**MAJOR PROJECT REPORT**

**DATASCIENCE**

**Topic: FAKE CURRENCY DETECTION**

**Group 10**

**-**

**Divyesh Cirikonda, Mansi Sapariya, Nishna Kumar, Supriya Vadali,**

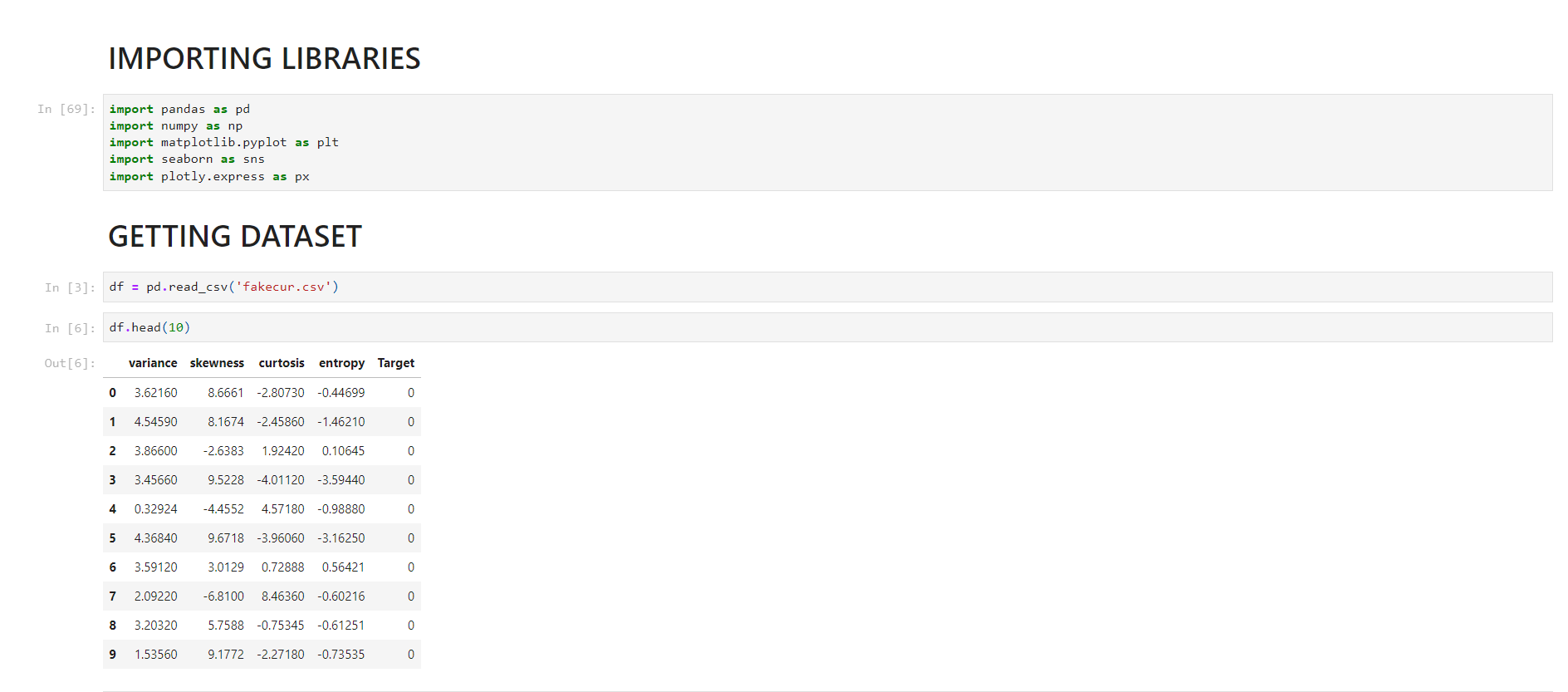
**Vaishnavi Chintha, Adil Zafar** and **Sriman Upadhyayula**

**Introduction:**

This report is based on EDA, Pre-processing and model optimization of the given dataset by running the files using jupyter notebook installed with the essential libraries. Converting models into user-interface webapp using streamlit.

**Data Description:**

Fake currency dataset provided to the group was analysed and then models were applied on the same. In the below picture the details of the dataset are shown which shows that there are 10 rows and 6 columns. We were asked to predict the ‘Target’ column of the dataset.

