International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Feature Selection Techniques for Wood Density Prediction in Forest Dataset

Jeyabharathy Sadaiyandi[1], Dr.Padmapriya Arumugam[2]
[1]Research Scholar, Department of Computer Science
Alagappa University
Karaikudi,India
bharadhina@gmail.com
[2]Professor, Department of Computer Science
Alagappa University
Karaikudi, India
padmapriyaa@alagappauniversity.ac.in

Abstract— Feature selection becomes important, especially in data sets with a large number of variables and features. It will eliminate unimportant variables and increase classification accuracy and performance. Because of the increasing growth of data in numerous industries, some data is high-dimensional and contains essential and complex hidden linkages, posing new problems to feature selection: i) How to extract the underlying available relationships from the data, and ii) How to apply the learnt relations to improve feature selection? To address these issues, we use the six feature selection approach as a pre-processing step in the analysis to avoid over fitting and potential model under performance. Which can learn and apply the underlying sample relations and feature relations for feature selection. This study compared six feature selection approaches (Pearson Coefficient, Correlation matrix, Variable Importance, Forward selection, and Backward Elimination) for determining the decomposition level of forest trees. Our trials clearly provide a comparative evaluation of the Wrapper approach from several angles. Furthermore, we compare the dataset result with critical attributes to obtain the highest percentage accuracy. The experimental results show that the wrapper technique outperforms all other methods in all experiment groups.

Keywords- Feature Selection, Wrapper method, Machine Learning, Prediction

I. INTRODUCTION

Machine learning is a present-day topic in computer science that is being used to develop a wide variety of decision support systems. When decision support systems are utilized in real-world applications, high-dimensional data is a prevalent issue. This issue is frequently referred to as the curse of dimensionality [1]. Classification systems function less effectively when there are unnecessary features in the data. Any decision support system development process starts with choosing the most important features.

Feature selection methods are used to select important features. To choose relevant features, a variety of feature selection techniques are available.

Applying a feature selection approach to a set of features results in the selection of various features. The features used affect the classification process' performance [2]. In order to extract essential features from the dataset, it is crucial to employ a suitable feature selection technique.

Feature selection methods are techniques used to choose a subset of variables or properties in a dataset that are relevant [5]. The purpose of feature selection is to increase model performance, decrease over fitting, and improve interpretability. Here are some commonly used feature selection methods:

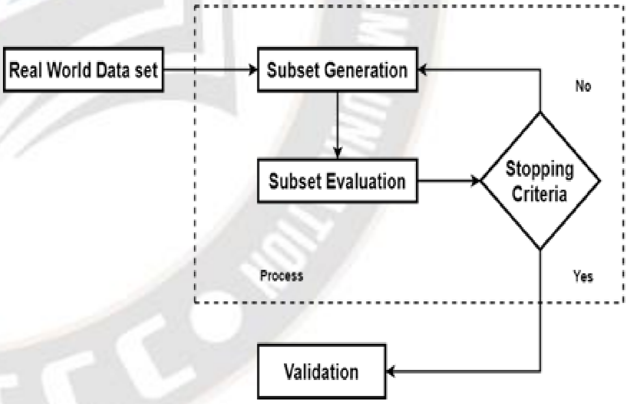

Figure 1. Feature selection Method

II. FEATURE SELECTION METHODS

The Feature selection methods help in the selection of relevant features, hence reducing the number of features. The system's accuracy improves when the number of characteristics is reduced. It also minimizes the system's complexity. Removing duplicate characteristics also aids in the reduction of computation time. [3].This section discusses the many feature selection techniques available, as well as their advantages and disadvantages. There is also a discussion of a few popular feature selection algorithms.

Feature selection methods are employed for several reasons:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

A. Improved Model Performance

Including unnecessary or redundant features in a model might introduce noise and raise the complexity of the learning task. Feature selection helps in identifying the most relevant and informative features, which can lead to improved model performance. By removing irrelevant or noisy features, the model can focus on the most discriminative aspects of the data.

B. Over fitting Prevention

Over fitting happens when a model becomes too complicated and learns to fit the training data too closely, resulting in poor generalization to unknown data. Feature selection reduces over fitting by lowering the amount of features and, hence, the model's complexity. By selecting the most relevant features, the model becomes less prone to fitting noise and specific patterns in the training data.

C. Enhanced Interpretability

In many domains, interpretability is crucial for understanding and trusting the models' predictions. Feature selection can help in creating simpler models with a reduced number of features that are easier to interpret and explain to stakeholders. By focusing on a subset of relevant features, it becomes clearer which aspects of the data are driving the model's decisions.

