

Social Media Sentiment Analysis Using Machine Learning

K.Solai priya and J. Malathika

Mahath Amma Institute of Engineering and Technology, Ariyur. Pudukkottai-622101.

ARTICLE INFO

Article history:

Received: 13 August 2023;

Received in revised form:
12 September 2023;

Accepted: 27 September 2023;

Keywords

Social Network,
Natural Language Toolkit.

ABSTRACT

In today's world, social networking website like twitter, Facebook, tumbler, etc. plays a very significant role. twitter is a micro-blogging platform which provides a tremendous amount of data which can be used for various applications of sentiment analysis like predictions, reviews, elections, marketing, etc. sentiment analysis is a process of extracting information from large amount of data, and classifies them into different classes called sentiments. python is simple yet powerful, high-level, interpreted and dynamic programming language, which is well known for its functionality of processing natural language data by using nltk (natural language toolkit). nltk is a library of python, which provides a base for building programs and classification of data. nltk also provide graphical demonstration for representing various results or trends and it also provide sample data to train and test various classifiers respectively.

© 2023 Elixir All rights reserved.

Introduction

Sentiment Analysis is process of collecting and analyzing data based upon the person feelings, reviews and thoughts. Sentimental analysis often called as opinion mining as it mines the important feature from people opinions. Sentimental Analysis is done by using various machine learning techniques, statistical models and Natural Language Processing (NLP) for the extraction of feature from a large data. Sentiment Analysis can be done at document, phrase and sentence level. In document level, summary of the entire document is taken first and then it is analyze whether the sentiment is positive, negative or neutral. In phrase level, analysis of phrases in a sentence is taken in account to check the polarity. In Sentence level, each sentence is classified in a particular class to provide the sentiment. In our thesis we use Python as our base programming language which is used for writing code snippets. NLTK is a library of Python which plays a very important role in converting natural language text to a sentiment either positive or negative. The main objective of this thesis work is to perform the sentiment analysis on Indian Political Parties like BJP, INC and AAP, such that people opinions about these parties progress, workers, policies, etc. Which are extracted from Twitter. Thus to achieve this objective we build a classifier based on supervised learning and perform live sentiment analysis on data collected of different political parties.

People make judgments about the world around them when they are living in the society. They make positive and negative attitudes about people, products, places and events. These types of attitudes can be considered as sentiments. Sentiment analysis is the study of automated techniques for extracting sentiments from written languages. Growth of social media has resulted in an explosion of publicly available, user generated text on the World Wide Web. These data and information can potentially be utilized to provide

real-time insights into the sentiments of people. Blogs, online forums, comment sections on media sites and social networking sites such as Facebook and twitter all can be considered as social media. These social media can capture millions of peoples' views or word of mouth. Communication and the availability of these real time opinions from people around the world make a revolution in computational linguistics and social network analysis. Social media is becoming an increasingly more important source of information for an enterprise. On the other hand people are more willing and happy to share the facts about their lives, knowledge, experiences and thoughts with the entire world through social media more than ever before. They actively participate in events by expressing their opinions and stating their comments that take place in society. This way of sharing their knowledge and emotions with society and social media drives the businesses to collect more information about their companies, products and to know how reputed they are among the people and thereby take decisions to go on with their businesses effectively. Therefore it is clear that sentiment analysis is a key component of leading innovative Relationship Marketing focused enterprises. Moreover for businesses looking to market their products, identify new opportunities and manage their reputation. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and take appropriate action upon it. Many are now looking to the field of sentiment analysis. In the era which we live today, sometimes known as information age, Customer Experience Management and Customer knowledge society; having access to large quantities of information is no longer an issue looking at the tons of new information produced everyday on the web. In this era, information has become the main trading object for many enterprises. If we can create and employ mechanisms to search and

