

## HW7 Report

This report describes the analysis of the dataset from the CSV file 'data\_banknote\_authentication.txt'. It generates various graphs to examine and analyze data, and provides insights into the dataset.

### Data Loading:

The code starts by importing necessary libraries, including NumPy, Matplotlib, Pandas, Seaborn, and scikit-learn. It loads a dataset from the file 'data\_banknote\_authentication.txt' using Pandas. The dataset is assumed to be a CSV file with a comma delimiter.

### Data Exploration:

The code separates the independent variables (features) into 'x' and the dependent variable (target) into 'y'. It prints the first 10 rows of the dataset to inspect the data structure. It also checks the data type of the 'dataset' object and prints the last 5 rows. The dimensions of the dataset are displayed using 'dataset.shape'.

### Data Quality:

The code identifies and prints the number of duplicate rows in the dataset.

### Data Splitting:

It splits the dataset into training and testing sets using the 'train\_test\_split' function from scikit-learn, with a 75% training and 25% testing split ratio.

### Decision Tree Regression:

A DecisionTreeRegressor model is created with a specified maximum depth and random state. The model is trained on the training data, and predictions are made on the testing data. The mean squared error between the actual and predicted values is calculated and printed using 'mean\_squared\_error'.

### Logistic Regression:

A LogisticRegression model is created and fitted with the training data. Predictions are made on the testing data. The mean squared error and accuracy score of the logistic regression model are calculated and printed.

Support Vector Machine (SVM):

An SVM model with a linear kernel is created and trained with the training data.

Predictions are made on the testing data.

The mean squared error and accuracy score of the SVM model are calculated and printed.

Naive Bayes:

A Gaussian Naive Bayes (GaussianNB) model is created and trained with the training data.

Predictions are made on the testing data. The mean squared error and accuracy score of the Naive Bayes model are calculated and printed.

This code appears to load a dataset, perform data exploration, split the data, and evaluate various machine learning models, including Decision Tree Regression, Logistic Regression, Support Vector Machine, and Naive Bayes. It calculates and prints mean squared errors and accuracy scores for each model, making it suitable for model evaluation and comparison on a binary classification problem.

**Discuss possible problems you plan to investigate for future studies**

- Text Mining
- Clustering(UnsupervisedLearning) • NeuralNetworkandDeepLearning