

# Final Project

## CRISP-DM: Business Understanding

Aitore Issadykova  
Data Science  
Nazarbayev University  
Nur-Sultan, Kazakhstan  
aitore.issadykova@nu.edu.kz

Assem Kussainova  
Data Science  
Nazarbayev University  
Nur-Sultan, Kazakhstan  
assem.kussainova@nu.edu.kz

Zhaniya Koishybayeva  
Data Science  
Nazarbayev University  
Nur-Sultan, Kazakhstan  
zhaniya.koishybayeva@nu.edu.kz

**Abstract**—CRISP-DM is most used and effective methodology for all Data related projects. CRISP-DM methodology consists of 6 steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. Business Understanding part is crucial phase to determine organizations' objectives and constraints in order to reach the solution. The goal of this stage of the process is to uncover important factors that could influence the outcome of the project. Neglecting this step may lead to the excessive costs during the project stages. This project aims to predict the marketing campaign results of the local bank for their current customers. This paper provided Business Understanding stage of the CRISP-DM process for current project. Four aspects of Business Understanding phase were described in detail in separate sections.

**Index Terms**—CRISP-DM, Data Science, Data Mining, Business Understanding, Marketing Campaign, Bank

### I. INTRODUCTION

Business Understanding is first fundamental part of CRISP-DM process for any projects. Business Understanding phase helps to understand the business problem, determine customer needs and assess business resources and constraints, as well as it provides some guide for data mining goals and provide some detailed project plan structure for future implementations. This stage is crucial for businesses, since it should provide detailed cost-benefit analysis, as well as the timeline of the whole project.

This paper will perform Business Understanding process for the domestic commercial bank "The Bee Bank", which would know to predict the results of the marketing campaign for their customers with new term deposit product. The Business Understanding process for "The Bee Bank" process will consists of four steps: Determination of Business Objectives, Situation Assessment, Determination of Data Mining Goals and rough Project Plan for future Steps.

### II. BUSINESS OBJECTIVES DETERMINATION

#### A. Business Background

The commercial bank "The Bee Bank" is domestically founded Tier 2 bank which have several offices in each city of the country. "The Bee Bank" known in the country as the most reliable bank and have a great loyalty program

for its customers. The major bank issues as well as bank development is solved on the Board of Directors meeting. The daily operational issues under consideration of the bank CEO, which was selected by Board of Directors of the bank. The organizational structure of "The Bee Bank" can be seen on the Fig.1, where participants in the current project were highlighted by turquoise color.

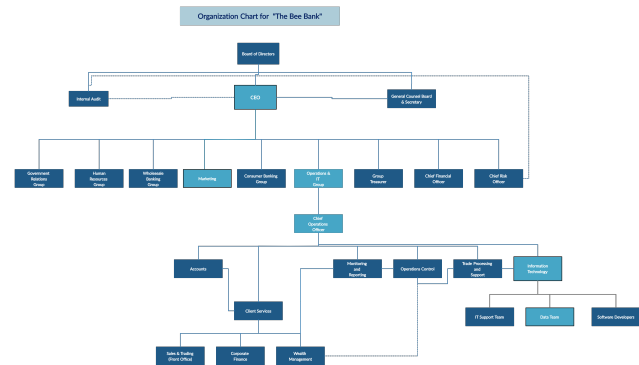


Fig. 1. Organizational Structure

#### B. Business Objectives

"The Bee Bank" introduced new bank product - term deposits. As part of their loyalty program, term deposits are currently available only for their current customers. The main goal of this project is to predict the certain marketing campaign result on the particular client. Results of this process will directly affect Marketing and Client Services Departments. By obtaining the prediction model of the campaign based on the general customer data, it will decrease costs of marketing campaign because it will be targeted on the particular group of clients, and Marketing and Client Services Departments will reach only this group of clients; maintain current bank reputation by not disturbing all clients and will obtain general model for future campaigns, since the data of customers may just be added to the particular model. Moreover, the successful marketing campaign will be awarded by the Board of Directors and will increase communication on the experience sharing with National Bank and other banks.

