

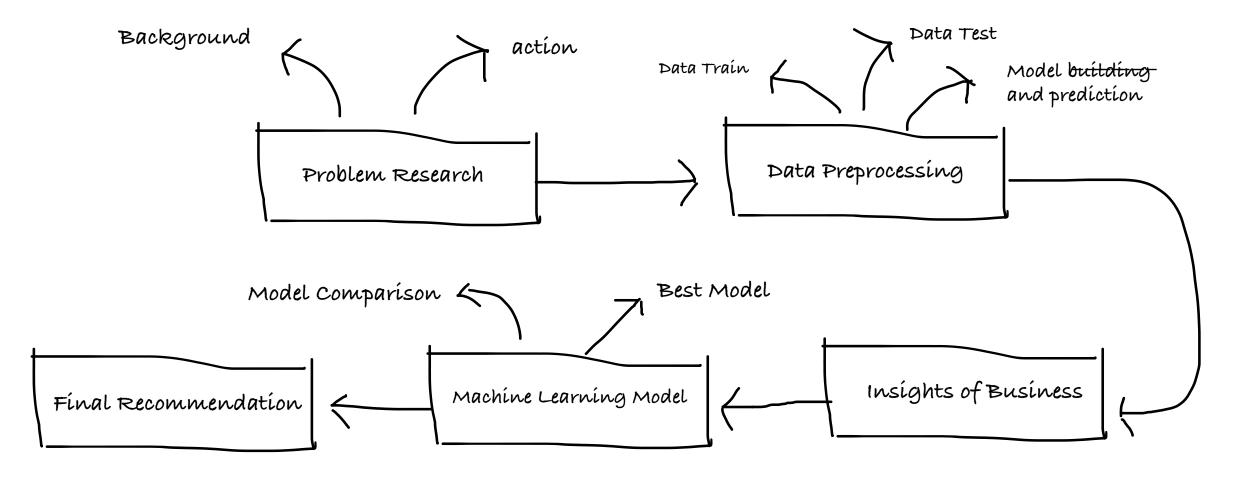


Home Credit Machine Learning Scorecard Model

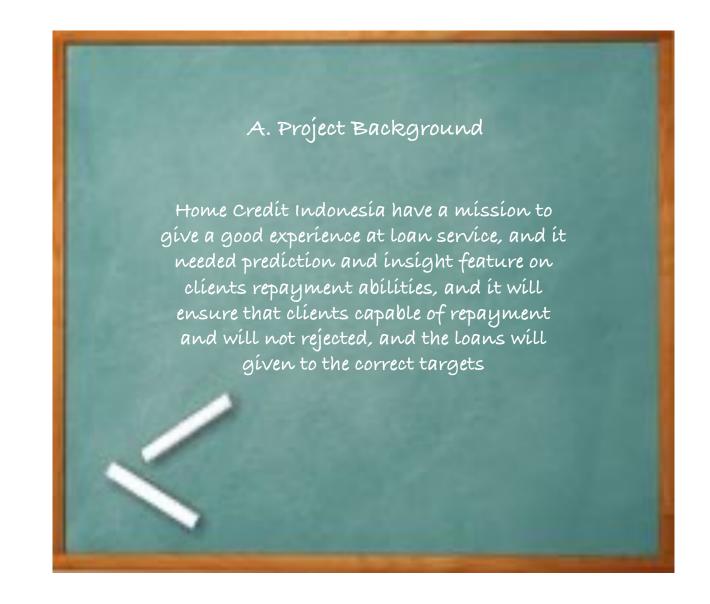
Presented by: Faurel Gema



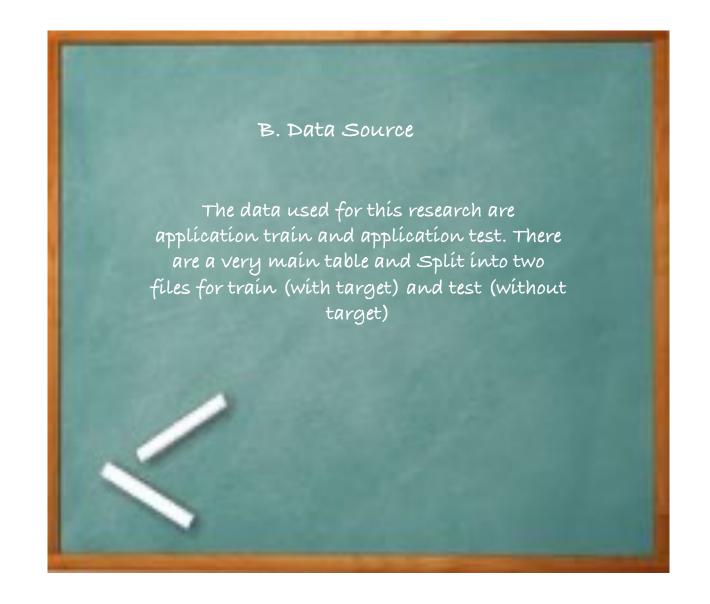
Flow of Content



- · A. Project background
- · B. Data Source
- · c. Objectivity
- D. Action



- · A. Project background
- · B. Data Source
- · c. Objectivity
- D. Action



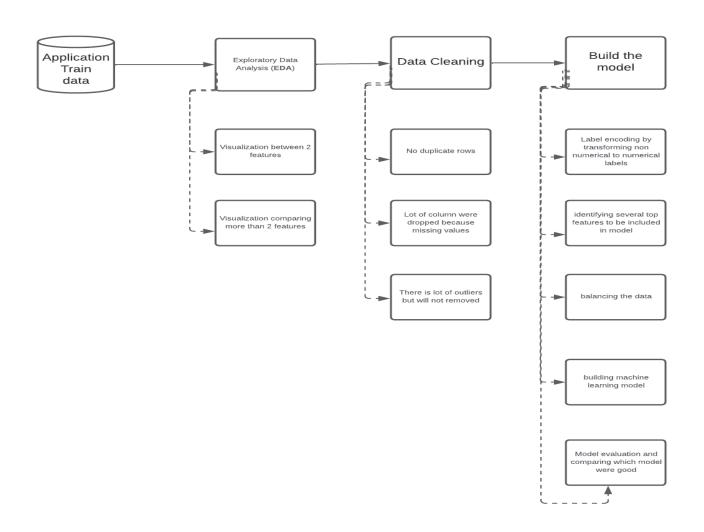
- · A. Project background
