Sentiment Analysis of Social Media Reviews Using Machine Learning and Word Embedding Techniques

Perla Anitha[1], K. Uma Rani[2] N. Selvamuthukumaran[3] Suneetha Bobbillapati[4] , K. Dhivya[5] Amit Gupta[6]

School of Engineering , Department of ECE, Department of CSE , Department of ECE, Department of CSE, Department of AI & ML

Department of Data Science, Malla Reddy , Nalla Malla Reddy , Narasaraopeta, Nalla Malla Reddy, Nalla Malla Reddy

Malla Reddy University , Engineering College, Engineering College , Engineering College, Engineering College, Engineering College

Telangana (India), Telangana (India), Telangana (India), Andhra Pradesh (India), Telangana (India), Telangana (India)

anithakrishna77p@gmail.com[1],karanamumarni@gmail.com[2],selvamuthukumaran.cse@nmrec.edu.in [3]

suneetha.bobbillapati@gmail.com[4], erkdivyapandiyan@gmail.com[5],dramitguptacv@gmail.com[6]

Abstract— The enormous quantity of information originated and distributed on societal communications each and every day highlights the need for knowledge-extraction processes to be automated. Sentiment investigation is an energetic area of information extraction investigate that faces numerous problems.

Any company should be eager to hear from its customers. Customers primarily rely on reviews to decide where to eat. Sentiment analysis is critical in categorizing restaurant reviews as positive, negative, or disinterested in organize to assess whether the cuisine is good, secure, and worth choosing over other restaurants. In this study, the Yelp dataset for restaurant reviews is used to examine various word embedding algorithms, similar as Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), GloVe,Word2Vec, and Doc2Vec. Supervised Machine Learning (ML) approaches such as Logistic Regression and Support Vector Machine are assessed using performance metrics similar as F1-Score, Accuracy, Recall, and Precision. Comparable results show that combining a Support Vector Machine and TF-IDF word embedding technique yielded results with a 98% accuracy. We investigate several pre-processing strategies and apply various features and classifiers.

We present results on additional generally popular topics, such as movie and product reviews, in addition to the social media dataset we produced. We anticipate that this paper will not only broaden existing sentiment analysis research to another language family, but will also foster competitiveness, potentially leading to the development of high-end commercial solutions

Keywords: Bag of words, GloVe, Logistic Regression (LR), Restaurant Reviews, Sentiment Analysis, Support Vector Machine(SVM), TF-IDF, Word2Vec and Doc2Vec.

I.INTRODUCTION

The disseminate of Internet engineering science has aided in the development of social communications and e-commerce stages . Furthermore, consumers represent becoming more habitual to conveying their thoughts and feelings to others via these platforms, which use script or mixed media data [1,2,3,4]. As a result of this phenomenon, a large amount of data has been gathered and generated that may be analyzed to evaluate sentiment.

Analyzing sentiment is beneficial to both individuals and corporations, especially given the vast volume of data created [5].

However, as indicated in [6,] identifying, continuously monitoring, and filtering data on social media applications to study sentiment might be problems. Among the challenges include unstructured data, linguistic differences, a compass of localities and social media platforms, and varied information on people's ideas.

Dang et al. [7] proposed three fundamental techniques to sentiment investigation hybrid, lexicon- conventional and machine learning-based. Corpus-based and dictionary-based lexicon-based approaches are the two types. Sentiment studies are performed using machine learning algorithms to categorize text data and determine the polarity of online product reviews (using datasets from Amazon, IMBD, and other sources), sometimes combining machine learning and deep learning approaches. [8, 9] investigate sentiment investigation using machine learning algorithms.

Our framework is able to take advantage of the potent capability of pre-formation terminology examples like GloVe and eliminate numerous of the difficulties affiliated through the supervised learning models thanks to a new unsupervised approach that E. S. Alamoudi reported on [10]. The elements of food, service, ambiance, and price have been divided into categories based on the sentiment surrounding them.

In the digital age, more than 4.62 billion [1] people use social media, generating massive amounts of data. Users can share suggestions or comments on social networking networks such as Facebook, Instagram, Twitter, LinkedIn, Amazon, Flipkart, Zomato, Yelp, and others. Individuals and businesses alike value this quick, low-cost, and effective mass communication.
One of the benefits of social media is the ability to communicate ideas and opinions, sometimes known as online reviews.

