Project Proposal: Predictive Drug Classification Using Machine Learning

1. Project Title

Predictive Analysis for Drug Classification Based on Patient Attributes

2. Introduction

The healthcare industry often faces the challenge of determining the appropriate class of drugs for patients based on various factors like age, blood pressure, and cholesterol levels. The proposed project aims to develop a predictive model that classifies drugs into categories using patient data, enabling better decision-making in prescribing medications.

3. Objective

To build a machine learning model that predicts drug classifications based on patient attributes (Age, Sex, Blood Pressure, Cholesterol, etc.).

4. Dataset Description

The dataset contains features such as:

- Age: Age of the patient.
- Sex: Gender of the patient.
- BP: Blood pressure levels categorized as 'Low', 'Normal', 'High'.
- Cholesterol: Levels of cholesterol (Normal/High).
- Na_to_K: Ratio of sodium to potassium in the blood.
- Drug: Target variable, representing the drug class.

5. Methodology

The project will use the following steps to develop the predictive model:

- **Data Preprocessing:** Handle missing data, perform feature encoding for categorical variables (e.g., 'BP', 'Cholesterol', 'Sex').
- **Exploratory Data Analysis (EDA):** Explore relationships between features, analyze correlations, and visualize the dataset to gain insights.
- **Model Development:**
  - Use supervised learning algorithms (e.g., Decision Trees, Random Forest, Logistic Regression).
  - Split data into training and test sets for model evaluation.
  - Perform hyperparameter tuning to improve model performance.
- **Evaluation Metrics:**
  - Accuracy, Precision, Recall, F1-score, and AUC-ROC will be used to assess model performance.
- **Model Deployment:** (Optional) Develop a simple web application or use a Jupyter notebook interface for interaction with the model.

6. Expected Outcomes

- A highly accurate predictive model that can classify drugs based on patient characteristics.
- Insights into how different patient attributes affect drug classification.

7. Tools and Technologies

- Python (for data processing, visualization, and model building).
- Libraries: Pandas, Scikit-learn, Matplotlib/Seaborn for EDA, and machine learning.
- Jupyter Notebook for development.

8. Conclusion

This project aims to leverage machine learning to predict drug classification based on patient attributes, potentially improving decision-making in healthcare settings.