Sentiment Analysis on Facebook comment using Machine Learning

Nivedita Mishra(1812010069)
Institute of Technology and Management Gida, Gorakhpur

Priya Srivastava (1812010053)
Institute of Technology and Management Gida, Gorakhpur

Shikha Gupta(1812010102)
Institute of Technology and Management Gida, Gorakhpur

Saumya Srivastava (1812010094)
Institute of Technology and Management Gida, Gorakhpur

Mr. Mahendra Sonker
(Assistant Professor, CSE Department)

ABSTRACT

Sentiment Analysis stands for outstanding studies subject matter in call for beneathneath the Natural language processing (NLP). The essential goal of this studies subject matter is to identify out the feelings and opinion of the consumer or clients via textual content basis. Even via numerous models. Sentiment evaluation nevertheless taken into consideration as hard hassle with such a lot of struggle to solved. Some of the demanding situations are because of the slag words, spelling mistakes, grammatical and new accents etc. The paper plans to make a overview the usage of one-of-a-kind gadget getting to know algorithms with numerous data. The modern literature overview temps to survey almost 20 contributions, which covers one-of-a-kind forms of programs getting used for sentimental evaluation. At first, the evaluation specializes in illustrating the contributions of every paintings and observes the sort of gadget getting to know algorithms used .The evaluation additionally concentrates at the become aware of the non-saturated utility for which the sentimental evaluation in wanted maximum in upcoming studies.

Keywords: *Sentiment Analysis; Application-oriented Analysis; Machine Learning Algorithms; performance Analysis; Research Gaps and Challenges*

1. INTRODUCTION

There is a active development inner side the micro strolling a blog net webweb webweb sites similarly to social networks. One of the critical internet locations to the customers is micro strolling a blog net webweb webweb sites, which can be useful for expressing the person's attitudes, critiques, and mind concerning numerous contexts [1] [2]. The maximum used social networking offerings and the micro strolling a blog platform is twitter, which gives greater facts. At gift, for the sentiment evaluation of the person's critiques at the product, occasion, or context, researchers employ social facts. Moreover, the opportunity call for sentiment evaluation is opinion mining, that is the exquisite NLP project. This sentiment evaluation defines orientation of sentiments associated with textual content as every neutral, excellent, or lousy [3] [4]. Moreover, sentiment evaluation represents the textual content analytics, computational linguistics, and NLP implementations for spotting and categorizing the critiques of the person. In general, the principle intention of the sentiment evaluation is to outline the author's trouble of view regarding the same context or the complete file's contextual polarity. The view may be every a person's

judgment or evaluation, affective america of the us or the deliberated verbal exchange of emotion. In general, the form of textual content expressions in supply substances into facts and critiques is executed via the sentiment evaluation. Facts are the aim expressions concerning the activities and their attributes similarly to entities. The critiques are the subjective expressions of sentiments, feelings, feelings, activities and attributes, and attitudes. This want to be specific that no longer all of the aim sentences embody no critiques and no longer all subjective sentences embody critiques. Thus, for sentiment evaluation, it is exquisite for spotting and extracting the facts and critiques from supply substances. However, this appears to be pretty complicated for reaching precisely. In modern times, important techniques are associated with tool getting to know, rule-primarily based totally definitely absolutely, and the mixture of each strategies. Machine getting to know fashions encompass traditional techniques like deep getting to know and conditional random situation techniques, while the rule-primarily based totally definitely absolutely fashions encompass lexicon-primarily based totally definitely absolutely method. Object detection [5] [6], community optimization [8], picturegraph popularity [7], device security [12],sensor networks[9] [10] [11], and transportation [13] are primarily based totally definitely totally on deep getting to know strategies, which can be withinside the maximum critical carried out in excellent fields. Several researchers have mixed deep getting to know similarly to tool getting to know algorithms into textual content sentiment evaluation via sentiment lexicon components and incredible results are received [14]. The maximum critical aim of the sentiment lexicon-primarily based totally definitely absolutely version is to broaden a sentiment lexicon, that is executed via selecting appropriate lousy phrases, sentimental phrases, and diploma adverbs. For the built sentiment lexicon, sentimental polarity and depth are marked. Once the textual content is given as enter, the phrases are matched with the sentiment phrases gift withinside the sentiment lexicon, and people phrases are weighted and delivered for obtaining the enter textual content's sentiment value, consequently the power of thoughts of soppy polarity is executed as constant with the sentiment value. However, there are few techniques for obtaining the capabilities of phrase vector associated with the textual content like Glove, Word2Vec, and FastText automatically. However, the traditional tool getting to know fashions notwithstanding the reality that require the emotional function extraction of the installed statistics from the enter textual content via textual content vectorization, human intervention, and later that algorithms are carried out for categorizing the sentiment of the textual content capabilities [15]. The maximum critical contributions of this paper are portrayed as follows. To go through a critical assessment of sentiment evaluation beneathneath excellent programs. To perform the excellent assessment of numerous sentiment evaluation fashions primarily based totally definitely absolutely at the tool getting to know algorithms, sorts of facts, equipment, and excellent widely wide-spread general overall performance measures. To formulate the treasured studies gaps and annoying situations primarily based totally definitely totally on the winning contributions beneathneath sentiment evaluation. The assessment on sentiment evaluation type is designed withinside the following way: Section II specifies the literature assessment on traditional sentimental evaluation in social media. Section III describes numerous tool getting to know algorithms for sentiment evaluation along side widely wide-spread general overall performance measures. The evaluation on excellent sorts of facts used and equipment for sentiment evaluation is given in Section IV. The studies gaps and annoying situations of sentiment evaluation the usage of tool getting to know algorithms are demonstrated in Section V. Section VI specifies the belief of the complete paper.

