Exploring value creation through web mining: a case study on the online weather forecast business

Jun Che

Department of Mechanical Engineering
Blekinge Institute of Technology
Karlskrona, Sweden
2015

Thesis submitted for completion of
Master of Sustainable Product-Service System Innovation (MSPI)
Blekinge Institute of Technology, Karlskrona, Sweden.

Abstract: The rapid progress of internet business makes organizations and companies to accumulate vast amounts of web data. It is a tremendous challenge to extract valuable information from vast amounts of these data. Generally, traditional data analysis tools and techniques are unable to process of such a large amount of web data effectively and accurately. In addition, given the complexity of web data even using traditional methods to process the relatively small data sets, it is difficult to obtain a noticeable effect. In this context, data and mining are emerging technologies to support the extraction of valuable data from vast amounts of web data. This thesis investigates the benefits of data and web mining for the internet-based weather service companies by finding gaps in the customer’s journey to improve the user experience, and outlining the potential of new business opportunities. A literature review and industry analysis was undertaken to understand the current state of online weather industry and identify the main touch points that impact the user. A case study was conducted to demonstrate research results applying into practice. The final outcomes are possible solutions to improve user experience and a new business model based on data mining to exploit new profit for online weather service companies.

Keywords: Business Model, Customer’s User Experience, Data Mining, Internet, Online Weather Service, Personalized Service, Web Data, Web Mining
Acknowledgement

I would like to express my gratitude to all the people who have been helped me along the way and all those who gave me powerful support through all the difficulties of the transition of the thesis.

Special thanks to my supervisor, Alessandro Bertoni for his patience and guidance in helping me finish my thesis. His valuable suggestions and constructive criticism helped me move forward. Whilst I appreciate him for his valued time and abundant experience guided and enriched the research I undertook.

I would like to express my gratitude to Matteo Savoldelli who gave me great help to understand the case company and industry.

Thanks to Christian M. Johansson, Marco Bertoni, Massimo Panarotto for giving me valuable feedback in the check-in presentations. Their participation has been a great perspective on the thesis.

I would like to thank all my fellow students who have shared learning experiences with me and given me supporting help during this thesis work.

Finally, thanks to my family and friends for their patience, understanding and supports throughout the thesis process.
Summary

Introduction
This thesis outlines potentials of web mining for online weather service, by identifying the benefits of web mining for company and customer in order to improve the customer’s user experience. This thesis also explored the new business opportunities generated by web mining for online weather service company. The thesis addresses two research questions; which are:

RQ#1: How can the current user experience of the customers of online weather service companies be improved by the use of available data and web mining?
RQ#2: How can weather service companies based on the web mining exploit new business model to gain new profit?

Methods
In this thesis a literature review was used to acquire an essential understanding of data and web mining. The industry research enabled an analysis of the online weather service development and clarified the current user experience of the online weather service company. Both the literature review and the industry analysis enabled the author to address the first research question. A case study was used to address the second research question.

Results and Discussion
This thesis investigates the current online weather service business and the current customer journey of the online weather service. This thesis identifies data and web mining technologies as possible means to create new value by improving the customer user experience. It proposes a new business model in order to provide evidence about the potential of web mining for online weather service companies. The related implementation plan of new business model is also formulated. Finally, an overview of current customer user experience improvement and the implementation challenge of the business model renewal are discussed.

Conclusion
This thesis shows how web usage mining and related methods can be used to improve the customer user experience. A new business model based on web mining is suggested in order to help the company to explore new profit.
Glossary

**Association Rules Mining:** Association rules mining is a data mining method. It discovers the meaningful relationships that are hidden in the large database set, and all the objects in these relationships are unordered.

**Business Model Canvas:** The business canvas is a visible and structured chart for describing existing business model or developing new business model. It is used to build a shared and common understanding of a business model throughout the company and all the involvers.

**Classification:** Classification is a widely used method in data mining and the aim of classification is to maps diverse data items into several predefined categories.

**Clustering:** The clustering is one data mining technique to discover the intrinsic attribute of data. It intends to gather the data items which shared similar characteristics into similarity groups.

**Conversion:** In the online business, conversion refers to the act of the website visitor complete a company expected activity.

**Customer Journey Mapping:** A process that present the user’s service experience recording the customer’s journey at each step of their interaction. It provides a map of the interactions and emotions that take place throughout the journey.

**Customer Loyalty:** Customer loyalty is the result of consistently positive emotional experience, physical attribute-based satisfaction and perceived value of an experience, which includes the product or services.

**Customer Perception:** The customer perception refers to the customer’s feeling, impression and awareness about the company and its offering.

**Customer’s User Experience:** Encompasses all aspects of the end-user's interaction with the company, its services, and its products.

**Database Management System (DBMS):** It refers to large-scale software to manipulate and manage the database, and it use for establishing, use and maintenance of the database.
Data Mining: Data mining is a process of extracting the hidden, unknown and potential valuable knowledge and information from the massive, incomplete, and random data.

Hypertext Transfer Protocol (HTTP): It is the underlying protocol used by the World Wide Web.

Touch Point: A touch point is any time a potential customer or customer comes in contact with the product or service of a before, during, or after they use something from the company.

Uniform Resource Locator (URL): It is the global address of documents and other resources on the World Wide Web.

Web Data: Web data is the data are generated on the internet which including web statistics data, web server log data, web content data and web structure data.

Web Mining: Web mining is the application of data mining technology in web environment.

World Wide Web (WWW): World Wide Web is an information system of interlinked hypertext documents that are accessed by internet.
# Table of Contents

Acknowledgement .......................................................................................................................... ii

Summary ......................................................................................................................................... iii

Glossary ............................................................................................................................................ iv

Table of Contents ........................................................................................................................ vi

List of Figures and Tables ........................................................................................................... ix

1 Introduction .......................................................................................................................... 1

1.1 Background ....................................................................................................................... 1

1.2 Research Problem and Objectives .................................................................................... 2

1.3 Thesis Structure .................................................................................................................. 3

2 Methodology .......................................................................................................................... 4

2.1 Literature Review ............................................................................................................. 4

2.2 Industry Analysis .............................................................................................................. 6

2.2.1 Web Review .................................................................................................................. 6

2.2.2 Customer Journey Map .............................................................................................. 7

2.3 Case Study ........................................................................................................................ 7

2.3.1 Semi-structured Interview .......................................................................................... 8

2.3.2 SWOT Analysis and Stakeholder Analysis ................................................................. 9

2.3.3 Business Model Canvas .............................................................................................. 9

2.3.4 Network Pictures ........................................................................................................ 11

3 Literature Review .................................................................................................................. 13

3.1 Data Mining Definition ...................................................................................................... 13
3.1.1 Data Mining Technical Definition ................................................... 13
3.1.2 Data Mining Commercial Definition ................................................ 14
3.1.3 The Tasks of Data Mining ............................................................... 14
3.1.4 The General Structure of Data Mining System ................................ 15
3.2 Web Mining ............................................................................................. 16
  3.2.1 The Data Sources of Web Mining ................................................... 18
  3.2.2 Web Mining Classification .............................................................. 19
  3.2.3 The Main Data Mining Methods and Algorithms Used in Web Mining  21
  3.2.4 Web Mining Processes Model ......................................................... 26
4 Result .................................................................................................................... 28
  4.1 Industry Analysis ...................................................................................... 28
    4.1.1 The Development of Weather Service Industry in Internet Environment ................................................. 28
    4.1.2 Trends of Online Weather Services ................................................. 29
    4.1.3 Key Indicators of Website Design ................................................... 32
  4.2 Customer Journey of a Weather Service Website ...................................... 33
    4.2.1 General Phases of Customer Journey on the Weather Service Website .......................................................... 34
    4.2.2 Important Touch-Points ................................................................. 36
  4.3 The Proposed Solution of Web Mining for Customer Experience Improvement .......................................................... 39
  4.4 A Case Study in the Internet-based Weather Service Company ............... 44
    4.4.1 The Background of Case Company ................................................ 44
    4.4.2 Business Analysis .......................................................................... 45
List of Figures and Tables

Figure 1.1 Thesis Structure ................................................................. 3
Figure 2.1 Literature Review Methods: Point of Departure (Liston, 2011) ......5
Figure 2.2 Business Model Canvas (Osterwalder, Pigneur, 2010) .......... 10
Figure 2.3 ARA model (Haakansson 1990) ........................................ 12
Figure 3.1 A Typical Data Mining System Structure (Han et al., 2011) ......15
Figure 3.2 the Architecture of WWW ................................................. 17
Figure 3.3 Two Basic Web Link Structures ....................................... 18
Figure 3.4 An Example of Web Server Log ....................................... 19
Figure 3.5 Web Mining Classification ............................................... 20
Figure 3.6 the Fundamental Process of Classification ........................... 22
Figure 3.7 the Example of Decision Tree ......................................... 23
Figure 3.8 the Example of Clustering Result ...................................... 24
Figure 3.9 the Web Mining Process Model ........................................ 26
Figure 4.1 US Time Spent Using the Internet by Device (comScore, INC, 2014) ...................................................................................................................... 31
Figure 4.2 Steps of Weather Service Website Customer Journey ........ 34
Figure 4.3 Important Touch Points of Weather Service Website .......... 37
Figure 4.4 the Classification of Important Touch Points ....................... 38
Figure 4.5 the Example of Clustering Result ...................................... 41
Figure 4.6 the Example of Customer Correlation Analysis ................. 42
Figure 4.7 the Current Business Model of 3Bmeteo ........................... 47
Figure 4.8 The SWOT Analysis of 3Bmeteo ....................................... 49
Figure 4.9 the New Business Model of 3Bmeteo ................................. 50
Figure 4.10 Analytics Salary/Income by Region and Employment type (KDnuggets, 2013) ...................................................................................................................... 52
Figure 4.11 the Two Stages of Revenue Stream in New Business Model ....53
Figure 4.12 Networking Picture of New Business Model ..................... 56
Figure 4.13 the Implementation Plan of New Business Model .............. 59
Figure 5.1 Challenges of Implementation Plan ..................................... 71

Table 1.1 Thesis Objectives .................................................................. 2
Table 3.1 the Comparison of Different Types of Web Mining ..................21
Table 4.1 Key Indicators of Web Design ........................................... 32
Table 4.2 the Proposed Solution of Web Mining .................................. 43
Table 4.3 the Characteristics of 3Bmeteo .......................................... 44
Table 4.4 Stakeholders of 3Bmeteo .................................................... 46
Table 4.5 The Scenarios of Return on Investment Example .................. 55
Table 5.1 Obstacles of Current Customer’s User Experience of Online Weather Service

67
1 Introduction

1.1 Background

Nowadays, internet influences every aspects of people’s daily life, allow people to better understand the world and to explore it more deeply. Through the internet people easily access a huge amount of information when they need irrespectively of where they are. The internet provides a wide range of choices for the people. On the one hand, with the improvement of consumer knowledge level and the change of consumption conception, people desire more diversified products, more personalized services and lower cost [Feng, 2006]. On the other hand, companies are using the internet to get a deeper understanding of customer needs and to establish closer ties with customers. The internet enables companies to offer better product and more convenient services and to enhance the interaction with customers. Thus, more and more individuals and organizations use the internet. Under such circumstances a lot of data are generated intentionally or accidentally and most of these data are recorded and stored.

Along with the continuous development of technology, people's thinking is also in constant revolution. Technical feasibility makes people break through previous barrier to process massive and complicated data. With the rise of cloud computing and computing technology progress, it’s possible to utilize these massive data to create new value. Nowadays, companies or individuals shall have the ability to process these data and recognize the value of the data from a fresh perspective. For instance, a company could utilize data to predict future market changes as well as it could develop better products and attract more users by using the available data. For this result, an increasing number of companies pay attention on data mining. Data mining, a sub discipline of computer science, is the process of discovering knowledge in large data sets. Generally, data mining aims to discover knowledge from massive data and transform it into an understandable result for further use [Trevor et. al., 2009]. In recent years, data mining has been used widely in multiple areas. In business, data mining is used to discover hidden knowledge. The knowledge refers to patterns and trends which can be used to reveal unknown strategic business information [Brien & Marakas, 2010].

The weather service is closely bound with human life, for some people checking the weather forecast is an essential daily activity. Along with the popularity of the internet, the weather service also extends to the internet. The flexibility of online weather services has attracted more and more users. Like for most of the other internet-based services, the user generates a large amount of data while using an online weather
forecast platform. The data mining may help the online weather service company to find new solutions to improve their products and services. Also it could help the online weather service company to gain new profit.

### 1.2 Research Problem and Objectives

This thesis intends to explore the potential benefits for the company by the implementation of data mining in order not only improves the user experience but also to gain more opportunities to create new profit channels in long term. It intends to exploit new business models in relation to data mining for an internet-based weather service company. The overall objectives of this research are shown in table 1.1.

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Thesis Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research question #1</strong></td>
<td>Gaining insight into the data mining application in the internet environment; understanding the current customer’s experience of online weather service company; identifying the benefits of data mining for company and customer.</td>
</tr>
<tr>
<td><strong>Research question #2</strong></td>
<td>Exploring the potential of data mining for an internet-based weather service company.</td>
</tr>
</tbody>
</table>

These two research questions were formulated to address the thesis objectives; which are:

**RQ#1**: How can the current user experience of the customers of online weather service companies be improved by the use of available data and web mining?

**RQ#2**: How can weather service companies based on the web mining exploit new business model to gain new profit?
1.3 Thesis Structure

The structure of this thesis is illustrated in a visual way to provide the reader with clear outline to follow the development process of this thesis. This thesis includes six chapters which are introduction, methodology, literature review, result, discussion and conclusion.

![Figure 1.1 Thesis Structure]

The first chapter states the research background and introduces the research topic, questions and objectives. The research methodologies used in this study are described in the chapter 2, namely literature review, as-is analysis, web analysis and case study. The chapter 2 explains how the data are obtained for each research question and also how the analysis is conducted. The literature review in Chapter 3 provides the background knowledge for the understanding of data mining and related applications in the internet environment. More specifically, this chapter introduces definition, characteristics and process of web mining. Chapter 4 elaborates on the results from the research and included four more detailed subchapters. Chapter 5 discusses the key points of this research. Chapter 6 presents the findings and conclusions addressing the research questions and the objective of this thesis, moreover, it states the limit of this research and suggests the possible future research directions.
2 Methodology

In this section, the methods chosen for conducting research, analysis and data collection are presented. Each method chosen is briefly described and the reason for applying the methods is explained. Based on the research questions and research timeline, there are three main research phases contained in this thesis and each phase consists of several methods. The first phase aims to acquiring a comprehensive understanding of the current state of data mining and web mining through literature review. Then the industry research enabled an analysis of the internet-based weather service development and clarified the current user experience of the internet-based weather service company. Both of the literature review and industry analysis enabled the author to propose a new business model renewal in a case study, provide evidence of the potential of data mining for the internet-based weather service company.

2.1 Literature Review

The literature review is a research activity that summarizing and analyzing the available relevant research and non-research literature on a particular topic [Hart, 1998]. The purpose of literature review is to gain an overview of existing studies in a particular topic [Cornin & Coughlan, 2008] and present the reader the current literature on the topic [Colling, 2003]. In this study, a literature review was conducted for the first two months at the beginning of thesis process. The literature review should contain a clear search and selection strategy (Carnwell & Daly, 2001). The figure 2.1 presents the framework of the process of the literature review.
Based on the Liston’s (2011) point of departure framework, the literature review of this research can be divided into four phases. As the primary interest for the research are online data utilization and data mining in the internet environment. The concentrated understanding of research field was developed in the initial literature review. The initial literature review is used for providing useful guide for future research explorations [Liston, 2011]. In this phase, the author focused on finding resources related to interest topic in order to gain a theoretical understanding of the data utilization and data mining. Meanwhile, the initial research questions were proposed. Afterward, the related research area was explored in an exploratory phase. The exploratory phase aims to outline and summarize topic related theories, models and limitations [Carnwell & Daly, 2001]. Also, the author identified categories relevant to research. This phase the focuses of the literature review is data mining theory, concept and application in the business domain. In order to analyze the key findings of previous literature review, the author conducted a focused literature review to study the application of data mining in the web environment. The focused literature review contributes to identify potential gaps, areas unexplored as well as define the scope of research [Alexander, 2012]. As the result, a detailed study of the web mining and related researches has completed in the third phase. Moreover, the author reviewed and analyzed the impact of web mining for online business, the method of realization of web mining and the related tools of web mining in the refined literature review. The research questions were refined during this phase as well. At the last stage of the literature review, a refined literature review is conducted in order to organize the literature to ensure that it supports the research and finalize the research questions [Liston, 2011]. During this process, the final research questions...
were proposed and the author organized the framework of the literature review writing.

