# CREDIT RISK ASSESSMENT

Bassant Mohamed
Doaa Samir
Nureen Ehab
Zeinab Mohamed

# PROJECT OVERVIEW

The main objective is to accurately predict the creditworthiness of borrowers, distinguishing between those likely to default on loans or credit obligations and those who are expected to fulfill their repayment obligations. This classification project aims to develop robust predictive models that can effectively categorize borrowers into different risk categories based on their credit profiles and financial characteristics.

# INTRODUCTION

Credit risk assessment is crucial for financial institutions to prevent losses from loan defaults. It involves evaluating borrowers' creditworthiness to manage lending risks effectively. This assessment guides informed lending decisions, ensuring funds go to reliable borrowers. It also promotes responsible lending, efficient capital allocation, and compliance with regulations, enhancing the financial system's stability. Additionally, it builds trust among stakeholders, supporting the credibility of financial institutions. Overall, credit risk assessment safeguards lenders' interests, facilitates access to credit, and fosters economic growth and financial stability.

# PREPROCESSING

### Missing Values

We handled any missing values in the dataset and removed it using the dropna() function.

### Duplicates

We also handled any duplicate rows in the dataset and removed them, using the drop_duplicates() function.

### Data types

We changed some data type in the dataset with its suitable one.

### Descriptive statistics

We used the describe() function to summarize the statistics for numerical and categorical data.

# PREPROCESSING

### Balance

Check the balance of the target column 'loan_status' using value_counts() function.

### Encoding categorical features

Converting each category of a categorical feature into its corresponding expected value.

### Scaling the numerical features

Using MinMaxScaler() tool to rescale the features to a fixed range between [0, 1]. This can help reduce the variance and skewness of the data.

### Splitting the dataset

:

We split the dataset into 75% training and 25% testing

# PREPROCESSING

## Handling the imbalance

We handled the target column 'loan_status' by resampling using the over-sampling technique.

## Correlation

We analyzed the correlation between numerical features and the target variable 'loan_status'

## Relationships

Displaying the scatterplot for pairs of numerical variables with KDE plots along the diagonal, providing the additional insights into the distribution of each numerical variable

## Outliers

We identified the outliers in our dataset and removed them using IQR method.
We first identified the outliers in numerical columns by plotting a series of box plots, then, handled the outliers using IQR method

