Group 7 - Topic Proposal

Forest Covertype Prediction

Subject: Machine Learning 1
Professor: Amir Jafari
Team Members: Yashwant Bhaidkar, Akhil Bharadwaj

Overview:
Forests are playing a vital role in the environmental lifecycle and it has a significant impact on the living species. It is very important to conserve the forest as we are facing multiple problems like deforestation,climate change and wildfire.To maintain the proper information of the forest is a critical task and there are several surveys conducted to find out the various factors about the forest. Using this information, we can predict various things about the forest and one of the prediction tasks is Forest covertype prediction.
Forest classification is an important tool for ecological management, conservation, resource management, and land use planning. It helps in understanding the ecological characteristics of different forest types, identifying areas of high conservation value, developing sustainable resource management strategies, and making informed decisions about land use.

About Dataset:
There are various techniques like Field Sampling,Remote Sensing,Geographic Information Systems (GIS),Climate data to collect the information for the analysis and prediction.
Data collection for forest type classification involves collecting information on the ecological characteristics of the forest, such as tree species, tree density, soil characteristics, topography, and climate. For this project, we are using UCI covertype dataset.

http://archive.ics.uci.edu/ml/datasets/covertype

Models and performance Index:

- This is a multiclass classification problem as we are classifying the data points into 7 different categories.We will use some of the advanced machine learning models like Random forest, Ensemble techniques and deep learning models like MLP classifier.
- We will use the sklearn library for all the above mentioned models.
- We will check the accuracy of prediction for model performance(how many data points are classified correctly). Also,For the quality of the model, we will use the Confusion matrix for all the classes(7X7 matrix).

References:
https://www.tandfonline.com/doi/full/10.1080/08839514.2020.1771523#:~:text=The%20scientific%20importance%20of%20classifying,in%20the%20soil%2C%20among%20others
analysis of forest land use, forest land cover and change at policy-relevant scales | Forestry: An International Journal of Forest Research | Oxford Academic (oup.com)

Project timeline:
Data cleaning and imputation(if required): till 16th April
Exploratory Data analysis,General analysis,feature selection and feature engineering: till 22nd April
Modeling and deployment: till 30th April