CHAPTER 1

INTRODUCTION

1.1 Problem Background

Unlike traditional media, social media serves as a more open platform. However, the rise of fake users and internet trolls has become a significant issue that is threatening data quality. The platform is often inundated with irrelevant data, while the internet is filled with non-compliant, emotional, repetitive, and meaningless content. Furthermore, some users distort their own views of the perspectives of others.

Meanwhile, web surveys hardly verify whether respondents are real individuals, and the results are easy to teamper. In contrast, social media serves as a more reliable and genuine source of data.

Real pulic reactions to various topics provide insigths that can help enterprises and governments make more informed decisions in response to real-world changes and challenges. At the same time, these organizations can decrease the costs with traditional questionnaires, which benefits their financial standing.

1.2 Problem Background

The weaknesses of the current working model highlight the need for new data sources and a more efficient, sophisticated approach. Utilizing data science and

machine learning techniques to analyze topic-based social media posts will help enterprises and governments identify valuable data to address these challenges.

1.3 Problem Statement

Social media posts are presented in various data formats such as text and emojis often mixed with irrelevant content and influenced by emotions. Traditional survey methods frequently produce false or unreliable results, which often deviate significantly from the actual situation.

1.4 Research Goal

Collecting and filtering social media (X) posts by topic, and using meachine learning techniques to analyze them, yield valuable and authentic data about the topic.

1.4.1 Research Objectives

The objectives of the research are :

(a) To identify significant relationships between the content of posts and the topic

(b) To build and develop analytical models that capture the topic inclination of posts

(c) To measure public reactions to the topic by summarizing the analysis results