The relevant books, journals, articles, academic databases, dissertations and statistic from internet were reviewed in this step. Research from journals was from various different countries that most of the literature was in English and some of them were in Chinese.

## 2.2 Industry Analysis

To create a good starting point to address the first research question, an industry analysis was conducted. The intention of industry analysis was to identify and clarify the current situation of the internet-based weather service industry, and to define the present customer journey of the internet-based weather service website. Moreover, this analysis helped to integrate the finding of literature review into the established scenario.

### 2.2.1 Web Review

The number of published studies associated with online weather service is limited, implying that the most of relative information had to be gathered through a web review. This process is mainly to search useful information and data through the internet which related to online weather service potential and its development trend.

In order to clarify the potential of online weather service, an important step of the web review is to estimate the scale of online weather business. However, there is no comprehensive statistic can indicate exact data of the online weather service. The author has to read some articles and blogs from economic forums and online weather website to find evidence to estimate the scale of online weather business.

From another aspect, the author analyzed several online weather service companies (which including AccuWeather [AccuWeather, 2015], the weather channel [weather, 2015], and the climate corporation [Climate Corporation, 2015] and Moji weather [Moji China, 2015]) in order to discover and explore the trend of the online weather service development. During this process, the author studied the characteristics of those companies and the emphasis of their development. Based on the finding, the author identified several trends of online weather service development.

Furthermore, the author reviewed and analyzed the public account of several online weather service companies on the social media. The users’ feedback, comment and
question are the focus of the review. It plays a significant role for construct customer journey map and it helps the author to recognize the users’ feelings and thoughts about the online weather service. The main public accounts involved in this process are 3Bmeteo’s Facebook [3Bmeteo Facebook, 2015] account and AccuWeather’s Facebook account [AccuWeather Facebook, 2015].

### 2.2.2 Customer Journey Map

As a common challenge of services, the emotional distance between service provider and its customers is reflected in the low customer satisfaction. To close this gap, services “need to be understood as a journey or a cycle—a series of critical encounters that take place over time and across channels” [Parker and Heapy, 2006]. The customer journey map is a tool that helps keeping focus on the customer experience before, during and after using a product or service. For this purpose, the customer journey mapping can be a good tool for analyzing the value that the customer gets throughout the experience, and to generate possible solutions to add value or reducing activities that do add anything to the user.

In order to gain a deeper understanding of the offering of internet-based weather service and visualize the user experience. The user experience was mapped out in a customer journey map. Based on the finding of web review, a horizontal comparison of weather website was conducted. The customer journey map was constructed by following steps [Oxford Strategic Marketing, 2009]:

1. Confirm the whole journey, identify and define the customer;
2. Identify key journey steps;
3. Identify the goals, activities, feeling and thoughts for each step;
4. Clarify and classify the important touch points of the customer journey;
5. Define the key point in the journey where customers may pause and evaluate the experience or make a crucial decision.

### 2.3 Case Study

Case study is a research method of qualitative research that it is used to analyze a particular case in a practical environment condition [Johansson, 2003]. It intends to explain the mechanism, outcome of the particular case as well as provides the
experience as reference for other related case [Baxter & Jack, 2008]. In this thesis a case study was conducted which aims at evaluating the usefulness of data mining into practice. Generally, a case study consists of four main components: case study design, data collection, data analysis and reporting case studies [Yin, 2003]. The key elements of case study design are establishing study’s question and its unit of analysis [Yin, 2003]. In this research, the case study’s question is based on the second research question and the unit of analysis is one Italian online weather service company. In order to collect evidence to support this case study, the data collection of this case study has been achieved using following five means: web review, semi-structured interview, SWOT analysis, and stakeholder analysis and business model canvas. The data analysis of this case study is inspired by United State General Accounting Office’s OTTR principle, which stands for “observe”, “think”, “test” and “revise”. The concept of OTTR suggests that during and after observations, the researchers should think about the meanings of information collected in data collection step. This thinking leads to ideas about new types of information required in order to confirm existing interpretations. During the test phase the researchers collect additional information which may lead to revisions of initial interpretations [Baškarada, 2013]. The case study writing includes background description, specific problems description and analysis (the two business models), results analysis and summary (in discussion section).

2.3.1 Semi-structured Interview

Semi-structured interview is a social science research method which widely used in qualitative research [Meehan, 2014]. Compared with structured interview, the semi-structured interview is more flexible and opens [Whiting, 2008]. Similar with structured interview, the interviewer need to develop an interview guide which includes a list of questions and topics in a particular order. The interviewer follows the guide, but is able to follow topical trajectories in the conversation that may stray from the guide when the interviewer feels this is appropriate [Cohen & Crabtree, 2008]. Moreover, the questions are standardized, but the interviewer could adjust the order in which questions are asked in the process of the interview [Bjornholt, 2012]. This kind of interview focus on collects the detailed information from the conversational and it often used when the researcher intends to explore into a topic deeply and to understand the answers provided completely [Clifford et. al., 2010].

Generally, the semi-structured interview consists of two parts: preparation and physical meeting. The physical meeting is an ideal way to touch the interviewees. For the reason of geography, the author was unable to meet the case company in person. Thus, a teleconference takes the place of the physical meeting in this research. The preparation part has taken three weeks, in this process the author gained a general
understanding of the case company and built a shared understanding with the case company through the email. Then a teleconference with the targeted case company was convoked. The intention of this teleconference was to reinforce the understanding of the case company’s current situation. And it clarified the specific details of the current business model and business process. Moreover, it defined the current challenges that are being faced by the case company. The main contact person is the IT technical consultant of the case company.

2.3.2 SWOT Analysis and Stakeholder Analysis

In order to identify the influence of company that is caused by the external factors, and to understand how the internal factors affect the current business process. The author conducted a SWOT analysis and stakeholder analysis with the case company. SWOT analysis is a method which can be used to summarize and evaluate the strengths, weaknesses, opportunities and threat of a business or a company [Ayub et. al., 2013]. Stakeholder analysis is the method of definition and identifying the individuals or organizations that are likely to affect the business or the company [Jepsen & Eskerod, 2009]. The results of these two analyses contribute to build the current business model and the business model renewal. In addition, it helps the author to evaluate the current situation of the case company from the company’s perspective. These two analyses have been completed after the web review and semi-structure interview. According to the web review results, the author created the initial SWOT map and stakeholders of case company. Then the author verified and improved them during the semi-structure interview with the case company. Based on the findings, it enables the author to formulate appropriate development strategies and plans and can help company to concentrate their resources and activities on the strengths or where have more opportunities.

2.3.3 Business Model Canvas

The business canvas is a visible and structured chart for describing existing business model or developing new business model [Barquet et al., 2011]. It is used to build a shared and common understanding of a business model throughout the company and all the involvers. It vividly exhibits the process that the company delivers their value proposition to customers. The business model canvas consists of nine important building blocks, which describes the business model in an organized way. The nine blocks focus on four main areas of a business [Osterwalder & Pigneur, 2010]:

9
Offering

- Value propositions: it refers to the value is created by the company provided product or service for the customer. The value aims at satisfy the customer or potential customer needs and achieves the purpose of profit. Besides, the value proposition is proposed to examine or analyze the dynamic relationship between benefit, cost and customer value [Value Proposition, 2010].

![Business Model Canvas](image)

*Figure 2.2 Business Model Canvas (Osterwalder, Pigneur, 2010)*

Customer

- Customer segment: the different groups individuals and organizations that a company intent to reach and serve. The company segments its potential customers into different groups in order to better meet customer needs. And it allows company to rationally allocate resources and improve the quality of services.

- Channels: the channel refers to how does a company reach and communicate with the customers and how does the company deliver the value propositions to its customer.

- Customer relationship: it describes the types of relationship the company established with different customer segments.

Infrastructure
• Key activities: the most important activities required that in order to be able to deliver the value proposition.

• Key resources: it includes the essential assets required to operate the business and the most important resources to underpin the company be able to deliver the value proposition to customers and maintain the customer relationship.

• Key partners: the most important partners required that in order to be able to deliver the value proposition.

Finances

• Cost structure: all the costs of operating a business model, which including fixed costs, variable costs, economies of scale and economies of scope [Kuo, 2014].

• Revenue stream: the income of a company that was created by different customer groups. There are two mainly types of revenues: transactional revenue (one-time deal) and recurring revenue.

These nine building blocks can be divided into two groups as follows:

The first group that related to value creation: value proposition, customer segments, customer relationships, channels and revenue Streams.

And the second group that related to internal efficiency: key Activities, key resources, key Partnerships and cost structure.

2.3.4 Network Pictures

A network picture was used to visualize the case company’s business process in the future scenario and elaborate the activities between each participant as well as the interaction between case company and external world in the new business model. This network picture used in this thesis was inspired by ARA (actors, resources and activities) model which has been developed by Haakansson [1990]. The ARA model including three components and the relationships between each component are shown in figure 2.3. This model describes actor bond, activity link and resources ties and their corresponding inter-organizational couplings [Prenkert, 2013]. The ARA model aims to capture or illustrate views that specific actors have of the network environment within which they operate [Ford & Ramos, 2006]. In the ARA model, actors can they interact with each other in their relationships in different ways. And
activities and resources in two different relationships can complement or compete each other [Haakansson & Snehota, 1989].

Figure 2.3 ARA model (Haakansson 1990)
3 Literature Review

3.1 Data Mining Definition

Simply stated, data mining refers to extracting knowledge from large amounts of data [Han et al., 2011]. The knowledge including hidden relationships, unknown patterns and potential trends, and the knowledge can be applied widely in some field - such as building decision support models, generating predictive decision making methods. The data mining technology is usually used by business intelligence, biological sciences, medical science and engineering field [Tan et al., 2005].

3.1.1 Data Mining Technical Definition

Data mining is a process of extracting the hidden, unknown and potential valuable knowledge and information from the massive, incomplete, and random data [Han et al., 2011]. From a technology perspective, the term “data mining” contains several layers of meaning:

(1) The data source has to be authentic and abundant;

(2) The purpose of data mining is discovering user interested knowledge;

(3) The discovered knowledge is acceptable, comprehensible and utilizable;

(4) The discovered knowledge can only be used in some special situations without having to apply to all situations [Su, 2006].

The result of data mining is to discover knowledge and the knowledge usually refers to the relationships, correlations and patterns in the domain of data mining [Su, 2006]. The data analyst takes the data as the basis to extract the knowledge from database, the raw data can be structured - such as data in a relational database; it also can be the semi-structured - such as text, image and graphics data; it even can be special-shaped structured [Cochran, 1999]. The knowledge can be extracted by using different ways - such as mathematics-based methods, deductive reasoning methods and inductive reasoning method. The knowledge also can be applied in a variety of fields - such as information management, query optimization, decision supporting, process control and data maintenance [Linoff & Berry, 1997]. Thus, data mining is an interdisciplinary technology which containing database management, artificial
intelligence, machine learning, mathematical statistics, data visualization, parallel computing, etc. It provides a powerful ability to process a huge amount of data, which elevates the data application from a basic level to a wider field. It is important to note that data mining is not going to discover new natural sciences theorems, mathematics formulas or mechanical theorem proving [Gorunescu, 2011]. In fact, the discovered knowledge is applied in specific field and solving particular problems. All the discovered knowledge is relative which is constrained by a particular premise.

3.1.2 Data Mining Commercial Definition

With the development of information technology and the increase of customer requirement, the application of data mining technology in the commercial area is increasingly broad, especially in the banking industry, the telecommunications industry, the insurance industry and the retail industry [Battiti & Brunato, 2011]. From a business standpoint, data mining is a new business information processing technology. Its essential feature is using model-based methods to process a large number of business data in the database, and extracting valuable knowledge from database [Choudhary et al., 2009]. In other word, data mining is a kind of deep level data analysis method. With the popularity of information technology in various industries, companies of various industries produced a large amount of data but most of these data produced are not for a purpose of analysis, they are just the result of the operation of the business process [Su, 2006]. Analysis of these data can help to discover valuable information to support business decisions in order to gain more profit or reduce cost. Examples of the commercial application of data mining are conducting market analysis to identify new product, uncovering the root cause of manufacturing problems and profiling customer needs to acquire new customers [Tian, 2004].

3.1.3 The Tasks of Data Mining

Data mining aims to discover different patterns from a vast amount of data and there are kinds of data patterns can be mined. Thus, there are different data mining tasks are used to find the specified patterns. In general, the data mining tasks can be divided into two categories: predictive tasks and descriptive tasks [Kamath, 2009]. The predictive mining tasks are used to perform inference on the existing data in order to make predictions. The descriptive mining task refers to characterize the general attributes of the data in the database [Gargano & Raggad, 1999]. Essentially, the data mining task and the types of mining data is associated. Some mining tasks can only be used on a particular data type, but some of them can be used in a variety of data [Lee & Siau, 2001]. Data mining generally has the following four main tasks: predictive
modeling, association analysis, cluster analysis and anomaly detection [Pang, 2005]. The predictive modeling includes two sub-tasks classification and regression. The classification is used to predict discrete variables and the regression is used to predict continuous variables [Fayyad et al., 1996a]. The details and related method of the first three mining tasks will be discussed in a later section. The anomaly detection task identifies the data which has characteristics significantly different from other data [Pritscher & Feyen, 2001]. It usually be used to uncover fraudulent usage of credit cards by detecting purchases of extremely large amounts for a given account number in comparison to regular charges incurred by the same account [Kwong & Fan, 1999].

### 3.1.4 The General Structure of Data Mining System

Data mining is a technology which combined multidisciplinary knowledge and approach into a completed system. Basically, a typical data mining system may contain the following major components (see Figure 3.1) [Han et al., 2011].

![Figure 3.1 A Typical Data Mining System Structure (Han et al., 2011)](image)

The database, data warehouse or other information repository is one or a set of database, data warehouse or any other kind of database. It is the raw data sources of the whole data mining system. According to the data mining requirements, database or data warehouse server is used to extract relevant data from the database or data warehouse [Zbigniew, 1996]. Data mining engine is the core part of the data mining system. A set of functional modules compose the data mining engine that each functional module consists of several data mining algorithms and rules [Yaginuma, 2000]. Knowledgebase stores the domain knowledge that is used to guide the data
mining process, and provides the required information for pattern evaluation module [Hjørland & Albrechtsen, 1995]. Pattern evaluation module validates the results of data mining that generally interact with data mining engine in order to help the mining process focus on valuable patterns [Liu & Guo, 2002]. The graphical user interface communicates between the data mining system and the user that allows the user to interact with the system by specifying a data mining query or task. The user also can browse database schemas or data structures, evaluate mined patterns, and visualize the patterns by using the interface [Witten & Frank, 2005]. The data mining structure also states the general data mining process; a typical web mining processes model will be presented in the section 3.2.4.

3.2 Web Mining

Web mining is the application of data mining technology in web environment, but it is not only a simple application of traditional data mining. Web mining has its unique characteristics which are different with traditional data mining. Compare with the general data mining, web mining has the following characteristics [Su, 2006]:

1. The source of web data mining is available internet-related data which including text, image, video, web link, hyperlink, log file and user profile, etc.

