International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

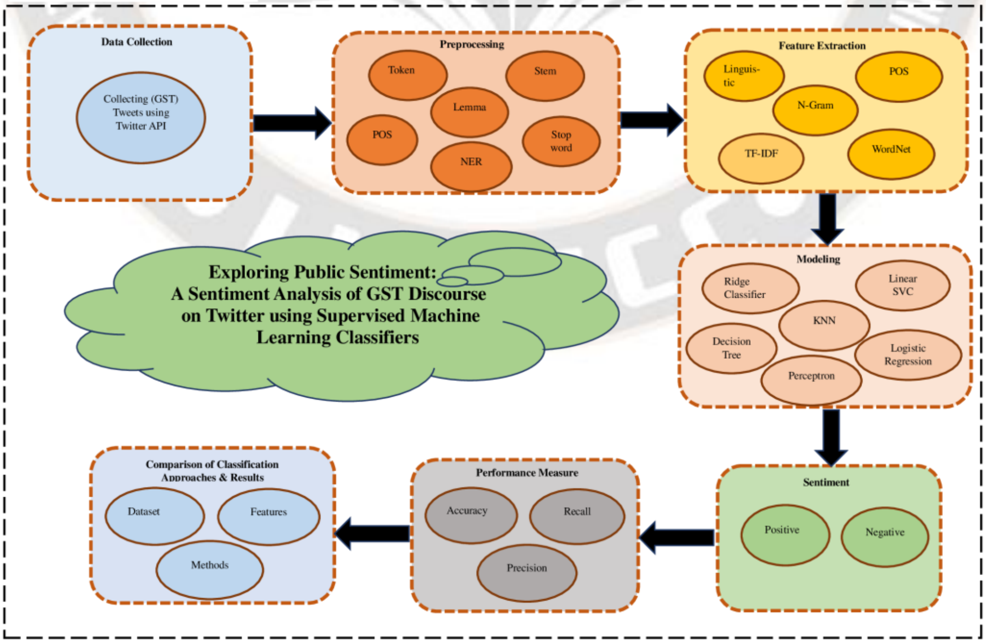
Exploring Public Sentiment: A Sentiment Analysis of GST Discourse on Twitter using Supervised Machine Learning Classifiers

Anima Srivastava[1], Amit Srivastava[2], Tanveer J. Siddiqui [3]
Department of Electronics and Communication, University of Allahabad, Prayagraj INDIA
anima24.jk@gmail.com[1],aksrivastava@allduniv.ac.in[2],tjsiddiqui@allduniv.ac.in[3]

Abstract— A key economic move that resulted in heated disputes was India's introduction of the Goods and Services Tax (GST). Social media channels offered a widely used forum for the people to express their views on the GST, providing insightful data for gauging mood and guiding next revisions. The emotion of 5629 GST-related tweets was assessed using the VADER lexicon after being obtained using the Twitter Developer API. The tf-idf feature was used for text vectorization, with 80% of the data going toward training and the remaining 20% going toward testing. In this study, six well-known classifiers—the Ridge Classifier, Logistic Regression, Linear SVC, Perceptron, Decision Tree, and K-Nearest Neighbor—were thoroughly compared to evaluate their performance in a range of circumstances. Accuracy, precision, recall, f-score, training, and testing times were all included in the performance measurements. The study presented novel pre-processing methods and examined the training/testing times before coming to the conclusion that the Ridge Classifier performed better than the others in terms of accuracy, precision, and efficiency. In this study, six well-known classifiers—the Ridge Classifier, Logistic Regression, Linear SVC, Perceptron, Decision Tree, and K-Nearest Neighbor—were thoroughly compared to evaluate their performance in a range of circumstances. Accuracy, precision, recall, f-score, training, and testing times were all included in the performance measurements. The study presented novel pre-processing methods and examined the training/testing times before coming to the conclusion that the Ridge Classifier performed better than the others in terms of accuracy, precision, and efficiency.

Keywords- GST-tweets, Sentiment Analysis, Machine Learning, Ridge Classification.

