A ROBUST TECHNIQUE OF FAKE NEWS IDENTIFICATION USING ENSEMBLE FEATURE SELECTION

R. Sandrilla

Research Scholar, Department of Computer Science,
Periyar University,
Salem, Tamilnadu, India.
E-Mail: sandrilla@shctpt.edu

M. Savitha Devi

Assistant Professor & Head, Department of Computer Science,
Periyar University Constituent College of Arts and Science,
Harur, Dharmapuri (Dt), India
E-Mail: savithasanma@gmail.com

Abstract

In the modern era, selecting the best feature from the high dimensional dataset with multiple variables has become a challenging task. It has become very prominent to train the model with relevant features eliminating the un-necessary feature. The traditional method feature selection methods are performing their excellence in the field of selecting the feature by creating a subset of feature from the dataset. Though it performs well, sometimes they are not helpful in learning the model by selecting single feature with single classifier. This overfits the model and leads to unnecessary confusion. Therefore, this work implements a Robust Framework of Ensemble feature selection Technique. The ensemble learning combines two or more outputs in which they may be the same type or different types, and they may or may not have been trained on the same training data. This study aims to extract the false news sub features by combining multiple subsets of features using ensemble technique. To get an efficient feature on the fake news opinion pool the Feature Score, Recursive Feature Selection and Elasticnet Feature selection has been used. Finally, feature importance has been created for each feature acts as an aggregator to select the final subset of features. The performance has been analysed with 5 classification algorithms of SVM RF, Logistic Regression, Gradient Boosting Classifier and Ridge classifier. The overall performance with accuracy has been evaluated with each classifier. The best classifier is determined by the highest accuracy rate. Our proposed implemented framework determines that Random Forest acquires a better performance in Accuracy.

Keywords: **Feature Score; Recursive Feature Elimination; ElasticNet; Feature Selection; Support Vector Machine; Random Forest; Logistic Regression; Gradientboostingclassifier; Ridge Classifier.**

1. Introduction

In machine learning everything is dealt with data. Enormous Data are streaming for every fraction of second. It is very important to lay them down into a pattern. It is necessary to keep the useful ones and remove the unnecessary ones and come up with a permanent option. The researchers specialize in conducting experiments to address these problems of high-dimensional variables and data.[1] They also want to extract significant characteristics from these high-dimensional data and variables.

It's vital to remember that we don't always use all of the features while training our model. There are lots of possibilities to improvise our model by feature correlation and non-redundant data. For this Feature selection is must and play its vital role in selection of feature in high dimensional data. It is imperative that we use fewer training samples to solve problems so Feature Selection Technique is one of the apt models to generate a subset of features to produce a better output. Feature Selection Technique has become a crucial task in the field of Machine Learning. It has become very important to achieve a best subset of feature in a huge variation of corpora. The method of selecting a subset of feature in a given dataset has become a preliminary step in any classification or regression problem. Feature Selection technique removes all irrelevant features and trains the model only with relevant subset of feature. This reduces the high rate of risk by choosing the correct set of features to train the model. Sometimes this may also lead to confusion in choosing the right subset of features, because the subset of feature chosen for some set of problem may be irrelevant to other set of problems. On the other hand, in some cases, they may discard the most relevant features in the restricted area because this match is overwhelmed by

their incompatibility Everywhere else. Many research studies are focused on finding solutions to these problems through experiments. The statistical technique of elimination of noise and redundant data was also used to identify key characteristics from high-dimensional variables and data. There is a possibility that different algorithm-based feature selections for the same dataset may yield different results, leading to different accuracy results. With ensemble-based feature selection, classification accuracy is increased since consistent feature sets are selected. On various sets of data, some models performed better and others performed worse. The predicted performance is summarized by a model's average accuracy or error. Improved model average performance resilience or reliability has also been observed as a notable benefit of ensemble approaches [17].

Since many of the previous Feature Selection methods did not work with high dimensional data, they are no longer adequate today. The motivation for writing articles on Feature Selection topics is driven by the desire to propose effective methods. The investigators examine wide range or testing processes while working on a predictive analysis project pick the ones that performed well or optimum as our final model. [2]. In comparison, locating or rating all potentially important variables is a difficult task. When developing a model, choosing the most relevant factors is frequently inefficient, especially if the factors are duplicated.[3]

Ensembles can be created by building the model many times on the training datasets and then aggregating the predictions using a reduction in the quantity like a mean for regression or a mode for classification. There must be a few key differences between each model, whether it is the stochastic learning algorithm, the differences in the training data, or the differences in the model itself [8][9]. As a result, the model's predictions will be more accurate.

Using a given dataset DT, it is first possible to select different subsets of features, say sf1, sf2, sf3… etc. Using these, one then develops an optimal feature (OF) subset from these subsets using sorted feature and feature Importance. Although the worst- and best-case performance will be moved closer to the mean performance, the mean performance will most likely be approximately the same. In effect, it smooths out the model's projected performance [4]. This is known as the "robustness" of the model's expected performance and is a minimal benefit of utilising an ensemble technique. Although an ensemble may or may not increase modelling performance over a single contributing member, it should narrow the dispersion in the model's average performance. Predictive performance improves when the variance portion of the prediction error is reduced. We explicitly employ ensemble learning to improve prediction performance, such as lower regression error or high classification accuracy. Combining various selection methods that identify traits that really are inferior individually but powerful together can boost the effectiveness of classifiers. However, most of this work only focuses on the stability of single feature selection techniques, an exception being the work of [5] which describes an example combining multiple feature selection runs.

