

# Chapter 30

## Sentiment Analysis of Tweets Using Naïve Bayes, KNN, and Decision Tree

**Kadda Zerrouki**

*Higher School of Computer Science May 8, 1945, ESI Sidi Bel Abbes, Algeria*

**Reda Mohamed Hamou**

 <https://orcid.org/0000-0002-0388-1275>

*GeCoDe Labs, University of Saida Dr Moulay Tahar, Algeria*

**Abdellatif Rahmoun**

*Higher School of Computer Science May 8, 1945, ESI Sidi Bel Abbes, Algeria*

### ABSTRACT

*Making use of social media for analyzing the perceptions of the masses over a product, event, or a person has gained momentum in recent times. Out of a wide array of social networks, the authors chose Twitter for their analysis as the opinions expressed there are concise and bear a distinctive polarity. Sentiment analysis is an approach to analyze data and retrieve sentiment that it embodies. The paper elaborately discusses three supervised machine learning algorithms—naïve bayes, k-nearest neighbor (KNN), and decision tree—and compares their overall accuracy, precision, as well as recall values, f-measure, number of tweets correctly classified, number of tweets incorrectly classified, and execution time.*

## INTRODUCTION

Twitter is a popular micro blogging service where users create status messages (called “tweets”). These tweets sometimes express opinions about different topics. We propose a method to automatically extract sentiment (positive or negative) from a tweet.

Sentiment Analysis is the process of finding the opinion of user about some topic or the text in consideration. It is also known as opinion mining. In other words, it determines whether a piece of writing is positive or negative.

Sentiment analysis is a process where the dataset consists of emotions, attitudes or assessment which takes into account the way a human thinks, as noted by Feldman Ronen (Feldman, 2013). In a sentence, trying to understand the positive and the negative aspect is a very difficult task. The features used to classify the sentences should have a very strong adjective in order to summarize the review. These contents are even written in different approaches which are not easily deduced by the users or the firms making it difficult to classify them.

This task has received a lot of interest from the research community in the past years. The work is regarded the manner in which sentiment can be classified from texts pertaining to different genres and distinct languages, in the context of various applications, using knowledge-based, semi-supervised and supervised methods, as noted by Liu Bing (Liu, 2011). The result of the analyses performed have shown that the different types of text require specialized methods for sentiment analysis, as, for example, the sentiments are not conveyed in the same manner in newspaper articles and in blogs, reviews, forums or other types of user-generated contents, as noted by Balahur Alexandra and al (Balahur, Steinberger, Kabadjov, Zavarella, Van Der Goot, Halkia & Belyaeva, J. 2013).

The Sentiment found within comments, feedback or critiques provide useful indicators for many different purposes and can be categorized by polarity, as noted by Kalaivani and Shunmuganathan (Kalaivani & Shunmuganathan, 2013). By polarity we tend to find out if a review is overall a positive one or a negative one. For example:

- **Positive Sentiment in Subjective Sentence:** “I loved the movie Mary Kom”: This sentence is expressed positive sentiment about the movie Mary Kom and we can decide that from the sentiment threshold value of word “loved”. So, the threshold value of the word ‘loved’ has positive numerical threshold value;
- **Negative Sentiment in Subjective Sentences:** “Phata poster nikla hero is a flop movie” defined sentence is expressed negative sentiment about the movie named: “Phata poster nikla hero” and we can decide that from the sentiment threshold value of a word: “flop”. So, the threshold value of a word: “flop” has negative numerical threshold value;
- **Sentiment Analysis is of Three Different Types:** Document level, Sentence level and Entity level (Kiritchenko, Zhu, & Mohammad, 2014).

The difficulties in Sentiment Analysis are an opinion word which is treated as positive side may be considered as negative in another situation. Also the degree of positivity or negativity also has a great impact on the opinions. For example: “good” and “very good” cannot be treated same. Although the traditional text processing says that a small change in two pieces of text does not change the meaning of the sentences (Kalaivani & Shunmuganathan, 2013). However the latest text mining gives room for

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/sentiment-analysis-of-tweets-using-nave-bayes-knn-and-decision-tree/308507](http://www.igi-global.com/chapter/sentiment-analysis-of-tweets-using-nave-bayes-knn-and-decision-tree/308507)

## Related Content

---

### Ranking of Evaluation Targets Based on Complex Sequential Data

Shigeaki Sakurai (2017). *International Journal of Data Warehousing and Mining* (pp. 19-32).

[www.irma-international.org/article/ranking-of-evaluation-targets-based-on-complex-sequential-data/188488](http://www.irma-international.org/article/ranking-of-evaluation-targets-based-on-complex-sequential-data/188488)

### Pattern Discovery from Biological Data

Jesmin Nahar, Kevin S. Ticklead A. B.M. Shawkat Ali (2010). *Dynamic and Advanced Data Mining for Progressing Technological Development: Innovations and Systemic Approaches* (pp. 168-223).

[www.irma-international.org/chapter/pattern-discovery-biological-data/39643](http://www.irma-international.org/chapter/pattern-discovery-biological-data/39643)

### Improved Approximation Algorithm for Maximal Information Coefficient

Shuliang Wang, Yiping Zhao, Yue Shuand Wenzhong Shi (2017). *International Journal of Data Warehousing and Mining* (pp. 76-93).

[www.irma-international.org/article/improved-approximation-algorithm-for-maximal-information-coefficient/173707](http://www.irma-international.org/article/improved-approximation-algorithm-for-maximal-information-coefficient/173707)

### A New Outlier Detection Algorithm Based on Fast Density Peak Clustering Outlier Factor

ZhongPing Zhang, Sen Li, WeiXiong Liu, Ying Wangand Daisy Xin Li (2023). *International Journal of Data Warehousing and Mining* (pp. 1-19).

[www.irma-international.org/article/a-new-outlier-detection-algorithm-based-on-fast-density-peak-clustering-outlier-factor/316534](http://www.irma-international.org/article/a-new-outlier-detection-algorithm-based-on-fast-density-peak-clustering-outlier-factor/316534)

### Preserving Privacy in Time Series Data Mining

Ye Zhu, Yongjian Fuand Huirong Fu (2011). *International Journal of Data Warehousing and Mining* (pp. 64-85).

[www.irma-international.org/article/preserving-privacy-time-series-data/58638](http://www.irma-international.org/article/preserving-privacy-time-series-data/58638)