Classifying Social Media Users' Stance: Exploring Diverse Feature Sets Using Machine Learning Algorithms

Kashif Ayyub†, Muhammad Wasif Nisar†, Ehsan Ullah Munir†, and Muhammad Ramzan††

† Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt., Pakistan
†† Department of Computer Science and Information Technology, University of Sargodha, Sargodha, Pakistan

Summary
The use of the social media has become part of our daily life activities. The social web channels provide the content generation facility to its users who can share their views, opinions and experiences towards certain topics. The researchers are using the social media content for various research areas. Sentiment analysis, one of the most active research areas in last decade, is the process to extract reviews, opinions and sentiments of people. Sentiment analysis is applied in diverse sub-areas such as subjectivity analysis, polarity detection, and emotion detection. Stance classification has emerged as a new and interesting research area as it aims to determine whether the content writer is in favor, against or neutral towards the target topic or issue. Stance classification is significant as it has many research applications like rumor stance classifications, stance classification towards public forums, claim stance classification, neural attention stance classification, online debate stance classification, dialogic properties stance classification etc. This research study explores different feature sets such as lexical, sentiment-specific, dialog-based which have been extracted using the standard datasets in the relevant area. Supervised learning approaches of generative algorithms such as Naïve Bayes and discriminative machine learning algorithms such as Support Vector Machine, Naïve Bayes, Decision Tree and k-Nearest Neighbor have been applied and then ensemble-based algorithms like Random Forest and AdaBoost have been applied. The empirical based results have been evaluated using the standard performance measures of Accuracy, Precision, Recall, and F-measures.
Keywords:
Sentiment analysis, Stance classification, Emotion, Opinion Mining, Feature set, Machine Learning.

1. Introduction

The use of social networking sites is rapidly increasing. The large amount of unstructured data available on the social media website, where people share their views, opinions and sentiments. These social media networks are medium of sharing thoughts in form of words, audios, videos and images. People express their opinions, feeling and sentiments through text. Opinion mining and sentimental analysis has gained attention of researchers. Sentiment Analysis (SA) is a process of identifying the emotions , sentiments of individuals to an entity [1]. It also aims to detect the sentiment polarity of a text by classifying into positive, negative, or neutral classes [2]. Sentiment analysis is usually carried out at three levels of document, sentence or aspect. Sentiment analysis examines people's feelings, opinions, attitudes, emotions and sentiments regarding topics, individual, events, product and their attributes.

Stance classification is the sub-domain of sentiment analysis [3, 4]. Stance classification is a process to determine from the text whether the writer of text is in favor, against or neutral about the given target which can be a product, social issue, individual or some policy. This research field is very challenging because the social data available include informal language such as emojis, misspellings, hashtags, repeated characters and slang words. Then we do not have to find the sentiment but the stance towards the target in given source. Stance detection in many fields is becoming increasingly important. Stance studies, for example, can help detect democratic problems and understand how public attitudes are shaped. For better explanation, the Table 1 shows the tweet examples in terms of target to detect stance.

Table 1: Sample examples of content and implicit Stance related Class

Target	Tweet	Stance
Feminist Movement	We live in a sad world when wanting equality makes you a troll...	Favor
Donald Trump	Donald Trump....you're fired, fired, & fired again!	Against
Hillary Clinton	I'm going to miss the 5pm interview so I'm depending on you all to live tweet it. Haha.	Neutral

The data provided in Table 1 is labeled as a target Subject, a related Tweet, and the Tweet's Stance against the target. Detection of the stance from the piece of text determines the favorability of a given target. In the above example, the given target is Donald Trump, but the target of opinion in tweet is expresses negative opinion, so the opinion of tweet is unfavorable towards the Donald Trump. The aims and objectives of the research study include exploring the different feature sets such as lexical, sentiment-specific features in stance detection.Our objective is to evaluate the performance of various machine learning algorithms such as Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM),

Manuscript received February 5, 2024
Manuscript revised February 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.2.9

KNN, AdaBoost and Random Forest (RT) for stance detection. In this research study, main research contributions include that we have extracted and computed diverse feature sets such as tweet-specific, emotion-based and linguistic features. The various different types of machine learning algorithms are applied to classify the stance classification. The machine learning algorithms include Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), l-Nearest Neighbor (KNN), AdaBoost and Random Forest (RT). The standard performance evaluation measures accuracy, precision, recall, and f-measures are used to assess the performance of different methods. The experiments are performed on two datasets of SemEval2016 task 6 which is specific for stance detection and Stance detection dataset.

The rest of the research paper is presented as follows: Section 2 presents formal research problem statement. Section 4 presents the proposed research framework, Section 5 Presents the experimental setup which presents the datasets used and the performance evaluation measures applied. Section 6 discusses the results and then conclusion is presented.

2. Problem Statement

The social web users share their reviews, opinions and sentiment on a target through social media websites. In user reviews contains information related to specific topic. The user opinion, views can be in favor against or neutral on specific topic. Formally, a user $u \epsilon U$ can give review $r \in R$ on specific target $t \epsilon T$. A r can be against or in-favor of the t. Our aim is to classify each r into the target multi-classes.