Considering the "reliability" supplied by an ensemble as the variation of anticipated scores produced by a model on a production to ensure, such as repeated k-fold cross-validation. Instead of simply shrinking the spread of the distribution, an ensemble that minimises the variance in the error will shift the distribution. When compared to a single model, this can result in a greater overall performance. Only when there is disagreement on some inputs is it beneficial to combine the outputs of numerous predictors. Clearly, adding numerous identical predictors yields no benefit.[6]. [7] showed that utilising strategies similar to those used in ensemble methods for supervised learning, the robustness of feature ranking and feature subset selection might be increased by creating ensemble feature selection procedures.

This paper presents a Robust Technique of Ensemble Feature Selection which combines the outputs of three selection methods to generate the best attribute subset for performing binary classification on Fake News datasets. Several conversations were undertaken in order to incorporate a variety of notions into the choice of the critical metric. Finally, the system examines the key features of various machine learning models, including Random Forest (RF), Support Vector Machine (SVM), Logistic Regression, Gradient Boosting Classifier, and Ridge Classifier. Different data categorization models will have varying strengths, affecting classification outcomes.

2. Literature Review

In 2018, the novel MRFES algorithm [8] was presented. The study not only created an MCCM model for selecting features in a large dataset with multiple relevant feature sources, but it also provided a unified consistency framework achieved through co-evolving memeplexes participating in cooperative feature ensemble selection. In 2018,[9], offers a framework for determining the most insightful spectral characteristics for plant phenotyping applications utilising an ensemble feature selection tool. The ensemble uses six feature selection methods as its foundation to rate spectral features. ReliefF, SVM-RFE, and random forest as an ensemble had the best results, decreasing the hyperspectral data set from 215 to 15 features and improved the precision of discriminating salt-treated vegetation pixels from control pixels by 8.5 percent.Both in mono and bi-objective variants, an exploration of three metaheuristic-based optimization methods (ant colony optimization, particle

swarm optimization, and genetic algorithm) for choosing feature subsets in ensemble systems has been addressed. Therefore, in research, researchers used 11 classification datasets. We discovered that PSO-based ensembles had the highest rate of accuracy in both cases; additionally, the bi-objective versions of the ensembles had higher accuracy levels than their mono-objective counterparts. Furthermore, ensembles based on metaheuristics were more accurate than ensembles based on all attributes (NFS) or random feature selection (NFS).[11] At the end of 2017, an ensemble of filters and embedded methods was used instead of a single approach. The objective was to develop the stability of the feature selection process while also contributing diversity to capitalise on the positive aspects of the independent selectors and recognise their weaknesses. Using a Support Vector Machine (SVM) [12] classifier, based on how well the data was dispersed and the extracted features used,In 2017 [13] used an ensemble of unigrams, semantic, psycholinguistic, and statistical elements, as well as data mining techniques, to propose a novel automated approach for figurative content detection. To assess satiric news and ironic customer feedback, the ensembled function subset is fed as an input to a variety of binary classifiers. Furthermore, we gleaned fascinating facets of satiric news and ironic feedback., As a result, they're satirical or sarcastic. We also explored the common characteristics of humour and irony.

In 2017, [14] proposed a unique ensemble-based feature selection technique using a bi-objective evolutionary algorithm and a dynamic mating pool. Because the suggested genetic algorithm, which can quickly categorise objects, uses rough set theory and knowledge theory to build objective functions, the method provides the most accurate and informative feature subsets in 2018 [15]. Based on RNA and protein sequence knowledge, we present RPIFSE, a new ensemble approach for predicting RNA-protein interaction. The technique generates numerous data sets by selecting whether the convolution neural network is fine-tuned and the characteristics of different weights. These data sets are forwarded to the ELM base classification system. These data sets are sent to the ELM base classifier for classification, which then uses a weighted voting method to pick the most probable categories as the final prediction result.

In 2019[16] experimented Several ensemble FS methods based on voting aggregation were proposed. To aggregate the performance of simple FS methods, these methods use voting schemes like the Borda count, STV, or plurality voting. In addition, the paper ran a series of tests to assess the efficiency of FS methods using three different metrics: When all performance measures were considered, the latest FS approach based on clustered Borda count outperformed other approaches. In terms of Sen rate and prediction efficiency, the ensembles (C)E-Borda, (C)E-WpowerB, (C)E-WstairB, (C)E-min, and CE-WstepB outperformed the others. However, some of them struggled in terms of prosperity. STV schemes such as E-STV and C-STV performed poorly, suggesting that they are not appropriate for constructing ensembles.

