



SENTIMENT ANALYSIS AND OPINION MINING ON PRODUCTS, MOVIES, TOURISM REVIEWS USING MACHINE LEARNING TECHNIQUES

Qureshi Mahek Fatima

3rd Year Btech (CSE)

P.E.S College of Engineering, Aurangabad,
Maharashtra, India

Sayyed Mustakim

3rd Year Btech (CSE)

P.E.S College of Engineering, Aurangabad,
Maharashtra, India

Abstract— Sentiment Analysis and Opinion Mining are one of the fast - emerging fields. Today our world is technologically advanced. There is extensive use of social media among the people. People express their opinions about their concerned domain on social networking sites or online websites. Sentiment Analysis helps to analyze people's reviews and opinions. This helps businesses, organizations, hotels, institutions, tourism industry, to know people's needs and try to modify their products and services according to it to yield the profits. Sentiment Analysis and Opinion Mining is performed in various domains like product reviews, movie reviews, tourism, education, politics, sports, etc. Some of the domains and the techniques used to analyze the Sentiment Analysis and Opinion Mining are discussed in this paper.

Keywords— Sentiment Analysis, Machine Learning, Naïve Bayes, SVM (Support Vector Machine).

I. INTRODUCTION

We live in a technologically advanced world. With the development of social media applications and its enormous use by the people, the world has been converted into an online world. Sentiment analysis and Opinion Mining helps to classify peoples opinions and reviews (e.g. a positive or negative opinion) within the text, whether a whole document, paragraph, sentence, or clause[1]. Today it is essential to know about the feedback, reviews, emotions of the users for businesses, movie makers, organizations, schools, colleges, tourist places, etc. This helps the concerned authorities to know what are the users need, what they expect from their products or services. Due to the advancement in social media people are expressing their opinions on various things quite openly. For people who want to buy some products or invest in some services from online websites, it is a cherry on the cake to know about the opinions, suggestions or, ratings of the people who have pre-owned it. From these reviews, they can

understand and decide whether it will be a bonus to buy that product or invest in that service or not. Likewise for the travel enthusiast who loves to explore new places, can go through online reviews on that particular place, which will help them to know whether that place would be a worth visit or not. Sentiment Analysis is an equally beneficial technique for the hotel industry. As today we can book hotels online. Before visiting some unknown places now people can go through online reviews of the people to know which hotel would be best to reside in. For the people making some unique cannot and posting it on social media, reviews analysis helps a lot to learn people's expectations from them and tailor their content according to their expectations or opinions. The School or college authorities too have a website where there is an option of feedback. Students, Faculty or, the Parents can write their opinions or suggestions for any particular thing they feel about. This helps the Institution authorities to work towards the suggestions in the feedback and improve the quality of the institute. Politics is another domain that people are interested in. As we all live in a democracy. Time to time there are certain elections. There are certain pools on social networks, people write their opinions and all these reviews and opinions help the people to guess who can win the upcoming elections. Sentiment Analysis or Opinion Mining is an emerging field. There are various steps through which SA and OP are performed to classify the reviews. Firstly data is extracted from a concerned site, mostly Amazon, Flipkart, IMBD, Twitter, etc. [1][2][3][4][5][6][7][9][16][17]. Later the data is pre-processed using various techniques for removal of stop words, punctuation marks, etc. using POS Tagger, etc. [2][4][5][6][9]. The next step is the feature extraction of data using various feature extraction techniques like Unigram and Bigram[1][7]. In some researches, there is a calculation of polarity score (positive or negative) using the available analyzing tools like SentiWordNet[6][9]. Lastly, the classifiers in Machine Learning are used to classify the reviews in to negative or positive and calculate the accuracy and precision of the model. The most commonly used classifiers are Naïve Bayes and



Support Vector Machine (SVM)[1][2][3][4][5][6][7][8][14][17]. In Section 2 there is a discussion on Sentiment Analysis on Product Reviews. In Section 3 there is a discussion on Sentiment Analysis On Movie Reviews. Section 4 discusses Sentiment Analysis In Tourism and Section 5 discusses the Conclusion.

