**Twitter Sentiment Analysis**

**Archit Jain1, Vaibhavi Kamalapurkar 2, Manali Pundkar 3, Aaryan Suthar 4, Devyani Bonde 5** U.G. Student, Department of Computer Engineering, Marathwada Mitra Mandal Institute of Technology PuneMaharashtra, India1

U.G. Student, Department of Computer Engineering, Marathwada Mitra Mandal Institute of Technology Pune,Maharashtra, India2

U.G. Student, Department of Computer Engineering, Marathwada Mitra Mandal Institute of Technology Pune,Maharashtra, India3

U.G. Student, Department of Computer Engineering, Marathwada Mitra Mandal Institute of Technology Pune,Maharashtra, India4

Associate Professor, Department of Computer Engineering Marathwada Mitra Mandal Institute of Technology Pune,Maharashtra, India5

***Abstract*—The world is changing quickly with the innovations happening these days. The Internet is now a must-have entity for everybody with the Web being utilized in every field. With the rapid increase in social networking applications, people started using these platforms to voice up their opinions with regard to various problems. Analyzing peoples’ reactions toward buying a product, public services, and many other relevant things have now become important. Sentiment analysis (or opinion mining) is a common dialogue preparation task that aims to extract the sentiments behind the opinions in texts and varying other forms. In recent years, researchers in the field of sentiment analysis have been working on analyzing opinions on new and different topics such as web series, commercial products, political views and many other daily societal issues. Twitter is an enormously popular microblogging site on which users voice their opinions and suggestions. Opinion evaluation of Twitter data is a field that has gained much attention over the last decade and involves dissecting “tweets” (comments) and the content of their expressions. As such, this paper explores the sentiment analysis applied to Twitter data and their outcomes.**

***Keywords—; Sentiment analysis; Twitter data; opinion mining; Classification; Naïve’s Bayesian algorithm;***

**I. INTRODUCTION**

Sentiment analysis is also known as “opinion mining” or “emotion Artificial Intelligence” and calls to the utilization of natural language processing , text mining, computational linguistics, and bio measurements to methodically recognize, extract, evaluate, and examine emotional states and subjective information. Sentiment analysisis generally concerned with the voice in client materials; for eg surveys and reviews on the Web and web-based social networks.

Sentiment analysis attempts to determine the disposition of a speaker, essayist, or other subjects in terms of theme via extreme emotional or passionate responses to an archive, communication, or occasion. The disposition might be a judgment or assessment, full of emotion (in other words, the passionate condition of the creator or speaker) or an expectation of enthusiastic responses (in other words, the impact intended by the creator or buyer). Vast numbers of client surveys or recommendations on all topics are available on the Web these days and audits may contain surveys on itemssuch as on clients or fault-findings of films, and so on. Surveysare expanding rapidly, on the basis that individuals like to provide their views on the Web. Large quantities of surveys areaccessible for

solitary items which make it problematic for clients as they must peruse each one in order to make a choice.Subsequently, mining this information, distinguishing client assessments and organizing them is a vital undertaking. Sentiment mining is a task that takes advantage of NLP and information extraction (IE) approaches to analyze an extensivenumber of archives in order to gather the sentiments of comments posed by different authors [1, 2]. This process incorporates various strategies, including computational etymology and information retrieval [2]. The basic idea ofsentiment investigation is to detect the polarity of text documents or short sentences and classify them on this premise. Sentiment polarity is categorized as “positive”, “negative” or “impartial” (neutral). It is important to highlight the fact that sentiment mining can be performed on three levelsas follows [3]:

∙ Document-level sentiment classification: At this level, adocument can be classified entirely as “positive”, “negative”, or “neutral”.

∙ Sentence-level sentiment classification: At this level, each sentence is classified as “positive”, “negative” or unbiased.

∙ Aspect and feature level sentiment classification: At this level, sentences/documents can be categorized as “positive”, “negative” or “non-partisan” in light of certain aspects of sentences/archives and commonlyknown as “perspective-level assessment grouping”.

The main purpose of this paper is to study the existing sentiment analysis methods of Twitter data and provide theoretical comparisons .Thepaper is organized as follows: the first two subsequent sections comment on the definitions, motivations, and classification techniques used in sentiment analysis. A number of document- level sentiment analysis approaches and sentence-level sentiment analysis approaches are also expressed. Various sentiment-analysis approaches used for Twitter are described .Finally, discussions and comparisons of the latter are highlighted.

**II. DEFINITION AND MOTIVATION**

Sentiment analysis is a strategy for checking assessments of people or groups; for example, a portion of a brand’s followers or an individual customer in correspondence with a customer supports representative.

