SentimentAnalysisofprintedandsocialmediausingNLP

GovindaPatil[1],VivekChaplot[2]

[1]ComputerScienceDepartmentBhupalNoblesUniversity,Udaipur-RajasthanMedi-CapsUniversity Indore, India, Email: patil.govinda1976@gmail.com
[2]ComputerScienceDepartmentBhupalNoblesUniversity,Udaipur-Rajasthan, Email: vivek@bnuniversity.ac.in

Received:05.04.2024 Revised:10.05.2024 Accepted:20.05.2024

ABSTRACT
During The pandemic many stores have changed their mode into online mode that is E-marketing andnot only that because of thismuch increase in online shopping it has become hard to trust which is a reliable product and which is not, to overcome these types of problems comments and star ratings helpus a lot but the star rating cannot be accurate some people have misconception like 1 star for good 5 for bad.The comments which they post don't have this type of ambiguity so if there is a model which can understand and analyse a huge number of comments and give a correct opinion then it will help a lot.Thismodel willhelpus to understand humanlanguageandsotheapplicationcanalsobeusedonprinted media whereby analysing the headlines we can come to know how a year, a month or a day went. Analysing various newspapers will help us to know which paper is imposing negative or positive or neutral thoughts on people.

Keywords: RoBERTa,WebScrapping,Positive,Negative,Neutral,OpticalCharacterRecognition(OCR)

INTRODUCTION

In today's world, with the vast amount of data generated on the internet, it has become increasingly important to be able to analyse and understand the sentiment behind the data. Sentiment analysis is the processofidentifyingandextractingsubjectiveinformationfromapieceoftextanddeterminingwhether it is positive, negative, or neutral. With the help of sentiment analysis, we can understand the opinions and emotions of people towards a particular product, service, or event.

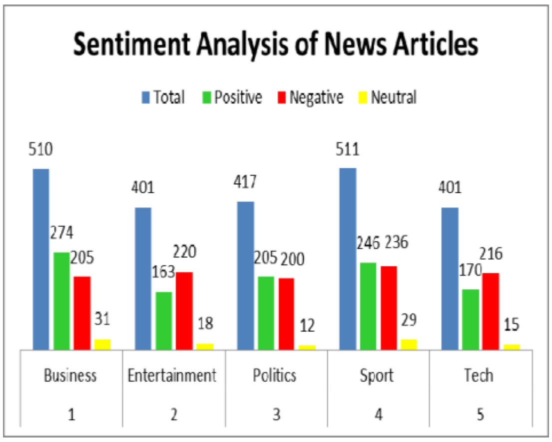
When it comes to product reviews, star ratings can often be misleading. A person might give a product a high rating but write a negative review, or vice versa. This is where sentiment analysis comes in. By analysing the text of the review, we can understand the true sentiment of the reviewer towards the product. Web scraping is a technique used to extract data from websites. By using web scraping, we can extract product reviews and comments from various websites and social media platforms. This data can then be stored in a CSV file for further analysis.

Machine learning models can be trained to extract headings and other important information from web pages. These models can identify patterns in the structure of web pages and extract relevant information such as headings, product names, and prices. Natural Language Processing (NLP) is a field of study that focuses on the interaction between computers and human language. NLP techniques can be used to analyse and understand human language and classify it into positive, negative, or neutral sentiments. By using NLP well defined in [1], we can create machine learning models that can accurately classify reviews and comments into their respective sentiment categories.

RELATEDWORK

Hanif Bhuiyan, Jinat Ara, Rajon Bardhan and Dr. MD Rashedul Islam [2] has developed this project named "Retrieving YouTube Video by Sentiment Analysis on User Comment". Their analysis aids in locating the most pertinent and well-liked YouTube video for the given search. Four phases make up their suggested procedure.Thefirststep pullsdata(comments) froma particularYouTubevideoanddoessomelinguistic pre-processing in order to set up the subsequent steps. The text is then put through an NLP-based approach to create data sets. They then use the data sets to apply their sentiment classifier to get the positivity and negativity ratings. The rating result is then determined using the Standard Deviation. 75.435% was the best accuracy, and 51.21% was the lowest.

Antony Samuels, John Mcgonical [3] has developed this project "News Sentiment Analysis". They concentrated on news stories about incidents that include emotions—good, terrible, and neutral. To examinehumanemotions(i.e.,sentiments)foundintextualdata,sentimentanalysisisused.They conducted sentiment analysis on news stories using a lexicon-based methodology. The studies were run using the BBC news dataset, demonstrating the applicability and validity of their chosen methodology. Data collecting was the first of five processes in the approach. This experiment makes use of the BBCNews dataset. Pre-processing the gathered data was the following stage in order to lessen dataset discrepancies. The wordNet lexical dictionary was then used to determine the word polarity in the gathered news items. The overall emotion score was then computed, with "-1" denoting a negative, "+1" a positive, and "0" denoting neutral. Using the total sentiment score of news stories as a criterion, three classes—positive, negative, and neutral—were identified. The program's lone flaw is that sentiment analysisisonlyusedonEnglishnewsarticlesfromasinglesource.Theydiscoveredthatwhilemanyitems in the entertainment and technology categories reflected negative views, the majority of news stories in thebusinessandsportscategoriesoftendisplayed favourableopinions.Almostequallyasmanyarticlesin the topic of politics expressed favourable and negative opinions.



