SYNOPSIS

The project used simulated data from a ground-based atmospheric Cherenkov gamma telescope to venture into the field of machine learning. The project explored many facets of machine learning through practical investigation, providing priceless insights into preprocessing techniques, model selection, evaluation metrics, and data visualisation.

Extensive testing of various classification algorithms was one of the project's highlights. Through this investigation, participants gained a broader comprehension of model selection and were able to identify the advantages and disadvantages of each algorithm in relation to the given task.

When evaluating the performance of the model, evaluation metrics were crucial. Metrics like accuracy, precision, recall, F1-score, confusion matrix, and classification report guided subsequent iterations and improvements by offering insightful information about the models' effectiveness.

On 11th May 2024, I downloaded my dataset and turned it into .csv format. I read several study materials regarding different classifiers and made the choice to include- Decision Trees, Naive Bayes, Random Forest, Logistic Regression, K-Nearest Neighbours, and Support Vector Machines in my project. I watched several youtube videos regarding the correct way of approaching my project.

By the end of the day I had successfully imported my dependencies, imported my dataset, done data preprocessing i.e. checked for missing values and converted categorical data into numerical data, made bar-plot and heatmaps and split the data.

On 12th May 2024, I learned how to apply various classification models and how to compare them. By the end of the day I had successfully standardised the data, predicted training and testing data accuracy and printed the confusion matrix and the classification report for each of the 6 models.

I also developed a predictive model, was able to visualise the distribution of the values of attributes and was able to successfully compare the 6 models using bar charts.