- · B. Data Source
- · c. Objectivity
- D. Action



- · A. Project background
- · B. Data Source
- · c. Objectivity
- D. Action



2. Data Preprocessing (Application Train Data)

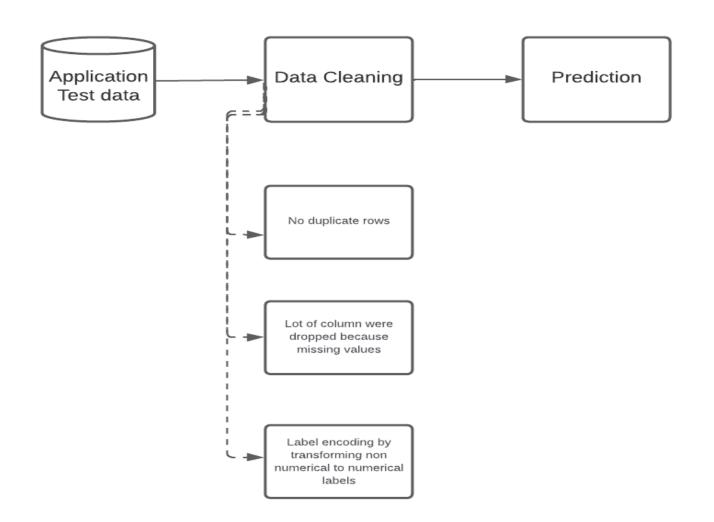


Application Train Data

> 122 Number of Columns

307,511 Number of Rows

2. Data Preprocessing (Application Test Data)



Application Train Data

> 121 Number of Columns

48,744 Number of Rows



How's the Insight of the Business?