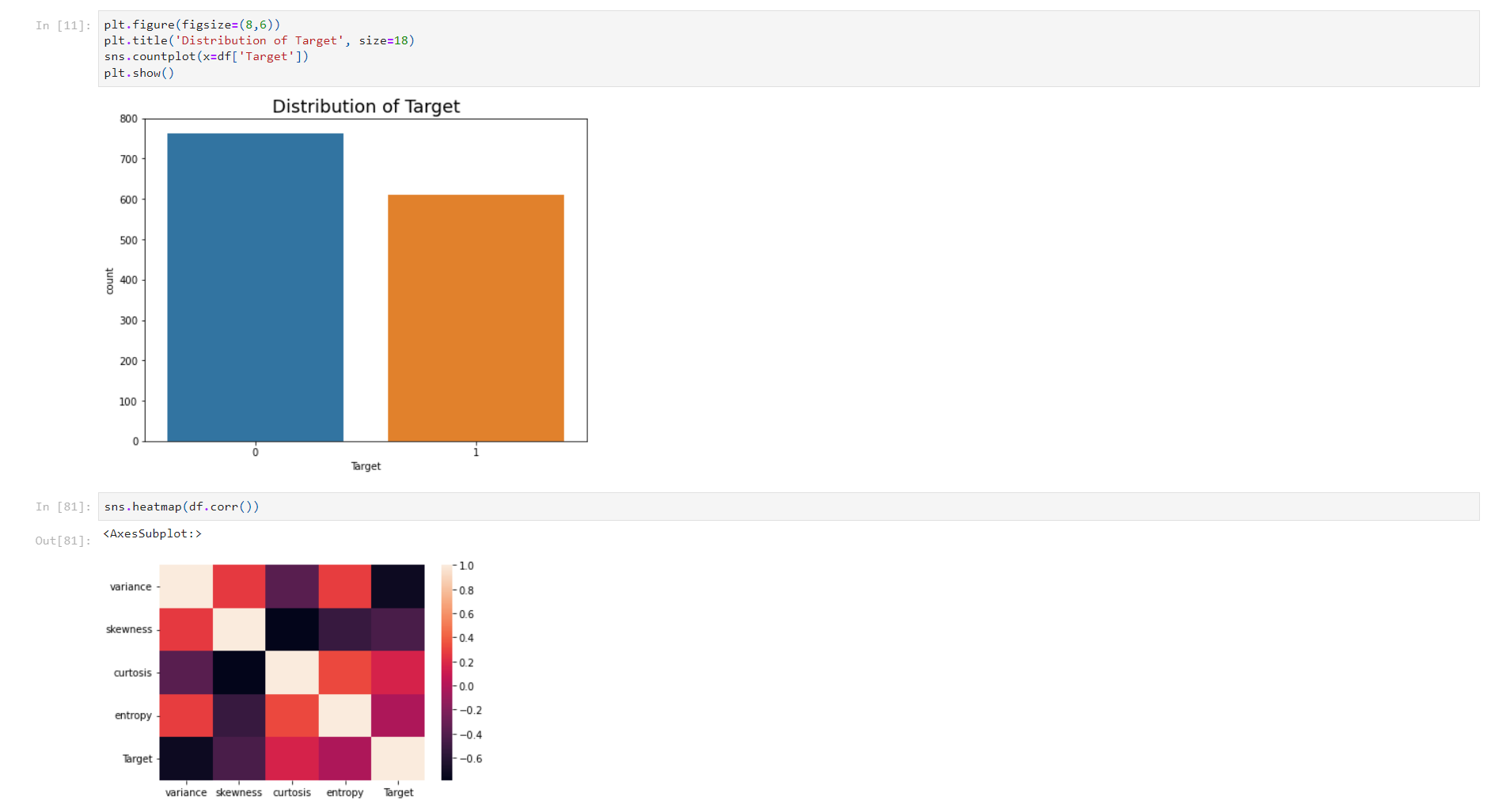
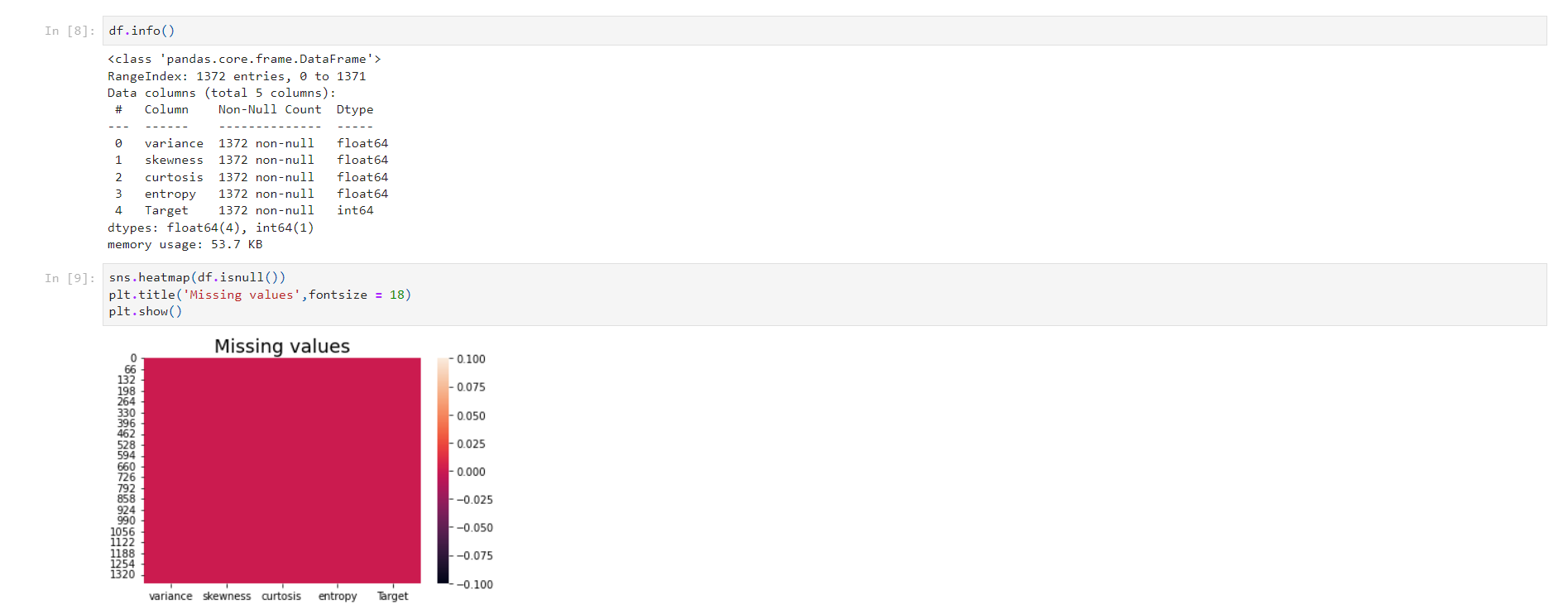
