Tech Science Press

An Ensemble Based Approach for Sentiment Classification in Asian Regional Language

Mahesh B. Shelke[1], Jeong Gon Lee[2,*], Sovan Samanta[3], Sachin N. Deshmukh[1], G. Bhalke Daulappa[4], Rahul B. Mannade[5] and Arun Kumar Sivaraman[6]

[1]Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, 431004, India
[2]Division of Applied Mathematics, Wonkwang University, 460, Iksan-daero, Iksan-Si, Jeonbuk, 54538, Korea
[3]Department of Mathematics, Tamralipta Mahavidyalaya, Tamluk, West Bengal, 721636, India
[4]Department of Electronics and Telecommunication Engineering, AISSMSCOE, Pune, Maharashtra, 411001, India
[5]Department of Information Technology, Government College of Engineering, Aurangabad, Maharashtra, 431005, India
[6]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600127, India
*Corresponding Author: Jeong Gon Lee. Email: jukolee@wku.ac.kr

Abstract: In today's digital world, millions of individuals are linked to one another via the Internet and social media. This opens up new avenues for information exchange with others. Sentiment analysis (SA) has gotten a lot of attention during the last decade. We analyse the challenges of Sentiment Analysis (SA) in one of the Asian regional languages known as Marathi in this study by providing a benchmark setup in which we first produced an annotated dataset composed of Marathi text acquired from microblogging websites such as Twitter. We also choose domain experts to manually annotate Marathi microblogging posts with positive, negative, and neutral polarity. In addition, to show the efficient use of the annotated dataset, an ensemble-based model for sentiment analysis was created. In contrast to others machine learning classifier, we achieved better performance in terms of accuracy for ensemble classifier with 10-fold cross-validation (cv), outcomes as 97.77%, f-score is 97.89%.

1 Introduction

In this digital age, millions of people are connected to one another through Web 2.0 and social networking. This allows for a new technique of exchanging knowledge with other people. Social networking sites, e-commerce websites, blogging, and other similar platforms allow users to instantly generate creative content, thoughts, and opinions, leading in the development of huge amounts of data every day. Sentiment analysis and opinion mining have grown as a challenging and dynamic field of research for both resourced and under-resourced languages. The term sentiment refers to a broad concept that encompasses sentiment, evaluation, appraisal, or attitude toward a piece of information that demonstrates the author's point of view.

Opinion mining or emotional intelligence are terms used to describe sentiment analysis. Sentiment analysis is the systematic process of extracting useful knowledge from unstructured and unorganized text information in various social platforms and online sources, such as chats on Twitter, WhatsApp, and Facebook, as well as online blogs and comments. Opinion mining includes establishing automated systems that employ any of the machine learning methods to accomplish opinion mining.

The number of Marathi internet users and web content has grown tremendously. Because Marathi is still an under-resourced language in the field of sentiment analysis, there have been few attempts to perform SA in Marathi. Users express their opinions in a variety of methods, including bilingual text, transliterated words, emoticons, spelling variations, incorrect linguistic structures, and many others [1]. This makes sentiment analysis a difficult field for research, particularly with Indian languages. This allows for the development of Marathi resources and research in the field of sentiment analysis.

The major contributions of this research work are the development and evaluation of lexical resources for sentiment analysis in Marathi, because there are minor lexical resources, libraries, and lexical Corpus available for Marathi, indicating that Marathi has not been explored in the field of sentiment analysis. In this research, we present an ensemble-based model for predicting the sentiment of Marathi texts through integrating the output of Machine Learning-based models. And for developing benchmark dataset, we manually annotated the Twitter dataset with the help of human annotators (domain experts), who are senior researchers in Marathi, and for analysis of these annotators' performance, we used Fleiss' kappa as performance measurement matrices, and lastly, all classification algorithms are also evaluated and discussed. In addition, an annotated dataset of Marathi tweets with positive, negative, and neutral sentiment orientations was created.

2 Related Work

In recent years, only a few Indian languages have been studied, including Hindi, Telugu, Tamil, Telugu, and Bengali. However, as Indian people's digital literacy grows and technology becomes easier to utilize for creating content in Indian languages, countries like India will be capable of creating content in regional languages on the Internet.

