



#### SDAIA Academy T5C04 Bootcamps: Data Science Classification Module Project Presentation



#### **Predicting the Risk of Loan Default**

**Presented by Nada Rambu** 



### **Problem and Solution**

Project background and objectives

#### Objective

Predict whether a customer's loan would be approved or denied depending on their risk of defaulting

## Data Processing

Getting the data from the its source to be processed



#### **Data**







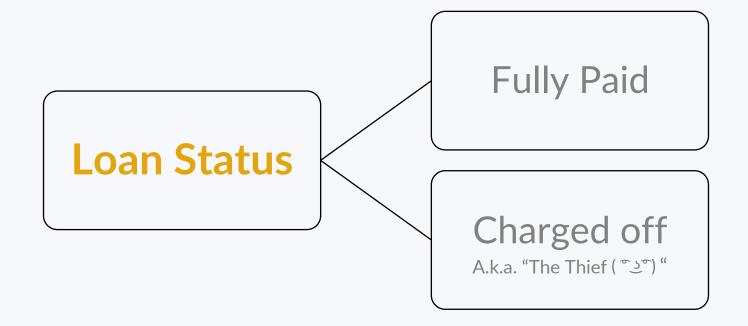
#### **Features**

Home Open Accounts
Home Ownership
Current Loan Amount
Bankruptcies Bankruptcies Annual Income

Customer ID

Loan ID Number of Credit Problems rs of Crears in current job Maximum Open Credit Score
Tax Liens