Ali Al-Sabbagh, Abdulrahman Alrumaih, Harith Kharrufa, Ruaa Alsabah, Kacheru, G. [4] have published this paper "Sentiment analysis of comments in social media". In "Tweets" and comments received in Arabic on Twitter during the 2018 World Cup, this study examines the emotional orientation of textual elements and emoji-based components. To get the results of sentiment analysis for texts and emojis separately, the data is gathered from hundreds of extracted tweets. Emojis and shorthand language are frequently used by people while communicating on computers. Heart symbols, cheerful emoticons, and sobbing expressions indicate the sentimentality of the speaker and frequently replace words in texts or strengthen their emotional impact. Because of this, it's important to understand how emojis function in any sentiment analysis of computer-mediated interaction. The purpose of their study is to look at how Arabiccommentspostedon Twitterduringthe WorldCuparesentiment-orientedin termsofbothtextual and emoji-based elements. It assesses if the analysis of the text without the emojis would yield different findings. This paper suggests that if text and emoji are used in sentiment analysis, an analysis of any Arabic remark would be more accurate. Additionally, it is capable of analysing comments on any other social networking site, including Facebook and Instagram.

Dahab Galal, Mohamed AbdelFattah, Doaa S. Elzanfaly, Nada Hassan, Greg Tallent [5] have published this paper "A Sentiment Analysis Tool for Determining the Promotional Success of Fashion Images on Instagram". Sentiment analysis (SA), also known as opinion mining, is the technique of examining natural language texts to uncover an emotion or a pattern of attitudes about a certain product in order to reach a judgement about that thing. SA is covered by the disciplines of text mining, natural language processing (NLP), and web mining. Instagram is one of the apps that generates a lot of comments and feelings on social media. This study's main goal is to evaluate the replies provided by the top 50 fashion brands on Instagram based on the 20 most popular photos. This is done in order to determine the proposed social value of each image using a sentiment analysis programme. Each image's comments will be input and analysed. Each remark will go through a preprocessing stage in which a lexicon will be used to categorise each word as good or negative. The vocabulary assigns each comment a score value to represent its amount of positivity, negativity, or if it has no effect on the social value. The quantity of likes and shares would be added to these findings to further quantify the image's worth. The social worth of a picture is then calculated as a sum of the individual results. Instagram is used to gather the information needed for thesuggestedapplication.Theinformationgatheredconsistsofthefeedbackleftbyacertainbrand's fans. We can now begin by giving each remark an emotional value and defining the social worth of each brand afterprocessingourdataandremovingunwelcomecomments.Thebrandevaluator, pictureevaluatorand comment analyzer components together make up this module's three stages of functioning..The project's goal is to create a self-contained application that can advise users on the proper elements to put in their images in order to better suit the tastes of their various social media followers.

Haruna Isah, Paul Trundle, Daniel Neagu[6] have published this paper "Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis". applying the recommended framework for brand analysis on Facebook comments and Twitter data, as well as developing a framework for gathering and analysing consumer input on medications and cosmetics using text mining , machine learning and sentimentanalysis,are someofthe contributions madeinthispaper.ApplicationProgrammingInterfaces (APIS) are currently created by the majority of organisations and enterprises, including social media platforms, in order to share data. An API call is made to the Facebook Graph and Twitter APIs for data extraction and authentication throughout the text collecting and cleaning process. Pages, status updates, and comments that hint to user opinions and experiences with medications and cosmetics may be obtained usingthe Facebook GraphAPI.Pre-processingincludes takingoutdelimiters,changingallwords to lowercase, eliminating stop words and digits, stemming words back to their origins, and adjusting certain aspects particular to an application or domain. The sentiment analysis employs both machine learning-based sentiment classification and lexicon-based sentiment classification.They have shown how machine learning methods may be applied to social media data to infer emotions, revealing the opinions and experiences of consumers of pharmaceutical and cosmetic items.

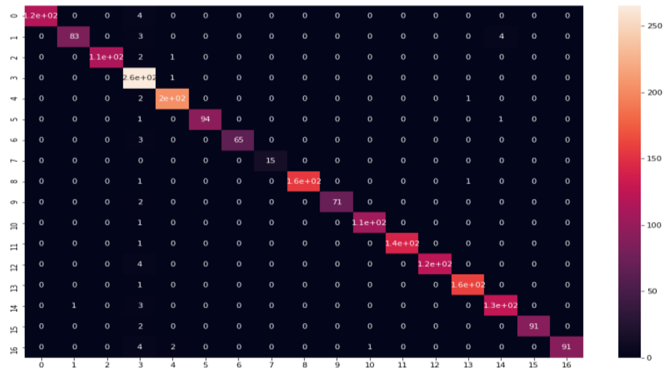
Elton Shah Aly and Dustin Terence van der Haar[6] have published this paper "Slang-Based Text SentimentAnalysisinInstagram".Thevarietyofwaysthatideasmay beexpressedmakesitchallengingto do text sentiment analysis in certain fields. The usage of slang or colloquialisms raises the level of complexity even further. In order to analyse the sentiment of Instagram comments relating to fashion, particularly sports shoes, this study compares the performance of numerous classifiers, including the Naive Bayes, J48, lexicon, and random forest, with that of a slang-based dictionary classifier. Popular Instagram accounts for fashion were used to create the dataset for the benchmark. Images of sneakersand the comments associated with them are pulled from Instagram using the Instagram API. All comments, as well as training and testing data, were collected from Sneakernews, a well-liked Instagram account.A totalof803 comments wereusedtotraineachsupervisedclassifier(NaiveBayes,J48decisions tree, and random forest model), and a total of 398 comments were used to assess the performance of the system. These comments were carefully chosen and gathered from Instagram because there are no publiclyavailable datasetspertinent toslang and basketballfootwear. Pre-processingonthetraining data involves removing stop words, removing punctuation and emoticons, normalising the data, and spell checking. Four classifiers are used to categorise and rate comments that primarily use slang. Followingthe modifications, the lexical classifier and the random forest produced the best results. The downside of thelexiconclassifieristhatitwouldperformbadlyinotherenvironmentssinceitsdictionariesaresowell suited for this particular environment. The least work was necessary to implement the J48 and random forest classifiers, and model training takes only a short while.

