REPORT OF DRUG CLASSIFICATION USING MACHINE LEARNING

PROJECT OF EXTERNSHIP

BY

TEAM:

1. MADIPADIGE SAICHARAN (19UK1A0520)

2. MAMIDALA VAMSHI (19UK1A0516)

3. NALLAM AMULYA (19UK1A0512)

4.KASU RAHUL(19UK1A0530)

5.BAIREDDY ABHINAVI(19UK1A0567)

SUBMITTED TO SMARTBRIDGE

PROJECT LINK :
https://github.com/smartinternz02/SI-GuidedProject-90777-1658311367

ABSTRACT

The aim of drug classification is to make sure you use a drug safely and get the greatest possible benefit. Every time you take a drug, your body chemistry changes.

An essential quality of drug classification systems is the ability to assign medications to a structured hierarchy for categories such as mechanism of action, physiological effects, and therapeutic indications. No single classification system can meet all of these needs; however, there should be consistency among those that group by the same underlying principals. We discovered discrepancies in how drugs with multiple therapeutic indications are classified among four widely used schemas.

CONTENTS

1. INTRODUCTION

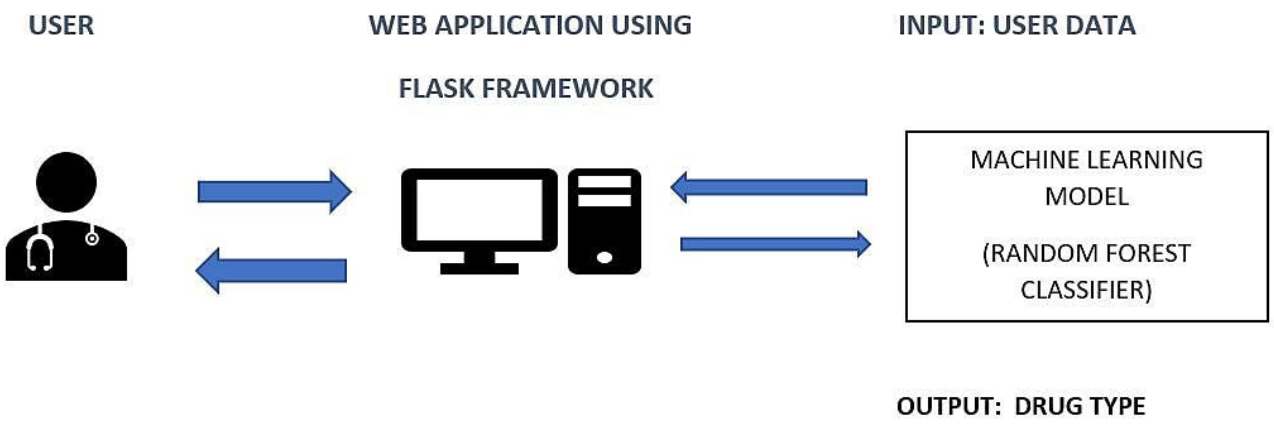
The majority of medications work by interacting with proteins in the host or pathogen. Drug targets include a variety of proteins, and the name receptor is only used when the interaction results in a signal transmission cascade.
A receptor is a molecule or polymeric structure that identifies and binds an endogenous substance on the surface or inside a cell. When a substance elicits a detectable physiological or pharmacological response characteristic of the receptor, it is said to be an agonist. Some medications maybe unable to initiate any action on their own after binding to a receptor site, but they can prevent the action of other agonists. These are called Antagonists.
Humans are predisposed to taxonomy and consequently there is no single and unified system of drug classification.

Depending on the context, the following classifications may prove useful.

2. THEORITICAL ANALYSIS

Different drugs can have different mechanisms of action.Random forest classifier algorithm is used in this project. The drug dataset consists of 'Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K' and 'Drug' features. The target feature is 'Drug'. 'Age' feature represents patient age (numerical), 'Sex' represents patient sex (categorical), 'BP' represents blood pressure of patient (categorical), 'Cholesterol' represents patient cholesterol level (categorical), 'Na_to_K' represents patient sodium to potassium ratio in the blood (numerical) and 'Drug' feature represent the type of drug (categorical).