#### **Features**



#### **Preprocessing**



Handling Missing values



**Data Type Conversion** 



**Dataset Resampling** 



**Encoding Categorical Features** 



**Removing Outliers** 







#### Preprocessing

38560 rows × 17 columns

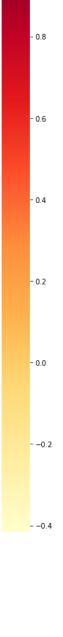
#### **Preprocessing**

38560 rows × 17 columns

### **Data Visualization**

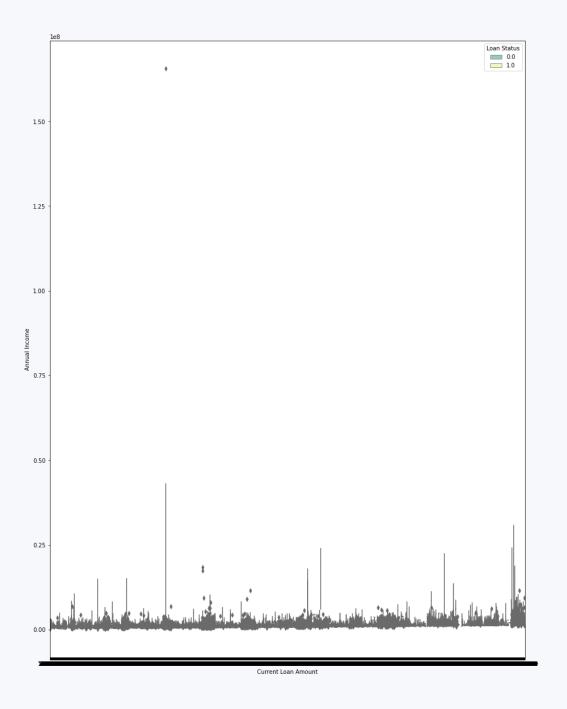
Gaining insights from the data

Loan Status -	1	0.041	0.11	-0.075	-0.053	0.0014	0.065	0.015	0.0045	-0.027	-0.02	0.011	0.0051	-0.017	-0.013	-0.0036	0.0098
Current Loan Amount	0.041	1	0.44	-0.21	0.27	0.094	-0.2	-0.13	0.44	0.16	-0.016	0.2	-0.071		0.043	-0.1	0.024
Term ·	0.11	0.44	1	-0.42	0.073	0.048	-0.12	-0.057	0.17	0.044	-0.0092	0.085	-0.029	0.11	0.011	-0.033	-0.00056
Credit Score	0.075	-0.21	-0.42	1	-0.023	-0.0093	-0.033	-0.077	-0.087	0.063	-0.038	-0.029	-0.032	-0.014	0.013	-0.02	-0.019
Annual Income	0.053	0.27	0.073	-0.023	1	0.04	-0.14	-0.0016		0.11	0.011	0.11	-0.0047	0.22	0.036	-0.034	0.038
Years in current job	0.0014	0.094	0.048	-0.0093	0.04	1	-0.19	-0.055	0.098	0.23	0.035	0.038	0.049	0.082	-0.00019	0.051	0.017
Home Ownership	0.065	-0.2	-0.12	-0.033	-0.14	-0.19	1	0.079	-0.2	-0.19	-0.014	-0.13	-0.0064	-0.16	-0.025	0.0006	-0.009
Purpose ·	0.015	-0.13	-0.057	-0.077	-0.0016	-0.055	0.079	1	-0.07	-0.022	-0.0064	-0.077	-0.0092	-0.043	-0.00049	-0.027	0.013
Monthly Debt	0.0045	0.44	0.17	-0.087		0.098	-0.2	-0.07	1	0.19	0.0069	0.41	-0.044	0.46	0.046	-0.072	0.026
Years of Credit History	-0.027	0.16	0.044	0.063	0.11	0.23	-0.19	-0.022	0.19	1	0.088	0.12	0.06	0.2	0.03	0.065	0.016
Months since last delinquent	0.02	-0.016	-0.0092	-0.038	0.011	0.035	-0.014	-0.0064	0.0069	0.088	1	0.041	0.077	-0.07	-0.013	0.051	0.031
Number of Open Accounts	0.011	0.2	0.085	-0.029	0.11	0.038	-0.13	-0.077	0.41	0.12	0.041	1	-0.014	0.22	0.033	-0.021	0.0054
Number of Credit Problems	0.0051	-0.071	-0.029	-0.032	-0.0047	0.049	-0.0064	-0.0092	-0.044	0.06	0.077	-0.014	1	-0.1	-0.012	0.74	0.59
Current Credit Balance	-0.017		0.11	-0.014	0.22	0.082	-0.16	-0.043	0.46	0.2	-0.07	0.22	-0.1	1	0.15	-0.11	-0.012
Maximum Open Credit	-0.013	0.043	0.011	0.013	0.036	-0.00019	-0.025	-0.00049	0.046	0.03	-0.013	0.033	-0.012	0.15	1	-0.014	-0.00023
Bankruptcies ·	0.0036	-0.1	-0.033	-0.02	-0.034	0.051	0.0006	-0.027	-0.072	0.065	0.051	-0.021	0.74	-0.11	-0.014	1	0.04
Tax Liens	0.0098	0.024	-0.00056	-0.019	0.038	0.017	-0.009	0.013	0.026	0.016	0.031	0.0054	0.59	-0.012	-0.00023	0.04	1
	Loan Status –	Current Loan Amount –	- Erm -	Credit Score -	Annual Income -	Mears in current job -	Home Ownership -	Purpose -	Monthly Debt -	Years of Credit History –	Months since last delinquent –	Number of Open Accounts -	Number of Credit Problems -	Current Credit Balance -	Maximum Open Credit –	Bankruptcies –	Tax Liens –









## Model Development

Train and test the model to estimate results

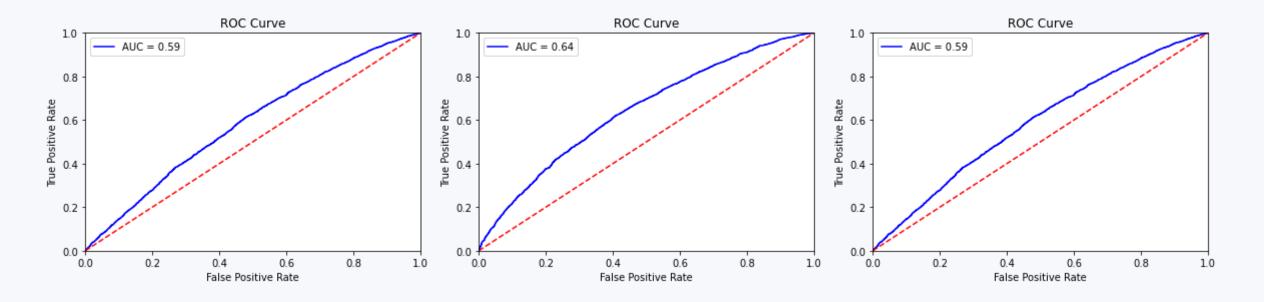
#### Models

Algorithm	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.57	0.58	0.57	0.56
Random Forest	0.60	0.64	0.62	0.60
<b>Gradient Boosting</b>	0.56	0.60	0.58	0.56

#### Models

Algorithm	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.57	0.58	0.57	0.56
Random Forest	0.60	0.64	0.62	0.60
<b>Gradient Boosting</b>	0.56	0.60	0.58	0.56

#### Models



Logistic Regression

Random Forest

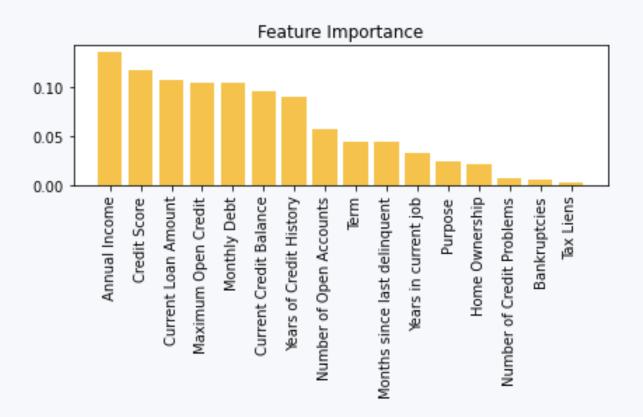
**Gradient Boosting** 



# (05) Conclusion

Some recommendation to solve the problem

#### **Insights**



#### According to the bar chart, these features:

- Annual Income
- 2. Credit Score
- 3. Current Loan Amount
- 4. Monthly Debt
- 5. Current Credit History

Are the most important features in explaining the risk of defaulting.







Thank You!