[17] In 2020 On HDLLSS samples, researchers tested just how ensemble feature selection performs. Further specifically, they would like to understand that ensemble selection of features improves single feature selection, and if so, whichever combination strategy is preferable for ensembles selecting features, parallel or Sequential forward selection and Support vector machine (SFS-SVM) based ensemble strategy is utilised to pick optimal subset of features in terms of reducing computational time and eliminate inappropriate and chaotic information in 2020 [18]. In 2020 [19] experimented ensemble approach is used in feature selection and model creation. In the feature selection stage, five feature selection methods are used to select features, and the final features are selected by voting to create our dataset. The year 2019[20] is dedicated to a novel feature selection method called HEFS, current filter measures are used to discover an adequate set of features used in machine learning-based phishing detection. Authors define two different ensemble designs based on ranking methodologies in 2019 [21]. The fundamental distinction around them is the sequence wherein the combining and adaptive threshold stages are carried out.

3. Ensemble Feature Selection

The optimal machine learning problem approach is to take a dataset, perform extensive EDA on it, and understand many to most of the important properties of the predictors before getting as far as seriously training models on these variables. However, this is not always possible. Sometimes the dataset has too many variables. Datasets may easily have hundreds or even thousands of variables, quickly outrunning human comprehension. Other times there just isn't enough time.

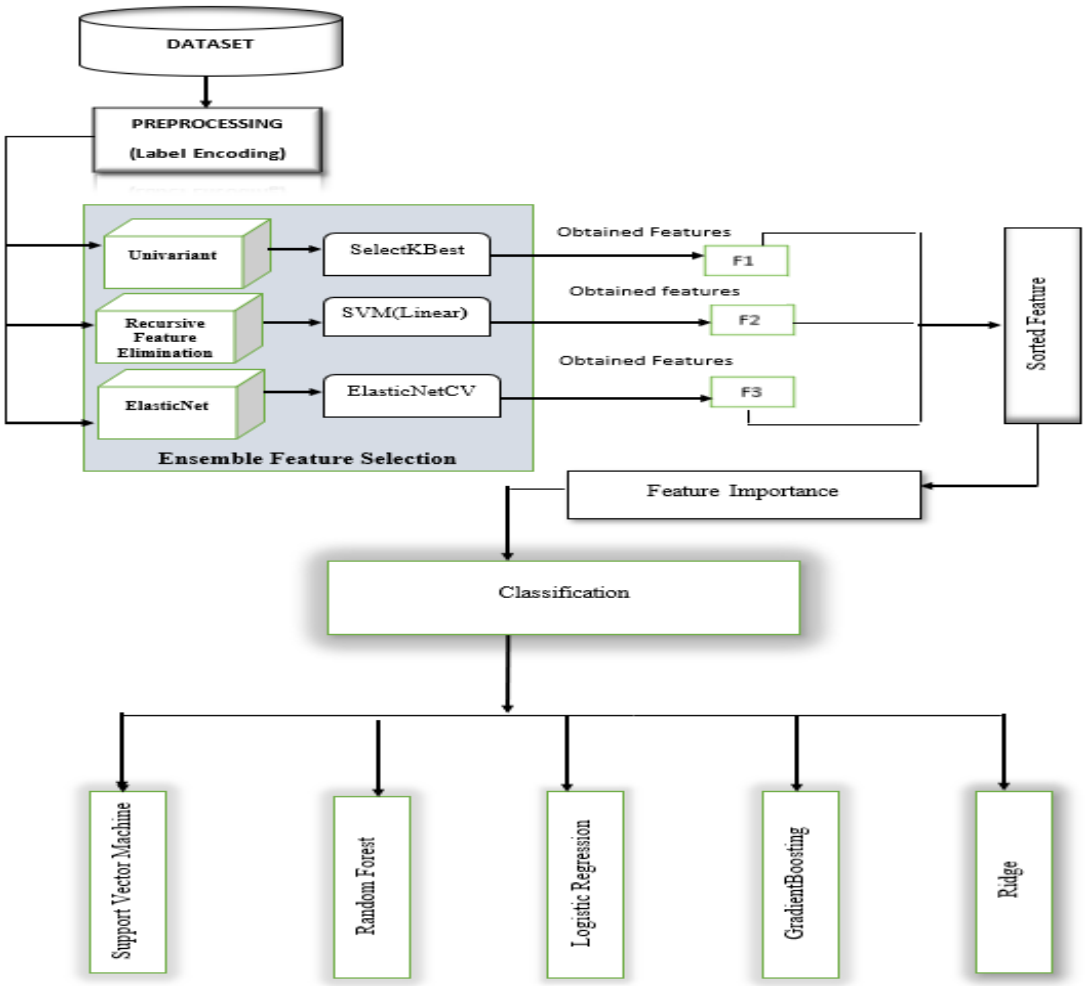

Fig. 1 . Architecture of the Proposed Ensemble Feature Selection

Feature selection is the process of turning down the number of predictor variables used by the models you build. For example, when faced with two models with the same or nearly the same score, but with the latter model using more variables, your immediate instinct should be to choose the one with fewer variables. That model is simpler to train, simpler to understand, easier to run, and less likely to be leaky. Tuning the number of parameters in a model is a natural part of data science in practice, and is something that comes naturally as part of the model-building process. While the number of features is small or you have time to sit down and consider them all, then feature selection can be a hand-driven process. In scenarios where the number of variables is overwhelming, or time is limited, automated or semi-automated feature selection can speed things up. The above Fig.1. describes the proposed architecture of the Ensemble feature selection. The novelty of the architecture is that it utilizes the Ensemble Techniques in Feature Selection phase. This Architecture considers the combination of three best feature selection Methods Univariate feature Selection, Recursive feature Elimination and ElasticNet, where each selection methods are discussed below in the sub sections. Thus, Combining the different selection methods can identify traits that really are inferior individually but powerful together which can possibly boost the effectiveness of classifiers.

