**Project Name:**

ASTRA

(Autonomous Statistical Toolset for Research & Analysis)

Description

ASTRA seamlessly integrates statistical inference and machine learning methods to explore data, visualize insights, and optimize model selection for comprehensive research and analysis.



**Group Members**

Abdullah Arif 22L-7764

Yousha Saibi 22L-7482

Tahmooras Khan 22L-7484

Kabir Uddin Shahab 22L-7486

National University Of Computer and Emerging Sciences

Department of Computer Science

Lahore, Pakistan

# Project Overview

ASTRA is a user-friendly tool designed to streamline data analysis processes in academia, business, and research. It offers a user-friendly interface for uploading, cleaning, and performing statistical analysis, predictive analytics, and comparing machine learning models. The project aims to expedite research workflows and provide actionable insights from data. By simplifying complex tasks, ASTRA allows researchers and students to focus on interpreting results and making informed decisions, accelerating innovation and discovery.

# Problem Statement

In today's data-driven world, the process of preparing, analyzing, and modeling datasets is often time-consuming and complex. Many users face challenges in handling datasets efficiently, from cleaning and visualizing data to selecting the most appropriate machine learning models for their tasks. Iit requires programming skills and a lot of time to perform data munging and robust analytics. This complexity can hinder users from effectively utilizing the potential insights hidden within their data. Therefore, there is a pressing need for a desktop application that streamlines the data preprocessing, analysis, and model selection processes, while also providing intuitive visualizations and comprehensive record-keeping capabilities.

# Solution to the Problem

To address the challenges outlined in the problem statement, we propose the development of a comprehensive desktop application named "ASTRA." ASTRA will provide users with a user-friendly interface to upload datasets, perform data cleaning operations (such as handling null values and removing duplicates), automatically visualize key statistical summaries, record cleaned datasets into a database, offer a selection of machine learning models for analysis, visualize training results, compare model accuracies, and maintain separate records for each user. Through these features, ASTRA will streamline the entire data analysis workflow, empowering users to efficiently preprocess datasets, explore insights through visualizations, select the most appropriate machine learning models, and evaluate model performances accurately.

# Scope:

ASTRA is a desktop-based data analysis tool that primarily focuses on basic statistical summaries and box plots, without extending to real-time processing or complex visualization. It includes essential data cleaning functionalities but does not cover advanced data preprocessing techniques or feature engineering. ASTRA facilitates model training and comparison but does not cover custom machine learning algorithms or deep learning models. It primarily integrates existing libraries and frameworks for model training and evaluation. ASTRA does not aim to provide advanced database management features or support for large-scale data storage and retrieval optimization.

# Basic Features of ASTRA

1. **Dataset Upload:**

With ASTRA, users may easily upload datasets, which simplifies the integration of labeled data for supervised learning tasks such as regression and classification.

1. **Dataset Cleaning:**

To make sure that datasets are consistent and appropriate for supervised learning analysis, ASTRA offers data cleaning operations including removing null values and duplicate entries.

1. **Automatic Visualization of 5-Point Statistical Summary:**

Before training a model, ASTRA facilitates data exploration and understanding by visualizing the five-point statistical summary for every numerical variable in the dataset

1. **Dataset Recording in the Database:**

For every numerical variable in the dataset, ASTRA offers a visual representation of the five-point statistical summary. This helps with understanding and data exploration before training the model.

1. **Selection of Supervised Learning Models:**

Users can choose the most appropriate supervised learning model for their particular research from ASTRA’s selection of models for classification ( such as decision trees, random forests, and support vector machines) and regression ( such as linear regression, polynomial regress, and gradient boosting).

1. **Visualizing of Training Results:**

In order to make the assessment and comprehension of some supervised learning models easier, ASTRA provides visualizations of the training procedure and outcomes, such as accuracy metrics, loss curves, and model performance graphs.

1. **Accuracy Comparison with Other Models:**

Researchers may choose the best-performing model for their classification or regression tasks by comparing the accuracy and performance of several supervised learning models applied to datasets using ASTRA.

1. **Feature Selection and Importance:**

By assisting users in determining the most pertinent variables for classification and regression tasks, ASTRA’s feature selection and feature importance analysis tools enhance the interpretability and performance of models.

1. **Integration of Interactive Dashboard:**

With the help of ASTRA’s interactive dashboard interface, users may interactively examine data trends, see model predictions, and have a greater understanding of their study findings.

# Project Plan

ASTRA's development process is divided into four iterations, each focusing on key functionalities. The first iteration sets up the project framework, followed by adding features like user authentication and data cleaning. The third iteration integrates a database and refines the user interface. The final iteration includes comprehensive testing, bug fixing, and documenting project details for a polished end product.

| Iteration | Stories / Modules |
| --- | --- |
| 0 (current) | - Scrum formulation and meeting  - Planning the development process  - Building a rough estimate of the prototype |
| 1 | - User authentication system  - Functionality to upload datasets  - Basic data cleaning features (e.g., removing null values)  - Statistical analysis functionalities (e.g., five-point summary) |
| 2 | - Develop UI components for model selection  - Model training and evaluation  - Integrate database for storing datasets  - Initial model training functionality |
| 3 | - Integrate database for analysis results  - Refine user interface based on usability testing and feedback  - Conduct comprehensive testing and bug fixing  - Document project details and user guides |

# Team Roles

| **Name** | **Roles** |
| --- | --- |
| Abdullah Arif | Lead/Developer |
| Kabir Uddin Shahab | Developer/Architect |
| Tahmooras Khan | Tester |
| Yousha Saibi | Requirement Engineer |

# 

# Prototype

A prototype is made with a Pencil tool and is attached with this document.