According to research [2,] 90% of buyers look for product reviews or comments before making a purchase. Customers can quickly learn everything they need to know about a product by reading reviews. As a result, in the age of social media and digital marketing, client feedback has become increasingly important.

Global social media adoption has had a significant impact on both digital marketing and our daily lives. Social media has an impact on fashion, fitness, health, money, restaurants, and other businesses. Over 61% [3] of consumers read internet restaurant reviews, well exceeding the national average.

Online restaurant reviews have an equal impact on both customers and restaurant owners. When making restaurant reservations, the majority of clients use online research to learn more about a number of factors such as the food's quality, cost, cleanliness, service, and ambiance.

Restaurant managers use customer feedback to improve customer service and business operations, which can increase sales and reputation. Restaurant reviews can be found on Yelp, Google My Business, and a variety of other websites. Yelp is a well-known restaurant review and reputation management website that foodies use to research restaurants before dining out. Yelp boasts 142 million monthly users, with the vast majority of its reviews for restaurants.

People frequently upload text-based product and service evaluations, photos, and five-star ratings on this website. The Yelp dataset has two fundamental flaws: first, the massive volume of customer evaluations makes going through each text-based review tedious; and second, the five-star rating is ambiguous, making it impossible to discern the consumer's reasons if they gave the same star rating.

Sentiment analysis is the procedure of extraction speech from a text, deleting unnecessary terms, and finding essential phrases in order to understand the client's attitude and feelings. This is performed through the utilize of a amalgamation of Natural Language Processing (NLP) and machine learning (ML) approaches. Simple sentiment analysis may be used by the application to identify whether a text conveys positive, negative, or neutral emotions. It is reliable.

The goal of this study is to do a Sentimental Analysis on Yelp restaurant reviews. Data preprocessing, tokenizing using word embedding approach similar as Bag of Words (BOW), TF-IDF, Glove, Word2Vec, and Doc2Vec), feature selection, and extraction are the main approaches used in this study. Later, employ ML techniques like as LR and SVM. The ML model's performance is evaluated using performance metrics based on a classification report that includes a Precision, disturbance matrix, correctness, recall, and F1-score.

The following is the work's structure: Section II contains related work. Section III explains how to create the model development dataset. Section IV employs machine learning algorithms to assess the efficacy of various word embeddings. Section V summarizes the paper and discusses future work.

II. RELATED WORK

Machine Learning-based approaches have been used in a number of investigations. The authors of (Vairetti et al., 2020) suggested a amended interpretation of the support vector machine (SVM) algorithm. The benefaction of each division of a review, including the title and body Word embedding and ML approaches have been applied to Yelp restaurant datasets in significant scholarly research.According to Eman S. A. et al., opinion mining has greatly enhanced comprehension and decision-making. The text of online reviews, as well as their rankings, was analyzed in this study.

To evaluate restaurant reviews on the Yelp website, two sentiment classifications, binary (positive or negative) and ternary (positive, negative, and neutral), were utilized. Predictive models have been used in conjunction with machine learning, deep learning, and transfer learning, among other techniques.

The real-valued sentiment intensity ratings from the enlarged Affective Norms of English Words (E-ANEW) sentiment lexicon are employed in Yu et al.'s (2018) proposal for a word vector refinement model to improve already-trained word vectors. Jianqiang et al. (2018) used the AFINN lexicon to illustrate the GloVe-DCNN (GloVe-deep convolution neural networks) word embedding approach. The study provided by Rao et al. (2018) focused on one of the issues in using long short-term memory networks (LSTM) for document-level sentiment categorization, specifically the modelling of semantic relationships between words. This study can help with the deployment of deep learning sequence models. The main method in [4] is to use an SVM model to extract the sentiment tendency of each review from its word frequency in the Yelp restaurant dataset.

TF-IDF and BoW were used to extract features. The SVM models' word scores are then translated into a polarity index, which represents the weight of each word for the various restaurant categories. The model has an accuracy rating of 88.9%.

This [5]study makes use of Yelp Challenge and Amazon Fine Food Reviews online review datasets. Two separate techniques are employed in this work to classify binary sentiment. First, a non-neural BoW method based on SVM and Multinomial NB is described. A Long Short-Term Memory (LSTM) is used in the second method. The LSTM approaches utilised a variety of embeddings, including Word2vec and Glove embeddings. The results show that LSTM approaches using GloVe and Word2vec embeddings outperform NB and SVM, which mandated a BoW strategy.

