id/x partners

Credit Risk Analytics in Financial Services Firm Study Case: ID/X Partners

Written by Novrizal Roynanda

id/x partners

Table of Contents

id/x partners

Problem Research

Problem Research

Business Understanding

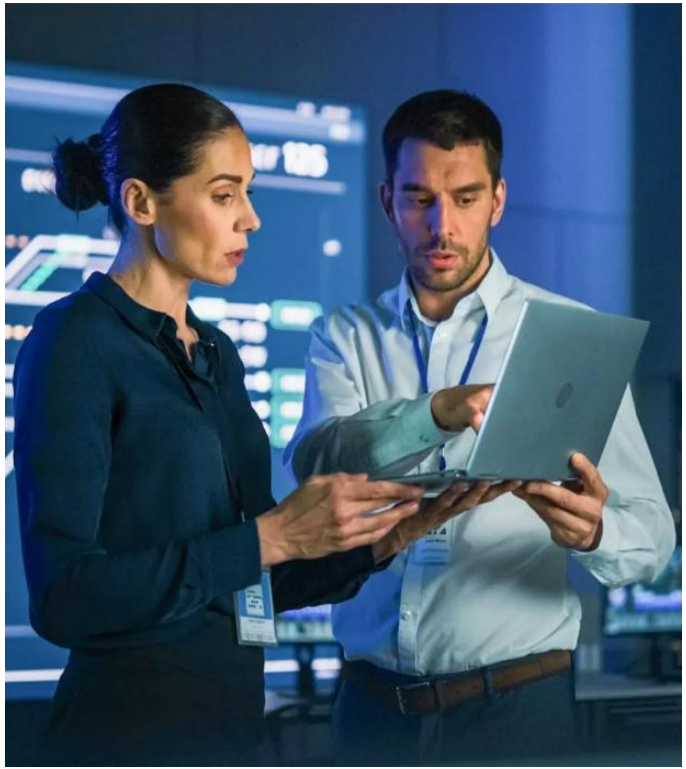
The data contains the information about past loans of applicants and whether they labeled as a good risk or bad risk.