3.1. Univariant Feature Selection

This is a type of statistical tests, used to find the relationship with the predictor variable. From the selected features. The most effective characteristics that support univariate statistical testing are chosen in univariate feature selection. To evaluate if there is a strong association between every attribute and the target variable, we assess them. We omit the opposite qualities after analysing the relationship between a feature and the target variable. That's why it's referred to as univariate. Each feature has a test score associated with it. Finally, all of the testing results are analysed, and the traits with the highest scores are chosen. Here the Univariant feature Selection utilizes the SelectKBest class with chi squared test (Ch^2) to find the best feature from the Fake News dataset. The Univariant-SelectKBest [18] algorithm is given as follows;

1. Create feature(F) and target variable (T).
2. Select class SelectKBest and set the score function Ch^2.
   (Where, k indicates the no of features to be selected)
3. Fit and transform the model.
4. Compare the newly achieved features with feature set values, then set values for K.
5. Choose the best values obtained in K Select features according to the k highest scores.

3.2. Recursive Feature Elimination

RFE is famous because it's simple to set up and use, and it's good at identifying which features (columns) in a training dataset are either more significant in determining the desired variable. When utilising RFE, there are two crucial configuration options: the number of features to choose from and methodology used to help select features. Each of these hyperparameters can be investigated, albeit their correct configuration does not have an important impact on the method's performance. In this study, the presence of the hyperparameter in the fake news database has no effect on the RFE feature selection efficiency. RFE is an approach that includes additional features. To select the relevant features, various machine learning algorithms are encapsulated by the wrapper style Recursive elimination method. This method works similar to filter-based method where each selected feature is assigned with a score with either largest or smallest. And then chooses a score based on the value. This is done specifically to select the relevant feature and remove the redundant feature in order to achieve the desired result. The desired targeted feature is obtained by utilising the ML algorithms in the model's core, assigning a score based on the feature importance property, and eliminating the least important features. The RFE-SVM algorithm [19] is given in four steps as

1. Train an SVM on the training set;
2. Arrange features using the weights of the derived classifier;
3. Remove features with the smallest weight;
4. Repeat the procedure with the training set constrained to the remaining features.

3.3. ElasticNet

The penalties from both the lasso and ridge approaches are used to regularise regression models in ElasticNet linear regression. The strategy incorporates the lasso and ridge regression techniques by benefiting from their flaws to better statistical model regularisation. The elastic net proposed to overcome lasso's constraints, including when high-dimensional data requires only few instances. The elastic net approach allows "n" variables to be included until saturation is reached. If the variables are highly connected groups, lasso will usually pick one from each group and ignore the others. The elastic net incorporates a quadratic expression ($||\beta||2$) in the penalty to overcome the restrictions of lasso, if used alone becomes ridge regression. The high tackles quadratic form causes the error function to become exponential. The lasso and regression strategies are used in 2 different stages of this process for determining the elastic net strategy's forecasting model. This first uncovers the ridge regression coefficients from the media dataset features, and then performs the second step by shrinking the coefficients with a lasso on a high-dimensional dataset. As a result, the coefficients are subjected to two kinds of shrinkages using this method. The double shrinkage caused by the naive edition of the elastic net results in low precision and high bias. To account for these impacts, the coefficients are resized by multiplying by (1+2). The ElasticNet-ElasticNetCV algorithm [18] is broken down into four steps:

1. Select Dataset(D), distinct F - feature and J - label, then split them into the train and test parts.
2. Define ElasticNet model by fitting alpha and train it with F and L data as F-train and L-train.
3. Cross Validate with ElasticNetCV to obtain best feature from multiple alpha value.
4. Predict the best features

3.4 Feature Importance

The method of selecting features in a sample those provide the most to determine the exact variable is known as feature importance. The relevance of each feature on the fake news data set was evaluated in this investigation. Working with a subset of characteristics rather than all of them helps in the prevention of over-fitting, reduces errors, and cuts learning time can be reduced. Feature importance assigns a value to each one of the data's features; the greater the score, the more essential or relevant the feature is to your target values. There are various types of feature significance metrics. Model-specific vs. model-agnostic features significance measurements are a meaningful distinction. Another major contrast is the importance of global vs. local features. Local measures are concerned with the contribution of features to a single prediction, whereas global measures consider all predictions. The first is relevant, for example, when we want to explain why a particular news item in the news dataset was faked.

4. Proposed Algorithm

The Algorithm of Ensemble Feature Selection Method has been described below.

Input.

D: The Input Data

F: Subset of Features f1, f2, fn.

Output:

F_{imp}: The important/relevant selected subset of Features.

