International Conference on Innovative Research in Engineering, Management and Sciences
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 4 | Issue 9 | ISSN : 2456-3307

Review on Feature Selection for the Analysis of Human Activities and Postural Transitions on Smart Phone.

Kamatchi Priya*[1], Punyasri H[2]

*[12]Computer Science and Engineering, New Horizon College of Engineering, Bangalore, India
priya.1211@gmail.com[1]
sripunya9966@gmail.com[2]

ABSTRACT

Most of the data in real world used for prediction have many features which are relevant and irrelevant. While performing prediction with large number of features, it will depreciate the performance in the terms of accuracy, space and time. To address this, features which influence the target prediction has to considered. Features which are irrelevant and redundant has to be eliminated. For the purpose, there are many algorithms. For high dimensional data like smartphone based recognition of human activities and postural transitions, requires feature selection. Many feature selection methods are applied and compared to get the best performance in terms of accuracy. It is found that Recursive feature elimination outperform others.
Keywords : Recursive feature elimination, feature selection, irrelevant, redundant, prediction, accuracy.

I. INTRODUCTION

Feature selection is called Variable selection or Attribute selection. It is referred as automatic selection of the attributes in data that are most relevant to the predictive output. Feature selection is the process of obtaining subset from the original feature set as per certain feature selection criterion, which selects all the relevant features. It also plays role in compressing the data processing scale, where all the redundant and irrelevant features are removed from data that do not contribute to the accuracy of predictive model or which may decrease the accuracy of the model. Redundant features are the features which add no relevant information to your other features, because they are correlated or because they can be obtained by combination of other features. Having such features may not affect information wise but affects the training and classification some times. Feature selection methods help in mission to create an accurate predictive model. They also help in selection of features that give better accuracy. Good FS results reduces learning time, improve learning accuracy, and simplify learning results[1]. Three main objectives of feature selection are: (1) Feature selection reduces the overfitting i.e., less redundancy of data, (2) Feature selection also improves the accuracy i.e., less misleading data, (3) Feature selection reduces the training time i.e., if data is less then algorithms train faster. There are many different types of methods of feature selection like Boruta, Feature Importance, Genetic Algorithm(GA), Information Value And Weight Of Evidence(IV and WOE), Lasso regression, recursive feature elimination(RFE), principal component analysis(PCA), Simulates Annealing(SA), Stepwise Forward And Backward Selection, Univariate Selection.

In this paper , Section 1 is a literature survey on the methods of feature selection and classifiers.

Section 2 describes in detail about the cause to choose feature subset selection over feature extraction, it also explains about the RFE algorithm and logistic regression classification method. Section 3 is all about the results where accuracies of experiments of each FS method is compared and the one which is efficient among the methods are considered.

II. LITERATURE SURVEY

In this section, it is mainly discussed on three methods that is feature subset selection, feature extraction and classification.

A. Feature-Subset-Selection

Feature subset selection(FSS) is a technique used to select variables from huge data which improve the accuracy of the model. In addition, the best FSS method can also reduce the cost of feature measurement. FSS plays major role in data mining and machine learning fields. A good Feature subset selection algorithm can efficiently and effectively remove all the irrelevant and redundant features and take in consideration of only those features which are important in the prediction of the target.