Problem Statement

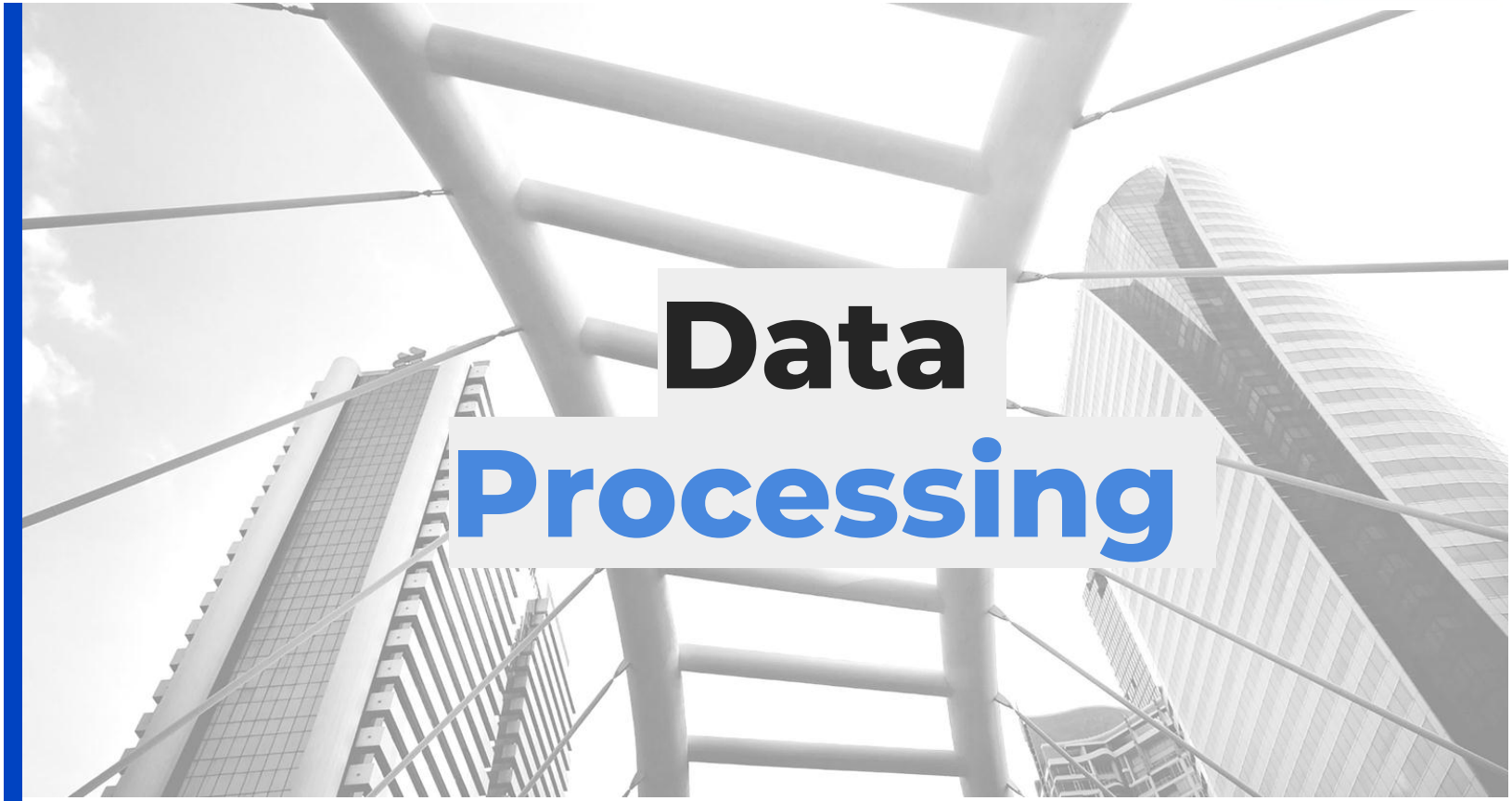
Lending loans to bad risk applicants is the largest source of financial loss. Credit loss is the amount of money lost by the lender when the applicant refuses to pay or runs away with the money owed.

