Twitter Sentiment Analysis For Feature Extraction Using Support Vector Machine (SVM) With TF-IDF

K. Brindha[1]*, Dr. E. Ramadevi[2]

[1]Research Scholar, NGM College, Tamilnadu, India.
[2]Associate Professor, Dept of Computer Science, NGM College, Tamilnadu, India.

***Corresponding Author: -** K. Brindha
*Research Scholar, NGM College, Tamilnadu, India.

Abstract - The goal of this research is to perform sentiment analysis on Twitter data by extracting relevant features using the Hybrid Support Vector Machine (SVM) with TF-IDF (term frequency-inverse document frequency) weighting scheme. Sentiment analysis involves analyzing the emotional tone of a piece of text, and in this case, is specifically interested in classifying tweets as positive, negative, or neutral. To achieve this, using a yahoo stock market dataset of tweets using the Twitter API and preprocess the data by removing stop words, special characters, and URLs. Proposed the SVM-TF-IDF to extract features from the preprocessed tweets calculates the importance of each word in a tweet by taking into account both its frequency within the tweet and its rarity across the entire stock dataset. The research aims to demonstrate the effectiveness of the SVM-TF-IDF for feature extraction in sentiment analysis tasks and to provide insights into the sentiment trends and patterns present in Twitter data.

Keywords: Stock Market, Machine learning technique, Twitter, Sentiment analysis, Feature extraction, Tweet annotation.

1. Introduction

The age of the web has had an impact on the manner in which people express their ideas. The social networking websites depend largely on the user-generated content. Typically, when people intend to purchase a product, they browse through a lot of websites to gain some information about the products before they make their purchase. They take into consideration the available reviews and ratings of these products on these websites before making purchases. The amount of data is inordinate for an ordinary individual to dissect it utilizing naive techniques. Hence, to make this interaction effective and to robotize it, a few opinion examination procedures are utilized. Symbolic techniques or knowledge based approach and machine learning procedures are commonly used to foster such models. In knowledge-based approach, research requires a comprehensive data set that contains pre-characterized feeling data and an efficient knowledge representation for identifying sentiments. Machine learning approach utilizes training informational index to recognize the feelings of each word accurately. Thus, this technique does not require the database of words like knowledge-based approach and therefore, is better. Using a mixture of both this techniques, predominantly machine learning technique; build a hybrid model which will be capable of analyzing sentiments of almost any text. A detailed overview of various techniques is discussed in this paper.

The Yahoo stock market dataset of tweets obtained through the Twitter API is a collection of text data that contains information about the sentiment of Twitter users towards Yahoo's stock prices. The dataset includes tweets that were posted during a specific period of time and contain keywords related to Yahoo's stock market performance. Preprocessing of this dataset is an important step in preparing the data for analysis. The preprocessing stage involves cleaning and transforming the data to remove any noise and make it suitable for further analysis. Some of the preprocessing steps that can be applied to this dataset include text cleaning, which involves removing stop words, punctuation marks, and other irrelevant characters. Tokenization is another important preprocessing step, which involves breaking the text data into individual words or tokens.

Twitter Sentiment Analysis

Due to increase in contents over social media such as Twitter, Facebook, and Trip advisor, express opinion about products, services, or any government policy among others. Twitter having 336 million1 active users monthly is now a main source of feedback for government, private organization, and other service providers. On Twitter around 500 millions tweets are produced per day2, generating huge measure of unstructured text information. Sentiment analysis (opinion Mining) is utilized to retrieve the knowledge data from the tweets posted by clients. Twitter sentiment is used to classify the tweets into neutral, positive, or negative. Many researchers have presented classification methods in sentiment analysis

With the advent of social media platforms, in the last 15–20 years, their popularity and the number of active users on these platforms are only increasing. And rightly so, after-all they provide efficient means to connect to family, friends, people care about, mentors, recruiters and everyone else on the platform, with access to internet. Now its noticeable phenomenon that when two or more people get together scope for crisis to arise. And here are talking about millions, even billions, of people. So naturally, almost every day, there's some user, somewhere on earth, posting or commenting or tweeting not-so-pleasant words. Or in some other case, a customer may write a review of a product recently purchased. That product itself may have reviews from thousands of customers and more on their way.

