Analysis of business case based on CRISP-DM methodology

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

***Abstract*— Data is integral to business, and software can help you mine and make sense of the information you collect. Data mining techniques have a lot of benefits and add value to business by providing possibilities to optimize marketing campaigns, detect possible fraud and make better business decisions. This paper provides a description of business case based on the first phase of the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology Business Understanding. In this case, an online platform for selling goods of Korean company introduced as a business project and then analyzed using the given methodology. Using the analysis, it is possible to study in detail all aspects of the project in order to avoid mistakes and unforeseen circumstances in the future.**

Keywords—CRISP-DM, business understanding, data analysis, data mining

I. Introduction

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs.

In this paper, creation of online platform for Korean women clothing sales company is described, based on the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology. This paper is organized as follows. Firstly, the theoretical background, including the definition of CRISP-DM methodology is overviewed. Then description of business is provided according to Business Understanding phase of CRISP-DM, concerning business objectives, possible risks, expected results and future plan of actions. Next, paper provides a conclusion of the business analysis. Finally, resources that were useful for the paper are listed.

II. Theoretical Background

Any good project starts with a deep understanding of the customer's needs. Data mining projects are no exception and CRISP-DM methodology most commonly used for these purposes. CRISP-DM methodology is a standard methodology for performing standard Data Mining process, which does not depend on a particular industry or a particular tool [1]. An important feature of the methodology is paying attention to the company's business goals. This allows management to perceive data analysis projects not as a playground for experimentation, but as a full-fledged element of the company's business processes. According to CRISP-DM, an analytical project consists of six main stages, performed sequentially: Business understanding, Data understanding, Data preparation, Modeling, Evaluation, and Deployment.

Business Understanding stage is described as defining project objectives and business requirements. This knowledge is then converted into a data mining task statement and a preliminary plan for achieving project goals. Figuring out the business understanding for the data mining effort helps to ensure mutual understanding among all team members before spending valuable resources. This phase includes 4 stages: Determining Business Objectives, Assessing the Situation, Determining Data Mining Goals and Producing a Project Plan [1]. Following is the analysis of the business according to each of the stages.

III. Determining Business Objectives

Initially, it is needed to decide on the goals and scope of the project. Conflicting goals and constraints are often set, so it is needed to formulate them correctly. The analyst's goal is to highlight important factors that can affect the results of the project. Neglecting this step can waste a lot of effort answering the wrong questions.

A. Background

Consider as a business a Korean women clothing company that wants to open its own online platform for selling merchandise. This includes different types of products like clothing, shoes, hats, accessories and more. At the moment the company operates using its stores in rented premises throughout some city. However, the company wants to expand its operating area to other cities in order to attract more buyers and be competitive with other companies. The company has made the following organizational structure:

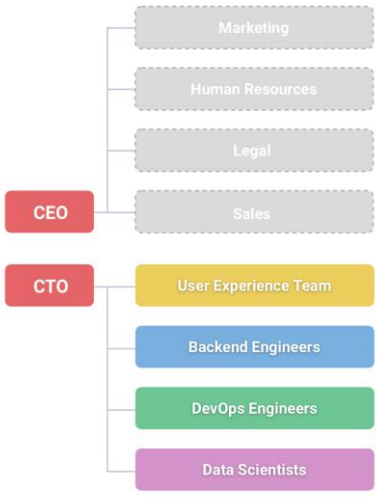

Fig. 1. Organizational structure of company

B. Business Objectives

The main goal of the business is to increase sales of offered goods and retain current customers. All layers of the organizational structure related to the CTO will be involved in the creation of the online store platform and Marketing and Sales work will also be required to achieve the main business goals. Achieving these core goals will also require developing existing customer relationships to maximize the value of every existing customer in the company. Thus, the project is complemented by the following goals: improving related sales through better recommendations and increasing customer loyalty through more personalized service.

C. Business Success Criteria

The project will be assessed as successful under possible conditions when the number of related sales will increase by 10-15%. In addition, if action is taken to address customer preferences, such as providing requested products, customers will spend more time and browse more pages on the platform than on competitors. An important criterion for the success of the project is also the completion of the planned work on time and within the budget.

IV. Assessing the Situation

This task includes a more detailed study of resources, constraints, assumptions and other factors that should be considered when determining the purpose of data analysis and drawing up a project plan. In the previous task, a goal was to quickly understand the essence of the task at hand, and now details should be highlighted.

A. Resources

The company is allocating $2 million to build an online platform. From human resources, technical experts from the CTO are involved who will be responsible for the creation of the site and data analysis. The data is provided by Marketing and Sales staff, which contains information about customers, purchases and their preferences. In terms of physical resources, technical experts will be provided with the required equipment with sufficient characteristics to complete the tasks, purchased for part of the budget.

B. Requirements, assumptions and constraints

Among the requirements is the confidentiality of customer and company data, which cannot be disseminated by the project participants. In addition, the database and online platform need to be protected to prevent information theft. This will be monitored by the company's CTO itself. It is assumed that the goods provided will need to be expanded with those of competitors if they are the preferences of the company's customers. Among the constraints of the project, it is assumed not to exceed the established budget for the implementation of the goals, as well as the aforementioned limitations on data access.