Initialize the input Dataset(D) for training

Step 1: Pre-process the inputted data by label encoding.
Step 2: Apply the filter univariant on the encoded dataset (D), (Ua),
Step 3: Apply the RFE method on the dataset(D), (Rb)
Step 4: Apply the method ElasticNet on the dataset(D), (Ec)
Step 5: The subset of features is determined using the in-build class selectkbest, svm(linear), elasticnet.
Step 6: The results of the 3 methods Ua, Rb, Ec are sorted using the property sorted-feature to obtain the subset of features.
Step 7: The Sorted attribute (S) is scored with Feature importance.
Step 8: Assign a value to each one of the sorted attributes (S), using feature importance property.
Step 9: Select those attributes (F_{imp}), whose score is greater than or equal to 50% of threshold value.
Step 10: Here, the most important features ($F_{imp)}$ (where threshold=10) are selected.
Step 11: Apply the subset of feature F_{imp} to the classification model

5. Results and Discussion

The fake news data set of Facebook fact check has been used from the Kaggle repositor. In this dataset, it supports the original story of "hyper partisan Facebook page which are publishing false and misleading information. The dataset has been published in the year 2016. It contains 2283 instance and 12 features.

S.no	Feature-name
1	Page
2	Post Type
3	Rating
4	Debate
5	Post URL
6	Category
7	comment_count
8	account_id
9	reaction_count
10	Date Published
11	post_id
12	share_count

Table.1. Dataset Feature Names

The figure describes the architecture of the proposed work, in this work the fake news features have been selected with Python script by computing feature importance using ensemble feature selection model of univariate statistical analysis, recursive feature elimination, and elastic net. In this investigation, heterogeneous-based ensemble feature selection has been employed. The 3-feature selection method of univariate a statistical analysis, recursive feature elimination, and elasticnet feature selection subsets created by selectkbest, Support vector classifier with the kernel, and ElasticNetCV. The fake news features have been sorted with a subset using the sorted feature. All features are not useful to all the algorithms. So, finding out the feature importance is as much necessary to reduce the complexity of a model and for clear interpretation. Choosing the best subset increases the performance accuracy of the model. Here the N no of features obtained from the sorted feature are collected. The N no of features is stored in a Data frame. The model is trained with the stored feature and the targeted variable. The average feature importance score is set for each feature. Finally removes the features lowest to the targeted feature. The higher the score is considered as the most relevant feature from the subset. These have been selected

by using feature importance property. The performance of the proposed model has been efficiently measured by using 5 different classifier models of SVM RF, Logistic Regression, Gradient Boosting Classifier, and Ridge classifier. The sorted feature subset has been described in the table. Based on the subset creation and the sorted feature model the feature importance has been created.

S.no	Feature
1	Platform
2	Post Type
3	Rating
4	Debate
5	Post URL
6	Category
7	Comment_count

Table.2. Sorted Feature Model

The feature importance of the proposed model has been described in the Table.2. The highest importance has been created for the feature of page, post Type, rating and Debate. The least importance has been created for share count.

Feature	Importance
Platform	8
Post Type	8
Rating	8
Debate	8
Post URL	7.7
Category	7
comment_count	7
account_id	6.8
reaction_count	6.7
Date Published	6.4
post_id	3.6
share_count	2.8

Table.3. Result of Feature Importance

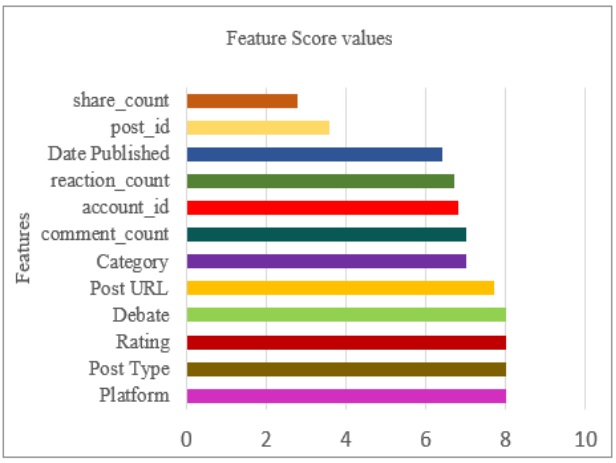

Fig.2. Top 10 features selected from data

The overall accuracy has been measured for each classifier. The final performance with each iteration is discussed below with the corresponding table and figures. The Fig.3. and Table.4. describes the Classifier RandomForestClassifier(n_estimators=300), with the final performance: 0.8749 (+/- 0.0265).