II. SENTIMENT ANALYSIS ON PRODUCT REVIEWS

In case Today, as technology has developed enormously, with the development of social media and its extensive popularity among the people, the world has transformed into an online world. From small businesses to big companies everyone has directed their focus in the marketing and development of their businesses online. There are various e-commerce websites like Amazon, Flipkart, Myntra, etc., which sells a variety of products. The business owners feel the need to know people's opinions and feedback regarding their products, services, brands, etc. This helps them to modify their products, services, and brands according to the user's needs. Likewise, the users/people before buying a product or investing in some company's services needs to know the reviews of the people who have pre-owned it. This helps them to make an appropriate decision whether to buy or invest in it or not. The reviews on online sites help people to know other people's opinions. Sentiment Analysis and Opinion Mining is a technique to mine people's opinions or reviews from a text and classify them into positive, negative, or neutral[10]. For example., if someone wants to buy a certain mobile can be from Amazon. Firstly that person will see the reviews and ratings on that product whether the people who have bought it found it's performance satisfactory or not. This will help people to invest their money in reliable things. In Sentiment Analysis or Opinion Mining first step is to extract or mine the data from the concerned online site. Mostly data from Amazon[1][2][3][4], Flipkart[1][5], Epinions, Cnet[4] have been used to perform the Sentiment evaluation. The data extracted can be labeled or an un-labeled format. In the Second step pre-processing of data is done using various available techniques, mainly POS tagger is used to extract nouns, adjectives, adverbs, and the common linguistic characteristics from the reviews[2][4][5][6][9]. There is the removal of stop words, emojis, punctuation marks which don't have a certain meaning. After pre-processing by using available analyzing tools like SentiWordNet, etc. there is a generation of sentiment score to the dataset[6][9]., For the extraction of features, some researchers used Unigram and Bigram feature extraction methods[1][7]. After all these steps the Machine Learning approaches are used to classify the reviews into positive or negative. Most commonly used ML classifiers are Naïve Bayes [1][3][4][5], SVM(Support Vector Machine)[1][2][3], Maximum Entropy[1], etc. Refer Table 1

III. SENTIMENT ANALYSIS ON MOVIE REVIEWS

This is an era of technological advancement. Everyday someone in any corner of the world invents something. If we take ourselves back a decade, we can understand how far we have come in the technological domain. One such evolution has happened in the entertainment industry. Before the technology took over us, it was a usual practice to watch a recent movie in a cinema hall or whenever it used to telecast on television. Dvd players were also available to watch a movie by inserting the desired cd. But the availability of the cd uses to take quite a lot of time. Now, this is not the scenario. A wide range of movie applications and websites have been developed that contain thousands of movies in it. For example Netflix, Amazon Prime, etc. The people can now just download that application, pay a certain subscription charge, and watch as many movies or whichever movie they want anytime, anywhere. As everyone loves entertainment and is a bit choosy of the genre of the movie they like, people want to invest their valuable time watching something useful. So, the user's opinions and reviews play an important role. By going through the reviews on a certain movie, people can get an idea if it's a great choice to watch it or not. For the moviemakers to reviews are necessary, which will help them to understand are the people loving their content or not and what are their expectations. Here Sentiment Analysis or Opinion Mining plays an important role. Sentiment Analysis or Opinion mining extracts or mines people's opinions and reviews and classify them into positive, negative, or neutral [10]. Firstly step for sentiment analysis is the extraction of data, mostly for movie reviews IMBD [6], Bo Pang[8], and Amazon[7] website reviews are used. The second step is preprocessing of data, using POS Tagger[2],[4][5][6][9] and there is stop word removal. Common feature extraction techniques used are Unigram and Bigram[1][7]. The next thing done is generating a sentiment score using available analyzing tools like SentiWordNet, WordNet, Opinion Lexicon[6][9]. Now the classifiers are used to classify positive or negative, mostly Machine Learning approaches are used like Naïve Bayes, SVM, Genetic Algorithm, KNN, Decision tree, etc [1][2][3][4][5][6][7][8]. Refer Table 2

IV. SENTIMENT ANALYSIS ON TOURISM

In the modern era of technology and development, the biggest platforms for the interaction of people to express their feelings, Sentiments, and their Opinions are Social Media, Blogs, Websites, etc. For the tourism sector, it has been a massive change from being offline to online on the internet so they can easily analyze the customer's feedback concerning their facilities, services, and management. Different tourism companies and hotels always try to analyze the feedback of their clients. Analyzing the reviews and opinions of consumers help the merchandise to modify the services, management, staff selection, etc. Similarly, for the user, he also tries to figure out where we can plan a trip for his holidays, stay in hotels, visiting some historical places, etc. Based on reviews opinions and comments from the previous consumers of that particular merchandise. The user will get an idea of whether he wants to



go with it or not. Sentiment analysis or opinion mining is the area deals with judgments, responses as well as feelings which are generated from text [11]. If we take an example of a person who wants to visit the historic place of Egypt pyramids for his holidays so he can go through different tourism websites like booking.com, airbnb.com tripadvisor.com, etc. Where he can gain some information about that place and also reviews of people who visited that place before. Then he would be able to decide whether he wants to go there or not. In this process of analysis different data sets have been used for analyzing the sentiments and opinions of people. Mostly social media blogs, tourism websites have been used for the extraction process of the data. Firstly the data collection has been done using different automated programs [12]. Here comes the second step which includes processing of data which results in spelling corrections, normalization, filtering the text, case folding, etc [12]. In the third step, it aims to obtain the classification of data using different techniques like text blog python library, text mining approach, Fuzzy c-means clustering, etc. By which the data can be categorized into positive, negative, and neutral stages [12] [13] [14]. The last step comes up with the classification model which has been made using techniques like K-nearest Neighbor, Support Vector Machine, Logistic Regression, Random Forest Classifier, Naïve Bayes, etc [14]. Refer Table 3

