Data Science Final Project
Job Connector

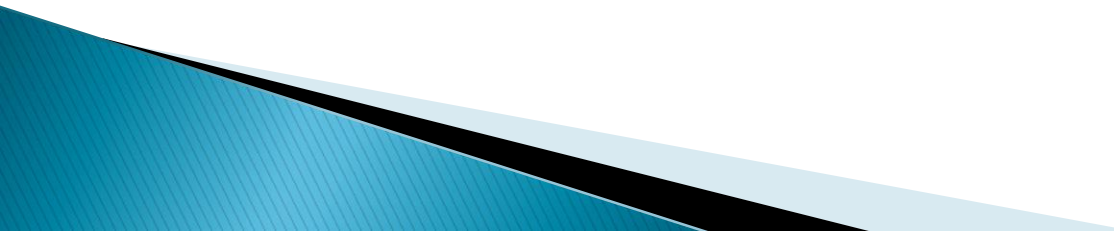Study Case : Bank Marketing

By: Muhammad Rizqa Aulia

Source Data : https://www.openml.org/d/1461

Goals:

To predict if the client will subscribe a term deposit or not.

To Do List:

- Which column will be changed?
- Which column will be used?
- What model could represent this data?
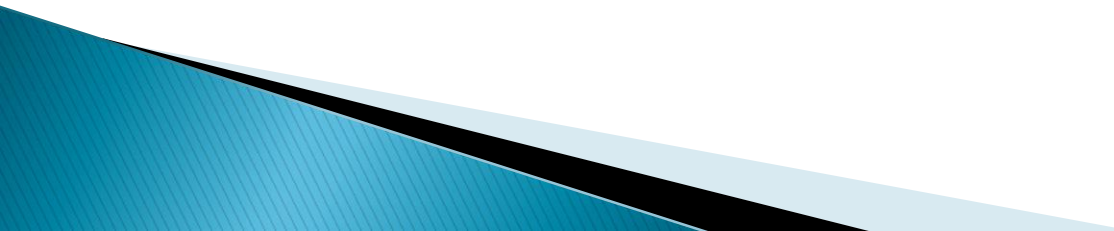- Hyperparameter, changed or not?

Pros and Cons of the Data

▶ Pros:
◦ Almost all of columns are categorical data, and none of the continuous data need to be filled.

▶ Cons:
◦ 4 Columns has 4 unknown values, decided to replace 2 unknown values from job and education and retain 2 others value unknown from contact and poutcome
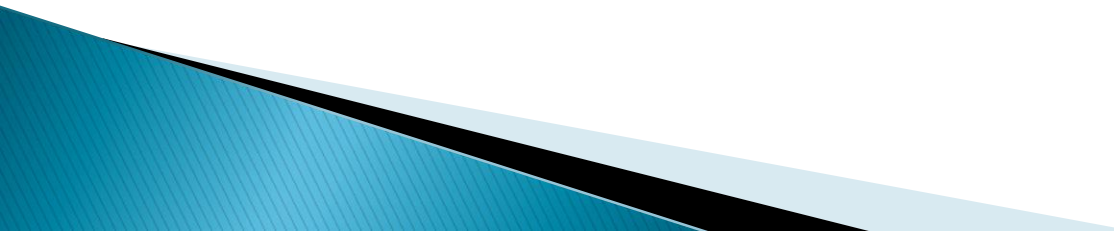◦ Target imbalance ratio : 11% – 13%, 39922 rejected and 5289 accepted

Timeline

- Knowing datasets
- EDA
- Cleaning and preprocessing
- Modelling data
  - Oversampling data target
  - Changing hyperparameter
- Conclusion

Data client from Bank

- Age
- Job
- Marital
- Education
- Default Credit
- Balance
- Housing Loan
- Personal Loan

Data related with Bank

- Last Contact by Type
- Last Contact Day
- Last Contact Month
- Last Contact Duration
- Contact Campaign
- Previous Contact Days
- Previous Contact Performed
- Previous Campaign Outcome

EDA

- Jump to Dash Plotly

Data Cleaning and Modelling

- Jump to Jupyter Notebook

Conclusion

▸ To determine people who wants to subscribe loan from bank, one should look at when do they received the latest info from bank (counting from duration, days and last contact days), how much balance do they have, and age could represent the necessity of loan itself.