Z. Desai et al. [6] used NLP to classify review sentiments from a dataset using Deep Learning -LSTM and BERT and ML techniques such as LR, Decision Tree(DT), Stochastic Gradient Descent(SGD), NB, and SVM. Business intelligence (BI), notably Microsoft PowerBI, assists businesses in selling products, streamlining procedures, and improving customer satisfaction. The research [7] addressed the topic of predicting restaurant review rating using multi-class classification using ML approaches on Yelp restaurant reviews.The star rating is the class label. Nearly sixteen models were developed using four different feature extraction procedures (indexing, Unigrams, Bigrams, Latent Semantic Indexing, and trigrams) and four supervised machine learning (ML) algorithms (NB, LR, SVM, and perceptrons). The prediction performance of LR on 10,000 Bigrams and Unigrams features is superior.

Lak et al. [8] created a system to assess the efficacy of sentiment analysis on reviews from various service and product websites. The findings of sentiment analysis and star ratings are compared. We want to know if sentiment analysis results can be used in place of star ratings when such ratings are available, and, more crucially, if they can be used in place of star ratings when such ratings are not accessible.Lexalytic, a sentiment analysis application, allows you to choose which reviews to read and highlight sections of the reviews to obtain the essential correct data.

Authors provide an overview of multi-lingual analysis utilizing a lexicon-based approach. The input text is decoded and converted into reference language. The new target lexicon is assigned to the sentiment scores. On a Movie Review dataset, Pang et al. (2002) tested unigrams (presence of a specific word, frequency of terms), bigrams, part-of-speech (POS) tags, and adjectives.

Martineau and Finin (2009) investigated several unigram weighting techniques based on the TFIDF model (Manning et al., 2008) and suggested delta weighting for a binary situation (positive, negative). Paltoglou and Thelwall (2010) presented more improvements in delta TFIDF weighting based on their technique.

III. METHODOLOGY

Figure 1 depicts the proposed work's model development steps. On Yelp Restaurant Review datasets, the model employs multiple classifiers, word embedding approaches, and hyper-parameter tuning with gridsearchcv. On supervised ML algorithms, performance measures assess the various word embedding strategies. The following section describes the many stages of model development:

A. Dataset

This paper utilizes a subset of Yelp: Yelp restaurant reviews from Kaggle [10]. More than 44,611 reviews on Yelp consist of text, images, and organization details. The Yelp Restaurant Review dataset consists of the following attributes:

- Business id- Business identification number.
- Date- the review posted date.
- Review id - Review identification number.
- Star- 1 to 5 grading for the business.
- Text- the review text.
- Type- the type of text.
- User id- the user’s identity who posted the review.
- Cool, useful, and funny are the comments on the review given by another user.

To increase the trustworthiness of the material used for training, this work considers 20,000 text reviews on 2000 eateries in four different categories, including a cafe shop, barbeque, Chinese eateries, and fast food, as indicated in Table I. As shown in Table II, each review contains approximately 80 English phrases such as good, poor, like, similar, and so on.

B. Data Preprocessing

1) Handling Missing rows/columns: Two ways to handle missing values are deleting specific rows or columns containing NULL values when the dataset has sufficient examples. This work removes irrelevant columns from the Yelp dataset. Second, utilizing statistical methods like mean, mode, and median to replace a missing value. The mode and mean are used in this study to substitute missing categorical variables and numerical values, respectively. The figure1 illustrated that work flow diagram of identification of Sentiment Analysis of Social Media in different process..