S. no	Training Accuracy	Testing Accuracy
1	0.9000	0.9109
2	0.9100	0.8218
3	0.9018	0.8911
4	0.8800	0.8600
5	0.8900	0.8900
6	0.8810	0.9000
7	0.9078	0.8384
8	0.8000	0.8788
9	0.0820	0.8889
10	1.0000	0.8749

Table.4. Training and Testing Accuracy of Random Forest

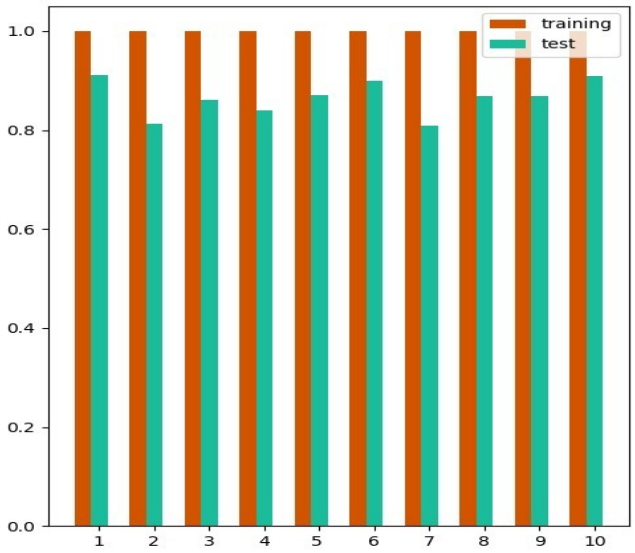

Fig. 3. Accuracy Graph of Random Forest

The Fig.4.and Table.5. describes the Classifier GradientBoosting(n_estimators=300), with final performance: 0.8687 (+/- 0.0360).

S.no	Training Accuracy		Testing Accuracy
1	0.8129		0.9010
2	0.8392		0.8119
3	0.8263		0.8713
4	0.8376		0.8900
5	0.8701		0.8700
6	0.8732		0.9100
7	0.7789		0.7879
8	0.8690		0.8788
9	0.8867		0.8687
10	0.8679		0.8687

Table.5. Training and Testing Accuracy GradientBoostingClassifier

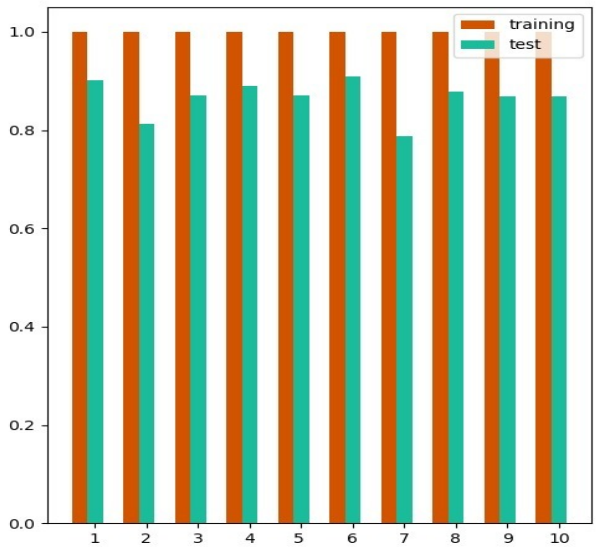

Fig. 4. Accuracy Graph of GradientBoostingClassifier

The below Fig.5.and Table.6. describes the Classifier Logistic Regression

S.no	Training Accuracy	Testing Accuracy
1	0.8497	0.8515
2	0.8530	0.8416
3	0.8441	0.8317
4	0.8376	0.8800
5	0.8632	0.9100
6	0.8476	0.8100
7	0.8500	0.7374
8	0.8389	0.8788
9	0.8533	0.8384
10	0.8444	0.8283

Table.6. Training and Testing Accuracy of Logistic Regression

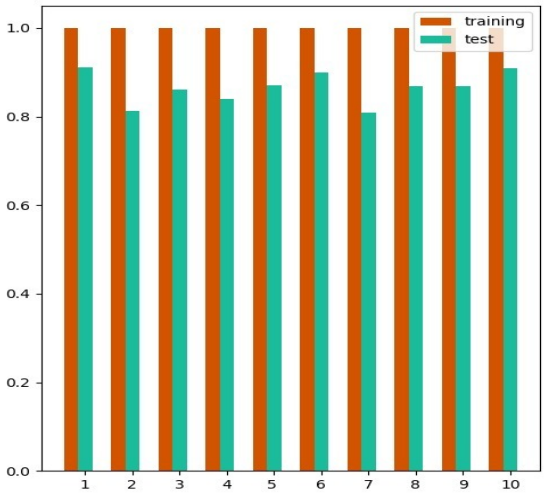

Fig. 5. Accuracy Graph of Logistic Regression

For the classifier Logistic Regression, the final performance: 0.8283 (+/- 0.0445), The Fig.6. and Table.7. describes the Classifier RidgeClassifier

S.no	Training accuracy	Testing Accuracy
1	0.8229	0.8614
2	0.8352	0.8218
3	0.8263	0.8020
4	0.8276	0.8300
5	0.8409	0.8500
6	0.8309	0.7900
7	0.8367	0.7576
8	0.8333	0.8687
9	0.8322	0.8586
10	0.8333	0.8182

Table.7. Training and Testing Accuracy of Ridge classifier

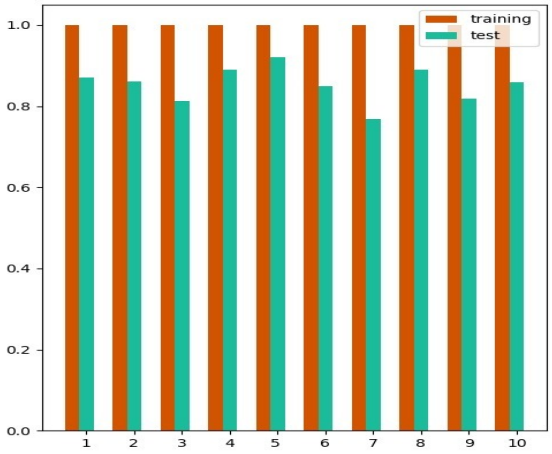

Fig. 6. Accuracy Graph of Ridge Classifier

For the classifier Ridge, the final performance:0.8182 (+/- 0.0337), The Fig 7.and Table.8. describes the Classifier Support Vector Machine (SVM).

