

# A Survey on Sentiment Analysis and Topic Modeling

Liyansi Patel<sup>1</sup>, Vimal Rathod<sup>2</sup>

<sup>1</sup>PG Research Scholar, Computer Engineering, Government Engineering Collage Modasa, Gujarat, India

<sup>2</sup>Assistant Professor, Information technology Department Government Engineering Collage Modasa, Gujarat, India

## ABSTRACT

Sentiment Reason Mining is an emerging research area in this era of social media. Sentiment Reason Mining aims to resolve two problems: first is finding the reason of a sentiment, and second is interpreting sentiment variations. Time and Event where sentiment is being changed is also an important factor. Aspect-Based methods, Supervised Learning, Topic Modeling, and Data Visualization etc. can be used for finding the reason of a sentiment. VADER Sentiment Classifier can be used for sentiment of tweets. LDA is topic Modeling algorithm. In this research paper we have reviewed some the research work performed for this purpose. We have reviewed various research work which have used social media content as dataset. TF/IDF feature extraction is used in most of the work. Sentiment Detection tools VADER and Text Blob are also discussed in our work.

**Keywords :** Topic Detection, interpreting sentiment variations, opinion reason mining, Sentiment Analysis, Sentiment Reasoning, Sentiment Spikes, Topic Model, Artificial Intelligence, Machine Learning, LDA, VADER

## Article Info

Volume 9, Issue 2

Page Number : 149-154

## Publication Issue :

March-April-2022

## Article History

Accepted : 20 March 2022

Published: 30 March 2022

## I. INTRODUCTION

Opinions related to products, services or any topic on social media and other platforms are very useful tools for stock holders of the businesses as well as other entities. A huge quantity of reviews and format of reviews are main challenge for analysis from these reviews. Topic modeling from the text is also an important research subdomain for this purpose. To finalize the topic from big amount of short text is a challenging task. Other challenge in this research is unstructured data and multiple languages. Algorithms like LDA[1] can be used for topic modeling in this case.

Many researchers have opted for various versions of LDA.

## II. SENTIMENT ANALYSIS

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP)" [1]. There are various methodology to measure sentiment analysis; among them most of work using TF/IDF. Mostly it works using classification techniques. It classify the text into multiple class like 'Positive', 'Negative' and 'Neutral'. Following Figure shows the example of Sentiment Analysis. Mostly sentiment

analysis extracts text from its source eg. Social Media, preprocess like stop word removal etc. tokenize the remaining text and then apply algorithm for classification.

LDA

The Linear Discriminant Analysis or LDA is a margin, column width, column spacing and other features.



Figure 1 Example of Sentiment Analysis

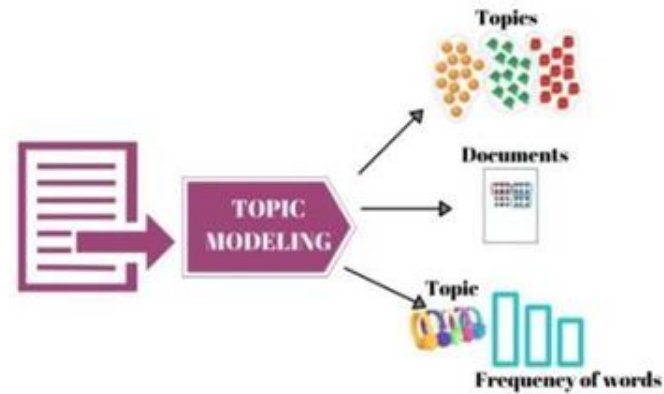


Figure 3. Topic Modeling

Various techniques are used for sentimental analysis. As per research work[1] sentiment analysis can be performed using any of method (supervised, unsupervised or semi supervised) in machine learning