GRAPHICAL ABSTRACT



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

I. INTRODUCTION

Sentiment Analysis is the process of analysing text automatically to determine people's feelings [1] sentiments, attitudes, and emotions towards certain products, services, events, organizations, individuals, etc. Nowadays, social media websites have become a hub of opinionated content [2][3]. According to the statistics published on statista.com,[1] the number of social network users in India in 2016 was 168 million and the prediction is that it will reach 258 million in 2019[1]. People share their thoughts, experiences, views, and emotions on these websites on all kinds of topics on a regular basis. The opinions expressed on these websites provide valuable feedback on products, policies, services, movies, individuals, etc. [4]. This information is quite useful for companies, service providers, individuals, policymakers, government, political parties, and celebrities. However, analysing this huge volume of opinionated content manually is a herculean task. This has made automatic sentiment analysis or opinion mining a hot topic of research. Both machine learning and knowledge-based approaches have been used to automatically analyse textual data to know its polarity [5][6]. This paper focuses on the analysis of tweet data related to GST (Goods and Service Tax) to identify the polarity of the sentiments expressed in it.

Figure 1 shows the mechanism of the GST levied. The GST mitigates the inadequacy of indirect taxes and improves tax compliance which in turn reduces the heavy taxes imposed on end customers by its cascading effect. Consequently, the GST is levied on manufacturers, wholesalers, retailers, and consumers out of which only the consumer has to pay 6% in GST and the rest of the lot claims it back. The GST is a single tax [7] that replaces all indirect taxes charged by the central and state [2]government of India (GST Council, 2019) [8]. GST was levied on manufacturers, wholesalers, retailers, and consumers (Figure 1). It aims to combat the inadequacies of indirect tax and to improve tax compliance. However, the induction of GST invited a lot of criticism from a section of society that blame GST for the slowdown in the economy. Consequently, certain reforms have been made and the government is open to future reforms. Knowing the sentiments of the general public may be of great interest to the government in shaping future reforms. Some earlier works reported on GST sentiment analysis [9][10].

[1] https://www.statista.com/statistics/278407/number-of-social-network-users-in-india/

[2] http://gstcouncil.gov.in/

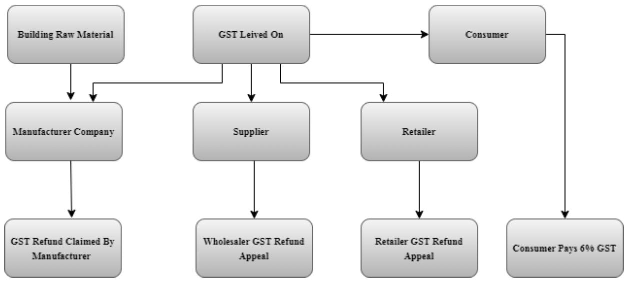


Figure 1. Mechanism of GST

II. RELEVANT WORK

Earlier work on GST sentiment analysis includes [11, 12, 13, 14, and 15]. Ganguly and Roy [13] analyzed opinions expressed about GST using Twitter data. They have collected all the tweets and re-tweets about GST from the day of its announcement till one day later (July 1st, 2017, to July 2nd 2017). The sentiment polarity is computed using the method presented in (Barnaghi et al. 2016) [16]. A cut-off of 0.25 is used to categorize tweets as positive or negative. Tweets with a polarity score of less than 0.25 are considered negative. The study also investigates the social connection among users who have expressed their opinions by building a directed graph based on the data collected. In this graph, nodes correspond to users and a connection between two nodes tells those users have responded or retweeted posts. The clustering coefficient and length of the average path in the resulting network were found to be 0.103 and 1.109 respectively, indicating that most nodes are not connected but are closer together. According to the polarity analysis, 38 percent of people support GST and 62 percent oppose it.

Gautam & Yadav, [14] used WordNet-based semantic analysis [17] to improve the results of the supervised classifier. To classify product reviews, they used three distinct classifiers: Support Vector Machine, Maximum Entropy, and Naive Bayes. The maximum accuracy was observed using the NB classifier. The output of the NB classifier was then used to label semantically related words as positive and negative. The semantic relatedness was derived using WordNet. The effectiveness of the classifiers was measured based on accuracy, precision, and recall.