Business Objectives

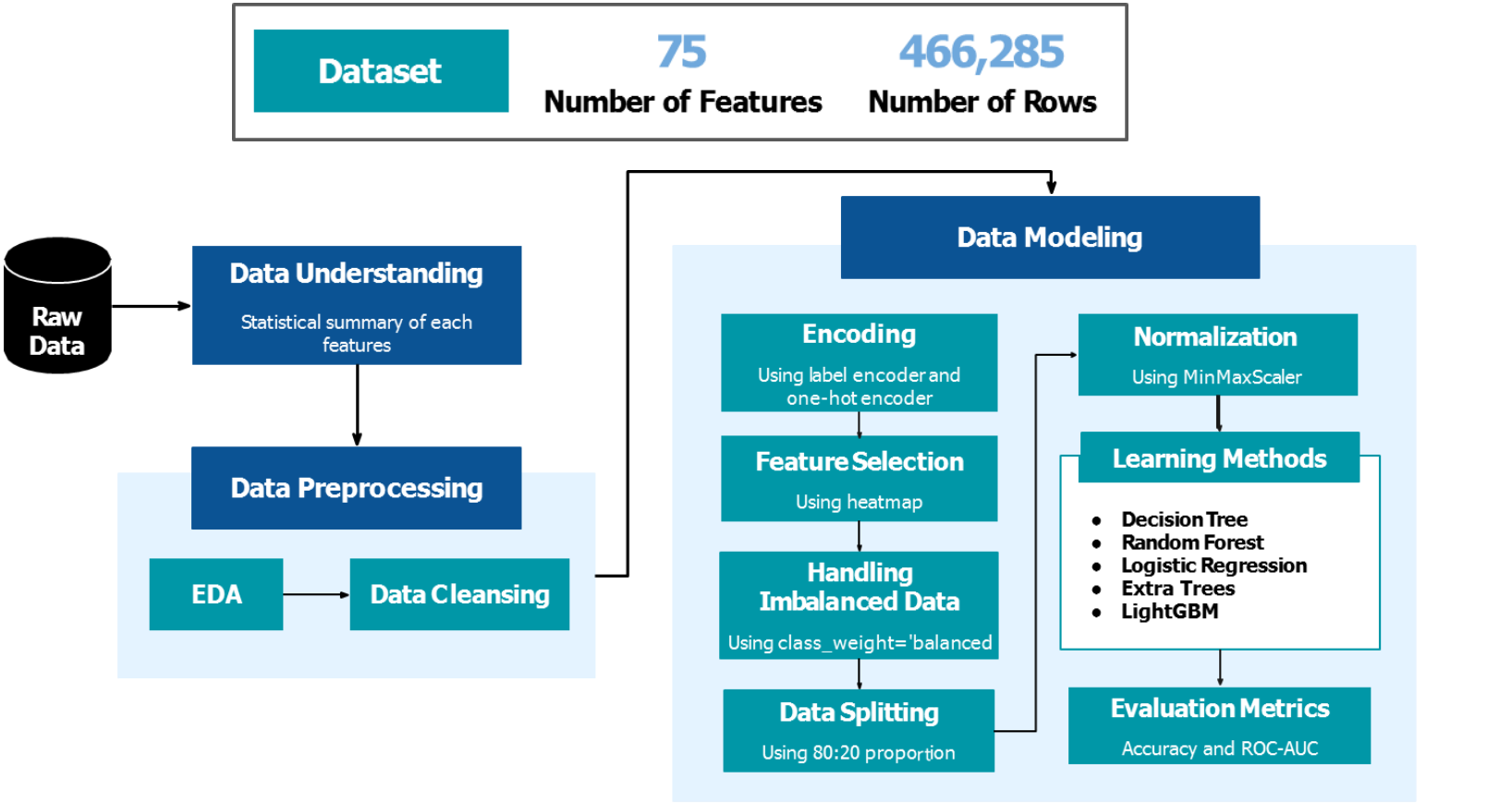
Identify patterns that indicate if a person is unlikely to repay the loan or labeled as a bad risk and implement machine learning algorithms to automatically predict whether the loan application submitted by the applicant will labeled as a bad risk or not.



id/x partners



Data Processing



id/x partners

id/x partners

Business Insights

id/x partners

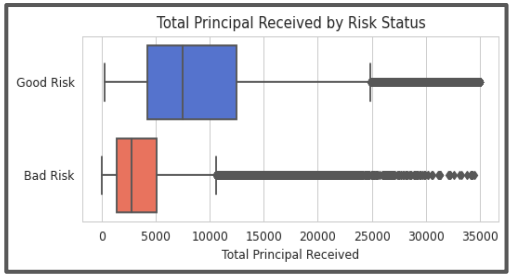
Total Loss Based on Loan Status

Loan Status	Total Loss	% Total Loss	Total Applicant	Average Loss
Charged Off	$574,356,330	83.49%	43,236	$13,284
Late	$102,293,296	14.87%	8,118	$12,600
Default	$11,299,446	1,64%	832	$13,581

The loan status of **Charged Off** is the biggest source of loss (**83%**) for the company with a total loss of **574 million** from **43,236 applicants**.

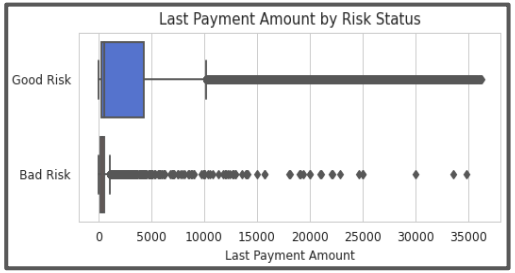
id/x partners

Business Insights



Total Principal Received by Risk Status

Bad risk loans tend to have lower principal received. It means applicants with low principal received are unlikely to repay the loan.