D. Reduced Computational Complexity

By selecting a subset of features, feature selection can significantly reduce the computational requirements of model training and inference. With fewer features, the model needs less time and resources for training, making it more efficient. This is especially crucial when dealing with high-dimensional datasets with a large number of features.

E. Data Understanding and Insights

Feature selection methods provide insights into the relationships between features and the target variable. By examining the importance or relevance of each feature, analysts can gain a deeper understanding of the data and identify key factors that influence the target variable. This knowledge can be valuable for feature engineering, domain understanding, and subsequent modelling efforts.

In general, feature selection is an important stage in the machine learning pipeline as it helps to improve model performance, prevent over fitting, enhance interpretability, reduce computational complexity, and provide insights into the data. However, it's important to note that not all problems may require feature selection, and the choice of method should be based on the specific properties of the dataset and the goals of the research.

Types of Feature Selection Methods

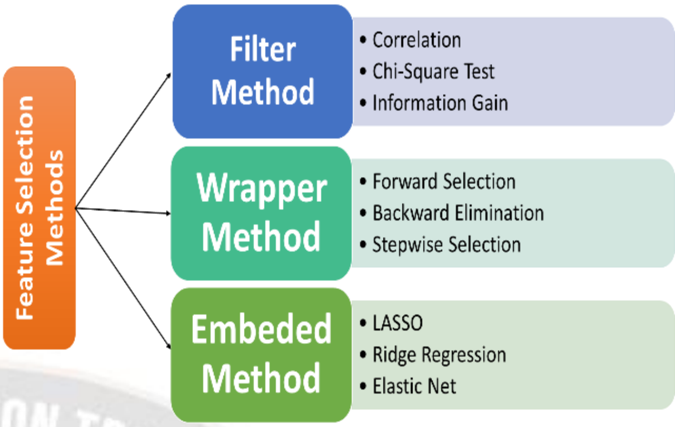

Figure 2. Types of feature selection methods

Filter Methods

The features are filtered by filter methods before being applied to the classification algorithm. Utilizing the fundamental properties of the data, they rate the features. This ranking is used to choose the features. These methods evaluate the relationship between each feature and the target variable independently of the machine learning algorithm. Figure 3 shows Common filter methods include.

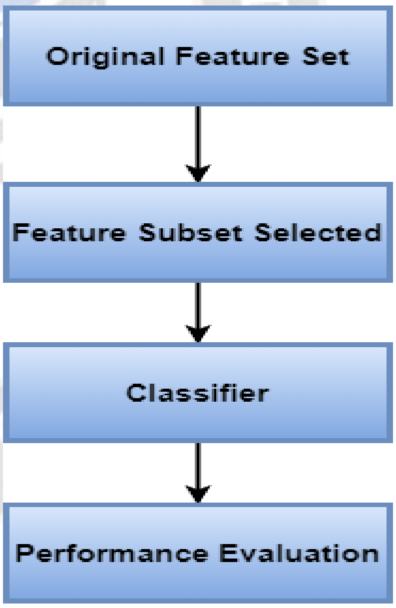

Figure 3. Filter method of feature selection

Advantages

i) The features are ranked according to criteria that are not dependent on the machine learning method.

ii) These methods are faster than Wrapper and Embedded methods. Therefore, for large datasets, these methods are the best.

iii) These methods have less risk of over fitting.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Disadvantages

i) Important features might not be chosen since the interplay between features is not taken into account while ranking them individually.

ii) The classifier used is not considered while selecting the features.

iii) There is no specific rule for selecting the threshold point to perform feature ranking.

Pearson's Correlation

Pearson's correlation statistics, often known as Pearson's correlation coefficient in statistical models, are also known as the R-value. It gives back a result that shows how closely any two variables are correlated. The Pearson correlation coefficient is determined by dividing the covariance of two variables by the sum of their standard deviations. Scale shifts in the two variables have no impact on the coefficient. Simply it Measures the linear correlation between features and the target variable [6].