3. Related Work

We find a number of related studies for stance detection which focus on conventional machine learning approaches for stance detection. Hawkes Processes for Continuous Time Sequence Classification Model [5] proposed to use Hawkes Processes to identify sequences of temporal textual data that manipulate both temporal and textual information. Their rumor stance classification experiments on four Twitter databases show the importance of using tweet temporal information along with text content. This paper uses a multinomial likelihood and a prior over label frequencies to model the document. The Naïve Bayes process is a special case of classification in a Hawkes process. Where 0 is set to a specific subset of parameters. paper used models from Random Forest as the outperformed regular logistical regression and SVM on the development set with linear kernels. They calculated the mean class probabilities across all trees to calculate the estimated probability from Random Forest.

Detecting Stance in Catalan and Spanish Tweets Model [11] described the stance and gender of tweets in Catalan and Spanish. In specific, they have made three classes of features; stylistic, structural and cultural-based, along with two new features that will exploit significant features expressed in tweets by URLs. This paper uses a combination of SVM, Random Forest, Logistic Regression, Decision Tree and Multinomial Naïve Bayes (MNB) classifiers and classifies tweets as one of the features mentioned in the paper. The dataset included tweets of common URLs in the dataset. Friends and Enemies of Clinton and Trump Model [12] describes a novel approach to the detection of tweet stance by considering the context surrounding and model training aimed at predicting the stance towards the targets listed. they became interested in investigating the 2016 election between Hillary and Donald's political debate on social media involving a comparison and it reveals the details between enemies and friends.

From Clickbait to Fake News Detection Model [13] demonstrates an approach to detecting the stance of headlines for fake news and particularly clickbait detection relative to the corresponding article bodies. This paper discusses the detection of fake news using machine learning algorithms for digital content creation. The authors used the Fake News Challenge dataset in this paper, where each example contains claim document pair with the following possible relationships between them: agree, disagree, discuss & unrelated. Improving Claim Stance Classification Model [14] faced two limitations on the classification of claim stance. That is to conclude, by providing a lexicon expansion method and a set of effective contextual features insufficient coverage of manually written sentiment lexicons and ignoring the context of the claim but challenge remains to accurate predictions of opposing targets. This system uses the SVM classifier that was trained for the lexicon expansion feature vector and labels for 200-dimensional word embedding. The study analyzed the trade-offs between accuracy and coverage Such trade-offs have been controlled by setting a minimum level of confidence to make a prediction. The average performance of the system is 51.9 percent.

Simple Open Stance Classification Model [15] presents an open-stance method for rumor and veracity in twitter. The approach benefits from a novel set of automatically identifiable problem-specific features to improve classifier accuracy and achieve on recent data sets above state-of - the-art results. This paper compares and trains Random Forest with other models on the RumourEval dataset and observed that Random Forest outperformed all other classifiers. Random Forest's accuracy measured was the highest 79.02 that was the lowest among Decision Tree and KNN. This paper used decision tree and random forest and KNN, but all other methods were outperformed by Random Forest, so authors do not tend to use this approach and go with Random Forest.

Shared Task on Stance and Gender Detection Model [16] used FastText, which is an extension of the classic bag of words, a classifier with 5 million tweets trained pre-trained embedding. In some tasks the model is outstanding but poor when computing on some that need improvement

Stance Classification of Context-Dependent Claims model [17] deals with the classification of claims with respect to a given topic. To break this complex task into simpler and well-defined subtasks, a model is proposed. I.e. the identification of open domain targets, the classification of emotions for each target, and the detection of open domain comparison between topic and claim targets.

In Toward Stance Classification Based on Claim Microstructures Model [18], A method for representing claims as microstructures is introduced , which describes the beliefs opinions and policies regarding particular concepts of the domain The authors used the SVM classifier with the RBF kernel in the method described in this paper are trained and evaluated the models on 803 claim instances using a 5x3 nested cross-validation with grid search to improve hyper parameters C and γ.

In A Biased Representation for a Biased Problem Model [19] the problem discussed is on Catalan independence in tweets has many characteristics and is exploited by developing a system based on pre-processing and representation. Only SVM and ANN show similar effects, but their difference will grow with more data to train. This model explores two authors ' systems. One is based on the Artificial Neural Network (ANN) and the other is based on Support Vector Machine (SVM). The SVM classifier method employs a multi-class one-vs-rest approach which uses a linear kernel. Stance Detection in Online Discussions Model [20] focuses on determining whether the comment of author is in support of the target or against it. The approach used the highest entropy classifier that uses features specific to the surface level, sentiment and domain. Designed for English tweets, this paper adapted this method to Czech news commentaries

Joint Named Entity Recognition and Stance Detection Model [21] explores the contribution of named entities to tweet-based stance detection tasks. The results of NER experiments and associated stance experiments will be published on a publicly available stance-annotated data set of tweets using named entities. The results improved when using unigrams to use named entities. The system uses SVM classifier with the following features namely unigrams, bigrams, and hashtags and unigrams. The results showed that the use of unigrams as features leads to favorable results and the use of unigrams and hashtags together further increases these results but results in poor performance with the use of bigrams.

