Chapter 15
Unlocking Insights: Ethical Considerations and Classifications of Data Analytics for Social Networks

Madhav Narayan Singh
Shell India Markets Pvt. Ltd., India

Shanmuga Raja B.
Department of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India

Anakath A. S.
Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India

R. Kannadasan
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Prabakaran N
https://orcid.org/0000-0002-1232-1878
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

ABSTRACT

The social media analytics, which relate to huge amount of data from various social media platforms, are used to understand an opinion from the written language such as tweets, chats, comments. In existing methods, the sentiment analysis on Xcorp (Twitter) was mostly used for emotion detection for the polling methods; and star ratings are used to see the response from people. In this model, using Twitter API to fetch the data and Naive Bayes model for classifying them. The tweets, retweets and comments are collected and processed with the positive, neutral, and negative responses from the user will be reflected for ethical considerations. The information obtained from this system is used in various applications like analysis of social media support for politicians, safety technology in social media, a review based on user response for a product, response from people or government for social or political issues (Hashtags), movies, etc. The system will help to visualize statistics to analyse people's responses to provide the most effective statistical tool for various industries.

DOI: 10.4018/979-8-3693-2679-4.ch015

INTRODUCTION

Social media platforms are one of the most utilized platforms by people. The enormous population uses it more often nowadays to share their thoughts fearlessly regarding various political, social, economic, and business issues. Social Media Analytics (SMA) refers to the process of social media platform data analysis. It involves the collection of a vast range of data from various media platforms like Twitter, Facebook, and Instagram sites, which are then pre-processed, analyzed, and visualized by applying various NLP and ML algorithms, and this paradigm is usually used to build various analytical tools. It gathers analyzed data of people and visualizes them as useful information. It also helps to understand and analyze the unique data and also shows how well social media is performing in day-to-day life.

Opinion mining, also known as sentiment analysis, involves analyzing and understanding the opinions, emotions, and attitudes expressed by users towards a given subject. It is a process that aims to extract and evaluate the sentiments conveyed in text data, enabling us to gain insights into people's perspectives and viewpoints. There are different types of sentiment analysis, which are Fine-grained analysis models, Intent analysis, Emotion-based models, Aspect analysis, etc. Our model seeks to implement Twitter sentiment analysis based on Emotion that enables us to know what's being said by people about any product, movie, politician, or celebrity, and the retweets being positive, neutral, and negative responses. And that helps the people to know about it more. The datasets from Twitter are collected using a Twitter API, which is a tool provided by Twitter for accessing data. Along with it, a Tweepy Python package is also used. The Pre-processing of collected texts can be done by various NLP methods like Tokenization, stop word removal, etc. Finally, to classify data based on emotions, the Naive Bayes classification approach is used here, which is known as a strong assumption of the features.

The Naive Bayes classification is a Supervised Machine Learning algorithm, and they are also a probabilistic classifier. This algorithm works based on the Bayes theorem. They are the popular baseline method category for the texts and judging the features. The foremost advantage of this is it requires only a minimum amount of data for the classification. This could correctly classify users' data of the given datasets as positive, neutral, or negative responses by using the Bayes probabilistic model. The model is built by training various datasets with their probability values and building a strong Bayesian model. The Input datasets are compared with the Model and classified accordingly as positive, neutral, or negative based on probability values. The basic logic of Naive Bayes is expressed as,

$$P\left(\frac{X_i}{C=C_k}\right)=P\left(\frac{X_i=1}{C=C_k}\right)*P\left(\frac{X_i=2}{C=C_k}\right)*\ldots.*\frac{X_i=m}{C=C_k} \tag{1}$$

$$P\left(\frac{X}{C}\right)=\pi_{i=1}^{n}\ P\left(\frac{X_i}{C}\right) \tag{2}$$