**Individual Reflective Report – Priya Arvind**

For this assignment, I and my colleague Ivan decided to adopt the Cross-Industry Standard Project for Data Mining (CRISP-DM). We agreed on a division of duties according to the different phases of the CRISP process. I was in charge of the Data understanding (EDA/PCA), Model Building phase for building the predictive model for the target feature-defaulter and contribute Conclusion/Next Steps with the team.

My contribution to the project is detailed as follows:

* Understand the dataset with the help of different visualization techniques. This included the following tasks:
* Data visualization techniques like pair-plot, histograms, heat map, Missingno, Correction Mapping and libraries were used.
* Applied PCA on the pre-processed data to understand the data distribution and correlation between the data points.
* Evaluate the data on Over-Sampling techniques and applied the Over-Sampling SMOTE technique on scaled Data points for Classification Model, since the target variables were highly imbalanced.
* Models building. This phase included the following tasks:
* Since we have 2 different predictions, I tried to apply different Models for the target value-defaulter and considered the best models based on the evaluation done by the team.
* Classification: I applied 2 different Classification models- Logistic Regression and Random Forest Model on the dataset. Then applied Gird SV to find the best hyper Parameter to apply the best score parameters to each classification Model.
* In the group report, I prepared the Introduction- Objectives, Data understanding-EDA, model building for defaulters (Logistic Regression and Random Forest) and equal team contribution in the conclusion section.
* In the slide deck, I prepared the Data understanding -EDA, Model building and equal team contribution in the conclusion section and Introduction section.
* In this project, I have developed a deeper understanding of EDA and how simple visualization techniques can be used to understand the data and learned how to use GitHub repository.
* Overall we did good teamwork and try to explore and help each other to develop this project better.

**Individual Reflective Report – Ivan Saravanja**

Working on this assignment helped me to get the better understanding of how machine learning methods work and to realise the importance of the data preparation process.

My role in this assignment involved finding appropriate dataset to use, finding, and loading required libraries, importing, and merging datasets, pre-processing the data (feature selection, removing NaN values, removing duplicates, converting negative to absolute values, and replacing categorical variables with numerical values), building two machine learning models and creating parts of the report and PowerPoint presentation.

Throughout the project, my primary focus was on data preparation and model building to gain insights into loan defaulters and contract status.

This reflective report aims to provide a simple and concise overview of everything I contributed during the assignment.

My first task was finding the right dataset that suited the project's objectives. It required thorough research and evaluation to ensure the data was relevant and would provide meaningful insights. This step emphasized the importance of data quality and its impact on the analysis. As dataset consisted out of two separated datasets, I had to merge them into one.

A screenshot of a computer

Description automatically generated with medium confidence

Data preparation was a crucial step in the project. This involved data exploration, dealing with missing values, removing the duplicates, converting negative values to absolute ones, and converting categorical variables into numerical ones. It taught me the significance of pre-processing data to ensure its integrity and reliability for modelling. I learned the importance of addressing missing values through strategies such as imputation or elimination, depending on the circumstances. I also used different plotting to spot easier features with missing values.

A screenshot of a computer

Description automatically generated with medium confidence  
A screenshot of a computer

Description automatically generated with low confidence

Additionally, removing duplicates ensured that each observation in the dataset was unique, eliminating any potential biases or inaccuracies.

A screenshot of a computer code

Description automatically generated with low confidence

As part of the data preparation process, I also executed techniques to handle negative values effectively. Converting negative values to absolute values helped maintain consistency and facilitated meaningful analysis. These steps were essential in ensuring the dataset was clean, reliable, and ready for further analysis.

A screenshot of a computer

Description automatically generated with low confidence

At the end, as a last part of the pre-processing the data, I replaced all the categorical variables with numerical values, to have dataset ready for the machine learning.

A screenshot of a computer

Description automatically generated with low confidence

With the pre-processed data, I began building predictive models to identify loan defaulters and predict contract status.

From my side, I did two models - namely the Decision Tree and KNN Classification models.

I used Decision Tree model on dataset, which was cleaned and pre-processed, however, we didn’t performed Scaling and PCA on it prior to building the model. Target variable for this model was column 'NAME\_CONTRACT\_STATUS'.

Also, I used ‘importance’ array to rank all the features based on their importance according to the RFE algorithm. By doing so, I was able to identify the most important features for further analysis and the accuracy of the model was almost 88%.

A picture containing text, screenshot, font, line

Description automatically generated

Second model I did was KNN Classification, which was the best performing model at the end.  
A picture containing screenshot, text, diagram, rectangle

Description automatically generated

Once we were done with our Jupyter notebook, I created a report template, explained the parts of the assignment I did in it and created my part of the PowerPoint presentation.

Collaboration and effective communication with my colleague played a big role in overcoming challenges and making the project successful.

To sum it up, this project gave me valuable insights into analysing loan applications, preparing data, and building predictive models. I learned the importance of focusing on data quality and implementing models effectively. This experience has deepened my understanding of the challenges faced in real-world data analysis.