Sentiment analysis is a means of assessing written or spoken languages to decide whether articulation is positive, negative or neutral and to what degree. The current analysis tools in the market are able to deal with tremendous volumes of customer criticism reliably and precisely. In conjunction with contents investigation, sentiment analysis discovers customers’opinions on various topics, including the purchase of items, provision of services, or presentation of promotions.

Twitter is a micro- blogging website where clients generate 'tweets' that are communicated to their devotees or to another client. At 2016, Twitter has more than 313 million dynamic clients inside a given month, including 100 million clients daily [7]. Client origins are widespread, with 77% situated outside of the US, producing more than 500 million tweets every day [8]. The Twitter site positioned twelfth universally for activity in 2017 [9] and reacted to more than 15billion API calls every day [10]. Twitter content likewise shows

up in more than one million outsider sites [8]. In accordance with this enormous development, Twitter has of late been the subject of much scrutiny, as Tweets frequently express client's sentiment on controversial issues. In the social media context, sentiment analysis and mining opinions are highly challenging tasks, and this is due to the enormous information generated by humans and machines [11].

**II. IMPORTANCE AND BACKGROUND**

In reality, organizations and associations dependablyneed to discover users’ popular sentiments about their items and services. Clients use different types of online platforms forsocial engagement including web-based social networking sites; for example, Facebook and Twitter. This kind of connection offers a remarkable open door for advertising knowledge. Individuals of everynationality, sexual orientation, race and class utilize the web to share encounters and impressions about virtually every featureof their lives. Other than composing messages, blogging or leaving remarks on corporate sites, a great many individuals utilize informal organization destinations to log opinions, express

feelings and uncover insights about their everyday lives. Individuals compose correspondence on nearly anything, including films, brands, or social exercises. These logs circulate throughout online groups and are virtual gatherings where shoppers illuminate and impact others. . To the advertiser,these logs provide profound snippets of insight into purchasers’ behavioral inclinations and present a continuous opportunity to find out about client emotions and recognitions, as they happen without interruption or incitement. Be that as it may, recent explosions in client-produced content on social sites are introducing unique difficulties in capturing, examining andtranslating printed content since information is scattered, confused, and divided [12].

Opinion investigation is a method of information mining that can overcome these difficulties by methodically separating and dissecting web-based information without causing delays. With conclusion examination, advertisers are able to discover shoppers’ emotions and states of mind continuously, in spite ofthe difficulties of information structure and volume. The enthusiasm in this study for utilizing sentiment analysis as an instrument for promoting research instrument is twofold.

Sentiment analysis critically encourages organizations to determine customers’ likes and dislikes about products and company image. In addition, it plays a vital role in analyzing data of industries and organizations to aid them in making business decisions.

**II. CLASSIFICATION TECHNIQUES**

In the machine learning field, classification methods have been developed, which use different strategies to classify unlabeled data. Classifiers could possibly require training data. Examples of machine learning classifiers are Naive Bayes, Maximum Entropy and Support Vector Machine [8] [9,10].These are categorized as supervised-machine learning methodsas these require training data. It is important to mention that training a classifier effectively will make future predictions easier.

***A. Naïve Bayes***

This is a classification method that relies on Bayes' Theorem with strong (naive) independence assumptions between the features. A Naive Bayes classifier expects that thecloseness of a specific feature (element) in a class is disconnected to the closeness of some other elements. For instance, an organic fruit might be considered to be an apple ifits color is red, its shape is round and it measures approximately three inches in

breadth. Regardless of whether these features are dependent upon one another or upon the presence of other features, a Naïve Bayes classifier would consider these properties independent due to the likelihood thatthis natural fruit is an apple. Alongside effortlessness, the Naive Bayes is known to out-perform even exceedingly modern order strategies.

The Naive Bayes is widely used in the task of classifying texts into multiple classes and was recently utilized for sentiment analysis classification.

***B. Maximum Entropy***

The Maximum Entropy (MaxEnt) classifier estimates the conditional distribution of a class marked a given a record b utilizing a type of exponential family with one weight for everyconstraint.

***C. Support Vector Machine***

The support vector machine (SVM) is known to perform well in sentiment analysis [13]. SVM investigates information, characterizes choice limits and uses the components for the calculation, which are performed in the input space [8]. The vital information is presented in two arrangements of vectors, each of size m. At this point, each datum (expressed as a vector) is ordered into a class. Next, the machine identifies theboundary between the two classes that is far from any place in the training samples [19]. The separate characterizes the classification edge, expanding the edge lessens ambivalent choices. As demonstrated in [20], the SVM has been proven toperform more effectively than the Naïve Bayes classifier in various text

classification problems.

When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support-vector clustering algorithm, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data.

