

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter introduces the combination of research methodologies applied to address the objectives of this study, which focus on analyzing topic-based social media posts through RNN and LSTM models.

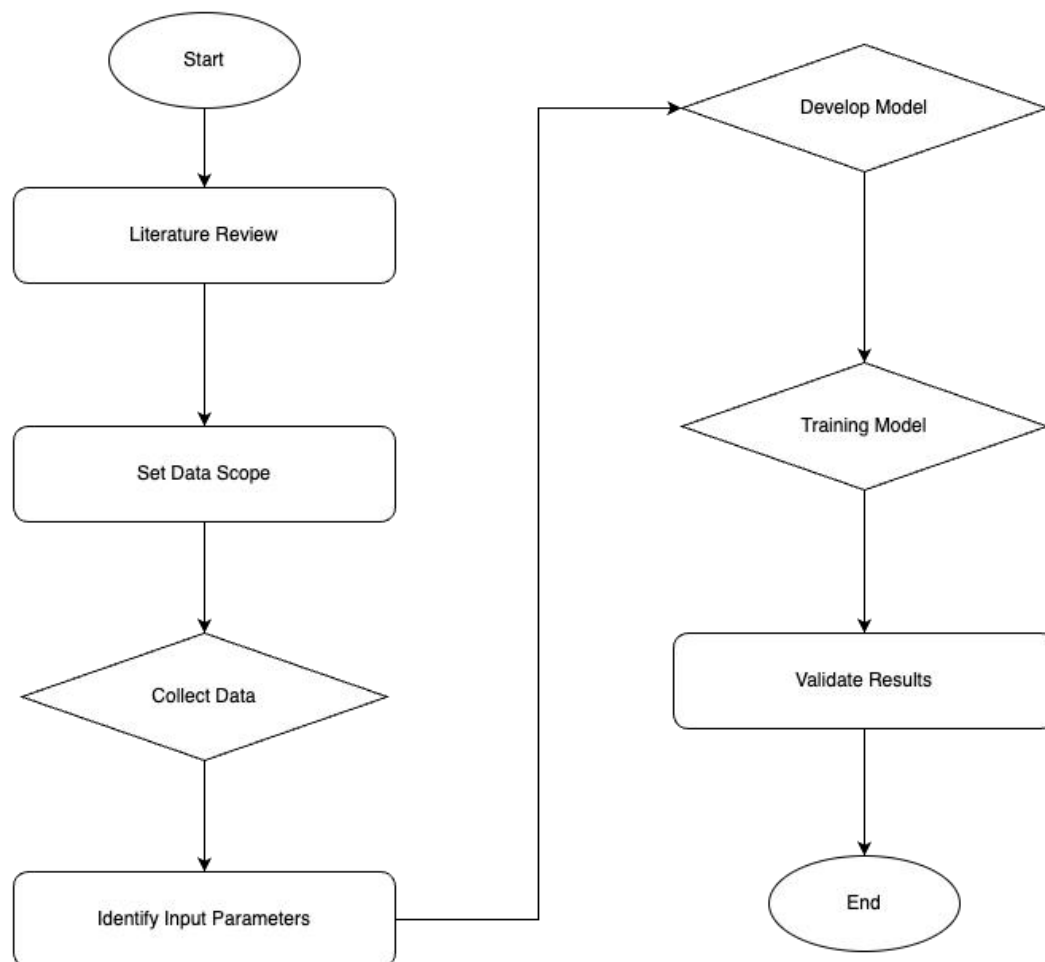
Using deep learning technology is the most common approach in this research field. RNNs and LSTMs can resolve and identify complex data from social media platforms. This chapter introduces the research framework and discusses the development of the models. It also elaborates on the input metrics and addresses the reliability and availability of the framework.

3.2 Research Framework

The full framework is split into steps. The study will begin by identifying the topics that serve as the data sources. High-quality data significantly improves the research's efficiency. As shown in the previous chapter, clearly defined topics set the boundaries of the research scope.

Building on the successful research of similar studies, the framework will be developed based on proven technologies, specifically RNN and LSTM models. The framework will be tested by adjusting parameters to achieve more accurate analysis results. It will also focus on the performance of the framework to ensure it has the ability to resolve various data types and sizes.

According to the scope of the data topics, data will be specifically collected from the Twitter platform, with data cleaning ensuring it can be accessed by the framework. A database is essential for storing the collected data.



3.3 Data Science Life Cycle

The data science life cycle includes several key stages that focus on the process of gaining insights from metadata.

3.3.1 Data Collection

This study uses an agent framework to subscribe to the social media platform, Twitter. Its responsibility is to collect data from the platform and identify specific topics, including the 2024 U.S. election, the Bitcoin market, and Tesla's stock price. Twitter provides an open API called Tweepy in Python, which retrieves the original data from the platform.

3.3.2 Data Pre-processing

Data pre-processing is a critical step in cleaning data that has been collected with noise and irrelevant information. This step eliminates junk data to ensure high data quality and usability. The cleaning process begins with filtering out spam, ads, and retweets. Next, emoji content is translated into text to ensure the framework can understand the posts. After that, the program removes stop words, performs word reduction (stemming or lemmatization), and removes punctuation.

This step also addresses missing data in the social media content, making the dataset more complete.

3.3.3 Build Analysis System

This study employs Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. LSTMs, a type of RNN, are designed to address the problem of analyzing longer sequences. Given the challenges of complex content and large context, LSTMs are the most popular solution.

The model is trained using an efficient framework that requires high-quality data and useful input parameters to generate meaningful insights. The project builds test cases to identify the best solution for the system. The results are validated using the provided dataset.

The cleaned data is saved into a PostgreSQL database to store the collected data and hold the results of the analysis.

3.3.4 Final Report

Based on the stored result data from the PostgreSQL database, the final report summarizes the results, including data visualizations, a summary highlighting the general sentiment, and providing recommendations. The final report helps researchers quickly gain insights from the project.

3.5 Chapter Summary

This chapter lays the essential foundation for the development and evaluation of the LSTM model and discusses how to build the project. It sets the data scope and size, aiming to address the objectives the framework is designed to achieve. The chapter identifies solutions for data collection and storage and emphasizes that pre-processing is crucial to the entire process. Pre-processing impacts both the model training and the analysis results. The framework also focuses on performance and input accuracy. The overall goal is to generate a useful report from the original social media data.