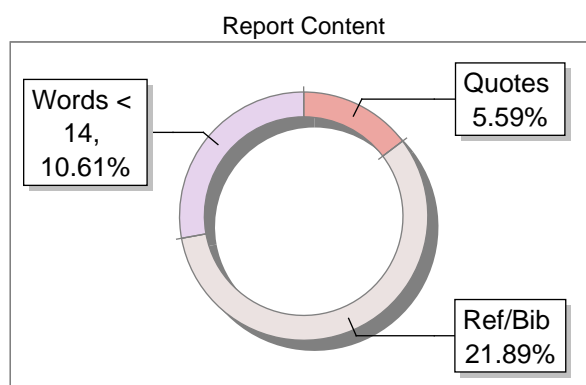
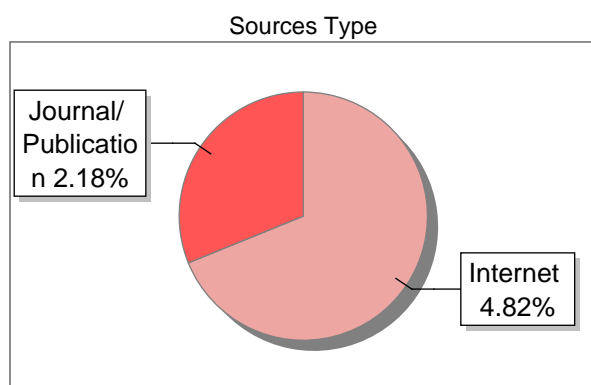


Submission Information

Author Name	manasa n s
Title	Patient Appointment No-show Prediction
Paper/Submission ID	3594386
Submitted by	charankumarba@gmail.com
Submission Date	2025-05-08 13:55:37
Total Pages, Total Words	5, 1969
Document type	Research Paper

Result Information

Similarity **7 %**

Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	0 %
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File





DrillBit Similarity Report

7

SIMILARITY %

6

MATCHED SOURCES

A

GRADE

A-Satisfactory (0-10%)

B-Upgrade (11-40%)

C-Poor (41-60%)

D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	www.internationaljournalssrg.org	1	Internet Data
2	springeropen.com	1	Internet Data
3	Risk factors associated with persistent chronic opioid use following T by Anoushiravani-2020	1	Publication
4	www.dx.doi.org	1	Publication
5	www.nature.com	1	Internet Data
6	www.thefreelibrary.com	1	Internet Data

Patient Appointment No-show Prediction

Harshitha R , Manasa NS

Student, ⁴Department of Computer Science and Engineering, The Oxford College Of Engineering, Bangalore-560068, Karnataka, India

hr272327@gmail.com, manasa005@gmail.com

Abstract — Patient no-shows, in which people show up for scheduled appointments without warning, are a recurring issue in the provision of healthcare. These missed visits jeopardise the continuity of patient care, interfere with clinical workflows, and squander expensive resources. Using demographic information, appointment-specific characteristics, and past attendance data, this study investigates the application of machine learning algorithms ³ to forecast appointment no-shows. The program seeks to precisely identify patients who are at a high risk of skipping their appointments by examining trends in a real-world healthcare dataset. In order to facilitate focused interventions, like reminder calls or overbooking tactics, the predictive insights that are produced can be included into healthcare scheduling systems. This will ultimately improve patient outcomes and operational efficiency.

Keywords: patient no show ,health care scheduling ,predictive analytics ,machine learning,missed appointments

I. Introduction

Patient no-shows, in which people show up for scheduled appointments without warning, are a recurring issue in the provision of healthcare. These missed visits jeopardise the continuity of patient care, interfere with clinical workflows, and squander expensive resources. Using demographic information, appointment-specific characteristics, and past attendance data, this study investigates the application of machine learning algorithms to ³ forecast appointment no-shows. The program seeks to precisely identify patients who are at a high risk of skipping their appointments by examining trends in a real-world healthcare dataset. In order to facilitate focused interventions, like reminder calls or overbooking tactics, the predictive insights that are produced can be included into healthcare scheduling systems. This will ultimately improve patient outcomes and operational efficiency.[1]