This paper describes a Sentiment Analysis test executed on over than one thousand Facebook posts approximately newscasts, evaluating the sentiment for Rai - the Italian public broadcasting service - withinside the route of the developing and additional dynamic personal company La7. This test maps test results with observations made thru way of method of the Osservatorio di Pavia, it surely is an Italian institute of studies specialised in media evaluation at theoretical and empirical level, engaged withinside the evaluation of political conversation withinside the mass media. This test takes furthermore in account the facts supplied thru way of method of Auditel concerning newscast audience, correlating the evaluation of Social Media, of Facebook in particular, with measurable facts. The posts had been accrued and analyzed thru way of method of the use of a content material cloth fabric allowing tool – iSyn Semantic Center - that offers deep semantic facts get proper of access to and dynamic type capabilities for massive portions of dispensed multimedia facts.

2. RELATED WORK

Authors of [15] officially said their task, and interpreted a way to mathematically comprise social context and topical context into the primary prediction version. They investigated the content-primarily based totally correlations the various topics, and calculated TCS to degree them. The assumptions approximately social context and topical context have been each corroborated through the speculation checking out over the Twitter records set they created. Finally, they performed experiments to assess the proposed ScTcMF framework, and the experimental outcomes verified that each social context and topical context can assist to enhance the overall performance for the person subject matter opinion prediction.

According to [16] in sentiment analysis, the overall performance of Bag of phrases occasionally stays restrained because of a few essential deficiencies in coping with the polarity shift trouble.So,to deal with this trouble for sentiment category they proposed a version referred to as twin sentiment analysis (DSA). They first proposed a singular records growth approach through growing a sentiment-reversed evaluation for every schooling and check evaluation. On this basis, they proposed a twin schooling set of rules to utilize authentic and reversed schooling critiques in pairs for mastering a sentiment classifier, and a twin prediction set of rules to categorize the check critiques through thinking about aspects of 1 evaluation. They additionally prolonged the DSA framework from polarity (positive-negative) category to 3-class (positive -negative-impartial) category, through taking the impartial emotions into consideration. Finally, they advanced a corpus- approach to assemble a pseudo-antonym dictionary.

Authors of [17] concentrated on engineering students' Twitter posts to understand problems and glitches in their educational experiences. They first conducted a qualitative analysis on samples taken from about 25,000 tweets associated to engineering students' college life. They found engineering students encounter problems such as deficiency of social engagement, heavy study load, and sleep deficiency. Based on these results, we implemented a multi-label classification algorithm to categorize tweets reflecting students' problems

They then used the algorithm to train a detector of student problems from about 35,000 tweets issued at the geo-location of Purdue University. This work, presents an approach and

results that show how casual social media data can provide insights into students' experiences.

