1.Name

-Md. Faysal Numin

2.Student ID (ID that is related to the course)

-24

3.Project Title

-Wine Quality Prediction

4.Provide a brief description of your project (2-3 sentences).

-Mainly it predicts the quality of wine depends on the physicochemical properties of the wine.

5.What is the primary problem your project aims to solve?

-Solve the problem to know the wine quality depending on the physicochemical properties

6.Who are the stakeholders or beneficiaries of your project? Data Understanding

-Wine Exporters,Consumer, Wine makers

7.What dataset(s) did you use? Provide the source(s).

- Red wine quality Prediction
- https://archive.ics.uci.edu/dataset/186/wine+quality

8.How did you preprocess and clean the data?

-Handling Missing Values

-Handling Outliers

-Handling Duplicates

9.Did you face any missing data or outliers? How did you handle them?

-No missing data but outliers . I have seen the stripplot of total sulfer dioxide and got some values out of range so i fixed the data range for this feature 0 to 159

10.What exploratory data analysis (EDA) techniques did you apply?

-Heatmap, ProfileReport, catplot , pairplot

11.Did you create any new features? If yes, explain their significance.

-Yes in the quality feature the target value is from 3 to 8 to i have make it binary where if the quality is greater or equal than 5 then it is 1 or 0 because the there is imbalance in the quality dataset

12.Which machine learning models or techniques did you use? Why?

I have used 3 models

1.Linear Regression

2.Decision Tree

3.Random Forest

Because mainly it a binary classification

13.Describe your model training and validation approach (e.g., train-test split, cross-validation).

-i have split the data ,training data is 80% and test data is 20% and i have tested the validation mean square error in the train and test data

14.What performance metrics did you use to evaluate your model?

-Mean Square Error

15.What were the key challenges in model selection and training?

- I have faced the prediction error in linear regression model which is making errors on the best and poor quality wines.

-In the Decision tree that the training error is 0, while the test error is 0.58 so it is an overfitted model.

-In random forest model it has less error in the test data

16.What were your model's final results (accuracy, precision, recall, F1-score, RMSE, etc.)?

-accuracy-0.965625

-precision-0.9685534591194969

-recall-0.9967637540453075

-F1-score-0.9824561403508771

17.Did you deploy your model (e.g., Flask API, Streamlit app)? If yes, provide details.

-Yes

18.Provide a link to your GitHub repository Interpretation and Insights

-https://github.com/md-faysal-numin/DSP-Wine-Quality-Prediction

19.What key insights did you gain from your project?

-I got the insights of how to visualize the data so that i can know how the features are connected and are relevant to the final output also i have got the insights of different models how they work on this type of data

20.What are the strengths and limitations of your approach?

-I have tried 3 models so that i can find the good result and my limitation is i have decrease the quality unique data from 6 to only 2

21.If you had more time, what would you improve or add to this project?

- I would try other models and increase dataset