Stance Detection Model [22] proposes a Turkish stance detection tweet data set consisting of stance annotation tweets for two popular sports clubs along with SVM classifier for each data set target where classifiers use unigram, bigram and hashtag features This analysis provides the initial data set for Turkish tweets. In Performing Stance Detection on Twitter Data using Computational Linguistics Techniques model [23] the authors use supervised learning approach to perform stance detection on Twitter Data. They started by extracting bag-of-words to perform classification, then trying to optimize the features to improve the accuracy of stance detection.

In stance and gender detection in Tweets on Catalan Independence Model [24] the authors proposed their gender and stance detection method in both Spanish and Catalan using character and word level features and classification SVM technique. In this model, the system uses Scikit-learn SVM implementation with a radial basis function kernel for classification after pre-processing the data set and extracting features. The authors also performed 10-fold validation for the system development on the given training dataset.

A Feature Selection and Machine Learning Based Model [25] detect stance on Russian texts in terms of labeling them for or against a topic of discussion. To achieve this goal, they used several Machine Learning algorithms. The system uses five methods of machine learning in this paper is to solve the problem of stance detection. A five-fold cross-validation procedure was applied to optimize the classifier parameters of a five-fold nested cross-validation procedure and to achieve objective estimates of the classification quality SVM with linear, RBF, and polynomial kernels, regularization coefficient, a Naïve Bayes classifier with a multinomial distribution, enabled the use of a dictionary composed of all lemmas of the text corpus to obtain a better classification quality but the result of this technique was weaker than other.

In Predicting Stances from Social Media Posts using Factorization Machines Model [26] the authors provide a method to identify a person's attitude to a topic based on their stance to other issues and social media posts. Factorizing devices are used to model user preferences to the social media data topics. This model presents a method for obtaining statements of stance from tweets and model the topic preferences from the statements of stance together with the user tweets. The experimental results show that posts from users are useful for model preferences of topics and thus predict silent user stances.
Concluding sentence

4. The Proposed Research Approach

The proposed framework will take two datasets to perform feature engineering. In feature engineering we will extract linguistic and emotional features then these extracted features are fed to different classifiers for classifying them in classes. And last step is to apply performance evaluation measures to measure performance of different feature engineering techniques and different classifiers. Figure 1 represents proposed framework.

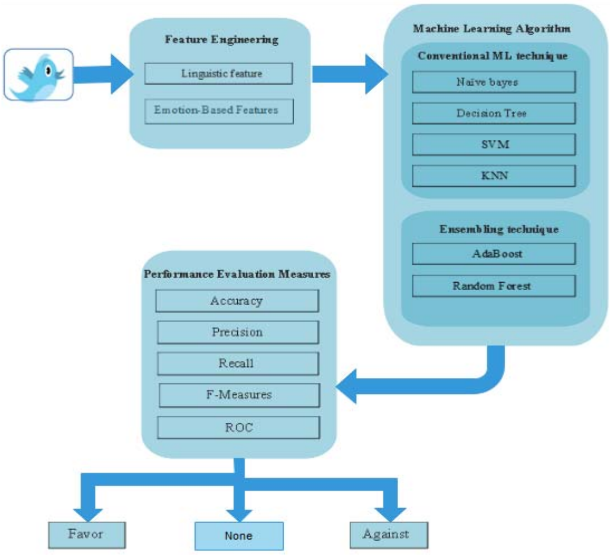

Figure 1: The steps of the Proposed Framework

4.1 Feature Engineering

The features are the characteristics of the data which are used as an input to the machine learning algorithms and the supervised learning algorithms learn from the set of features and then the target feature which is also known as class label is predicted. Let us here share the three different feature sets and the arguments behind choosing the various features in these sets.

4.1.1 Tweet-specific Features

The basic structure of the tweets is explored in research studies as this implicitly help us to analyze the overall characteristics of the source tweets. It includes the text length of the tweet. It has been considered after removing the hashtags, mentions and URLs from Twitter. Another feature is the question marks count, it is significant in this regard that usually the discussion or conversation in the tweets is carried out so this is more applicable in discussions where sentiment or stance is considered and it is not applicable in tweets where facts are provided. In addition to this, as the tweets overall length is less than the existence of question mark is also considered as Boolean feature. Moreover, the presence of hashtag is also considered as feature. The reason behind taking this as a feature is that hashtag depict topic. In stance-based tweets, topic may be more important and explicit as compared to objective tweets. Also, the number of URLs has also been considered as a feature exploring whether this plays important part in distinguishing stance based content or not.