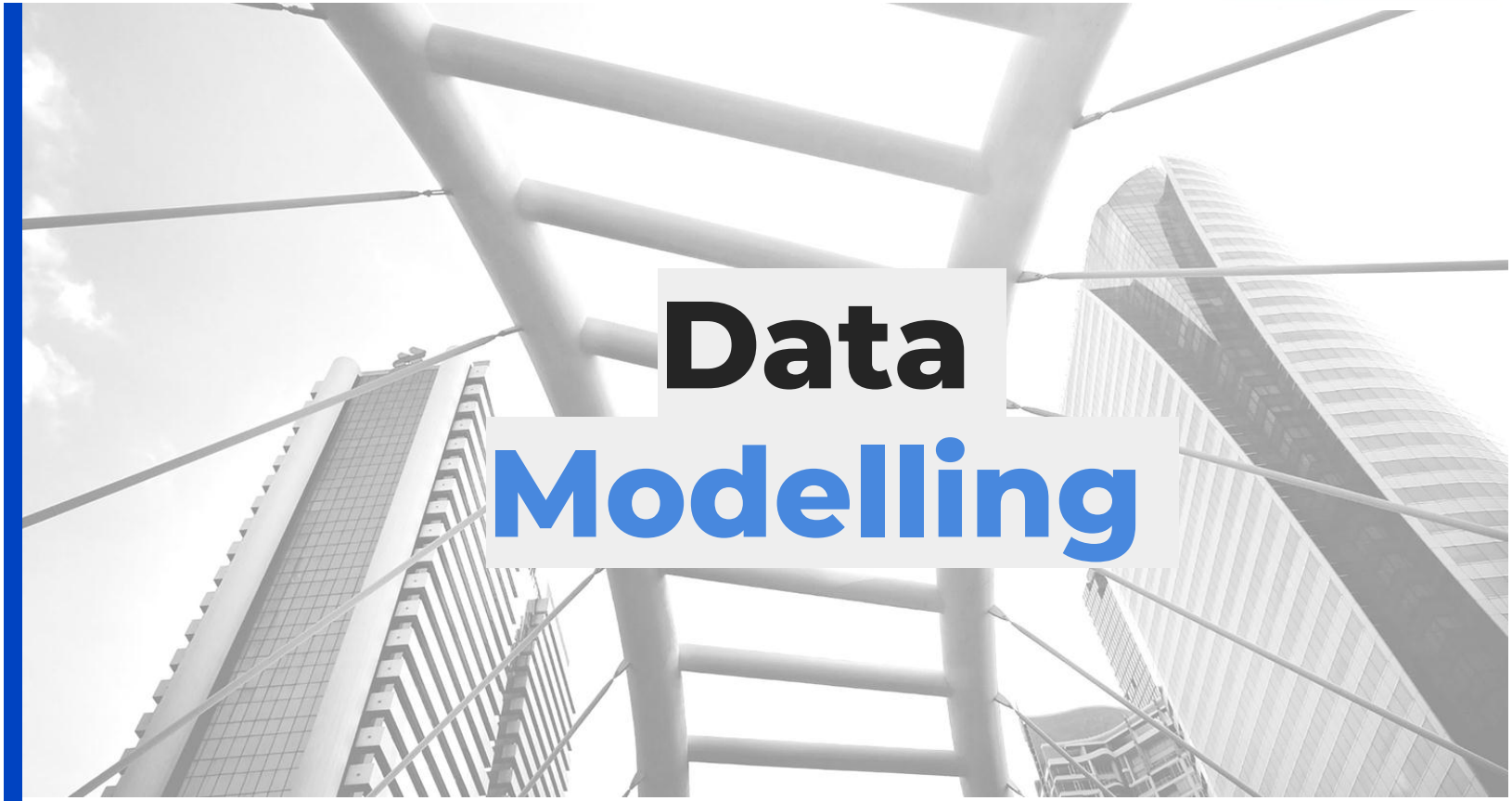
Last Payment Amount by Risk Status

Bad risk loans tend to have a lower amount of last payment. It means applicants with a low last payment amount are unlikely to repay the loan.