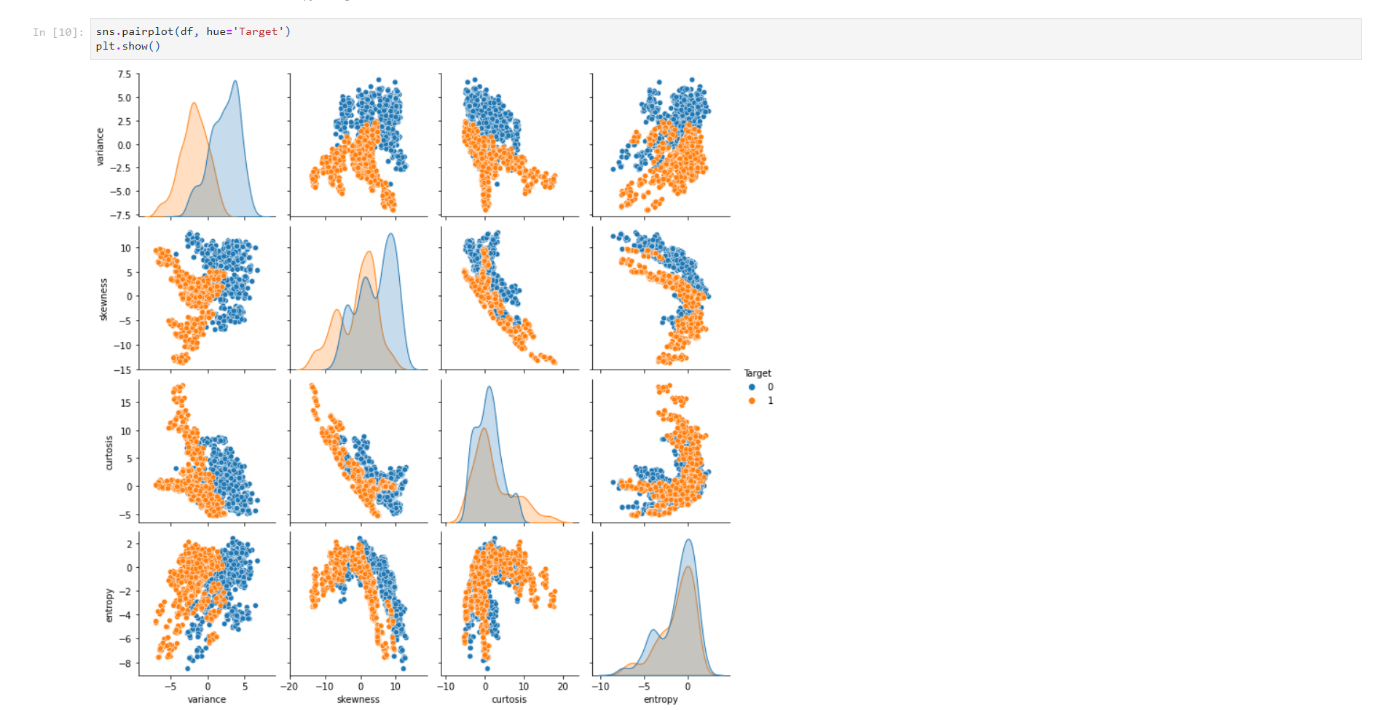
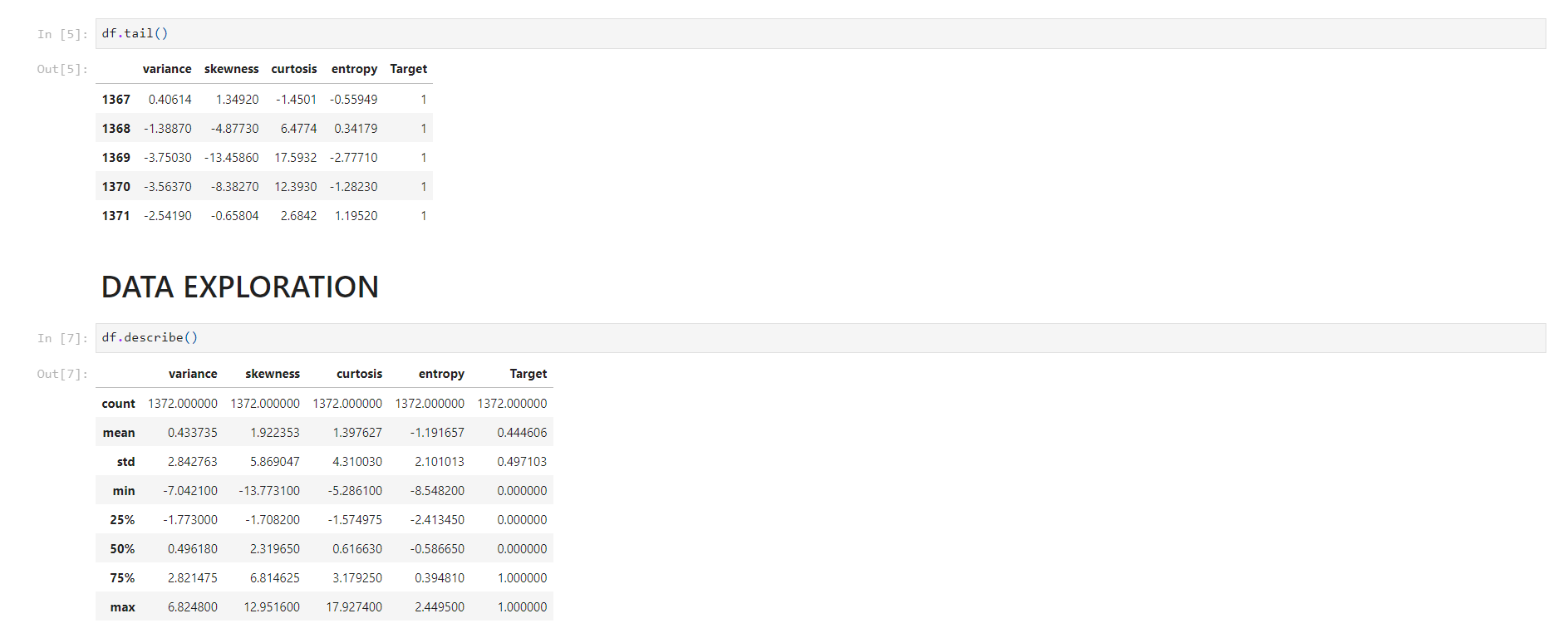


