DATA ANALYTICAL TOOLS TO EVALUATE SERVICE OF MOBILE COMMUNICATION PROVIDERS IN MALAYSIA USING R LANGUAGE

NURUL IZZAIMAH BINTI JAMALUDDIN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: DATA ANALYTICAL TOOLS TO EVALUATE SERVICE OF MOBILE COMMUNICATION PROVIDERS IN MALAYSIA USING R LANGUAGE

SESI PENGAJIAN : 2017

SAYA_____NURUL IZZAIMAH BINTI JAMALUDDIN_____

(HURUF BESAR)

mengaku membenarkan tesis (PSM/~~Sarjana/Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (/)

_____ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

_____ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

_____ TIDAK TERHAD

_____ _____

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat tetap: No.15 Jalan Meranti 3A/7
Bandar Utama, 44300 Batang Kali, Selangor

PM DR BURHANUDDIN BIN MOHD ABOOBAIDER
Nama Penyelia

Tarikh: 11/08/2017 Tarikh: _____

CATATAN:* Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

**Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

DATA ANALYTICAL TOOLS TO EVALUATE SERVICE OF MOBILE COMMUNICATION PROVIDERS IN MALAYSIA USING R LANGUAGE

NURUL IZZAIMAH BINTI JAMALUDDIN

This report is submitted in partial fulfillment of the requirement for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION AND TECHNOLOGY

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I hereby declare that this project report entitled

DATA ANALYTICAL TOOLS TO EVALUATE SERVICE OF MOBILE COMMUNICATION PROVIDERS IN MALAYSIA USING R LANGUAGE

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT: _____ Date:_____

(NURUL IZZAIMAH BINTI JAMALUDDIN)

I hereby declare that I have read this project report and found

this project report is sufficient in term of the scope and quality for the award of

Bachelor of Computer Science (Artificial Intelligence) With Honours.

SUPERVISOR: _____ Date: _____

ASSOC. PROF. DR. BURHANUDDIN BIN MOHD ABOOBAIDER

DEDICATION

To my beloved family, thanks for your support and caring during the development of this project.

To my beloved supervisor, Assoc. Prof. Dr. Burhanuddin bin Mohd Aboobaider, thank you for supervising and guiding me along the completion of this final year project.

To my beloved friends, UTeM students from other courses who are willing to share their knowledge with me during the development of this project.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Assoc. Prof. Dr. Burhanuddin bin Mohd Aboobaider for giving me this opportunity to developed this project. Thank you for supervising, guiding and giving me advice along the completion of this project. I appreciate all of your trust and effort towards me in developing this kind of project.

I would like to thank my family as well for the non-stop moral support as I have to sacrifice quite an amount of time with them for the development of the project

I also would like to thank all of my dearest friends who are very helpful and willing to share their knowledge with me.

ABSTRACT

Nowadays, one of the sector in Malaysia that has been significant growth in the market is telecommunications sector. Among major telecommunication companies available in Malaysia are TM Berhad, Celcom, Maxis, U-Mobile and etc. However, many telecommunication service providers faced with the increasing competition and decreasing profits. In Computer Science, we are able to apply analytical method in doing visualization in finding best telecommunication provider among telecommunication sector in Malaysia. This project embarks to identify the data mining techniques for communication application where the data itself gotten from social media of Twitter platform. Besides, to analyses how the system could process and interpret the big data into a simple visualization. Then, build the computerized tools and recommend the data analytic models based on the collected result. From prepping data for pre-processing until conducting analysis, this project is focusing on the process of data science itself where Cross Industry Standard Process for Data Mining(CRISP-DM) methodology will be use as a reference. This analysis will be developed by using R language and RStudio packages. In sum, this project might brings insights of how the telecommunication industry nowadays and exposes it to the application of the data analytics lifecycle. Upon completion of this project, Malaysia telecommunication industry will get the benefit by improving their customer satisfaction and business growth. Besides, it will give the awareness to the telecommunication user of updated review from the users.

ABSTRAK

Pada masa kini, salah satu sektor di Malaysia yang telah berkembang pesat dalam pasaran ialah sektor telekomunikasi. Antara syarikat telekomunikasi utama yang terdapat di Malaysia ialah TM Berhad, Celcom, Maxis, U-Mobile dan sebagainya. Oleh hal yang demikian, banyak syarikat perkhidmatan telekomunikasi menghadapi persaingan yang semakin sengit dan keuntungan yang semakin menurun. Di dalam bidang sains komputer, kami ingin menggunakan kaedah analisis untuk melakukan visualisasi dalam mencari syarikat telekomunikasi yang terbaik di Malaysia. Projek ini dimulakan dengan mengenal pasti teknik dan kaedah untuk mendapatkan data untuk proses yang berikutnya iaitu pembersihan data. Selain itu, untuk menganalisis bagaimana sesebuah sistem itu boleh memproses dan mentafsir data besar menjadi visualisasi yang mudah. Kemudian, menyediakan alat berkomputer dan cadangan model analitik data berdasarkan hasil yang dikumpulkan. Dari pembersihan data untuk pra pemprosesan sehingga menjalankan analisis, projek ini memberi tumpuan kepada proses sains data sendiri di mana metodologi khas untuk data mining akan digunakan sebagai rujukan. Analisis ini akan dibangunkan dengan menggunakan pakej bahasa pengaturcaraan R iaitu RStudio. Secara ringkasnya, projek ini mungkin dapat membawa gambaran tentang bagaimana industri telekomunikasi pada masa kini dan memperkenalkan aplikasi kitaran hayat analisis data. Setelah projek ini disiapkan, industri telekomunikasi Malaysia akan mendapat manfaat dengan menambah baik perkhidmatan kepada pelanggan mereka seterusnya mampu meningkatkan keuntungan perniagaan mereka.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