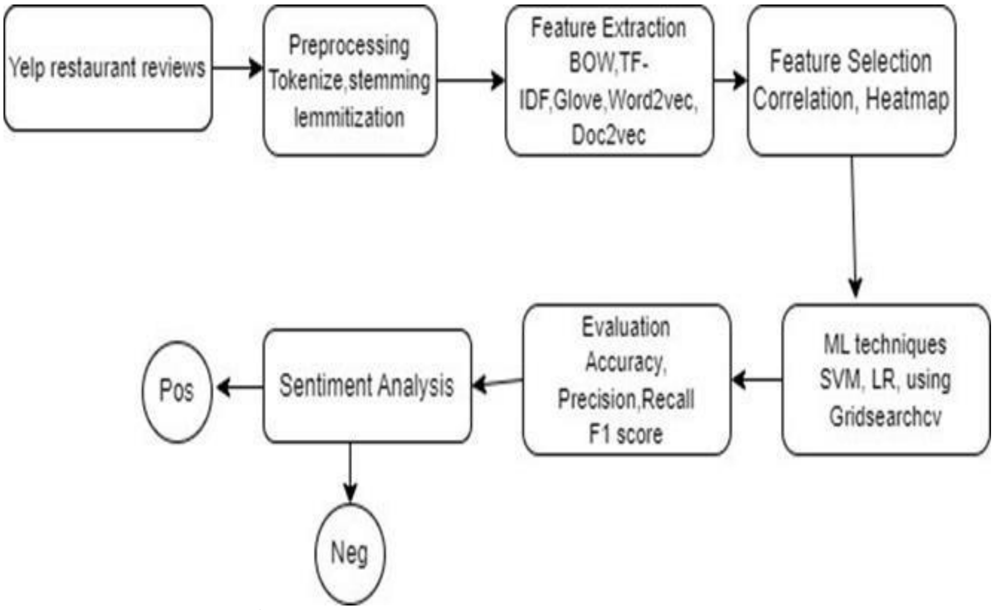

Fig.1 Work flow diagram in machine learning process

TABLE I :RESTAURANT REVIEW SAMPLE

Sl No.	Reviews
1	Fantastic atmosphere, and the cuisine is fantastic. The cuisine on the menu is really varied and superb, and the treatment is the highest.
2	The restaurant´s atmosphere was just acceptable. The fish was excellent, but the spaghetti was only averagely tasty. The ambiance and service are excellent.

2) Eliminating Outliers:. Outliers are caused by data entry error, often known as noisy data. Outliers in the restaurant review datasets are visualized using a boxplot. It provides the interquartile and median ranges, which aids in the examination of the distribution.

A. Tokenization, stemming, and Lemmatization: Tokenizing the text is the first stage in the NLP workflow. Tokenization is the process of dividing a document into words, punctuation, numerical values, and so on. Stemming is the process of developing morphological variations of a root or base word. For example, by deleting 'ed' and 'ing' from the terms served, serving, and serve, it simplifies them to the fundamental word serve.

Lemmatization is similar to stemming, except the meaning of the root word is preserved. Tokenization is performed in this study using the NLTK Python package, which is derived from the Snowball stemmer, and lemmatization is performed using WordNet Lemmatize

.

B. **Feature scaling**:Feature scaling is a technique for standardizing independent variables so that they fall inside a particular range. The the StandardScaler approach is used for feature scaling. After subtracting the mean and establishing a characteristic, it scales to unit variance.

C. Feature Extraction: This study's dataset is primarily text-based, hence certain pre-processing procedures are required before applying classification algorithms. Texts must be cleaned up and turned into numerical features before they can be used by prediction models. The following factorization methods were employed in this work:

- BoW –. is a straightforward and versatile technique for extracting components from documents. The literary representation of word repetition in a document is referred to as a "BoW." The emphasis is on word count rather than grammatical standards or word structure. The model records the term's occurrence but ignores its location.

- TF-IDF - The TF-IDF textual feature extraction technique uses an inverse proportion of the words in the whole text to determine the relative occurrence of terms in a given text.
- GloVe – GloVe is an unsupervised learning approach for creating word vector representations. It is a count-based model that takes into account both local and global statistics.
- Word2Vec - To process text, it is a two-layer neural net that vectorizes words. It takes in a text corpus and returns a set of vectors: feature vectors matched to corpus words.
- Doc2Vec - • Doc2Vec is a Word2Vec extension that applies to a whole document/review rather than individual words. This model seeks to generate a numerical representation of a document/review rather than a word representation. Doc2Vec is based on the idea that a word's meaning is influenced by the context in which it appears.

D. Feature Selection

As the number of attributes in the Yelp restaurant dataset grows, there is a chance that the threshold will rise, lowering the model's accuracy. The model with additional characteristics gets inefficient during training. To address this issue, a subset of the dataset's features is chosen. Several feature selection algorithms are available, including univariate selection, feature importance, and a correlation matrix with a heatmap.