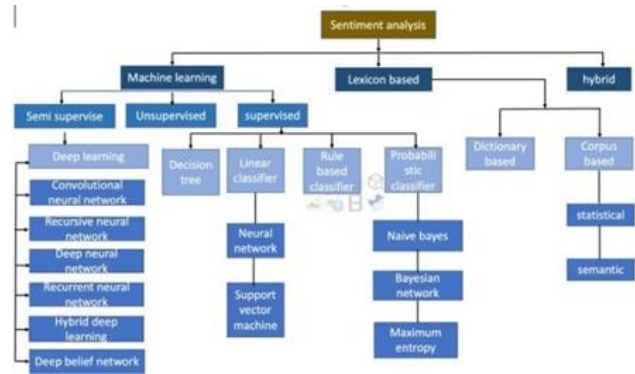


Figure 4. Main Sentimental Analysis Techniques [1]

TOPIC MODELING

Topic Modeling is a technique in NLP that is used for extraction of topic from various text or documents. In other word we can also say that it is clustering of text with similar content in particular subject. In topic modeling mostly tokenized words are applied. Various preprocessing techniques are also applied like Stop Word removal, non- ascii characters are removed etc. your final paper, check that the format conforms to this template. Specifically, check the appearance of the title and author block, the appearance of section headings, document topic modeling algorithm It is dimensionality reduction technique. This method is used as a pre-processing step in Machine Learning and applications of pattern and topic classification. The main goal of LDA is to extract the features in higher dimensional space onto a lower- dimensional space in order resources and dimensional costs. Following example shows relation how topic modeling is done using LDA.

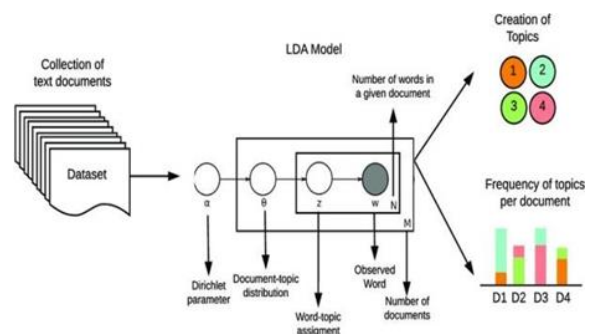


Figure 5. LDA Model[3]

## VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool. This tool is specifically attuned to sentiments expressed in social media. VADER is used for sentiment analysis of text which has both the polarities i.e. positive/negative. It does not require any training data. VADER can be used in multi domain also

## TextBLOB

TextBlob is a library for processing text content for sentiment extraction. This method provides a simple API for diving into common natural language processing (NLP) tasks. The NLP process contains part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. Polarity is calculated in TextBlob. For sentiment polarity score it uses Tokenization.

## SVM

SVM is a supervised method for machine learning algorithm. Use of SVM algorithm are in both classification or regression challenges. As we know classification of documents is predicting a label/group/topic and Regression is predicting a continuous value. SVM performs classification by finding the hyper-plane that differentiates the classes.

SVM can work in linear and non-linear challenges. Here  $x$  is a feature vector and  $w \cdot x - b = 1$  is hyper plane. Flow for SVM algorithm for Sentiment Analysis.