Vasudeva Varm, Pinkesh Badjatiya, Shashank Gupta and Manish Gupta [8] have published this paper "Deep Learning for Hate Speech Detection in Tweets". Their primary goal was to create a system that could analyse tweets that contained rudeness, harsh language, or other undesirable influences. For this investigation, they divided tweets into three categories: sexist, racist, and neither. Numerous classifiers were employed, including Deep Neural Networks, Logistic Regression, Random Forest, SVMs, Gradient BoostedDecisionTrees(GBDTs),andRandomForest(DNNs).Toobtainthe bestaccuracy,theycomputed the respective recall and precision for each of the three models.

	Method	Precision	Recall	F1
Part A: Baselines	Char n-gram+Logistic Regression [6]	0.729	0.778	0.753
	TFIDF+Balanced SVM	0.816	0.816	0.816
	TFIDF+GBDT	0.819	0.807	0.813
	BoWV+Balanced SVM	0.791	0.788	0.789
	BoWV+GBDT	0.800	0.802	0.801
Part B: DNNs Only	CNN+Random Embedding	0.813	0.816	0.814
	CNN+GloVe	0.839	0.840	0.839
	FastText+Random Embedding	0.824	0.827	0.825
	FastText+GloVe	0.828	0.831	0.829
	LSTM+Random Embedding	0.805	0.804	0.804
	LSTM+GLoVe	0.807	0.809	0.808
Part C: DNNs + GBDT Classifier	CNN+GloVe+GBDT	0.864	0.864	0.864
	CNN+Random Embedding+GBDT	0.864	0.864	0.864
	FastText+GloVe+GBDT	0.853	0.854	0.853
	FastText+Random Embedding+GBDT	0.886	0.887	0.886
	LSTM+GloVe+GBDT	0.849	0.848	0.848
	LSTM+Random Embedding+GBDT	**0.930**	**0.930**	**0.930**

The range of these approaches' standard deviations was 0.01 to 0.025. Gradient-boosted decision trees and embeddings discovered from deep neural network models were combined to create decision trees with the greatest accuracy values.

Kacheru, G. [9] has published this paper on “Language Detection Using Natural Language Processing”. The goal of this research is to recognise various languages. It could tell if the text was in Swedish, Portuguese, English, Greek, French, Spanish, Dutch, Russian, Japanese, Danish, Italian, Turkish, Arabic, Malayalam, Hindi, Telugu or Tamil. It could also tell if the text was in Danish, Italian, Turkish, or Swedish. It also had the ability to distinguish between Danish and Italian. It was able to recognise 17 distinct languages in all.It employed a multinomial NB model to train and evaluate the data on a dataset known as the Language Detection dataset. It provided an accuracy of 97.22% and was able to function well due of the NB model. This provides a solution for several applications in computational linguistics and artificial intelligence. These prediction algorithms are often used on robotics and electronic devices like laptops and mobile phones for machine translation. It also aids in monitoring and recognising papers in several languages.



Eric Holgate, Daniel Preot-iuc-Pietro. Isabel Cachola and Junyi Jessy Li[10] have published this paper “Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media”. The objective of their evaluation is to determine how explicit knowledge of vulgarity can aid in sentimentprediction. Tothis end,theyfirstcollectedthe data,thencreateddemographicvaluesandsome coding, from which these vulgar tweets were extracted and classified into several classes, including gender, age, education, income, faith, and political ideology. Then, the utterances were divided into neutral,positive,and negativecategoriesusingthesentimentmodel.MeanAbsolute Error (MAE),ametric they used to assess the model, was 1.068. This provided a report on which class receives more profanity and stuff that has been restricted.

Deepak Singh [11] has published this paper “Text Classification of News Articles”. This study focuses on classifying news sections so that readers may access the stuff they are interested in. The work was done manually to avoid human error, and the model aids in categorising the news content in accordance with headlines. The model was trained using BBC news.csv from the Kaggle website. KNN classifier, Decision tree, SVM, Naive Bayes, and other models were used to categorise the headers. They stayed with the classifierthat had the highest accuracy, which was often the random forest classifier. It had a 97.99% accuracy rate.

	Model	Test Accuracy	Precision	Recall	F1
0	Logistic Regression	97.09	0.97	0.97	0.97
1	Random Forest	97.99	0.98	0.98	0.98
2	Multinomial Naive Bayes	97.09	0.97	0.97	0.97
3	Support Vector Classifer	96.64	0.97	0.97	0.97
4	Decision Tree Classifier	83.22	0.83	0.83	0.83
5	K Nearest Neighbour	73.60	0.74	0.74	0.74
6	Gaussian Naive Bayes	76.06	0.76	0.76	0.76

Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena[12] have published this paper “Large-Scale Sentiment Analysis for News and Blogs”. Newspapers and blogs present thoughts regarding news organisations while covering recent occurrences (people, places, and things). They provide a method that rateseachuniquetextcorpusentityandstateswhetherthescoreisfavourableornegative.Theygiveeach word a polarity (positive or negative) and look for synonyms and antonyms; synonyms get the polarity of the parent word, whilst antonyms get the opposite. They annotate all sentiment terms and related things in our corpus using our sentiment lexicons. Every time a feeling word is followed by a negative, they flip its polarity. When a word is followed by a modifier, the polarity intensity is increased or decreased. As a result, poor = -1, good = +1, and excellent = +2.