Tomar et al. [15] used SVM to classify GST tweets. They experimented with two different models. The first model was trained on the IMDB dataset whereas the second model was developed primarily using a combination of datasets composed of the IMDB dataset, manually annotated tweets on GST. Both models were tested on the GST dataset collected from Twitter. They reported an accuracy of 73.28% using model two (IMDB+ domain-specific dataset). Implementation is done by using a modern open-source data platform Waikato Environment for Knowledge Analysis (WEKA) [18].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

By combining manually annotated GST-related tweets with the IMDB dataset's labelled reviews, domain and time-specific characteristics were used in the training dataset. They used two different models and evaluated accuracy, precision, recall, and f1-score. The model-1 was trained on the IMDB movie review dataset and tested on GST-related tweets. The model-2 was developed and validated using the IMDB dataset and Twitter dataset. GST-related tweets collected from Twitter microblogs.

Das & Kolya [11] used the NB classifier to tag tweets into one of the five categories: most positive, positive, normal, negative, and most negative. Emojis were also considered in the sentimental rating generation. They collected approximately 30,000 tweets from Twitter Streaming API and analyzed people's opinions about GST using the Naïve Bayes algorithm. The dataset comprises of 10 days tweets on GST during the implementation phase of GST in India. The sentiment rating for each of the five categories is reported on a 10-point scale (1 to 10).

Chaudhary and Paulose [12] proposed a new opinion-mining method and model using Stanford CoreNLP, on newspaper headlines [3] .Three different variants of support vector classification classifiers were used namely linear SVM, TF-IDF + linear SVM, and Stochastic Gradient Descent (SGD). They evaluate the performance of three different models: Model A, Model Band Model C. Model B with bigram feature secure (91.52%) highest accuracy among all models used.

III. MATERIALS & METHODS

We compare the performance of six different supervised classifiers on GST tweet data using tf-idf. The dataset consists of 5629 tweets. The best-performing case is compared with existing works on the GST dataset.

A. Machine Learning Classifiers

The objective of the machine learning classifier is to refine a methodology that enhances model performance using conventional training data. Various supervised machine learning models were trained to assess sentiment in a Twitter microblog [19] dataset related to GST. Subsequently, these machine learning algorithms underwent testing to evaluate their accuracy and precision in making predictions. The existing machine learning classifiers considered in previous research encompass Logistic Regression, Perceptron, Decision Tree, Linear Support Vector Classifier, K-Nearest Neighbor, along with the proposed approach with Ridge Classifier.

[3] http://www.indianexpress.com

1. Logistic Regression

One kind of analysis is logistic regression which is used to classify data and to figure out how different independent variables interact. It is a probabilistic classifier and uses a logistic function to model the probability that describes the possible outcome of a single trial. It works when the assumed variable is dual (binary two class- 0 or 1 classification), free from missing values and all predictors are independent of each other.

The outcome of logistic regression is determined by taking the event's log odds in (P/1P), where P is the probability of the event. As a result, P is always between 0 and 1. The equation (1) of logistic regression says that to find P, the exponential of a+bx is added to one (1) and is branched out with the exponential of a+bx.

$$P = \frac{1}{1 + e^{-(a+bX)}} \quad (1)$$

2. Perceptron

The Perceptron, an algorithm for linear classification suggests that it learns a decision boundary that splits two classes using a feature space line called hyper plane. As a result, it works well for situations where the classes can be efficiently divided by a line or linear model, referred to as linearly distinguishable problems. The model's coefficients, or input weights, are trained using the stochastic gradient descent optimization technique. For classification in binary format with two classes, the Perceptron method is machine-learning strategy. It is a member of a group of neural network models, arguably the most fundamental. It is composed of an individual node or neuron that determines the class from a sequence of incoming inputs. This is accomplished by calculating a bias and the weighted total of the inputs (set to 1). The model's activation is the weighted sum of its input as given in equation (2).

$$Activation = (Weights * Inputs) + Bias \quad (2)$$

The pseudo code represents the Perceptron binary prediction 0 and 1.