2. The above mentioned internet-related data is ruleless which need to be screened, scrubbed and converted.

3. These data must be processed based on the corresponding features of themselves and then exploited by targeted methods.

Generally speaking, web mining is mainly applying to perform web data analysis. The most important source of web data is the World Wide Web. World Wide Web (also known as Web or WWW), is an information system of interlinked hypertext documents that are accessed by internet. In this system, each useful resource is called a data object and every data object has its own ‘uniform resource locator’ (URL) [Lee & Fielding, 2005]. The data object is transmitted to the user through the hypertext transfer protocol and then the user receives the data object by clicking on the link (URL) [Wang & Zhou, 2003]. World Wide Web as an important data source cannot be ignored, it provides an abundant, worldwide online information services. The data objects are closely connected to facilitate interactive access in the World Wide Web. Users seeking information of interest traverse from one object via links to another. Such systems provide ample opportunities for data mining [Han et al., 2011]. It increased the difficulty of extracting information in a network environment. Based on this, data mining has been increasing applied in the internet environment, thus web
mining is becoming increasingly important for internet companies [Liu & Yu, 2009]. In figure 3.2, the architecture of WWW is presented. This figure illustrates the workflow between the user request and the server responses on the WWW.

In general, WWW consists of three parts: the client, proxy server and web server. We can collect the user’s online data from all the three parts, but the data collected from different part has different features [Khare & Jacobs, 2004]. Web server data includes the content of a website, the hyperlink’s structure, logbook data, user registration data and cookies’ data. The web server will record user request into logbook while the server respond user request. Client side records the whole online browse data of the single user and these data are stored in the user terminal - such as laptop and mobile. Between the two parts is the proxy server which receives the user request from client side and returns the relevant page to client side from web server [Su, 2006]. During this process, the proxy server may store parts of website content and user request data [Server.zzidc, 2014].

Figure 3.2 the Architecture of WWW
3.2.1 The Data Sources of Web Mining

A web page mainly contains three types of data: web content, link structure and web log [Spertus, 1997]. These three types of data are the primary data sources of web mining.

Web content refers to the content on the web page which used for user browse. There are many kinds of data on the web page - such as text, images, video and audio, but the text is generally treated as the main data source of web page [Srivastav & Cooley, 2000]. Although the image, video and audio data contains a lot of useful information, but due to multimedia analysis technology is not mature, so the text data is currently main the data source of web content.

Link structure is a descriptive data used to organize web content, mainly referring to the structure of hyperlinks between pages, including hypertext markup language (HTML) and extensive markup language (XML) tags within the page [Deshmukh & Garg, 2015]. The link structure of the web is composed of web pages and page-based hyperlinks. These structures are very useful and important resources that they reflect the domain knowledge of web designers, at the same time they provide a great help for the accurate analysis of the web page [Zhu et al., 2015]. There are two basic web link structures: line link structure and star link structure. As shown in figure 3.3, the line link structure is a one-to-one link structure that only one link between every two pages. On the contrary, the links of star link structure is a one-to-many link structure.

![Figure 3.3 Two Basic Web Link Structures](image-url)
The web log is the user usage data that reflects the user’s browsing behaviors - such as, IP address, browsing time, HTTP referrer [Goel, 2013]. Web log includes web server log, proxy server log and issue log, see an example of web server log in figure 3.4.

<table>
<thead>
<tr>
<th>IP address / Date</th>
<th>Time / Timezone</th>
<th>Request Methods</th>
<th>URL</th>
<th>Request Protocol</th>
<th>Status Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>202.119.32.6</td>
<td>28/Jul/2003</td>
<td>GET</td>
<td>/http</td>
<td>1.1</td>
<td>200</td>
</tr>
<tr>
<td>202.119.32.6</td>
<td>28/Jul/2003</td>
<td>GET</td>
<td>/mail</td>
<td>1.1</td>
<td>302</td>
</tr>
<tr>
<td>61.115.107.129</td>
<td>30/Jul/2003</td>
<td>GET</td>
<td>/adv.php</td>
<td>HTTP</td>
<td>1.0</td>
</tr>
<tr>
<td>61.115.107.129</td>
<td>30/Jul/2003</td>
<td>GET</td>
<td>/send.php</td>
<td>HTTP</td>
<td>1.0</td>
</tr>
<tr>
<td>202.119.32.6</td>
<td>03/Aug/2003</td>
<td>GET</td>
<td>/script/root.exe /c+dir HTTP</td>
<td>1.0</td>
<td>404</td>
</tr>
<tr>
<td>202.119.32.6</td>
<td>03/Aug/2003</td>
<td>GET</td>
<td>/MSADC/root.exe /c+dir HTTP</td>
<td>1.0</td>
<td>404</td>
</tr>
</tbody>
</table>

**Figure 3.4 An Example of Web Server Log**

In the web server log, IP address refers to the client address that sending the request to the web server. Time stamp (date, time and time zone) indicates the time of web server receive the request. The request includes request methods, URL and requested protocols. Request methods are GET, POST and HEAD, GET acquires target object from the server, POST sends a request to the server and HEAD retrieves the targeted object’s HTTP. URL is a static file or an executable program in the server file system. Generally, request protocol means HTTP which used to receive the requested web pages. Status field is used to indicate the response situation of the web server, as shown above, the code 200 to 299 indicates that the request getting a successful response from the web server. Code 300 to 399 indicates the request need to be redirected and code 400 to 499 indicates the error request [Jafsoft forum, 2005]. Proxy server log and issue log have the similar format with web server log, the issue log is mainly used to record the error or failed request - such as request time out and permission problem.

**3.2.2 Web Mining Classification**

Based on the data analysis objectives, web mining can be divided into three different types (as shown in the figure 3.5): web content mining, web structure mining and web usage mining, the comparison of different types of web mining as shown in table 3.1.
Web content mining: Web content mining refers to the discovery of knowledge from web page content. Its purpose is to realize the automatic retrieval of the web resource in order to improve the utilization rate of the web resource. Web resources are widely distributed on the Web which including File Transfer Protocol, Gopher, Digital Library, Electronic Commerce Website and numerous of invisible private data and dynamic query results. The forms of web resources are also varied, the forms of web resources are constituted by many webpage elements - such as text, image, audio, video, hyperlink, the directory structure of a website, user profiles.

Web structure mining: Web structure mining discovers the potential patterns from hyperlinks of a website. It analyzes the links of a website to build the hyperlink structure model. Web structure mining can be used for web page classification, web page correlation analysis and similarity analysis and authority site recommendation. The representative tools of web structure mining are PageRank algorithm and hyperlink-induced topic search (HITS) algorithm. PageRank algorithm is an algorithm used to rank the websites according to the hyperlink structure [Altman & Tennenholtz, 2005]. The PageRank algorithm is used by Google search and developed by one of the founders of Google Larry Page [Page, 1998]. The HITS algorithm is a link analysis algorithm also can be used to rank website which developed by IBM CLEVER project [Liu, 2009].

Web usage mining: Each information resources providing server has a structure record set called web access logs on the WWW. Whenever the server receives a request for access to resources, the server will record and store the user interaction data. Analysis of different web access logs can help to understand user behavior and website structure, in order to improve the website structure and to provide personalized service for users. Web usage mining can be divided into two types: general access pattern tracking and customized usage tracking. General access pattern tracking analyzes web log to understand the user access patterns and tendencies, in order to improve web structure; customized usage tracking analyzes the individual
user’s preferences and uncovers user’s access patterns. And these patterns determine which of customized services will provide to individual users.

Table 3.1 the Comparison of Different Types of Web Mining

<table>
<thead>
<tr>
<th></th>
<th>Web Content Mining</th>
<th>Web Structure Mining</th>
<th>Web Usage Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Data Source</td>
<td>Text document</td>
<td>The link of website</td>
<td>Serverlog</td>
</tr>
<tr>
<td>Data Characteristics</td>
<td>Unstructured/semi-structured data</td>
<td>Structured data</td>
<td>Interactive data</td>
</tr>
<tr>
<td>Mining Methods</td>
<td>Classification</td>
<td>Link analysis</td>
<td>Association rules mining Clustering</td>
</tr>
<tr>
<td>Main Algorithms</td>
<td>Decision tree algorithm</td>
<td>PageRank</td>
<td>Apriori K-means Clustering</td>
</tr>
</tbody>
</table>

3.2.3 The Main Data Mining Methods and Algorithms Used in Web Mining

Web mining evolved from data mining, thus the web mining inherited the majority of data mining methods. The main data mining methods used in web mining are classification, association rules and clustering [Liu & Yu, 2009]. In addition, compared with data mining, web mining has its unique features because there is no link structure in the relational database [Liu & Yu, 2009]. Thus web mining has its exclusive methods to perform the link mining task - such as link analysis [Liu, 2002]. In the following section, these methods and relevant algorithms have been presented.

Classification (Supervised learning)

Classification is a widely used method in data mining, it is also known as supervised learning or inductive learning [Liu & Yu, 2009]. The aim of classification is to maps diverse data items into several predefined categories [Han et al., 2011]. And it is called supervised learning because of the categories are clearly defined and this is compared to unsupervised learning which is described in the next section. Classification is often used to discover the commonality of data from the database in order to identify unknown data attribution or classes [Gehrke & Ramakrishnan, 2000]. In the data mining tasks, classification can be used for descriptive modeling and predictive modeling. Descriptive model is an interpretative tool to distinguish
different data’s classes and the role of the predictive model is to predict the classes of unknown data [Kosala & Blockeel, 2000]. Classification is well suited to describe or predict the nominal dataset, but it is less effective for ordinal dataset (for example, classification cannot sort different data classes into a sequence), because the classification does not consider the order relation between different data classes [Tan et al., 2005]. Classification applies to almost all the fields, including text and web fields. For web filed, using classification will help to divide user profile into the targeted user category [Choudhary et al., 2009]. It can help analyst to establish specific user overview profile, in order describe the characteristics of the user [Su, 2006]. In general, the fundamental process of classification (see in figure 3.6) contains two steps: training step and testing step. In the first step, a learning algorithm uses the training data to create a classification model. And the model uses the testing data to create new data. The new data as the result is used to evaluate the accuracy of the classification model. If the accuracy achieves the expected result, the model can be applied in the practical case [Pang et al., 2005].

![Figure 3.6 the Fundamental Process of Classification](image)

The one of most common technique used in classification is the decision tree and the corresponding algorithm has great advantage compared with other algorithms, especially in the accuracy and efficiency aspects [Liu & Yu, 2009]. The classification model was generated by these algorithm is shown as a tree which called decision tree. An example of decision tree was shown in figure 3.7. The decision tree includes two types of nodes, decision nodes (which in the block) and leaf nodes (i.e., sunny). Each decision node in the tree specifies a test for some attribute of the instance and each leaf node indicators a classification. The testing of decision tree starts from the root node (outlook) downward toward to the base (Yes/No). If a test sample finally reaches a leaf node, thus the class of this sample can be regard as the class of the leaf node [Liu & Yu, 2009]. The main decision tree algorithms are classification 4.5 (C4.5) algorithm and classification and regression tree (CART) algorithm [Wu et al., 2009]. Compared with C4.5 algorithms, the CART algorithm has higher efficiency. Because the CART algorithm always divide the sample set into two sub-sample sets, thus each decision node only has two branches. Therefore, the decision tree generated by CART algorithm is binary decision tree structure [Ning, 2009]. It contributes to the analyst to evaluate the results.
Clustering (Unsupervised learning)

Classification discovers the relational pattern between the data attribute and its classes attribute, and using the pattern to predict the classes attribute of unknown data items. And these data items usually represent predictive or classification tasks in the real world. However, in some cases, the classes attribute of data items is nonexistent [Liu & Yu, 2009]. It needs to explore the data to discover the intrinsic attribute of data items. The clustering is one technique to discover the intrinsic attribute of data. It intends to gather the data items which shared similar characteristics into similarity groups, and these groups called clusters. The data in the same cluster have similar attributes as well as the data in different cluster have different attributes. The clustering task aims to discover the different clusters which are hidden behind the data [Zhang, 2009]. The results of clustering can be used to help company changing marketing decision, building customer segmentation model or managing the document structure.
The figure 3.4 presents an example of clustering result. It is a visible example of clustering result. This is not an exact clustering result, just use to explain the concept. Assumptions, the clustering result generates 3 groups, members in each group have similar attribute. The member in these groups could be website users or web pages. If they are pages, the group A refers to pages which have high ad clicks, and the group C refers to pages which have low ads click. Then the data analyst can analyze these pages in group A and group C to find out the reason that cause pages in group A have higher ad clicks than pages in group C. Furthermore, the designer can improve pages in group C by these findings [Liu & Yu, 2009].

The fundamental of clustering is using distance function and corresponding algorithm to calculate the similarity of two data points and then generate different clusters. The k-means algorithm is one best known clustering algorithm. Its simplicity and efficiency perhaps makes it become the most widely used clustering algorithm. The principle of k-means algorithm is to preset the amount of clusters, and the algorithm using distance function group data points into clusters [Fayyad et al., 1996b].

Association rules mining

Association rules mining discovers the meaningful relationship which were hidden in the large database set, and all the objects in these relationships are unordered [Tan et al., 2005]. It is often used to discover the potential relationships and patterns, these
relationships or patterns called co-occurrence relationships, also as known as associations. These associations represent the causal relationships between data sets. It describes rules and patterns that certain attributions simultaneously appeared in a dataset. In other word, this method explores valuable connections between different data objects in a same dataset. Its main function is to analyze the probability that multiple data targets simultaneously occur in the same event [Cooley & Tan, 1999]. For an example, if A and B are two data targets in a same event, we can analyze the probability of A and B simultaneously occurring. The association rules mining is widely used in various fields - such as, business, internet security, education management system, mobile communication [Zhou & Xie, 2001]. Generally, association rules mining does not require specific data type, various types of data can be mining by this method [Yu & Aggarwal, 2001]. The famous application of association rules mining is the market basket analysis, it helps the market manager find the relationship between different products customer bought in the market [Gao, 2008]. An example of association rule is

\[
\text{Diaper} \rightarrow \text{Beer} \; [\text{support} = 15\%, \text{confidence} = 70\%]
\]

This association rule indicates that there are 15% customers bought diaper and beer at the same time. And for all the customers who bought diapers, 70% of them also bought the beer [Stanley, 2015]. The market manager uses this association rule to set out a new sales strategy so as to improve the product sales. For instance, according to the finding they can adjust the position of the goods (move the beer closer to the diaper). The method also used in the web usage mining for click streams analysis in server logs, in order to visualize the website structure and user behaviors [Zhang et al., 2013]. The common use algorithm of association rules mining is Apriori algorithm. It contains two step, one step for define the support indicators, and another for define the confidence indicators [Liu & Yu, 2009].

**Link analysis**

Link analysis is the exclusive task of web mining. The most important purpose of link analysis is to make up the shortage of classification techniques which applied in the search engines [Liu, 2002]. With the increasing amount of web page on the internet, that leads to classification techniques has no longer effectively support the development of search engine. During the period of 1997 to 1998, there are two important link analysis algorithms had been designed: PageRank algorithm [Brin & Page, 1998] and HITS [Kleinberg, 1999] algorithm. And the PageRank is used by Google search engine [Brin & Page, 1998]. Both of PageRank algorithm and HITS algorithm utilize the link structure of web page to analyze the relationship between different pages. And according to the prestige of webpage or the authority level to classify and rank the webpage [Linoff & Berry, 2002]. In this case, the prestige and
authority level refer to the frequency of link clicks of the webpage. The degree of prestige and authority level means that the value of the webpage [Liu & Yu, 2009].

### 3.2.4 Web Mining Processes Model

Web mining is the application of data mining technology in the web environment, thus, the web mining processes and data mining processes have something in common. The difference between them is the processing object and method (algorithm). Based on the general data mining process and the characteristics of web data, web mining can be divided into five steps [Kosala & Blockeel, 2000].