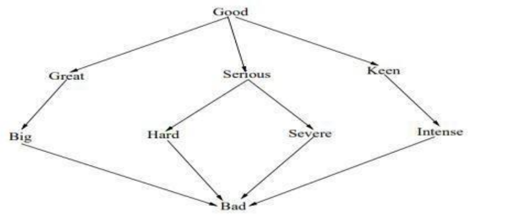
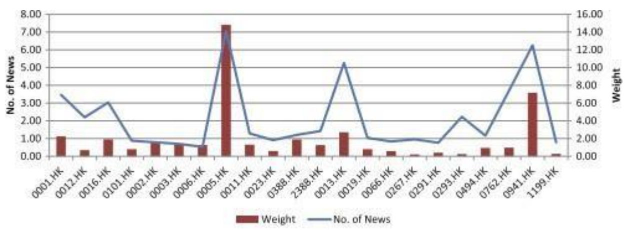


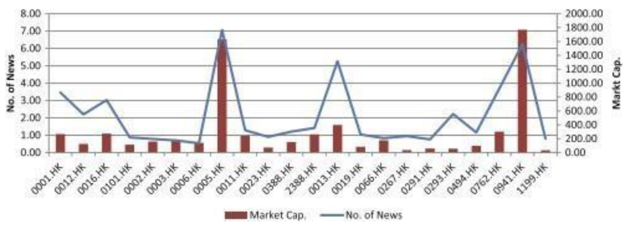
Fig. 1: *Four ways to get from bad to good in three hops*

Dimension	Seeds		Algorithmic		Hand-curated	
	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.
Business	11	12	167	167	223	180
Crime	12	18	337	337	51	224
Health	12	16	532	532	108	349
Media	16	10	310	310	295	133
Politics	14	11	327	327	216	236
Sports	13	7	180	180	106	53

Table 1: *Sentiment dictionary composition for adjectives*

Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang ,Xiaotie[13] Deng have published the paper "Newsimpact on stock price return via sentiment analysis". Financial news articles are supposed to affect the recovery of stock values. They use the Harvard psychology dictionary and the Loughran-McDonald financial sentiment lexicon to generate a sentiment space. Then, utilising objectively quantified text news items, the sentiment space is projected. The instance labelling approach is carefully examined and evaluated. They use a news archive from FINET, a significant provider of financial news in Hong Kong, for the stocks they are looking at. In the experiment, they assess and contrast six various strategies. They put up two sentiment analysis methods, one that makes use of SenticNet and the other that makes use of a bag-of-words. Additionally, they use two conventional methods that depend on the asymmetry in the polarity of the news feeling.





Hassan Raza, M. Faizan, Ahsan Hamza, et al. [14]. have published a paper "Scientific Text SentimentAnalysis using Machine Learning Techniques". On the data set created, experiments are run. The data set is split into training and testing sets in the proportion of 60:40. Various evaluation metrics, such as F- score andAccuracyscore, are usedtocalculate the classifiers' accuracylevels. The outcomes demonstrate that SVM outperforms alternative classifiers. Naive Bayes performs well after SVM. SVM classifier performsbestwhencomputingFscoreandaccuracymeasuresforthemacroaverage,whilerandomforest performs best when computing Fscore and accuracy measures for the micro average. The classifiers were abletoreachthehighest accuracyratings withthehelpoftheperformanceoftheuni-grams,bi-grams,and tri-gram features. Based on various features, they contrasted their results with those obtained during the experimental phase. As the lemmatization procedure was not employed, they combined the n-grams approach with it to minimise the data dimensions.

Setia Pramana, Wahyu Calvin Frans Mariel and Siti Mariyah.[15]in their paper “Sentiment Analysis to Predict Twitter Data using Machine Learning and Deep Learning”. Combined various feature extraction and classification approaches. They discovered that deep learning neural networks outperform SVM and Naive Bayes in terms of performance. Their analysis demonstrates that Bigram and the Deep Learning Neural Network offer the greatest results. The performance of sentiment classification using various combinations ofalgorithms andfeatureextractionmethodsproduces quitea range ofoutcomes.Basedon the aforementioned research and the entire provided dataset, the Deep Learning Neural Network algorithm performs overwhelmingly well. For every dataset, at least one of the feature extraction approach combinations in the Deep Learning algorithm receives the highest score. Additionally, unlike Naive Bayes and SVM, neither balanced nor unbalanced conditions for data composition have asubstantial impact on the modelling outcomes provided by Deep Learning Neural Network. The finalscore is almostalways very high.Bigram is typicallythe bestfeature extraction strategythat can raise the final modelling score for a mix of Deep Learning Neural Network algorithms (N-Grams). As a result, it can be inferred from this research that the Bigram approach in conjunction with the Deep Learning Neural Network algorithm is particularly effective for doing sentiment analysis on Indonesian text data.

Tariq Ahmad Lone, Kh. Muhammad Shafi, Muzafar Rasool Bhat [16] have published the paper "Sentiment analysis of print media coverage using deep neural networking". Any circumstance that exists anywherein the globe may be describedusinga variety of words and phrases.Situations can often be describedin a positive, negative, or neutral way. A deep neural network model (NSAM) is suggested in this study for analysing sentiment in print media coverage of common events throughout a country. A set of 1800 news stories from two regional and two national dailies that covered Jammu and Kashmir, India's most northern state, were used to train the recommended model. Based on accessibility and newspaper readership, local and national newspapers were chosen for the training set of a DNN model. The model consists of five thick layers, with a 20% dropout in the first four and a 0% dropout in the final layer. A MaxPooling layer, an embedding layer, a convoluted layer, and one layer are also included. The dataset featuresacrossvalidationsplitof6,2,and2,respectively,fortraining,validation,andtestdata. Thiswork has shown how sentiment analysis of news items using DNN may be extended to the sphere of printmedia coverage. The model can be expanded to categorise opinions that fall outside of the triplet of positive, negative, and neutral.

