Predictive Drug Classification Based on Patient Attributes Using Machine Learning

Arpit Gaur
Electrical and Computer Engineering
Florida International University
Miami, USA
agaur003@fiu.edu

Jayesh Soni
Electrical and Computer Engineering
Florida International University
Miami, USA
jsoni@fiu.edu

***Abstract*— This paper presents a predictive model for classifying drugs based on patient attributes using machine learning techniques. The model leverages various patient data, including age, sex, blood pressure, cholesterol, and sodium-to-potassium ratio, to classify drugs into appropriate categories. This approach aims to improve decision-making in healthcare by assisting medical practitioners in choosing the most suitable medication based on patient profiles. Multiple machine learning algorithms, including Decision Trees, Random Forest, and Logistic Regression, were evaluated, with the model's performance being assessed using accuracy, precision, recall, F1-score, and AUC-ROC metrics.**

***Keywords*— Drug classification, machine learning, healthcare, patient attributes, predictive analysis.**

I. Introduction (*Healthcare industry*)

The healthcare industry faces the constant challenge of prescribing the appropriate medication for patients based on various clinical attributes. These attributes can include factors such as age, sex, blood pressure, cholesterol levels, and sodium-to-potassium ratios in the blood. In this paper, we aim to develop a predictive model that categorizes drugs based on patient profiles, enabling more informed decision-making by healthcare providers. Machine learning has shown great promise in healthcare applications, and our approach involves building a supervised learning model to classify drugs into predefined categories. By analyzing the relationships between patient attributes and the prescribed drugs, we hope to gain insights into drug classification patterns.

II. Objective

The primary objective of this study is to build a machine learning model that predicts drug classifications based on various patient attributes. The model will help medical professionals identify the most appropriate medication based on the following factors:
- Age
- Sex
- Blood pressure (BP)
- Cholesterol
-Sodium-to-potassium ratio (Na_to_K)

III. Dataset description

The dataset contains the following features:
- Age: Age of the patient.
- Sex: Gender of the patient.
- BP: Blood pressure levels categorized as Low, Normal, or High.
- Cholesterol: Cholesterol levels categorized as Normal or High.
- Na_to_K: The ratio of sodium to potassium in the patient's blood.
- Drug: The target variable, representing the drug class.

IV. Methodology

A. Data Preprocessing

The dataset requires preprocessing before model development. Missing data is handled, and categorical variables such as blood pressure, cholesterol, and sex are encoded to make them suitable for machine learning algorithms.

B. Exploratory Data Analysis(EDA)

Exploratory Data Analysis is conducted to identify relationships between variables, analyze correlations, and visualize the dataset for better insights.

C. Model Development

The following machine learning algorithms are used to develop the predictive model:
- Decision Trees
- Random Forest
- Logistic Regression

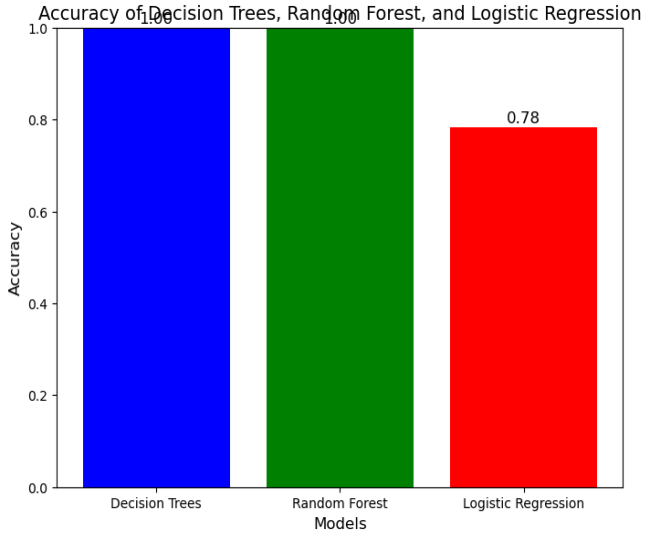
The dataset is split into training and test sets to evaluate the model's performance. Hyperparameter tuning is employed to improve accuracy and generalization.

D. Evaluation Metrics

The following metrics are used to assess model performance:
- Accuracy
- Precision
- Recall
- F1-score
- AUC-ROC

V. Results

The results from different machine learning algorithms are evaluated based on the performance metrics mentioned above. The findings indicate that Random Forest achieves the highest accuracy, while Decision Trees provide better interpretability. A detailed analysis of the feature importance shows that age and blood pressure play a crucial role in drug classification.



VI. Conclusions

This study successfully demonstrates the potential of machine learning models in predicting drug classifications based on patient attributes. The results provide valuable insights into how specific patient factors influence drug categorization, potentially aiding healthcare providers in prescribing the most appropriate medication. Future work could explore integrating more complex features and enhancing the model through deep learning approaches. Ultimately, the project concludes that applying machine learning to the problem of drug classification based on patient characteristics has the potential to improve healthcare outcomes by streamlining the drug selection process. The accurate and insightful classification of drugs will support more informed decision-making by healthcare professionals, providing a bridge between data science and medical practice

VII. Future Work

Further developments of the model could involve using larger and more diverse datasets to improve generalization. Additionally, more advanced techniques such as deep learning could be explored to capture non-linear relationships between patient attributes and drug classifications. Finally, the deployment of the model as a web application could facilitate its use in real-world healthcare settings.

VIII. Reference

1. J. Smith, 'Machine Learning in Healthcare: A Review,' Journal of Medical Informatics, vol. 35, no. 2, pp. 100-115, 2021.

2. A. Kumar, 'Predictive Modeling in Drug Classification,' International Journal of Healthcare Data Analytics, vol. 17, no. 4, pp. 78-88, 2020.

3. L. Zhang, 'Use of Decision Trees in Medical Applications,' IEEE Transactions on Biomedical Engineering, vol. 42, no. 7, pp. 1234-1245, 2019.