Authors have proposed ensemble-based model sentiment analysis of Persian text [2–4]. Sentiment analysis is performed using deep learning and shallow approaches. In experimentation, achieved accuracy is up to 79.68% [5]. Researchers proposed an ensemble-based recommender system for hotel reviews and also categorized aspects [6]. And used ensemble of binary classification known as BERT technique, with features as Word2Vec, subjectivity score and Term Frequency-Inverse Document Frequency (TF-IDF), achieved performance of model with 84% f-score and 93.26% accuracy [7]. In proposed ensemble model for feature extraction author has considered Information Gain (IG), Gini Index and Chi Square. And used machine learning algorithms as Sequential Minimal Optimization (SMO), Multi-nominal Naïve Bayes (MNB), and Random Forest (RF) and considered multi-domain dataset.

The researchers studied the use of Naive Bayes (NB) and Support Vector Machine for machine learning-based sentiment classification of movie reviews (SVM) [8–12]. Sentiment Analysis is a two-class classification problem comprising Positive and Negative classes; this kind of study may be used to classify textual information and feature selection affects classifier performance.

The Authors have performed comparative performance weight of each binary classifier in the training sample set is computed for enhanced one-vs-one (OVO) technique based on the K nearest neighbours and the class centre of each category in the training sample set about the classification algorithm [13]. The information gain (IG) approach is used to identify the key features for multi-class sentiment analysis; a binary SVM classifier is then trained on feature extraction training of every pair of sentiment categories.

Ensemble approaches, as alternative to using each of the individual learning algorithms alone, employ many learning algorithms to achieve greater efficiency [14]. Deep learning techniques' performance can be improved by combining them with standard approaches based on manually acquired features [15].

Machine Learning based techniques has played a significant role in Natural Language Processing [16]. Machine learning techniques are divided into two learning classes as supervised and unsupervised learning. For task of Sentiment analysis mostly preferred supervised algorithms as Support Vector Machine (SVM), Maximum Entropy and Naïve Bayes (NB) [17–19]. It includes feature-based sentiment analysis and summarization.

3 System Development

This section describes corpus creation process, pre-processing, manual annotation, and performance evaluation of human annotator with the help of Fleiss's Kappa [20]. And proposed ensemble-based model for sentiment classification.

3.1 Corpus Creation

3.1.1 Corpus Acquisition

We have extracted publicly available Marathi Tweets from twitter with the help twitter-API. Initially, we have collected generalized 1493 Marathi Tweets.

3.1.2 Data Preprocessing

Initially, pre-processed the data into the necessary forms, for which following steps are carried out:

- Identified and eliminated duplicate and irrelevant tweets manually.
- Identified and transliterated English words present in tweets into Marathi manually.
- Removed stop words.
- Performed lemmatization to find root word.
- Removed any incorrect punctuation, smileys, hashtags, or photo tags.
- Removed complicated sentences since they are inappropriate for performing sentiment analysis.

3.1.3 Data Annotation

We chose three domain experts who are senior scholars with a Ph.D. in Marathi to do manual data annotation with the help of human annotators. We asked them to tag the Marathi Tweets dataset with 1, 0, and −1 to represent the positivity, neutrality, and negativity of Marathi tweets.

3.2 Feature Extraction

Supervised Machine learning methods generates output for test data by learning from a pre-defined set of features in the training samples [21]. As Machine learning methods cannot directly works on raw text, as result feature extraction methods are required to transfer text into a vector of features. In this research work we are considering unigram with Term Frequency–Inverse Document Frequency (TF-IDF) for feature extraction. Mostly, unigram i.e., single words hold important opinions, emotion [22]. For example, "Camera of this mobile is good", here word "good" expresses opinion about camera. So, it becomes important for to consider Unigram + TF-IDF model for feature extraction.

The unigram word vectors obtained during initial stage are used to build a matrix containing all of the tweets, and the unigrams recovered from the matrix are handled as features. The TF-IDF feature matrix is constructed with the features as columns and tweets as rows. The Lexical TF-IDF is calculated by multiplying each feature column of the TF-IDF feature matrix by its sentiment score. This matrix is used to train supervised machine learning algorithms.

3.3 Sentiment Classification Approach

To learn and classify, machine learning algorithms employ various series. The names of the input feature vectors and their classes are included in the training set. Using this training set, a classification model was created to classify the input material into positive and negative class [23]. Extracted feature sets are applied to train the classifier to evaluate if the data set review is positive or negative. Ensemble techniques are a type of machine learning methodology that integrates numerous base models to create a single best prediction model.