A correlation matrix with a heatmap is employed in this paper. As shown in Figure 2, data is grouped using star ratings to see whether there is any association between attributes such as cool, useful, and amusing. The graph indicates that Useful is hyper-parameters have a significant impact on the model's performance.

the greatest ways to tweak the hyper-parameters. The dictionary is given all potential value combinations as input. The GridsearchCV is used to evaluate the model for each combination. Finally, the hyper-parameter combination with the highest accuracy/performance is selected.

param grids ='C': [0.1, 1, 10, 100, 1000],

- 'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
- 'Kernel': ['rbf,' 'linear', 'sigmoid']
- n jobs: set to -1 means all processes will run in parallel
- verbose: set to 1 to get the detailed printout while fittingthe data to GridSearchCV.

E. Performance Metrics

Confusion matrix, precision, recall, F1- score, and accuracy are used to evaluate the performance of positive and negative sentiments on the Yelp restaurant review dataset. The figure 2 illustrated that matrix for text data analysis.

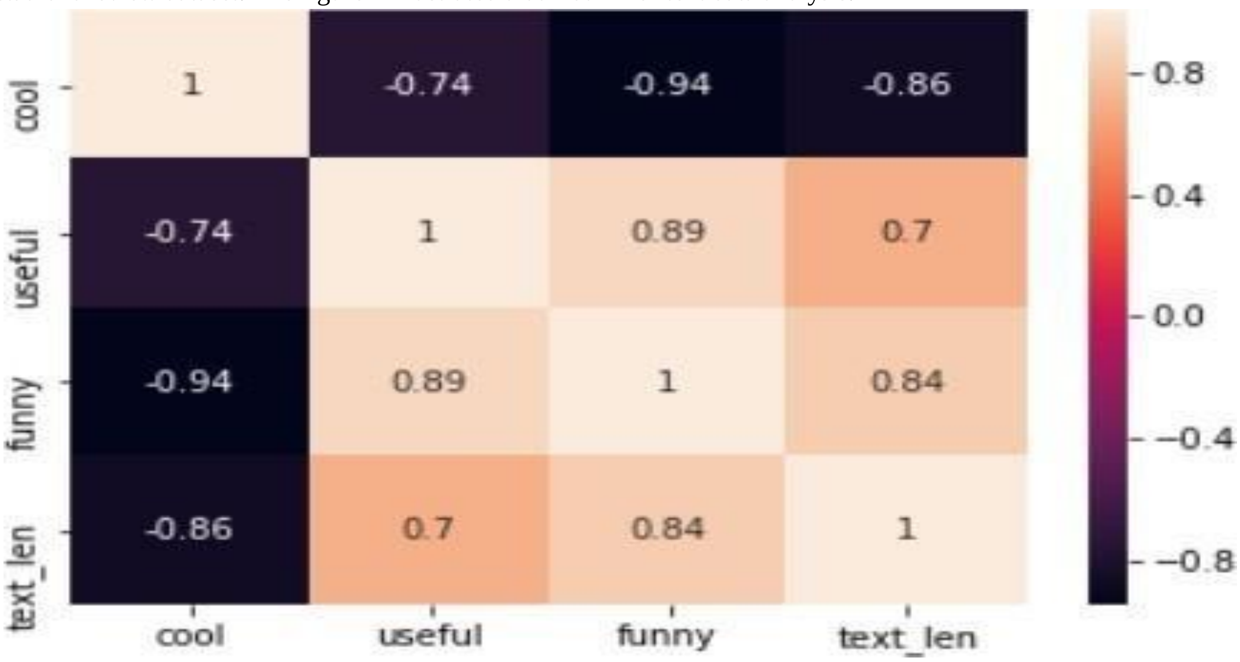

Fig. 2. Heatmap Loop matrix pf text Data

Text length is tightly related to Useful, and it is substantially associated with humorous. Furthermore, the association between cool and the other three traits is negative, implying that longer evaluations are more amusing and beneficial. Finally, for this job, seven features were evaluated (company id, date, review id, star, text, type, user id).

F. Machine Learning Algorithms

Different machine learning models used to perform sentiment analysis of Yelp restaurant reviews are: Precision - Among the positive predictions, precision computes the percentage of True Positive. It is characterized as a proportion of True Positive predictions to total positive predictions(either correct or incorrect predictions).