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^{2} \sum(y-\bar{y})^{2}}} \qquad (1)$$

- Σ indicates the total of the subsequent computations.
- x and y are the individual data points for the two variables.
- x̄ (x-bar) and ȳ (y-bar) are the means (average) of the x and y values, respectively.
- sqrt denotes the square root function

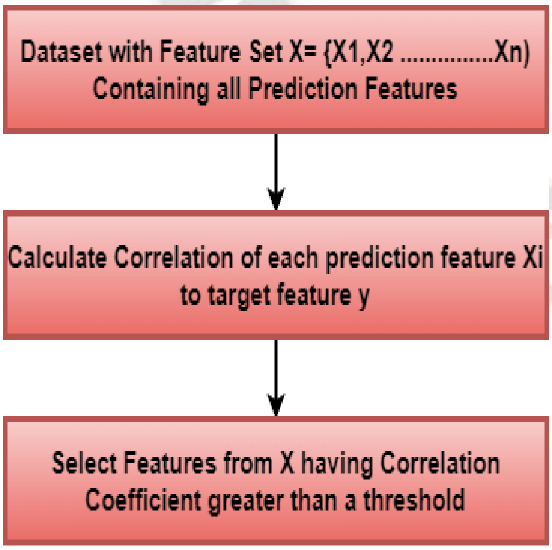

Figure 4. Pearson Correlation Algorithm

The linear correlation between the variables' strength and direction is quantified. The correlation coefficient, indicated by the symbol "r," is between -1 and +1, where:

- When r = +1, a perfect positive linear relationship is present.
- A perfect negative (inverse) linear connection is indicated by r = -1.
- r = 0 denotes the absence of a linear relationship (independence of the variables).

It helps identify features that are highly correlated (positively or negatively) with the target, suggesting their potential predictive power.

Correlation Matrix

In a table known as a correlation matrix, many variables' pairwise correlations are displayed. It is a handy tool for examining the connections between variables and comprehending the associations between them.

The diagonal members of a correlation matrix represent the correlation between each variable and itself, which is always 1. The matrix is symmetrical across the diagonal because no matter what order the variables are in, there is always a correlation between them.

Correlation coefficients can take on three possible values:

- Positive correlation: A number that approaches 1 denotes a strong positive association, which means that when one variable rises, the other rises generally as well.

- Negative correlation: If the value is near to -1, there is a strong negative association between the two variables, meaning that as one variable rises, the other tends to fall.

- No correlation: A value that approaches 0 indicates that there is little to no linear relationship between the variables. It's crucial to remember that two variables might still have a nonlinear relationship even when their correlation coefficient is very close to 0.

Here is the R Code to find Correlation Matrix

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

```
//Load the dataset
mydata<-read.csv("c:/dataset.csv")
//find the correlation for whole dataset
res<-cor(mydata)
//round the result to 2 decimal places
round(res,2)
```

Variable Importance

This technique detects the independent variable from the dataset which influence more on target variable. It returns the score for individual features in the dataset. The variables having higher score are included to train the model. It is simplest method to calculate the coefficient statistics between the independent and target variable

Wrapper Methods

Wrapping methods By iteratively training the model on several subsets of features, then selecting the best subset, one can choose features. The model inferences are used to select which feature subset. [4]. These techniques train and test the model with various feature combinations before evaluating the feature subsets[10][11]. Figure 5 displays the method used to choose features in the wrapper methods

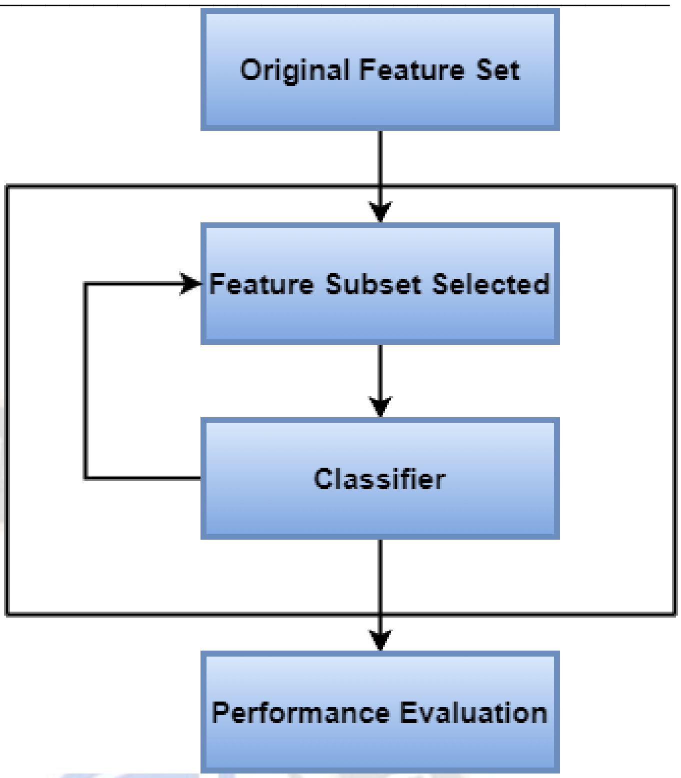

Figure 5. Wrapper method of feature Selection

Here's a general overview of how wrapper methods work:

a) Define a subset search space: List the potential feature subsets that will be taken into consideration throughout the selection process. This could be the power set of all features, including subsets of all sizes, or a search space that is constrained due to domain expertise or computing constraints.