![Figure 3.9 the Web Mining Process Model](image)

- Data sampling

Web can provide the data source including web page data (text, image, multimedia), hyperlink data, and web server logs. Data sampling extracts a subset of data related to exploration target from a large amount of data, this subset of data are providing materials and resources to support the rest of web mining processes. Specifically, in this step, a targeted sub dataset is extracted from a large number of web pages and the unrelated web pages will be filtered in order to ensure that all the sampling data are associated with the mining object. It can reduce the workload of data preprocessing as well as ensure the quality of the data. Then the dataset will be reviewed and analyzed to analyze whether it is suitable for the establishment of a data model [Fayyad et al., 1996b].
• Data preprocessing

Data preprocessing organize a variety of data into structured data which can be used for web mining. The results of data preprocessing directly affect the final results of web mining, so that data preprocessing is the key factor of web mining quality assurance. The main tasks of data preprocessing are data cleaning, data integration and data conversion. The intention of data cleaning is to eliminate the irrelevant data items from the sampling data. Data integration is used to integrate and classify the cleaned data and then the data will be converted into standardized data format according to the mining requirement [Florin, 2011].

• The establishment of web mining model

The main purpose of establishment of mining model is to extract potential, acceptable and valid rule and pattern from the result of data preprocessing. It is the most important step in the process of web mining. The mining objective and the characteristics of data determine the mining method used to establish the mining model [Kosala & Blockeel, 2000]. There are four major methods of web mining model, which are classification, clustering, association rules mining and link analysis. The main difference between web mining process and traditional data mining process is the mining method used in the mining model. For an example, the link structure analysis is the unique task of web mining, and its corresponding mining model will only appear in the process of web mining [Kroeze et al., 2003].

• Analysis and evaluation

Analysis and evaluation is an important step of web mining, it selects the discovered rules and patterns to convert into specific knowledge, and then uncovers valuable knowledge by pattern analysis. Moreover, the validation and reliability of the mining result need to be evaluated in this step. One evaluation way is using new data to evaluate the result in the practical environment [Jones & Gupta, 2006].

• Knowledge visualization

Knowledge visualization refers to deliver web mining results (rules, patterns or relationships) with the appropriate forms to user, in order to user acceptance and utilization. In other words, it uses visualization techniques to present the user interested knowledge by a graphical way.
4 Result

This section presents the analyzed results from the methods. It can be divided into the following three sections, 1) Literature Review, 2) Industry Analysis and 3) Case Study.

4.1 Industry Analysis

4.1.1 The Development of Weather Service Industry in Internet Environment

The weather forecast is the essential information which influences every aspect of human daily life. In the modern world, the progress of technology allows people to get information in diversified ways, among them, the weather service websites way is booming in last decade [Weather Forecast Development, 2010]. There are several facts that can prove the potential of weather service industry development in the internet environment. Compared with a decade ago, the total numbers of websites in the world have increased by 1800%, and the size of internet users have grown from 90 million to 3 billion [Internet Live Stats, 2015]. More and more people choose to obtain information from the internet, and an increasing number of companies start doing business online. These facts show that the internet provides a lot of opportunities and possibilities for the weather service industry. The weather forecasting system consists of several different components and these components can be divided into three parts: manpower resource, infrastructure and technology. The manpower resource refers to the staffs of online weather services company. One of the most important manpower resources is the meteorologist who is responsible for processing and analyzing raw weather data. Because the raw weather data is generated by computer simulation results and sometimes these data are not entirely accurate. The meteorologist needs to reprocess these raw data by their specialized knowledge and experience to provide user accurate weather forecast [Knowledge @ Wharton, 2013]. The infrastructure includes the weather station and the devices used to capture weather data. The technology used in online weather services mainly includes radio engineering, remote sensing technology, and information technology [Glickman, 2014].
4.1.2 Trends of Online Weather Services

The progress in science and technology allows the entire weather service industry develops diversely, and the internet has promoted the process progress greatly. Based on the web review results, the author found the following factors might drive the development of online weather service.

*Localized and detailed forecasting product*

The weather forecast is an application of meteorological science and technology to predict the weather condition, in particular, areas [Infoplease, 2002]. Nowadays the weather forecast cannot achieve 100% accuracy, but the accuracy of weather forecast has gradually improved in the last decades [Zastrau & Elsner, 2015]. In this context, the accuracy of forecasting is the trend of the entire weather service industry. But the trend of the development of weather forecast technology is not the focus of this study, so it will not be discussed. The emphasis is placed on how the existing technology can improve the accuracy by providing targeted product. The internet as a flexible and efficient medium is beyond the limitation of time and space, and essentially changes the broadcasting way of traditional forecasting. For this reason, the internet-based weather service can provide targeted products applied to small and medium size geographical system. On the one hand, the weather in different places in the same area may vary greatly, so that the localized forecasting will effectively improve the forecast accuracy. On the other hand, the conventional weather forecast has been unable to meet the growing needs of customers. Customers not only want to know the numerical value of weather parameters (e.g., temperature, precipitation). They desire more details can help them to understand and use these parameters. Take precipitation as an example, customers may know it will rain today, and go out with an umbrella [Fei, 2015]. But they do not know which value of precipitation means the rain will influence the traffic or causing a sports game canceled [Fei, 2015]. Therefore, the companies should pay attention to the details of forecasting in order to improve their competitiveness. In addition, time variation has a great influence on the accuracy of the forecasting, the real-time and continuous forecasting product has huge development potential and good application prospect [121 Net, 2011]. Today, many internet-based companies already noticed the potential of localized and detailed forecasting products [Chen & Qian, 2015]. But the human cost restricts the development of localized and detailed forecasting product. Because the accuracy of localized and detailed forecasting product depends on the analysis of the meteorologists, small companies do not have enough meteorologists to support it.
**Customized service and intelligent website**

For companies, the internet provides a timely and effective way to offer novel personalized service and to increase customer experience value for customers. The weather forecast closely correlated with customer’s daily life and work, and there are different needs and feelings of different customers. The standardized product and service help company to control the quality and to manage the resource consumption, but in the certain sense it ignores the specific needs of customers and their psychological feelings [Feng, 2004]. In addition, there is no geographical barrier in the internet environment. The internet-based weather service company has ideal conditions to develop customized services. And it has a positive influence to enhance the competitiveness of the company. For online weather service, the customized service might be:

Customized content refers to users can choose the specific weather content what they want to see or eliminate the content that they do not need. Today, the online weather service provides the user dozens of weather contents/parameters and these contents/parameters usually appear in a same page/screen. The users have to take the time to find out the information they needed, this reduces the efficiency and influence on the use experience. Thus, the customized content would be helpful to improve this problem.

Customer habit-based push service is a flexible way to push the weather service to users. For some weather service, the user can subscribe the weather notification service and receive the weather forecast automatically. But this service pushes the weather forecast in the regular time and way (e.g., on a Monday morning and by email) that the user cannot choose when and how to receive the weather forecast. The customized push service could be a useful way to solve this problem, the user could customize the receive time and way according to their daily routine and habit instead of change the activation and deactivation of push service every time.

The customized service must have appropriate technology as support, the most important technology is computer technology. In order to provide high-quality customized service, the companies needs to use advanced information technology to understand and capture the customer needs, at the same time, to establish special user interface and database system to manage user information [Wang, 2004]. The intelligent website is a website that can adapt user’s activities and information and build a model of user activity by analyzing the user usage data and profile data. Based on this, the website provides the specific content/information/presentation of information to the specific user in order to better address user’s needs [Brusilovsky et al., 2007]. These factors led to the development of online company move oriented to more intelligent way. On the one hand, the intelligent website help company to deliver their services in a better way, the customer can receive the services which they
expected automatically and they can get personalized system service (push services, advertising services, etc.). On the other hand, the intelligent website is an effective way to decrease the website operation costs as well as to save human resource costs.

**The web development for mobile devices**

The biggest change in the Web development over the past few years has been a marked increase in the mobile computing [Web Development, 2013]. The result of the statistical in 2013 highlights a few key statistics [Hepburn, 2013]:

- 91% of all people on earth have a mobile phone
- 56% of people own a smart phone
- 50% of mobile phone users, use mobile as their primary internet source
- 80% of time on mobile is spent inside apps
- 72% of tablet owners purchase online from their tablets each week

Mobile devices took up 55% of the internet usage in the United States in January 2014. The mobile applications made up 47% of the internet traffic and 8% of traffic came from mobile browsers [CNNMoney, 2014].

![US time spent accessing the Internet by device](image)

**Figure 4.1 US Time Spent Using the Internet by Device (comScore, INC, 2014)**

Web development for mobile devices is an unavoidable trend of an internet company development. Mobile devices have a great potential, but they also have some limitations. For instance, the slower JavaScript engine and less memory mean that the same Web page that runs quickly and smoothly on a desktop might be quite slow on a
mobile device. So it had to pay more effort when developing Web sites both for the
desktop and the mobile [Queue.acm.org, 2013].

4.1.3 Key Indicators of Website Design

There is a mass of user usage data are recorded and stored in web logbook and these
data can be transformed into different indicators to help website designer to judge the
quality and value of a website. These key indicators can be used to redesign the
website, track user’s movement and improve the user experience of the website.
According to different types of website, the relevant key indicators are different.
Based on the web review results, the author summarized applicable indicators for the
purpose of judging the quality and value of a weather forecast website, as shown in
Table 4.1. In this table, the key indicators including two categories of indicators: traffic
indicators and business indicators. The traffic indicators refer to the indicator which
can be used to measure the visitor’s activity. By analyzing these indicators, the
website designer can understand the current situation of the website and develop
improvement strategy [Lan, 2011]. Business indicators refer to the ratio of visitor
activity turn into business profits or the website promotion effect on the internet
[Song, 2012].

Table 4.1 Key Indicators of Web Design

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic</td>
<td></td>
</tr>
<tr>
<td>User Dwell Time</td>
<td>The time the user spent on the web page.</td>
</tr>
<tr>
<td>The Number of Register Member</td>
<td>The overall number of registered members of a website.</td>
</tr>
<tr>
<td>The Number of Active Users</td>
<td>The users who will use the website in a period of time. (Daily or monthly)</td>
</tr>
<tr>
<td>The Month Growth Rate of Membership</td>
<td>The membership growth rate of a website in every month.</td>
</tr>
<tr>
<td>The Number of Daily Log</td>
<td>The average number of users logged in daily.</td>
</tr>
<tr>
<td>The Number of membership published</td>
<td>The number of the information/file/article posted by membership daily.</td>
</tr>
<tr>
<td>The Number of Daily Update</td>
<td>The amount of the updated contents of a website daily.</td>
</tr>
<tr>
<td>The Number of Core Content Update</td>
<td>The amount of the updated core contents of a website daily.</td>
</tr>
<tr>
<td>Average Session Duration</td>
<td>The average time of each session of a website in monthly.</td>
</tr>
<tr>
<td><strong>Average Page Depth</strong></td>
<td>Average page depth is the average number of pages viewed during a session.</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Unique Visitors</strong></td>
<td>A person who visits a site at least once within the reporting period.</td>
</tr>
<tr>
<td><strong>Page Views</strong></td>
<td>A page view is a request to load a single HTML file (web page) of a website.</td>
</tr>
<tr>
<td><strong>Repeat Viewer</strong></td>
<td>A repeat visitor is someone who comes to your website more than once. Repeat visitors are important as it means that your site keeps its value for your customers.</td>
</tr>
<tr>
<td><strong>Bounce Rate</strong></td>
<td>Bounce rate represents the percentage of visitors who enter the site and then leave rather than continuing on to view other pages within the same site.</td>
</tr>
<tr>
<td><strong>The Number of Total Sessions</strong></td>
<td>The overall number of sessions of a website in monthly.</td>
</tr>
<tr>
<td><strong>The Number of New Visitors</strong></td>
<td>The number of daily new visitors in a website.</td>
</tr>
<tr>
<td><strong>The Ratio of New Users</strong></td>
<td>The new users account for the proportion of the total number of users in monthly.</td>
</tr>
<tr>
<td><strong>Conversion Rate</strong></td>
<td>The percentage of visitors who take a desired action. (i.e., click the advertisements or enter the purchase page)</td>
</tr>
</tbody>
</table>

**Business**

| **The Number of Cooperative Website** | The number of the collaborators of a website. |
| **The Number of Media Partner**      | The number of the media partner of a website. |
| **Keyword Ranking**                  | The keyword ranking in search engines. |
| **Landing Page**                     | A landing page is a website page that allows you to capture a visitor's information through a lead form. |
| **The Number of Feedback/Interaction** | The number of user feedback in monthly. |

### 4.2 Customer Journey of a Weather Service Website

The general customer experiences of weather service website have been mapped on the customer journey map. To view the Customer Journey Map please refers to appendix A. In order to fully understand customer’s user experience from their perspectives, this journey map including the whole journey of the customer from
when the customer intends to obtain weather service through to leaving a weather service website. In this case, the target customer segment is the people who rarely or never used an online weather service. This target was selected because the author intends to discover the gap and barrier which will impact on the attraction of the website for new customers. In order to find out the way help the website attract more new customers. In addition, these people need to have the ability and conditions to use the internet. This customer journey of a weather service website includes eight specific steps. There are inspiration, searching and planning, decision making, information achievement, website exploration, registration, App/Software downloading and value-added product/service. The journey map result visualized the customer activities, feelings, and thoughts, also it highlights the important touch-points and the customer overall experiences.

![Figure 4.2 Steps of Weather Service Website Customer Journey](image)

### 4.2.1 General Phases of Customer Journey on the Weather Service Website

#### Inspiration

Generally, this step is the first step of the whole journey of weather service website. Customers have the desire to know the weather forecast in a certain period of time. In this step, customers will look for an available and suitable medium - such as television, radio, internet. It depends on the habits of customers and the majority of customers may already know which one is their idiomatic medium. Or sometimes customers may come up with an idea to get weather service before searching and planning step. Thus, the inspiration step is not always the first step of online weather service journey. Besides family members’ request, friends’ suggestion, commercial, environment or other factors may influence customers’ behaviors in this step.

#### Searching and comparing

After the step of inspiration, the next step is searching and comparing. In general, this step only involves the customers who choose to get weather service online. The customers may explore several options and according to the searching results to find out the most optimal alternative. The searching results are largely dependent upon the
search language and key words when the customers use identical search method. (The same search engine or the same browser) The process of comparing happens quickly; customers will not spend a lot of time to compare different alternatives. There are two reasons lead to this result, one because the information of searching results is limited that customer can hardly gain enough information to help them judge the quality of a website. Another reason is that customers are often reluctant to spend too much time on this process. The online weather service customers pay close attention to efficiency that it is an important factor to affect customers’ behavior in the process of comparing.

**Decision making**

Customers tend to the make decision after comparing different options, sometimes there is no an absolute boundary between comparing and decision making since they usually happen at the same time. Sometimes, customers may enter several websites during the comparing step and it will increase the efficiency of decision making. Although the final decision is formed in a flash, but the entire process of decision making starts from the moment when customers decide to use weather service online. In another case, if the customer has ever used the online weather service before, their habits and previous experience will play an important role in decision-making step.

**Information achievement**

The role of the landing page is extremely important for the whole website; it affects the customer’s first impression of the website. The customer would like to quickly and easily obtain the expected content in the landing page when they enter in the website. Moreover, most of website shows their core content and main services on the landing page and it also represents the overall style of a website. In many cases, this is the only chance the website can get to impress the visitor and keep them continue browsing. In the process of obtaining information, the customers may have a lot of feelings and thinking, but they may not generate a lot of data. The reason is customers may get the information directly without any click that will not create a new session. Thus, there is nothing can be recorded on the server logbook. In this case, the only indicator can be used to analyze customers’ behaviors is browsing time. For customers, this step also determines whether they will continue to browse this website. There are various factors may keep customers to stay on the website - such as the good appearance and style of the website, concise website structure and layout, attractive features content.