Pseudo code

IF (Activation > 0.0)
THEN Predict 1
ELSE IF (Activation <=0.0)
THEN Predict 0

3. Decision Tree

The decision tree is considered among the most influential approach for supervised classes of machine learning. It is simple to understand and comprehend. It can be used for both categorical and numerical data. The output of the decision tree

ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

is expressed as a sequence of rules which are used for classification tasks. Sometimes, DT learning can produce a complex tree that does not generalize well. DTs can be unbalanced because little dissimilarity in the data might generate a completely different tree. The decision tree learning algorithm uses a measure called information gain to build a decision tree. Knowledge improvement is estimated in terms of entropy of the initial set and the split obtained after testing an attribute. The entropy of a sample S is mathematically defined in equation (3).

$$E(S) = \sum_{i=1}^{c} -p_i log_2\, p_i \qquad (3)$$

4. Linear Support Vector Classifier

A Linear Support Vector Classifier's goal is to categorize or split the data provided by returning the "best fit" hyperplane. We can then add specific characteristics to the classifier to get the "predicted" class after acquiring the hyperplane. The LSVC uses a linear kernel function to conduct classification and does well with a lot of samples. The LSVC includes more parameters as compared to the SVC model, such as the loss function and penalty normalization, which applies "L1" or "L2". The kernel method cannot be changed since LSVC depends on the kernel linear methodology.

5. K-NN

The supervised learning method serves as the foundation for the K-Nearest Neighbour algorithm. The K-NN algorithm operates under the presumption that similar items exist nearby. Because of this, the K-NN method uses attribute resemblance among additional data points and points in the training set (existing cases) to forecast the value of the target data points. In general, the K-NN approach determines the value of the most recent data point by comparing it to the values in the training dataset. Although the K-NN technique is applicable to both regression and classification issues, it is most frequently used for classification issues.

6. Ridge Classifier

The proposed approach with ridge classifier is described here briefly. In machine learning, ridge classification is an algorithm used to perform the analysis of linear discriminant models. It's based on the Ridge regression technique which converts the label data into the range -1 to 1 and solves with regression process. A modification of linear regression called ridge regression modifies the loss function to simplify the model. The ridge classifier is a technique for evaluating multiple regression data with multi-co linearity. Although least squares forecasts are unbiased in the context of multi-co linearity, their wide variances render them possibly erroneous. In order to reduce the standard errors, the ridge classifier slightly slants the regression estimates. This method is used when the independent variables are significantly linked. L2 regularization is carried out, and it entails a penalty proportional to the square root of the size of the coefficients as given in equation (4).

$$\text{Minimization goal} = \text{LS Oj} + * (\text{Sum of square of coefficients}) \ (4)$$

This change involves the addition of a compensation component equal to the square of the magnitude of the coefficients. The loss function is determined by adding ordinary least squares (OLS) and alpha (squared coefficient values). We must choose alpha as the parameter in the loss function shown above. Low alpha values can lead to over-fitting while high alpha values may lead to under-fitting. Scikit Learn's Ridge class is used to create a ridge regression model. To reduce the subsequent cost function, use the formula of equation (5).

$$(y - X\beta)^T (y - X\beta) + \lambda\beta^T\beta \qquad (5)$$

λ is a value given by user input (or by a grid search, or whatever). Note that here we use λ, sci-kit-learn uses α. β is a vector of weights, βi, assigned to each of the features to produce a finished model.

B. Methodology

In our methodology, we describe the methodical process used to analyze sentiment in GST-related tweets and assess the effectiveness of six well-known classifiers. This procedure includes data collection, sentiment analysis, text vectorization, dataset splitting, classifier comparison, and evaluation of performance indicators.

[1] Dataset Preparation

We collected Twitter messages discussing Goods and Service Tax (GST) or all in one tax via the streaming API in keyword tracking mode using python client Tweepy. The keywords used are: #gst, #CGST, #SGST, #gst tax, #gstbenefits, #onenationonetax, #dualgst. We dropped non-English words occurring in these tweets. Only micro blog messages in English were retained. The data thus obtained contains re-tweets as well. This increases the size of the data but no new information. Therefore, we remove all duplicate re-tweets. We obtained 500 KB tax.csv file comprising of 5629 tweets.