In feature subset selection there are different techniques ,(1)Boruta-Boruta is the technique which achieves supreme importance when a data set comprised of several variables is given for model building. Particularly when one is interested in understanding the mechanisms related to the variable of interest, this technique is used. . Boruta is a simple algorithm used for feature selection, it find all features which are either strongly or weakly relevant to the decision variable and it is an easy to use package as there aren't many parameters to remember. These make the advantages of Boruta. The main disadvantages of Boruta are data set with missing values should not be used to check important variables using Boruta, if so it will throw errors and Boruta can be used only on classification/regression. Boruta is well suited for biomedical applications where one might be interested to determine which human genes(features or attributes)are connected in some way to a particular medical condition(target variable or predicted output)[2]. (2)Feature Importance-In feature importance model we importance of feature is measured by calculating rise in the model's prediction error after permuting the particular feature. A feature is "unimportant" if shuffling its values leaves the model error unchanged, because in this case the model ignored the feature for the prediction. Feature importance gives highly compressed and global insight into the model's behavior. A positive aspect of using the error ratio instead of the error difference is that the feature importance measurements are comparable across different problems. The importance measure automatically takes into account all interactions with other features. Permutation feature importance does not require retraining the model. The disadvantages of feature importance are it is very unclear whether you should use training or test data to compute the feature importance, permutation feature importance is linked to the error of the model. You need access to the true outcome. If someone only provides you with the model and unlabeled data but not the true outcome you cannot compute the permutation feature importance. The permutation feature importance depends on shuffling the feature, which adds randomness to the measurement. When the permutation is repeated, the results might vary greatly. Repeating the permutation and averaging the importance measures over repetitions stabilizes the measure, but increases the time of computation. If features are correlated, the permutation feature importance can be biased by unrealistic data instances. Another tricky thing is adding a correlated feature can decrease the importance of the associated feature by splitting the importance between both features. Feature importance is used in fitting a support vector machine model to predict the number of rented bikes, given weather conditions and calendar information and random forest model is fitted to predict cervical cancer[3].(3) Genetic Algorithm(GA)- The main idea of genetic algorithm is to combine the different solutions generation after generation to extract the best genes (features) from each one. That way it creates new and more fit individuals. This algorithm work on population of the individuals in order to produce better approximations. At every generation new population is produced/created by selecting individuals as per to their level of fitness. And these are recombined together using operators from natural genetics and the offspring may also undergo mutation. This algorithm leads to the evolution of population that suits their environment in better way than those individuals that they were created from just as in nature adaptation. Here for the prediction of model accuracy fitness values are used. The advantage of this technique over others is that it allows the best solution to emerge from the best of the prior solutions. An evolutionary algorithm which improves the selection over time. One issue with using GAs for feature selection is that the optimization process can be very aggressive and there is potential for the GA to overfit to the predictors. GA is applied in Hyper-tuning parameters, finding the maximum (or minimum) of a function, or searching for the correct neural network architecture (neuro evolution)[4]. (4)Information Value And Weight Of Evidence-Weight of evidence (WOE) and Information value (IV) are simple, yet powerful techniques to perform variable transformation and selection. These concepts have huge connection with the logistic regression modelling technique. The weight of evidence tells the predictive power of an independent variable in relation to the dependent variable. Since it evolved from credit scoring world, it is generally described as a measure of the separation of good and bad customers. "Bad Customers" refers to the customers who defaulted on a loan. and "Good Customers" refers to the customers who paid back loan. This algorithm Handles missing values and handles outliers. The transformation is based on logarithmic value of distributions. This is aligned with the logistic regression output function. Here there is no need for dummy variables. By using proper binning technique, it can establish monotonic relationship (either increase or decrease) between the independent and dependent variable. Also, IV value can be used to select variables quickly. IV and WOE have demerits like loss of information (variation) due to binning to few categories. It is a "univariate" measure so it does not take into account correlation between independent variables. It is actually easy to overfit or manipulate effect of the variables according to how categories are created. IV and WOE is widely used in credit scoring to measure the separation of good versus bad customers[5]. (5) Lasso Regression-LASSO refers to least absolute shrinkage and selection operator regression is a type of regularization method. It basically imposes a cost to having large weights (value of coefficients). And it is called L1 regularization, because the cost added, is proportional to the absolute value of weight coefficients. As a result, in the process of shrinking the coefficients, it eventually reduces the coefficients of certain unwanted features all the to zero. That is, it removes the unneeded variables altogether. So effectively, Lasso regression can be considered as a variable selection technique as well. LASSO is that it is better than the usual methods of automatic variable selection such as forward, backward and stepwise - all of which can be shown to give wrong results. The results from LASSO are much better. The biggest disadvantage of LASSO is that it is automatic; therefore, it has problems. The biggest problem is that it lets you (the data analyst) avoid thinking. Other, lesser problems. It can also produce models that make no sense. It ignores nonsignificant variables that may, nevertheless, be interesting or important. It doesn’t follow the hierarchy principle. Lasso regression are powerful techniques generally used for creating parsimonious models in presence of a 'large' number of features[6]. (6)Principal component analysis- It is a basic technique well-suited for this problem which is called as PCA which tries to find the directions of most variation in your data set. PCA gives us the transformed feature set. Assume that the dimensionality of the feature set is larger than just two or three. Using PCA we can identify what are the most important dimensions and just keep a few of them to explain most of the variance we see in our data. Hence we can drastically reduce the dimensionality of the data and make EDA(exploratory data analysis) feasible again. Moreover, it will also enable us to identify what the most important variables in the original feature space are, that contribute most to the most important PCs. Intuitively, one can imagine, that a dimension that has not much variability cannot explain much of the happenings and thus is not as important as more variable dimensions. PCA removes correlated features where after implementing the PCA on your dataset, all the Principal Components are independent of one another. There is no correlation among them. With so many features, the performance of your algorithm will drastically degrade. PCA is a very common way to speed up your Machine Learning algorithm by getting rid of correlated variables which don't contribute in any decision making. The training time of the algorithms reduces significantly with less number of features.