**Website exploration**

In this step, customers already achieve their primary purpose and they decide to spend more time on the website. The customer may discover useful or interesting things on
the website, and get a deeper understanding of the website. During this process, the 
customer will create a lot of sessions while they look around the website, and there 
are a lot of data are recorded on the server logs in the process of interaction between 
customers and website. The behaviors of customers can reflect a lot of things to help 
designers to enhance the website in this step.

Registration

Customers may become the member of the website after website exploration. For 
most of websites, customers usually need to create an account first in order to use 
extra products or valued-added services. The registration step is an effective and 
convenient way to collect customer information, and it also is an important step of the 
website to establish a relationship with the customers. Furthermore, the difficulty of 
the registration determines the success rate of user registration, generally, customers 
prefer simple registration process.

App/software downloading

Customers may download the website app/software in order to visit the website 
flexibly. Sometimes, it happens on the previous step if customers want to download 
the software of the desktop. Otherwise, customers may download the apps after them 
leaving the website. This step also is the first step of the customer journey of the 
app/software.

Value-added product/service

The value-added product/service can be classified into two types: free service and 
fee-based service/product. In many cases, the free service including push service, 
subscription service, and others. And there are many kinds of fee-based service and 
product - such as consulting, prediction, equipment leasing. No matter free or not, this 
step is an effective way to collect the customers’ feedback (by customer email and 
after-service), and the data (including customer profile data, usage data, and feedback) 
created in this step is valuable for understanding customers.

4.2.2 Important Touch-Points

Identifying the touch points is a necessary step to create a customer journey map, and 
it is an important way to improve customers’ satisfaction. A touch point is a contact or 
interaction between a website and potential customers or existing customers in the 
before, during or after use process [Spengler & Wirt, 2009]. Each step in the customer 
journey contains a number of touch points, they not only represent the interaction of
website with customers, but also reflect customers’ feelings and thoughts. These feelings and thoughts can be linked to the customer’s mental activities, which could help the author of this thesis learn how each interaction builds or destroy value in the customer’s eyes.

26 touch points were identified in the weather service website customer journey. (See figure 4.3) For different customers, these touch points may not all occur in their journey, but there will be at least one touch point in each step. Take the inspiration step as an example, some customers may come into contact with the website through the social media, others may be inspired by the people around them. (Word of mouth) And the same touch points may occur in different steps - such as response speed, but they may represent the different feelings or thoughts of customers.

![Figure 4.3 Important Touch Points of Weather Service Website](image)

According to the characteristic of the touch points, they can be classified into three groups: marketing, design, and technology. This classification reflects the influence degree of the touch points on the customer feelings. (As shown in figure 4.4) In this research, the author uses these touch points to identify where the highest-impact moments occur in the customer journey. These key points represent the critical steps that affect the customer experience in the whole journey. By analyzing these key points, the author recognized the potential key points which can be optimized through data mining technology. These potential key points are website structure, advertisement and the content of the website. There are three reasons the author picks these points as potential key points: 1. these points mainly occur in the use phase; the customer may evaluate the experience or make a critical decision in this phase. 2. The
A company can completely improve these points by their internal efforts. 3. These points and the web usage data has a strong correlation, the designer can evaluate and improve these points by analyzing the web usage data.

Figure 4.4 the Classification of Important Touch Points

Marketing group includes the important external factors that are involved in customer's journey of the website. It contains almost all the external environmental factors - such as the reputation of the website, the support of partners and the feedback of customers. It reflects the popularity of the website in the market. Furthermore, these touch points are most likely to affect the feelings of the customer.

Design group mainly contains the appearance and style of the website. Most design touch points have been concentrated in the process of customers browse the website. These touch points help customers to judge the visual experience of the website and also to help them evaluate the quality of this website. In order to retain customers, the website designers need to pay close attention to these touch points. The feelings of the customer are also easily affecting by these touch points. For example, customers very concerned about the position of the core content, they expect to see their expected content or information in the landing page. It almost determines whether the customer will visit this website again.
Technology touch points are the intangible objects that can be improved by technology. These touch points have a minor impact on customer feelings, but this does not mean that these touch points do not affect the customer’s feelings.

4.3 The Proposed Solution of Web Mining for Customer Experience Improvement

Nowadays, companies start to realize the potential of creating value from data - such as, some online video sites use the data mining results to discover the user’s preferences in order to guarantee the popularity of TV plays [Carr, 2013]. This fact implies the possibility to involve data mining technology more in the process of customers use the website. Besides acting as variables for quantizing customer information and recording customer activities, data could contribute more in online business. Based on the findings of customer’s journey, the web mining solution is proposed in this section that can contribute to improving customer’s user experience of the weather service website. Web mining provides effective technology to support the website designer to modify the website to better address the user’s needs. The following text elaborates how web mining as a technical method can be used to help the designer to optimize the website. In the context of weather service, web mining mainly contributes to recommendation bar and customized service. The more details of recommendation bar can be found in section 4.4.5 and the further discussion of customized service can be found in the section 4.4.6.

Web usage mining

The customer is likely to judge the quality of the website through the content of the website. The distribution of the content, the value of the content and the efficiency of the content acquisition greatly influence the customer’s user experience. The website offers targeted contents to users and constantly adjusts the content to adapt the user’s behavior pattern. It will effectively improve the user experience as well as attract the customer to stay longer on the website. The web usage mining provides a possible solution to help designers to achieve this goal.

For the online weather service, there are two tasks of the web usage mining might contribute to improving the customer’s user experience. There are visitor segmentation and customer correlation analysis. Both of these two tasks aim to discover the user’s behavior pattern and preferences in order to help the designer get insights into the customer’s user experience.
**Visitor segmentation:** the visitor segmentation is a common application of clustering technology used in the web usage mining. The clustering is a data mining technology which intends to gather the data items which shared similar characteristics into similarity groups. The objects of visitor segmentation are users’ behavior data which record and store on the website server. The designer imports these data into data mining software and performing related algorithm model to get the clustering result (groups). The clustering algorithm has been described in the literature review. Different groups refer to different types of visitor which have similar intrinsic attributes. The most important benefit of visitor segmentation is to reveal the internal relations between different visitors in order to discover shared characters of them. The designer can use visitor segmentation to provide the user of the same group (cluster) personalized content. The purpose of visitor segmentation is to build groups of users which exhibiting similar behavior patterns. These groups and user’s behavior patterns are especially useful for providing personalized web content to the users with similar interests. Furthermore, the clustering based on web usage data can be used to establish user community in order to reflect similar interests of users. The user’s behavior patterns may contribute to developing dynamic recommendations of web personalization service. For instance, the figure below presents a clustering result. The cluster 0, 1, 2 refer to three user group, and the A, B, C, D, E, F mean the page view. In the cluster 1, the entire user viewed the page B and F, 75% of them viewed the page A. Thus, we can get a user behavior pattern that the user view page B and F, the user is likely to interest in the page A. As a result, the system will recommend the page A to the user who has viewed page B and F (cluster 1). Based on this method, the website is able to provide users personalized content which accord with their interests and behavior (providing different contents to the users in the different cluster). Also, it can be used to improve the advertisement content that the website might base on the user’s interests to push corresponding advertisements (presenting different advertisements for the users in the different cluster). It can reduce the negative effect of advertisements for the users.
Customer correlation analysis: The customer correlation analysis can discover pages which are visited together or keywords which are searched together. The main method of customer correlation analysis is association rules mining which as mentioned in the literature review. The association rules mining could help the designer more effectively organize the website content. For an example (see the figure 4.6), there are two web pages A and B on the landing page of website C. The website designer does not provide the link to connect page A and page B, if the user visited page A and then the user wants to visit page B, the user has to return landing page first. Assuming that the customer correlation analysis sees that there is an association rule: $A \rightarrow B$, it implies the majority of users visit page A then they will visit page B. The designer can utilize this finding to offer the direct link from page A to page B. That will help the user get the expected information or product effectively.
Data mining software is the key tool for implementing data mining. In the data mining process, the data mining software is mainly used to establish the mining model for analysis and processing the data. The mining model is the foundation of data mining software which includes algorithm module, modeling module, and the processing module. The most commonly used data mining software can be divided into two groups: the classic suites software and the open source software [Wang, 2013]. The classic suites software commonly used in the business application, it provides the entire data mining required task module, including the data preprocess module, data mining module (diversified algorithm and relevant model), and analysis and evaluation module. Generally, the classic suites software is easy to learn and use, that the user can use it without the knowledge of programming and algorithm [Chen, 2012a]. But the classic suites software has shortages in flexibility and extensibility, the user cannot develop new algorithm or model. On the contrary, many open source software only contains data mining module and the knowledge of the programming is indispensable to the user [Chen, 2012a]. Compared with the classic suites software, the open source software is more suitable for academic applications - such as algorithm development, data model development [Wallace & Hobbs, 2006]. The classic suites software focuses on the analysis of data and the result evaluation and visualization. Thus, the classic suites data mining software is more suitable in the scenario that in order to support the improvement of customer’s user experience of the online weather service. Commonly used classic suites software are Angoss Knowledge Studio, IBM SPSS Modeler, SAS Enterprise Miner and StataCorp Stata.
This software has different characteristics and this thesis has not focused on the comparison of the software. Thus, the data mining software will not be further discussed in this section.

Table 4.2 the Proposed Solution of Web Mining

<table>
<thead>
<tr>
<th>Solution</th>
<th>Web Usage Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Source</td>
<td>Web Server Log:</td>
</tr>
<tr>
<td></td>
<td>1. Web server access logs</td>
</tr>
<tr>
<td></td>
<td>2. Application server logs</td>
</tr>
<tr>
<td>Types of Data</td>
<td>Usage data: the navigational behaviors of visitors, each behavior corresponding to an HTTP request.</td>
</tr>
<tr>
<td></td>
<td>Content data: the text file and media file of the website.</td>
</tr>
<tr>
<td></td>
<td>Structure data: the tree structure of HTML and XML.</td>
</tr>
<tr>
<td></td>
<td>User data: the user profile information.</td>
</tr>
<tr>
<td>Process of Web Usage Mining</td>
<td>Web mining process model (as mentioned in literature review)</td>
</tr>
<tr>
<td>Task</td>
<td>Visitor segmentation</td>
</tr>
<tr>
<td>Contribution</td>
<td>Customer correlation analysis</td>
</tr>
<tr>
<td>Method</td>
<td>Personalized Web Content</td>
</tr>
<tr>
<td></td>
<td>Better advertisement experience</td>
</tr>
<tr>
<td>Algorithm</td>
<td>K-means clustering</td>
</tr>
<tr>
<td>Tool</td>
<td>Data Mining Software</td>
</tr>
</tbody>
</table>
4.4 A Case Study in the Internet-based Weather Service Company

In this section, a case study regarding the opportunities of data mining in an internet-based weather service company is presented below. The details of the current business model of the case company are described, and the business analysis result is presented in this section as well. Afterward, the future business model of the case company is proposed in order to illustrate how data-mining technology can be used as a support tool for internet-based weather service company development, which might contribute to exploiting new profit channels.

4.4.1 The Background of Case Company

3Bmeteo is an Italian weather service company founded in 2002, as the statistical results of January 2015, from the perspective of the number of users 3Bmeteo is the second-placed weather service website in the Italian market. The main service channel of 3Bmeteo is the website (www.3bmeteo.com), and their main products are mobile apps and software. They also provide their weather data to their corporate customers, and they offer their weather service to the radio, local TV broadcasters and local and national newspapers. Currently, the main profit source of 3Bmeteo is advertising revenue, and there is a small number of customers pay for the fee-based service. In the below table, the characteristics of 3Bmeteo have been described.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>3Bmeteo</th>
</tr>
</thead>
<tbody>
<tr>
<td>The types of products and services</td>
<td>Weather forecast, weather news, consultation.</td>
</tr>
<tr>
<td>The pertinence of products and services</td>
<td>Public weather forecasting</td>
</tr>
<tr>
<td>Customers</td>
<td>Individual customers, other websites</td>
</tr>
<tr>
<td>Founding and operating location</td>
<td>Italy</td>
</tr>
<tr>
<td>Features</td>
<td>Application</td>
</tr>
<tr>
<td>Social networking services</td>
<td>3B+ community</td>
</tr>
<tr>
<td>For-profit or non-profit</td>
<td>For-profit</td>
</tr>
</tbody>
</table>
4.4.2 Business Analysis

In order to get an understanding of case company, the business analysis is essential. The business analysis is including stakeholder analysis, SWOT analysis, and interview. The stakeholder analysis is an efficient way to analyze individuals or organizations associated with the company’s interests. It provides a way to scan case company from the different perspectives. The stakeholders have a profound influence to the company’s value orientation. The SWOT analysis has been conducted to analyze and summarize the company current reality [Robèrt et al., 2010]. This analysis covers the company current situation both from the external world (the outside environment - such as trends of industry, market environment) and the interior world (the components of an organization - such as staffs, development strategy). The SWOT analysis is a useful method to understand the company’s advantages and disadvantages and to identify the opportunity the company has and the threat the company faced [Mind tools, 2013].

Stakeholder analysis

3Bmeteo as an internet-based weather service company, their main products are the website and mobile apps. The weather forecast closely associated with people daily life and social production activities as well as more and more people use online service. Thus, the stakeholders of 3Bmeteo cover a wide range of groups. This analysis identified six important stakeholders and they are shown in below table:
<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>Influence/Power</th>
<th>Interest</th>
<th>Potential</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>High power to promote the brand of 3Bmeteo.</td>
<td>High interests because 3Bmeteo provides good weather forecast module service for company customers.</td>
<td>Increase referral traffic.</td>
<td>Direct and regular.</td>
</tr>
<tr>
<td>Individual</td>
<td>High influence because the number of individual customers is an important way to attract advertisers.</td>
<td>Medium interest in free use weather service.</td>
<td>Attract new visitors.</td>
<td>Direct and irregular.</td>
</tr>
<tr>
<td>Advertiser</td>
<td>High power because it the currently mainly source of income.</td>
<td>High interests because 3Bmeteo has low bounce rate and considerable users.</td>
<td>Interaction with users.</td>
<td>Direct and indirect.</td>
</tr>
<tr>
<td>The media</td>
<td>Medium power on market promotion.</td>
<td>Medium interest due to post weather news to attract readers.</td>
<td>Can enhance company brand.</td>
<td>Indirect.</td>
</tr>
<tr>
<td>Suppliers</td>
<td>High influence because the weather forecast quality largely depends on weather data.</td>
<td>Low interest in weather forecast development and weather data sharing.</td>
<td>Key factor of product and service development.</td>
<td>Direct and indirect.</td>
</tr>
<tr>
<td>Competitors</td>
<td>Medium influence in maintaining existing users.</td>
<td>Low interest because the development of 3Bmeteo may promote the improvement of competitors.</td>
<td>Opportunities of cooperation.</td>
<td>Indirect.</td>
</tr>
</tbody>
</table>
The results of business analysis and interview provide information to generate the current business model. This business model canvas includes nine important aspects, which describes the current business model in a graphical way. The model is shown in figure 4.7.

![3Bmeteo Business Model](image)

**Figure 4.7 the Current Business Model of 3Bmeteo**

The value proposition is clear and definite for 3Bmeteo currently. They enable customers to gain accurate weather forecast and diversified weather services by providing the convenient and efficient website, and mobile apps. Moreover, the company provides customers the personalized weather product. (The details of current personalized weather service will be discussed in section 4.4.6) The purpose of 3Bmeteo is to deliver high-quality weather services for both individual customer and company customer all the time by continuous improvement.

The company can segment their customers into four groups: internet users, mobile device owners, companies and advertisers. In all these customers, the largest number of customers is internet users and the second is mobile device owners. Sometimes,
internet users and mobile device owners are overlapped. The mostly advertisers are local companies and Google AdSense\(^1\). The case company relates to customers with virtual means mostly. Thus, their main channel is the internet, which consists of web search engines, online app store, and partner’s website. In many cases, customers come into contact with the case company with search engines or online app stores. Besides, customers may get to the company through the website of 3Bmeteo’s company customers. In this business model, the case company built three customer relationships. The membership and community are used for maintaining relationships with individual customers. They also intend to build a good brand image by their community. 3Bmeteo connect with their company customers by technical support.

