these factors on appointment adherence and overall healthcare outcomes.

REFERENCES

Ayele, T. A., Alamneh, T. S., Shibru, H., Sisay, M. M., Yilma, T. M., Melak, M. F., Bisetegn, T. A., Belachew, T., Haile, M., Zeru, T., Asres, M. S., &amp; Shitu, K. (2022, October 4). Effect of covid-19 pandemic on missed medical appointment among adults with chronic disease conditions in northwest Ethiopia. PLOS ONE. Retrieved April 23, 2023, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0274190

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and Regression Trees.

Car, J., Sheikh, A., Wicks, P. et al. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. BMC Med 17, 143 (2019). https://doi.org/10.1186/s12916-019-1382-x

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794. https://dl.acm.org/doi/pdf/10.1145/2939672.2939785

Filip, R., Gheorghita Puscaselu, R., Anchidin-Norocel, L., Dimian, M., & Savage, W. K. (2022). Global Challenges to Public Health Care Systems during the COVID-19 Pandemic: A Review of Pandemic Measures and Problems. Journal of personalized medicine, 12(8), 1295. https://doi.org/10.3390/jpm12081295

Huang, Y., & Hanauer, D. A. (2014). Patient no-show predictive model development using multiple data sources for an effective overbooking approach. Applied clinical informatics, 5(3), 836–860. https://doi.org/10.4338/ACI-2014-04-RA-0026

Kheirkhah, P., Feng, Q., Travis, L.M. et al. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 16, 13 (2015). https://doi.org/10.1186/s12913-015-1243-z

Marbouh, D., Khaleel, I., Al Shanqiti, K., Al Tamimi, M., Simsekler, M. C. E., Ellahham, S., Alibazoglu, D., & Alibazoglu, H. (2020). Evaluating the Impact of Patient No-Shows on Service Quality. Risk management and healthcare policy, 13, 509–517. https://doi.org/10.2147/RMHP.S232114

Mubaraki, A. A., Alrabie, A. D., Sibyani, A. K., Aljuaid, R. S., Bajaber, A. S., & Mubaraki, M. A. (2021). Advantages and disadvantages of telemedicine during the COVID-19 pandemic era among physicians in Taif, Saudi Arabia. Saudi medical journal, 42(1), 110–115. https://doi.org/10.15537/smj.2021.1.25610

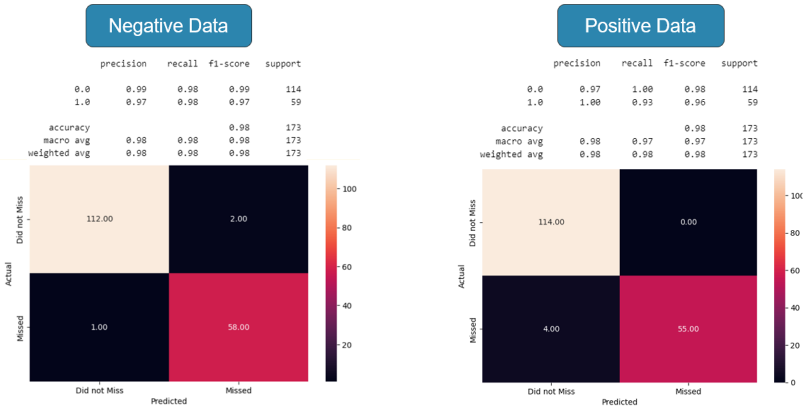
Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. The New England journal of medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Ranney, M. L., Griffeth, V., &amp; Jha, A. K. (2020, April 30). Critical supply shortages - New England Journal of Medicine. Retrieved April 23, 2023, from https://www.nejm.org/doi/full/10.1056/NEJMp2006141

Peiffer-Smadja, N., Lucet, J. C., Bendjelloul, G., Bouadma, L., Gerard, S., Choquet, C., Jacques, S., Khalil, A., Maisani, P., Casalino, E., Descamps, D., Timsit, J. F., Yazdanpanah, Y., & Lescure, F. X. (2020). Challenges and issues about organizing a hospital to respond to the COVID-19 outbreak: experience from a French reference centre. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 26(6), 669–672. https://doi.org/10.1016/j.cmi.2020.04.002

APPENDIX

Appendix 1: Decision Tree



Negative Data

```
              precision    recall  f1-score   support

         0.0       0.99      0.98      0.99       114
         1.0       0.97      0.98      0.97        59

    accuracy                           0.98       173
   macro avg       0.98      0.98      0.98       173
weighted avg       0.98      0.98      0.98       173
```