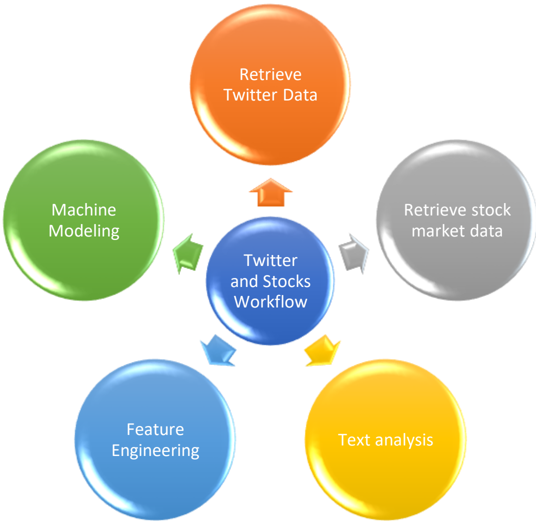

Figure 1.Twitter and Stock Workflow

Microblogging is, nowadays, one of the most well known specialized services among Internet clients. Consistently, millions of clients share opinions on different aspects of life. Micro blogging clients expound on their life, share opinions on variety of topics and discuss recent concerns. Due to the free format of these messages and the wide availability of microblogging platform clients in mobile and other devices, internet users tend to move from traditional communication apparatuses, like traditional blogs, to miniature publishing content to a blog services.

Twitter is a microblogging platform and social networking service that lets its users to post real time messages, called tweets. Tweets have many unique characteristics the most important of which is that they are limited to 140 characters in length. Tweet sentiment refers involve additional classes such as the neutral sentiment, attribute multiple sentiments to a single instance, or have different magnitudes of sentiment instead of binary classification. Sentiment Analysis on Twitter data is more complicated than for large reviews / posts because the tweets are very short and mostly contain slangs, emoticons, hash tags and other twitter language specific characters.

Opinion mining

Opinion mining is a natural language processing technique used to decide if data is positive, negative or neutral. Labels of positive, negative or neutral are contextual and subjective, but in essence they carry their literal meaning. Automating this process can have numerous advantages.

Since tweets are basically short texts, tweet classification methods borrowed many techniques from text classification. Text examination is traditionally performed with Normal Language Processing (NLP) techniques. However, NLP tools cannot always deal with tweets since the latter often do not follow even the simplest and most basic syntactic rules. On the other hand, in some occasions NLP tools are still necessary for tweet analysis as in the case of tweet author's gender identification. Nevertheless, the type of writing in tweet posts is so particular that it has been proven through many experiments that NLP alone is not enough if want optimal results.

As referred earlier, one of the ways twitter uses NLP is by analyzing and classifying all the tweets as positive, negative or neutral. This helps in keeping the platform positive overall and provide a pleasant user experience. As a result, it is the responsibility of the platform to manage these unappealing texts, in order to serve as a medium for people's connection and not a medium of propagandizing hate or to analyze the review written by a valuable customer to let them know of their worth in the company. Additionally thusly breaking down client analysis, similar to feelings in overview reactions and web-based entertainment discussions, licenses brands to acknowledge what satisfies clients or frustrated, so that they can tailor products and services to meet their customers' needs. But searching through and reading billions of text is a laborious task. This is where Natural Language Processing (sentiment analysis) comes into play.

2. Related Work

2.1 Linear Matrix Inequality (LMI)

D. Deepa et.al proposed Sentiment Analysis using Feature Extraction and Dictionary-Based Approaches. This paper is worried about the remarkable stability examination for time-defer systems. Introductory, two new weighted integral inequalities are introduced in view of the helper capability based integral inequalities. In the new weighted integral inequalities, not at all like past examinations, dramatically weighted integral vectors are utilized to find the lower limits of the weighted integral quadratic terms. Then, by utilizing the new weighted integral inequalities, one more linear matrix inequality (LMI) still up in the air for the dramatic stability of the considered time-defer structures. At last, the mathematical models are directed to approve the viability of the new LMI condition. The model results show that the not set in stone in this paper is less moderate than existing ones in breaking down dramatic stability of the thought about systems.

