

Universiti Teknologi MARA

**Mining Textual Terms for Market
Prediction Analysis Using Financial
News**

Mohamad Asyraf Safwan Bin Ab Rahman

**Thesis submitted in fulfillment of the requirements for
Bachelor of Information Technology (Hons.)
Intelligent Systems Engineering
Faculty of Computer and Mathematical Sciences**

August 2017

ACKNOWLEDGEMENT

“In the name of Allah, the most Gracious and the most Merciful”

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor, Assoc. Prof. Dr. Shuzlina Abdul Rahman, for her continuous support and guidance throughout the semesters in order to complete this project.

Special appreciation also goes to The Edge Markets for their cooperation on the use of their articles as the dataset for this research. Without their help, this research cannot be completed successfully.

Besides that, special thanks go to both of my family, especially my beloved mother, who continuously support me for 23 years. Without their support and caring, this final year project cannot be accomplished.

Many thanks also go to the both CSP600 and CSP650 lecturers, Dr. Azlin Ahmad, Pn. Nurzeatul Hamid and Assoc. Prof. Dr. Shuzlina Abdul Rahman for their support, knowledge and motivation throughout the semesters. Last but not least, this sincere thank would be given to all members of CS 243 6A and friends for their laughs and bad jokes. Finally, we have finished this final year project successfully.

ABSTRACT

Stock market is uncertain and it is not entirely predictable. Financial investors use fundamental analysis and technical analysis to study the movement of stock prices. Fundamental analysis makes use of quantitative and qualitative factors such as financial news to drive investment decision. But, it is tedious to digest textual articles from numerous news sources. Besides, Bursa Malaysia sees little studies on the use of algorithms to predict stock prices. This study focuses on the use of machine learning algorithms to construct a model that is able to predict Bursa Malaysia stock prices. In this research, we concentrate to identify linguistics terms from financial news that can contribute to the movement of stock prices and to develop a prototype that can classify sentiment of financial news for investment decision. We experimented with five companies from different industries of the top market constituents in FBMKLCI. First, financial news of these companies from The Edge Markets amounted to 14,999 articles were crawled to be used as the dataset. In addition, the dataset was processed and transformed into TF-IDF representation. SVM was used to train and test the dataset, and the accuracy recorded was measured to be at 56%. Hopefully, the findings of this research can be used to assist investors in investment decision making.

TABLE OF CONTENTS

CONTENTS	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi

CHAPTER ONE: INTRODUCTION

1.1	Background of Study	1
1.2	Problem Statements	2
1.3	Research Questions	3
1.4	Project Objectives	3
1.5	Research Scope	3
1.6	Project Significance	4
1.7	Outline of Thesis	5
1.8	Summary	6

CHAPTER TWO: LITERATURE REVIEW

2.1	Stock Market	7
2.2	Text Mining	8
2.3	Machine Learning	9
2.3.1	Support Vector Machines	11
2.3.2	k Nearest-Neighbor	12
2.3.3	Artificial Neural Network	12

CHAPTER 1

INTRODUCTION

This chapter discusses an overview on the background of study of the project. It also includes the problem statement, project motivation, research questions, objectives, research scope, significance and the outline of the thesis.

1.1 Background of Study

Public companies trade their shares in either stock exchanges or over-the-counter markets. Shares can be defined as a claim on the company's assets and earnings. Stock prices get determined by the supply and demand of that particular shares. It is known that investor's trading that drives the supply and demand of stocks. S are focused on two approaches which are the fundamental and technical analysis. The former focuses on the use of quantitative and qualitative factors for prediction. However, the latter is focused on the use of charts and tools on historical data primarily on the price and its volume significantly.

Shumaker and Chen (2006) argued that the dynamic nature of stock market makes prediction on stock market had not been entirely successful. Shynkevich, McGinnity, Coleman and Belatreche (2015) and Fama (1965) identified that news and unemotional investors can affect stock prices significantly. Hagenau, Liebmann and Neumann (2013) associated stock price movement to the release of financial news resulted by the investors' sentiments. Previous studies have supported this relationship between the release of a news to the movement of a stock price (Fung, Yu, & Lu, 2005).

However, it is difficult to predict stock prices from textual data (Shumaker & Chen, 2006). Fundamentally, these articles usually come verbosely and lengthy in textual representation thus extracting patterns