ALL VIDEO

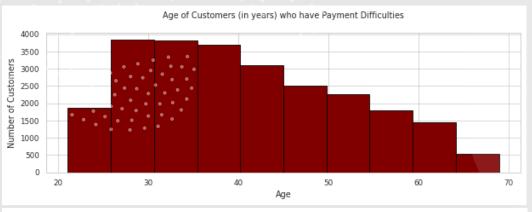
PICTURE

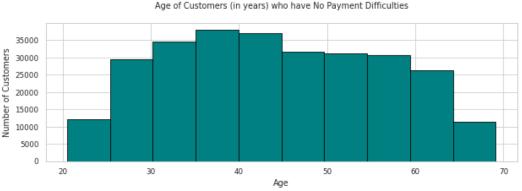
MAPS

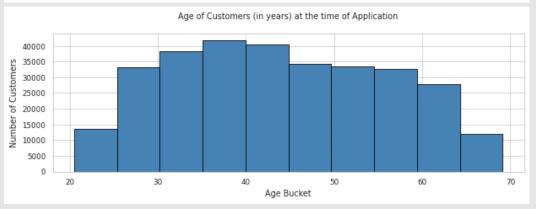
-Goodluck for the result.

3. Insight of Business

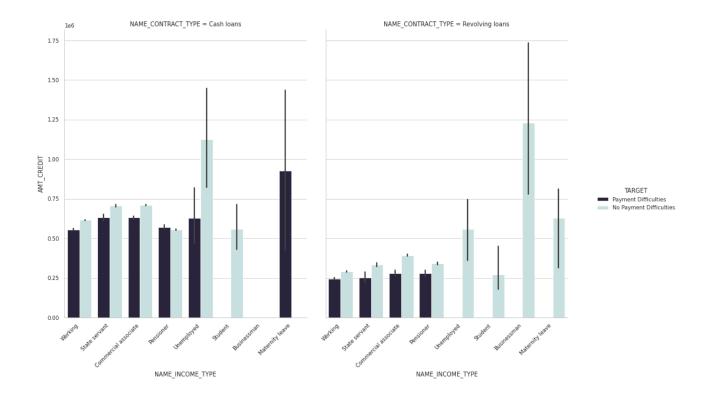
- 1. 35-40 years old customers are the most number of customers who apply for loan
- 2. <25 and >65 years old of group were the lowest apply for loan
 - 3. 25-35 years old group have a payment difficulties issues
- 4. 35-45 years old group have no payment difficulties

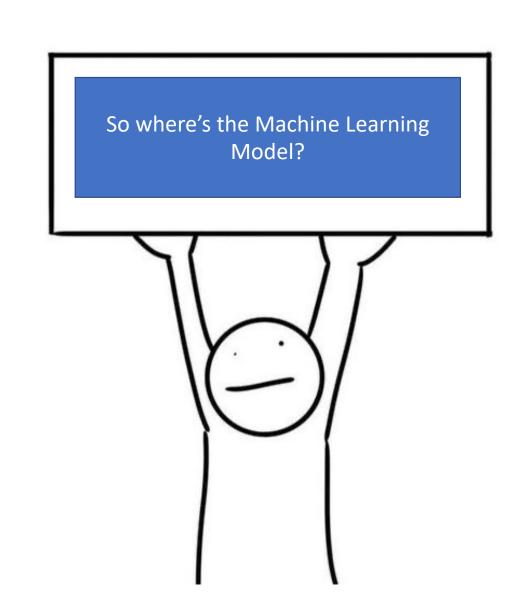






- All student have no difficulty to repaying the loans, both cash or revolving loan (low to medium credit)
- For the unemployed customers with cash loans, more than 50% have medium amounts kind of repaying loans issue. But there is no difficulty at revolving loans.
- Income type of maternity leave with cash loans, all customers have problems repaying loans for medium credit, but revolving loans have no difficulty repaying loans