retrieve relevant data and information and mine them to transfer it to knowledge with accuracy and timeliness, that is where we get the exact usage of this large volume of information available to us. However, in many cases these relevant data and information are not found in structured sources such as tables or databases but in unstructured documents written in human language. Human languages are ambiguous and the same sentiment can be used to express two different ideas in two different contexts. Moreover some people use different jargon, slang communications and short forms of the words for their ease. Therefore, it is difficult to gauge and measure the sentiments accurately in terms of their polarity such as positive, negative or neutral and the subjectivity of sentiments. Most solutions in the market today rely on simple Boolean terms to express sentiment about a post, tweet, Facebook wall post etc. But this is not enough to address the above mentioned problems in the area of sentiment analysis and it will not generate precise and timely knowledge for aggregate sentiments. In order to get accurate knowledge after analyzing a sentiment, it should thoroughly consider solving the issues mentioned above. Most other systems that try to give solutions for these issues are still on research level, some systems also try to analyze sentiments from multiple languages and few systems which address some of the above mentioned drawbacks are available commercially also. This paper reveals an approach which is implemented as a tool that can analyze sentiments on twitter social media addressing above issues and then develop an application to generate knowledge that can be useful for business environments using people's attitudes about their products and services.

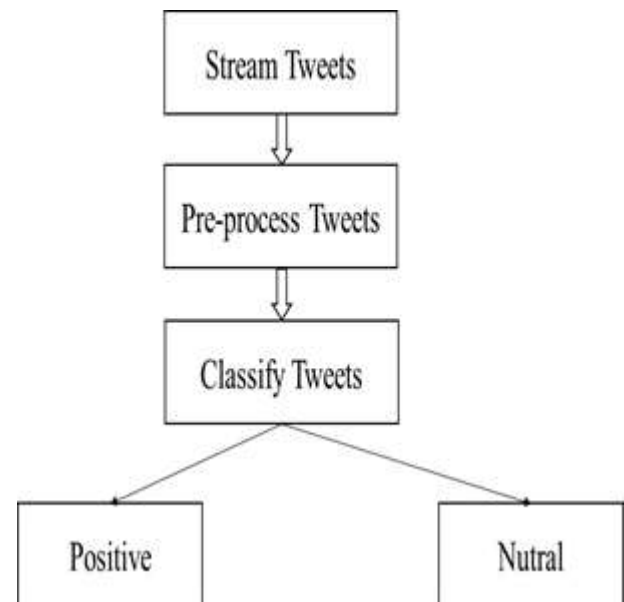
System Analysis

Data collection is not a simple task, as it may seem. Various decisions have to be made for collecting data. For our thesis we maintain dataset for training, testing and for twitter sentiment analysis. Here, we are going to study how data is collected, stored, processed and classified. Before discussing these process and different dataset, let us discuss our proposed architecture. As our goal is to achieve sentiment analysis for data provided from Twitter. We are going to build a classifier which consists of different machine learning classifiers. First we are going to stream tweets in our build classifier with the help of Tweepy library in python.

Advantages of Proposed System

- Social media is the current environment for data collection and analysis of sentiments of people.
- People can share and comment on everything, from personal thoughts to common events or topics in society.

The access to social media also can provide more information in the form of hidden metadata. Creating a working module for social media sentiment analysis involves a combination of data preprocessing, machine learning, deployment, and ongoing maintenance. The specific technologies and tools you use may vary based on your project's requirements and constraints. A thorough study of existing approaches and techniques in field of sentiment analysis. Collection of related data from Twitter with the help of Twitter API Pre-processing of data collected from Twitter so that it can be fit for mining. To build a classifier based on different supervised machine learning techniques. Training and testing of build classifier using large datasets Computing the result of different classifier using dataset collected from Twitter. Comparing results of each classifier and plotting a graph that show the trend of positive and negative sentiment for different political parties.



1. Data Collection

Collect data from social media platforms using their APIs or web scraping tools. Popular social media platforms like Twitter provide APIs for accessing data.

2. Data Preprocessing

Clean and preprocess the collected data. This includes tasks like removing special characters, stopwords, URLs, and performing tokenization and stemming or lemmatization.

3. Labeling

Label the data with sentiment labels (e.g., positive, negative, neutral) either manually or using rule-based methods. Alternatively, you can use pre-labeled datasets for supervised learning.

4. Feature Extraction

Transform the preprocessed text data into numerical features suitable for machine learning models. Common techniques include TF-IDF or word embeddings like Word2Vec or GloVe.