2.2 Corpus-based sentiment analysis (CBSA)

F. Alqasemi et.al proposed an upgraded highlight extraction technique for further developing sentiment analysis in Arabic language. Sentiment analysis (SA) is a cutting edge message mining disciplinary that acquired notable situation due its different application in social networks (SN) and numerous internet domains. Since, it is used for discovering audience directions, and impressions about products or any subjects discussed in the internet via social media Personal feelings availability in SN focused entirely on discussions producers on Current Corporation. This paper showed five states of the SA corpus, and an implementation of four classification operations on them, with various features selection. The main significant results depicted the enhancing of accuracy that based on merging terms after pruning them. Then selecting features that depend on lexicon-based SA without any other features weights had been tested by utilizing other features with it. Also, this had asserted that using these weights alone is better, especially with SVM classifier.

2.3 Self-Attention and Gated Convolutional Networks

J. Yang et.al proposed Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks. Aspect based sentiment analysis (ABSA) is a fine-grained sentiment analysis task, whose fundamental objective is to recognize the sentiment extremity of an aspect in a sentence. A sentence could contain different aspects, all of which could have different sentiment polarities. Considering the momentum investigates around here, ABSA can be separated into two subtasks: aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA). This paper proposes an efficient convolutional neural network consolidating self-consideration component and gating instrument for ABSA errands. In any case, they found the amazing execution of self-consideration calculation and gating component in some other related explores. Thusly, to tackle the issues above, form another model based on gating instrument, joined with convolutional neural networks (CNN) and self-consideration component. In the first place, they utilize self consideration regarding separate the underlying component of the info, and coordinate it with the highlights of the first sentence extricated by CNN. On such premise, they further join the aspect-category or aspect-term of the info sentence to frame the last sentiment highlight. Probes SemEval datasets show the exhibition of research models and the adequacy of the model are demonstrated.

2.4 Support Vector Machine and co-reference resolution

M. H. Krishna et.al proposed a feature based approach for sentiment analysis using SVM and co-reference resolution. Online shopping is one of the most comfortable ways to shop in this new era of technology and people buy online products frequently and post their reviews about the products they have used. The viewpoint of the user will be in the form of tweets or product reviews which they post in an e-commerce site. These reviews will also help the manufacturers to improve the features of the product as required but it is a very difficult task to manually read the reviews and assign sentiment to them. This problem can be solved by creating an automated system in which they can analyze the reviews posted by the users and extract the user's perception about a particular feature. In this paper have proposed a combination of SVM and co-reference resolution to improve the accuracy of feature based sentiment analysis. Stanford dependency parser has been used to extract opinion words about the features and SentiWordNet has been used to assign scores to these opinion words.

2.5 Pearson correlation

F. R. Saputra Rangkuti et.al proposed Sentiment Analysis on Movie Reviews Using Ensemble Features and Pearson Correlation Based Feature Selection. Microblogging has become the media information that is very popular among internet users therefore; the microblogging became a source of rich data for opinions and reviews especially on movie reviews. They proposed, sentiment analysis on movie review using selection Features Pearson's Correlation to reduce the dimension of the feature and get the optimal feature combinations. Use the feature selection is done to improve the performance of the classification, reducing the dimension of the feature and get the optimal feature combinations. In this study, feature selection is employed for ensemble features based sentiment analysis. The method used is Pearson correlation. One advantage offered by feature selection is obviously reducing computational complexity of the task without declining the system accuracy. The experiment results show that the method used can retain the classification accuracy. The best result is obtained when using threshold 20% with accuracy 82%, precision 86%, recall 79, 62% and f-measure 81, 9%. This performance equals to the result obtained without feature selection.

3. Proposed Methodology

In this paper, research extracts aspects or features for building a classification model. The extracted features are in a format suitable to sustain directly to machine learning algorithms from datasets containing raw data of different formats, such as text, a sequence of symbols, and images. Therefore, research uses certain techniques to extract features from the tweets, such as count vectorizer and term frequency-inverse document frequency (TF–IDF). In this research use SVM-TF-IDF to extract the feature matrix that reflects the importance of terms to the corpus in a text.

3.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification tasks. It works by finding a hyperplane in a high-dimensional feature space that maximally separates the two classes. In the case of Twitter sentiment analysis, the two classes are positive and negative sentiments.