4.1.2 Emotion-based Features

As stance detection lies under the main umbrella of sentiment analysis, therefore, we consider the emotion-based features. As stance in favor or against is very much related to sentiment, so we take negative sentiment as a feature as this may be helpful for prediction of against class. The analogous assumption may be considered for other class. Also, in addition to sentiment polarity, we take emotion of Fear as well as this is strong negative emotion which is helpful for stance detection. The fear has been considered based on a list of words which have been considered for the same purpose from the Plutchik model [27, 28]. In addition, we consider two specific features, one which is related to positive mood which is categorized as three sub-perspectives of Happy, Active and Imaginative [29] while the other is the core analysis of sentiment related model known as VAD model [30].

4.1.3 Linguistic Features

The use of lexicons is very common for the sentiment related studies and stance detection is not an exception. Linguistic Inquiry and Word Count (LIWC) [31] is a well-known source and widely used in the relevant literature [32-36], for vast and diverse categorization of words into various sentiment are emotions. In this study, we consider as these are helpful for us in this study for features related to emotion which indirectly relate to stance such as support, question opinion and sad. LIWC provides features in Boolean form in our case, where the presence of the stance related points have been considered as one if found else zero.

5. Experimental Setup

Let us discuss the dataset to be used, algorithms to be used and performance evaluation measures to be applied.

5.1 Dataset

We used two datasets for experimentation purposes. First is the SemEval2016[1] for task 6 which is for stance

[1] http://alt.qcri.org/semeval2016/task6/

detection. This dataset is applicable for stance classification. It consists of 1956 tweets in total. The dataset has two subtasks. The data has attributes like ID, target, tweets, stance. The size of dataset is 98 KB and available publicly for research purpose. Another data is Stance Detection [2] dataset which consists of 75385 headlines. It has different attributes like bodyID, articleBody, headlines, stance. The dataset has been released on Jan 20, 2019.

5.2 Machine Learning Techniques

The applied machine learning are briefly elaborated here.

5.2.1 Conventional Machine Learning Techniques

We use two Conventional Machine Learning methods for our proposed work. These methods are Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM) and KNN.

Naïve Bayes (NB)
Naive Bayes is one of the most likely machine learning and data mining algorithms. For classification of text, NB is commonly used. NB simply based on Bayes theorem. The Bayes theorem formula is:

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (1)$$

$$P_{NB}(c|d) = \frac{(P(c))\sum_{i=1}^{m} P(f|c)^{n_{i(d)}}}{P(d)} \quad (2)$$

Where f is a feature, feature count (fi) is labelled as n (i(d)) and is present in d as a tweet. Here, m indicates no. of features.

Decision Tree (DT)
The decision tree is supervised algorithm of machine learning and use for classification problem by using tree representation. Two entities in the decision tree, are nodes and leaves. The leaves are the final results and the nodes of the decision are where the data are divided. The decisions are selected in such a way that the tree is as small as possible while aiming for high classification.

Support Vector Machine (SVM)
SVM is the supervised algorithm for machine learning and classifies text effectively. SVM evaluates the data, defines the boundaries of the decision and uses the computing kernels in the input space. Any input value that was a vector is categorized into a class and then the margins between the class are defined. SVM optimization is calculated as,

$$\vec{\alpha} = \operatorname{argmin}\{-\sum_{j=1}^{n} \alpha_j \sum_{k=1}^{p} \sum_{k=1}^{p} \alpha_i \alpha\, y_i y(\vec{z}_j \quad , \vec{z}_k)\} (3)$$

$$\sum_{j=1}^{n} \alpha_i y_i = 0; 0 \le \alpha \le C \quad (4)$$

K-Nearest Neighbor (KNN)
This algorithm is supervised machine learning algorithm which is used to solve problems of classification and regression. Classification is carried out by majority decision to its nearest neighbors. The data is classified to the class with the nearest neighbors. As increase the number of nearest neighbors, the k value can increase the accuracy.

5.2.2 Ensemble based Technique

Ensemble approach includes multiple algorithms to achieve better performance. In our proposed work, we use two ensemble-based techniques; AdaBoost and Random Forest.

AdaBoost
AdaBoost is a machine learning meta-algorithm. It aims to create a strong classifier for several weak classifiers. Each instance is weighted in the training dataset. The initial weight is set to:

$$\text{weight(ki)} = \frac{1}{X} \quad (5)$$

Where ki is the i'th instance of training and X is the number of training instances. The misclassification rate for the trained model is calculated as:

$$error = \frac{(\text{correct}-X)}{X} \quad (6)$$

Another equation to use the learning instances weighting:

$$error = \frac{sum(k(i)^{*}\ terror(i))}{\text{sum(k)}} \quad (7)$$

In the equation (7), k is the weight of training elements i and terror is the error prediction for element i when terror is 1 mistake detection exists if terror is 0 elements are properly classified.

Random Forest
Random forests are a classification and regression method for group learning and Ensemble based learning. Decision trees tend to learn unusual patterns. Using this process, when deep trees learn the same part of the training sample RF takes an average of the variance of its value. Training set P = p1, ..., pn with R= r1, ..., rn, repeated bagging (K times) selects a random sample to replace the training set and fits trees to the samples:
For k = 1, ..., K:

[2] https://www.kaggle.com/ad6398/stance-detection#__sid=js0

Replaced samples, R training examples from P, R; call these P_k, R_k. Regression tree is trained f_k on P_k, R_k or more vote for Decision Tree.

$$\frac{1}{K}\sum_{k=1}^{k} f_k\,(\dot{R}) \tag{8}$$

5.3 Performance Evaluation Measures

To check how accurate classifiers classified the Stance Classification accuracy, precision, recall and F-measure are considered which are elaborated as follows.