Overfitting mainly occurs when there are too many variables in the dataset. So, PCA helps in overcoming the overfitting issue by reducing the number of features.

PCA transforms a high dimensional data to low dimensional data so that it can be visualized easily. The disadvantage of PCA are after implementing PCA on the dataset original features will turn into Principal Components. Principal components are not readable as original features. Data must be standardized before principal component analysis. If we don't select the number of components properly it may miss some information. Data compression, Image processing, visualization, exploratory data analysis, pattern recognition and time series prediction is done using PCA. PCA can be viewed as a special scoring method under the SVD algorithm. [7]. (7) Simulated Annealing- Simulated annealing is a global search method that makes small random changes (i.e. perturbations) to an initial candidate solution. If the performance value for the perturbed value is better than the previous solution, the new solution is accepted. From this, a sub-optimal solution can be accepted on the off-change that it may eventually produce a better solution in subsequent iterations. Simulated annealing is known for the better behavior than the naive local search algorithm because it enables us to leave the local optima to find the best answers. A disadvantage is that the SA algorithms are computation intensive. There exist faster variants of the Simulated annealing, but these are not as quite easily coded and widely used[8]. (8) Step forward and Step backward feature selection- In step forward feature selection starts each individual feature is evaluated, and selects the one which results in the best performance. Further, the second feature is selected by all possible combinations of that selected feature and by evaluating the subsequent feature, and so on, until the required predefined number of features is selected. Step backward feature selection is closely related, and as you may have guessed starts with the entire set of features and works backward from there, removing features to find the best subset. The primary advantage of stepwise regression is that it's computationally efficient. It's faster than other automatic model-selection methods. It provides vital information about the quality of predictors by observing the order in which variables are added or removed. Both the methods are potentially very computationally expensive. These methods may take too long to be at all useful, or may be totally infeasible. That said, with a dataset of accommodating size and dimensionality, such an approach may well be your best possible approach. Have to keep in mind that an optimized set of selected features using a given algorithm may or may not perform equally well with a different algorithm. Collinearity is usually a major issue. Excessive collinearity may cause the program to dump predictor variables into the model. Some variables (especially irrelevant variables) may be removed from the model, when they are deemed important to be included. These can be manually added back in[9]. (9) Univariate Selection- Statistical tests can be used to select those features that have the strongest relationship with the output. CHI SQUARED test is one such statistical test which is used to select the best features. This algorithm examines each and every feature separately to determine the strength of relationship of feature as per the response variable. Univariate feature selection is in general best to get a better understanding of the data, its structure and characteristics. This method gives better understanding of the data. There are lot of different options for univariate selection. Univariate filter methods are ideal for removing constant and quasi-constant features from the data. The disadvantages of univariate selection are it can work for selecting top features for model improvement in some settings, but since it is unable to remove redundancy (for example selecting only the best feature among a subset of strongly correlated features), hence this task is better left for other methods. One of the major disadvantage of univariate filter methods is that they may select redundant features because the relationship between individual features is not taken into account while making decisions. Univariate selection is mainly used in image processing. It has industrial applications and text mining[10]. (10) Recursive Feature Elimination- The Recursive feature elimination is a method which recursively removes the attributes. And build a model on the attributes that remain and makes use of the accuracy of the model in order to find which attributes contribute the most in predicting target. It can be an effective and relatively efficient technique for reducing the model complexity be removing irrelevant predictors. The principal drawback of RFE is the huge time consumption. It is used in Microarray datasets like colon cancer, lymphoma3 cancer, and Cancer of Unknown Primary (CUP). RFE is also applied mainly in gene-related studies, specially gene recognition and disease diagnosis. The above are the methods or techniques of feature subset selection where some of them outperform others and some may not[11].

