FOREST COVER TYPE PREDICTION

ABSTRACT

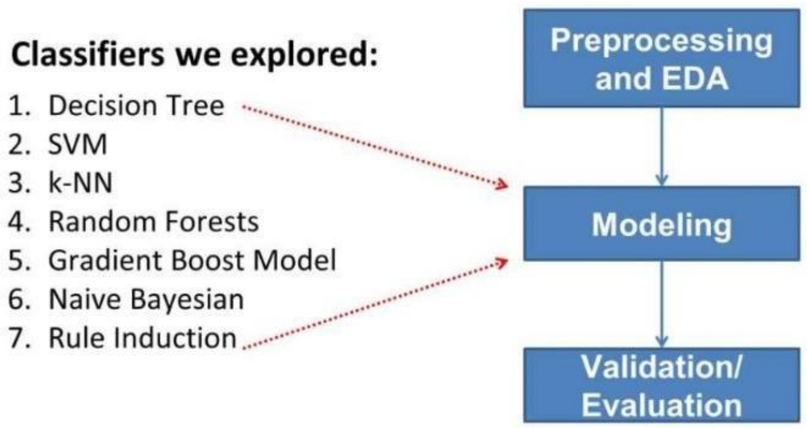
To predict forest cover type from descriptive variables. Each observation (30*30 meter cell) was determined by the US Forest Service.

OBJECTIVES

- The studied area includes 4 wilderness areas located in the Roosevelt National Forest in Colorado. These areas represent forests with minimal human-caused disturbances, so forest covers are more a result of ecological processes than forest management practices
- We will be analyzing various machine learning algorithm on Forest Cover Type. We need to predict the cover type using the given data. We will also analyze various parameters of the applied algorithm and view the effect on the dataset. Parameter tuning using Grid search is done. The dataset is asking one to predict the type of forest cover from given cartographic variables.

PROCEDURE

The data is already clean and there are no null values. We applied Random Forest Classifier algorithm as there are seven types of forests. We have got the initial accuracy of 0.859. In the next step we have calculated the feature importance. We have also dropped the features which contributed less to the accuracy. Now we got an accuracy of 0.862 after re-classifying the data again. We have added few new attributes using the already existing attributes. Calculating the accuracy score again, we got it as 0.90. We tried using boosting algorithms as Adaboost which gave us an accuracy of 0.91.Like wise we do for all other algorithms

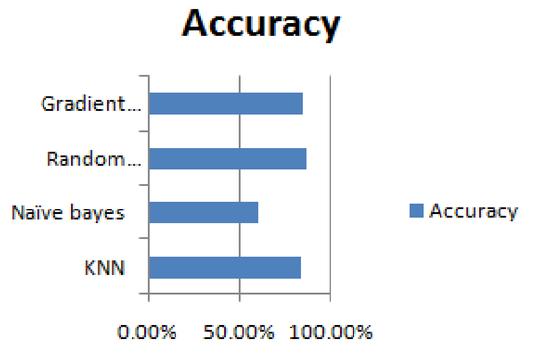


FUTURE SCOPE

- ❑Semi Supervised Learning Methods to increase the training data size.
- ❑Using two-way classification approaches to distinguish between majority class groups with minority.
- ❑Feature Engineering using principal component Analysis.
- ❑Apply of advanced classifiers and boosting methods.

RESULT:

Model	Accuracy	Error
KNN	83.90%	16.10%
Naïve bayes	60.10%	39.90%
Random Forest	86.40%	13.60%
Gradient boosting	84.50%	15.10%



PROJECT MEMBERS

T. Supriya	18MIS1064
N. Pranitha	18MIS1095
B. Priyanka	18MIS1097