5. Model Selection

Choose a machine learning model for sentiment analysis. Common choices include Naive Bayes, SVM, Random Forest, or deep learning models like RNNs or Transformers.

6. Training the Model

Split your labeled data into training and testing sets. Train the selected model on the training data using the extracted features.

7. Evaluation

Evaluate the model's performance on the testing data using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

8. Hyperparameter Tuning

Optimize the model's hyperparameters to improve performance. You can use techniques like grid search or random search.

9. Deployment

Deploy the trained model to analyze real-time social media data. This can involve creating a web application or API to interact with the model.

10. Real-time Analysis

Integrate the module with social media data streams to perform real-time sentiment analysis.

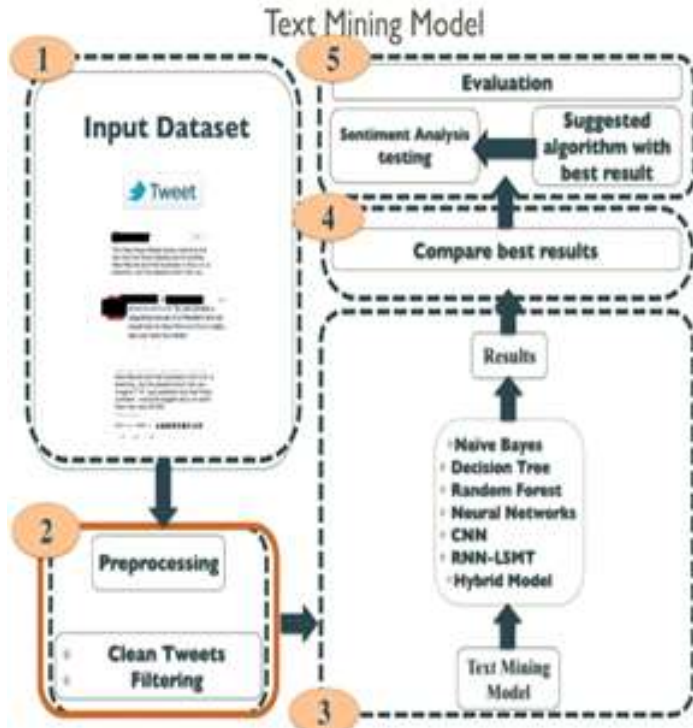
11. Monitoring and Maintenance

Continuously monitor the model's performance and retrain it periodically to adapt to changing trends and language usage on social media.

12. Scaling

If needed, scale your system to handle large volumes of social media data efficiently. This might involve distributed computing or cloud-based solutions.

Creating a working module for social media sentiment analysis involves a combination of data preprocessing, machine learning, deployment, and ongoing maintenance. The specific technologies and tools you use may vary based on your project's requirements and constraints.



Result

Social media sentiment analysis is the process of using natural language processing (NLP) and machine learning (ML) to identify and measure the emotions and opinions expressed by online users. It can help customer service teams to monitor customer feedback, improve satisfaction, and respond to issues more effectively. Twitter sentiment analysis determines whether a tweet is positive, negative, or neutral. You can do it manually by analyzing each tweet and evaluating whether it is positive or negative. But it is a time-consuming process. Collecting large amounts of unstructured data from various sources. Tracking real-time customer feedback and sentiment about an organization's brand, products and services. Providing feedback on ways to improve products, services and customer experience.



Conclusion

Sentiment analysis is used to identifying people's opinion, attitude and emotional states. The views of the people can be positive or negative. Commonly, parts of speech are used as feature to extract the sentiment of the text. An adjective plays a crucial role in identifying sentiment from parts of speech. Sometimes words having adjective and adverb are used together then it is difficult to identify sentiment and opinion. To do the sentiment analysis of tweets, the proposed system first extracts the twitter posts from twitter by user. The system can also computes the frequency of each term in tweet. Using machine learning supervised approach help to obtain the results. Twitter is large source of data, which make it more attractive for performing sentiment analysis. We perform analysis on around 15,000 tweets total for each party, so that we analyze the results, understand the patterns and give a review on people opinion. We saw different party have different sentiment results according to their progress and working procedure. It is general classifier. It can be used for any purpose based on tweets we collect with the help of keyword. It can be used for finance, marketing, reviewing and many more.