[2] Recommended Approach

The classification tasks encompass various undertakings involving machine learning classifiers for sentiment analysis of recent tax measures (such as GST), utilizing term-frequency

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

and inverse-document frequency techniques. The following are steps followed for recommended approach:

i. Data Collection and Pre-processing:

- Gather Twitter text data from Twitter API.
- In pre-processing eliminating numbers, special characters, and punctuation, also converting the tweets or text data to lowercase.
- Tokenize the text into words and remove stop words.

ii. Evaluate Sentiment:

- Import SentimentIntensityAnalyzer from Vader lexicon of nltk. sentiment to evaluate sentiment score. [20]
- On the basis of sentiment score assign sentiment labels (positive or negative) for each tweet.
- TextBlob lexicon of nltk can also be used where sentiment polarity is used for assigning sentiment labels for each tweet.

iii. Feature Extraction:

- Utilize TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to transform textual data into numerical feature vectors.
- TF-IDF assigns values to words by considering their occurrence frequency within a document and their significance across the entire corpus.

iv. Data Splitting:

- Divide the dataset into sets for training and testing. 75% are selected for training and 25% for testing.

v. Classifier Training & Prediction:

- Pipelined each classifier (Ridge Classifier (RC), Logistic Regression (LR), Linear SVC (LSVC), Perceptron (P), K-Nearest Neighbors (KNN), & Decision Tree (DT)) with tf-idf feature which helps to ensure that these steps are executed in a consistent and organized manner, making it easier to manage the entire workflow from model training to prediction.
- Predict sentiment labels for the testing data using the trained classifier.

vi. Performance Evaluation:

- Assess the performance using the subsequent metrics: accuracy, precision, recall, F1-score, as well as training and testing time.

IV. EXPERIMENT AND RESULT ANALYSIS

In this experiment a train-test split on the data frame's X and Y components. GST Twitter dataset, as explained in the section data preparation section, and the train test split () method were used to divide our data into train and test sets for each of the six classifiers. A set of data was used to fit the model. The training dataset is what it is called the data set that the model was trained on. The model notices and takes note of this information. Our data must first be divided into features (X) and labels (y) for analysis. The data frame's components are divided into X trains, X tests, and Y trains, Y tests. The model is fitted and trained using X and Y train sets. The model is assessed using the X test and y test sets to see if it correctly predicts the outcomes and labels. The train and test sets' dimensions can be directly tested. The test sets should be less extensive than the training ones. In our study, 25% of the data were used for testing, and 75% were used for training tests.

i. Assessment of the machine learning classifier's parameters

There are certain measuring parameters that are used to evaluate the performance of various machine learning models are accuracy, precision, recall, f1-score, training and testing time are the measurement parameters employed in the suggested study.

The ability to quantify how accurately a machine learning algorithm predicts outcomes is known as accuracy. This refers to the proportion of correctly predicted observations out of the total observations. The formula for the accuracy is given in equation (6) as:

$$\text{Accuracy (A)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

Precision measures of how accurate a classifier. In order to calculate precision, divide the overall number of positive predictions (P) by the proportion of correct positive predictions. The highest precision is 1.0, while the lowest is 0.0. The precision formula in equation (7) is as follows:

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

Recall indicates the classifier's ability to correctly classify positive samples. A higher recall value suggests that the classifier predicts more positive samples. It is determined as the sum of the true positive samples and the false negative samples divided by the true positive samples. The true positive rate, or recall (R), is also known as True positive rate (TPR). The Recall is calculated using the formula in equation (8) as follows:

$$Recall(R) = \frac{Rate}{Recall} = \frac{TP}{TP + FN} \tag{8}$$

The F-score is commonly known as the F-Measure or F-1Score. It's the harmonic mean of the precision and the recall. Its formula is given in equation (9) as follows:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

$$F - Measure(F1) = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (9)$$

Where FP is the overall number of incorrectly made positive predictions, FN represents the overall number of incorrect negative predictions, TP stands for the total number of correctly predicted positive outcomes, and TN denotes the overall number of accurately predicted negative outcomes. Each classifier was tested using tf-idf and bi-gram. With GST dataset the best result was obtained using tf-idf. The best performing results of existing work on same dataset in Table-II and proposed classifier algorithm is reported in Table I. The highest accuracy of 96% was obtained using our proposed algorithm Ridge classifier with our dataset. Figure 2 shows comparison graph of recall of proposed classifier with existing classifier. Figure 3 depicts the comparison graph of precision for proposed and existing classifier algorithms