3.3.1 Logistic Regression (LR)

Logistic regression estimates probabilities using a logistic function, which is the cumulative logistic distribution, to assess the association between a categorical dependent variable and one or more independent variables [24–28]. Logistic regression is a linear approach; however, the logistic function is used to modify the predictions. It is a statistical technique for assessing a dataset that has one or more independent variables that affect the outcomes.

Instead of fitting a regression line, we fit a "S" shaped logistic function that predicts two maximum values in logistic regression (0 or 1). Logistic regression starts with a conventional linear regression and then adds a sigmoid to the linear regression result. Regression is expressed Eq. (1) and logistic function in Eq. (2).

$$= w_0x_0 + w_1x_1 + w_2x_2 + \cdots w_Nx_N \tag{1}$$

where, w_0 indicates weights and x1 represents independent variables.

$$h(z) = \frac{1}{1 + exp^{-z}} \tag{2}$$

3.3.2 Stochastic Gradient Decent (SGD)

Stochastic Gradient Descent (SGD) is a straightforward but highly efficient method for fitting linear classifiers and regressors to convex loss functions. SGD has been effectively used to large-scale and sparse machine learning applications, such as text categorization and NLP. Given the sparsity of the data, the classifiers in this module can efficiently scale to problems with more than training instances and more than features. The class SGD Classifier provides a simple stochastic gradient descent learning process that supports various classification loss functions and penalties. The decision boundary of an SGD Classifier trained with the hinged loss, which is comparable to a linear SVM.

3.3.3 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a well-known supervised machine learning model for categorization and prediction of different datasets. Several studies claim that SVM is a fairly accurate approach for text categorization. It is also often used in sentiment analysis.

For example, if we have a dataset with data that has been pre-labelled into two categories: positive and negative labels in Fig. 1, we may train a model to classify real time data into these two categories. This is precisely how SVM operates. We train the model on a dataset so that it can evaluate and classify unknown data into the categories that were present in the training set.

3.3.4 Naive Bayes (NB)

The Naive Bayes classifier is a prominent supervised classifier that allows you to express positive, negative, and neutral sentiments in content. To classify words into their respective categories, the Naive Bayes classifier employs conditional probability. The advantage of using Naive Bayes for text classification is that it just requires a minimal dataset for training. The raw data is pre-processed, with removal of stop words, punctuation marks, extra spaces, transliteration of other language words and

special symbols removed. Human annotator performs the manual tagging of words with labels of positive, negative, and neutral tags.
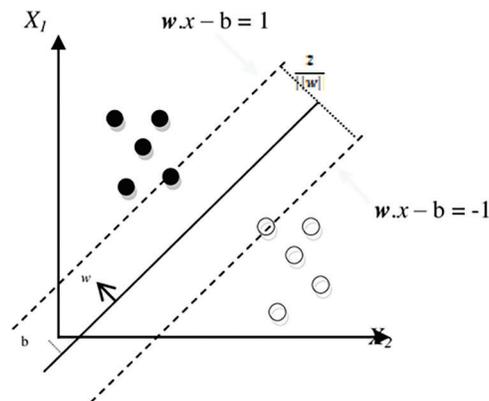

Figure 1: Optimal hyperplane

It can be beneficial for determining the likelihood of each statement using sentiment. In this technique, each attribute helps to selecting which labelling should be allocated to the emotion value of each phrase. The Naive Bayes classifier starts by computing the prior probability of each labelled sentence, which is derived by examining the occurrence of each labelled statement in the training data set. Following Eq. (3) describes bayes rule.

$$P\ (A|B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{3}$$

where, A is Particular class, B sentence which needs to be classified, $P(A)$ and $P(B)$ are Prior probabilities, and $P(A \mid B)$ and $P(B \mid A)$ are Posterior probabilities.

3.3.5 Nearest Neighbour

Nearest Neighbours (KNN) is an important classification technique in Machine Learning. It is a supervised learning algorithm that is widely used in text classification. It is extensively applicable in real-world circumstances since it is non-parametric, which means it makes no underlying assumptions regarding data distribution. We are provided some previous data (also known as training data) that classify locations into categories based on a characteristic.