S.no	Training accuracy	Testing accuracy
1	0.8641	0.8713
2	0.8675	0.8614
3	0.8608	0.8119
4	0.8576	0.8900
5	0.8643	0.9200
6	0.8521	0.8500
7	0.8689	0.7677
8	0.8578	0.8889
9	0.8633	0.8182
10	0.8622	0.8586

Table.8. Training and Testing Accuracy of SVM.

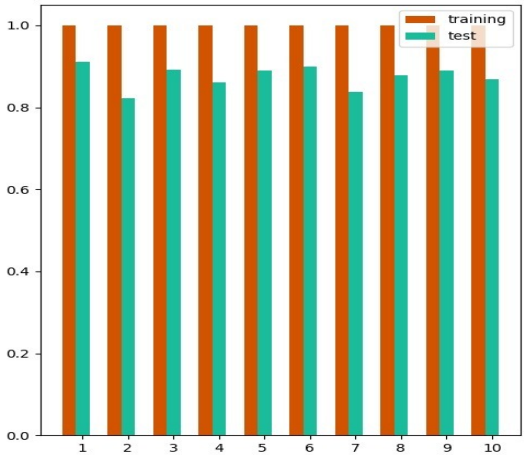

Fig. 7. Training and Testing Accuracy of SVM

The Classifier Linear Support Vector Machine (SVM), performance is predicted and the obtained final performance is 0.8586 (+/- 0.0422)

Classifier	Final Performance
GBC	0.8687 (+/- 0.0360)
RF	0.8749 (+/- 0.0265)
LR	0.8283 (+/- 0.0445)
Ridge	0.8182 (+/- 0.0337)
SVM	0.8586 (+/- 0.0422)

Table.9. Overall Accuracy Performance

The performance was evaluated using Support Vector Machine Random Forest, Logistic Regression, GradientBoosting Classifier, and Ridge classifiers. The overall performance in terms of accuracy was evaluated for each classifier using Feature importance. The result shows that the overall performance has been higher of 87.49% for Random Forest with the proposed Ensemble Feature Selection.

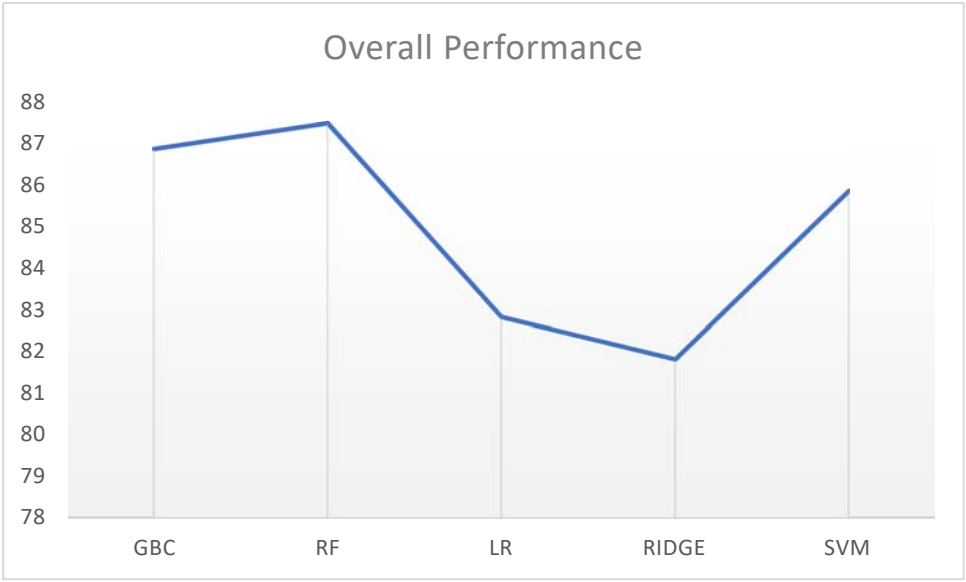

Fig. 8 . Overall Performance Analysis

6. Conclusion

A machine learning ensemble is a model that combines the predictions of two or more models. Ensemble members serve as role models and contribute to the ensemble's success. They might be of the same or different types, and they could have been trained on the same or different sets of data. The false news sub features were derived using ensemble feature selection in this study. Feature selection methods such as elastic net, recursive feature selection, and feature score were all used. If embedded feature selection is necessary, a wrapper technique for feature selection aggregation on the ensemble model has been investigated, such as feature importance. To develop an efficient feature on the fake news opinion pool, this study used the feature score, recursive feature selection, and elastic net feature selection. The feature importance has been computed for each feature in the dataset. The performance was evaluated using Support Vector Machine, Random Forest, Logistic Regression, Gradient Boosting, and Ridge Classifiers. The overall performance in terms of accuracy was evaluated for each classifier and feature importance. The result shows that the overall performance has been higher of 87.49% for random forest with the proposed ensemble feature selection.