Interest Rate by Risk Status

Bad risk loans tend to have higher interest rates. It means applicants with the high interest rate have a high chance of not being able to repay the loan.

id/x partners

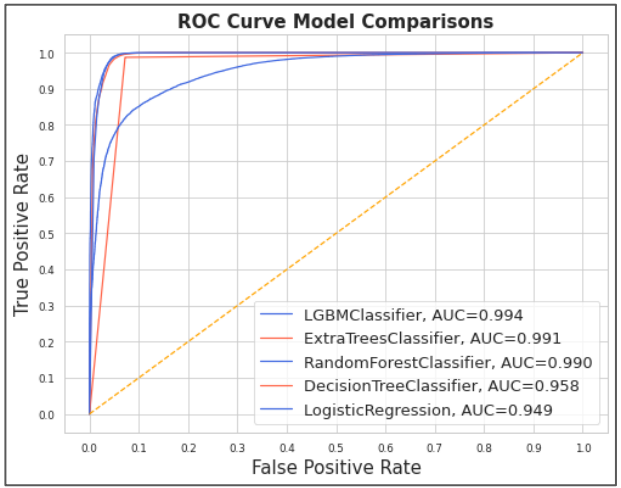


Data Modelling

id/x partners

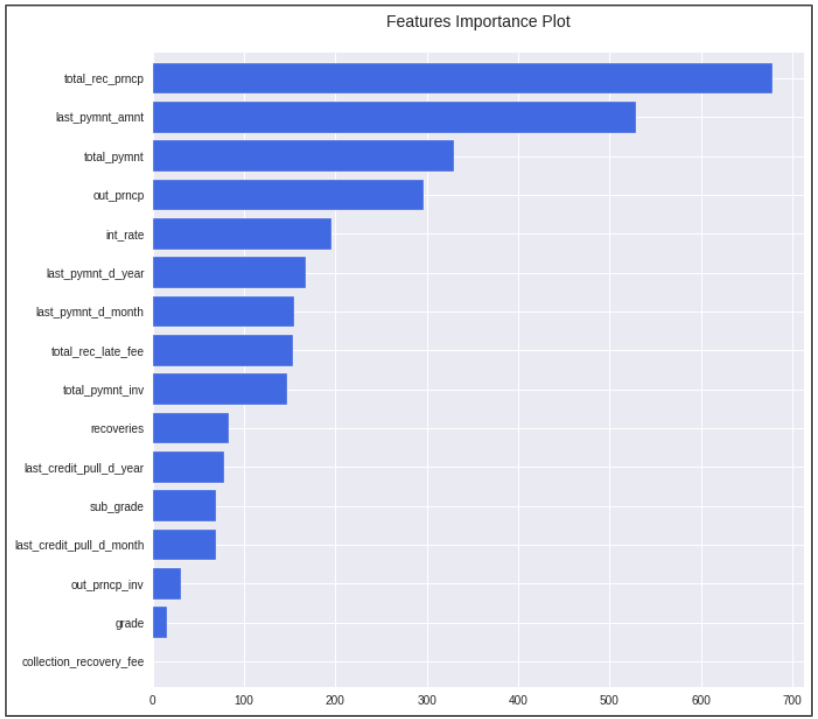
Model Results

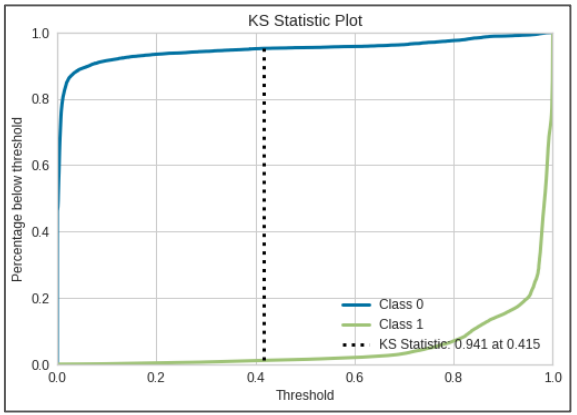
Algorithms	Training Accuracy Score	Testing Accuracy Score	Error Margin
Decision Tree	99.93%	97.99%	1.94%
Random Forest	99.95%	99.03%	0.92%
Logistic Regression	87.26%	85.16%	2.10%
Extra Trees Classifier	99.93%	98.86%	1.07%
LightGBM Classifier	98.49%	98.20%	0.20%



- The best model to predict the risk status of loan applications is **LightGBM Classifier**.
- The difference in accuracy between training and testing sets on the random forest model is smaller than in the **LightGBM Classifier** model.
- Although, **Random Forest** has the highest accuracy scores, but LightGBM Classifier has the lowest error margin and has the highest ROC-AUC scores with **0.994**.