Ortigosa and Alvaro [17] have published this paper “ Survey on Sentiment Analysis Techniques on Social Media Data ” in which they have shown the trend of social networking sites.They have suggested a fresh approach to sentiment research on Facebook, Twitter, and other social media sites. It begins when users write messages and aids in identifying the sentiment polarity of social network users, which can be positive, negative, or neutral and is revealed from their posts; and (ii) modelling normal sentiment polarity of users and exploring their significant emotional swings. This approach has been used by the author in Sent Buk, a Facebook application that is also mentioned in the study. In order for a system to draw conclusions based on facts, views, and feelings of the users, it must collect and retain this data. By directly requesting people to fill out surveys so that their opinions can be taken into account, the most significant way for gathering information and data on user opinions may also be used. However, for certain users, this task may be too slow, laborious, and inefficient.

Raktim Kumar Dey, Debabrata Sarddar, Indranil Sarkar, Rajesh Bose, Sandip Roy [18] have published the paper “Sentiment Analysis Techniques Involving Social Media And Online Platforms”. On a multitude of different web forums, blogs, product review websites, and social networking sites people produce and trademillions ofbytesofdataintheInternet realm,whichencompassespracticallyeveryknownareaand field of human endeavour on Earth. The sentiment analysis (SA) concept consists of three levels. Document-based is the initial layer. The second is a sentence-based approach. The third level is aspect, which is also referred to as word- or phrase-based. Techniques for classifying sentiment can be divided into three groups. These include lexicon-based, machine learning, and hybrid techniques. There are numerousclassificationmethodsthathavebeenutilised,includingmachinelearning approaches,decision tree classifiers, and others.This in-depth study covers the uses and difficulties of sentiment classification in both intra- and inter-domain contexts. Subjectivity analysis, negation handling, and polarity classification must be finished prior to applying SA. Additionally, this paper looks at a few notable supervised and unsupervised methods.

Kacheru,G.[19]havepublishedthepaper“SurveyOnTextCategorizationUsingSentiment Analysis”.Theyproposedaground-breakingmethodtoclassifysentimentintobinaryandternarytraining, which may be the most prevalent classes. Separate the text that you extracted from Twitter into seven subclassesaswell.TheycreatedtheSENTAtooltodothis,whichenablestheusertoselectfunctions before running types through seven different classes (fun, happiness, love, neutral, sadness, anger, hate). TocategorisetweetemotionsthetheyusedtextualTwitterdataandaRandomForestclassifier.Theirternary class accuracy was 70.1%, compared to seven training class accuracy of 60.2%. The highest sentiment classification accuracy is produced by the machine learning algorithms Random Forest, Support Vector Machine,NaiveBayes,andLogisticRegression, with respectivevaluesof81.3%,77.99%,79%, and85.1%. In this survey study, a summary of the most current advancements in sentiment analysis techniques and applications was given. Thesehave beenlooked at,and it isclearthat more research isneeded toenhance sentiment classification.

Mr.Karimuzzamanand ArafatHossain"TextMining and Sentiment Analysis of NewspaperHeadlines" is a paper written by Md. Moyazzem Hossain and Azizur Rahman [20]. This essay's goal is to look at the word patterns that appeared on the front page of The Daily Star, a well-known daily English newspaper published in Bangladesh, in 2018 and 2019. Word patterns were also used to try to explain the likely social andpolitical backgroundof thattime period. The three popularand modern textmining techniques usedin the study are wordclouds,sentiment analysis,andcluster analysis.However, a recent assessment found that the majority of sentiment analysis-based research is based on the usage of lexicons and analyses social media and microblogging websites.. The Daily Star, the most widely read daily English newspaperinBangladesh,is thesubjectofthis essaybecauseitsfront-pageheadlinesdeal withimportant topics. After pre-processing and creating a list of terms, the word cloud was utilised because of how well its appearance aids in precise analysis and word placement, size, and boldness. The top 15 terms that appeared more than 40 times were further illustrated by a bar diagram, despite the fact that both the bar and word cloud serve the same purpose.

Ubale Swati, Chilekar Pranali,Sonkamble Pragati[21] Have Published This Paper “Sentiment Analysis Of News Articles Using Machine Learning Approach”. Sentiment analysis's main objective is to determine a writer's attitude toward a certain subject or the overall mood of a work. News analysis may be used toplot the firm's behaviour over time, giving valuable strategic information about the organisation. Themost pertinent research compares different Machine Learning (ML) techniques and algorithms for text analysisandsentimentanalysis.Positive,negative,andneutralclassesareusedtocategorisethetext.This research demonstrates that naive bayes classifiers can be used well for sentiment analysis. As compared to otherclassifiers usedforsentimentanalysis, it provides superioraccuracy. Maximumentropy, decision trees, winnow, and c4.5 classifiers are further classifiers used for comparison.The algorithms utilised not only produce better results than the alternatives, but they also cut down on processing time. Due to the employment of the quick and precise naive bayes classifier, which ensures user happiness and offers a cost-effective methodology, the results obtained will be optimised and obtained more quickly.

Nirag T. Bhatt, Asst. Prof. Saket J. Swarn deep et al. [22], intheirpaper “NLP Based Review Categorization: A Survey”. They provide straightforward explanations of sentiment analysis, including its types, degrees, applications,andbenefitsanddrawbacks.They focusalotonvariousfeatureextractionmethodsandhow to use various machine learning algorithms. This study uses lexicon-based, machine learning (NaiveBayes,SVM,RandomForest, LogisticRegression),anddeeplearningtechniquestoassessthesentimentor polarity of the dataset. The accuracy of these methods was 88.38%, 92.31%, 92.39%, 91.32%, and67.46%, respectively (LSTM, CNN).It is evident from the visualisation of all the accuracy percentages that machine learning outperforms deep learning and lexicon-based approaches in terms of precision. Sentiment extraction serves a practical purpose in a wide range of industries, including education,politics, entertainment, and health care. Future improvements to this model are possible as deep learning did not produce satisfying results inthis experiment. In this case, we selected anepoch size of10, but you may raise it to acquire a higher accuracy rate.