For the classification of sentiment applying sentiment classifier trained results in poor performance because words that occur in the train (source) domain might not appear in the test (target) domain so, To overcome the feature mismatch problem in cross-domain sentiment classification [18] proposed a cross-domain sentiment classifier using an automatically extracted sentiment sensitive store. They have done the comparisons against the SentiWordNet, a lexical resource for word polarity. They show that the created sentiment-sensitive store accurately captures words that express similar sentiments [18].

Authors observed that previous research mainly focused on modeling and tracking public sentiment so, they moved one step further to interpret sentiment variations. They worked on twitter dataset. They observed that emerging item (named foreground item) within the sentiment variation periods are highly related to the actual reasons behind the variations. These foreground item can help to interpret the sentiment variations .Based on this observation, they proposed a Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to dig out foreground item and filter out background item. To further improve the readability of the mined reasons, they ranked them with respect to their "fame" within the variation period using Reason Candidate and Background LDA (RCB-LDA) method

[19].

Authors presented a joint sentiment-topic model and a reparameterized version of JST called Reverse-JST when most of the existing approaches to sentiment classification favor supervised be in train. Unlike supervised approaches to sentiment classification which often fail to produce acceptable performance when shifting to new domains, the weakly supervised environment of JST makes it highly convenient to other domains [20].

Authors have explored the predictive electricity of opinions the use of the film area as a case study, and studied the trouble of predicting income overall performance the use of sentiment facts mined from opinions [21]. They suggest Sentiment PLSA (S-PLSA), wherein a overview is taken into consideration as a file generated via way of means of some of hidden sentiment factors, if you want to seize the complicated nature of sentiments. Then they suggest ARSA, an Autoregressive Sentiment-Aware version for income prediction. Then they look for in addition development withinside the accuracy of prediction via way of means of thinking about the first-rate factor, with a focal point on predicting the first-rate of a overview withinside the absence of user-provided indicators, and gift ARSQA, an Autoregressive Sentiment and Quality Aware version, to apply sentiments and first-rate for predicting product income overall performance.

It is feasible for the inventory charge of a few groups to be expected with a mean accuracy as excessive as 76.12%.They proposed a way to mine Twitter information for solutions to the questions like if the charge of a choice of 30 groups indexed in

NASDAQ and the New York Stock Exchange can sincerely be expected via way of means of the given 15 million facts of Twitter message [10]. We have summarized the survey in Table I for the unique kind of paintings carried out withinside the Sentiment Analysis field. R* is the reference number. The purpose in the back of choosing the columns of the desk is to simply to investigate the paintings carried out withinside the sentiment evaluation field.

.

3. PORPOSED MODELLING

3.1 Incremental Approach:

Analysis of actual time statistics isn't always one time operation. Whenever statistics is brought we want to do evaluation then why have to now no longer we use the preceding evaluation end result. Incremental technique lets in an present end result to be up to date the usage of most effective new character statistics instances, while not having to re-method beyond instances. This can be beneficial in conditions wherein the complete dataset isn't always to be had while the statistics adjustments over time.

3.2 Parallel Computing For Massive Data

If we divide the computation into duties or strategies that may be completed simultaneously, then there may be an improvement withinside the velocity via the usage of parallelism, it's miles vital to gain this in sentiment evaluation for huge facts of social media, wherein huge immediate messages are posted each day in order that we are able to make use of the general computing power.

3.3 Credibility/Behavior/Homophile

Behaviors in social media are best discovered with the aid of using the strains they go away in social media. We not often examine the using elements that purpose those behaviors; nor are we able to interview people concerning their behaviors. Even if a conduct is analyzed on social media and associated styles are gleaned, it is hard to confirm the validity of those behavioral styles. Evaluation will become even extra hard for industries wherein crucial choices are to be made primarily based totally on observations of man or woman conduct.

3.4 Grammatically Incorrect Words

There are many tactics that examine sentiments however hardly ever any paintings achieved on grammatical mistakes. The effects of sentiment evaluation may be advanced if those sorts of mistakes may be mapped to accurate words.

3.5 Sarcasm

Sarcasm may be used to harm or offend or may be used for comedian affect.It method fake positives for eg."Children sincerely enliven a household - they by no means flip the lighting off". Detecting sarcasm from the expressions and locating out the ideal context associated sentiments is a tough task. It is an ironic or satirical statement that appears to be praising a person or some thing however is sincerely taunting or cutting.