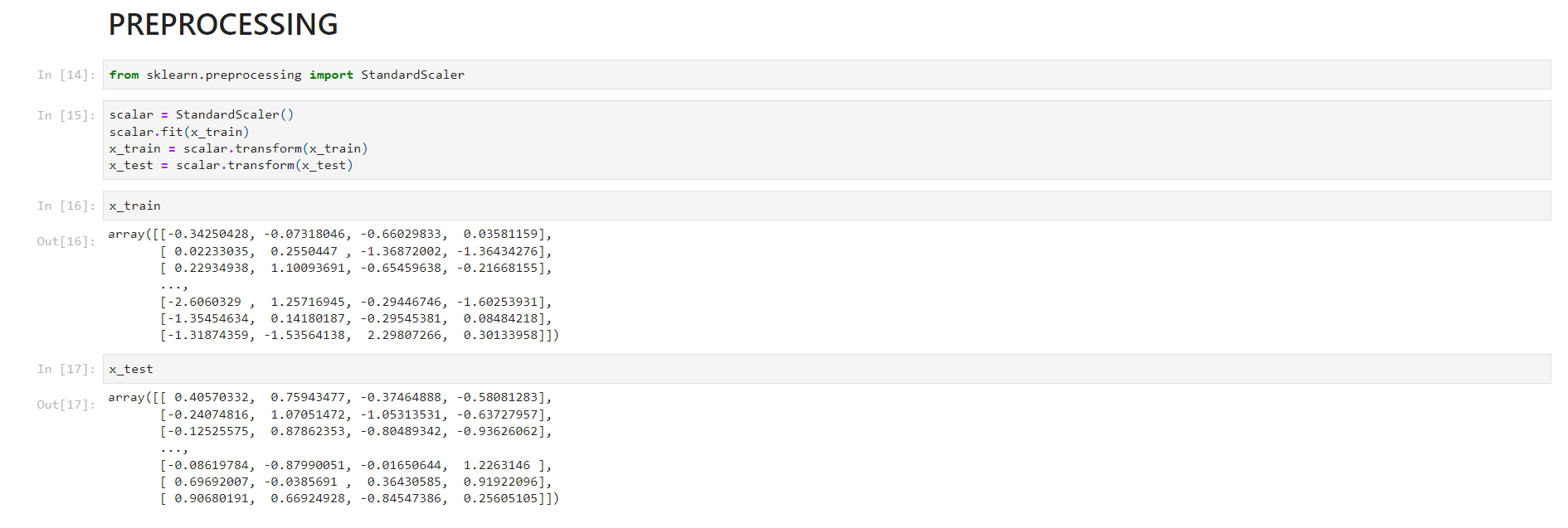
**Approach:**

Data has been analysed to find the null values in the dataset and a specific plot for the same has been made. Correlation of all the features have been analysed. Pair plots and other required visualization has been carried out. Data pre-processing and then models were applied later. For data visualization and user interface web app, streamlit application has been used.

**Visualization:**

In visualization, the process of finding null values, correlation matrices, pairplots have been obtained. Below figures represent the same.