5.3.1 Accuracy

Accuracy is the most commonly used measure of performance. In general, it is the ratio of the number of correctly predicted observation over the total observation. The formulae to calculated accuracy is given below:

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{9}$$

where, TP is the True Positive and TN is True Negative, FP is False Positive, and FN False Negative.

5.3.2 Precision

Precision is the ratio of measurements that are predicted to be correctly positive from the total positive predictive observation [37]. The formulae to find precision is given below:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

5.3.3 Recall

Recall is the ratio to measures the correctly predicted positive observation to actual observations. The formulae to find recall is given below:

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

5.3.4 F-Measures

It combines precision and recall in one performance metrics and keeps a balance between them. F-measures is the harmonic mean of precision and recall. To calculate any F-α, we can use equation 12:

$$\text{F} - \text{measure} = \frac{(\alpha 2\ +\ 1) *\ \text{precision}\ *\ \text{recall}}{\alpha 2\ *\ \text{precision}\ +\ \text{recall.}} \tag{12}$$

The F-score is equal if $\alpha = 1$. when $\alpha > 1$ It favors precision, and otherwise recall.

6. Results and Discussions

For classification purposes, many supervised learning algorithms are used. Such learning algorithms are divided into two main categories that are algorithms based on conventional machine learning and ensemble algorithms including Random Forest and AdaBoost with commonly used conventional machine learning algorithms including Support Vector Machine, Naïve Bayes, Decision Tree and KNN.

Although we classified our classifiers based on their method. In this chapter, we will address separately the results obtained on both datasets by each classifier. In this research study two features which are linguistic based, and emotion based along with conversational and structural features are used. In this section we will discuss the results obtained by using conventional machine learning algorithms to classify datasets. For classification purposes, Conventional machine learning algorithms including Support Vector Machine, Naïve Bayes, Decision Tree and KNN are used and then the standard classification performance evaluation measures of Accuracy, Precision, Recall, F-measures and ROC are applied to evaluate the performance.

Results obtained after applying these machine learning algorithms on each feature set show different behavior on different datasets. Table 3 presents accuracy, precision, recall, f-measures for the SemEval 2016 Task 6 dataset. These results have shown that F-measures is higher when conventional machine algorithms are applied on all feature sets. In case of structural features best results are obtained by Decision Tree (F-measures = 94). Overall the decision tree gives the best results on all feature sets in terms of Accuracy, precision, recall and F-measures. The highest accuracy is obtained by using decision tree when applied on all features set. KNN also performs better result on all features sets except structural features. There is no result is obtained from conversational features because in conversational features we are using two features; text similarity to source tweet and tweet level. On each class of text similarity to source tweet is zero and for tweet level is 1 that's why the training is not possible.

Table 3: Results of Conventional ML Algorithm on Semeval2016 Dataset

Feature Set	ML Algorithm	Accuracy	Precision	Recall	F-Measure
All Features	NB	53	40	43	74
	SVM	69	50	79	90
	KNN	71	65	68	81
	DT	**80**	**75**	**77**	**83**
Structural Features	NB	57	33	40	70
	SVM	57	33	69	67
	KNN	44	37	37	51
	DT	**58**	**38**	**57**	**94**
Linguistic Features	NB	54	35	39	73
	SVM	57	34	76	89

	KNN	55	34	39	77
	DT	**57**	**33**	**72**	**88**
Emotional Features	NB	57	.33	40	69
	SVM	57	33	69	89
	KNN	70	64	67	79
	DT	**76**	**70**	**73**	**80**

Figure 2 give comparison of all algorithms applied for classification by applying proposed techniques for feature engineering in-terms of F-measures. This figure represent that conventional machine learning algorithms gives best results. KNN and DT best results are obtained by using structural, linguistics and emotion-based features. Overall the decision tree gives the best results on all feature set in terms of F-measures. Overall best performance on SemEval2016 is obtained by conventional machine learning algorithm is by using DT as classifier by using structural, linguistics and emotion-based features.