3.6 Review Author Segmentation

Opinion in the direction of a goal can be specific with the aid of using many folks that may be referred to as as evaluate authors. Depending at the commenting fashion of those authors, they need to be classified in order that credibility assessment could be easy. In selection making this credibility assessment is helpful.

3.7 Refinement of existing Lexicons or Updating/Down- Dating Lexicons

Many humans comments, the Performance of sentiment analyzer depend upon the correctness of the lexicon. Finetuning of present lexicons is needed to deal with new phrases and spoil the phrases which might be no greater used for higher results. Lexicon growth thru the usage of synonyms has a disadvantage of the wording loosing it number one which means after some recapitulation.

3.8 Handling Noise and Dynamism

Social media statistics are enormous, noisy, unstructured, and dynamic in nature, and hence novel demanding situations arise, introduces consultant studies issues of mining social media. Identifying and elimination of noisy statistics is a tough task.

.

4 RESULTS AND DISCUSSIONS

R*	Approach	Tools/Techniques	Experiment	Language Dependency	M/c Learning/ Lexicon Based(ML*/LB*)	Data Scope	Data Source
1	User-Topic opinion prediction (2013)	Social context and Topical context incorporated Matrix Factorization (ScTcMF)	To predict the unknown user-topic opinions.	Yes	LB*	Twitter	Tweets
2	Polarity shift in sentiment classification(2015).	Dual sentiment analysis(DSA)	Polarity classification task	No	LB*	Multi-domain sentiment English dataset. two Chinese dataset	Amazon.com, ChnSentiCorp corpus
3	Qualitative analysis and large-scale data mining techniques(2014)	Naïve-Bayes multi-label classification algorithm	Show how informal social media data can provide insights into students' experiences.	Yes	ML*	Twitter	Tweets
4	Cross-domain sentiment classification(2013)	Corpus based	To evaluate the benefit of using a sentiment sensitive thesaurus for crossdomain sentiment classification	Yes	LB*	Product reviews	Amazon.com
5	To interpret sentiment variations (2014)	Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), generative model called Reason Candidate and Background LDA (RCBLDA	To mine possible reasons of public sentiment Variations.	Yes	ML*	Twitter	Tweets
6	Sentiment and topic detection (2012)	Weakly supervised joint sentiment-topic (JST) model based on	To detect sentiment and topic simultaneously from text	Yes	ML*	Product reviews, Movie reviews	Amazon.com, IMDB movie archive

		latent Dirichlet allocation (LDA, Reverse-JST					
7	Hashtag-level sentiment classification(2011)	SVM classifier	To automatically generate the overall sentiment polarity for a given hashtag in a certain time period, which markedly differs from the conventional sentence-level and document-level sentiment analysis.	Yes	ML*	Self-annotation manner to label the dataset, Twitter	Tweets
8	Sentiment polarity classification and sentiment strength detection (2012)	Hybrid approach(lexicon based +M/c learning)	To classify polarity and detect sentiment strength	Yes	ML* and LB*	Software reviews and movie reviews	CNET, IMDB
9	Sales prediction (2012)	Sentiment PLSA (S-PLSA, ARSQA, an Autoregressive Sentiment and Quality Aware model	To Predict Sales Performance	Yes	ML*	Movie reviews	IMDB
10	Predicting Stock Price Movements (2014)	NLP techniques	To determine if the price of a selection of 30 companies listed in NASDAQ and the New York Stock Exchange can actually be predicted by the given 15 million records of tweets	Yes	ML*	Twitter	Tweets

5 CONCLUSION

Masses of users share their feelings on social media, making it a valuable platform for tracking and exploring public sentiment. Social media is one of the biggest platforms where massive instant messages are published every day which makes it an ideal source for capturing the opinions towards various curious topics, such as products, goods or celebrities, etc. The main goal of this paper is to give an overview of latest updates in sentiment analysis and classification methods and it includes the brief discussion on the challenges of sentiment analysis for which the work needs to be done. We also found that most of

REFERENCES

[1] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series.," Icwsm, vol. 11, no. 122–129, pp. 1–2, 2010.

[2] M. A. Cabanlit and K. J. Espinosa, "Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons," in Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on, pp. 94–97, 2014.

[3] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proceedings of the 20th international conference on Computational Linguistics, pp. 1367, 2004. classification", Comput. Mater. Continua, vol. 55, no. 2, pp. 243-254, 2018.