### C. Success criteria

The successful prediction model of the marketing campaign results will rise funds available for the banks, since targeted customer will open term deposits, save costs on the introducing campaign for not interested customer and will not lose customers due to enormous activity from marketing campaign and in the long run it will attract new customers. It will result in rising of the operational turnover, customer loyalty and bank reputation. Moreover, successful model will attract other banks for additional agreements as well as new communication with National Bank.

## III. ASSESSING SITUATION

### A. Resource Inventory

1) *Personnel*: The project will be mostly lead by Data Team Department of "The Bee Bank". Other participants in this project are Marketing Department and CEO. Marketing Department employees will communicate with Client Services Department and will help with data collection from their customers. CEO will help with communication with the Board of the Directors as well as perform as auditor on the current project.

2) *Data*: The Marketing Department provided the data of the bank current customers, which included the age, job, phone number, marital status, education level, housing availability and loan existence. The Marketing team will expand the data by introduce to the some part of the clients the new term deposits product and then will record whether customer is interested to get term deposit in the bank. In addition, data analysts in the Data Team will perform some each consumer data analysis and provide some data about consumer price index, consumer confusion index and possible outcome.

3) *Hardware resources*: The bank have their own Data Team Department, which includes data engineers, data scientists and data analysts. The Data Team have sufficient hardware resources such as new multicore computers with several monitors and was granted partial access to the customer database.

### B. Requirements, Assumptions, and Constraints

There are several requirements for this project that should be assessed first.

1) *Legal Requirements*: According to the Law, Bank employee should not threat the bank customers and use the personal information of the customers for their personal goals. The customer should be informed that his records will be included in the bank database and customer should be called by Marketing team only within working hours.

2) *Security Requirements*: Since the customers data is sensitive, the Marketing and Data Team Departments should complete and sign the non-disclosure information form. Moreover, there should be some security actions performed for storage of the data and passwords to the databases.

3) *Assumptions*: There no costs related to the data collection, since the data of the bank customers was collected before project initiation. The data assumed to be sufficient for the Data Science Project. CEO and Board of the Directors expected to receive the resulted model with some explanation of model selection and results. In addition, bank may sell the model of the marketing campaign company for other banks.

4) *Constraints*: The legal and security constraints will be completed as it was mentioned in the requirements part. The financial constraint of this project is 2 million budget. This budget should include the payments of the whole team during project and costs related to launch the result of this project.

### C. Risks and Contingencies

1) *Data*: We already assumed that data quality will be sufficient for model analysis, however it can be resulted that data amount is not enough or there are some required missing data. The bank addressed this risk by asking Marketing team to conduct the introduction of campaign to the part of the customers. All data collection processes conducted by phone by addressing legal requirements.

2) *Scheduling*: The risk related with data quality can affect the scheduling, but the contingency plan of data collection will reduce the risk of scheduling issue. Moreover, the Bank is actively recruit new Data Team employees, which will speed up the time of the whole project.

3) *Financial*: One of the main risk of this project is over budget. This risk will be assessed by CEO of the bank, which will deliver the additional raised costs for the Board of the Directors.

4) *Results*: This project aims to increase operational turnover of the bank and rise its funds for the investments. These factors will increase the profit of the bank as well as reputation of the bank among other banks.

### D. Terminology

The terminology will include only some data science terms and banking related terms.

*Model Selection* - choose the best model between the proposed ones.

*Classification* - the systematic grouping of the data points by their common features

*Regression Analysis* - the statistical analysis which shows the relations between input and output data

*Term deposit* - fixed-term investment that includes the deposit of money into an account at a bank. [1]

*Loan* - type of credit vehicle in which a sum of money is lent to another party in exchange for future repayment of the value or principal amount. [2]

### E. Costs/benefits

The potential costs of this project are additional data collection, additional staff employed and the costs related for creation prediction model, i.e., service server payments for data storage memory increasing, payments for providers of the data modeling software and new equipment. The benefit

of this project is to increase indirectly the bank profit without losing in the bank reputation and customer loyalty scores.