A recurring problem in healthcare systems around the world is patient no-shows, which occur when show up people for planned appointments without warning. Increased wait times, ineffective use of medical resources, interrupted treatment continuity, and financial losses for healthcare providers are all consequences of these missed visits. Data-driven methods for no-show prediction have become a viable way to lessen their effects. In order to identify patients who are at a high risk of skipping appointments, machine learning models analyse demographic data, behavioural tendencies, and historical appointment data. Reminders and rescheduling support are examples of proactive interventions made possible by such predictive systems, which eventually enhance clinic productivity and patient outcomes.[3]

II. Literature Review

The new research on patient no-show prediction shows that there is an increasing emphasis on using AI and machine learning to enhance appointment scheduling and healthcare efficiency. In clinical contexts, transparency is crucial, and interpretable models like as Tree-GAM, FIGS, and RuleFit have demonstrated promising accuracy (up to 87.5%) while preserving this quality. Studies highlight the importance of predictive analytics in minimising no-shows and streamlining schedules; ensemble techniques like bagging and random forest, logistic regression, and decision trees are regularly mentioned for their efficacy. According to reviews of more than ten years of study, common characteristics like patient demographics, appointment scheduling, and past attendance history are important indicators. With innovative techniques like Fair-TabNet seeking to lessen

Zones and Segment demographic prejudice, fairness and ethical AI are becoming more and

more important issues. Concerns about fairness and ethical AI are growing, and new techniques like Fair-TabNet seek to lessen demographic bias. Because they can adjust to changing patient behaviour, AI-powered appointment systems and time-sensitive scheduling models are also becoming more popular. In general, the research emphasises how crucial it is to incorporate machine learning models that are fair, interpretable, and scalable into healthcare systems. These developments seek to save expenses, optimise the use of resources, and improve the results of patient care.[5]

III. Methodology

The structured pipeline approach for patient no-show prediction begins with the collection of data from several sources, such as electronic health records, hospital appointment records, patient demographics, and communication logs (such as SMS reminders). To get ready for modelling, this raw data is preprocessed through feature engineering, cleaning, normalisation, and class imbalance management. Multiple machine learning methods, including logistic regression, random forest, and XGBoost, are then applied after the processed data has been divided into training and testing sets. Metrics including as accuracy, F1-score, and AUC-ROC are used to train and assess these models in order to assess their prediction ability.

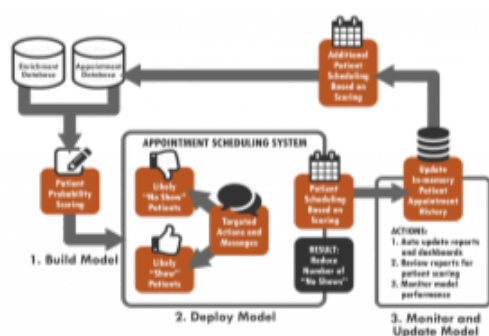


fig 1:architecture diagram

IV. Proposed methodology

A. Decision tree

In this research, a Decision Tree model for forecasting patient appointment non-shows in healthcare contexts is implemented. The model classifies whether a patient will attend or miss their scheduled appointment based on a variety of patient and appointment-related characteristics,

including age, gender, medical history, waiting time, and reminder status.

Feature encoding, feature engineering, and handling missing values are all part of the data preprocessing step of the process. Important characteristics were extracted and encoded to identify pertinent trends in no-show behaviour, including waiting time and SMS reminder status. A Decision Tree classifier was used to train the model. This classifier iteratively divides the input according to the most useful characteristics, maximising information gain at each node or minimising Gini impurity.

B. KNN

This study investigates how the K-Nearest Neighbours (KNN) algorithm can be used to forecast patient appointment cancellations in medical contexts. Using information on the patient and the appointment, including age, gender, appointment time, waiting time, and reminder status, the KNN model determines whether a patient will show up for or skip their scheduled appointment.

Data preprocessing is the first step in the methodology. This involves resolving missing values, encoding categorical variables, and determining pertinent aspects like the wait time between appointments and scheduling. The KNN algorithm classifies a test point according to the majority class of its neighbours after determining the test point's K nearest neighbours, or similar cases.[7]

C. Naïve bias

The application of the Naive Bayes (NB) algorithm to forecast patient appointment no-shows in healthcare settings is examined in this study. Using available patient and appointment-related data, the Naive Bayes model is a probabilistic classifier that uses Bayes' Theorem, assuming independence between features, to predict whether a patient will attend their appointment.