3.2 Term Frequency–Inverse Document Frequency (TF– IDF)

TF–IDF is a statistical measure and a term-weighting scheme that provides the bag-of-words model with information on word importance. Different aspects are extracted from the processed dataset, such as verbs, adjectives, and nouns. Subsequently, these aspects are used to calculate the sentiment polarity in a sentence to determine the opinion of individuals by using models, such as unigrams, bigrams, or n-grams.

TF–IDF is used to evaluate the significance of a word to a document in a dataset. Each word is assigned a weight in the document. Research that TF–IDF vectorizer Python module of Scikit-learn. A TF–IDF vectorizer extracts features based on word count, providing less weight to frequent words and more weight to rare words.

The TF– IDF weight of a term within a document is Refer to ‘‘(1)’’, where t is a term appearing in document d:

$$\boldsymbol{TF - IDF\ (t, d) = TF(t, d).IDF(t)}$$

SVM-TF-IDF finds the weight of each feature in a document using the product of term frequency (TF) and inverse document frequency (IDF). TF is the frequency of a feature in document and depends on the length of the document. It can be defined as

$$\boldsymbol{TF_{t,d} = \frac{Count_{t,d}}{totalcount_d}}$$

Where count (t, d) is the number of term t in the document d and totalcountd is the total number of all terms in the document d. IDF measures the extent of a term t being informative in a document for model training. It can be computed as

$$\boldsymbol{IDF = \frac{N}{DF_t}}$$

$$\boldsymbol{IDF(t) = log\frac{N}{DF(t)}}$$

Where TF (t, d) is the number of times the term t appears in a particular document d and IDF is the total size of document N divided by the number of documents in the entire dataset D, which contains the term t.

Where N is the number of reports in the corpus and Dft is the quantity of records that contain the term t. IDF measures the weight of a term t low when term t occurs frequently in many documents. For instance, stop words have low LMI value. Finally, SVM-TF-IDF can be defined as

$$\boldsymbol{TF - IDF = TF_{t,d} * log(IDF)}$$

Word ngrams

Unigram and bigram are extracted for each word in text without any stemming or stop-word removing, all terms with occurrence less than 3 are removed from the feature space.

Negation Features

The standard based algorithm presented in Christopher Potts Sentiment Discussion Instructional exercise is carried out. This algorithm appends a negation suffix to all words that show up inside a negation scope which is determined by the negation key and punctuation. All these words have been added to the feature space.

Twitter dictionary

All terms presented in a text and in the twitter dictionary presented in 3.3 are mapped to their corresponding terms in the dictionary and added to the feature space.

Z score

Z score can distinguish the importance of each term in each class; their performances have been proved and assume as in the mentioned work that the term frequencies are following a multi-nomial distribution. Thus, Z score can be seen as a standardization of the term frequency using multi-nomial distribution. Research compute the Z score for each term t_i in a class C_j (t_{ij}) by calculating its term relative frequency tfr_{ij} in a particular class C_j , as well as the mean $(mean_i)$ which is the term probability class C_j , and standard deviation (sd_i) of term ti as per the fundamental corpus and tried unique thresholds for choosing the words which have higher Z score.

$$\boldsymbol{Zscore(ti)\ =\ \frac{tfr_{ij} -\ mean_i}{sd_i}}$$

Thus, research added the number of words having Z score higher than the threshold in each class positive, negative and neutral, the two classes which have the maximum number and minimum number of words having Z score higher than the threshold. These 5 features have been added to the feature space.

Semantic Features
The semantic representation of a text may bring some important hidden information, which may result in a better text representation and a better classification system.

Brown Dictionary Features
Each word in the text is planned to its group in Brown, 1000 features are added to features space where each feature addresses the number of words in the text mapped to each cluster.

Topic features
Latent Dirichlet association is configured with 10 topics and the training data is used for training the model, then for each sentence in the test set, the trained model estimates the number of words assigned to each topic.

Semantic Role Labeling Features
Encode semantic role labeling features in SVM classifier. Research system also extracts two types of features, the names: the whole term which represents an argument of the predicate and the tags: the type of each argument in the text (A0 represents the subject of predicate, A1 the object, AM-TMP the time, AMADV the situation, AM-loc the location). These encodings are defined by the tool which used (Senna). Research that predicates contentions can comprise a multi-word expression which might be useful in Sentiment Classification.

