

DMLL Group 3 Lab 1 Report

Date: 22-09-22

Members:-

Gowresh
Yehia
Sheetal
Gunjan
Nasir

Summary of Discussion:-

Forest Fire Dataset

The dataset chosen for our lab work is generally obtained from satellites Tegra and Aqua using an instrument called MODIS based on hotspots of forest fire in Turkey. There are totally 36,011 records/instances in the dataset with each row pertaining to the satellite readings within a specific time interval. The columns represent 15 attributes/features. The target feature is the 'confidence' which helps to gauge the quality of the hotspot/fire. The 'confidence' attribute is logged in terms of percentage, which can be represented as a nominal data with three categories Low, Nominal/Medium and High. This dataset can be used to train a model to predict whether the confidence of the forest fire is Low, Nominal or High.

The following are the each attribute name with their data types and attribute types:

Latitude (metres)
float (ratio)
Longitude (metres)
float(ratio)
brightness temperature (Kelvin)
float(ordinal)
scan (metres)
float(ratio)
track (metres)
float(ratio)
acq_date
string(interval)
acq_time
int(interval)
satellite
String(nominal)
instrument
String(nominal)
confidence (%)
int(can be represented as nominal)
version
float(nominal)
brightness 31 channel temperature(Kelvin)
float(ordinal)
frp (MW)
float(ratio)

type (inferred hotspot type)
int(can be represented as nominal)
day night
string(nominal)

This dataset has a good combination of categorical and ordinal data that will help us to best understand the ML Practices. There are no missing values in the dataset which may benefit us with us accuracy of prediction. The end outcome we are expecting from using this dataset is to predict the confidence levels of the hotspot given a set of attributes. The target feature can be used to create categorical data that will help us to understand the normalisation and binning concepts effectively.

Other Datasets chosen:

Mobile Robot Floor Prediction -

A dataset containing 500,000 records containing 10 of quaternion, acceleration and velocity features that all are ratio attributes and a target feature 'floor type' that is nominal

Hand Gesture Detection -

A dataset containing 778 training images of hand gestures and 6 different classes to predict. Additionally labels are provided defining the bound boxes of the classes in each image.