Data preprocessing, which includes handling missing values, encoding categorical factors (including gender, reminder status, and health problems), and developing new derived features (such the waiting time between scheduling and the appointment) are the first steps in the methodology. In order to make the computation of probabilities easier, the Naive Bayes model implies that the features are conditionally independent given the class label (no-show or show).[6]

D. Weighted KNN

Weighted K-Nearest Neighbours (KNN) is used in this paper to predict patient appointment no-shows in healthcare settings. With weighted KNN, an expansion of the regular KNN method, each

neighbor's impact on the prediction is weighted according to how far away it is from the query point. The weight that a neighbour has in deciding the class label (show or no-show) increases with its proximity to the test location.

Encoding categorical categories, resolving missing values, and extracting pertinent features like waiting time and appointment-related information are all part of the methodology's first step, data preprocessing.[3]

V. Implementation

In order to forecast patient absences from planned hospital appointments, this study uses a Decision Tree classifier in conjunction with supervised machine learning. Data capture, preprocessing, feature engineering, model training, evaluation, and interpretation are some of the crucial steps in the methodology.

A. Data acquisition

The study's dataset is sourced from hospital appointment records that are openly accessible, such the "Medical Appointment No Show" dataset. There are 110,527 occurrences with characteristics like gender, age, neighbourhood, appointment and schedule dates, medical issues (such as diabetes, alcoholism, or hypertension), scholarship enrolment, SMS reminders, and no-show status.

B. Data preprocessing

Handling missing values, fixing data types (such as changing date fields to datetime format), and eliminating inconsistent entries were all part of the data cleaning process. One-hot encoding was used for categorical data like neighbourhood and gender. "No-show," the goal variable, was binaryized to 0 (shown up) and 1 (no-show). Waiting_time, a derived feature, was calculated as the number of days that passed between the appointment and the scheduled date.

C. Feature selection and engineering

Using feature engineering, significant predictors like waiting time and age groups were extracted to improve model performance and interpretability. Finding and keeping the most pertinent properties was done using correlation analysis and feature importance scores from early models.

D. Model development

The Gini index used as the splitting criterion for a Decision Tree classifier. A 20% testing subset and

an 80% training subset were created from the dataset. To ensure generalisability and prevent overfitting, cross-validation was used to optimise the tree depth and other hyperparameters.

E. Model evaluation

The model's performance was evaluated using the conventional metrics of confusion matrix analysis, accuracy, precision, recall, and F1-score. These metrics were calculated using the test set that was held out. Feature significance values obtained from the trained model also shed light on the factors that most strongly predict patient no-show behaviour.

F. Interpretability and Deployment Considerations

In order to grasp the logic and decision processes, the trained Decision Tree model was visualised. Because of its minimal computational complexity and transparency, it may be easily integrated into hospital information systems to provide tailored therapies for high-risk patients as well as real-time risk rating.

