CHAPTER I

INTRODUCTION

1.1 Background

Twitter is an online *social networking* service that enables users to send and read short 140-character messages called "tweets". Twitter was created in March 2006 by Jack Dorsey, Evan Williams, Biz Stone and Noah Glass and launched by July 2006. The service rapidly gained worldwide popularity, with more than 100 million users who in 2012 posted 340 million tweets per day [1]. As twitter use grows, the influence and reputation of a person or business entity are increasingly associated with their number of Twitter followers. A recent study [2] found that nearly 80% of consumers would more likely be interested in a company because of its brand's presence on social media. It is therefore not surprising that 77% of the Fortune 500 companies have active Twitter accounts and 70% of them maintain active Facebook accounts to engage with their potential customers. On the third party services like *Klout* will estimate the influence of accounts ranging from normal users to celebrities and politicians [3] based on a series of features such as the number of followers and the frequency with which content is re-shared.

To aim the approach and investigate the use of both unsupervised and supervised learning methods for target audience classification on Twitter with minimal annotation efforts. Topic domain were automatically discovered from contents shared by followers of an account owner using *Twitter Latent Dirichlet Allocation (LDA)* [4]. A *Support Vector Machine (SVM)* ensemble was then trained using contents from different account owners of the various topic domains identified by Twitter LDA [5]. In their attempt to shut down market operations, Twitter blocks the OAuth applications that are used by such schemes. However, pyramid merchants overcome this problem by periodically creating new OAuth applications and using the user credentials to authorize such applications [6]. Furthermore, to hide their involvement in any followers market, customers who purchase followers typically add these followers slowly. In fact, some follower markets advertise that it can take up to one month for a customer to add 3.000 followers.

Twitter accounts that actively interact with followers markets, essentially building the ground-truth data on the customers who bought followers from the markets who were compromised by the markets and traded as followers. My data collection includes two steps. First, we locate popular follower markets. Second, we collect accounts of customers as well as legitimate Twitter users for further analysis.

1.2 Studies Issues

According to the background that has been presented, it can be a statement of problems in this case, as follows:

1. How to analyze and implement *Support Vector Machine* method to deliver followers to their customers to distinguish between the real followers and benign website?
2. How to analyze the result from *Support Vector Machine* method?
3. How to analyzing the factors that will help Twitter to keep its stabilized position, if it is reached?

Limitation problems on this research are:

1. Using the data set from the most popular account on Twitter every week
2. Using the data set from the most popular Twitter markets

1.3 Research Objectives

From several existing problems, the goals of this research are:

1. To analyze and implement *Support Vector Machine* method to deliver followers to their customers to distinguish between real followers and benign website
2. To analyze the result from *Support Vector Machine* Method
3. To analyze the influence of followers growing on Twitter

1.4 Hypothesis

The hypothesis from this research is predicting the popularity and effect on Twitter's ecosystem and also to predict which one is the best twitter followers market on social media who provide the real followers with *Support Machine Methods* with prediction market lead to 70% accuracy [7].

1.5 Systematics Writing

This research has a few methods to solve the problem, such as:

1. Introduction
   This chapter was discusses about the background, studies issue, research objective, hypothesis, and solving problem.
2. Literature
   This chapter explains the brief description of twitter, online social network, klout, oauth application, twitter latent dirichlet allocation, support vector machine, pyramid merchant, and sentistrength.
3. System Design
   This chapter discusses the system design analysis such as, data set, inputting data, raw data, preprocessing data, filtering data user account from twitter market, and support vector machine classification.

4. Testing and Analysis

   This chapter discusses the inner system design analyzing an outcome from *Support Vector Machine* method that generated the performance from the architecture that have been designed. Then, we conclude from a comparison of the data result and implementation system with actual data are available.

5. Conclusion

   This chapter contains the conclusion of the final assignment result and suggestion for further development.