Barakat AlBadani, Ronghua Shi, et al. [23] in their paper “A Novel Machine Learning Approach for Sentiment Analysis on Twitter ” SVM and fine-tuning were combined to construct a deep learning model. They tested the accuracy of their model using three datasets. For Twitter US Airlines, IMDB, and the GOP debate, they received accuracy scores of 99.78%, 99.71%, and 95.78%, respectively. They thought about document-levelsentimentanalysisintheirwork butnotaspect-levelanalysis.Notmanycompaniesutilise sentimentanalysisofthiskind.A book'schaptersorpagescanbecategorisedasgood,negative,orneutral using this method. Both supervised learning techniques and unsupervised learning approaches can be used to categorise the material at this level. The two largest problems with document-level sentiment analysis are cross-domainand cross-language sentiment. (Saunders, 2021) It has beendemonstrated that domain-specific sentiment analysis can reach exceptional accuracy while remaining extremely domain- sensitive. The feature vector in these tasks consists of a set of restricted and domain-specific terms.

Peng Cen, Kexin Zhang, and Desheng Zheng et al. [24] published a paper “Sentiment Analysis Using Convolutional Neural Network” They analysed the sentiment of movie reviews from the IMDB dataset usingRNN,LSTM,andCNNintheirwork.RNNandLSTMarelessaccuratethanCNN.Theaccuracyof

CNN, RNN, and LSTM are, respectively, 88.22%, 68.64%, and 85.32%. RNN and LSTM models are frequently used for sentiment analysis, while CNN is frequently used for image identification. They ranan experiment and found that the CNN neural network was still effective, processing a sequence of data with an accuracy of 88.22%. The accuracy provided by previous research utilising various machine learning approaches was compared to the accuracy achieved, and it was discovered that the recommended deep learning techniques (CNN, RNN, LSTM) performed better than SVM, RNTN. We want to improve the influence of movie reviews and other emotional evaluations in the future, build stronger experimental models,andcreateintegrationutilising thesuperpositionmodeltechnique. Thedata preparationhasabig impact on model recognition and feature extraction.

Pasumpon Pandian et al. [25], published a paper "Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis"They employed 6 datasets in their study, which show algorithms that combine automated feature extraction and manual feature separation. They plan to expand their work and test it using additional languages in the future. Statistical analysis is used to determine the information compiled from many studies and the attributes necessary to outperform sentiment classification performance. With regard to deep learning approaches, the suggested work covers the fundamental framework necessary to characterise the sentiment analysis that is already available based on conventional research methodology.Analysis reveals that the performance of the suggested work significantly outperforms that of the already used methods. This serves as an illustration of how fusing data from several sources, such as effect word vectors, generic features, and surface characteristics, might improve sentiment analysis jobs. Finally, this study will offer recommendations for testing methods that could be used to boost deep sentiment analysis' efficiency. This approach may eventually be applied consistently to other languages as well. To include the proposed work with reference to emotion analysis, research is being done.

PROPOSEDSYSTEM

The proposed system consists of three main components: data extraction, sentiment analysis, and visualization. The first component, data extraction, involves extracting comments from a given post or extracting headlines from printed media. The system utilizes web scraping techniques to extract the comments from the given link. The extracted data is then cleaned and pre-processed to remove any irrelevant information.

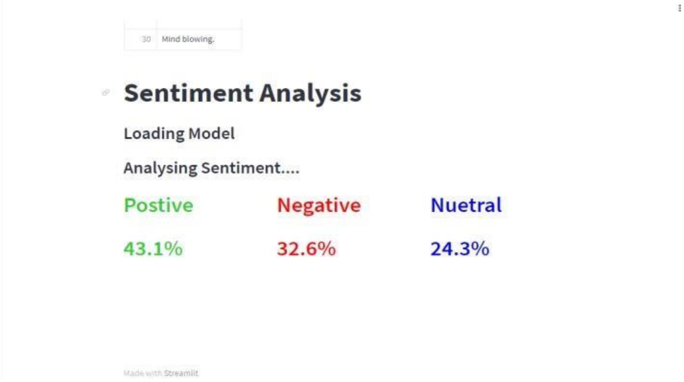
The second component of proposed system involves using machine learning algorithms and NLP techniques to perform sentiment analysis on the extracted data. For this we will utilize RoBERTa model, which is a transformer-based language model. The RoBERTa model is trained on large corpus of text data and can learn the contextual relationships between words and phrases. The RoBERTa model has several advantages over traditional machine learning models, such as its ability to capture complex semantic relationships between words and phrases and its high accuracy on various NLP tasks.

Thefinalcomponent, Visualization, involvesvisualizing theresultofthesentiment analysis.Theproposed system utilizes a dashboard to display the sentiment analysis results in a user-friendly format. The dashboard displays the overall sentiment of the comments or printed media, as well as the sentiment distribution across different topics or categories. Additionally, the system can provide insights into the most common positive and negative sentiments expressed in comments.

SETUP

Our setup is simple and user friendly where he/she is provided with three options Amazon, social media and printed media. As shown in below figure:

Where the user is asked to enter the URL of the amazon product and from there our scraping code gets intoactionandsavesallthecommentswiththeirdates.Foranyothersocialmedialikeyoutube, Instagram,Twitteretc. awebsitenamedexportcommentsisusedtoextractcommentsandsavethemInto a csv file where as for printed comments we used OCR to extract text and save them into a CSV file. After the data is collected our model is applied and then we generate results as shown below.