3.3.6 Ensemble Classifier

The purpose of Ensemble techniques is to integrate the predictions of numerous base estimators with a specific learning algorithm to increase the classifier's accuracy and resilience. The idea behind the Voting Classifier is to integrate conceptually distinct machine learning classifiers and forecast the class labels using a majority vote or the average projected probability (soft vote). Such a classifier can be effective for balancing out the individual flaws of a set of similarly highly performing classifiers.

Fig. 2 shows An Ensemble based Sentiment classification approach using supervised Machine Learning algorithms. And Algorithms are Support Vector Machine (SVM), Nave Bayes (NB), k-Nearest Neighbour (KNN), Neural Network, Decision Tree (DT), Logistic Regression (LR), Stochastic Gradient Decent (SGD), and the proposed Ensemble-based Model are implemented in research work.

Algorithm: Sentiment analysis using proposed ensemble based algorithm

---

Input: An annotated Marathi Tweet Dataset

A list of tweets, **T** contains positive and negative sentences

$T = \{ T_{p1}, T_{P2}, \ldots\ldots, T_{pi}, T_{n1}, T_{n2}, \ldots\ldots, T_{ni}\}$

Where, T_{pi} is number of positive sentences, and T_{ni} is number of negative sentences

Pos_count, **Neg_count** is Positive count and negative count respectively.

Output: Pos_score$_i$, Neg_score$_i$ contains sentiment score

Sentiment_polarity_score(dataset)

    1. **for each** T_i, **in** T **do**

    2.   *Pos_count$_i$* = 0;

    3.   *Neg_count$_i$*= 0;

    4.   **for each** Classifier_C$_i$ **in** ensemble_classifier **do**

    5.   **if** C$_i$ predict_positive **then**

    6.   *Pos_count$_i$* += 1;

    7.   **else**

    8.   *Neg_count$_i$* += 1;

    9.   **[if-else end]**

    10.   **[for end]**

    11. **[for end]**

    12. **for each** Classifier_C$_i$ **in** ensemble_classifier **do**

    13.   Weight$_{Ci}$ = accuracy_C$_i$ / (Sum of all learning

        classifiers in ensemble classifier)

    14. **[for end]**

    15. **for each** T_i, **in** T **do**

    16.   *Pos_count$_i$* = 0;

    17.   *Neg_count$_i$*= 0;

    18.   **for each** Classifier_C$_i$ **in** ensemble_classifier **do**

    19.   **if** C$_i$ predict_positive **then**

    20.   *Pos_score$_i$* += Weight$_{Ci}$ * prob(pos$_i$);

    21.   **else**

    22.   *Neg_score$_i$* += Weight$_{Ci}$ * prob(neg$_i$);

(Continued)

Algorithm: (continued)

23.   **[if-else end]**

24.   **[for end]**

25.   **return** Pos_score_i, Neg_score_i

26. **[for end]**

Calculating Probability:

$$prob(pos_i) = \frac{Pos\_count_i}{Pos\_count_i + Neg\_count_i}$$

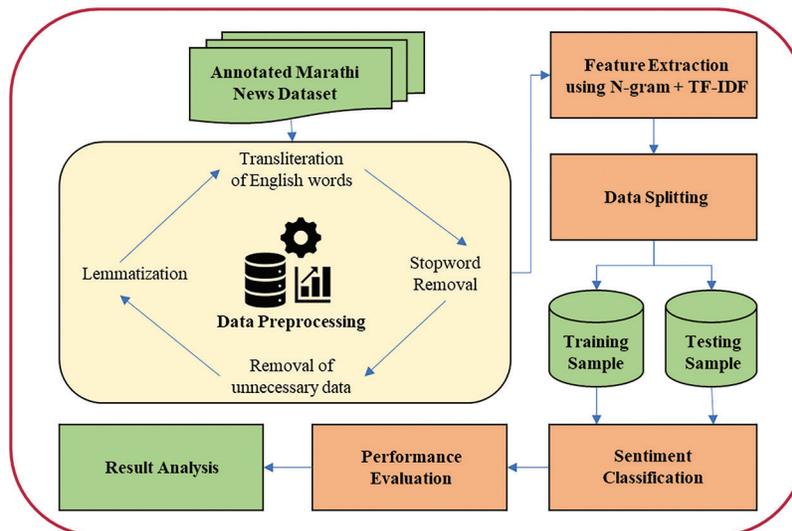$$prob(neg_i) = \frac{Neg\_count_i}{Pos\_count_i + Neg\_count_i}$$

