AI-Based Patient No-Show Prediction using Machine Learning

Author: Md. Khalid Hasan
Department of Information and Communication Technology (ICT)
Bangladesh University of Professionals (BUP)
March 2026

Abstract: Patient no-shows are a significant challenge in healthcare systems, leading to inefficient resource utilization and increased operational costs. This study applies Machine Learning techniques to predict whether a patient will miss an appointment using the Medical Appointment No-Shows (Brazil) dataset. After data preprocessing and feature selection, a Logistic Regression model was developed, achieving approximately 80% accuracy. The results demonstrate that ML-based prediction systems can support healthcare providers in taking preventive actions, improving scheduling efficiency, and reducing no-show rates.

Keywords: Machine Learning (ML), Patient No-Shows, Healthcare Analytics, Logistic Regression, Predictive Modeling

1. INTRODUCTION

Healthcare systems often suffer from patient no-shows, where patients fail to attend scheduled appointments. This leads to inefficient use of medical resources and negatively impacts both healthcare providers and patients. Predicting such behavior in advance can help hospitals take preventive actions such as sending reminders or rescheduling appointments. With the advancement of Artificial Intelligence, Machine Learning provides powerful tools to analyze historical data and make accurate predictions. This research aims to explore whether Machine Learning can effectively predict patient no-shows using real healthcare data. This study focuses on developing a predictive model to identify patients at risk of missing appointments using Machine Learning techniques.

2. PROBLEM STATEMENT

Patient no-shows are a common issue in modern healthcare systems, leading to inefficient use of medical resources and increased operational costs. Missed appointments also result in longer waiting times and reduced service quality for other patients.

The key challenges associated with patient no-shows include:

a. Wasted doctor time
b. Increased operational costs
c. Longer waiting times for other patients
d. Reduced healthcare efficiency

Despite the availability of historical data, many healthcare systems lack effective methods to predict patient no-shows in advance. This limits the ability to take preventive actions such as sending reminders or optimizing appointment schedules. The main objective of this study is to develop a Machine Learning-based model to predict whether a patient will miss an appointment and improve overall healthcare efficiency.

3. Dataset Description

The dataset used in this research is the **Medical Appointment No-Shows (Brazil)** dataset. It contains information about patient appointments, including:

a. Age
b. Gender
c. Scholarship (financial support)
d. Hypertension
e. Diabetes
f. SMS reminders
g. Appointment status (No-show or Show)

This dataset enables the analysis of patient behavior and supports the development of predictive models. It provides valuable insights for identifying patterns associated with missed appointments.

4. Methodology

This section outlines the steps used to prepare data and build the prediction model.

4.1 Data Preprocessing: Categorical variables were converted into numerical format, and relevant features were selected for model training. The dataset was then split into training (80%) and testing (20%) sets. Finally, data normalization was applied to improve model performance.

4.2 Model Used: A Logistic Regression model was used for prediction due to its simplicity, efficiency, and effectiveness in classification tasks. It is computationally efficient, performs well for binary classification problems, and provides easily interpretable results, making it suitable for this study.

4.3 Training: The model was trained using historical appointment data to learn patterns associated with no-shows.

5. Results

The proposed model was trained and evaluated using the prepared dataset. The experimental results show that the model achieved an accuracy of approximately **80%**. This level of accuracy indicates that the model is capable of effectively predicting patient no-shows in most cases. The performance of the Logistic Regression model demonstrates its suitability for binary classification problems of this nature. Furthermore, the results suggest that meaningful patterns exist within the dataset, which can be leveraged to improve decision-making in healthcare systems. Overall, the findings highlight the potential of Machine Learning techniques in addressing appointment management challenges.

6. Discussion

This section analyzes the performance of the proposed model and highlights key findings derived from the results. The outcomes indicate that Machine Learning techniques can effectively identify patterns in patient behavior related to appointment attendance. The analysis reveals several important insights. Patients who did not receive SMS reminders were found to have a higher likelihood of missing appointments. Additionally, certain health conditions and demographic factors were observed to influence patient attendance. Despite the promising performance, there is scope for further improvement. The model can be enhanced by employing more advanced algorithms such as Random Forest and XGBoost, incorporating additional relevant features, and applying hyperparameter tuning techniques to optimize performance. These findings emphasize the importance of data-driven decision-making in healthcare systems. By leveraging predictive models, hospitals can proactively manage appointments and improve overall operational efficiency.

7. Conclusion

This research demonstrates that Machine Learning can be a useful tool in predicting patient no-shows. With approximately 80% accuracy, the model shows promising potential for real-world applications such as automated appointment management, smart reminder systems, and resource optimization in hospitals. This work represents an initial step toward integrating AI into healthcare systems. Future work can focus on improving model performance using advanced algorithms and larger datasets. Additionally, integrating this system into real-time hospital management platforms can further enhance healthcare efficiency and patient satisfaction.

The implementation of this work is available on GitHub for reproducibility and further development: https://github.com/khalid67-ux/patient-no-show-ml

8. Future Work

Future improvements may include the use of advanced machine learning models to enhance prediction accuracy, as well as the development of a real-time prediction system for practical deployment. Integration with hospital management software can further support automated decision-making and efficient appointment scheduling.

Additionally, the development of a web-based application can make the system more accessible to healthcare providers. The use of larger and more diverse datasets, along with real-world implementation, can further improve model performance and generalization.

9. Tools & Technologies

The following tools and technologies were used to develop and implement the prediction model: Python, Pandas, Scikit-learn, and Google Colab.

10. References

[1] **Kaggle (2016)**, *Medical Appointment No Shows Dataset (Brazil)*. Available at: www.kaggle.com

[2] **Pedregosa, F. et al. (2011)**, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, 12, 2825–2830

[3] **Python Software Foundation (2024)**, *Python Language Reference*. Available at: www.python.org

[4] *Google Colaboratory Documentation*. Available at: colab.research.google.com