**Drug Detection Machine Learning Model**

**Problem Statement:** This project aims to develop a classification model to predict the name of the drug based on their features and analyse the effectiveness and prescription patterns of various drugs based on patient demographics and health metrices.

**Libraries Used:**

* Pandas: Data manipulation and analysis
* Matplotlib and seaborn: Data visualization
* Numpy: Mathematical operations
* Sklearn: Machine learning library
* Scipy: Scientific computing library
* Mixed\_naive\_bayes: Custom implementation of Mixed Naive Bayes
* Function Transformer, Power Transformer: Preprocessing tools
* ConfusionMatrixDisplay: Display confusion matrix

**Data collection:** I have collected this data from Kaggle containing 200 datapoints with 6 features in the form of a csv file.

**Analysing the raw data**: In the process of analysing the data, I have observed that the data is totally imbalanced as it contains almost 45% of the data is belonging to the drug “DrugY” class label. And in this dataset it doesn’t have any null or duplicated values in the data. It is also contains both numerical and categorical features in the data

**Data Cleaning**: The dataset is cleaned by mapping categorical values to numerical ones for the Sex, BP, Cholesterol, and Drug columns. The target variable, Drug, is transformed into numerical labels. The data is now ready for modeling.

**Data Splitting:** The dataset is split into training and testing sets using the train\_ test\_ split function from sklearn. Stratification is applied to ensure a proportional distribution of classes in both sets.

**Data Transformation:** Probability plots and kernel density plots are used to visualize the distribution of numerical features like Age and Na\_ to\_ K. Transformation methods like square root, logarithm, and power transformations are applied to handle non-normality in the data.

**Pipeline:** Here, I have created a pipeline for all this function to make things easy to transform and share the code to any where.

**Model Training:** A custom implementation of Mixed Naive Bayes is used for classification, considering mixed types of features (numerical and categorical). The model is trained on the transformed training data.

**Model Evaluation:** The model is evaluated on both the training and testing sets. Training error and testing error are calculated. Confusion matrix and accuracy score are used to assess the model's performance on the testing set.

**Results:** The trained model shows promising results with low training and testing errors. Confusion matrix analysis indicates that the model is effective in predicting drug classes for patients.

**Conclusion:** After completing all the preprocessing and training i deployed my model into streamlit so that i can share with some other people.This project demonstrates the development of a Mixed Naive Bayes model for drug classification based on patient attributes. The pre-processing steps, data transformation, and model training contribute to the accuracy of predictions. Further optimization and fine-tuning of the model could enhance its performance.