Best Algorithm Finder

Semester Project

Group Id : 02

Course :

Data Mining

Course Teacher:

Dr. Salahuddin

Lab Teacher:

Engr.Ramsha Mashood

Group Name: DM Project Group

Group Members:

- Mahnoor Gohar (57239)
- Qaiser Abbas (57245)
- Gul Saba (57227)

Department Of Software Engineering

MOTIVATION:

- To Implement DM Concepts and to learn more about Machine Learning Algorithms.
- To enhance our understanding and to test our skills.
- To introduce something that is beneficial for finding best algorithms for any dataset and to find which one is more accurate.

INTRODUCTION

- The goal of this project is to examine several ML algorithms for classification and clustering that could adjust to any Dataset and attempt to find which one is more accurate.
- Our goal with this is to perform some initial data visualization and to determine which Algorithm handles any dataset the best.

TASK DISTRIBUTION

REG. No	MEMBERS NAME	TASKS
57239	Mahnoor Gohar	(Front end & Implementation of Clustering Algorithms)
57245	Qaiser Abbas	(Front end &Implementation Of Classification Algorithms)
57227	Gul Saba	(Front end &Implementation Of Classification Algorithms)

TECHNOLOGY

- Python language is used including following libraries: pandas, matplotlib, flask, numpy, sklearn, classifier, clusterer, scipy, preprocessor, algo_data.
- PyCharm is used as IDE to check the output and response.
- Datasets are available on Kaggle:
- https://www.kaggle.com/aljarah/xAPI-Edu-Data
- https://www.kaggle.com/uciml/iris

- For Classification Bagging Classifier, Gradient Boosting, AdaBoost Classifier, Ridge Classifier, Decision Tree Classifier and Extra Tree Classifier are used.
- For Clustering K-Means Clustering Algorithm and Agglomerative Clustering Algorithm are used.

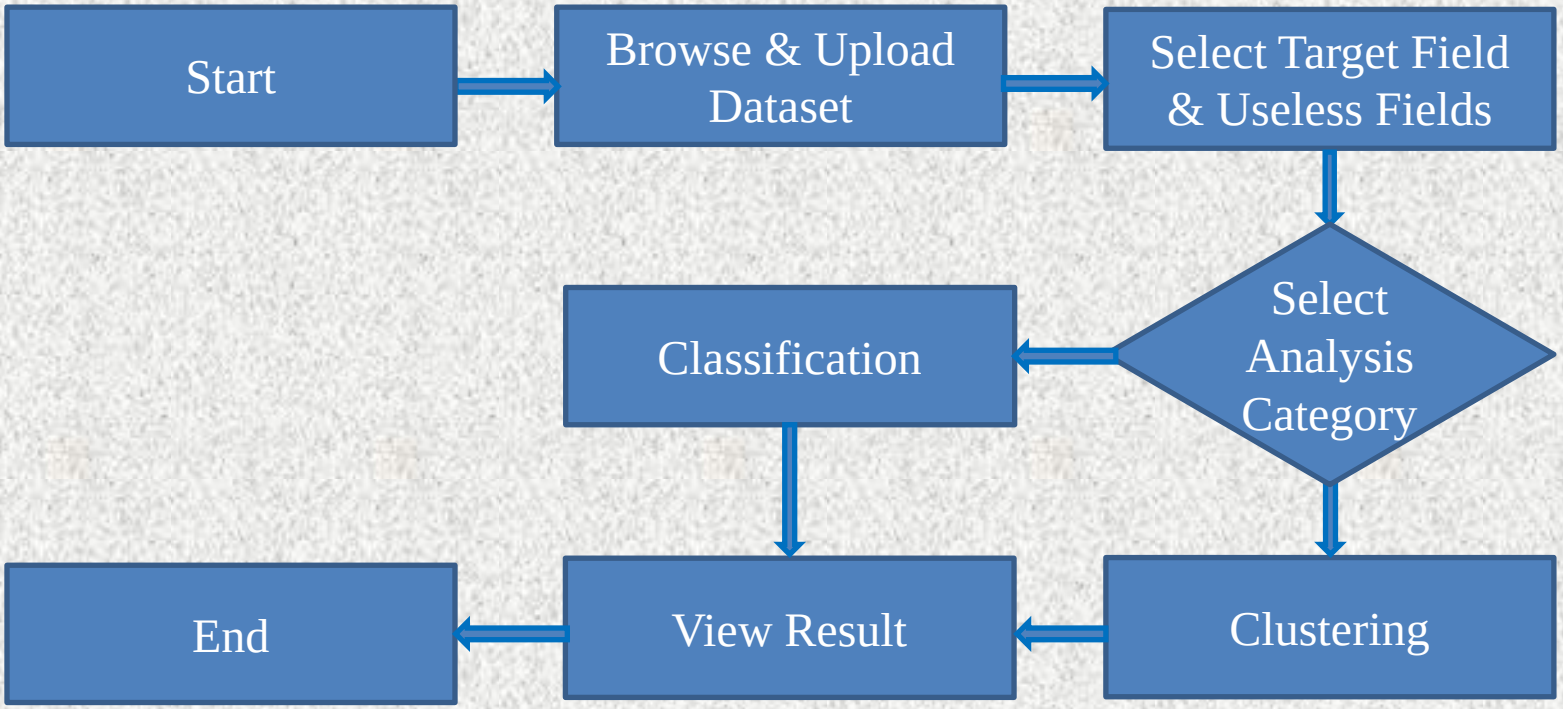
TARGET MARKET

- In this project **Best Algorithm Finder**, we have used different algorithms to predict the accuracy of the given dataset for multiple features in order to determine which algorithm handles the dataset better. So, using this Project, we can overcome the problems that we can face during implementation of algorithms on any dataset and hence can reduce the chances of inaccurate results or less accurate results.

FEATURES

- We can read data from csv, txt, tsv, xlsx file.
- We can perform analysis on data through classification and clustering.
- We can select target column and exclude useless data fields from any dataset.
- We can determine which ML Algorithm is best and more accurate for our dataset.

OVERVIEW
(Flow Diagram)



OVERVIEW
(Details)

- First, we will browse and upload any dataset file in csv, txt, tsv, xlsx format in order to read that dataset.
- Then we will choose analysis category(i.e., classification and clustering) in order to perform analysis on data.
- After that we will select target column and exclude useless data fields from dataset.
- Then we will view analysis results and will be able determine which ML Algorithm is best and more accurate for our dataset.

PROS

- It will be beneficial to determine best and more accurate algorithm for any dataset.
- It will save time as we don’t have to implement and test multiple algorithms one by one in order to find accurate results.

ENHANCEMENT

- Further Enhancement of this project will be to implement other ML Algorithms for Classification and clustering in order to predict more accurate results.
- We can also try other datasets with more deep information.


Thank You