The key activities of the case company are website maintain and operation. In order to retain existing customers, 3Bmeteo invested the most resource in website maintain and operation. It contains website optimization, build new product/service and improve existing product/service. The company also pays close attention on weather product/service innovation to drive more traffic to its website. In consideration of the products that 3Bmeteo deliver to their customers are weather forecast and weather services. The quality of these products largely depends on the weather data, so that 3Bmeteo sees weather sources and weather station as key resources. Besides, the professional meteorologist plays a significant role in weather service business, thus meteorologist is actually the foremost resource that they have. The loyal and active users contribute to publishing interesting content and to attract other users, so that they are a key resource as well, take online video site as a reference, the active users upload much interesting content that make a lot of users spend more time on the website. Considering partners, the key partners of the case company can be divided into two categories. There are weather source suppliers (i.e., the European center for medium range weather forecasts) and company customers. In many cases, the company partners also played the role of the channel. The internet users may find 3Bmeteo from the website of company customers. Thus, the company customers have a double role in the business model.

For now, the advertising revenue accounts for most of the total revenue. There are only a few customers pay for the fee-based service. The costs structure of 3Bmeteo can be divided into three parts: data center costs, employees’ salary, and R&D costs. The most of the costs come from the data center (Amazon web services).

---

\(^1\) https://www.google.com/adsense/
**SWOT analysis**

The SWOT analysis mainly focuses on the company strengths, weaknesses, opportunities and threats as well as its comparison with competitors. In the following figure, the company’s strength, weakness, opportunities and threats have been enumerated, it presents the company current reality, especially how the companies influenced by the external environment.

### SWOT Analysis

<table>
<thead>
<tr>
<th>Helped to achieve the objectives</th>
<th>Harmful to achieve the objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Origin: Attributes of the Group</strong></td>
<td><strong>External Origin: Attributes of the Environment</strong></td>
</tr>
<tr>
<td>Concise website design and layout.</td>
<td>The increase of mobile device owners.</td>
</tr>
<tr>
<td>Support a variety of mobile devices.</td>
<td>The growth of customer personalized demand.</td>
</tr>
<tr>
<td>The number of users was the second placed in Italian market.</td>
<td>A broad market space of weather business.</td>
</tr>
<tr>
<td>Professional meteorologists.</td>
<td>The potential of mobile internet.</td>
</tr>
<tr>
<td>A small number of advertisement.</td>
<td>The development of other media.</td>
</tr>
<tr>
<td>Free for use.</td>
<td>The business has attractively low barriers to entry.</td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
<td><strong>Customers unwilling to pay for existing product.</strong></td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
<td><strong>Opportunities</strong></td>
</tr>
<tr>
<td>Unitary profit pattern.</td>
<td>The development of other media.</td>
</tr>
<tr>
<td>Cannot connect to customer directly.</td>
<td>The business has attractively low barriers to entry.</td>
</tr>
<tr>
<td>Lacking new way to retain users visit.</td>
<td>Customers unwilling to pay for existing product.</td>
</tr>
<tr>
<td><strong>Threats</strong></td>
<td><strong>Opportunities</strong></td>
</tr>
</tbody>
</table>

*Figure 4.8 The SWOT Analysis of 3Bmeteo*

The main outcomes of SWOT analysis highlighted following facts which in relation to 3Bmeteo current business. 3Bmeteo’s main business is focused on the internet and the most noticeable advantage of 3Bmeteo is their application support a variety of mobile device. They invested a lot of efforts on mobile application development. It plays a positive role to maintain existing customers and attract new customers. Moreover, one big problem of current business is they excessive dependence on advertising revenue. In the short term, advertising revenue can bring considerable income and it is an effective way to achieve profitability. But in the long term, along with the saturation of the market and increasing competition, user growth rate is gradually slowing until it turns to decline [He, 2014]. Meanwhile, the advertising revenue may not be continuous growth. Under such circumstances, 3Bmeteo have to take the exploitation of new profit channel into account. 3Bmeteo have a considerable number of users, they should further discover the user’s value to explore new profit channel.
### 4.4.4 Future Scenario

Based on the previous study and the semi-interview with 3Bmeteo, a new business model has been created in order to explore the opportunities of data mining for 3Bmeteo. In other words, the business model renewal wants to help the case company get the bigger market share and to create new business based on data mining. The change of the business model covers every aspect of the case company. Every block in the current business model canvas has changed. The new business model of 3Bmeteo is depicted in Figure 4.9. The cost structure components were not accurate at this stage. The differences between the current situation and future scenario have highlighted in red.

![Figure 4.9 the New Business Model of 3Bmeteo](image)

The new value proposition contains the improved personalized service and the valuable customer data. The improved personalized service aims to enhance the current customer’s user experience for individual customers. The purpose of it is to provide the individual customers better service in order to attract new customer and maintain existing customers. The user’s behavior pattern discovered by the web usage mining is the basis of improved personalized user experience. The case company
could redesign the current personalized service based on the finding of web mining. Also, they might use the user’s behavior pattern to develop preference-based recommendation bar. In the new business model, the customer’s usage data has great value for the case company. The case company might use the usage data to discover customers’ behavior pattern by the web usage mining. The data mining result is valuable for some companies. In this case, the data mining result may contain user’s behavior pattern and interest point - such as daily travel habits of users, the user preferences of travel destination, the weather parameters of users concern and the activities habit of the user. The case company could discover the tourism industry market trend or user consumption trend through analyzing and evaluating these pattern and interest points. As a result, 3Bmeteo could provide data mining results and related evaluation reports to some tourism related companies which business is easy to be affected by these results. Based on this, 3Bmeteo is able to shift their business into the B2B market. In this process, the company plays the role of counselors and key products include data mining results and related evaluation report.

According to the change of value proposition, the case company has increased new customers in the future. The companies of the service industries are involved in the new business model. Companies of the service industries include two types of companies. One of them with business activities mainly concentrates on the offline markets - such as travel agents and holiday resort. The other is the company which provides the tangible product or service - such as restaurants and car-rental firms. This model focuses on Italian local small and medium size companies, which business activities concentrate in Italy. In general, these companies do not have the capacity of implementation of data mining. They either do not have enough amounts of data, or does not have enough experience on the internet business. Moreover, weather factors have a significant impact on these companies’ business activities. An example of the travel agent is that weather changes directly influence on the experience of a series of tourism products - such as hiking, fishing and driving [Zhong et al., 2011]. The user usage data of 3Bmeteo may help these companies find out the customer’s potential interest or intention. It is similar with the user clustering. The users in the same cluster may have similar interest or intention. If a new user’s profile matches the features of the cluster, the potential interest of this user can be found. These companies can adjust the business strategy or manage the resources in order to gain the expected income by use the finding of 3Bmeteo. In order to reach these company customers, the case company should strengthen the promotion of the website and the market's competitive position. The social media provides a low-cost and extensive platform to achieve this goal.

The key activities can be derived from the value proposition. As the main value proposition is ‘the valuable customer data’, it stands out that the case company needs to discover the valuable customer pattern from customer data and deliver them to interested companies. Thus, the most important activity to achieve this goal is
customer data mining. The customer data mining can be divided into five steps: 1. Understanding and defining the business goal of data mining; 2. Data extract, transform and load (data sampling and preprocessing); 3. Use data mining software to establish the data mining model; 4. Analyze and evaluate the data mining results 5. Visualize the data mining results. The rest of the key activities have been proposed as the support of customer data mining. Developing hardware and software will enhance the interaction between individual customer and case company. In addition, the company customers might bring new users to 3Bmeteo’s website via their own website. This business model needs considerable initial investment and technical support for data mining. Therefore, the new model adds several new partners.

In this new business model, the cost structure includes two main parts: the additional employee’s salary and the cost of purchase for the data mining software. The case company needs more technical staffs in order to run the data mining business and enhance their personalized service. Thus, 3Bmeteo may put more cost into the employee’s salary. According to the statistics in 2013, the full-time data analyst’s annual salary is between 59.6K to 90.6K US dollars in Western Europe [KDnuggets, 2013]. The following graph shows the overall landscape of analytics salary:

4.10 Analytics Salary/Income by Region and Employment type (KDnuggets, 2013)

In order to support data mining, the data mining software is another significant cost in this model. As previously mentioned, the classic suites software suits this scenario. Based on the current data size of the case company and the size of the case company, one data mining software of classic suites has been proposed. The IBM SPSS Modeler is the data analysis software provided by IBM. It is easy to learn and use, that the data analysts can use it without the knowledge of the programming and algorithm. It provides the entire data mining required task module, including the data preprocess module, data mining module (diversified algorithm and relevant model), and analysis and evaluation module. As business solution offerings, the cost of SPSS is acceptable
for small and medium enterprises. The cost of SPSS is 25,000 USD, and the case company does not need to cost additional money to upgrade and maintain the software [IBM, 2015]. In general, the SPSS is suitable data mining software for the company which first time step into the data mining domain.

Given the above, the initial investment of the new business model consists two parts: one-time expense for the data mining software purchase and ongoing costs for the data analyst’s salary.

As regards the revenue streams, the case company can gain new profit from increasing advertisements income and company product sales. According to the implementation plan, the revenue stream can be divided into two stages. In the first stage, the case company starts to use the data mining to improving their website and the new revenue only generated by the increasing advertisement income. After a specific period (e.g., six months), the data mining obtains a good effect in improving the individual users of the website. The case company may start to run the data mining business with company customer. As a result, the case company will get new profit from company product sales. These two stages show in figure 4.11.

*Figure 4.11 the Two Stages of Revenue Stream in New Business Model*

In order to make the case company could evaluate the financial feasibility of the new business model. The author created the following equality to help the case company to analyze the feasibility of the data mining business in the first stage. Thus, this equality does not consider the company product sales income as well as payment service revenue. This equality is inspired by Break Even Point (BEP) calculation formulas. The Break Even Point is the point at which the total revenue and total cost are equal in a new investment [Dayananda et al., 2002]. When a company reaches this point which mean there are no net loss or gain, and the company has broken even [Tucker, 1963]. See the Break Even Point equality below:

\[
\text{Total Revenue} = \text{Total Cost}
\]

In addition, the current cost only includes the employee’s salary and the data center cost. The author does not take the uncertain costs of R&D and the cost of capital assets into account in order to facilitate the calculation.
\[ cp \times x\% = a \times t + b \]

In this equality:

- \( cp \) = the total advertisement revenue of company in last year
- \( t \) = the time in the first stage (monthly)
- \( x\% \) = the total advertisement revenue growth rate
- \( a \) = the cost for data mining analyst (monthly)
- \( b \) = the cost of data mining software

In the above equality, we could regard as when the company reaches the break even point that they will start get new profit. After simplification:

\[ x\% = \frac{a \times t + b}{cp} \times 100\% \]

Based on this equality, the case company could calculate how much growth rate they should achieve in order to achieve profitability (in the first stage).

In the following text, the different usage and adoption scenarios of the equality will be discussed in the detail. In order to make these scenarios are easy to be understood, the author uses imaginary data in the following example and scenarios.

Assumption: The current annual advertisement revenue of the case company is 1,000,000 Euros and the cost for data mining analyst is 5,000 Euros per month. The cost of data mining software is 25,000 Euros and the case company decides to run the first stage in twelve months. As a result:

\[ x\% = \frac{5000 \times 12 + 25000}{100000} \times 100\% = 8.5\% \]

The result shows that if the company wants to achieve the break even point, their advertisement revenue should be increased by 8.5% in the first stage.

If the total advertisement revenue growth rate of the company achieves to 8.5% in twelve months, the fixed cost (software cost) will be recovered and the company will start to get new advertisement profit in the thirteenth months. Based on the increased net profit and the total investment of this project, the return on the investment is presented in the following three scenarios.
The scenario A is the preferable scenario, the total growth rate increases to a satisfactory level. Under such situation, the company may recover their investment and obtain extra profit. Moreover, the return on investment rate could help company realize the potential of the data mining and they could consider to continuing this project, and starting the second stage.

On the opposite of the scenario A, the total revenue increase is not as optimistic as expected. Sometimes, the growth rate could not reach the break even point, which means the company will lose money on this project or cannot recovery the investment in the short term. Moreover if the growth rate is very unsatisfactory in a few months after the beginning of the project and the situation of the company capital is not good. The company needs to be more serious consideration of the data mining project. The company could consider stopping this project to limit the losses.

In the third case, the growth rate exactly reaches the break even point. Based on the company’s finance condition and the current operation status, the company could consider cutting down the investment or appropriately extending the first stage.

4.4.5 Network Picture of New Business Model

A network structure of the new business model is illustrated in Figure 4.10. It showed 3Bmeteo’s business process in the future scenario. The network picture visualizes the activities between the participants. It describes the interaction between 3Bmeteo and external world in the new business model. It highlights the dynamic relationship between each participant. In addition, this picture indicates that the role of participant and data in this scenario. In this picture, the end user refers to the individual customer and the industrial customer means the company customer. The end user stream and data stream represent the movement direction of the end user and data. The profit generation shows the new profit channel for 3Bmeteo. The following text will present the detail of each activity as well as explain the specific actions of each actor.
Reach: This is the first step that end users reach to the new business model. The 3Bmeteo website as the entrance of end users, it plays an irreplaceable role in the whole business process. 3Bmeteo offers valuable product and service to attract end users to visit their website and use their apps. In order to further explore end user’s value, 3Bmeteo could lead end users to the industrial customer’s website which conforms to end user’s preference. As the main action in this step, it is an improved advertisement experience. The purpose is not only making end users to visit industrial customer’s website but also inspire them to explore it. It can be regarded as a part of the personalized content. As aforementioned, the web usage mining may provide the practical guide to help user’s preferences. Based on this, 3Bmeteo could utilize the finding of web usage mining to deliver recommended information to end users. The information will drive end users to visit industrial customer’s website. For example, the relationship between visit time and locations may imply the end user is planning a trip. Based on this, the case company can deliver travel-related information to customers. If the information is useful for customers, then they will go to the industrial customer’s website. For the industrial customer, this may bring potential consumers to them. It also could be a new profit channel for 3Bmeteo. They can charge to the industrial customer by driving the end user to visit the industrial customer’s website. And the fee depends on a number of end users. End users will create a lot of data during this step, both of visit 3Bmeteo website and industrial customer’s website.
**Conversion:** In the online business, the conversion refers to the act of the website visitor complete a company expected activity [E-commerce, 2014]. For instance, the company converts a free user into a paying customer by offering attractive service. This act can be called one conversion. For the different business purpose, the company may consider different indicators to measure a conversion - such as, the user dwell time, registration, feedback, clicks of advertisements. For this networking picture, the conversion of 3Bmeteo represents the following indicators: the time of the users stays on the industrial customer’s website and the registration/order quantity submitted by the user. (The user here refers to the individual customer comes from the 3Bmeteo)

After above activity, end users come to the industrial customer’s website from 3Bmeteo. End users can get detailed information in the industrial customer’s website. They may make the decision to choose the industrial customer’s product and service based on the information. Once the end user submits the registration or the order, it will be marked as one conversion. During this process, the industrial customer will obtain new end users and increase the sale. In addition, the new business model picks the company of service industries as the primary industrial customer, so the end user will be directed to the offline environment after this step. An example of this, based on the user behaviors 3Bmeteo considers this user maybe plan a travel. As a result, 3Bmeteo recommend travel information which in relation to this user interesting location. And it drives the user to a travel agent’s website and this recommendation conforms to the user’s expectation. Then the user may choose this travel agent as the travel service provider.