Figure 2: Ensemble based sentiment classification approach

4 Performance Evaluation

4.1 Data Annotation: Inter-annotator Agreement Score

We employed the Fleiss' Kappa inter annotator agreement score to evaluate manual data annotation evaluation between annotator. Fleiss' kappa score is calculated using the formula below (Wik21).

$$k = \frac{\bar{P}_x - \bar{P}_x}{1 - \bar{P}_x} \qquad (4)$$

Where, the factor $1 - \bar{P}x$ represents the degree of agreement that can be obtained other than by chance, The degree of agreement that was achieved above chance is given by $\bar{P}_x - \bar{P}_x$. and if the evaluators are totally in agreement, Kappa k = 1 and k = 0 if there is no agreement among the evaluators (other than what would be expected by chance). And for Marathi Tweets dataset the inter-annotator agreement score is k = 0.957, which is almost perfect agreement. Tab. 1. Inter-Annotator agreement score shows Inter-Annotator agreement score and Tab. 2. The statistics for Marathi tweets dataset after preprocessing and data annotation. shows the

statistics for Marathi tweets dataset after preprocessing and data annotation. Inter- Annotator agreement score and the statistics for Marathi tweets dataset after preprocessing and data annotation are shown in graphical manner in Figs. 3 and 4. respectively.

Table 1: Inter-annotator agreement score

Annotator (i with j)	Fleiss's Kappa score
A_{12}	0.953
A_{23}	0.965
A_{13}	0.954
Avg. agreement score	**0.957**

Table 2: The statistics for marathi tweets dataset after preprocessing and data annotation.

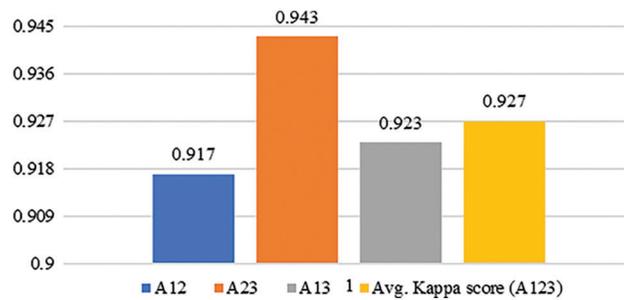| Sr. no.	Statistics	No. of tweets
1	Initial	1493
2	After Preprocessing	1069
3	Positive	627
4	Negative	430
5	Neutral	12

Figure 3: Graphical representation of inter-annotator agreement score (fleiss's kappa)
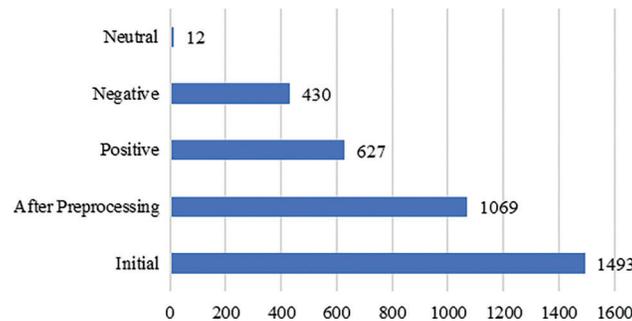

Figure 4: Graphical representation of statistics for marathi tweets dataset

4.2 Performance of Sentiment Classification Approach

We concentrated on three sorts of class problems in the experiment: positivity, neutrality, and negativity. Using the Twitter API, we retrieved Marathi tweets. Furthermore, the Marathi Tweets dataset is classified into three groups depending on the sentiment represented in the statements. If the expressed attitude indicates positivity, then labelled as 1, if it is neutrality then labelled as 0, and if it is negativity then labelled as −1.

The dataset is partitioned into 75:25 ratios for training and testing datasets. The dataset is subjected to different preprocessing methods, including data cleaning, URL and Hashtag removal, unnecessary blank spaces, emojis, removal of Stopword, and lemmatization. k-fold cross validation with k = 5 and k = 10 was also employed.