Positive Data

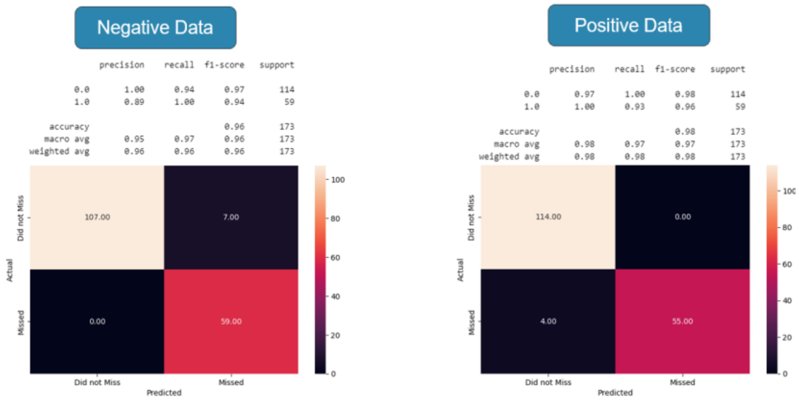
```
              precision    recall  f1-score   support

         0.0       0.97      1.00      0.98       114
         1.0       1.00      0.93      0.96        59

    accuracy                           0.98       173
   macro avg       0.98      0.97      0.97       173
weighted avg       0.98      0.98      0.98       173
```

The performance of Decision Tree using negative and positive datasets before resolving data leakage.

Appendix 2: Random Forest



Negative Data

```
              precision    recall  f1-score   support

         0.0       1.00      0.94      0.97       114
         1.0       0.89      1.00      0.94        59

    accuracy                           0.96       173
   macro avg       0.95      0.97      0.96       173
weighted avg       0.96      0.96      0.96       173
```

Positive Data

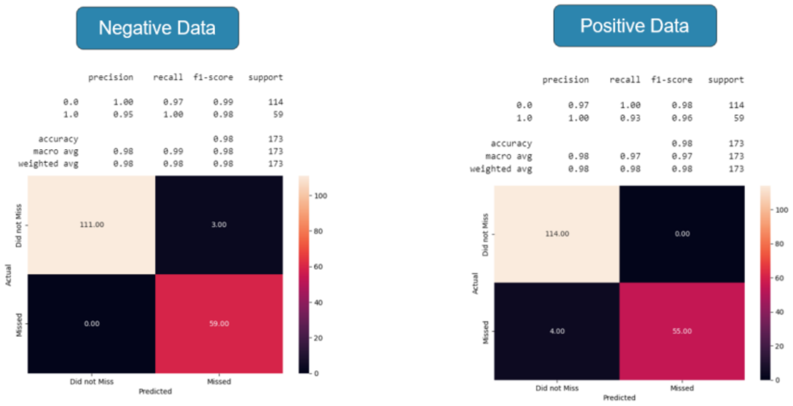
```
              precision    recall  f1-score   support

         0.0       0.97      1.00      0.98       114
         1.0       1.00      0.93      0.96        59

    accuracy                           0.98       173
   macro avg       0.98      0.97      0.97       173
weighted avg       0.98      0.98      0.98       173
```

The performance of Random Forest using negative and positive datasets before resolving data leakage.

Appendix 3: XGBoost



Negative Data

```
              precision    recall  f1-score   support

         0.0       1.00      0.97      0.99       114
         1.0       0.95      1.00      0.98        59

    accuracy                           0.98       173
   macro avg       0.98      0.99      0.98       173
weighted avg       0.98      0.98      0.98       173
```

Positive Data

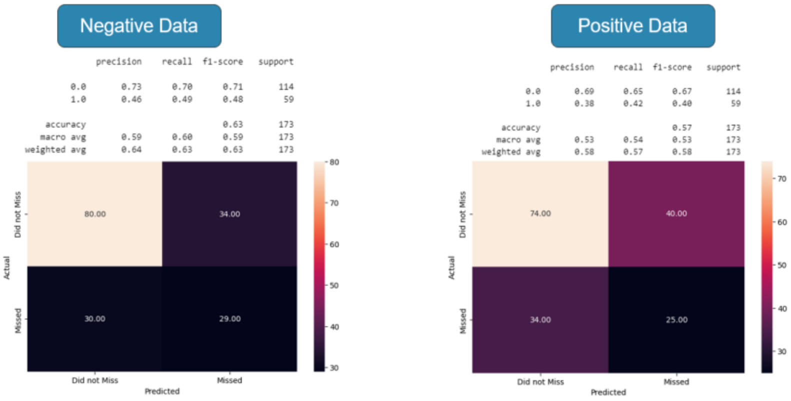
```
              precision    recall  f1-score   support

         0.0       0.97      1.00      0.98       114
         1.0       1.00      0.93      0.96        59

    accuracy                           0.98       173
   macro avg       0.98      0.97      0.97       173
weighted avg       0.98      0.98      0.98       173
```

The performance of XGBoost using negative and positive datasets before resolving data leakage.