# BUILDING THE MODEL WITH TENSORFLOW

Train the model

Evaluate the model

Generate classification report

Plot confusion matrix

# EVALUATE THE MODEL

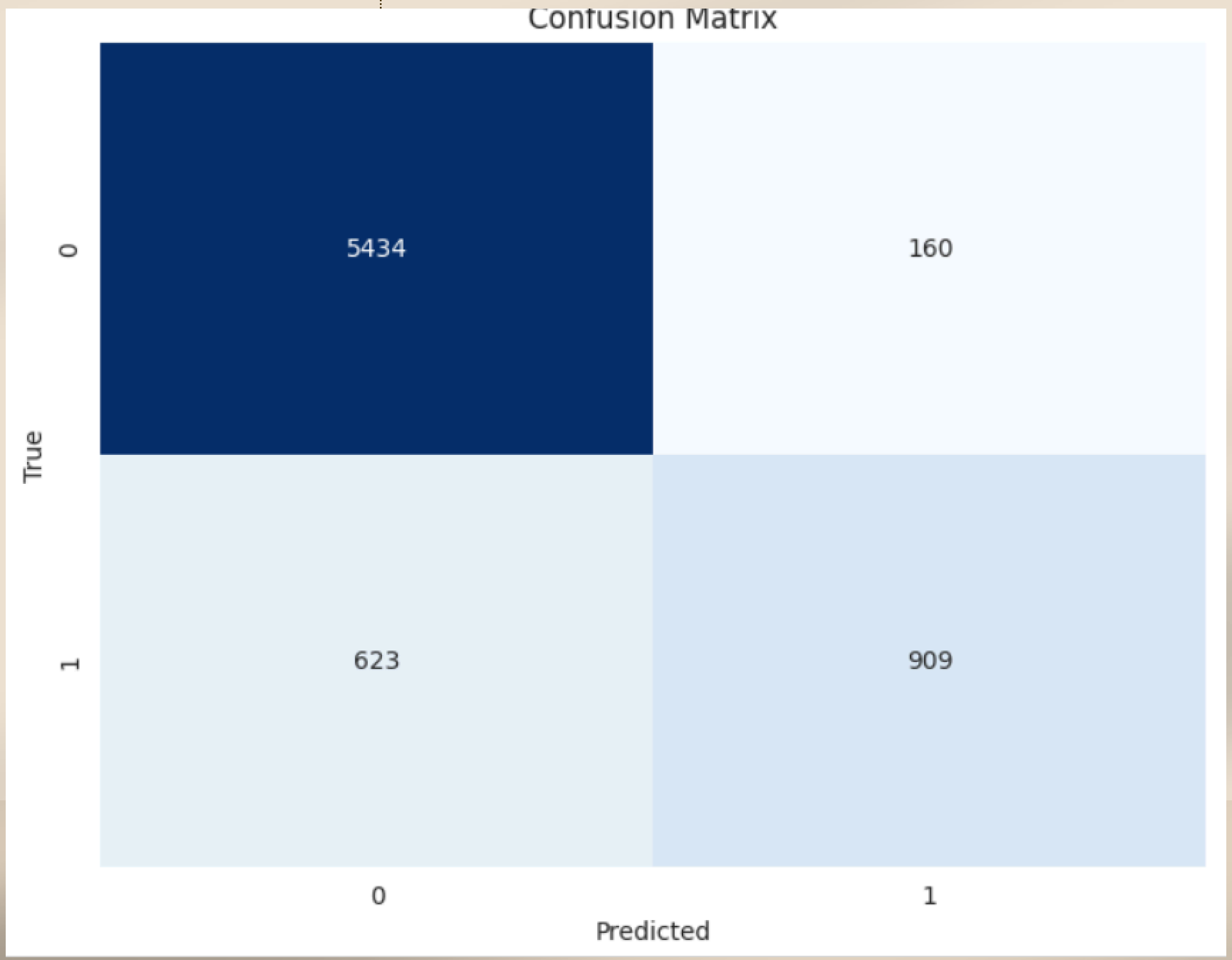
```
223/223 [==============================] - 0s 2ms/step - loss: 0.3265 - accuracy: 0.8901
Test Accuracy: 0.8901206851005554
```

# GENERATE CLASSIFICATION REPORT

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.90 | 0.97 | 0.93 | 5594 |
| 1.0 | 0.85 | 0.59 | 0.70 | 1532 |
| accuracy | | | 0.89 | 7126 |
| macro avg | 0.87 | 0.78 | 0.82 | 7126 |
| weighted avg | 0.89 | 0.89 | 0.88 | 7126 |

# PLOT CONFUSION MATRIX



# ENHANCING THE MODEL WITH TENSORFLOW

Padding

Max pooling

Global average pooling

Early stopping

# EARLY STOPPING

```
Epoch 1/50
134/134 [==============================] - 3s 24ms/step - loss: 0.2537 - accuracy: 0.9093 - val_loss: 0.2581 - val_accuracy: 0.9062
Epoch 2/50
134/134 [==============================] - 4s 26ms/step - loss: 0.2491 - accuracy: 0.9102 - val_loss: 0.2441 - val_accuracy: 0.9146
Epoch 3/50
134/134 [==============================] - 3s 21ms/step - loss: 0.2473 - accuracy: 0.9120 - val_loss: 0.2491 - val_accuracy: 0.9151
Epoch 4/50
134/134 [==============================] - 3s 22ms/step - loss: 0.2462 - accuracy: 0.9116 - val_loss: 0.2684 - val_accuracy: 0.9050
Epoch 5/50
134/134 [==============================] - 3s 22ms/step - loss: 0.2448 - accuracy: 0.9145 - val_loss: 0.2449 - val_accuracy: 0.9146
```

# CLASSIFICATION REPORT

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.92 | 0.98 | 0.95 | 5594 |
| 1.0 | 0.89 | 0.68 | 0.77 | 1532 |
| accuracy | | | 0.91 | 7126 |
| macro avg | 0.90 | 0.83 | 0.86 | 7126 |
| weighted avg | 0.91 | 0.91 | 0.91 | 7126 |