**Consumption:** When the end user moves to the offline environment, they will use the industrial customer’s product and service. 3Bmeteo can involve this part by providing weather-related service. As the example mentioned above, 3Bmeteo can offer the real-time weather service associate with the travel destination for the travel customer. For 3Bmeteo, this step increases the channel to reach end users. The industrial customer’s existing end users may choose 3Bmeteo’s service when they use the industrial customer’s product. It brings a chance for 3Bmeteo to gain new end users.

**Feedback:** After the end user finished consumption activities, they may submit their feedback to industrial customers or 3Bmeteo. Actually, 3Bmeteo should encourage end users to share their experience and feelings with others. One effective way is the beneficial incentive that may attract end users to share their feedback. An easy-to-understand example of the beneficial incentive is discount or free for use. For 3Bmeteo, they can offer the discount of paying service for the end user who submitted feedback to the 3Bmeteo community. For instance, the end user can get one week of radio weather forecast for free once submit the feedback. Because for the internet-based company it is the uncommon customer data which collected from the offline environment. It not only contributes to customer needfinding and user experience
improvement, but also increase the diversity of data. The end user’s feedback is different from the server log data. It contains the user’s comment, viewpoint and a series of subjective feeling. Analyzing these data, 3Bmeteo can know what activities the end user did in the real world. It reflects the result of the online behavior of end users. Based on this, 3Bmeteo can improve the accuracy of the result of data mining. In addition, the web content mining provides an ideal solution for feedback analysis [Liu & Yu, 2009].

**Web mining:** The last part of this network is web mining. In the previous activities, a large amount of data was created and these data are the basis for the implementation of web mining. The data involved in the web mining including the data from the web server log and the feedback of individual customers. On one hand, the finding of web mining depends on data’s quantity and quality. Also the methods, the algorithm is very important. The case company needs adequate technical support to run the web mining business. In this part, the partner refers to the technology vendors and investment. On the other hand, web mining aims to find useful pattern or relationship to support industrial customer’s business activities - such as, the previously mentioned travel industry, car-rental companies. Besides, it cannot be ignored that to define the appropriate key pattern for the different industry. For instance, the web mining result may indicate the influence of weather on the flight ticket. It is valuable for a travel agent but may be not useful to a restaurant. Analyzing and evaluating the key pattern may uncover market trend, the user's consumption patterns or associated activities. As a result, it is an important product of 3Bmeteo in the new business model. 3Bmeteo may gain new profit by offering the web mining result to industrial customers.

**4.4.6 Implementation Plan of New Business Model**

After a comprehensive strategy has been formulated, an implementation plan has been proposed. The implementation plan specifies the tasks and objectives that are required to achieve in the new business model. All these tasks were divided into the short, medium and long term to create a time-based action plan. And the author decided on merging activities and placing them on the time line according to following two factors. 1. The complexity of the activities. 2. The needed resources for implementation (time, knowledge and financial)
Figure 4.13 the Implementation Plan of New Business Model

Phase I

Defining and identifying the possible potential company customers have great influence on the success of data mining. The company customer determines the data mining direction and the task. The possible potential company customers have been identified for the new business model. The characteristics of these potential company customers are: 1. involving the service industries, 2. the customer’s consuming behaviours of the company will be influenced by the weather factors, 3. the weather factors may influence their business activities. Based on the above aspects, the potential company customers are: restaurant, car-rental company, travel agent and holiday resort.

Data mining is the core task of the new business model that throughout all phases. The data mining has two main missions in this model, one is to develop recommendation bar to deliver the personalized content to the user in phase I. As previously mentioned, the main method of this mission is the web usage mining, and the data required is the data of web server log. Another mission is to discover the value of the data. The main methods include classification, association rules mining and the clustering. For this mission, the main mining objects are web server log data and users’ feedback.

Ideally, 3Bmeteo would have a considerable data volume after phase I. This means that they have the prerequisites for the implementation of data mining. In this phase, the case company could start data mining while building the database management system. Considering the technical and cost factors, the technology vendors and investors need to be settled at the beginning of this phase.
Theoretically, the database management system (DBMS) refers to large-scale software to manipulate and manage the database, and it uses for establishing, use and maintenance of the database [Ullman & Widom, 1997]. At present, the case company uses the Amazon web service to operate their website and to store data. In the future scenario, the Amazon web service is able to provide technical support for data mining [AiLab, 2012]. The Amazon web service can provide resizable compute capacity in the cloud and flexible data storage solution [Amazon web service, 2015]. The company does not have to concern about whether the data amount will over the capacity of the hardware in computing or storage aspects. The other option is to choose the third-party developers as technical support. 3Bmeteo can choose suitable technology vendors according to the cost, time and the size of DBMS.

Once the DBMS has been settled, the case company might run their data mining immediately. In this business model, the case company expects to exploit data value based on their weather service business. And their potential customers are the company which business activity may be influenced by the weather factors. Thus, the key pattern they expected should be within the following aspects. 1. The relationship between weather condition and customer activity (what the user doing in the real world - e.g., sports, relaxations or works.) 2. The relationship between customer behaviors (how the user browsing the website and what they did on the website - e.g., searching a location, pay attention to a certain parameter, likes pollen) and their activities. 3. The relationship between customer preferences (what the user intend to do or would like to do in the real world - e.g., planning a travel) and their activity. 4. The relationship between customer behaviors and their preferences. 5. The relationship between weather condition and customer preferences.

This phase is the advancement of the current business. The main purpose of it is to increase the number of users and to collect a wide variety of data. In order to achieve this goal, the case company should deliver the improved personalized user experience as above-mentioned. It not only refers to the efforts to improve the product and service, but also implies to the concerns to respect the user’s preferences. A good user experience is embodied with the quality of product or service as well as the concern of customer feeling.

The case company needs to redesign the existing personalized service. The personalized service is contained within the 3Bmeteo’s community service which called 3B+. 3B+ is a community service that provides registered users personalized weather service and a platform for the information exchange of the users. The registered user can subscribe the weather forecast of interested locations and share the weather video, picture or article. The redesign task put the emphasis on improving the personalized service. In the beginning, the user can find the personalized service on the navigation bar. It is a valuable function for customers, but 3Bmeteo sees that not many customers use this service. So that 3Bmeteo put it into their community service
after a website revision. Generally speaking, there are three main reasons reduced the customer’s interest in online service market [Nielsen, 2008] [Demaria, 2013]. 1. Poor customizable elements, it only offers two customizable modules for users that are location and activities. The user can an select interested location and choose the activities (sports, relaxations or habit) they may do in the interested location. 2. The monotony of push service, the user will regularly receive a weather forecast of the interested location by email. And it is one-page weather forecast report for the next week. It is almost no different with the ordinary weather service. The customer can get the weather forecast on the website as well. 3. The interaction lacks flexibility. When the user was chosen preferred activities and time, this service will inform recommended or possibly activities to the user depends on the weather condition. It is very useful for the user, but 3Bmeteo only gives an extremely simple result to the user that only tell the user the activity is recommended with a smiling symbol or the activity is not possible with a sad symbol. It will reduce the user’s interest of this service.

Current personalized service is more like a subscription service. It saved some time for users and meets their segmental needs, but this is far from enough. The following suggestions may help case company to improve their personalized service.

The case company needs to increase the types of customizable elements. Except the location and activity, the user also concerns about the time (e.g., hourly weather forecast), season (e.g., seasonal food or seasonal diseases) and local event (e.g., local festival). 3Bmeteo needs to consider adding these elements to the personalized service. In addition, different customers may concern different weather parameters. If the website allows the user to add the concerned or preferred weather parameters into the user profile. It will increase the user’s satisfaction for this service. For instance, pollen allergy patients care about the pollen issues. If they provided this information into their user profile that the website will categorize them into the specific user group. Then the website sends the pollen warning to these users by email or app. The benefit for users is that they do not need to check this information on the website every time. The pollen notification can be combined with other customizable elements. This notification may interact with users by different ways which depend on the season, user’s location and activities. The details should contain different pollens for the different season, the impact of weather changes on the spread of pollen and suggestions of outdoor activities in pollen season. In addition, the user’s location will determine the time and duration of the pollen season. Pollen concentrations will vary from day to day as well as hour to hour. Users would like to provide more information into their user profile in order to track pollen situation in their area. 3Bmeteo provides local pollen information to users that will be increasing the user’s willingness to register to the website and use their app. Further, when they consider regional pollen service that may combine with weather forecast to offer users instant pollen forecast. Moreover, the user intends to participate in some activities and the information also
registered in the user profile. If the destination of the activity is the high pollen concentration area that 3Bmeteo could warn the user about this condition.

The push service is another important factor for this service. The user will not be willing concentrate on a monotonous weather forecast report. They can try to avoid using heavy text to express the weather forecast and to utilize figure and diagram to highlight the core content for the push service. Connecting the personalized service and 3Bmeteo’s app that the user can flexibility receive the result of the service instead of regularly get this information from the email.

Then the customers desire more interaction of this service. The user no longer wants to just know the weather or be informed whether the activity is recommended. The personalized service should provide more than these. The practical suggestion and possible alternative activities will inspire the customer’s motivation. For an example, the service alert pollen allergy this location with the high concentration of pollen. It not only informs the user this area is not suitable of jogging, but also should provide possible alternates. It can suggest the user the nearest available area for jogging or recommend the user to play related indoor sports.

Phase II

The main task in this phase is involving the smart hardware. With the improvement of customer’s health consciousness, the weather condition of indoor environment is more and more valued by customers. The main application scenario of the indoor weather stations mentioned in this section is the private residence. The indoor environment different from the outdoor environment, the indoor weather condition is not complex. For the indoor environment, the user concerns about humidity, concentrations of carbon dioxide and temperature. According to these parameters, the user could know whether they should open the window, turn down the heating system or increase indoor humidity. These customers concerned parameters can be easily collected by low-cost smart hardware. Theoretically, the built-in sensors of the smart phone will be able to collect these data. The basic component of the indoor weather station is temperature transducer which can be used to detect the air humidity and temperature. It has the potential for 3Bmeteo that can increase the customer data both of quantity and quality as well as raise the brand awareness. Also, it can bring certain economic benefits for 3Bmeteo by selling the indoor weather station to individual customers. The data of indoor weather station is different with the website server data; it can provide particular data which cannot get on the internet. These data refer to the usage data of the hardware which including the frequency of utilization, the daily utilization time and the specific time of using. These data may reflect the user’s routine and living habit. The case company can combine these data with the server log data in order to discover more comprehensive customer behavior pattern both in the online and offline world. In the other word, the usage data might help the case company
uncover what the user did online and in certain extent the indoor hardware could help the case company know what the user did in the real life. These data can make the case company to gain more insight into the users' points of view, thus providing better service. 3Bmeteo could either develop their own devices or cooperate with third party hardware manufacturer. The role of the company in the indoor weather system is more like a software vendor than a hardware developer. The most important thing is not to develop a powerful device that the function of the device can be simple. The case company should put emphasis on the interaction between the user, the device, and their app.

Phase III

Furthermore, the case company should pay enough attention on establishing the connection with the potential company customer in the third phase. The good connection between the case company and the potential customer is the essential factor of the new business model. Through continuous contact with potential customers, 3Bmeteo may convert some of them into the real customer. The case company might reach these potential customers by offering them the weather service on their website. Based on this, they will build the primary connection. The interaction between the case company with the potential customer is also contributing to building the connection. For instance, 3Bmeteo can utilize the social network to communicate with the potential customer and help them realize the value of the 3Bmeteo's offering.
5 Discussion

The discussion section aims to connect the research questions with the findings. It can be divided into three parts: 1) the overview of current customer user experience, 2) the validation with the case company and 3) the implementation challenge of the business model renewal.

The first part summarizes the current customer overview experience and highlights the obstacles of the customer journey. Thereafter it discusses the key aspects of data mining technology in relation to the current user experience improvement.

The second part describes the validation of case study result with the case company, the process, outcome of this validation are discussed.

The third part points out the possible challenges of the implementation plan that will hinder the new business model development. Meanwhile, it discusses the effects of these challenges.

5.1 The Overview of Current Customer User Experience

The user experience refers to "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [ISO 9241-210:2010, 2010]. In other word, the user experience reflects the customer’s subjective feeling. The user experience influences every step of the customer journey. Thus, a user-centered and people-oriented design concept is fundamental to improve the customer’s user experience [ISO 9241-210:2010, 2010]. Based on the customer journey map results, the followings list presents the key points of the customer journey of online weather service.

1. For the customer who is for the first time using the online weather service, the company's brand awareness plays a more important role than the product. The company’s brand awareness contains a wide range of aspects and an important role of it is to familiarize the customer with the company. In this case, there are two aspects that affect customers’ experiences which are: the previous user’s feedback and the website ranking on the search engine.

2. Today, most of the online weather websites rely on advertising to make profits and the user realizes the inevitability of advertising presence. But not all the
advertisements would undermine the user experience. On the contrary, the reasonable distribution of advertising position may contribute to making website structure clearer as well as to increase user satisfaction. Thus, on the one hand, the website which has a well-organized page layout and reasonable distribution of advertising should be attracting the attention of the user. The website should put more effort into the content of the advertisement. In recent years, a new form of advertising has become an important trend of the development of online advertising which named native advertising [Yahoo Emarketing Tumblr, 2014]. The concept of native advertising was proposed by Interactive Advertising Bureau in February 2012, it is a form of online advertising that matches the content and the features of the website in order to create extra value for users [Greenberg & Navin, 2013]. In other words, the native advertising is blending advertisements into the website content instead of displaying them separately on the page. Take online weather services, for instance, the weather website post weather related articles to attract users. If one article talks about the trend of pollen, the website could put some pollen related advertisement in the article that increases the pertinence of the advertisement.

3. The visual experience is the most direct experience of the user. The first impression of the website appearance is an essential factor that determines whether the user will continue to browse the website or not. The visual experience of a website not only is the appearance of the website, it also includes the design style, color matching, page layout, page size, font size, the logo of the website. This is very critical to delivering to the customer the core value of a website through the visual experience. There are two design trends of visual experience cannot be ignored: responsive web design and reasonable content rendering [Tina, 2015]. The responsive web design is a website design approach that aims to apply this website for different devices (from desktop computers to mobile devices) in an optimal viewing and interaction experience [Marcotte, 2010]. Along with the popularity of mobile devices, more and more website develops the mobile device version website. Thus, it is important that make sure users on the mobile devices can also have good visual experience. The responsive web design provides a good solution for this problem. The reasonable content rendering refers to concise text content and rich media content [Birch, 2014]. The test content is the basic carrier of internet information, but the text content is not the most effective information carrier. Compared with text content, the media content (picture and video) is more intuitional and concise. And the media content contains a greater amount of information than text content on the same page space [Chen, 2012b]. For online weather services, these two trends should be considered as the premise of the future web design.

4. Users expect to interact with the website in a variety of way, rather than just browsing the information in a static mode. A website that interacts with the customer will help the connection between the customer and the company. Furthermore, the diversified interaction may help users to explore the website more deeply and to use
the service of the website in a personalized way (The online weather services company how to interact with users in a personalized way can be found in section 4.4.5 and 4.4.6). In addition, the user interaction experience of the website is not only embodied in the services, but also reflects in the interaction with the web page [CDC, 2014]. For example, when the user browses web pages, the different activities may trigger different image or audio effect that makes the interaction between user and website become more lively and interesting.