Model Evaluation:
LightGBM Classifier





- The higher KS value the higher significant good performance model. LightGBM Classifier KS value is **0.941**.
- Top 3 important features is **total principal received**, **last payment amount**, and **total payment**.

id/x partners

Business Recommendation

id/x partners

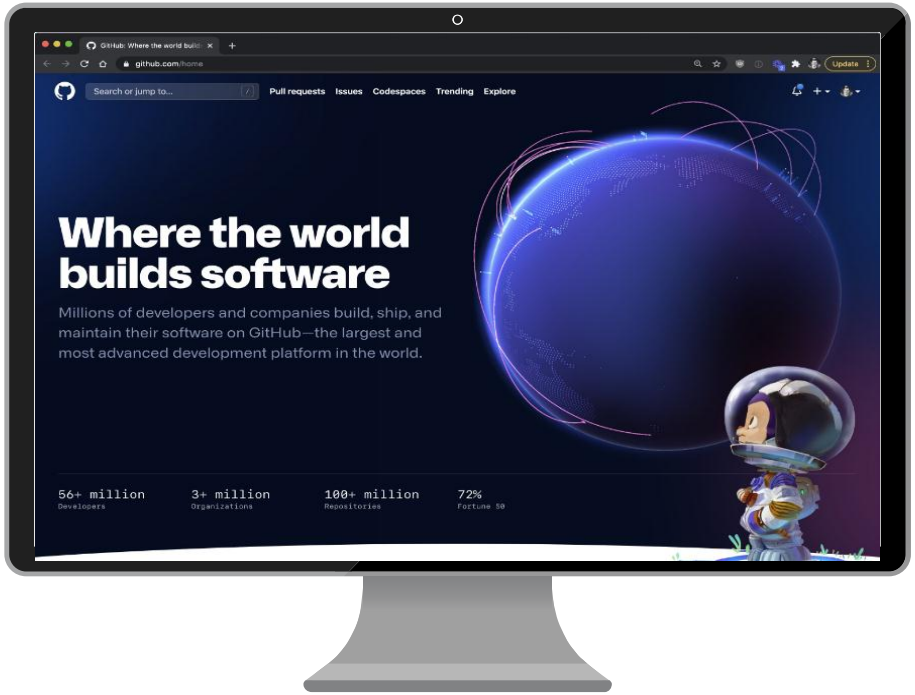
Business Recommendation

Monitoring Evaluation

Total principal received, **last payment amount**, and **total payment** is the most important features to identify whether the applicant has the possibility of not repaying the loan. The company needs to monitor these indicators to reduce the risk of loss.

Risk Assessment

If there are applicants with those indicators, then the company can take action such as rejecting their loan, reducing the amount of the loan, or lending at a higher interest rate to avoid and reduce the total loss suffered by the company.



Visit my github!

You can see the entire project documentation here from my github @novrizalrnd