Table I. Performance Evaluation of LSVC, LR, P, KNN, DT, RC

Sl. No.	Classifier	Accuracy	Precision	Recall	F-1 Score
1.	Linear SVC	0.92	0.90	0.95	0.94
2.	Logistic Regression	0.90	0.71	0.94	0.83
3.	Perceptron	0.89	0.94	0.92	0.95
4.	K-Nearest Neighbor (K-NN)	0.80	0.82	0.83	0.88
5.	Decision Tree (DT)	0.76	0.89	0.80	0.91
6.	**Ridge Classifier**	**0.96**	**0.97**	**0.98**	**0.94**

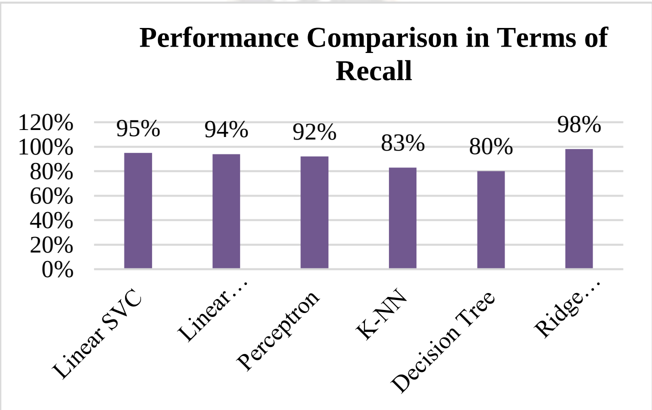


Figure 2. Performance Comparison in Terms of Recall

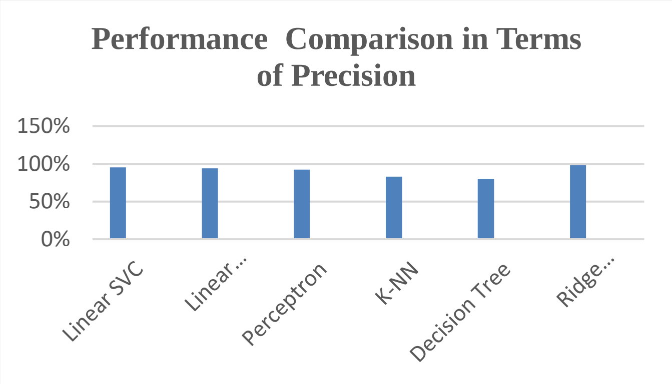


Figure 3. Performance Comparison in Terms of Precision

The classifiers exhibit varying levels of novelty and efficiency. The Ridge Classifier stands out with a high accuracy of 96% shown in Figure 4, coupled with rapid training (0.0553 seconds) and testing times (0.0037 seconds), making it suitable for accurate and efficient predictions. Training and testing time comparison for all classifiers are visualized in Figure 4, Logistic Regression offers a balanced approach, with moderate training (0.7005 seconds) and quick testing (0.0039 seconds). The Perceptron demonstrates simplicity and efficiency, achieving 89% accuracy with fast training (0.2244 seconds) and testing (0.0053 seconds). LSVC maintains a balance with an accuracy of 92%, showing moderate training (0.6081 seconds) and efficient testing (0.0076 seconds). Decision Trees provide interpretability, though at the cost of longer training (0.9985 seconds) and testing times (0.1002 seconds) for 76% accuracy. KNN's adaptability yields 80% accuracy, yet requires longer training (10.3873 seconds) and testing times (4.0020 seconds) for capturing local patterns. The choice depends on accuracy, efficiency, and interpretability needs.

In our proposed approach, the Ridge Classifier (RC) significantly outperformed existing methods, demonstrating improved accuracy, precision, recall, and notably, the shortest training and testing times among the five machine learning algorithms considered.

