Sentiment Analysis and User Behavior Prediction in Social Networks

LIU MINGJIE

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 3

INTRODUCTION

3.1 Introduction

This research chapter unfolds the methodologies applied to studying user sentiment analysis as well as behavior prediction in social networks. The research design, data sources, processing methods, and analytical tools are outlined in detail to provide a structured and reproducible framework for the study. Each step in the data science project lifecycle is discussed to ensure clarity and rigor..

3.2 Research Design

The research design entails the use of two techniques, namely explorative and hypothetical methods. In these models, data is computed and adjusted for both qualitative and quantitative results, and the sentiment analysis and behavior prediction is run simultaneously. Figure 3.1 illustrates the framework for the plan as described in this paper.

The design aims to:

1.Identify the research problem and propose possible solutions.

2. Ensure a choice of appropriate methods and tools for data acquisition and processing.

3. Implement intelligent machine conjoint models to maintain the psychiatrists' objectives concerning the analysis.

4. Assess the outcome and its implication for formulating such decision proposals.

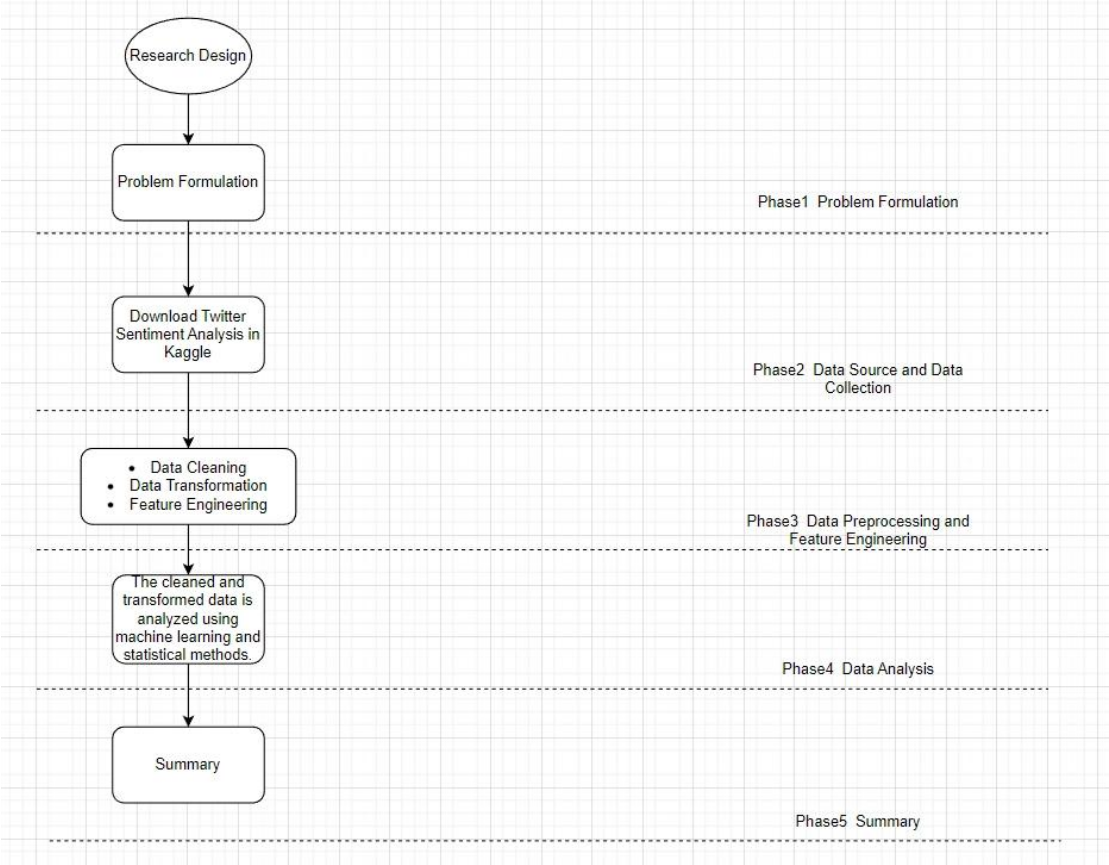

Figure 3.1: Research Framework of Proposal

3.3 Data Science Project Life Cycle

This stage of the study is based on the traditional data science project life cycle, which consists of posing the problem, data collection, preparation, analysis, as well as evaluation.

3.3.1 Problem Formulation

The book is to walk you through the most common problem in building interpretable and accurate models for sentiment analysis and user behavior forecasting based on social network data. Sub-questions include:

1. Option or options that work better at predicting the sentiment and the behavior of users.
2. What is the effect of different cultural and linguistic relationships on the model's performance?
3. Can explainable AI techniques help in achieving better model interpretability?

3.3.2 Data Source and Data Collection

The data is gathered from several social media networks, such as, e.g., Twitter or Reddit. Many of these procedures include:

1. Identifying Relevant Data: Filtering posts and comments related to predefined topics.
2. Ensuring Data Quality: Referring to spam, duplicate entries, and irrelevant information.
3. Ethical Considerations: Defendants have the right to privacy and should comply with platform rules..

3.3.3 Data Pre-Processing

1 Data Cleaning:

Cleaning up the missing and tinted data.

Several data types like types, punctuation, and some symbols can introduce noise. Filters can help remove such noise.

Text Normalization, which includes the transformation of text into some format, e.g. Case Normalization, Lemmatization.

2 Data Transformation:

Word tokens and phrases can be used as a text.

Categorical variables are usually encoded with two possible methods (verbose description goes here). Word embeddings, such as Word2Vec or TF-IDF, can be used for this conversion.

3 Feature Engineering:

Extract of linguistic features that gives the idea about sentiment score as well as n-gram.

The second thing is the identification of features that depend on the behavior of users, e.g. interaction patterns and levels of activity.

The last thing is developing some domain-specific features because of the research context which we have been talking about..

3.4 Data Analysis

The data is pre-processed and analyzed by using statistical and machine learning techniques. Key steps include:

Model Training: The application of models for classification, regression, and sequence projections.

Model Evaluation: Considers as performance metrics accuracy, precision, recall, F1, Mean Squared.

Cross-Validation: The establishment of a train/test dataset splits helps in ensuring a generalizability of models.

Exploratory Data Analysis (EDA): Generating data plots that illustrate the way data is distributed, correlate between each other, as well as trends, with the purpose of accumulating knowledge.

3.5 Summary

his chapter explicitly explained in detail the research approach and methodology regarding the problem statement, the primary data collection, and preprocessing the data. It indicated the application of a clear process of getting ready the data for sentiment analysis and human user behavior prediction. The detailed data processing steps, consequently, provide a thorough and accurate investigation, the end product being new information and novel perspectives emerging from the subsequent chapters.