References

- [1] P Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, *Social Media Mining – An Introduction*, Cambridge University Press, Publisher Location, 2014.
- [2] Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S.Li, L. T., Shoemaker, D. J., Xie, L. *Social media use by the government: from the routine to the critical*. *Government Information Quarterly*, 29(4), 2012.
- [3] *Twitter Usage Statistics – Internet Live Stats*. (n.d.). Retrieved October 22, 2018, from
- [4] Liu, B., *Sentiment analysis and opinion mining*. *Synthesis lectures on human language technologies*, 2012.
- [5] Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, *Social media analytics*. *Business and Information system Engineering* (2014), 89-996.
- [6] Stieglitz, S., & Dang-Xuan, L., *Social media and political communication: a social media analytics framework*. *Social Network Analysis and Mining* 3(4), 2013, 1277-129.
- [7] Ratkiewicz, J., Conover, M., Meiss, M., & Gonçalves, B., *Detecting and Tracking Political Abuse in Social Media*. *ICWSM*, 2011.
- [8] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. *A System for Real Time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*. *Proceedings of 50th Annual Meetings of the Association for Computational Linguistics*, 2.12, 115-120.
- [9] Tayal, D. K & Yadav. *Sentiment Analysis on Social Campaign “Swachh Bharat Abhiyan” using unigram method*. *AI & Society*, 2016.
- [10] Barbera, P. *Birds of the Same Feather tweet together. Bayesian ideal point estimation using twitter data*. *Political Analysis*, 2015.
- [11] Barbera, P., & Rivero, G. *Understanding the Political Representativeness of Twitter Users*. *Social Science Computer Review* (2015), 712-729.
- [12] Farha Nausheen, Sayyada Begum, *Sentiment analysis to predict election results using python*, *Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018) IEEE Explore Compliant*, 2018, 978-1-5386-0807-4.
- [13] K. Schouten, O.V. Weijde, F. Frasinca, and R. Dekker, *Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with CoOccurrence Data*, *IEEE Transactions on Cybernetics*, 2017.
- [14] E. S. Chifu, T. Leția, and V. R. Chifu, *Unsupervised aspect level sentiment analysis using Ant Clustering and Self-organizing Maps*, In *Speech Technology and Human-Computer Dialogue (SpeD)*, 2015 *International Conference on*, pp. 1-9. *IEEE*, 2015.
- [15] K. Paramesha, and K. C. Ravishankar, *Exploiting dependency relations for sentence level sentiment classification using SVM*, In *Electrical, Computer and Communication Technologies (ICECCT)*, 2015 *IEEE International Conference on*, pp. 1-4, *IEEE*, 2015.
- [16] *Tags Google Sheet Template, Version 6.0*, from.
- [17] *Consumer Affairs*, <https://consumeraffairs.com/travel/uber.html>.
- [18] Sneha Passerate, Rajashree Shedge, *Comparative Study on Feature Extraction Techniques used Sentiment Analysis*, *international conference on innovation and challenges in cyber security (ICICCS)*, 2016.
- [19] Y. Yuan, X. Li, Y. Pang, X. Lu, and D. Tao, *Binary sparse nonnegative matrix factorization*, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 772–779, 2009.
- [20] Quoc Lec, Tomas Mikolov, *Distributed Representations of Sentences and Documents*, Google Inc.
- [21] Sebastian Roschka, *Naïve Bayes and Text Classification*, Corner University Library, 2014.
- [22] Kamal Nigan, John Lafferty, Andrew McCallum, *Using Maximum Entropy for Text Classification*, *IJCAI workshop on Machine Learning*.
- [23] Qing-Yun Dai, Chun-Ping Zahang and Hao Wu, *Research on Decision Tree Classification Algorithm in Data Mining*, *International Journal of Database Theory and Application* vol.9, 2016.
- [24] Eesha Goel, Er. Abhilasha, *Random Forest: A review*, *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, 2017.