3.3 Proposed Hybrid Support Vector Machine (SVM) with TF-IDF
SVM is a popular machine learning algorithm for text classification tasks. In this approach, the TF-IDF vectorization method is used to convert the Twitter text into feature vectors, and then SVM is used to classify the tweets into positive or negative sentiments.
Here's the proposed methodology for using SVM with TF-IDF for Stock Market Twitter sentiment analysis:
Feature Extraction Convert pre-processed text data into numerical feature vectors using the TF-IDF vectorization method. The TF-IDF score of a term in a document is calculated as follows:

$$\boldsymbol{TF(t,d) = (Number\ of\ times\ term\ t\ appears\ in\ document\ d)\ /\ (Total\ number\ of\ terms\ in\ document\ d)}$$

$$\boldsymbol{IDF(t) = log_e(Total\ number\ of\ documents\ /\ Number\ of\ documents\ with\ term\ t\ in\ it)}$$

$$\boldsymbol{TF - IDF(t,d) = TF(t,d) * IDF(t)}$$

Train/Test Split: Split the Stock Market data into training and testing sets to evaluate the model's performance.
Training: Train an SVM model on the training Stock Market data using the TF-IDF feature vectors. SVM tries to find a hyperplane that maximizes the margin between the two classes while minimizing the classification error. The SVM optimization problem can be formulated as follows:

$$\boldsymbol{minimize\ \left(\frac{1}{2}\right)\left|\left|w\right|\right|^2 + C * Sum_i = 1^n\ max(0, 1 - y_i(w^T x_i + b))}$$

where ||w|| is the Euclidean norm of the weight vector w, C is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error, x_i is the feature vector for the i-th training instance, y_i is its corresponding class label (1 for positive, -1 for negative), b is the bias term, and n is the number of training instances.
Testing: Evaluate the performance of the trained SVM model on the testing Stock Market data by calculating various evaluation metrics such as accuracy, precision, recall, and F1 score.
Fine-tuning: Fine-tune the SVM model by adjusting the hyperparameters (such as the regularization parameter C and the kernel function) to optimize performance.
Prediction: Use the trained and fine-tuned SVM model to predict the sentiment of new Stock Market Twitter data by converting its text data into TF-IDF feature vectors and applying the SVM model to classify it into positive or negative sentiment.

Algorithm: SVM-TF-IDF
Procedure Feature Extraction (clean_tweet):
Extract Features using (SVM-TF-IDF) in a format suitable to machine learning algorithms
End Procedure
Procedure Balancing and scoring (Features):
Calculating polarity degree of the Stock Market tweets, ‘balancing and labeling the tweets
End Procedure
Procedure Sentiment Classification (Feature):
Classify tweet using machine learning technique (SVM)
Input: Pre-processed Stock Market Twitter data (text and corresponding sentiment labels)
Output: Predicted sentiment labels (positive/negative/neutral) for new Twitter data
Convert pre-processed text data into numerical feature vectors using the TF-IDF vectorization method.
Split the data into training and testing sets.

Train an SVM model on the training data using the TF-IDF feature vectors.
Evaluate the performance of the trained SVM model on the testing data.
Fine-tune the SVM model by adjusting the hyperparameters (such as the regularization parameter C and the kernel function) to optimize performance.
Use the trained and fine-tuned SVM model to predict the sentiment of new Twitter data.
End Procedure
End Until
End

Here, the TF-IDF feature extraction technique helps in converting text data into numerical feature vectors that can be used for training an SVM model. SVM is a powerful machine learning algorithm that can classify the Stock Market Twitter data into positive or negative sentiment labels. Fine-tuning the SVM model helps in optimizing its performance on the testing data. Finally, the trained and fine-tuned SVM model can be used for sentiment prediction of new Twitter data.

4. Experiment Results

1. Precision

Dataset	LMI	CBSA	Proposed SVM-TF-IDF
50	66.45	74.12	87.76
100	69.78	71.89	90.89
150	74.91	67.35	92.41
200	79.33	68.98	95.56
250	86.86	65.33	97.12

Table 1.Comparison tale of Precision

The Comparison table 1 of Precision Values explains the different values of existing LMI, CBSA and proposed SVM-TF-IDF. While comparing the Existing algorithm and proposed SVM-TF-IDF, provides the better results. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and proposed SVM-TF-IDF values starts from 87.76 to 97.12. The proposed method provides the great results.