References

[1]. Wang XD; Chen RC, Yan F; et al. Fast adaptive K-means subspace clustering for high-dimensional data, IEEE Access. 2019;7:42639–51.
[2]. V. Bolón-Canedo; N. Sánchez-Maroño; A. Alonso-Betanzos; Feature selection for high-dimensional data, Springer, 2015.
[3]. I. Guyon; A. Elisseeff; An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[4]. Dunne; K. Cunningham; P. Azuaje, F. Solutions to instability problems with sequential wrapper-based approaches to feature selection, Technical report TCD-2002-28. Dept. of Computer Science, Trinity College, Dublin, Ireland (2002)
[5]. P. Cunningham; J. Carney, Diversity versus quality in classification ensembles based on feature selection, in: R. Lpez de Mntaras, E. Plaza (Eds.), Proc. European Conference on Machine Learning (ECML), LNAI 1810, 2000, pp. 109–116.
[6]. D. Opitz; Feature selection for ensembles, in: Proc. 16th Nat. Conf. on Artificial Intelligence, AAAI Press, 1999, pp. 379–384.
[7]. Y. Saeys; T. Abeel; Y. Van der Peer, Robust feature selection using ensemble feature selection techniques", in: W. Daelemans, et al. (Eds.), Proc. European Conference on Machine Learning (ECML PKDD), LNAI 5212, 2008, pp. 313–325
[8]. Ding, Weiping; Chin-Teng Lin; and Witold Pedrycz. Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus MapReduce, IEEE transactions on cybernetics 50, no. 2 (2018): 425-439.
[9]. Du; Xudong; Wei Li; Sumei Ruan; and Li Li. CUS-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection, Applied Soft Computing 97 (2020): 106758.
[10]. Moghimi; Ali; Ce Yang; and Peter M. Marchetto. Ensemble feature selection for plant phenotyping: A journey from hyperspectral to multispectral imaging, IEEE Access 6 (2018): 56870-56884.
[11]. Ravi; Kumar; and Vadlamani Ravi. "A novel automatic satire and irony detection using ensembled feature selection and data mining", Knowledge-based systems 120 (2017): 15-33.
[12]. Santana, Laura Emmanuella A. dos S; and Anne M. de Paula Canuto. Filter-based optimization techniques for selection of feature subsets in ensemble systems, Expert Systems with Applications 41, no. 4 (2014): 1622-1631.
[13]. Seijo-Pardo; Borja; Iago Porto-Díaz; Verónica Bolón-Canedo; and Amparo Alonso-Betanzos; Ensemble feature selection: homogeneous and heterogeneous approaches, Knowledge-Based Systems 118 (2017): 124-139.
[14]. Wang; Lei; Xin Yan; Meng-Lin Liu; Ke-Jian Song; Xiao-Fei Sun; and Wen-Wen Pan. Prediction of RNA-protein interactions by combining deep convolutional neural network with feature selection ensemble method, Journal of theoretical biology 461 (2019): 230-238.
[15]. Tsai; Chih-Fong; and Ya-Ting Sung; Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches, Knowledge-Based Systems 203 (2020): 106097.
[16]. Ahmad; Ashfa;, Shahid Akbar; Maqsood Hayat; Farman Ali, and Mohammad Sohail. Identification of antioxidant proteins using a discriminative intelligent model of k-spaced amino acid pairs-based descriptors incorporating with ensemble feature selection, Biocybernetics and Biomedical Engineering (2020).
[17]. G. Brown; Ensemble learning; in: "Encyclopedia of Machine Learning", Springer, 2011, pp. 312–320.
[18]. Kaggle-Feature Selection and ElasticNet Retrieved from https://www.kaggle.com/cast42/feature-selection-and-elastic-net
[19]. Guyon, Weston, Barnhill, and Vapnik, "Gene selection for cancer classification using support vector machines," MACHLEARN: Machine Learning, vol. 46, (2002).

Authors Profile

Ms. R. Sandrilla: completed her Master of Computer Applications from Mount Carmel College, Bangalore and Master of Philosophy in Computer Science from KMG Arts and Science College, Gudiyattam. Currently she is working as Assistant Professor in the Department of Computer Science at Sacred Heart College (Autonomous) Tirupattur and pursuing her Ph.D. degree in Periyar University, Salem. She is having 11 years of Teaching experience and 6 years of Research experience. Her research areas include Web Mining, Machine Learning, Natural Language Processing, Artificial Intelligence.

Dr.M. Savitha Devi: Currently working as an Assistant Professor and Head, Department of Computer Science, Periyar University Constituent College of Arts and Science, Harur. She has 18 years of Teaching experience and 10 years of Research experience. She has published more than 25 research papers in National, International journals and Conference proceedings. She has published books related to network security. Her research areas include E-Content Development, Network Security, Web Mining and Machine Learning. She is currently acting as a University Swayam Mentor.