- Recall - Recall gauges how well a model can identify positive samples. It measures the proportion of accurate positive predictions to all the positive examples found in the dataset.

F1-Score - The harmonic mean of precision and recallinto a single metric.Support Vector Machine: It is an algorithm for supervised machine learning. The categorization of a sample is predicted using SVM [12], using labeled input data separated into two classes by a margin. The input is directly translated into a higher dimension, anda support vector classifier is used as a threshold to most accurately differentiate the two classes (or hyperplane).

G. Grid SearchCV

Hyperparameter tuning is a technique to recognize the bestvalues for a definite model. The value of the Confusion Matrix - A N x N matrix assesses the efficiency of a classification model, where N is the number of target classes. The matrix compares the predicted values of the machine learning model with the actual target variables. In this work, target values aregood or bad reviews. Hence, the target class(N) value is2, and Table III denotes a 2X2 confusion matrix.

TrPos - True Positive exactly predicts the positive class.

TrNeg - True Negative exactly predicts the negative class.

TABLE II CONFUSION MATRIX

Confusion Matrix		Actual Values	
		Positive (1)	Negative(0)
Predicted Values	Positive(1)	TrPos	FaPos
	Negative (0)	FaNeg	TrNeg

- FaPos - False Positive inaccurately predicts the positiveclass.
- FaNeg - False Negative incorrectly predicts negativeclass.

IV. EXPERIMENTAL SETUP AND RESULTS

The model is written in Python and runs on an Anaconda navigator using the jupyter notebook IDE environment. This platform enables the creation of models using various machine learning and deep learning methods. In this setup, several Python libraries for machine learning must be installed. Kaggle was used to choose the Yelp restaurant reviews dataset. The dataset is separated between train and test sections in an 80:20 ratio. To handle the given text, data preparation techniques are used. Several word embedding approaches, including as BoW, TF-IDF, GloVe, Word2Vec, and Doc2Vec, are used in feature extraction. To pick essential features, a correlation matrix with a heatmap is employed. Several machine learning classifiers, such as SVM and LR, are classified as positive or negative after feature extraction and selection.The results of the experiment are tabulated Table III The performance metrics is calculated for LR and SVM classifiers using BoW, TF IDF, GloVe, Word2Vec and Doc2Vec word embedding techniques. From the tables Support Vector Machine with TF-IDF embedding techniques outperforms with the highest accuracy of 98%.

Table III. Techniques Using LR Algorithm

Logistic Regression (LR)					
Word embeddings	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Confusion Matrix**
BOW	0.974	0.987	0.960	0.973	[[1376 17] [55 1338]]
TF-IDF	0.975	0.990	0.960	0.975	[[1380 13] [55 1338]]
Glove	0.895	0.900	0.889	0.894	[[1256 137] [153 1240]]
Word2Vec	0.925	0.929	0.921	0.925	[[1295 98] [109 1284]]
Doc2Vec	0.902	0.899	0.906	0.902	[[1252 141] [131 1262]]

Fig. 3 illustrated that different technique used for logistic regression method for accuracy,precision,recall,F1score.

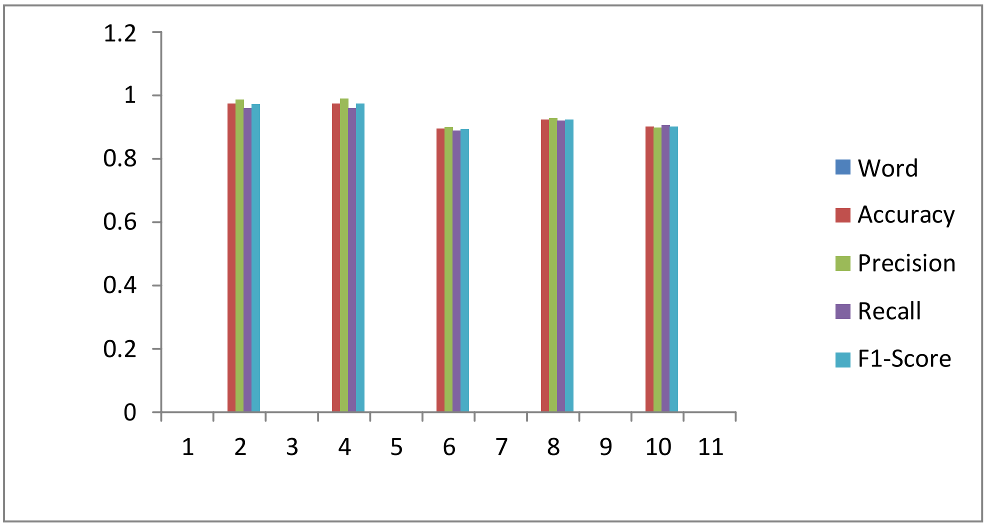

Fig. 3. Different Technique used for data analysis

TABLE IV. Performance Metrics of Different Word Embedding Techniques Using SVM Algorithm