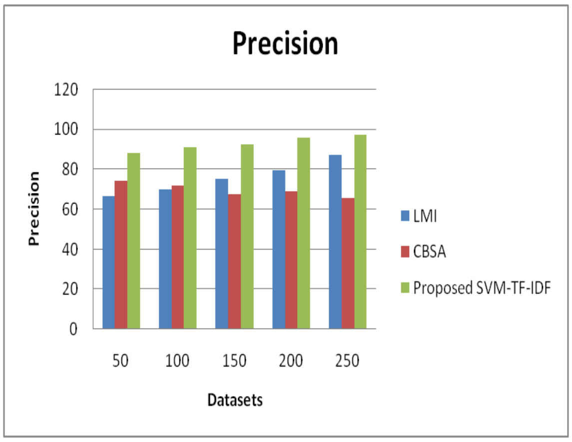

Figure 2.Comparison chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing CBSA, LMI and proposed SVM-TF-IDF. X axis denote the Dataset and y axis denotes the Precision ratio. The proposed SVM-TF-IDF values are better than the existing algorithm. The existing algorithm values start from 66.45 to 86.86, 65.33 to 74.12 and proposed SVM-TF-IDF values starts from 87.76 to 97.12. The proposed method provides the great results.

2. Recall

Dataset	LMI	CBSA	Proposed SVM-TF-IDF
20	0.62	0.72	0.83
40	0.66	0.65	0.87
60	0.70	0.59	0.90
80	0.72	0.62	0.94
100	0.75	0.59	0.96

Table 2.Comparison tale of Recall

The Comparison table 2 of Recall Values explains the different values of existing LMI, CBSA and proposed SVM-TF-IDF. While comparing the Existing algorithm and proposed SVM-TF-IDF provides the better results. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and proposed SVM-TF-IDF values starts from 0.83 to 0.96. The proposed method provides the great results.

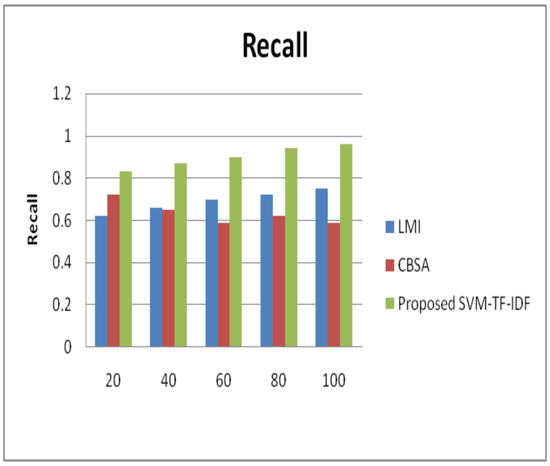

Figure 3.Comparison chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing CBSA, LMI and proposed SVM-TF-IDF. X axis denote the Dataset and y axis denotes the Recall ratio. The proposed SVM-TF-IDF values are better than the existing algorithm. The existing algorithm values start from 0.62 to 0.75, 0.59 to 0.72 and proposed SVM-TF-IDF values starts from 0.83 to 0.96. The proposed method provides the great results.

3. F-Measure

Dataset	LMI	CBSA	Proposed SVM-TF-IDF
100	0.89	0.72	0.98
200	0.85	0.70	0.96
300	0.86	0.67	0.95
400	0.84	0.64	0.93
500	0.82	0.61	0.92

Table 3.Comparison tale of F -Measure

The Comparison table 3 of F -Measure Values explains the different values of existing LMI, CBSA and proposed SVM-TF-IDF. While comparing the Existing algorithm and proposed SVM-TF-IDF, provides the better results. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed SVM-TF-IDF values starts from 0.92to 0.98. The proposed method provides the great results.