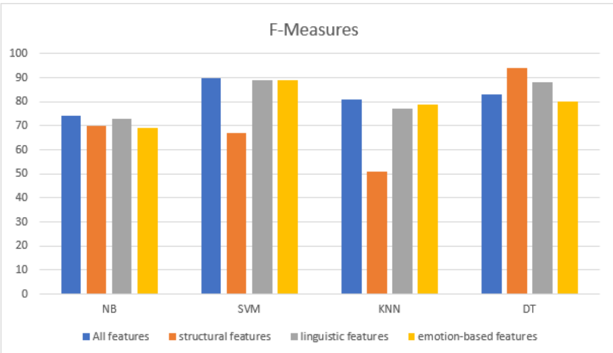

Figure 2: Comparison of Conventional ML Based Algorithm on SemEval2016 Dataset

Table 4 shows the results obtained by using five features set by applying machine learning algorithms. These results are represented for the Stance detection dataset in terms of accuracy, precision, recall, F-measures. Results obtained from this dataset indicate different behaviors in terms of F-measures from the first used dataset. Accuracy gives better results when conventional used in all set of features. The best result is obtained by NB (Accuracy = 70) in terms of structural features. Overall the KNN gives the better results on all feature sets in terms of Accuracy. This Stance detection dataset does not perform as good as SemEval2016 dataset.

Table 4: Results of Conventional ML Algorithm on Stance Detection Dataset

Feature Set	ML Algorithm	Accuracy	Precision	Recall	F-Measure
All Features	NB	40.5	38.1	42	32
	SVM	53.5	45.1	49	45
	KNN	**68**	**48**	**40**	**42**
	DT	59	38	38	38
Structural Features	NB	**70**	**46**	**40**	**42**
	SVM	55.3	45	48	45
	KNN	68	48	40	42
	DT	59	38	38	38
Linguistic Features	NB	**68.4**	**44**	**49**	**45**
	SVM	54	45	48	45
	KNN	68	45	39	40
	DT	68	46	39	40
Emotional Features	NB	55	38.3	45.3	37.5
	SVM	53.5	45	48.9	44.6
	KNN	**68**	**45.4**	**38.3**	**39.8**
	DT	67.1	44.8	38.7	40.3

Figure 3 gives comparison of the algorithms applied for classification by applying proposed techniques for feature engineering in-terms of accuracy, but this dataset does not perform as good as semEval2016 dataset. Overall the best result is obtained by NB. KNN gives better result from conventional learning algorithms in terms of all sets of features. After KNN, DT is also performed better in linguistic and emotion-based features.

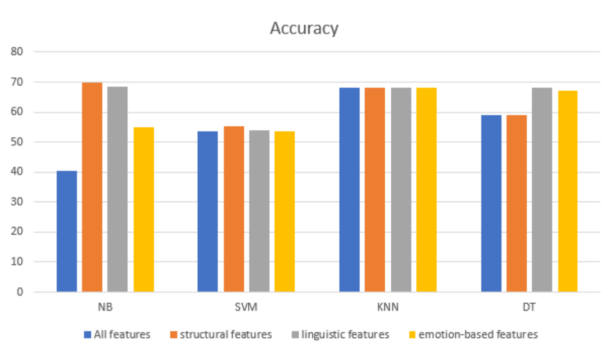

Figure 3: Comparison of Conventional ML Algorithm on Stance Detection Dataset

5.1 Results of Ensemble-based Technique on each Data Set

In this section, the empirical results of classification on basis of Ensemble based methods are discussed for both datasets by each classifier. In this research study two features which are linguistic based, and emotion based along with conversational and structural features are used. In this section we will discuss the results obtained by using Ensemble based machine learning algorithms to classify datasets. For classification purposes, Ensemble based machine learning algorithms including AdaBoost and Random Forest are used and then the standard classification performance evaluation measures of Accuracy, Precision, Recall, F-measures and ROC are applied to evaluate the performance.

Table 5 shows the results obtained by using five features set by applying Ensemble based machine learning algorithms. These results are represented for the Stance detection dataset in terms of accuracy, precision, recall, F-measures. Results obtained from this dataset indicate different behaviors in terms of F-measures from the first used dataset. Accuracy gives better results when all features are used. The better result is obtained by AdaBoost (Accuracy = 52.4) in terms of structural features in terms of accuracy, precision, recall, f-measures. This Ensemble based algorithms on Semeval 2016 dataset does not perform as good as conventional machine learning algorithm on semEval2016 dataset.

Table 5: The Results of Ensemble-based Algorithms on SemEval2016 Dataset

Feature Set	ML Algorithm	Accuracy	Precision	Recall	F-Measure
All Features	AdaBoost	**53**	**50**	**36**	**29**
	Random Forest	51	49	50	50
Structural Features	AdaBoost	**52.4**	**50**	**52**	**50**
	Random Forest	42	41	43	41
Linguistic Features	AdaBoost	5	39.4	39.9	39.6
	Random Forest	5	45	45.6	46.4
Emotional Features	AdaBoost	5	39	39.9	39.6
	Random Forest	**48.7**	**45.8**	**47.5**	**46.4**

Figure 4 give comparison of all algorithms applied for classification by applying proposed tech-niques for feature engineering in-terms of Accuracy. This figure represents that Ensemble based machine learning algorithms gives best results on all features, structural features and emotion-based features. AdaBoost and RF best results are obtained by using all features, structural and emotion-based features. Overall the AdaBoost gives the better results on all feature set and structural feature, and RF on emotion-based feature in terms of accuracy.