B. Feature extraction

Feature extraction is defined as process of reduction in the data dimension where starting set of raw data is reduced to more manageable groups or sets. A characteristic of these large data sets is a large number of variables which need lot of resources for computing. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set.

When there is need to reduce the number of resources needed for processing without the loss of precious data then the process of feature extraction is very advantageous. Feature extraction can also reduce the amount of redundant data for a given analysis. Also, the reduction of the data and the machine's efforts in building variable combinations (features) facilitate the speed of learning and generalization steps in the machine learning process.

Practical uses of Feature Extraction are Autoencoders: The purpose of autoencoders is unsupervised learning of efficient data coding. Feature extraction is used here to identify key features in the data for coding by learning from the coding of the original data set to derive new ones. Bag-of-Words: A technique for natural language processing that extracts the words (features) used in a sentence, document, website, etc. and classifies them by frequency of use. This technique can also be applied to image processing. Image Processing: Algorithms are used to detect features such as shaped, edges, or motion in a digital image or video. There are many varieties of methods for managing texture are developed The other methods which also come under feature extraction are independent component analysis, isomap, kernel principal component analysis, latent semantic analysis, partial least squares, principal component analysis, multifactor dimensionality reduction, non-linear dimensionality reduction, multilinear principal component analysis, multilinear subspace learning, semidefinite embedding, autoencoder[12].

C. Classification

Classification algorithms can be applied on unstructured or structured data. Classification is a technique where data is categorized into number of classes. Identifying the class to which a new data will fall under is the main aim of classification problem.

There are different types of classification methods. (1) Logistic regression : Logistic Regression is a machine learning classification algorithm i.e., CLASSIFIER. Produces results in a binary format .So the outcomes are discrete/categorical such as: 0 or 1, yes or no ,true or false, high and low. Logistic regression is designed for this purpose (classification), and is most useful for understanding the influence of several independent variables on a single outcome variable. Works only when the predicted variable is binary, assumes all predictors are independent of each other, and assumes data is free of missing values. (2) Naïve Bayes: Naive Bayes algorithm is based on Bayes' theorem. With the assumption of independence between every pair of features Naive Bayes classifiers work well in many real-world situations i.e., spam filtering and document classification. These are fast compared to more sophisticated methods. It requires small amount of training data to estimate the necessary attributes. The main disadvantage of Naive Bayes is it is known to be a bad estimator. (3) Stochastic gradient descent: It is a simple and very efficient method to fit linear models. When the number of samples is very large it is specifically very useful. It supports different loss functions and penalties for classification. The advantages of this algorithm are efficiency and ease of implementation. The disadvantages are it requires a number of hyper-parameters and it is sensitive to feature scaling.(4) K-Nearest Neighbours: Neighbours is a lazy learning. It is so called because it simply stores instances of the training data. And does not attempt to construct a general internal model. From a simple majority vote of the k nearest neighbours of each point classification is computed. This algorithm is simple to implement, robust to noisy training data, and effective if training data is large. The disadvantages are need to determine the value of K. Since it needs to compute distance of each instance to all the training samples cost of computation is high .(5)Decision Tree: Decision tree is simple to understand. It is also easy to visualize. Decision tree requires little data . The disadvantage of this algorithm is tree can create complexity. Created tree can be unstable because of small variations in the data. This can lead to the creation of completely different tree sometimes.(6)Random forest: Random forest classifier is named as meta-estimator. It is so called because it fits a number of decision trees on different sub samples of datasets. It uses average to improve the accuracy prediction of model. And hence controls over fitting. The size of sub sample is same as the size of original input always. But the samples are drawn with replacement. In terms of over fitting Random forests outperform decision trees in most of the cases. This classifier has slow real time prediction. It is difficult to implement and complexity of algorithm is high.(7) Support vector machine: SVM is a represents the training data as points in space separated into categories by a clear gap that is as wide as possible. Further new examples are mapped into that same space and then predicted to belong to a category based on side of the gap they fall. This classifier is better in high dimensional spaces. And then it uses a subset of training points in the decision function so that it is also memory efficient. The algorithm does not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation[13].