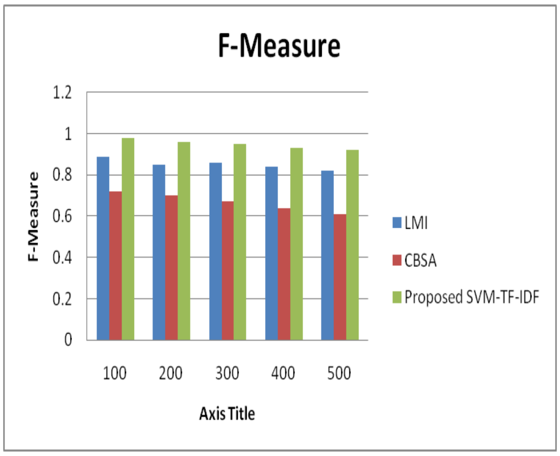

Figure 4.Comparison chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing CBSA, LMI and proposed SVM-TF-IDF. X axis denote the Dataset and y axis denotes the F -Measure ratio. The proposed SVM-TF-IDF values are better than the existing algorithm. The existing algorithm values start from 0.82 to 0.89, 0.61 to 0.72 and proposed SVM-TF-IDF values starts from 0.92to 0.98. The proposed method provides the great results.

4. Accuracy

Dataset	LMI	CBSA	Proposed SVM-TF-IDF
10	69	79	89
20	76	82	91
30	79	85	93
40	81	80	95
50	85	89	98

Table 4.Comparison tale of Accuracy

The Comparison table 4 of Accuracy Values explains the different values of existing LMI, CBSA and proposed SVM-TF-IDF. While comparing the Existing algorithm and proposed SVM-TF-IDF, provides the better results. The existing algorithm values start from 69 to 85, 79 to 89 and proposed SVM-TF-IDF values starts from 89 to 98. The proposed method provides the great results.

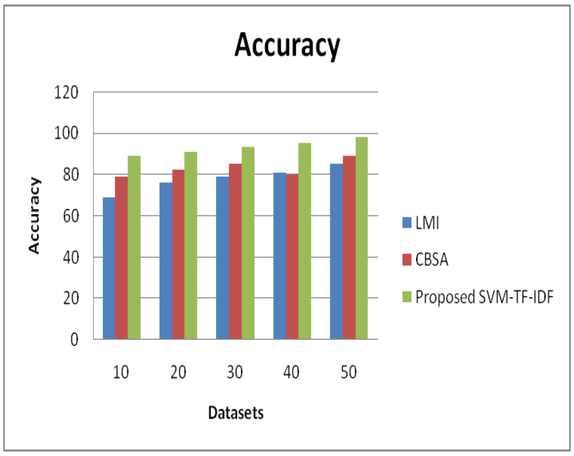

Figure 5.Comparison chart of Accuracy

The Figure 5 Shows the comparison chart of Accuracy demonstrates the existing CBSA, LMI and proposed SVM-TF-IDF. X axis denote the Dataset and y axis denotes the Efficiency Measure ratio. The proposed SVM-TF-IDF values are better than the existing algorithm. The existing algorithm values start from 69 to 85, 79 to 89 and proposed SVM-TF-IDF values starts from 89 to 98. The proposed method provides the great results.

5. Conclusion

In this paper tested the impact of combining several groups of features on the finance yahoo stocks and publicly available Twitter data. This research performs a point by point sentiment investigation of Stock tweets based on ordinal regression utilizing machine learning techniques. In conclusion, the use of SVM-TF-IDF for feature extraction in Twitter sentiment analysis has shown promising results. Through the pre-processing of Stock data and the application of SVM-TF-IDF, were able to extract important features that significantly contributed to the classification of tweets as positive, negative, or neutral. The Support Vector Machine (SVM) model performed well in classifying stock tweets with an overall accuracy. Moreover, this research has highlighted the potential of sentiment analysis in uncovering trends and patterns in Twitter data. By analyzing the sentiment of tweets can gain insights into the public perception of events, products, and services, among others.