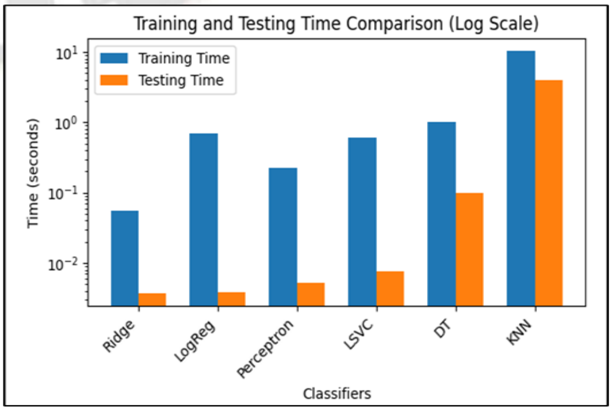


Figure 4. Training and Testing Time Comparison for RC, LR, P, LSVC, DT, KNN

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8179
***Article Received:** 25 June 2023 **Revised:** 26 August 2023 **Accepted:** 05 September 2023*

Table II. Comparison of existing work on Twitter GST data with Machine Learning Classifiers

S. No.	Author	Methods	Features	Dataset	Accuracy	Precision	Recall
1.	Das & Kolya [11]	NB	Particular feature from the class of features	Twitter API(GST)	Calculate Range, TF, CF, Zipf Score		
2.	Ganguly& Roy [13]	Polarity Method	Structural Features	Twitter API(GST)	38% +ve, 62% -ve	---------	------
3.	Gautam & Yadav [14]	**NB***, ME, SVM	Unigram Feature Technique, **WordNet***	Product Review	89.9% (WordNet)	88.3%	44.3%
4.	Tomer et. al. [15]	SVM	Linguistic Based Feature	IMDB+(GST) Twitter*, IMDB	73.28%	73.09%	73.67%
5.	Go et al. [21]	NB, **ME***, SVM	Unigram, Bigram, **Unigram Bigram***, Unigram+ POS	Twitter API (Product/ Brand)	83.0%	---------	------
6.	Our Proposed Approach	**RC***,LR,P, LSVC,DT,KNN	**TF-IDF*,** Bi-gram	Twitter API(GST, All-in-one Tax)	96.0%	97.0%	98.0%

V. CONCLUSION AND FUTURE SCOPE

The understanding of the public opinion on the new taxation system is crucial for shaping future reforms. The classifiers that were assessed produced distinct performance traits. The 'Ridge' classifier won the competition with a decent precision of 0.90 and a stunning accuracy of 0.96. 'LSVC' and 'Perceptron' also demonstrated competitive accuracy and precision, emphasizing their efficacy. As opposed to 'DT', which lagged behind with an accuracy of 0.76, 'LogReg' and 'KNN' both displayed respectable accuracy levels of 0.90 and 0.80, respectively. Notably, 'Ridge' demonstrated the quickest training and testing times, in contrast to 'KNN' and 'DT', which showcased relatively longer times. In the future, employing ensemble techniques and researching dimensionality reduction strategies could improve classifier performance even more while reducing the lengthier training times.

REFERENCES

[1] C. Bhadane, H. Dalal, and H. Doshi, “Sentiment analysis: Measuring opinions”, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), Procedia Computer Science. Elsevier, 45: 808 – 814, (2015).https://doi.org/10.1016/j.procs.2015.03.159

[2] T. J. Siddiqui, “Utilizing sentiments in online contextual advertising. In online multimedia advertising: Techniques and Technologies”, IGI Global, 32-37, (2011). https://doi.org/10.4018/978-1-60960-189-8.ch003

[3] S. K. Singh, and M.K. Sachan, “SentiVerb system: Classification of social media text using sentiment analysis”, Multimedia Tools and Applications India, 78(22), 32109–32136, (2019).doi. /10.1007/s11042-019-07995

[4] A. Pak, and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining”, Proceedings of the 17th conference on International Language Resources and Evaluation (LREC’10 European Language Resources Association (ELRA), Valletta, Malta. 10, No. 2010, 1320-1326, (2010). https://lexitron.nectec.or.th/public/LREC-2010_Malta

[5] B. Liu, “Sentiment analysis and opinion mining”, Human Language Technologies (Synthesis Lectures). Morgan & Claypool Publishers, 5(1):1-67, (2012). https://www.cs.uic.edu/~liub/FBS/liub-SA-and-OM-book.pdf

[6] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques”, In Empirical methods in natural language processing EMNLP ’02. ACL, arXiv preprint cs/0205070, (2002). https://doi.org/10.48550/arXiv cs/0205070