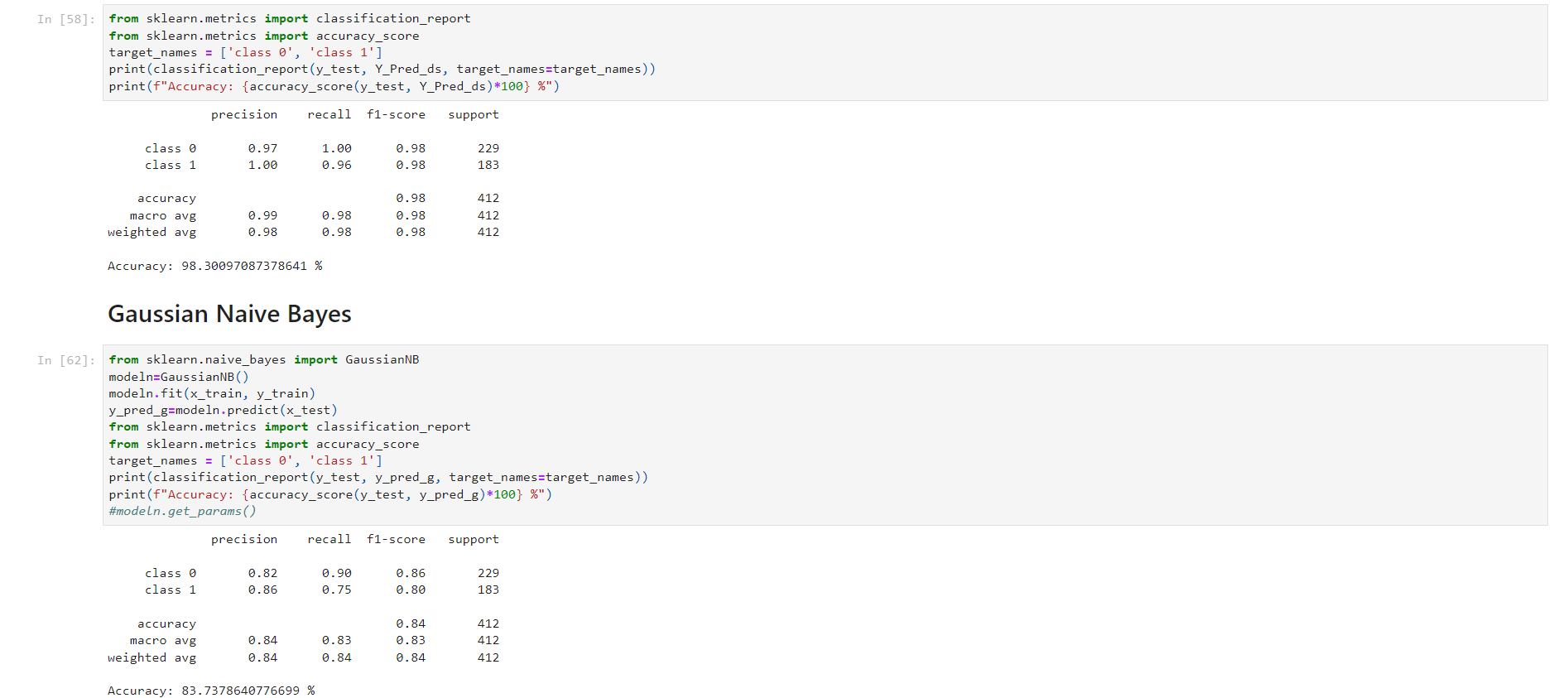
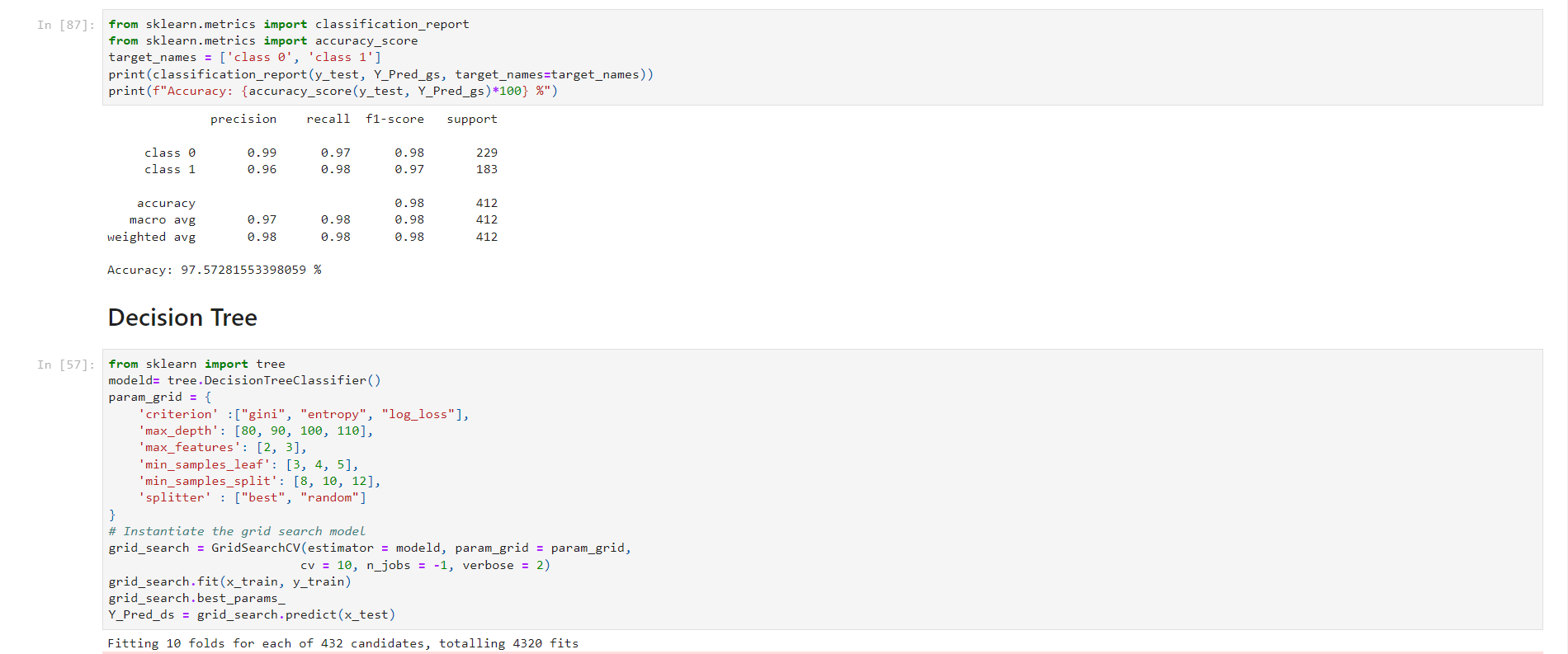