**II. SENTENCE LEVEL SENTIMENT ANALYSIS APPROCHES**

This analysis focuses on classifying sentences into categories according to whether these sentences are positive, negative, or neutral. Twitter sentiment analysis is considered anexample of sentence-level sentiment analysis. The next sectionexplores Twitter sentiment analysis approaches. Machine learning approaches utilize classification methods to classify text into various categories. There are mainly two types of machine learning strategies: supervised learning and ensemble.

There are four basic Twitter sentiment analysis approachesincluding supervised machine learning based, ensemble methods, lexicon-based, and hybrid. These four approaches are described are as follows:

*A. Twitter Sentiment Analysis using Supervised MachineLearning Approaches*

It depends on labelled datasets that are given to machine learning models during the training process. These marked datasets are utilized to train these models to obtained significant outputs. In machine learning systems, two datasets are required: training set and test set. Machine learning approaches such as classifiers can be utilized to detect the sentiment of Twitter. The performance of Twitter sentiment classifiers is principally relying upon the number of training data and the features sets are extractors. Twitter sentiment analysis strategies that rely on machine-learning methods are more popular, especially SVM and NB classifiers. Fig. 2 illustrates the procedure of supervised machine learning approaches for Twitter sentiment analysis.

The Twitter sentiment analysis process consists of three steps. First, the classifier is trained using datasets comprising positive, negative, and unbiased tweets. Examples of tweets are shown below:

∙ Positive tweets:

∙ Negative tweets:

∙ Unbaised tweets:

Tweets can containvaluable information expressing opinions on any topic.However, they may also include specific characters that are not helpful in detecting sentiment polarity; hence, it makes sense to preprocess tweets. This second step consists of converting all tweet texts to lower case. In addition, tweets should be cleaned by removing URLs, hashtag characters (such as #Trump) or user mentions (such as @Trump) as Twitter sentiment-analysismethods are not concerned with these characters. The preprocessing step includes filtering out stop words that are considered unusual discriminant features [11].After preprocessing, predictions are performed. In this phase, various prediction algorithms, such as the SVM, Bayesian classifier, and Entropy Classifier, can be used to decide the sentiment polarity of tweets.

**II. CONCLUSION**

In this paper, various techniques for Twitter sentiment analysis methods were discussed, including machine learning, ensemble approaches and dictionary based approaches. In addition, hybrid and ensemble Twitter sentiment analysis techniques were explored. Research outcomes demonstrated that machine learning techniques; for example, the SVM and MNB produced the greatest precision, especially when

multiple features were included. Machine learning algorithms, such as The Naive Bayes, Maximum Entropy, and SVM, achieved an accuracy of approximately 80% when n-gram and bigram model were utilized. Ensemble and hybrid-based Twitter sentiment analysis algorithms tended to perform better than supervised machine learning techniques, as they were able to achieve a classification accuracy of approximately 85%.

In general, it was expected that ensemble Twitter sentiment-analysis methods would perform better than supervised machine learning algorithms, as they combined multiple classifiers and occasionally various features models. However, hybrid methods also performed well and obtained reasonable classification accuracy scores, since they were able to take advantage of both machine learning classifiers and lexicon

based Twitter sentiment-analysis approaches.

**II. REFERENCES**

[1] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," ACM Comput. Surv., vol. 49, no. 2, pp. 1-41, 2016.

[2] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Business horizons, vol. 53, no. 1, pp. 59-68, 2010.

[3] A. Abirami and V. Gayathri, "A survey on sentiment analysis methods and approach," in Advanced Computing (ICoAC), 2016 Eighth International Conference on, 2017: IEEE, pp. 72-76. [4] K. P. Murphy, "Naive bayes classifiers," University of British Columbia,vol. 18, 2006. [5] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," Comput. Linguist., vol. 22,no. 1, pp. 39-71, 1996.

[6] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support vector machine," Teori dan Aplikasinya dalam Bioinformatika, Ilmu Komputer. com, Indonesia, 2003.

[7] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: A survey of techniques," arXiv preprint arXiv:1601.06971, 2016.

[8] A. Harb, M. Plantié, G. Dray, M. Roche, F. Trousset, and P. Poncelet, "Web Opinion Mining: How to extract opinions from blogs?," in Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, 2008: ACM, pp. 211-217.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002: Association for Computational Linguistics, pp. 79-86.

[10] J. Khairnar and M. Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," International Journal of Scientific and Research Publications, vol. 3, no. 6, pp. 1-6, 2013.

[11] C. Wu, L. Shen, and X. Wang, "A new method of using contextual information to infer the semantic orientations of context dependent opinions," in Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on, 2009, vol. 4: IEEE, pp. 274-278.

[12] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon- based methods for sentiment analysis," Computational linguistics, vol. 37, no. 2, pp. 267-307, 2011.