Table -1 Comparative Analysis On Product Reviews

Sr.no	Paper Title	Algorithm Used	Datasets	Outcomes
1.	Bordoloi et al(2018)[10]	Naïve Bayes, SVM, Maximum Entropy	Amazon, Flipkart reviews	The ML approaches of classification are used and in all the 3 used datasets NB outshines SVM and MaxEnt with the 81.33% accuracy in one dataset, 80% in the 2 nd dataset, and 78.24% in the 3 rd dataset.
2.	Jabbar et al (2019)[2]	SVM	Amazon	After the pre-processing of data, the machine learning method of the SVM classifier is used to classify the reviews into

				Positive, negative, and neutral. By using this proposed method with SVM the precision recorded was 87.88%.
3.	Jagdale et al (2017), [3]	SVM , Naïve Bayes	Amazon	For the camera reviews, the Naïve Bayes gets an accuracy of 98.17% and SVM gets 93.54%.
4.	Jeyapriya et al (2015)[4]	Naïve Bayes	Amazon, Epinions, Cnet	By using the Naïve Bayes algorithm the reviews are classified into positive and negative ones. Aspect extraction has given the accuracy of 80.36% using frequent itemset mining.
5.	Venkata et al (2015)[5]	Naïve Bayes	Flipkart	After Pre-processing Product score was generated by examining the rating, reviews, and all the aspects. Naïve Bayes classifier was used to classify the reviews. There has been given a comparison between two mobile phones by displaying star rating, polarity rating,



				reviews per month, and the grand score to help the user choose a better mobile
--	--	--	--	--

Table -2 Comparative Analysis On Movie Reviews

Sr. No	Papers	Classifiers Used	Dataset used	End Result
1.	Sahu et al (2016)[6]	Decision Tree, Bagging , Naïve Bayes, COCR , KNN.	IMBD reviews	Random forest gives the best accuracy of 88.95%. The rest classifiers have the accuracy as follows, Decision Tree 87.53%, COCR 82.53%, Bagging 88.57 %, KNN 88.56%, and Naïve Bayes 54.77%. By comparing this proposed model with the previous models. This gives the highest accuracy of 88.95%.
2.	Mr. B. Narendra et al (2016) [7]	Naïve Bayes, Proposed model- Apache Hadoop	Amazon site	Naïve Bayes classifier with stop word removal and bigram feature extraction method and got the accuracy of 81.6%. The proposed method using Apache Hadoop obtained the precision 98%. This proposed method outshined the Naïve Bayes Classifier technique
3.	M.Govindarajan et al (2013)[8]	Naïve Bayes,	Bo Pang	The accuracy obtained by

		Genetic Algorithm, The hybrid method of NB-GA		Naïve Bayes is 91.15%, for the Genetic Algorithm is 91.25% and the Proposed hybrid model of NB-GA is 93.80%.
4.	Sharma P et al (2016)[9]	Used available Analyzing tools like SentiWordNet, WordNet, Opinion Lexicon.	IMBD	The proposed system is compared to three analysis tools WordNet, SentiWordNet, and Opinion Lexicon. Scores are evaluated using all these tools. The accuracy obtained for the WordNet is 57%, SentiWordNet 63%, Opinion Lexicon is 70% and the proposed model is the highest accuracy with 81%.

Table -3 Comparative Analysis On Tourism

Sr.no	Paper Title	Algorithm Used	Datasets	Outcomes
1.	Aitor García, et al (2012)[15]	Linguistic tools based on an annotated lexicon	Lexical Dataset	The obtained accuracy in mere polarity detection was 80%
2.	Sarah Anis et al (2020) [14]	Naïve Bayes, K-nearest neighbor, Support Vector Machine, Logistic regression, Random forest classifier	Reviews from the Kaggle website.	The support vector machine achieved 86.3%, logistic regression achieved 85.9%, Random forest classifier achieved 84.6%, K-nearest neighbor, and Naïve Bayes achieved 83.8% and



				7.8% respectively.
3.	Puteri et al (2017)[12]	Text mining program called Parmeswari v1.4.0	Tripadvisor.com website	The classifier model used in this study had an average accuracy of 85%.
4.	Vallikannu et al (2019)[16]	Domain-specific on combined with lexicon-based approach conceptual and semantic analysis incorporating with the machine learning method Naïve Bayes	Twitter Tweets	Using the conceptual semantic analysis method performance of the sentiment analysis improved with precision 85.54%.
5.	Taweesak Kuhamaneet al (2017)[17]	Decision Tree, Support Vector Machine, Artificial Neural Network, and Naïve Bayes.	Twitter Tweets	Decision Tree achieved 79.83 % accuracy, SVM achieved 80.11% accuracy, ANN achieved 80.33 % accuracy and Naïve Bayes achieved 55.66% accuracy.