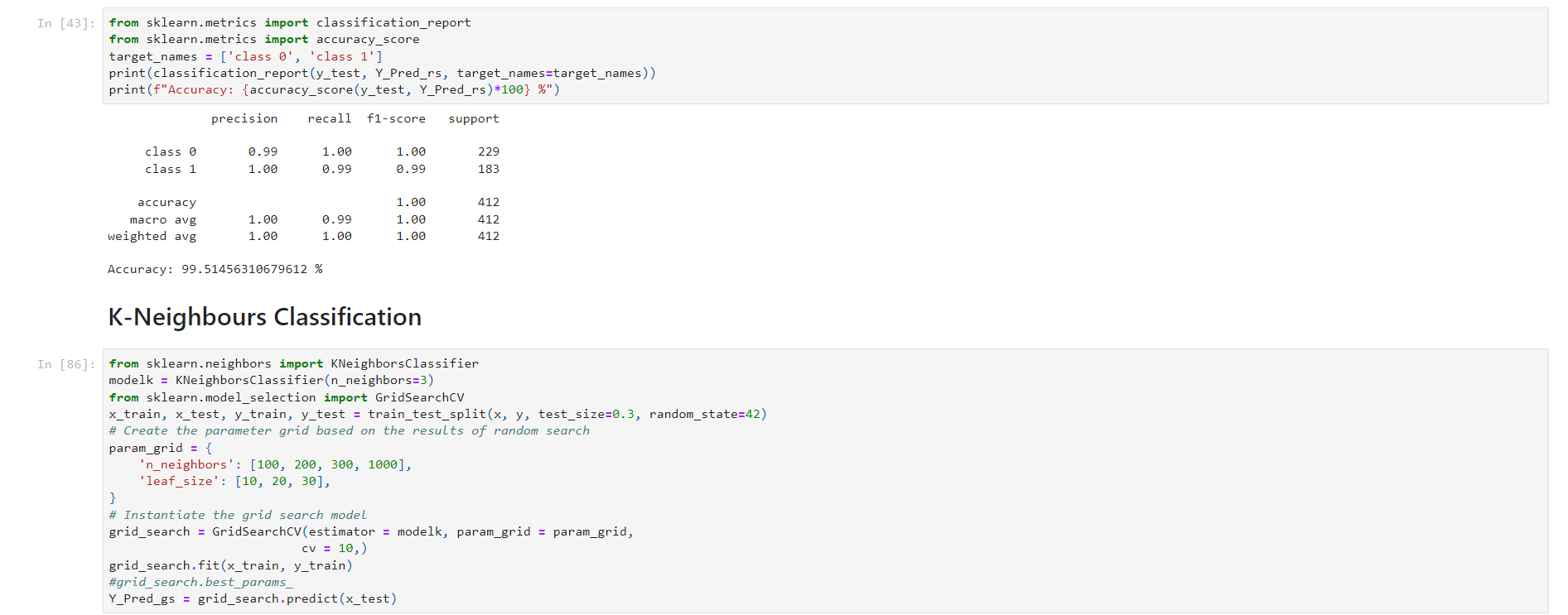
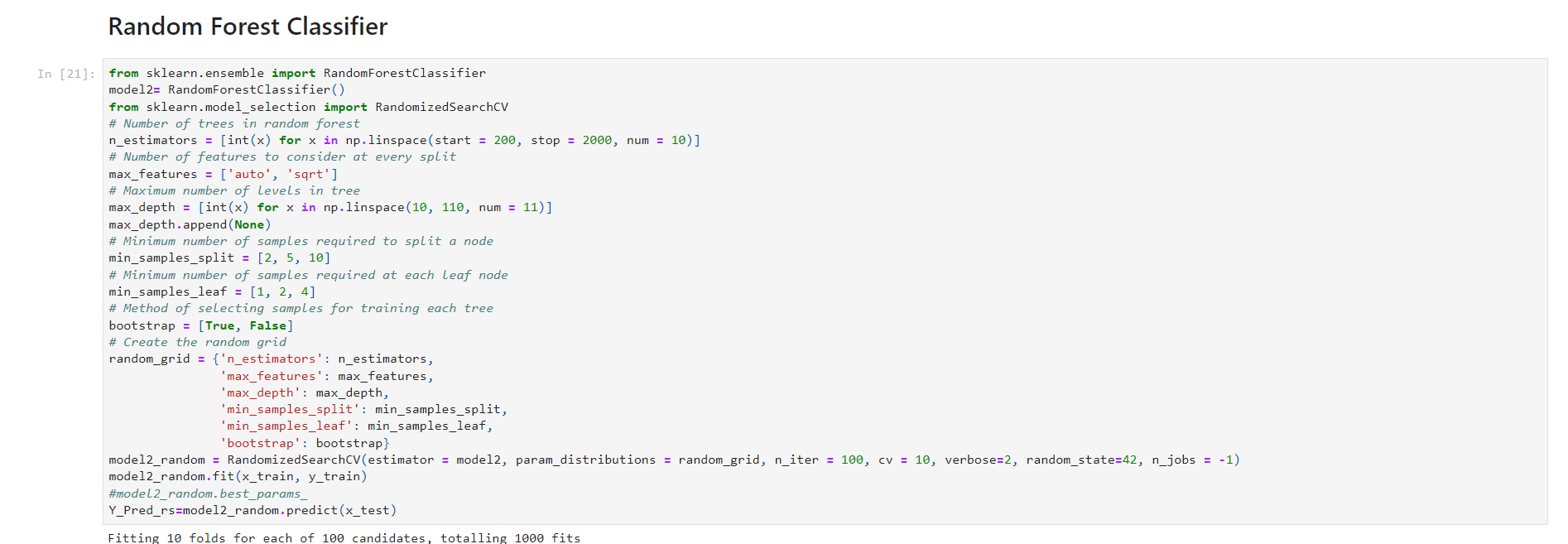