4. Machine Learning model

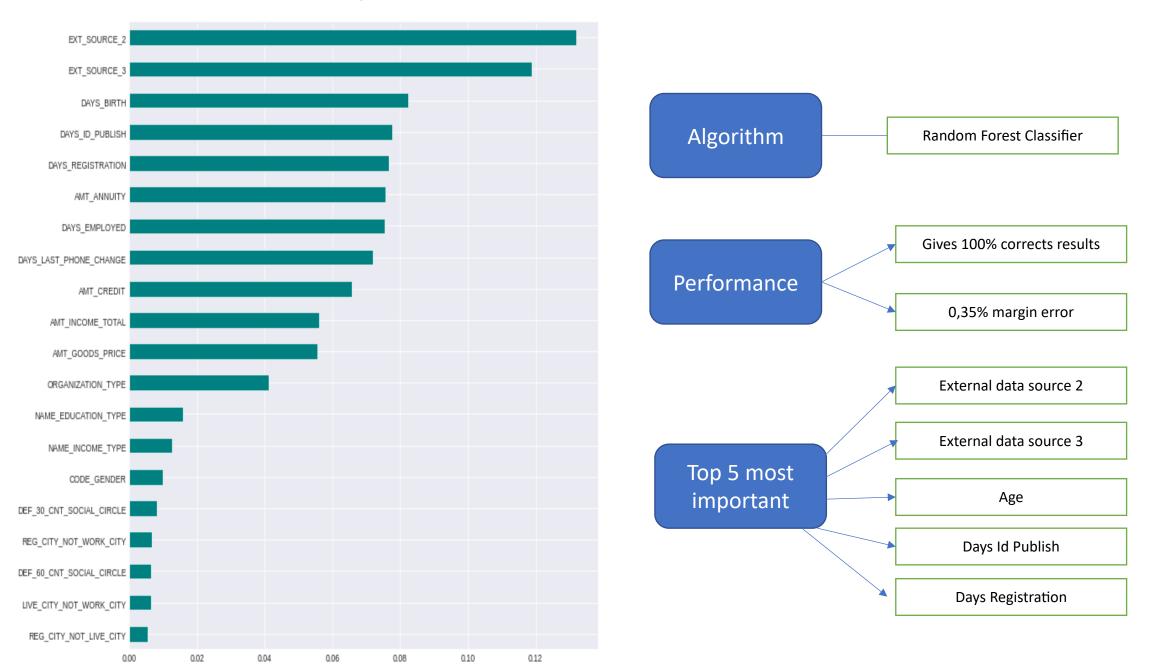
Model Comparison

Models	Training	Accuracy	Score	Testing	Accuracy	Score	ROC	Score
--------	----------	----------	-------	---------	----------	-------	-----	-------

Random Forest	1.000000	0.996500	0.996500
Logistic Regression	0.671600	0.672900	0.672900
Gaussian Naive Bayes	0.602400	0.603900	0.604000

Random Forest model is an excellent model which have no underfitting and overfitting. So it's choosen as the best one for customer repayment abilities predicition.

Features Importance Plot



5. Business Recommendation (List of Case)

- 1. <u>Customers with income type student is capable to repaying loans bot cash or revolving loan (100% approved) but only came 0.005% the application from student</u>
- 2. <u>Customers with income type accountant is capable to repaying loans bot cash or revolving loan (95% approved) but only came 3,19% the application from student and the rest of high skill tech staff and manager too.</u>
- 3. <u>Maternity leaves and cash loans type of customers have 100% rejected</u> applications, but maternity leaves with revolving loans approved
- 4. <u>Unemployed customers 50% have problem of repaying loans if using cash loans, but able to repaying if using revolving loans</u>