In order to help the designer effectively improve the customer’s user experience of online weather service. Table 5.1 highlights obstacles which hinder the customer to gain a better experience during the customer journey of online weather service. Moreover, the proposed solution of data mining for part of obstacles can be found in section 4.3 and 4.4.
<table>
<thead>
<tr>
<th>Obstacle</th>
<th>Detailed Description</th>
<th>The Impact on User Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confused page layout</td>
<td>The page layout is important to influence user dwell time. Also it is the first step that the website to establish connections with the customer. Mixing the content and the advertisement, complex layout or the incompatible theme will directly reduce the user's interest.</td>
<td>The sensory feeling of customer; Reducing the efficiency.</td>
</tr>
<tr>
<td>Advertisement</td>
<td>The excessive amount of advertisements and the improper management will direct damage the customer’s user experience. And it even causes the user to leave the website with a bad impression. The attention of designer should be centered on the following four aspects: content, size, position and frequency.</td>
<td>Increasing the response times; The reading and browsing experience.</td>
</tr>
<tr>
<td>The availability of content</td>
<td>The availability refers to not only the quality of the content but also the structure of the content. It includes the link between relevant contents, the efficiency of getting content and the distribution of contents.</td>
<td>The customer’s reading and browsing efficiency.</td>
</tr>
<tr>
<td>The usability of website</td>
<td>The usability of a website refers to the navigation bar and the embedded search engine of the website. The main interaction between the customer and the website is implemented through them. The inefficient navigation bar and the embedded search engine will hinder the user to explore the website deeply.</td>
<td>The interaction with the website/company; The search ability of the information.</td>
</tr>
<tr>
<td>The usability of mobile website</td>
<td>Some customers prefer browsing the website through a mobile device. They expect to gain the same experience as well as using a computer. The designer should not ignore the influence of the mobile device user. It is a considerable factor that causes a loss to customer churn.</td>
<td>The customer loyalty; The flexibility of usage.</td>
</tr>
</tbody>
</table>
5.1.1 The Key Aspects of Data Mining Applied in The Website

This section discusses the key aspects of data mining technology application in relation to the improvement of the current customer experience. In the website analysis, it is not difficult to get a lot of data. The difficulty is to establish valid evaluation system in order to achieve high-quality analysis and effective results [Kaushik, 2014]. This is the basis of digitizing website operation and promotion [Texun Tech, 2012]. In order to build valid and effective website analysis system, it is an essential step to select suitable key indicators (the details of the mainly key indicators of website design can be found in table 4.1). In this section, the key indicators will be discussed which can be used to match the data mining applied in the website design. The importance of the key indicators in web analysis is not because of its form but depends on the user activities and trends reflected by the data. Therefore, it must comply with the following criteria when selecting key indicators [Lan, 2011].

1. The key indicator must be a ratio or percentage of average. The key indicator should not be a raw data because the raw data does not contain any background information [Henneberry, 2013]. Thus, the designer is hard to find the valuable information from raw data. 2. The key indicators must have a clear definition and boundary in order to allow all the designers to have a clear understanding of the role of the key indicators. 3. The change of the key indicators must be able to drive a specific activity. The key indicator is used to analyze and optimize the website. So that it must ensure that there is at least one corresponding activity can be improved when the key indicator changes [Think Metrics, 2014]. According to these, the author outlines several key aspects in order to help the designer to apply data mining effectively and accurately.

The valid measurable indicators and the effective goal

The value of a data mining result depends on the goal of the data mining. In other word, the company that blindly runs the data mining for their business will waste resources and cannot gain the expected results. Before starting a data mining, the company should know what they want to find from the data. For an example, if an online weather service company intends to run data mining for their business, they should understand the relationship between the data they owned and their business goal. In order to establish an effective goal of the data mining, it should first develop valid measurable indicators to help the company to evaluate their business. The indicators can help the company understand the relationship between their business and different data. (The measurable indicators of the online weather website can be found in section 4.1.3.) Then they can use these indicators to establish the effective goal to discover valuable information by data mining. The effective goal refers to not only clear and organized data mining objectives but also the data mining goal should
respect the principle of data mining technology. For example, it should not set a goal to uncover the relationship between IP and the accuracy of forecasting, because there are not inevitable logical connections between them from the data mining perspective.

**Active user and churned user**

The churned user represents the customer who has used the product or service but for different reasons is no longer using the products or services - such as registered users, app downloader. Analyzing the ratio of active and churned user has a significant impact on the website usage data. Because a website may have millions registered users with only half of being the active users. If the designer regards these two types of data as the same, it will influence the result of the data analysis which would not accurately reflect the current customer’s user experience, thereby affecting the user experience improvement. Additionally, the data of churned users may help the designer to understand the shortage of the website. Thus, the data of churned users is useful and the designer should carefully separate them from the active user’s data.

**Page view (PV) /unique visitor (UV) ratio**

The customer’s user experience is hard to measure, and the statistics of users may not directly represent the user satisfaction. The PV/UV ratio may help the designer to understand the user satisfaction of the website, and the designer might base on this ratio to decide the improvement strategy and direction. If this ratio was significantly decreased while the UV has inconspicuous change, it shows that the numbers of visit pages of the unique user are reducing. This case might suggest that the content or the structure of the website needs to be improved. The structure of the website refers to the hierarchical relationships between different pages of the website. It has a very important impact on the interaction of the website and user experience. The PV/UV ratio as a dynamic indicator that is easy to monitor, it provides the designer a convenient way to judge the current user experience.

**The judgment of the effect of the user experience improvement**

The effect of the user experiences improvement is much concerned by the designer. There are different factors that can be used to judge the effect of the user experience improvement. The designer may continually monitor the hourly traffic trend of the website or to analyze the bounce rate of the homepage. It will help designer evaluate the effect of the improved user experience. But the most important factor is the time period, the designer should pay enough time to observe the change of the user. Because, at the first time, the customer may not accustom to the improvement of the website, even more they may not enjoy the improved user experience. The designers need enough time to inspect the anomalous behavior of customers and then to judge whether the effect of the improved user experience satisfied the customer’s needs.
Security and privacy issue

The purpose of the data mining is to uncover the customer’s inner activities based on their behaviors. It aims to explore the same behavior patterns and relevancies between different user data, as opposed to discovering the customer’s individual privacy and information. Thus, in order to regain the confidence of the customers, the company should maximize the protection of customer information and avoid involving the detailed customer information during the re-design process.

5.2 Validating with Case Company

In order to validate the case study result, the case company has been consulted in the later stage of the thesis work. The target of this activity was to present the case study results to the 3Bmeteo and to gather their feedback on the case study results. The current business model, the business model for the future scenario, the networking picture, and the implementation plan have been iterated with 3Bmeteo. The author communicated and shared the work results with 3Bmeteo by email. According to the feedback of 3Bmeteo, the author has found that one of the cost structures in the current business model had been overlooked. That is the important effort that 3Bmeteo has also paid in search engine optimization and social network management. For the new business model, 3Bmeteo has confirmed an interest into use data mining. Thus, the scenario for the future might comply with their development strategy in a certain extent. In addition, the core business of 3Bmeteo is transforming also into B2B focus. For the implementation plan of the new business model, 3Bmeteo sees the integration of new hardware as a challenge in a short term, but nevertheless they do not deny the possibility of this to happen in the long run.

5.3 The Challenges of Implementation Plan

In this section, the challenges of the new business model implementation plan are described. The company developing the new business model into practice will encounter multiple challenges in each phase. These challenges are categorized into three aspects that need to be addressed in order to the company to achieve the predefined goal. These challenges may arise throughout the entire implementation plan, but one of these challenges may become the main barrier of different phase. In figure 5.1, the main barriers of different phase are presented.
5.3.1 The challenge of the market

The main goal of phase I is to increase the number of customers and to collect a wide variety of data. In this phase, the main challenge of 3Bmeteo is that they will face larger pressure from the market. It contains the customer perception and the competition.

Customer perception

The customer perception refers to the customer’s feeling, impression, and awareness about the company and its offering [Blan, 2014]. The feeling and awareness is the first step of the customer experiencing a product or service. It is the foundation for the customer to further exploring the product. In some cases, the customer may not have a clear perception of some products on their cost and benefit. In phase I, the main activities of the company are to launch the improved personalized service and indoor weather device. It needs the customer to take the time to learn the new features and functions. Meanwhile, customers are required to update more personal information, in order to use the personalized service or indoor weather device. In this process, the customer may show resistance to change towards personalized service or toward the use of indoor weather devices. This creates the risk of declining the customer satisfaction of the company’s existing products. Therefore, the company should pay enough attention on how to provide the customer adequate guidance to enhance their perception of the improved personalized service and indoor weather device.

The competition

For any company, competition is inevitable and the competition from the same industry will not hinder the development of the implementation plan. But the company will encounter the impact from external industry while the company starts to develop the smart weather device. It makes the company to face larger pressure in a
completely new context. The main challenge comes from the smart home manufacturer and wearable device developer. With the progress of the information technologies and the development of the sensor, it is not difficult to capture the indoor weather parameters [MassThinker, 2014]. The smart home manufacturer and wearable device developer also pay respectable attention to this field [Reinisch et al., 2010]. Thus, in this regard the online weather service company does not have special advantages compare with other industries. Moreover, the other industry’s product has relatively more interaction with the customer according to the current situation. In general, in this phase, the company should take these factors into account. 3Bmeteo is a growing company that they may have not enough abilities to design and produce a smart device by themselves. Their advantages are they have a completed weather service website and plenty of experience in application development. One solution is to collaborate with wearable device developer or smart home manufacturer. The manufacturer develops the device and 3Bmeteo integrates their service and application into the device.

5.3.2 The Technological Barriers

The technological barriers of the implementation plan concentrate on the process of web mining. Generally, the technological barrier is a widespread problem of web data mining. In this section, there are two barriers are identified that will influence the progress of the new business model developing.

Semi-structured data environment

The data in the web environment is different with the data in the traditional database. The traditional has organized data model that the developer can characterize the specific data according to the data model. The data in the web environment is unorganized and it does not have a particular model to help designer characterize the data. Therefore, the data in the web environment is semi-structured. The semi-structure is the most important feature of web data that will interfere the accuracy of the result of web mining.

Language issue

The web data mining also involves the language issue. In general, the conventional data mining is aimed at processing one kind of language. For example, data mining can handle either Italian or English in a processing time, but it cannot handle two kinds of language in the same processing time. Therefore, it greatly increases the complexity and difficulty of data mining.
5.3.3 The Challenge of New Business

The last phase of the implementation plan involves different participants. The biggest challenge for this phase is how the company builds a good relationship with both of the environment inside and outside of the company.

Employee resistance

Under some circumstances, the 3Bmeteo’s employee may resist the change of new business model. The reason is that they are accustomed to the current business model as well as they understand their expectations and their role within the current business model. Thus, the employee may feel confused when the company makes significant changes of the business model. It leads to the employee cannot learn new knowledge about the new business or change the way of work to adapt the new tasks. The company should offer enough support to their employees and provide appropriate training to them in order to enable them to understand the new responsibility. It will help the company transforming its business model at a steady rate.

Lack of consensus

Building the connection with potential company customers depends on sharing the same understanding of the new business on both sides. The company fails to reach a consensus with their customers that will lead to the customer cannot effectively participate the new business. It may adversely affect the progress of the new business. The different business goals of partners and the potential conflicts of interest between different partners that make it difficult to share the same picture with all partners. In addition, based on the consideration for the protection of commercially sensitive information, the partner may not be willing to participate in this business. Therefore, in order to implement the new business model, it is an inevitable task that to establish a shared vision with partners in the future scenario.
6 Conclusion

This thesis has investigated the current customer journey of the online weather service and how the customer user experience can be improved by using web mining technology. Meanwhile, the potential of web mining for an online weather service company has been explored. It has exploited the possible solutions of business model renewal with the aim of applying data mining in practice. The answers to the research questions are presented in the following section as well as the main points of the results were discussed. Also, the suggestions for the further study were proposed.

6.1 Answer to Research Questions

Research question 1: How can the current user experience of the customers of online weather service company be improved by the use of available data and web mining?

Generally, web mining provides an effective way to help people (especially for the decision-makers who have no related technology knowledge) understand the customer online behaviors by discovering user behavior patterns. These behavior patterns can be used to understand the user experience of customers. Using traditional methods (e.g., data analysis, statistic analysis) it is hard to capture the online customer’s user experience, also it cannot help designers provide personalized service for different users. On one hand, the online weather service company lacks effective communication methods that enable them to obtain the customer’s user experience. The surveys and statistics can help companies to obtain the customer’s user experience, but they require a lot of human resources and time. The feedback review could help to improve the customer’s user experience, but its usefulness is limited. The reason is that the majority of customers’ feedback is subjective and emotional. The company needs to invest a lot of time in order to find useful information from the feedback. On the other hand, the judging criteria of customers are fuzzy and diverse. The same criteria may represent different meanings for the different customer. Under such circumstances, data mining technology provides appropriate and efficient methods to support the company to gain a better understanding of online customer’s user experience. Based on the findings of data mining, the designers could improve the current user experience and provide more personalized services. Web usage mining methods provide several useful tools to support customer’s user experience improvement. The clustering techniques can help the company to segment different visitors and to offer them personalized content according to user’s characteristics. The association rule mining contributes to optimizing the link structure of the website and
to manage the content of the website. It increases the efficiency of browsing of users and the usage experience of the website.

**Research question 2: How can weather service company based on the web mining exploit new business model to gain new profit?**

For the second research question, this thesis has proposed a new business model for an online weather service company based on web mining. The new business model aims to achieve the goal of gaining new profit by using web mining technology. In some cases, people’s daily activities (i.e., travel plan and trip mode) will be affected by the weather as well as the purpose of the activities will be changed by weather condition. For an example, the customer might change or cancel the travel plan according to the weather condition of the travel destination. The change of customer’s activities may impact on the business activities of some companies - such as, travel agent, car rental company. The user’s online behaviors could reflect the change of their activities. The case company might utilize web mining to discover the user’s behaviors from individual customer’s data. It can uncover the valuable business information from the user’s behaviors. The case company could offer the evaluated data mining results to company customers in order to gain profit. It can contribute to the company customer make effective business strategy in order to increase sales and to adapt to market changes. An implementation which includes three phases has been proposed to support the new business model as well. It has clarified the goal and specific activities for the different phase of the new business model. In order to implement the new business model, the company needs to increase the quality and quantity of individual customer’s data. The data quality and quantity is the foundation of the realization of web mining that does not imply the web mining cannot be conducted with existing customer data. But the detailed data may increase the accuracy of data mining results. In order to achieve this goal, the case company has two main activities in the implementation plan. One of them is to redesign and improve the current personalized service to provide a better user experience to users. The new personalized service aims to provide the user personalized content (e.g., notification for pollen allergy patients) and the improved advertisement experience. Another activity is related to indoor weather devices. The case company could work with the hardware manufacturers to jointly develop indoor weather device. These activities could be new channels used to attract more registration to 3Bmeteo platform and make more data available in the future. Moreover, in order to help the company could evaluate the financial feasibility of the new business model as well as analyze the return on investment, the author created an equality based on Break Even point calculation. Using this calculation, the company could draw a conclusion that implies the relationship between the financial feasibility and the total revenue growth rate.
6.2 Suggestions for Future Work

This research has identified the important touch points of the current customer journey. There may be more potential factors that would affect the customer’s user experience. In order to get a full-scale understanding of the customer journey, the further work would focus on the more interactive method to collect customer information. Surveys and questionnaires are feasible methods that can be used for collecting customer information. Moreover, the further study could formulate the technical details of data mining which can be applied to the customer’s user experience improvement. The feasibility of data utilization technology for online weather service industry development also could be discussed in the future work.

Based on the new business model proposed in this thesis, there are several aspects that could be explored in the further study. The web mining requires that specific customer data are identified and defined. Thus, future research can focus on how to analyze the needs and the expectation of the weather related company. Identifying the specific potential key customers and finding out the information on which they are interested is also an aspect to be considered in the future from an industrial standpoint. Furthermore, the financial feasibility of the implementation plan needs to be discussed as well as the cost of data mining.
References


Yahoo Emarketing Tumblr (2014), *The evolution of online advertising: native advertising*, Available from: http://yahoo-emarketing.tumblr.com/post/120177445511/2015%E6%95%B8%E4%B8%AD%E5%85%AD%E5%8E%9F%E7%94%9F%E5%BB%A3%E5%91%8A-%E5%AE%8C%E7%BE%8E%E8%9E%8D%E5%85%A5%E5%86%8D%E9%80%B2%E5%8C%96 (accessed 9 October 2015).