**Algorithms:**

The given problem statement is meant for classification and hence models like:

1. Random Forest Classifier
2. K-Neighbours Classifier
3. Decision Tree
4. Gaussian Naïve Bayes
5. SVM Classification

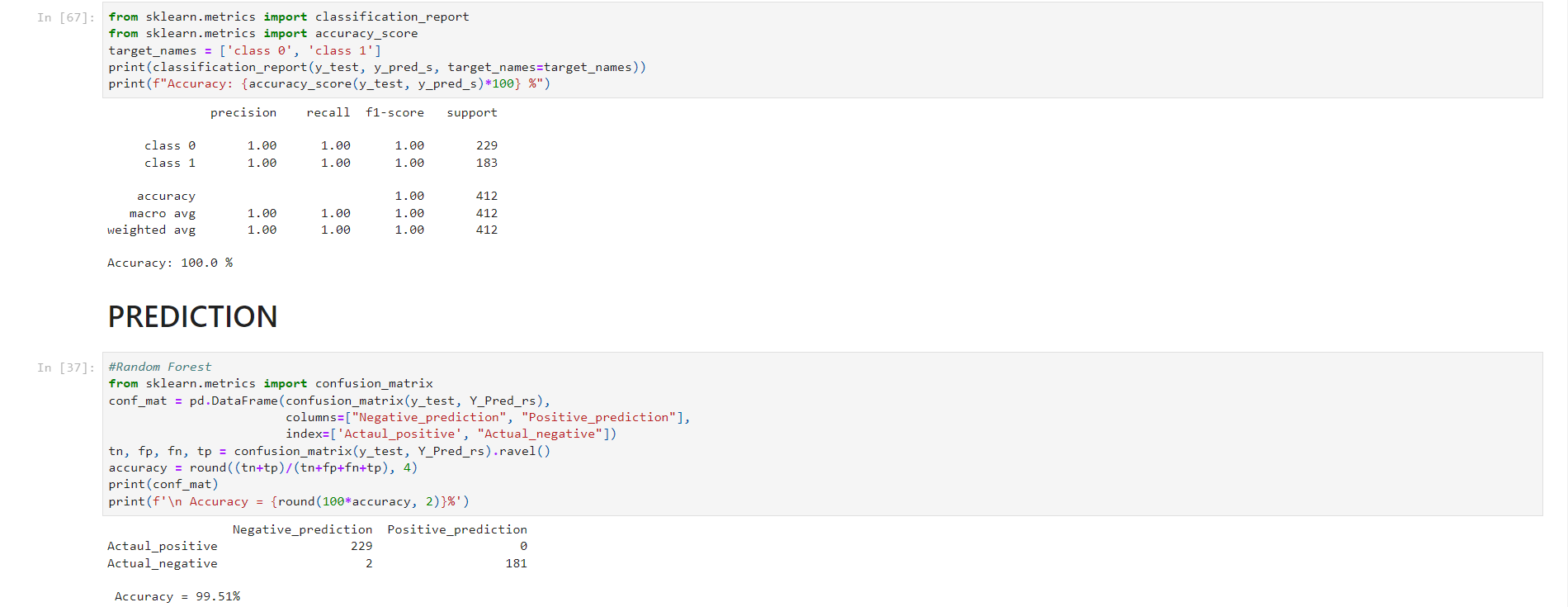
Have been implemented. All the metrics calculations have been done and compared for all the models.