[4] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 625–631, 2005.

[5] J. Zhang, Y. Wu, W. Feng and J. Wang, "Spatially Attentive Visual Tracking Using MultiModel Adaptive Response Fusion," IEEE Access, vol. 7, pp. 83873-83887, 2019.

[6] Yuantao Chen, Jin Wang, Runlong Xia, Qian Zhang, Zhouhong Cao, and Kai Yang, "The visual object tracking algorithm research based on adaptive combination kernel", Journal of Ambient Intelligence and Humanized Computing, vol. 10, pp. 4855-4867, 2019.

[7] Yuantao Chen, Jin Wang, Songjie Liu, Xi Chen, Jie Xiong, Jingbo Xie, and Kai Yang, "Multiscale fast correlation filtering tracking algorithm based on a feature fusion model", Concurrency Comput., 23 October 2019.

[8] Y. Tu, Y. Lin, J. Wang, J.-U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification", Comput. Mater. Continua, vol. 55, no. 2, pp. 243-254, 2018.

[9] J. Wang, Y. Gao, W. Liu, A. K. Sangaiah, H.-J. Kim, "An intelligent data gathering schema with data fusion supported for mobile sink in wireless sensor networks", Int. J. Distrib. Sensor Netw., vol. 15, no. 3, Mar. 2019.

[10] J. Wang, Y. Gao, W. Liu, A. K. Sangaiah, H.-J. Kim, "Energy efficient routing algorithm with mobile sink support for wireless sensor networks", Sensors, vol. 19, no. 7, pp. 1494, 2019.

[11] J. Wang, X. Gu, W. Liu, A. K. Sangaiah, H.-J. Kim, "An empower hamilton loop based data collection algorithm with mobile agent for WSNs", Hum.-Centric Comput. Inf. Sci., vol. 9, no. 1, pp. 18, Dec. 2019.

[12] Z. Tang, X. Ding, Y. Zhong, L. Yang, K. Li, "A self-adaptive Bell–LaPadula model based on model training with historical access logs", IEEE Trans. Inf. Forensics Security, vol. 13, no. 8, pp. 2047-2061, Aug. 2018.

[13] J. Zhang, W. Wang, C. Lu, J. Wang, A. K. Sangaiah, "Lightweight deep network for traffic sign classification", Ann. Telecommun., 2019.

Authors

[14] Y. Chen, J. Xiong, W. Xu, J. Zuo, "A novel online incremental and decremental learning algorithm based on variable support vector machine", Cluster Comput., vol. 22, no. S3, pp. 7435-7445, May 2019. Fuji Ren,Ye Wu," Predicting User-Topic Opinions in Twitter with Social and Topical Context", *IEEE* Trans. on affective computing, vol. 4, no. 4, October-December 2013.

[15] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li,"Dual Sentiment Analysis:Considering Two Sides of One Review", *IEEE* Trans.on Knowledge and Data Engineering, 2015.

[16] Xin Chen,Mihaela Vorvoreanu, and Krishna Madhavan," Mining Social Media Data for Understanding Students' Learning

Experiences" *IEEE* trans. on learning technologies, vol. 7, no. 3, JulySeptember 2014.

[17] Danushka Bollegala, David Weir, and John Carroll," Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus", *IEEE* trans. on knowledge and data engineering, vol. 25, no. 8, August 2013.

[18] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, ,Chun Chen, and Xiaofei He," Interpreting the Public Sentiment Variations on Twitter", *IEEE* trans. on knowledge and data engineering, vol. 26, no. 5, May 2014.

[19] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, ,Chun Chen, and Xiaofei He," Interpreting the Public Sentiment Variations on Twitter", *IEEE* trans. on knowledge and data engineering, vol. 26, no. 5, May 2014

[20] Chenghua Lin, Yulan He, Richard Everson, Member, IEEE, and Stefan Ru¨ger," Weakly Supervised Joint Sentiment-Topic Detection from Text", *IEEE* trans. on knowledge and data engineering, vol. 24, no. 6, June 2012

[21] Xiaohui Yu,Yang Liu, Jimmy Xiangji Huang, and Aijun An,," Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain", *IEEE Trans.On* knowledge and data engineering, vol. 24, no. 4, April 2012