And evaluation metrics used are F-score and Accuracy which are calculated as below.

$$Recall = \frac{TP\_Sentiment}{TP\_Sentiment + FN\_Sentiment} \tag{5}$$

$$Precision = \frac{TP\_Sentiment}{TP\_Sentiment + FP\_Sentiment} \tag{6}$$

$$F - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7}$$

$$Accuracy = \frac{TP\_Sentiment + TN\_Sentiment}{TP\_Sentiment + FN\_Sentiment + FP\_Sentiment + TN\_Sentiment} \tag{8}$$

Analyzed comparative results from base classifiers, majority voting ensemble, and developed ensemble classifier. The proposed ensemble classifier's performance is compared to that of the individual conventional classifier and the majority voting ensemble classifier. Tab. 3. displays the results. On Marathi datasets, the suggested ensemble classifier outperformed the stand-alone classifier and the majority voting ensemble classifier.

Table 3: Performance evaluation of individual classifier with k-fold validation

Sr. no.	Classifier	Unigram + TF-IDF ($k = 5$)		Unigram + TF-IDF ($k = 10$)	
		Accuracy	F-score	Accuracy	F-score
1	SVM	92.46%	96.00%	91.89%	95.65%
2	Multinomial Naïve Byes	90.76%	95.15%	89.53%	94.44%
3	K-Nearest Neighbour	91.98%	95.70%	89.63%	94.35%
4	Neural Network	93.40%	96.44%	92.83%	96.01%
5	Decision Tree	91.71%	91.84%	90.90%	92.94%
6	Logistic Regression	90.76%	95.15%	88.97%	94.15%
7	Stochastic Gradient Decent	95.47%	97.37%	96.13%	97.62%
8	Ensemble Classifier	**96.77%**	**98.73%**	**97.77%**	**97.89%**

A classification model may be assessed using a variety of metrics, the most basic of which is accuracy and f-score. Tab. 3. shows the performance evaluation of individual classifier with k-fold validation. Graphical representation of performance evaluation of individual classifier with k-fold validation is shown in Figs. 5 and 6.

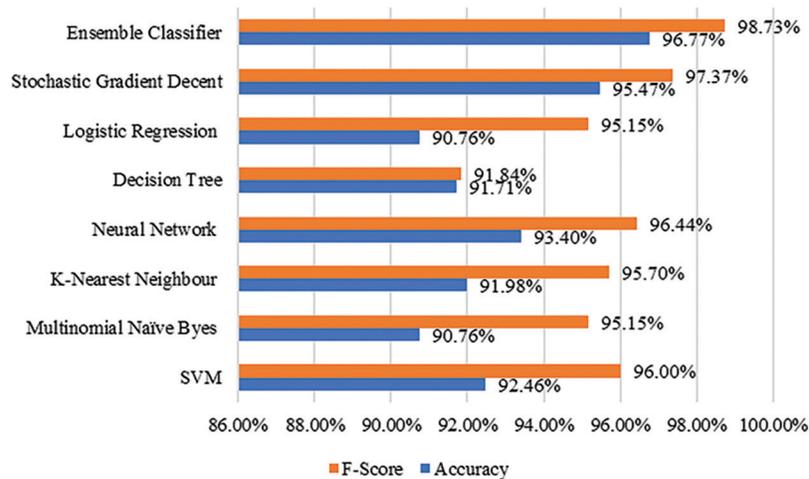**Figure 5:** Graphical representation of performance evaluation of individual classifier with 5-fold validation
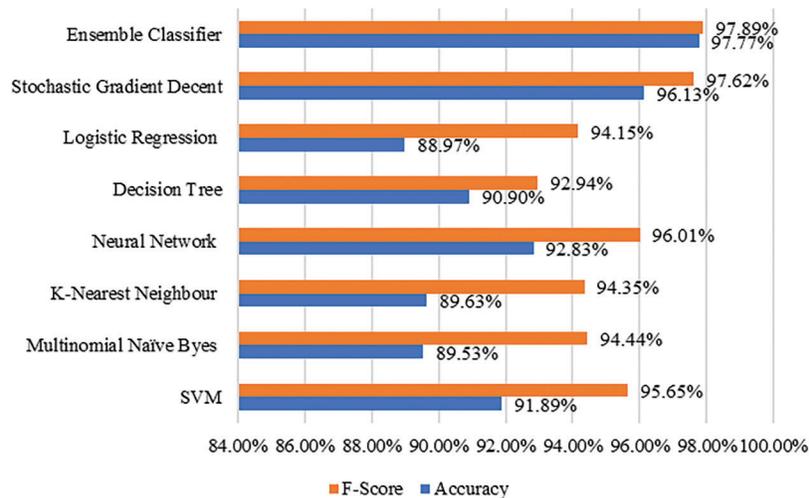