1. INTRODUCTION

1.1 Introduction

Nowadays, one of the sector in Malaysia that has been significant growth in the market is telecommunications sector. The primary regulator of telecommunications in Malaysia is the Malaysian Communications and Multimedia Commission(MCMC). Regulatory reforms and market liberalization play a big part in creating competition and structural reforms in the industry. Among major telecommunication companies available in Malaysia are TM Berhad, Celcom, Maxis and U-Mobile.

The Malaysian telecom industry growth has been exceptional in recent years. Mobile penetration grew from 98.9% in 2008 to 141.2% in 2013. Not to mention, mobile subscribers also grew to 41.9 million at a rate of 54% during the same period.

However, with the growing demand in the field of Big Data, utilization of structured and unstructured data leads to worthy information for telecom industry in Malaysia to grow exponentially. To utilized this structured and unstructured data, it requires some critical and analytical method and this method is what this report are all about.

In this report, we would like to indicate the application of several machine learning method that is used in the industry today in gain the competitive advantages by analyze the social media. Methods such as machine learning based analysis will be discuss further in the next chapter.

1.2 Problem Statement

Today we face a problem to find out the best telecommunication service provider, since there is too many choices in the market. Though data storage costs are dropping, volume and velocity are increasing at astronomical rate, and mobile use has permeated the entire world and is set to expand even further.

Many telecommunication service providers mainly faced with the increasing competition and decreasing profits. Many telecom company has devised plan to introduce new services to their customers by creating an alliance with a satellite provider as well as applying practices that will increase profitability and cut costs. All these elements were not being utilise completely and this leads to the failure growth for telecommunication sector to increase brand awareness.

And from the perspective of Computer Science, we are trying to apply analytical method in doing visualization in finding the best telecommunication service provider in Malaysia telecommunication sector.

1.3 Objective

The project embarks on the following objective:

- To identify the data mining techniques for communication application.
- To analyses how the system could process and interpret the big data into a simple visualization
- To build the computerized tools to mine the big data
- To recommend data analytic models based on the collected result.

1.4 Scope

From prepping data for pre-processing until conducting analysis, the scope of this project is focusing on the process of data science itself where Cross Industry Standard Process for Data Mining, commonly known by its acronym CRISP-DM is the methodology that will be applied in this project.

1.5 Project Significance

This project brings insights of how the telecommunication industry such as Celcom, Maxis, U-mobile and others work nowadays and while exposes it to the application of the data analytics lifecycle. Upon completion of this project, this project will benefit the Malaysia telecommunication industry in using the data in more flexible and innovative way while giving them a benchmark in brand awareness to rely upon.

1.6 Expected Outcome

Based on the project the expected output would be as the following:

1. Understands the process of Data Science application in industry
2. Data analysis on demographics for industry performance
3. Interpret the data analytic tools for telecommunication industry.

1.7 Conclusion

This research paper has been divided into 6 chapters and each chapter will be discussed on specific topics. As for chapter 1, it has discussed on the background of the study and in chapter 2, we will be discussing more about the literature review. Next, in the chapter 3, we will discuss more on the methodology approach in completing this project so that the project is in line with the proposed milestone that have been planned before in the proposal. Then, in the chapter 4, we will explain regarding the implementation phase for data analytics have been described and in the chapter 5, we will be focusing more on the analysis of the data to retain the result of this analysis. Lastly, in the chapter 6, all the weakness and strength of the project is identified and suggestion for project improvement is also given in order for the system to be fully functional for telecommunication service provider company to utilize all this information for the sake of their business organization.

CHAPTER II

2. LITERATURE REVIEW

2.1 Introduction

This part gives a review of past exploration on several related topics such as marketing campaigns, business intelligence, data analysis and etc. It presents the idea on how the industry work nowadays and how machine learning can be applied in improving and optimizing the business decisions that decides the success or failure of a telecommunication industry.

This part also talks about several machine learning techniques that is used in the industry today for the sole purpose of indication what techniques will be used in this project.

