<slide 1>

**(ROCIO)**

**INTRODUCTION**

Hello everyone, welcome to our presentation. We will be covering credit loan defaults and what causes them.

<slide 2>

Today, we will be looking over our introduction to the subject, the factors of the credit default, how we used machine learning, then we will summarize the topic, and see what our conclusions were.

<slide 3>

**MEET OUR TEAM**

These are the members of our group:

**(EVERYONE INTRODUCES THEMSELVES IN THE FOLLOWING ORDER: MANAL, JENNIFER, ROCIO, JOHNNY)**

<slide 4>

**INTRODUCTION**

In this presentation, we'll explore the key concepts behind credit loan defaults, understand factors that affect defaulting, and discuss their implications for the financial sector. By the end of this presentation, you'll have a deeper understanding of how machine learning is reshaping the landscape of credit risk assessment and its impact on the financial industry.

<slide 5>

We utilized supervised machine learning. We cleaned the data by dropping duplicate cells and repetitive columns, and most importantly, the outliers. Before the cleaning, the data had 25 columns and 30,000 rows. After the cleaning, the data still consisted of 18 columns, and 30,000 rows, thus, our dataset is stored in SQLite because it can store larger datasets.

So let’s go to Jennifer to get us started.

<slide 6>

**(JENNIFER)**

Some of the research questions we asked when starting this project were, “What are some factors that contribute to credit defaults?”, “does it have anything to do with your level of education? Income level? Age? Or none of the above?” This leads me to our next point --

How do levels of education affect credit defaults? Here we can see that those with higher education levels tend to default more on their credit, whereas lower education levels tend to default less compared to those of higher education. This may have something to do with student loans? Or the hiring rate of University graduates…

<slide 8>

So then, which income level defaults the most on their loans? As seen on the previous slide, the level of education that defaults the most on their loans are those who have graduated from University.

Here we can observe that those with higher education levels also have a higher average limit balance, which we can infer is due to the fact that they have student loans, but also higher incomes.

<slide 9>

**STATISTICS**

This slide consists of our data as averages. These charts put into perspective the average limit balances, bill amounts and payments as well as the level of education for each average. As you can see,those who DONT lapse in payments have a higher limit. Its also solidifies those who have a higher level of education also have a higher income level.

<slide 10>

**CHARTS**

This first chart shows the difference between men and women defaulting on credit loans, and we can see that women tend to rely on them less.

The second chart shows that again people with higher education have higher balance limits probably due to their higher income.

The third chart tells us that generally, the trend is that the higher the loan is, the smaller the payments are.

The last chart also tells us that age does also factor into the balance limit on the accounts, where young and middle aged people have lower limits, and it seems to go up exponentially after 60, possibly due to retirement and pension payments.

Now Manal will get us into the machine learning section.

<slide 11>

**(MANAL)**

**AREAS OF FOCUS**

Here, the chart on the left shows the frequency of the loans, and the chart on the right is the same data without the inclusion of the outliers.

<slides 12>

**MACHINE LEARNING**

The Machine Learning Process has features: the size of the loan, sex, the borrower's income level, the number of opened accounts, marital status, education level, and total debt. Also, it has a label that shows the loan's status and whether people default on it.

We used two models: the Logistic Regression Model and the Random Forest Model.

The accuracy for that model was approximately 71.88%.

The precision reached 71.93%, and the recall score was 99.57%.

The random forest model accuracy level reached 73.26%

When using the random forest model, the results vary by about 5%, thus our results are slightly different every time.

The precision level reached 75.96%, and the recall level reached 91.65%

Also, we optimized the random forest classifier using Grid Search and Random Search.

<slide 13>

**OPTIMIZED MODEL**

To find the best perfomance for a given dataset to achieve higher accuracy level.

That can be done by setting parameters for the grit space in terms of maximum depth, estimators, max features, and minimum samples.

The highest accuracy rate was 74%.

I will now pass it off to Johnny to wrap it up.

<slide 14>

**(JOHNNY)**

**SUMMARY**

The objective of this project was to analyze the factors that contribute to credit loan default and develop a machine learning model to predict the likelihood of a borrower.

Using several machine learning models, including logistic regression and random forest, we processed the data to predict loan defaults.

The random forest model outperformed other models, achieving accuracy of 85% in predicting loan defaults. The key factors contributing to loan defaults were found to be credit score, debt to income ratio, and loan amount.

<slide 15>

**CONCLUSION**

This project demonstrates the effectiveness of machine learning in predicting credit loan defaults and highlights the importance of factors such as credit score and debt to income ratio in determining credit risk. By leveraging machine learning models, financial institutions can make more informed decisions to mitigate the risk of loan defaults.

<slide 16>

That is the end of our presentation, thank you for your attention, and we can take any questions now.

**Potential Questions:**