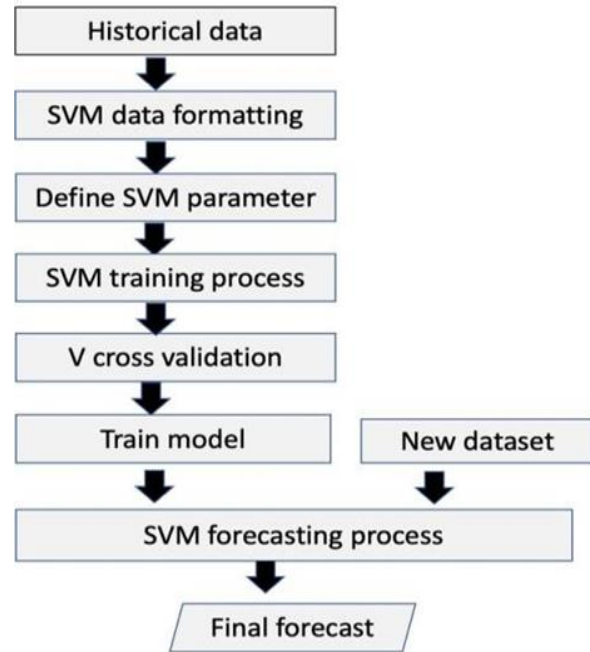
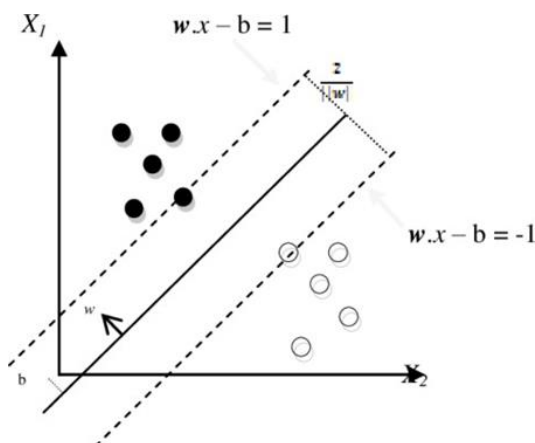


Figure 6. SVM Forecast for Classification Flow

## III. LITERATURE STUDY

In this research paper we have reviewed work related to sentimental analysis and topic modeling. Various researchers have used different techniques and approaches. Each technique has its own pro and cons. We have also discussed strong features and limitations of the same work.

In research work[1] Fuad Alattar and Khalid Shallan have presented a Filtered-LDA (FLDA) method of interpreting sentiment variations on Twitter. Their approach utilizes cascaded LDA Models[1] with multiple tuning of hyper parameters to get candidate reasons those are cause of sentiment changes. After that they applied a filter to remove content that is about old topics. At last they extract Emerging Topics that are interpretable by a human. Finally, a novel Twitter's sentiment reasoning dashboard is introduced to display the most representative tweet for each candidate reason. In their work they have applied VADER[16] for sentiment analysis. For training purpose they have applied Airlines Dataset from Twitter.





#### IV. COMPARATIVE STUDY

#### V. CONCLUSION

This section will go through the most commonly used methods and algorithms in sentiment analysis and topic modeling.

We have reviewed some of the research work related to topic modeling and sentiment analysis and found some basic issues related to optimized topic modeling. After reviewing the research work we can summarize that LDA with VADER can be a better option particularly when Social Media text is being used for sentiment analysis.

Table 2 Methods Overview

Method/Tools	Pros	Cons
VADER[1,3,15]	Better Option for Small Text like Social Media. Easy to Train Less Complexity	Not accurate for complex data Does not recognize context.
TF/IDF[1,2,4]	Similarity calculation is simple.	Position of word is not considered. No sentiment is measured.
LDA[1,3,4,5,6,15]	The LDA is a generative model, but in text mining, it introduces a way to attach topical content to text documents. Each document is viewed as a mix of multiple distinct topics.	Latent Dirichlet allocation when run on different datasets, LDA suffers from "order effects" i.e. different topics are generated if the order of training data is shuffled.
K-Means Clustering[8]	K-means is hard clustering model. It can be used with large dataset.	Can not provide better result when single context contain multiple topics.

As a part of future work Hybrid models of Sentiment analysis and topic model and transfer learning can be studied and applied.

#### VI. REFERENCES

- [1]. F. Alattar and K. Shaalan "Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media" in IEEE Access vol. 9 pp. 61756-61767 2021 doi: 10.1109/ACCESS.2021.3073657
- [2]. P. Gupta S. Kumar R. R. Suman and V. Kumar "Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter" in IEEE Transactions on Computational Social Systems vol. 8 no. 4 pp. 992-1002 Aug. 2021 doi: 10.1109/TCSS.2020.3042446.
- [3]. Christopher Whitfield Yang Liu and Mohd Anwar. 2021. Surveillance of COVID-19 pandemic using social media: a reddit study in North Carolina. In Proceedings of the 12th ACM Conference on Bioinformatics Computational Biology and Health Informatics (BCB '21). Association for Computing Machinery New York NY USA Article 43 1–8. DOI:https://doi.org/10.1145/3459930.3469550
- [4]. Xue J Chen J Chen C Zheng C Li S Zhu T (2020) Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS ONE 15(9): e0239441. https://doi.org/10.1371/journal.pone.0239441