b) Select an evaluation metric: Select a statistic that captures the efficacy or standard of the predictive model. Common evaluation criteria include recall, accuracy, precision, F1 score, and any other pertinent metrics dependent on the particular situation.

c) Train a predictive model: Choose a machine learning model to serve as the process' evaluation tool for choosing features. Decision trees, logistic regression, support vector machines, and any other model appropriate for the current situation can be used, depending on the particular challenge at hand.

d) Iterate through feature subsets: Progressively add or delete features from the search space after starting with an empty set of features. Only the chosen features should be used to train the prediction model for each subset, and the performance should be assessed using the chosen evaluation metric.

e) Select the best feature subset: Compare the model's performance for every subset of features. Pick the subset that, based on the evaluation metric, produces the best performance. The chosen feature set is this subsection.

f) Assess the selected feature set: As the wrapper method has determined the optimum feature subset, it is crucial to assess

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *05 July 2023* ***Revised:*** *25 August 2023* ***Accepted:*** *15 September 2023*

its effectiveness using the right methods, such as cross-validation or holdout validation. This ensures that the features chosen have good generalizability to new data.

They rely on a specific machine learning algorithm to provide an evaluation metric for feature selection.

Common wrapper methods include:

- Recursive Feature Elimination (RFE): Recursively removes features and assesses their impact on the model's performance.
- Forward Selection: Begins with a blank feature set and gradually adds features until the model is performing at its peak.
- Backward Elimination: Starts off with all features and iteratively removes the one that has the least impact on the model's performance.
- Embedded Methods: These techniques include feature selection in the model training phase. They use algorithms that inherently incorporate feature selection techniques.

FORWARD SELECTION METHOD

Forward selection is a feature selection method that incrementally builds a predictive model by iteratively adding features that contribute the most to the model's performance. The procedure begins with a blank feature set and gradually adds features one at a time in accordance with a predetermined criterion.

Here's a step-by-step explanation of the forward selection method:

```
//Forward Selection Method
    i)      Initialize an empty feature set
    ii)     Evaluate the Candidate features
    iii)    Select the best feature
    iv)     Iterate and add features
    v)      Termination condition
    vi)     Finalize the Feature set
    vii)    Build the Predictive Model using
            Machine Learning algorithms
    viii)   Validate and Evaluate the Model
```

Forward selection has the advantage of being a simple and intuitive method for feature selection. It progressively adds features based on their individual performance, which can lead to a subset of features that are highly relevant to the prediction task. However, it's important to note that forward selection does not consider interactions between features, and removing a previously added feature is not possible in this method. Additionally, it may be computationally expensive for datasets with a large number of features.

BACKWARD ELIMINATION METHOD

Backward elimination is a feature selection technique that begins with a model that contains all of the available characteristics and systematically eliminates the least important ones one at a time until a stopping criterion is satisfied. It's a wrapper technique frequently used in conjunction with statistical models to pick a subset of pertinent features. Figure.6 shows the Backward Elimination Process.

Backward elimination helps identify the most relevant features by iteratively eliminating those that are least significant according to the chosen significance level. By progressively removing features, the method aims to improve the model's performance and interpretability by reducing over fitting and removing irrelevant or redundant features.