2.1 ARCHTECTURE OF WEB APPLICATION

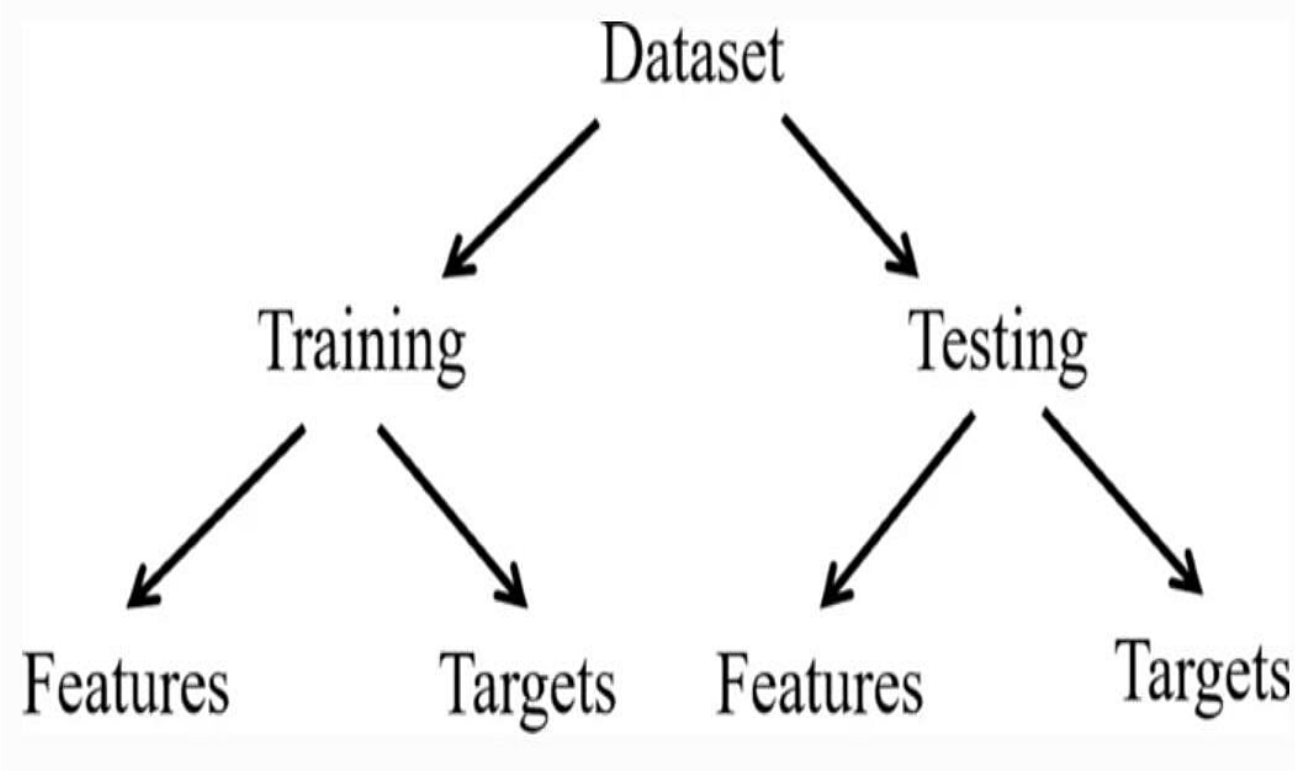


2.2 PRE-REQUISTIES

Python Web Frame Works,Exploratory Data Analysis,Data Preprocessing Techniques,Classification Algorithms,Python-Flask.