Figure 6: Graphical representation of performance evaluation of individual classifier with 10-fold validation

We performed 5-fold cross validation (cv) on dataset, for individual classifier Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), K- Nearest Neighbour (KNN), Neural Network (ANN), Decision Tree (DT), Logistic Regression (LR), Stochastic Gradient Decent (SGD), we obtained accuracy as 92.46%, 90.76%, 91.98%, 93.40%, 91.71%, 90.76%, and 95.47% respectively and obtained better performance in terms of accuracy for ensemble classifier as 96.77%, f-score is 98.73%. For 10-fold cross validation (cv) on dataset, individual classifier SVM, MNB, KNN, ANN, DT, LR, and SGD, we obtained accuracy as 91.89%, 89.53%, 89.63%, 92.83%, 90.90%, 88.97%, and 96.13%, respectively and we obtained better accuracy for ensemble classifier as 97.77%, f-score is 97.89% for Marathi tweets dataset.

4.3 Result Discussions

This is the first attempt to develop and evaluate a machine learning-based ensemble classifier for Marathi, and because there are no results for the same language, we compared our model with Hindi and Konkani for result analysis because these languages are considered for sentiment analysis using Machine Learning algorithms, and they are also in the Devanagari language family. The authors employed

machine learning techniques such as Naive Bayes, Decision Tree, and Support Vector Machine (SMO) using the Weka tool to reach accuracy of 50.95%, 54.48%, and 51.07% for the electronics product review dataset in Hindi [25]. In the case of Konkani, the authors used a dataset of Konkani poetry with Naive Bayes classification and attained an accuracy of 82.67% [26–28]. Furthermore, we have obtained better classification results for ensembled based classifier as 96.77%, 97.77%, for 5-fold and 10-fold cv respectively.

5 Conclusions

This research work presents a benchmarked technique for Sentiment Analysis of an Asian language "Marathi". For which we created an annotated corpus of Marathi Tweets, and performed manual data annotation with the help of domain experts with tweets labelled as positivity, neutrality and negativity polarity score that is 1, 0, and −1. And for performance evaluation of manually annotated corpus we used Fleiss's kappa (Inter-annotator agreement score) metrics and achieved average kappa score k = 0.957, which is almost perfect agreement between inter-annotator. For ensemble-based Sentiment classification experimentation, obtained better performance in terms of accuracy for ensemble classifier with 5-fold cross validation (cv) 96.77%, f-score is 98.73% and with 10-fold cross validation (cv), we obtained better accuracy for ensemble classifier as 97.77%, f-score is 97.89% for Marathi tweets dataset in comparison with another machine learning classifier.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] R. Biswarup, G. Avishek and S. Ram, "An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews," *Applied Soft Computing*, vol. 98, no. 17, pp. 106–119, 2021.

[2] K. Dashtipour, C. Ieracitano, M. Carlo, A. Raza and A. Hussain, "An ensemble based classification approach for persian sentiment analysis," in *Progresses in Artificial Intelligence and Neural Systems, Smart Innovation, Systems and Technologies*, Springer, Singapore, vol. 184, no. 3, pp. 207–215, 2021.

[3] M. Ghosh and G. Sanyal, "An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning," *Journal of Big Data*, vol. 5, no. 1, pp. 123–138, 2018.

[4] K. Sarkar and M. Bhowmick, "Sentiment polarity detection in bengali tweets using multinomial naïve bayes and support vector machines," in *IEEE Calcutta Conf. (CALCON)*, India, pp. 31–35, 2017.

[5] A. Kannan, G. Mohanty and R. Mamidi, "Towards building a sentiwordnet for tamil," in *Proc. of the 13th Int. Conf. on Natural Language Processing*, India, pp. 30–35, 2016.

[6] M. G. Jhanwar and A. Das, "An ensemble model for sentiment analysis of hindi-english code-mixed data," in *Workshop on Humanizing AI (HAI)*. Stockholm, Sweden: IJCAI, 2018.

[7] R. Gayathri, R. Vincent, M. Rajesh, A. K. Sivaraman and A. Muralidhar, "Web-acl based dos mitigation solution for cloud," *Advances in Mathematics Scientific Journal*, vol. 9, no. 7, pp. 5105–5113, 2020.