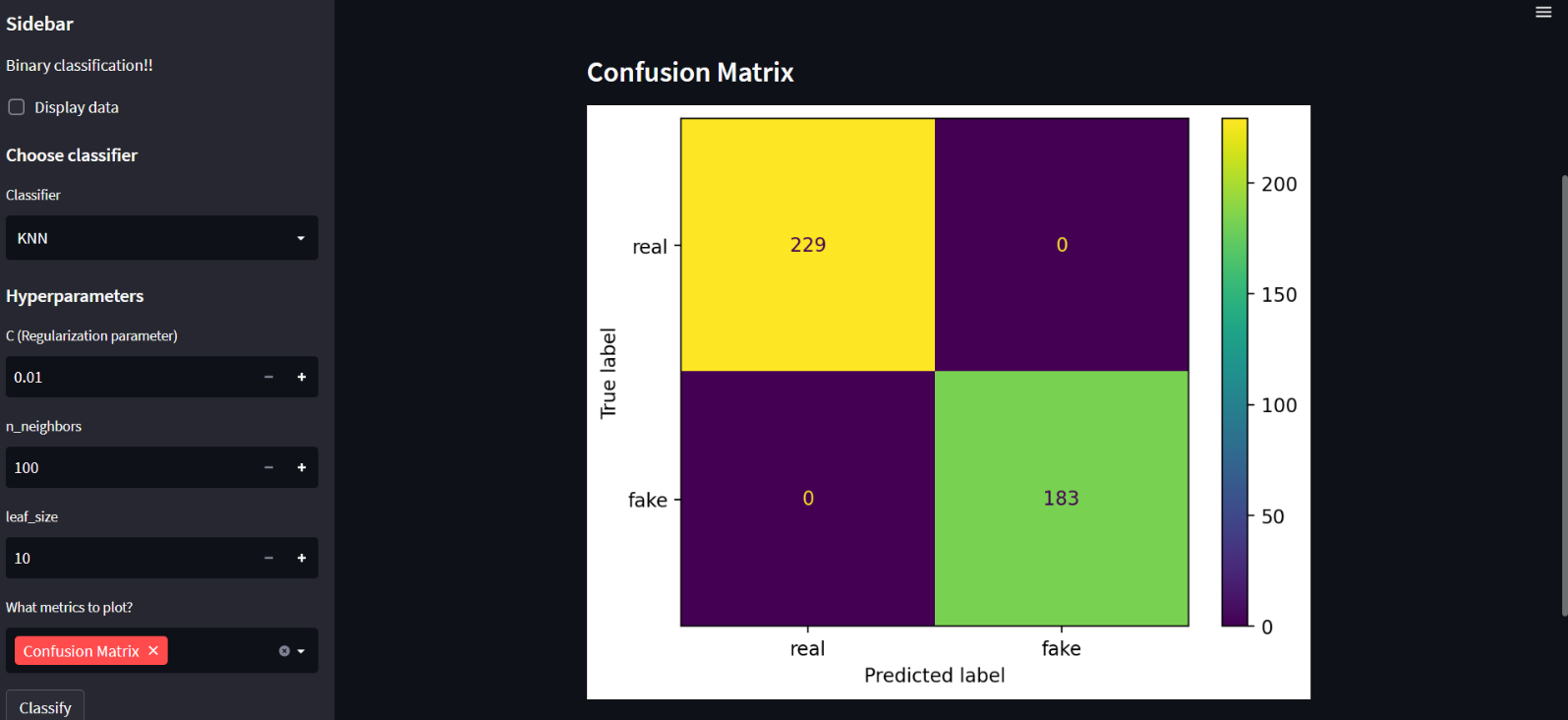
**Results/Predicitons and Discussion:**

All the metrics calculations have been performed on the models and compared. From the below predictions, KNN is the preferred model has there is less issue of overfitting being taken place. Most of the models have been optimized using RandomizedSearchCV and GridSearchCV.

Dash boarding has been done using the streamlit extension in python. The webapp screenshots have been displayed below. The dashboard provides 3 models where the users can give hyperparameter tuning values and get the required metric graphs.



Dashboarding





**Conclusion:**

From this project we were able to predict fake currency notes where K-neighbors Classifier gave proper accuracy keeping in mid the overfitting issue of the data from all the applied models. This project also gives insights about the data parameters being taken under consideration for predictions. Finally users can be able to view and visualize data and make changes to their model of choice and make predictions.

Fake currency is something which can tip the scales of a country’s economy from better to worse. People get exploited and leads to chaos and confusion. Hence, it becomes crucial to remove this dirt. Fake currency detection model can help predict the amount of cash being illegally deployed and also the amount of fake transactions being taking place in the society. Government can implement these models to detect the amount of fake money being transferred, track down the person responsible and take necessary actions and help improve the economy of the country.

**Future Work**:

More of the models for regression can be tried out and optimized to find the best r2 score from all the models. The model can be deployed into real world by using flash. Data interpretation can be made more user-friendly using streamlit application. There is always scope of improvement for any kind of machine learning. With experience on can truly gain a lot of knowledge and built good machine learning models.

**References:**

1. Scikit-learn documentation
2. Anaconda documentation
3. <https://stackoverflow.com/>
4. <https://docs.streamlit.io/>