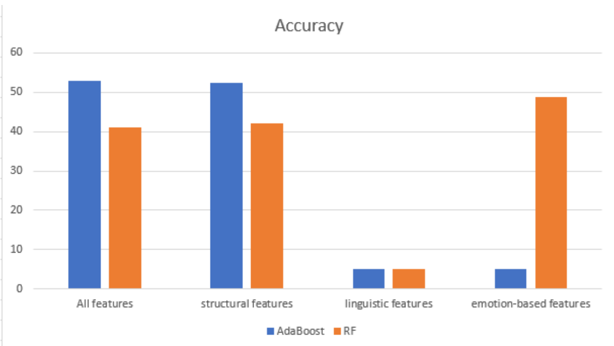

Figure 4: Comparison of Ensemble based Algorithm on SemEval2016 Dataset

Table 6 shows the results obtained by using five features set by applying machine learning algorithms. These results are represented for the Stance detection dataset in terms of accuracy, precision, recall, F-measures. Results obtained from this dataset indicate different behaviors in terms of Accuracy from the first used dataset. Accuracy gives better results when Ensemble based algorithms are applied. The best result is obtained by both Ensemble based algorithms; AdaBoost and RF on all feature set, structural, linguistic and emotion-based features. This Stance detection dataset performed well as compared to SemEval2016 dataset.

Table 6: Results of Ensemble based Algorithm on Stance Detection Dataset

Feature Set	ML Algorithm	Accuracy	Precision	Recall	F-Measure
All Features	AdaBoost	70	46	41	42
	Random Forest	70	51	42	44
Structural Features	AdaBoost	**70**	**46**	**41**	**42**
	Random Forest	69	47	41	42
Linguistic Features	AdaBoost	**70**	**46**	**39**	**40**
	Random Forest	69	46	40	42
Emotional Features	AdaBoost	**70.2**	**48**	**38.2**	**39.8**
	Random Forest	67	45	38.6	40

Figure 5 give comparison of all algorithms applied for classification by applying feature engineering in-terms of accuracy. This figure represent that incase of AdaBoost and RF best results are obtained by using structural, linguistics and emotion-based features. Overall best performance on Stance detection dataset is obtained by Ensemble based algorithm is by using AdaBoost as classifier along with structural, linguistic and emotion-based feature.

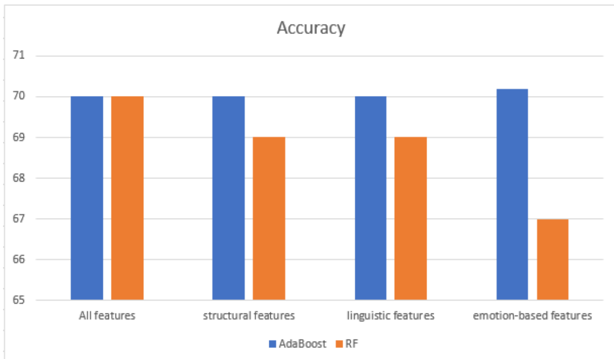

Figure 5: Comparison of Ensemble based Algorithm on Stance Detection Dataset

7. Conclusion

In this research study, the diverse features such as emotion-based and linguistic are used along with structural features which are used evaluate how these feature extractions affect performance of different classifiers by using conventional machine learning-based and Ensemble based algorithms in terms of accuracy, precision, recall and f-measures. In order to evaluate the performance of classifiers two learning algorithms are classified into two main categories which are conventional machine learning-based and Ensemble based algorithms. Ensemble based algorithms and conventional machine learning algorithms are compared using standard performance evaluation measures: accuracy, F-measure, recall, and precision. Classifiers are compared with each other on basis of emotional and linguistic features along with structural and conversational features. Results have shown that, there is no result is obtained from conversational features on two datasets because in conversational features we are using two features; text similarity to source tweet and tweet level. On each class of text similarity to source tweet is zero and for tweet level is 1 that's why the training is not possible. Overall best performance on SemEval2016 is obtained by conventional machine learning algorithm is by using DT as classifier by using structural, linguistics and emotion-based features. NB also performed good results in terms of structural features on stance detection dataset. Overall the KNN gives the better results on all feature sets in terms of Accuracy but Stance detection dataset does not perform as good as semEval2016 dataset when conventional ML algorithms is applied. For Ensemble based technique Adaboost obtained the better performance on semeval2016. Overall best performance is obtained by Ensemble based algorithms is by using AdaBoost along with structural, linguistic and emotion-based.

References

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5 (4): 1093–1113," ed, 2014.

[2] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 70-77.

[3] J. Ebrahimi, D. Dou, and D. Lowd, "A joint sentiment-target-stance model for stance classification in tweets," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2656-2665.

[4] J. Du, R. Xu, Y. He, and L. Gui, "Stance classification with target-specific neural attention networks," 2017: International Joint Conferences on Artificial Intelligence.

[5] M. Lukasik, P. Srijith, D. Vu, K. Bontcheva, A. Zubiaga, and T. Cohn, "Hawkes processes for continuous time sequence classification: an application to rumour stance classification in twitter," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 393-398.