RESULTSANDDISCUSSIONS

Wehaddeveloped anapplicationwhich can be used toanalyzevarious data, ourfocus was totally onhow well toanalyze anyproduct or service by people's feedback. This applicationcanaccurately tell how good the product is or how well the social media is influencing people in negative way or positive way. Fig-1 shows how a COVID-19 related video influenced people over a period of time. The graph likely shows changes in number of views or engagement with the video, and can help to determine whether the videois having a positive or negative impact on the audience. It could also reveal trends in people's interests or concerns related to pandemic.

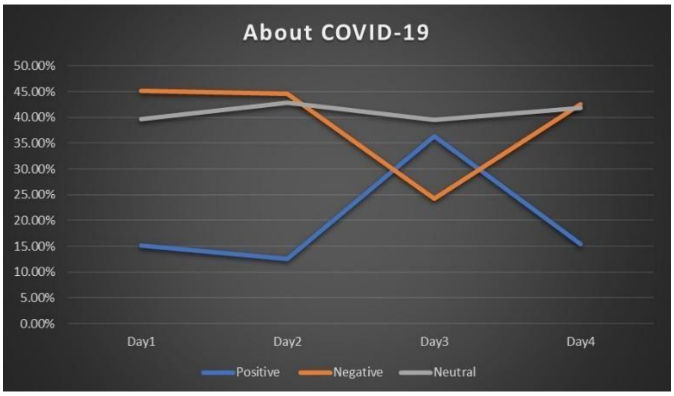


Figure1

Fig-2 shows how the Ukraine wars news influenced people. It would be interesting to know what specific aspects of the news were analyzed in this graph, such as whether it was related to the conflict itself or to political developments. This information could be useful for understanding how news coverage affects people's perceptions of world events.

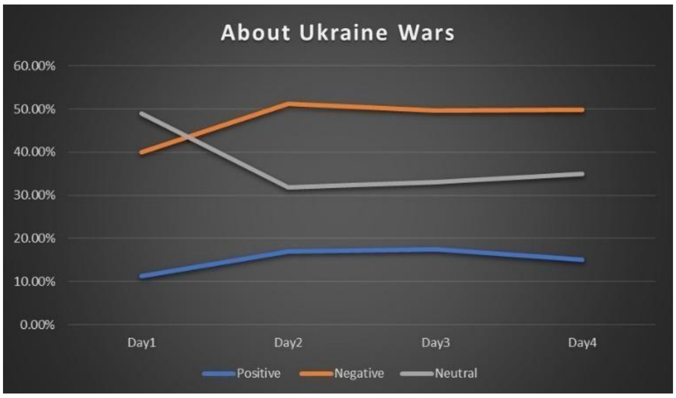


Figure2

Fig-3 tells us about an Amazon product which is a laptop brush, and the graph shows customer satisfaction over a period of time. This type of graph is useful for businesses to track how satisfied their customers are with their products, and to identify any trends or issues that need to be addressed.

Figure3

Fig-4 tells about an Instagram product how happy the customers are after buying this product van be analyzedthroughthisgraphandwecanalsoseewhenthenegativeinfluenceisincreasingbythisanalysis and prevent those actions.

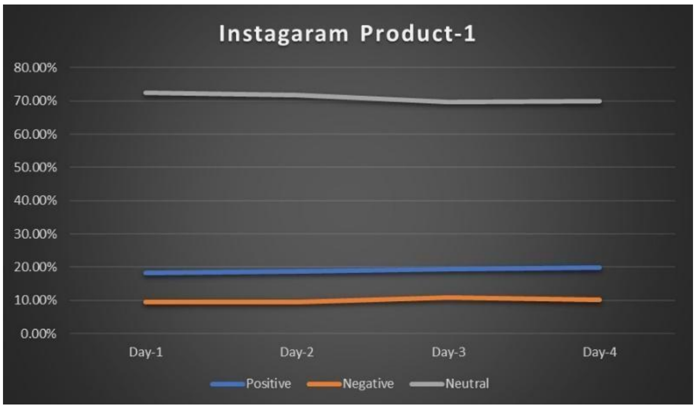


Figure4

Finally, Fig-5 is an analysis of printed comments from people who watched a match in a stadium, and the graphtellsabout theirexperience. Thistype ofanalysis canhelptoprovideinsight into whatfans liked or disliked aboutthe match,as well as anysuggestions for improvements. This information can be useful for event organizers or sports teams to enhance the fan experience in the future.

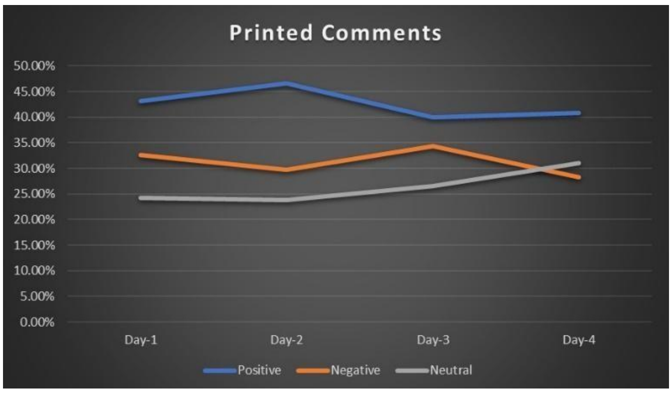


Figure5

CONCLUSION

In conclusion, sentiment analysis is an important field of natural language processing that allows us to understandandanalysehumanemotionsandopinionsexpressedintext.Throughtheuseofmachine learning algorithms and techniques, we can build models that can accurately classify text into positive, negative, or neutral sentiments.