Appendix 4: Decision Tree



Negative Data

```
              precision    recall  f1-score   support

         0.0       0.73      0.70      0.71       114
         1.0       0.46      0.49      0.48        59

    accuracy                           0.63       173
   macro avg       0.59      0.60      0.59       173
weighted avg       0.64      0.63      0.63       173
```

Positive Data

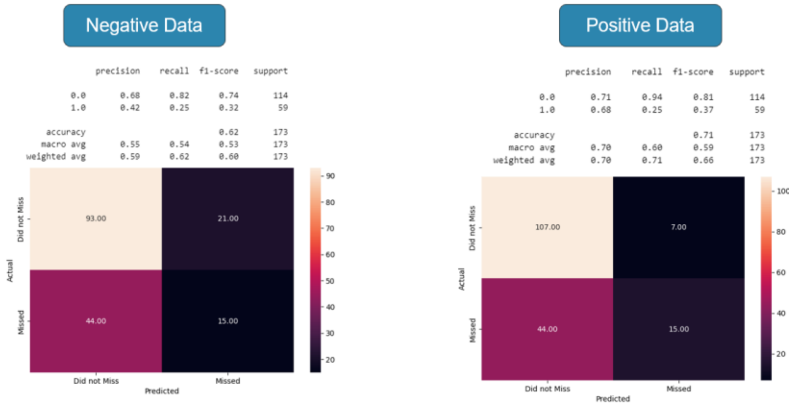
```
              precision    recall  f1-score   support

         0.0       0.69      0.65      0.67       114
         1.0       0.38      0.42      0.40        59

    accuracy                           0.57       173
   macro avg       0.53      0.54      0.53       173
weighted avg       0.58      0.57      0.58       173
```

The performance of Decision Tree using negative and positive datasets after resolving data leakage.

Appendix 5: Random Forest



Negative Data

```
              precision    recall  f1-score   support

         0.0       0.68      0.82      0.74       114
         1.0       0.42      0.25      0.32        59

    accuracy                           0.62       173
   macro avg       0.55      0.54      0.53       173
weighted avg       0.59      0.62      0.60       173
```

Positive Data

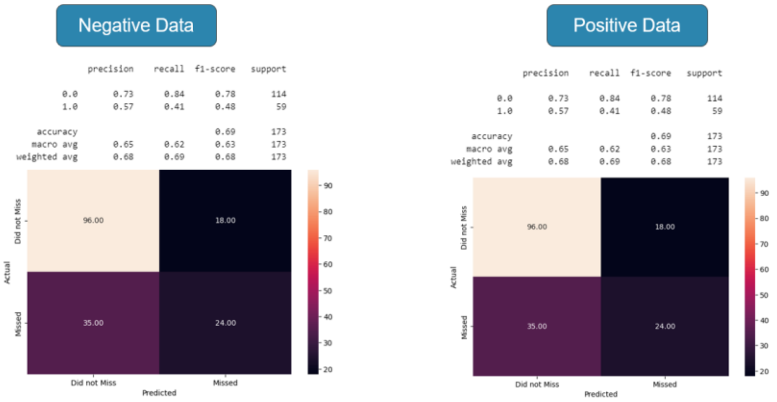
```
              precision    recall  f1-score   support

         0.0       0.71      0.94      0.81       114
         1.0       0.68      0.25      0.37        59

    accuracy                           0.71       173
   macro avg       0.70      0.60      0.59       173
weighted avg       0.70      0.71      0.66       173
```

The performance of Random Forest using negative and positive datasets after resolving data leakage.

Appendix 6: XGBoost



Negative Data

```
              precision    recall  f1-score   support

         0.0       0.73      0.84      0.78       114
         1.0       0.57      0.41      0.48        59

    accuracy                           0.69       173
   macro avg       0.65      0.62      0.63       173
weighted avg       0.68      0.69      0.68       173
```

Positive Data

```
              precision    recall  f1-score   support

         0.0       0.73      0.84      0.78       114
         1.0       0.57      0.41      0.48        59

    accuracy                           0.69       173
   macro avg       0.65      0.62      0.63       173
weighted avg       0.68      0.69      0.68       173
```

The performance of XGBoost using negative and positive datasets after resolving data leakage.