References

1. D. Deepa, Raaji and A. Tamilarasi, "Sentiment Analysis using Feature Extraction and Dictionary-Based Approaches," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 786-790, doi: 10.1109/I-SMAC47947.2019.9032456.
2. F. Alqasemi, A. Abdelwahab and H. Abdelkader, "An enhanced feature extraction technique for improving sentiment analysis in Arabic language," 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016, pp. 380-384, doi: 10.1109/CIST.2016.7805075.
3. J. Yang and J. Yang, "Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks," 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), 2020, pp. 146-149, doi: 10.1109/ICSESS49938.2020.9237640.
4. M. H. Krishna, K. Rahamathulla and A. Akbar, "A feature based approach for sentiment analysis using SVM and coreference resolution," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 397-399, doi: 10.1109/ICICCT.2017.7975227.
5. F. R. Saputra Rangkuti, M. A. Fauzi, Y. A. Sari and E. D. L. Sari, "Sentiment Analysis on Movie Reviews Using Ensemble Features and Pearson Correlation Based Feature Selection," 2018 International Conference on Sustainable Information Engineering and Technology (SIET), 2018, pp. 88-91, doi: 10.1109/SIET.2018.8693211.

6. X. Fan, X. Li, F. Du, X. Li and M. Wei, "Apply word vectors for sentiment analysis of APP reviews," 2016 3rd International Conference on Systems and Informatics (ICSAI), 2016, pp. 1062-1066, doi: 10.1109/ICSAI.2016.7811108.
7. Ahuja R, Chug A, Kohli S, Gupta S, Ahuja P. The impact of features extraction on the sentiment analysis. Procedia Computer Science. 2019 Jan 1;152:341-8.
8. Mandloi L, Patel R. Twitter sentiments analysis using machine learninig methods. In2020 International Conference for Emerging Technology (INCET) 2020 Jun 5 (pp. 1-5). IEEE.
9. Fouad MM, Gharib TF, Mashat AS. Efficient twitter sentiment analysis system with feature selection and classifier ensemble. InInternational conference on advanced machine learning technologies and applications 2018 Feb 22 (pp. 516-527). Springer, Cham.
10. A. Razzaq et al., "Text sentiment analysis using frequency-based vigorous features," in China Communications, vol. 16, no. 12, pp. 145-153, Dec. 2019, doi: 10.23919/JCC.2019.12.011.
11. D. Deepa, Raaji and A. Tamilarasi, "Sentiment Analysis using Feature Extraction and Dictionary-Based Approaches," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 786-790, doi: 10.1109/I-SMAC47947.2019.9032456.
12. L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learninig Methods," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154183.
13. V. Ramanathan and T. Meyyappan, "Twitter Text Mining for Sentiment Analysis on People’s Feedback about Oman Tourism," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1-5, doi: 10.1109/ICBDSC.2019.8645596.
14. S. Singh, K. Kumar and B. Kumar, "Sentiment Analysis of Twitter Data Using SVM-TF-IDF and Machine Learning Techniques," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 2022, pp. 252-255, doi: 10.1109/COM-IT-CON54601.2022.9850477.
15. E. S. Usop, R. R. Isnanto and R. Kusumaningrum, "Part of speech features for sentiment classification based on Latent Dirichlet Allocation," 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2017, pp. 31-34, doi: 10.1109/ICITACEE.2017.8257670.
16. R. I. Permatasari, M. A. Fauzi, P. P. Adikara and E. D. L. Sari, "Twitter Sentiment Analysis of Movie Reviews using Ensemble Features Based Naïve Bayes," 2018 International Conference on Sustainable Information Engineering and Technology (SIET), 2018, pp. 92-95, doi: 10.1109/SIET.2018.8693195.
17. W. P. Ramadhan, S. T. M. T. Astri Novianty and S. T. M. T. Casi Setianingsih, "Sentiment analysis using multinomial logistic regression," 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), 2017, pp. 46-49, doi: 10.1109/ICCEREC.2017.8226700.
18. P. Barnaghi, P. Ghaffari and J. G. Breslin, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment," 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService), 2016, pp. 52-57, doi: 10.1109/BigDataService.2016.36.
19. M. Bouazizi and T. Ohtsuki, "Opinion mining in Twitter: How to make use of sarcasm to enhance sentiment analysis," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2015, pp. 1594-1597, doi: 10.1145/2808797.2809350.
20. U. Makhmudah, S. Bukhori, J. A. Putra and B. A. B. Yudha, "Sentiment Analysis Of Indonesian Homosexual Tweets Using Support Vector Machine Method," 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2019, pp. 183-186, doi: 10.1109/ICOMITEE.2019.8920940.