III. METHODOLOGY

Feature subset selection indeed reduces the dimension whereas feature extraction adds on the dimension. Anyhow in this paper our motive is to classify and as well as select the prominent features which truly plays vital role in predicting the target. It improves the performance of learner's interpretability. Hence we choose to implement feature subset selection algorithm.

Although a large number of FSS algorithms have been proposed, there is no single algorithm which performs uniformly well on all feature selection problems. Many experiments have confirmed that there could exist significant differences of performance (e.g., accuracy) among different FSS algorithms for a given data set. That means for a given data set, some FSS algorithms outperform others. This raises a practical and very important question that which FSS algorithms should be picked up for a given data set? The common solution is to apply all candidate FSS algorithms to the given data set, and choose one with the best performance by the cross validation strategy. However, this solution is quite time-consuming especially for high dimensional data. For the purpose of solving this problem in a better way, in this paper, an FSS algorithm automatic recommendation method is proposed. Based on the study and comparing different feature selection algorithms we could find that RECURSIVE FEATURE ELIMINATION ALGORITHM(RFE) would help us to select the prominent features for predicting the target efficiently. So, in this paper let us learn about the Recursive Feature Elimination in a detailed manner.

A. RECURSIVE FEATURE ELIMINATION

Recursive feature elimination is a backward compatible way of feature selection. It makes use of the accuracy to find which attributes contribute lot for predicting the target attribute. Recursive feature elimination works by recursively removing predictors/attributes and then building a model on those predictors/attributes that remain. In this technique initially model is built based on the entire set of features(predictors/attributes) and compute score for each predictor. The least important predictors(features or attributes) are then removed, the model is re-built, and importance scores are computed again. In practice, the analyst specifies the number of predictor subsets to evaluate as well as each subset's size. Therefore, the subset size is a tuning parameter for RFE. [11].

Steps that tells how RFE works are as follows are initially start with all the features and then build a model. Then calculate the accuracy score for the model and check the accuracy score. In the next step, remove or eliminate one feature and build the model again. Now see how much of variation is caused in accuracy score. And then based on this decide whether to keep the feature or remove the feature. If the accuracy score has increased because of eliminating the feature then eliminate the particular feature. If the accuracy score has decreased because of eliminating the feature then retain the feature in the model.

Hence RFE can be an effective and relatively efficient technique for reducing the model complexity be removing irrelevant predictors.