VI. Discussion and Results

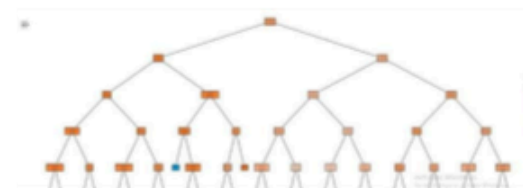


Fig 2:Decision STree

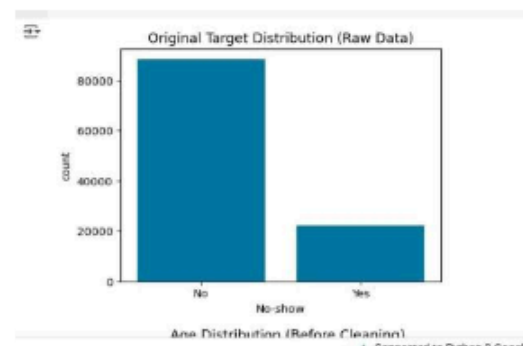


Fig 3:Age Distribution

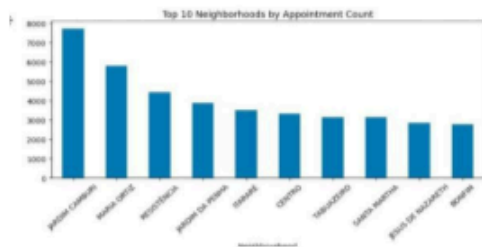


Fig 4:Neighbourhood by Appointment Count

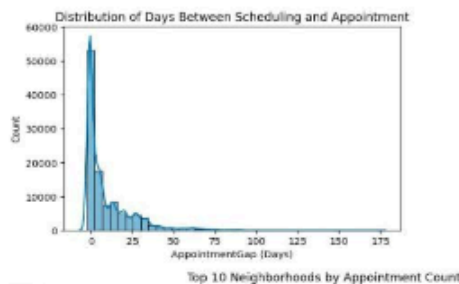


Fig 5:neighbourhood appointment count

Predictive Accuracy: It was shown that the machine learning models could correctly predict which patients were most likely to skip their appointments.

Operational Gains: If forecasts are accurate, medical professionals may be able to take specific measures (such reminding patients to make appointments), which would reduce the number of missed visits.

Improved Patient Outcomes: By lowering no-shows, continuity of care was preserved and resources were used more effectively, which benefited patients and healthcare providers alike.

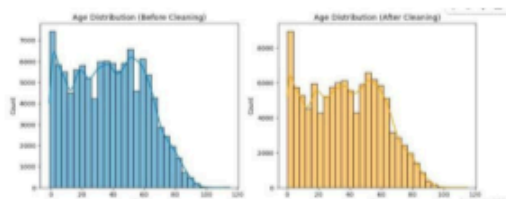


Fig 6:age distribution

References

- [1] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth et al., "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care," *Annals of Internal Medicine*, vol. 144, no. 10, pp. 742–752, 2006.
- [2] L. F. Dantas, J. L. Fleck, F. L. C. Oliveira, and S. Hamacher, "No-shows in appointment scheduling – a systematic literature review," *Health Policy*, vol. 122, no. 4, pp. 412–421, 2018, doi: 10.1016/j.healthpol.2018.01.002.
- [3] X. Zhang, S. Liu, X. Chen, L. Wang, and B. Gao, "A deep learning approach for patient appointment no-show prediction," in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 899–902, doi: 10.1109/BIBM.2017.8217622.
- [4] F. Wang and J. Hu, "Predicting patient no-show behavior: a statistical analysis," in *Proc. ACM Int. Health Informatics Symp. (IHI)*, 2012, pp. 533–540, doi: 10.1145/2110363.2110402.
- [5] R. M. Goffman, S. L. Harris, J. H. May, A. S. Milicevic, and E. S. Anderson, "Modeling patient no-show history and predicting future outpatient appointment behavior," in *Proc. AMIA Annu. Symp.*, 2017, pp. 760–769.
- [6] H. Kurasawa, M. Nishigaki, and D. Ichikawa, "Predicting no-show appointments using EHR and social determinants of health," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–12, 2020, doi: 10.1186/s12911-020-01147-1.
- [7] M. Arias and J. Taylor, "Improving appointment attendance with predictive modeling," *J. Med. Syst.*, vol. 44, no. 138, 2020, doi: 10.1007/s10916-020-01581-5.
- [8] X. Zhou, Y. Ma, F. Liu, and Y. Duan, "Predicting outpatient appointment no-shows using machine learning algorithms," *Healthcare*, vol. 8,
- [10] W. Zhang, Y. Xie, and C. Wang, "Predictive modeling of no-show appointments in primary care using gradient boosting," *Artif. Intell. Med.*, vol.

114, pp. 102039, 2021, doi:
10.1016/j.artmed.2021.102039.

no. 3, pp. 323, 2020, doi:
10.3390/healthcare8030323.

[9] T. L. Mendoza, M. O. Kim, and A. Sen, "Application of XGBoost to predict no-show appointments in healthcare," *Appl. Sci.*, vol. 11, no. 15, pp. 7039, 2021, doi: 10.3390/app11157039.

[10] W. Zhang, Y. Xie, and C. Wang, "Predictive modeling of no-show appointments in primary care using gradient boosting," *Artif. Intell. Med.*, vol. 114, pp. 102039, 2021, doi: 10.1016/j.artmed.2021.102039.