C. Risks and Contingencies

There is a risk that the project will take longer than originally planned, but this risk does not have a strong impact on the company, since the sale of goods will be carried out offline before the launch of the project. As for the situation, if the company exceeds the budget for the implementation of the project, additional resources from the sale of goods will be required to cover the additional costs. If there is not enough information to use as a project database, the company will have to conduct additional customer surveys to replenish the information.

D. Terminology

To ensure that the business and data mining teams understand each other a small glossary of technical terms is provided that might need clarification.

Data Mining - the process of examining a set of data to determine relationships between variables which could affect outcomes.

Data Modeling - is all about turning data into predictive and actionable information.

Data Visualization - any attempt to make data more easily digestible by rendering it in a visual context.

Unstructured Data - any data that does not fit a predefined data model.

E. Costs and Benefits

The cost of the project includes funds that may be necessary for additional data collection, for assembling an online platform (buying equipment, renting a server, linking payment transactions), as well as advertising activities to disseminate information about a new additional format for the company's sales. The benefits of the project include an expanded customer base of the company and improved service for existing customers, which will positively affect the growth of sales of goods.

V. Determining Data Mining Goals

A business goal defines a task in terms of a business. The purpose of data analysis defines the task in technical terms. This step describes the expected results of the project, which will lead to the achievement of business goals and determine the criteria for the successful outcome of the project.

A. Goals

To achieve the goal of increasing sales of goods, a recommendation system model will be built. The main task of the system is to offer new products based on purchased-viewed or selected ones. Several goals are pursued at once, but the main one is to offer a product to a buyer who is most likely to lead to a sale and satisfy his requests. This means that an informal recommendation system offers some ordered list of products based on the customer's background. Such model is implemented in different online platforms in various fields, and on Figure 2 the Amazon recommendation for customers is displayed [2].

To achieve the goal of providing a personalized service for each client, it is assumed to use ready-made mathematical models for profiling activity based on the received events. The online platform will need to define what different types of buyers are trying to find and then redesign the site to highlight those positions. Thus, each individual buyer will see a special main page of the platform where products are collected specially for him.

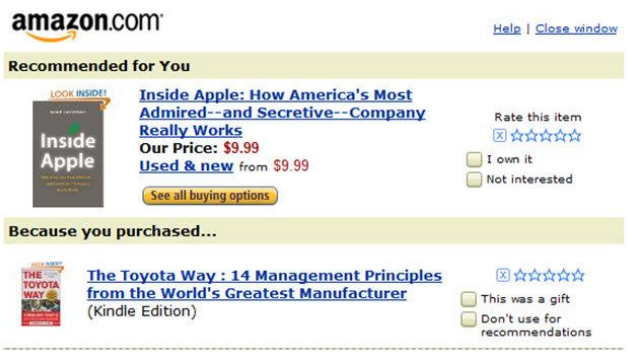

Fig. 2. Recommendation system of Amazon company website

B. *Success Criteria*

Any recommendation system helps to solve a specific business problem. And the result is measured in ways that are understandable for a business task - the number of visitors, sales, CTR, etc. The best approach is to try to design a metric directly for the business problem. The product of the proposed service is personal lists of recommended products, and the rating of the product is its assessment by the user. User feedback is compared with lists of items that have been recommended. If the film was recommended to a person, and there is a positive rating from him, then they got it for sure. It is these cases that are best for the system, so you need to maximize their number. If the recommendations contain all positive signals, the accuracy of the system is maximized and improvements are no longer required.

VI. Producing a Project Plan

Once all the basic questions are answered and the purpose of the project is clear, it's time to draw up a project plan. At this stage, a plan is described to achieve the goals of the data analysis and thus the business goals. The plan should contain the intended set of steps to be completed during the rest of the project, including the initial selection of tools and methods.

A. *Writing the Project Plan*

The overview plan for the project is as follows:

TABLE I. Project Plan

Phase	Time	Resources	Risks
Business understanding	1 week	All specialists	Financial problems
Data understanding	2 weeks	All specialists	Data and Technical problems
Data preparation	4 weeks	Data scientist, Marketing, Sales	Data and Technical problems
Modeling	3 weeks	Data scientist, IT specialists	Technical problems
Evaluation	2 weeks	All specialists	Problems with results and finance
Deployment	1 week	Data scientist, IT specialists	Problems with results and finance

B. *Assessing Tools and Techniques*

The project will also require various analysis and implementation tools. Data can be stored in SQL databases and for analysis can be retrieved and stored in Microsoft products that require a license purchase. To build the necessary models, the Google Collab online service can be used, which provides free services for all users.

VII. Conclusion

In conclusion, this paper described the analysis of provided business case at the Business Understanding stage of the CRISP-DM methodology. The report describes business case, its objectives, possible risks as well as data mining models for reaching chosen goals of project. After all of the business aspects, following all the points, useful information was obtained and future plan of actions was constructed.

References

[1] http://crisp-dm.eu/.

[2] https://www.amazon.com/