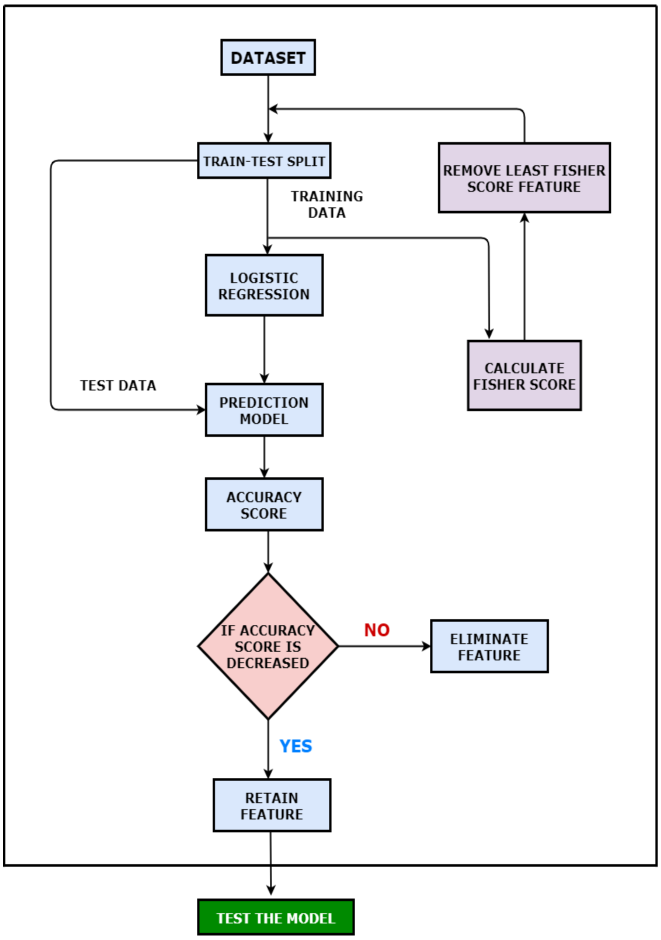

Figure.1 Recursive Feature Elimination

B. LOGISTIC REGRESSION

Logistic Regression is a machine learning classification algorithm i.e., CLASSIFIER. Produces results in a binary format .So the outcomes are discrete/categorical such as: 0 or 1, yes or no ,true or false, high and low. The assumptions of logistic regression are binary logistic regression requires the dependent variable to be binary. For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome. Only the meaningful variables should be included. The independent variables should be independent of each other. That is, the model should have little or no multicollinearity. The independent variables are linearly related to the log odds[1].

Figure 2 Logistic Regression

IV. RESULTS

In this paper, four different methods of feature selection are applied on Smartphone-Based Recognition of Human Activities and Postural Transitions data set and compared the accuracy of one with another. For each method ten experiments are performed and their Mean Accuracy(MA) is calculated and as well as Variance(V) and Variance Score(VS) are also calculated. Mean accuracy is the mean of accuracies of the ten experiments performed.

VarianceScore(VS)=Mean Accuracy(MA)/Variance(V)

In the dataset "smartphone-based recognition of human activities and postural transitions" :

No of predictors/attributes-7767
No of records-561

Table 1 Accuracy of FS Methods

METHODS	Number of Features before FS	Number of Features after FS	Number of Experiments	Mean Accuracy(%)	Variance	Variance Score
UNIVARIATE SELECTION	7767	7667	10	64	5.8	11.03
PRINCIPAL COMPONENT ANALYSIS	7767	4	10	86	3.3	26.06
RECURSIVE FEATURE ELIMINATION	7767	7444	10	94	2.2	47.72
FEATURE IMPORTANCE	7767	7588	10	75	4.4	17.04

Here the method which has high mean accuracy and less variance gives high variance score. Hence the method which satisfies this criteria outperform other methods.

In this paper, four methods of FS were compared and ten experiments for each method were performed and their mean accuracy is considered. Among the four, recursive feature elimination gives the maximum mean accuracy and minimum variance where its variance score has outperformed other methods. Since RFE gives the best and high variance score it can be said that Recursive Feature Elimination is a better method for the removal of irrelevant features and prediction of the target attribute rather than the other methods.