- [5]. E. De Santis A. Martino and A. Rizzi "An Inveillance System for Detecting and Tracking Relevant Topics From Italian Tweets During the COVID-19 Event" in IEEE Access vol. 8 pp. 132527-132538 2020 doi: 10.1109/ACCESS.2020.3010033.
- [6]. Zoya S. Latif F. Shafait and R. Latif "Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling" in IEEE Access vol. 9 pp. 127531-127547 2021 doi: 10.1109/ACCESS.2021.3112620.
- [7]. G. Tao Y. Miao and S. Ng "COVID-19 Topic Modeling and Visualization" 2020 24th International Conference Information Visualisation (IV) 2020 pp. 734-739 doi: 10.1109/IV51561.2020.00129.
- [8]. J. Rashid et al. "Topic Modeling Technique for Text Mining Over Biomedical Text Corpora Through Hybrid Inverse Documents Frequency and Fuzzy K-Means Clustering" in IEEE Access vol. 7 pp. 146070-146080 2019 doi: 10.1109/ACCESS.2019.2944973.
- [9]. M. Bahja and G. A. Safdar "Unlink the Link Between COVID-19 and 5G Networks: An NLP and SNA Based Approach" in IEEE Access vol. 8 pp. 209127-209137 2020 doi: 10.1109/ACCESS.2020.3039168.
- [10]. E. S. Negara D. Triadi and R. Andryani "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method" 2019 International Conference on Electrical Engineering and Computer Science (ICECOS) 2019 pp. 386-390 doi: 10.1109/ICECOS47637.2019.8984523.
- [11]. <https://doi.org/10.1007/s13042-018-0798-5>
- [12]. Capdevila J. Cerquides J. & Torres J. Mining urban events from the tweet stream through a probabilistic mixture model. Data Min Knowl Disc 32 764-786 (2018). <https://doi.org/10.1007/s10618-017-0541-y>
- [13]. Dahal B. Kumar S.A.P. & Li Z. Topic modeling and sentiment analysis of global climate change tweets. Soc. Netw. Anal. Min. 9 24(2019). <https://doi.org/10.1007/s13278-019-0568-8>
- [14]. "Machine Learning Algorithms" <https://www.ibm.com/cloud/learn/machine-learning>
- [15]. "Concepts of Machine Learning" <https://www.mygreatlearning.com/blog/what-is-machine-learning/>
- [16]. <https://monkeylearn.com/sentiment-analysis>
- [17]. [https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664#:~:text=VADER%20\(%20Valence%20Aware%20Dictionary%20of%20rIntensity%20\(strength\)%20of%20emotion.&text=The%20sentiment%20score%20of%20each%20word%20in%20the%20text.](https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664#:~:text=VADER%20(%20Valence%20Aware%20Dictionary%20of%20rIntensity%20(strength)%20of%20emotion.&text=The%20sentiment%20score%20of%20each%20word%20in%20the%20text.)
- [18]. <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- [19]. "Topic Modelling with LDA" <https://www.analyticsvidhya.com/blog/2021/07/topic-modelling-with-lda-a-hands-on-introduction/and> Second Author. 2002. International Journal of Scientific Research in Science, Engineering and Technology. (Nov 2002)

**Cite this article as :**

Liyansi Patel, Vimal Rathod, "A Survey on Sentiment Analysis and Topic Modeling", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 149-154, March-April 2022. Available at doi : <https://doi.org/10.32628/IJSRSET229221>  
Journal URL : <https://ijsrset.com/IJSRSET229221>