Support Vector Machine (SVM)					
Word embeddings	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Confusion Matrix**
BOW	0.973	0.981	0.965	0.973	[[1368 25] [48 1345]]
TF-IDF	**0.980**	**0.986**	**0.974**	**0.980**	**[[1374 19] [36 1357]]**
Glove	0.956	0.976	0.934	0.955	[[1362 31] [91 1302]]
Word2Vec	0.967	0.983	0.951	0.967	[[1371 22] [68 1325]]
Doc2Vec	0.907	0.914	0.898	0.906	[[1276 117] [141 1252]]

Figure 4 illustrated that support vector machine for method for accuracy, precision, recall &F1score.

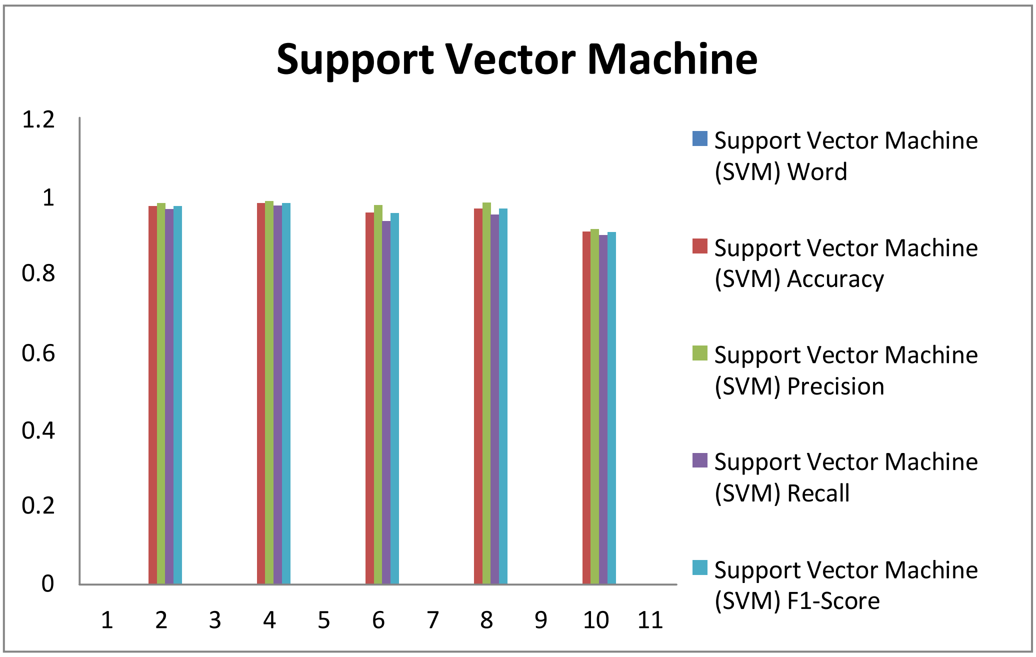

Fig. 4. Support Vector Machine used for accuracy, precision, recall& F1score

V. CONCLUSION

The majority of customers preferred hotel meals during the COVID-19 outbreak because reviews commended it as hygienic, healthy, and delicious. Based on the reviews, sentiment analysis can tell whether a reviewer has a favorable or unfavorable point of view. The goal of this study is to examine the performance metrics for several embedding of words, which are given in figure IV.The article provides an in-depth investigation of supervised machine learning methods for sentiment analysis in Czech social media. We created a huge Facebook dataset with 10,000 postings that were highly agreed-upon (Cohen's 0.66) by humans. The dataset is freely used for non-profit purposes.19 We thoroughly examined numerous state-of-the-art features, classifiers, and language-specific preprocessing techniques. The complete examination of supervised machine learning methods for sentiment analysis in Czech social media was provided in this research. We produced a huge.