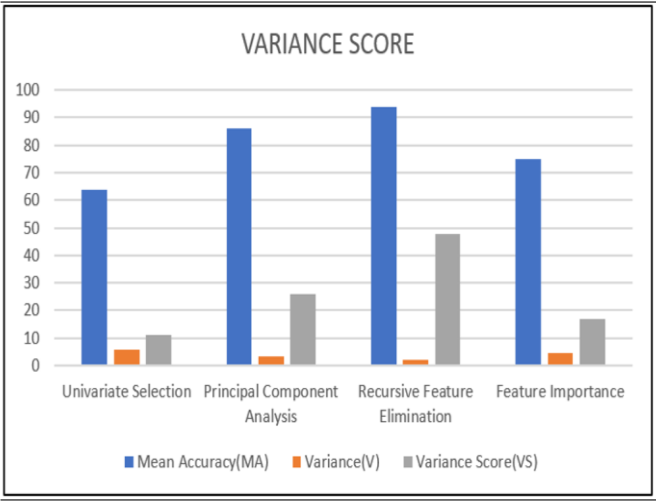

Figure 3 Variance Score

In the above graph, RFE has got the high mean accuracy and variance score compared to others which tells that RFE has the better performance.

V. CONCLUSION

Recognition of Human Activities and Postural Transitions can be done using Smartphone, but the data collected from the smartphone may or may not influence the Human Activities or Postural Transitions. To understand the features which influence Human Activities and Postural Transitions, it is important to identify the features which influence the classification process. For the purpose, two activities are performed viz., identifying relevant features and eliminating redundant features using

RFE algorithm. Ten experiments were conducted in each of the algorithm considered and it is found that the mean accuracy of RFE is better than others state of art methods. RFE is more reliable because the accuracy score across the experiments were consistent and is evident from the variance graph. Variance score is a performance measure which illustrates the best algorithm and RFE is proved to be better than others. The problem with RFE is it consumes more time since it iterates. To overcome this limitation, linear support vector machines can be combined with recursive feature elimination strategy.

VI. REFERENCES

[1] Cai, Jie, Jiawei Luo, Shulin Wang, and Sheng Yang. "Feature selection in machine learning: A new perspective." Neurocomputing 300 (2018): 70-79.

[2] How to perform feature selection (i.e. pick important variables) using Boruta Package in R ? GUEST BLOG, MARCH 22, 2016.

[3] Heaton, Je. "Feature Importance in Supervised Training." Predictive Analytics and Futurism (2018): 22.

[4] Feature Selection Using Genetic Algorithms (GA) in R by Pablo Casas Jan. 23, 19.

[5] Chen, Keqin, Kun Zhu, Yixin Meng, Amit Yadav, and Asif Khan. "Mixed Credit Scoring Model of Logistic Regression and Evidence Weight in the Background of Big Data." In International Conference on Intelligent Systems Design and Applications, pp. 435-443. Springer, Cham, 2018.

[6] He, Yaoyao, Yang Qin, Shuo Wang, Xu Wang, and Chao Wang. "Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network." Applied energy 233 (2019): 565-575.

[7] Khedher, Laila, Javier Ramírez, Juan Manuel Górriz, Abdelbasset Brahim, Fermín Segovia, and Alzheimer′ s Disease Neuroimaging Initiative. "Early diagnosis of Alzheimer′ s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images." Neurocomputing 151 (2015): 139-150.

[8] Barbu, Adrian, Yiyuan She, Liangjing Ding, and Gary Gramajo. "Feature selection with annealing for computer vision and big data learning." IEEE transactions on pattern analysis and machine intelligence 39, no. 2 (2016):272-286.

[9] Hwang, Jing-Shiang, and Tsuey-Hwa Hu. "A stepwise regression algorithm for high-dimensional variable selection." Journal of Statistical Computation and Simulation 85, no. 9 (2015): 1793-1806.

[10] Emura, Takeshi, Shigeyuki Matsui, and Hsuan-Yu Chen. "compound. Cox: univariate feature selection and compound covariate for predicting survival." Computer methods and programs in biomedicine 168 (2019): 21-37.

[11] Aich, Satyabrata, Ahmed Abdulhakim Al-Absi, Kueh Lee Hui, John Tark Lee, and Mangal Sain. "A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques." In 2018 20th International Conference on Advanced Communication Technology (ICACT), pp. 139-143. IEEE, 2018.

[12] Jia, Feng, Yaguo Lei, Liang Guo, Jing Lin, and Saibo Xing. "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines." Neurocomputing 272 (2018): 619-628.

[13] Mandy Sidana. “Types of classification algorithms in Machine Learning”.(2017)