# SURVEY

## Paper

Dataset: "German Credit Risk" dataset

Model: ML models

Link of the paper:

https://publications.eai.eu/index.php/IoT/article/view/5376

## Our Team

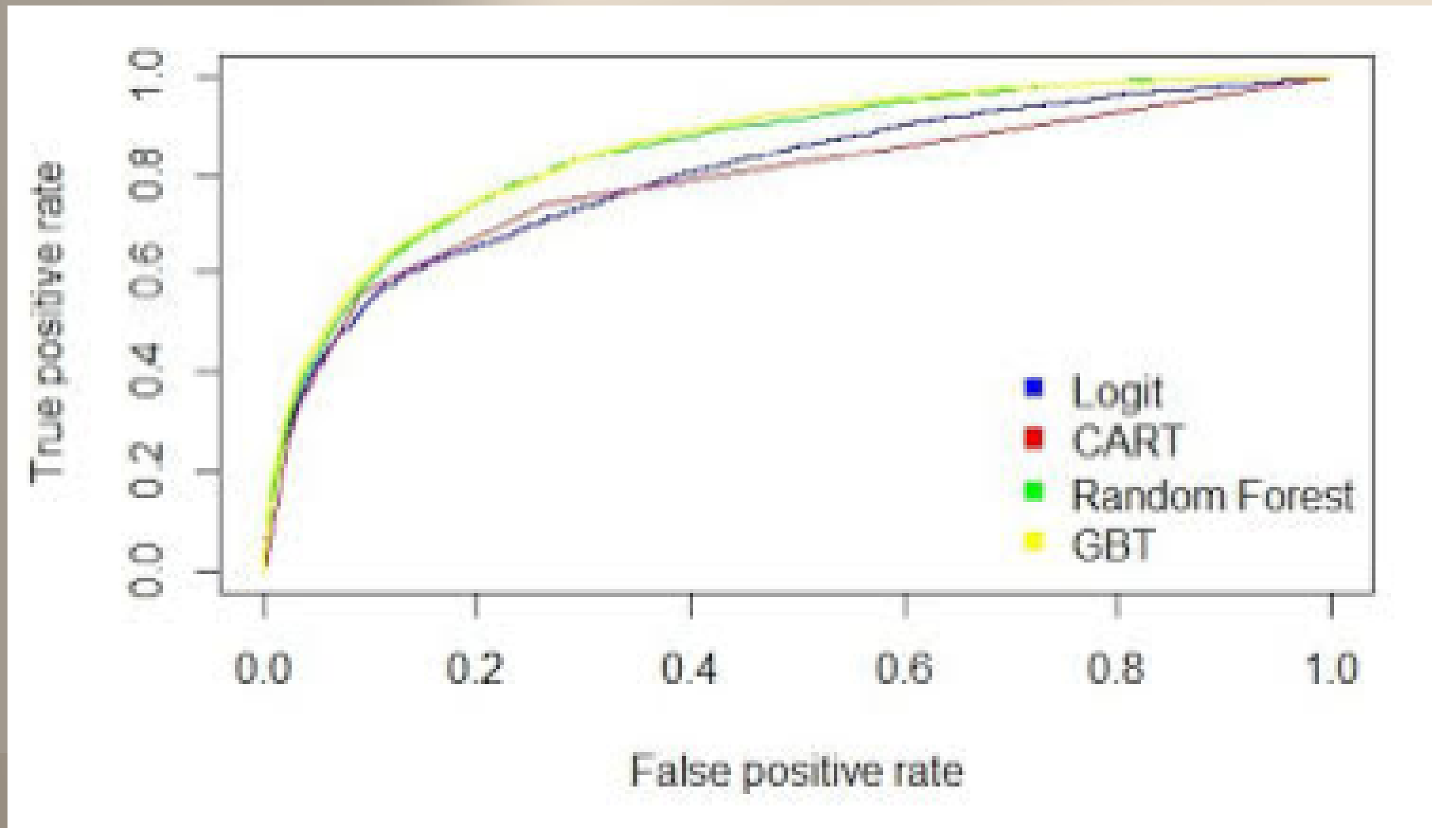
Dataset: "Credit Risk" dataset

Model: Neural Networks model

# SURVEY

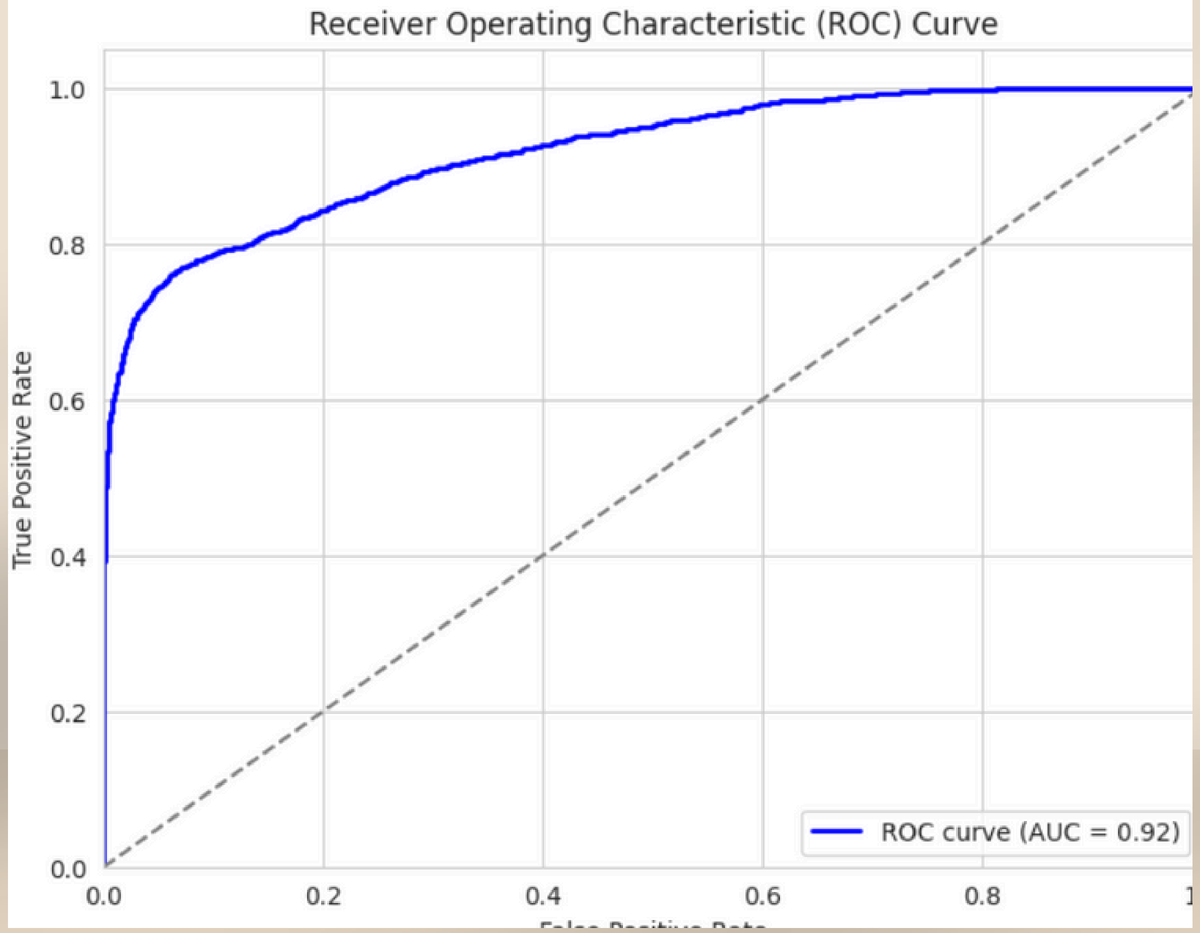
## COMPARING RESULTS WITH ROC CURVE:

### Paper:




Accuracy around 0.85

### Our Team:




Accuracy around 0.92

# CONCLUSION

Our project has successfully tackled the critical challenge of predicting borrower creditworthiness with precision and reliability. Through the development and refinement of robust predictive models, we have achieved a commendable accuracy level, as evidenced by an impressive Area Under the Curve (AUC) value of 0.92. This accomplishment underscores our ability to differentiate between borrowers likely to fulfill their repayment obligations and those at risk of defaulting on loans or credit obligations. By doing so, we empower financial institutions to make informed decisions, mitigating potential risks and optimizing lending strategies.

# THANK YOU