In this article, we have demonstrated the effectiveness of sentiment analysis on a dataset of consumer reviews for a product, their comments about it in social media and the headlines in printed media. We have shown by using various pre-processing techniques such as stop word removal, stemming, and lemmatization, along with a range of machine learning algorithms, we can achieve high accuracy in classifying sentiments.

However,therearestillsomechallengesthatneedtobeaddressedinsentimentanalysis,suchashandling sarcasm, irony, and other forms of figurative language. Nonetheless, sentiment analysis has a wide range of applications, including brand monitoring, customer feedback analysis, and social media monitoring. As such, sentiment analysis will continue to be a valuable tool in understanding and analysing human emotions and opinions in text.

REFERENCES

[1] Chirgaiya, Sachin et al. 'Analysis of Sentiment Based Movie Reviews Using Machine Learning Techniques'. 1 Jan. 2021 : 5449 – 5456.

[2] Hanif Bhuiyan, Jinat Ara, Rajon Bardhan and Dr. MD Rashedul Islam "Retrieving youtube Video by Sentiment Analysis on User Comment" MAY 2018, Conference: ICSIPA 2017 : IEEE International Conference on Signal and Image Processing applicationsicsipa 2017

[3] AntonySamuels,JohnMcgonical[2]hasdevelopedthisproject"NewsSentimentAnalysis"MAY 2007,UniversityofSouthernCalifornia,Caltech.

[4] Kacheru, G. (2018). Blockchain Technology: Architecture, Applications, and Challenges. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 9(3), 1433–1438.

[5] Ali Al-Sabbagh, Abdulrahman Alrumaih, Harith Kharrufa, Ruaa Alsabah, James Baldwin "Sentiment analysis of comments in social media". International Journal of Electrical and Computer Engineering (IJECE)Vol.10, No.6, December2020

[6] Dahab Galal, Mohamed abdelfattah, Doaa S. Elzanfaly, Nada Hassan, Greg Tallent "A Sentiment Analysis Tool for Determining the Promotional Success of Fashion Images on Instagram". British University in Egypt, Cairo, Egypt

[7] HarunaIsah,PaulTrundle,DanielNeagu"SocialMediaAnalysisforProductSafetyusingTextMining andSentimentAnalysis".20October2014,IEEE.

[8] EltonShahAlyandDustinTerencevanderHaar"Slang-BasedTextSentimentAnalysisinInstagram". 3rd JANUARY 2020, Part of the Advances in Intelligent Systems and Computing book series(AISC,volume 1027)

[9] Kacheru, G. "Aipowered test automationframeworks: choosing the righttools." International journal of artificial intelligence & machine learning (IJAIML) 3.02 (2024): 110..

[10] Eric Holgate, Daniel Preot-iuc-Pietro. Isabel Cachola and Junyi Jessy Li "Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media", Computer and Information Science, University of Pennsylvania

[11] DeepakSingh "TextClassificationofNewsArticles".DECEMBER2021,AnalyticsVidya.

[12] NamrataGodbole,ManjunathSrinivasaiah,StevenSkiena["Large-ScaleSentimentAnalysisforNews andBlogs".GoogleInc.,NewYorkNY,USA

[13] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang ,Xiaotie "News impact on stock price return via sentiment analysis". JUNE 2018

[14] Hassan Raza, M. Faizan, Ahsan Hamza, et al. "Scientific Text Sentiment Analysis using MachineLearning Techniques". 2019, (IJACSA) International Journal of Advanced Computer Science and Application

[15] Setia Pramana, Wahyu Calvin Frans Mariel and Siti Mariyah. "Sentiment Analysis to Predict Twitter Data using Machine Learning and Deep Learning". JULY 2022, International Journal of Engineering Research & Technology (IJERT)

[16] Tariq Ahmad Lone, Kh. Muhammad Shafi, Muzafar Rasool Bhat "Sentiment analysis of print media coverage using deep neural networking", JULY 2018.

[17] OrtigosaandAlvaro"SurveyonSentimentAnalysisTechniquesonSocialMediaData", SEPTEMBER 2018,InternationalJournalofRecentResearchAspectsISSN2349-7688

[18] RaktimKumarDey,DebabrataSarddar,IndranilSarkar,RajeshBose,SandipRoy"SentimentAnalysis Techniques Involving Social Media And Online Platforms", May 2020, International Journal Of Scientific & Technology Research Volume 9, Issue 05

[19] Kacheru, G. "The role of AI-Powered Telemedicine software in healthcare during the COVID-19 Pandemic." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 11.3 (2020)

[20] M.BouaziziandT.Ohtsuki"Survey OnText Categorization UsingSentiment Analysis",August2019, InternationalJournalOfScientific&TechnologyResearchVolume8,ISSUE 08

[21] Mr. Karimuzzaman and Arafat Hossain "Text Mining and Sentiment Analysis of Newspaper Headlines", AUGUST 2021, MDPI
[22] Ubale Swati, Chilekar Pranali, Sonkamble Pragati "Sentiment Analysis Of News Articles Using Machine Learning Approach", April 2015, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835
[23] Nirag T. Bhatt, Asst. Prof. Saket J. Swarn deep et al. [21], in their paper "NLP Based Review Categorization: A Survey", SEPTEMBER 2022
[24] Barakat albadani, Ronghua Shi, et al "A Novel Machine Learning Approach for Sentiment Analysis on Twitter ", JANUARY 2022, MDPI
[25] PasumponPandianetal."PerformanceEvaluationandComparisonusingDeepLearning Techniques inSentimentAnalysis",AUGUST2019,IEEE
[26] PasumponPandianetal. "PerformanceEvaluationandComparisonusingDeepLearning Techniques inSentimentAnalysis"2021,JournalofSoftComputingParadigm(JSCP)(2021)Vol.03/No.02