1. Introduction

Welcome to the Data Analysis and Preprocessing Application! This tool is designed to help you easily upload, preprocess, analyze, and apply machine learning models to your datasets. Whether you're dealing with missing data, outliers, or simply need to visualize your data, this app has you covered.

2. Getting Started

To begin using the application:

Launch the app.

You'll see a sidebar on the left with navigation options.

3. Data Upload

Home Page:

Navigate to the "Home" section.

Upload your dataset in CSV format using the file uploader.

The app will automatically display the first few rows of your data.

4. Data Preprocessing

After uploading your data, you can perform various preprocessing tasks:

Handling Missing Values

Check the "Handle Missing Values" option in the sidebar.

The app will fill missing values with the most frequent value for categorical columns and the mean for numerical columns.

Removing Duplicates

Check the "Remove Duplicates" option in the sidebar.

The app will automatically remove any duplicate rows from your dataset.

Data Encoding

Navigate to the "Data Analysis with Your Data" page.

Select the "Data Encoding" section.

Choose your preferred encoding method:

Label Encoding: Converts categorical labels to numeric values.

Ordinal Encoding: Converts categories into ordered numerical values.

One-Hot Encoding: Converts categories into a binary matrix.

Select the columns to encode and apply the encoding.

Feature Scaling

In the "Data Analysis with Your Data" page, go to the "Feature Scaling" section.

Choose between "Normalization" (scales values between 0 and 1) and "Standardization" (scales values based on z-scores).

Select the columns to scale and apply the scaling.

Outlier Detection and Handling

Select the "Outlier Detection and Handling" section.

Choose a column to analyze for outliers.

Apply either the IQR method or capping to handle outliers.

Visualize the results with updated box plots.

5. Exploratory Data Analysis (EDA)

In the "Exploratory Data Analysis" section, choose the type of chart you want to generate:

Bar Chart

Box Plot

Heat Map

Scatter Plot

Pie Chart

Histogram

Distplot

Select the column(s) for analysis and generate the desired plot.

6. Model Application

You can apply machine learning models directly within the app:

Regression Models

Select "Regression" under the "Model Application" section.

Choose the target column (the column you want to predict).

The app will automatically apply various regression models (e.g., Linear Regression, Ridge, Lasso) to your data.

Review the performance metrics for each model.

Classification Models

Select "Classification" under the "Model Application" section.

Choose the target column (the column you want to classify).

The app will apply classification models (e.g., Logistic Regression, SVM, Random Forest) to your data.

Review the performance metrics such as accuracy, precision, recall, and F1 score.

7. Downloading Processed Data

After you've completed your data preprocessing, analysis, or model application:

Go to the sidebar and select the "Download Analyzed Data" section.

Choose the format in which you want to download the processed data (CSV or Excel).

Click the "Download" button to save the file to your device.

8. FAQs and Troubleshooting

What file formats are supported for upload?

Currently, only CSV files are supported.

How can I revert changes made to the dataset?

Refresh the app and re-upload your dataset to start fresh.

What if the app crashes or shows an error?

Ensure that your data is clean and properly formatted. If issues persist, consider reloading the app or contacting support.