[7] Nguyen Thanh Tung, Luong Van Van. (2023). Effect of Braking Force on Wheel Load and Braking Efficiency of Tractor Semi-Trailer on A Roundabout Using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 428–433. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2689

[8] A. K. Deshmukh, A. Mohan, and I. Mohan, “Goods and services tax (GST) implementation in India: An SAP–LAP–twitter analytic perspective”, Global Journal of Flexible Systems Management, Springer, 23, 165-183, (2022). https://doi.org/10.1007/s40171-021-00297-3

[9] https://doi.org/10.1145/1656274.1656278http://gstcouncil.gov.in/ sites/default/files/01012019-GST-Concept-and-Status.pdf

[10] S. Cnossen, “Preparing the way for a modern GST in India”, International Tax and Public Finance, Springer Science Business Media New York. 20, 715-723, (2013). https://doi.org/10.1007/s10797-013-9281-0

[11] A. Madan, R. Arora, and N.R. Roy, “Sentiment analysis of Indians on GST”, International conference on recent developments in science, engineering and technology REDSET 2017: Gurgaon, India. Revised Selected Papers, Springer Singapore. 4, 568-575, (2017). https://doi.org/10.1007/978-981-10-8527-7_47

[12] S. Das, and A.K. Kolya, “Sense GST: Text mining & sentiment analysis of GST tweets by naive bayes algorithm”, Third international conference on research in computational

intelligence and communication network (ICRCICN). IEEE, 239-244, (2017). https://doi.org/10.1109/ICRCICN.2017.8234513

[13] J.R. Chaudhary, and J. Paulose, "Opinion mining on newspaper headlines using SVM and NLP", International Journal of Electrical and Computer Engineering (IJECE), 9 (3), 2152-2163, (2019).https://doi.org/ 10.11591/ijece. v9i3.pp2152-2163

[14] Dr. M. Varadharaj. (2019). Density Based Traffic Control System with Smart Sensing Of Emergency Vehicles. International Journal of New Practices in Management and Engineering, 8(02), 01 - 07. https://doi.org/10.17762/ijnpme.v8i02.75

[15] M. Ganguly, and S. Roy, "A social network analysis of opinions on GST in India within twitter", In Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking, Varanasi, India. Association for Computing Machinery, ACM, 1-2, (2018). https://doi.org/10.1145/3170521.3170539

[16] G. Gautam, and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis", In Proceedings of Seventh International Conference on Contemporary Computing (IC3), IEEE, 437-442,(2014).https://doi.org/10.1109/IC3.2014.6897213

[17] N. Tomar, R. Srivastava, and B. Ahuja, "Opinion mining of GST implementation using supervised machine learning", International Journal of Computer Applications, 180, 1-7, (2018). https://doi.org/ 10.5120/ijca2018917283

[18] P. Barnaghi, P. Ghaffari, and J.G. Breslin, "Opinion mining and sentiment polarity on twitter and correlation between events and sentiment", Second International Conference on Big Data Computing Service and Applications, (Big Data Service), IEEE. 52-57, (2016). https://doi.org/10.1109/BigDataService.2016.36

[19] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), arXiv preprint cs/0212032, (2002). https://doi.org/10.48550/arXiv.cs/0212032

[20] Waheeb , M. Q. ., SANGEETHA, D., & Raj , R. . (2021). Detection of Various Plant Disease Stages and Its Prevention Method Based on Deep Learning Technique. Research Journal of Computer Systems and Engineering, 2(2), 33:37. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/30

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, "The WEKA data mining software: an update", ACM SIGKDD Explorations newsletter.11 (1), 10-18, (2009).

[22] M.D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A comparative study on different approaches", Procedia Computer Science, Elsevier, 87, 44–49, (2016). https://doi.org/10.1016/j.procs.2016.05.124.

[23] S. Bird, E. Klein, and E. Loper, "Natural language processing with python – analyzing text with the natural language toolkit", O'Reilly Media, Inc. (2009), https://www.nltk.org/book/

[24] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford,1(12), p.2009, (2009). https://nlp.stanford.edu/courses/cs224n/2009/fp/3.pdf