

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.