2.2 Facts and Finding

2.2.1 Business Intelligence

Business leaders or managers deals with daily decision making that could make or break their company success. All the decisions that have been made is based on the reliable facts and information that they have or gain. This specific decision making is a critical process where it requires some critical reasoning and logic to ensure no loss gain after decision have been made. This is the classic way of how business works in an organization. In short, if you are a business leader and if you don't have any experience in doing business, you tend to loss more than gaining profit.

The term 'Business Intelligence' is coined by Richard Miller Devens in order to portray how a banker named Sir Henry Furnese increase profit by utilizing information about his business environment compare to his competitors. (Devens & Devens, 1865). And in the age of Information Technology globalization, Business Intelligence have been automated into a system originated the decision support systems(DSS) that began in the 1960s but is still based on Richard Miller Devens concept in portraying the utilization of data for increasing business performance.

Figure 2.1 Cyclopedia of commercial & business anecdotes

In general, the challenges in business intelligence is not about organizing and reporting data, but, it is beyond that, it is about understanding data and visualizing it ways that are relevant to contribute the greater insight and business acuity to the organizations.

Figure 2.2 Data analytics dashboard for visualizing data

Speaking of impact, here are few impacts of Business intelligence to business units such as want to empower the business users with better tools to increase productivity and come out with cost effective decision. (Sullivan, 2011). Then, want to provide the better support for data management for modelling and analytics. (Sullivan, 2011)

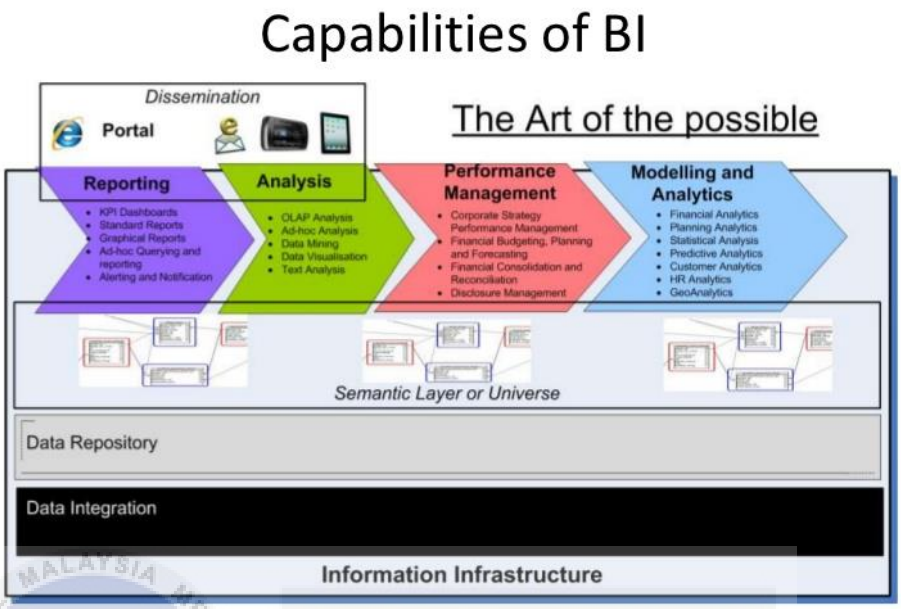

Figure 2.3 Impact of Business Intelligence

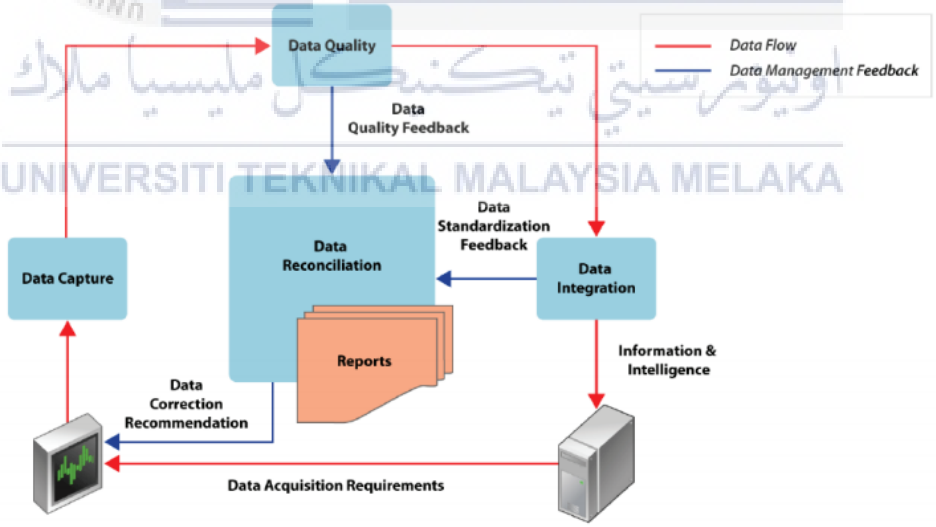

Figure 2.4 Business Intelligence Process Data Flow