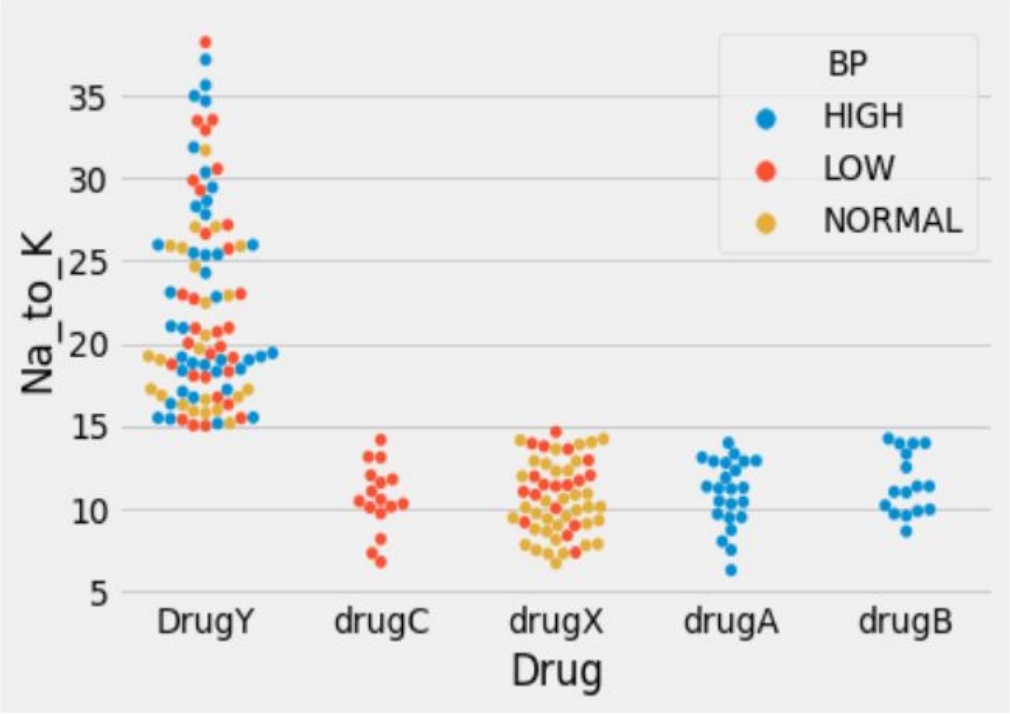
3.DATA PRE-PROCESSING

A detailed analysis of this dataset and the methodology used has been provided in a previous paper. The dataset is taken from kaggle. As shown, the given dataset is divided into the training dataset and the testing dataset. Furthermore, the training dataset and the testing dataset have been divided into the features dataset and the target dataset.



*The following is the **structure of the data set**.*

Variable Name	Description	Sample Data
Age	Patient Age	23; 47; ...
Sex	Gender of patient (male or female)	F; M; ...
BP	Levels of blood pressure (high, normal, or low)	HIGH; NORMAL; LOW; ...
Cholesterol	Levels of cholesterol (high or normal)	1.4; 1.3; ...
Na_to_K	Sodium to potassium ratio in blood	25.355; 13.093; ...
Drug	Type of drug	DrugY; drugC; ...



4. EVALUATION OF MACHINE LEARNING MODEL

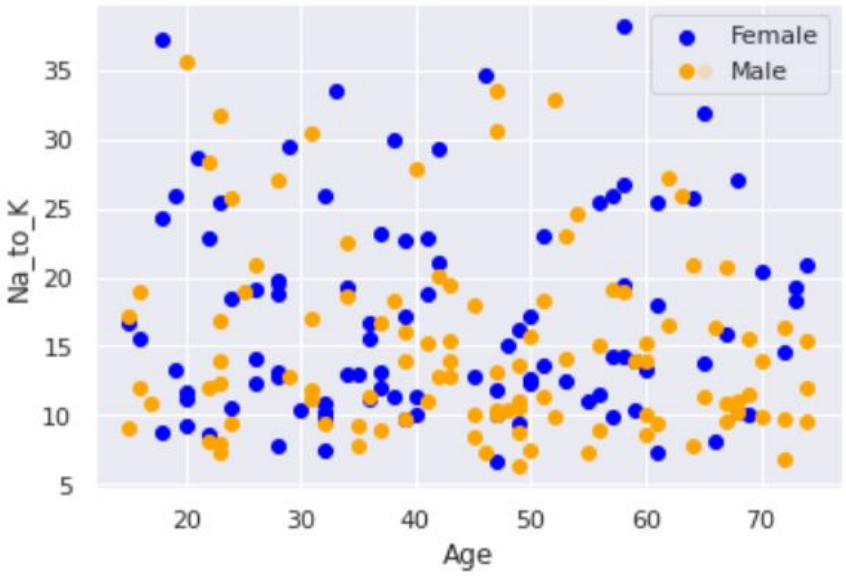
The accuracy of the machine learning model is evaluated by using the Random Forest Algorithm to solve regression problems, you are using the mean squared error (MSE) to how your data branches from each node.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2$$

Where N is the number of data points,
fi is the value returned by the model and
yi is the actual value for data point i.

5. RUNNING THE WEB APPLICATION

The testing dataset which is to be uploaded in the web application (drug200.csv) is shown. For each of the drug present in the testing dataset, the top MoA's for each of the drug is displayed. Along with it, a scatter plot is displayed for each of the drugs. In order to visualize the scatter plot, an ID is given to each of the class of drug. The first class of drug (i.e. 5-alpha_reductase_inhibitor) is given ID 1, the second class of drug (i.e. 11-beta-hsd1_inhibitor) is given ID 2 and so on. This list of ID's is stored in a python list (present in the flask application). The X-axis of the scatter plot represents the ID of the class of the drug, and the Y-axis represents the probability of the drug belonging to that class of the drug. The home page of the web application is shown in Fig. 11. The top classes of drug are shown .



6.OUTPUT SCREEN

7.CONCLUSION

AI technology is employed in pharmaceutical industries including ML algorithms and deep learning techniques in existence. In life science, ML models predict the trained data in an exceedingly known framework i.e., the compound structure can perform alternative tools like PPT inhibitors, macro cycles with traditional algorithms. Additionally, deep learning models are often considered the chemical structures and QSAR models from pharmaceutical data which was pertinent for molecules with appropriate properties, because to the forward success rate in clinical trials. AI technology has taken a forward step in stepping into computer-aided drug development to retrieve the powerful capabilities in data processing.

In the future, if a better-performing machine learning model is discovered, they can seamlessly integrate their machine learning model into the web application provided in this paper. Also, for a better visualization of the scatter plots, the web application can be extended to use interactive scatter plots using plot, bokeh or any other plotting packages.