#### IV. DATA MINING GOALS DETERMINATION

1) *Goals:* The main goal of this Data Science project is the prediction of the results on the marketing campaign. For this project two type of the models will be considered. First is classification models such as Decision Tree, Random Forest and others. Second model is logistic regression which will calculate the probability of customer to get term deposit. Both models will get as the input the customer data and will predict label "yes" or "no" with some additional threshold inputs.

These models will assess the business goal of effective marketing campaign on the bank products which will help bank to target only specific auditory and do not disturb other clients. It is expected that effective target marketing campaign will increase the operational profit of the bank.

Fig.2 is an example of effective target marketing by other companies. It shows the difference between traditional marketing campaign and targeted campaign for different sectors of the company [3].

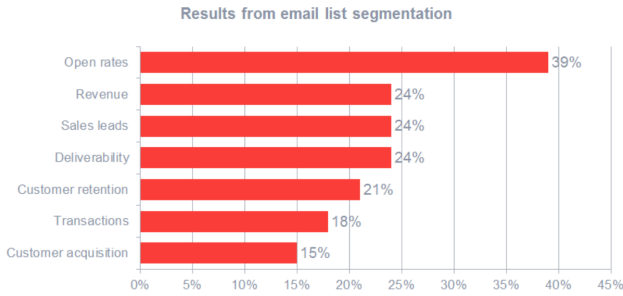


Fig. 2. Increase in outcomes for targeted marketing campaign [3]

2) *Success Criteria:* All models will be assessed with accuracy for classification by splitting initial available data on test and train. The benchmark of the classification model is accuracy and model will be selected if the accuracy will be higher than 90%, since the main goal is correct customer segmentation for the giving campaign. In case of more than two models will achieve the threshold, the additional analysis on the classification result will be conducted.

If results will outperform of expected results or will give same results as models provides after deployment of targeting marketing campaign, it will be considered as data mining project success and the data science team will be awarded by the Board of Directors and will be collaborate with other banks Data Science Teams.

#### V. PROJECT PLAN

##### A. Plan

The initial project plan with phases, time completion, resources used and risks are described in Table 1. Risks of each

phase as well as risk addressing techniques were covered in III-C section.

TABLE I  
INITIAL PROJECT PLAN

Phase	Time	Resources	Risks
Business Understanding	2 weeks	Data Team, Marketing and CEO	Financial
Data Understanding	2 weeks	Data Team and Marketing	Data and Scheduling
Data Preparation	4 weeks	Data Team	Data and Scheduling
Modeling	4 weeks	Data Team	Scheduling
Evaluation	2 weeks	Data Team and CEO	Results
Deployment	2 weeks	Data Team, Marketing and CEO	Financial and results

##### B. Initial Assessment of Tools and Techniques

For this project purposes SQL databases will be used for data extraction and Python supporting software such as PyCharm will be used for Data Preparation, Modeling and Evaluation phases. All licences were bought by bank before project launching.

#### VI. CONCLUSION

The CRISP-DM methodology is considered as the most effective strategy for leading Data Science projects. This paper provides the first phase of CRISP-DM process: Business Understanding for the local bank. "The Bee Bank" is added new bank product "term-deposits" and will introduce it to its customers. The Board of Directors of the Bank and CEO would like to introduce the targeting marketing campaign to maintain current bank reputation and customer loyalty as well as increase its operational profit. The four steps of Business Understanding phase: Determination of Business Objectives, Situation Assessment, Determination of Data Mining Goals and rough Project Plan for future Steps - were discussed in the separated sections. The next step of this project process will include Data Understanding phase as well as some part of Data Preparation phase of CRISP-DM.

#### REFERENCES

- [1] Chen J. "Term Deposit Definition.", Investopedia September 2020. Available: investopedia.com
- [2] Kagan J., "Loan Definition.", Investopedia, October 2020. Available: investopedia.com
- [3] "89% of Businesses Make the Same Email Marketing Mistake", Super-Office, October 2020. Available: superoffice.com