V. CONCLUSION

Sentiment analysis and opinion mining use a proper dataset to research the sentiment of individuals . Different algorithms are made but still, there's a niche between better and excellent to beat all problems. From the above data, we will conclude that Support Vector Machine and Naïve Bayes are the foremost frequently used algorithms for sentiment analysis and opinion mining. Although, many organizations are working for creating an ideal algorithm which will solve all the issues and difficulties. Still, it requires many more effort to form one. there's an enormous demand for this technique in our field of labor .

VI. REFERENCE

- [1] <https://monkeylearn.com/sentiment-analysis/>.
- [2] Jabbar. J , Urooj. I , JuSheng Wu , Azeem. N,(2019),"Real- time Sentiment Analysis on E-Commerce Application", (DOI- 10.1109/ICNSC.2019.8743331).
- [3] Jagdale. R , Shirsat . Vishal , Deshmukh Sachin,(2017),"Sentiment Analysis on Product Reviews Using Machine Learning Techniques", (10.1007/978-981-13-0617-4_61/AISC, Volume 768).
- [4] Jeyapriya . A, Selvi Kanimozhi . C.S , (2015), "Extracting aspects and Mining Opinions in product Reviews using Supervised Learning Algorithm", (10.1109/ECS.2015.7124967).
- [5] Venkata Rajeev, Smrithi Rekha ,(2015),"Recommending Product to Customers Using Opinion Mining Of Online Product Reviews and Features", (10.1109/ICCPCT.2015.7159433).
- [6] Sahu Tirath P, Ahuja Sanjeev , (2016), "Sentiment Analysis Of Movie Reviews : A Study On Feature Selection & Classification Algorithms , (10.1109/MicroCom.2016.7522583).
- [7] Mr. B Narendra , Mr. K Uday Sai, Mr.G Rajesh , Mr.K. hemanth , Mr M.V. Chaitanya Teja , Mr K. Deva Kumar ,(2016),"Sentiment Analysis On Movie Reviews: A Comparative Study Of Machine Learning Algorithms and Open Source Technologies, (10.5815/ijisa.2016.08.08).
- [8] M. Govindarajan , (2013), Sentiment Analysis of Movie Reviews using Hybrid Method of Naïve Bayes and Genetic Algorithm',(IJACR, Vol 3 issue 13, p139).
- [9] Sharma Pallavi , Mishra . N , (2016),"Feature Level Sentiment Analysis On Movie Reviews", (10.1109/NGCT.2016.78.77432).
- [10] Bordoloi Monali , Biswas Saroj Kr,(2018), Sentiment Analysis Of Product Using Machine Learning Techniques: A Comparison among NB , SVM and MaxEnt ",(118(18): 71-83).
- [11] www.google.com
- [12] Puteri Prameswari , Isti Surjandari , Enrico Loah, (2017),"Opinion Mining from Online Reviews in Bali Tourist Area", (10.1109/ICSTech.2017.8257115).
- [13] Gupta Gunjan , Parul Gupta , (2019), Twitter Mining for Sentiment Analysis in Tourism Industry", (101.1109/WorldS4.2019.8903940).
- [14] Anis Sarah , Saad Sally , Aref Mostafa , (2020),"Sentiment Analysis of hotel Reviews Using Machine Learning Techniques", (10.1007/978-3-030-58669-0_21).
- [15] Aitor Garcia, Sean Gaines , Maria Teresa Linaza,(2012),"A Lexicon Based Sentiment Analysis Retrieval System for Tourism Domain", (e-Review of Tourism Research(eRTR)Vol 10, No 2).



- [16] Vallikanu Ramanathan, T. Meyyappan ,(2019), “Twitter Text Mining for Sentiment Analysis on People’s Feedback about Oman Tourism”,(10.1109/ICBDSC.2019.8645596).
- [17] Taweesak Kuhamanee, Nattaphon Talmongkol, Krit Chaisuriyakul, Wimol San-Um, Noppadon Pongpisuttinun, Surapong Pongyupinpanich, (2017)“Sentiment Analysis of Tourists to Bangkok using Data Mining through Online Social Network”, IEEE 15th International Conference on Industrial Informatics (INDIN).