Facebook dataset includes 10,000 posts and human annotation with significant agreement (Cohen's 0.66). The dataset is publicly accessible for non-commercial use.19 We extensively assessed numerous cutting-edge features and classifiers, as well as language-specific preprocessing approaches.

VI. REFERENCES

[1]Behera, R.K.; Naik, D.; Rath, S.K.; Dharavath, R.: Genetic algorithm-based community detection in large-scale social networks. Neural Comput. Appl. 32(13), 9649–9665 (2020). https:// doi.org/10.1007/s00521-019-04487-0.

2. Zhang, Z.; Sun, R.; Zhao, C.;Wang, J.; Chang, C.K.; Gupta, B.B.:CyVOD: a novel trinity multimedia social network scheme. Multimed.Tools Appl. 76(18), 18513–18529 (2017). https://doi.org/10.1007/s11042-016-4162-z.

3. Gupta, S.; Gupta, B.B.:XSS-secure as a service for the platforms of online social network-based multimedia web applications in cloud. Multimed. Tools Appl. 77(4), 4829–4861 (2018). https://doi.org/10.1007/s11042-016-3735-1.

4. Wang, H.; Li, Z.; Li, Y.; Gupta, B.B.; Choi, C.: Visual saliency guided complex image retrieval. Pattern Recognit. Lett. 130, 64–72 (2020). https://doi.org/10.1016/j.patrec.2018.08.010.123

[5]Liu, B.: Text sentiment analysis based on CBOW model and deep learning in big data environment. J. Ambient Intell. Human Comput. 11(2), 451–458 (2020). https://doi.org/10.1007/s12652-018-1095-6.

6. Pathak, A.R.; Pandey, M.; Rautaray, S.: Empirical evaluation of deep learning models for sentiment analysis. J. Stat. Manag. Syst.22(4), 741–752 (2019). https://doi.org/10.1080/09720510.2019.1609554.

[7]. Dang, N.C.; Moreno-García,M.N.; De la Prieta, F.: Sentiment analysis based on deep learning: Acomparative study. Electronics 9(3),483 (2020). https://doi.org/10.3390/electronics9030483.

[8]. A. S. Rathor, A. Agarwal, and P. Dimri, ―Comparative study of machine learning approaches for Amazon reviews,‖ Procedia Computer Science, vol. 132, pp. 1552–1561, 2018.

[9]. J. Sadhasivam and R. B. Kalivaradhan, ―Sentiment analysis of amazon products using ensemble machine learning algorithm,‖ International Journal of Mathematical, Engineering and Management Sciences, vol. 4, no. 2, pp. 508–520, 2019.

[10] Eman Saeed Alamoudia and Norah Saleh Alghamd,‖ Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings‖, journal of decision systems,2021, VOL. 30, NOS. 2–3, 259–281.

[11] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational

[12]B. Yu, J. Zhou, Y. Zhang, and Y. Cao, –Identifying restaurant features via sentiment analysis on yelp reviews,‖ *arXiv preprint arXiv:1709.08698*, 2017.

[13]J. Barry, ―Sentiment analysis of online reviews using bag-of-words and lstm approaches.‖ in *AICS*, pp. 272–274, 2017.

[14]Z. Desai, K. Anklesaria, and H. Balasubramaniam, –Business ntelligence visualization using deep learning based sentiment analysis on amazon review data,‖ in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–7, 2021.

[15]N. Asghar, –Yelp dataset challenge: Review rating prediction,‖ *arXiv preprint arXiv:1605.05362*, 2016.

[16]P. Lak and O. Turetken, ―Star ratings versus sentiment analysis–a comparison of explicit and implicit measures of opinions,‖ in *Hawaii international conference on system sciences*, pp. 796–805, 2014.

[17]A. Sadia, F. Khan, and F. Bashir, ―An overview of lexicon-based approach for sentiment analysis,‖ in *2018 3rd International Electrical Engineering Conference (IEEC 2018)*, pp. 1–6, 2018.

[18]S.-B. Yang, S. Hlee, J. Lee, and C. Koo, ―An empirical examinationof online restaurant reviews on yelp. com: A dual coding theory perspective,‖ *International Journal of Contemporary Hospitality Management*, pp. 121–130, 2017.