[8] D. M. Mathews and S. Abraham, "Twitter data sentiment analysis on a malayalam dataset using rule-based approach," In: S. N., P. L., N. H., H. P., and N. N. (Ed.), *Emerging Research in Computing, Information, Communication and Applications*, Springer, Singapore, vol. 906, pp. 407–415, 2019.

[9] A. Oscar, C. P. Ignacio, J. Fernando and I. Carlos, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Systems with Applications*, vol. 77, no. 12, pp. 236–246, 2017.

[10] M. Ganga, N. Janakiraman, A. K. Sivaraman, A. Balasundaram, R. Vincent *et al.*, "Survey of texture based image processing and analysis with differential fractional calculus methods," in *Int. Conf. on System, Computation, Automation and Networking (ICSCAN)*, IEEE Xplore, Puducherry, India, pp. 1–6, 2021.

[11] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. of the ACL-02 Conf. on Empirical Methods in Natural Language Processing (EMNLP '02)*, USA, pp. 79–86, 2002.

[12] D. Kothandaraman, A. Balasundaram, R. Dhanalakshmi, A. K. Sivaraman, S. Ashokkumar *et al.,* "Energy and bandwidth based link stability routing algorithm for IoT," *Computers Materials & Continua*, vol. 70, no. 2, pp. 3875–3890, 2021.

[13] P. Sharma and T. S. Moh, "Prediction of indian election using sentiment analysis on hindi twitter," in *IEEE Int. Conf. on Big Data (Big Data)*, Japan, pp. 1966–1971, 2016.

[14] Y. Qiang, Z. Zhang and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6527–6535, 2009.

[15] A. Balasundaram, G. Dilip, M. Manickam, A. K. Sivaraman, K. Gurunathan *et al.,* "Abnormality identification in video surveillance system using dct," *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 693–704, 2021.

[16] S. Rushdi, "Experiments with SVM to classify opinions in different domains," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14799–14804, 2011.

[17] A. Q. Md, D. Agrawal, M. Mehta, A. K. Sivaraman and K. F. Tee, "Time optimization of unmanned aerial vehicles using an augmented path," *Future Internet, MDPI*, vol. 13, no. 12, 308, pp. 1–13, 2021.

[18] S. S. Prasad, J. Kumar, D. K. Prabhakar and S. Tripathi, "Sentiment mining: An approach for Bengali and Tamil tweets," in *Ninth Int. Conf. on Contemporary Computing (IC3 (pp. 1-4)*, Noida, IEEE, pp. 263–278, 2016.

[19] S. Rani and P. Kumar, "A journey of Indian languages over sentiment analysis: A systematic review," *Springer*, vol. 6, no. 2, pp. 1415–1462, 2018.

[20] V. C. Joshi and V. M. Vekariya, "An approach to sentiment analysis on gujarati tweets," *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 1487–1493, 2017.

[21] R. Gayathri, A. Magesh, A. Karmel, R. Vincent and A. K. Sivaraman, "Low cost automatic irrigation system with intelligent performance tracking," *Journal of Green Engineering*, vol. 10, no. 12, pp. 13224–13233, 2020.

[22] Y. Sharma, V. Mangat and M. Kaur, "A practical approach to sentiment analysis of hindi tweets," in *1st Int. Conf. on Next Generation Computing Technologies (NGCT-2015)*, Dehradun, India, vol. 19, no. 2, pp. 677–680, 2015.

[23] R. Vincent, P. Bhatia, M. Rajesh, A. K. Sivaraman and M. S. S. Al Bahri, "Indian currency recognition and verification using transfer learning," *International Journal of Mathematics and Computer Science*, vol. 15, no. 4, pp. 1279–1284, 2020.

[24] L. Yang, B. Jian-Wu and F. Zhi-Ping, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Information Science*, vol. 23, no. 8, pp. 38–52, 2017.

[25] M. S. Akhtar, A. Ekbal and P. Bhattacharyya, *Aspect based sentiment analysis: Category detection and sentiment classification for hindi*. Vol. 23. Cham: Springer, pp. 246–257, 2018.

[26] A. Rajan and A. Salgaonkar, *Sentiment analysis for Konkani language: Konkani poetry, a case study*. vol. 9. Singapore: Springer, pp. 321–329, 2020.

[27] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, "TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 32, no. 9, pp. 1–13, 2021.

[28] W. Sun, L. Dai, X. R. Zhang, P. S. Changa and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 83, no. 12, pp. 1–16, 2021.