Group Details:

Group Name	SoleDS
Name	Anju Paul
Email	anjupaul14@yahoo.co.in
Country	UK
College	Anglia Ruskin University
Specialization	Data Science

Problem Description:

To understand the persistency of a drug as per the prescription given by the physician is an important question faced by the pharmaceutical companies. The problem here is to build a classification model to understand the persistency (persistent or not) of a drug for the given dataset.

Model:

Various machine learning models has been built and compared and a best model has been selected. The models used are Logistic Regression (Linear Model), Support Vector Classifier, Naïve Bayes, K-Nearest Neighbors, Decision Tree, Random Forest(Ensemble Learning- Bagging) , Adaboost(Boosting) and XGB(Boosting). The comparison matrix for various models is as shown below.

1 to 8 of 8 entries Filter

Models	Sensitivity	Specificity	Accuracy	AUC Values	F1 Score Values
Logistic Regression	0.40437158469945356	0.9290953545232273	0.7668918918918919	0.7965315911125362	0.7446727939489631
Support Vector Classifier	0.366120218579235	0.9339853300733496	0.7584459459459459	0.7982951888519246	0.731491659766629
Decision Tree Classifier	0.48633879781420764	0.8166259168704156	0.714527027027027	0.6514823573423116	0.7099516198358813
Naive Bayes	0.4644808743169399	0.8801955990220048	0.7516891891891891	0.7804855238018892	0.7395148319266621
K-Nearest Neighbors	0.4207650273224044	0.9070904645476773	0.7567567567567568	0.7882279850895828	0.738338562781425
Random Forest CLassisifier	0.44808743169398907	0.9119804400977995	0.768581081081081	0.8031050008684382	0.7521116723854484
Adaboost	0.4098360655737705	0.9290953545232273	0.768581081081081	0.7972263417371437	0.7469214302459266
XGB	0.47540983606557374	0.8899755501222494	0.7618243243243243	0.7228546234318008	0.7495324808615949

From these values, Random Forest Classifier has a better matrix. Hence the final selected model is Random Forest Classifier.

A prediction has also been done on the final selected model using the prediction system.