[6] A. Mandya, A. Siddharthan, and A. Wyner, "Scrutable feature sets for stance classification," in *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, 2016, pp. 60-69.

[7] I. Persing and V. Ng, "Modeling stance in student essays," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 2174-2184.

[8] P. Sobhani, D. Inkpen, and X. Zhu, "A dataset for multi-target stance detection," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 2017, pp. 551-557.

[9] A. Misra, B. Ecker, T. Handleman, N. Hahn, and M. Walker, "Nlds-ucsc at semeval-2016 task 6: A semi-supervised approach to detecting stance in tweets," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 420-427.

[10] K. Joseph, L. Friedland, W. Hobbs, O. Tsur, and D. Lazer, "Constance: Modeling annotation contexts to improve stance classification," *arXiv preprint arXiv:1708.06309,* 2017.

[11] M. Lai, A. T. Cignarella, H. FARIAS, and D. IRAZU, "ITACOS at ibereval2017: detecting stance in Catalan and Spanish tweets," in *IberEval 2017*, 2017, vol. 1881, pp. 185-192: CEUR-WS. org.

[12] M. Lai, D. I. H. Farías, V. Patti, and P. Rosso, "Friends and enemies of clinton and trump: using context for detecting stance in political tweets," in *Mexican International Conference on Artificial Intelligence*, 2016, pp. 155-168: Springer.

[13] P. Bourgonje, J. M. Schneider, and G. Rehm, "From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 84-89.

[14] R. Bar-Haim, L. Edelstein, C. Jochim, and N. Slonim, "Improving claim stance classification with lexical knowledge expansion and context utilization," in *Proceedings of the 4th Workshop on Argument Mining*, 2017, pp. 32-38.

[15] A. Aker, L. Derczynski, and K. Bontcheva, "Simple open stance classification for rumour analysis," *arXiv preprint arXiv:1708.05286,* 2017.

[16] F. Barbieri, "Shared Task on Stance and Gender Detection in Tweets on Catalan

Independence-LaSTUS System Description," in *IberEval@ SEPLN*, 2017, pp. 217-221.

[17] R. Bar-Haim, I. Bhattacharya, F. Dinuzzo, A. Saha, and N. Slonim, "Stance classification of context-dependent claims," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 2017, pp. 251-261.

[18] F. Boltužić and J. Šnajder, "Toward stance classification based on claim microstructures," in *8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017.

[19] D. A. Garcıa and A. M. L. Flor, "Stance detection at IberEval 2017: A Biased Representation for a Biased Problem," *System,* vol. 2, p. 1, 2017.

[20] P. Krejzl, B. Hourová, and J. Steinberger, "Stance detection in online discussions," *arXiv preprint arXiv:1701.00504,* 2017.

[21] D. Küçük, "Joint named entity recognition and stance detection in tweets," *arXiv preprint arXiv:1707.09611,* 2017.

[22] D. Küçük, "Stance detection in Turkish tweets," *arXiv preprint arXiv:1706.06894,* 2017.

[23] G. G. Shenoy, E. H. Dsouza, and S. Kübler, "Performing stance detection on Twitter data using computational linguistics techniques," *arXiv preprint arXiv:1703.02019,* 2017.

[24] S. Swami, A. Khandelwal, M. Shrivastava, and S. S. Akhtar, "LTRC IIITH at IBEREVAL 2017: Stance and Gender Detection in Tweets on Catalan Independence," in *IberEval@ SEPLN*, 2017, pp. 199-203.

[25] S. Vychegzhanin and E. V. Kotelnikov, "Stance Detection in Russian: a Feature Selection and Machine Learning Based Approach," in *AIST (Supplement)*, 2017, pp. 166-177.

[26] A. Sasaki, K. Hanawa, N. Okazaki, and K. Inui, "Predicting stances from social media posts using factorization machines," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3381-3390.

[27] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Computational Intelligence,* vol. 29, no. 3, pp. 436-465, 2013.

[28] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist,* vol. 89, no. 4, pp. 344-350, 2001.

[29] C. Whissell, "Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language," *Psychological reports,* vol. 105, no. 2, pp. 509-521, 2009.

[30] C. E. Osgood and G. J. Suci, "& Tannenbaum, PH (1957). The measurement of meaning," *Urbana: University of Illinois Press,* vol. 335.

[31] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," *Mahway: Lawrence Erlbaum Associates,* vol. 71, no. 2001, p. 2001, 2001.

[32] K. Krippendorff, *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[33] J. Haidt, *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

[34] D. Riff, S. Lacy, F. Fico, and B. Watson, *Analyzing media messages: Using quantitative content analysis in research*. Routledge, 2019.

[35] J. Graham, J. Haidt, and B. A. Nosek, "Liberals and conservatives rely on different sets of moral foundations," *Journal of personality and social psychology,* vol. 96, no. 5, p. 1029, 2009.

[36] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," *Annual review of psychology,* vol. 54, no. 1, pp. 547-577, 2003.

[37] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process,* vol. 5, no. 2, p. 1, 2015.