Here's a step-by-step explanation of how backward elimination works:

```
// Backward Elimination Method
    i)      Select a Significance Level(e.g., p-value
            threshold) , this is often set to a fixed
            value, such 0.05 or 0.01
    ii)     Train the Initial Model using Regression
            or Statistical Model
    iii)    Perform the Feature Elimination Iteration
            a.  Fit the Model
            b.  Calculate the p-values
            c.  Identify the least significant feature
            d.  Remove the least significant feature
    iv)     Repeat the Iteration
    v)      Finalize the Model
    vi)     Validate the selected feature subset.
```

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

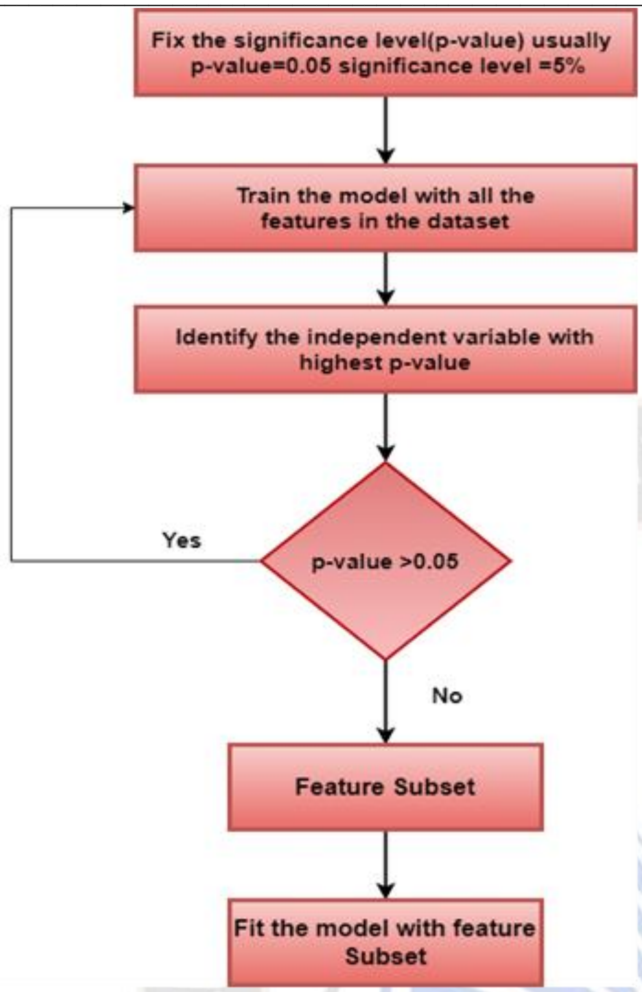

Figure 6.Backward feature elimination algorithm

III. METHODS AND MATERIALS

In general, Wrapper approaches require training and assessing the predictive model numerous times for various feature subsets, which can be computationally expensive. However, compared to filter approaches that only use statistical metrics, they can offer more precise feature selection. Wrapper methods are helpful when trying to maximize the performance of a particular model since they consider how the model behaves with various feature subsets [11]. Here, we've put the Wrapper method for feature selection for predicting a tree's wood density in a forest dataset. Figure.7 shows the work flow diagram.

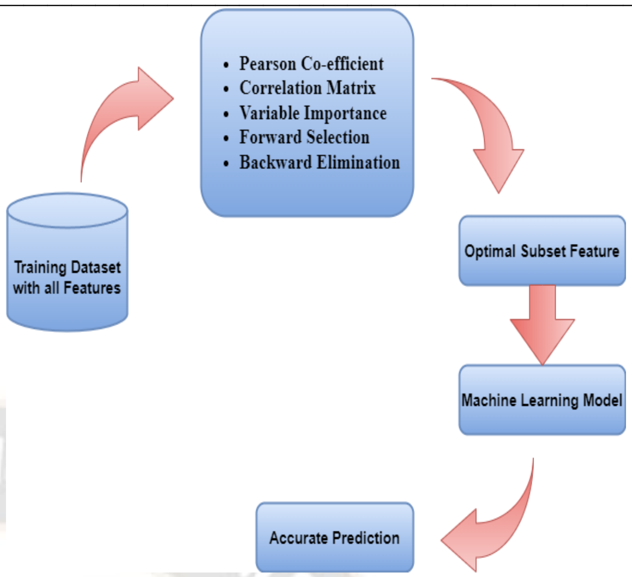

Figure 7. Work Flow Diagram

A. Dataset

The dataset taken from USDA repository. It has 21 attributes consisting of both relevant and irrelevant attributes. The irrelevant attributes are nothing but which does not support for our output. We obtain a subset features from raw dataset which contribute maximum to predict the wood density and with that results we predict its decay class of the tree.

Table 1. Forest Data set

Attributes	**Description**
Log num	It is a number given to every log
Species	Kind of the tree, Here 4 species
Time	Number of years for the tree
Year	Year Label
Subtype	Type of the trees such as Hard,Soft
Radpos	Position where measurement taken
D1	Diameter of the tree
D2	Diameter of the tree at various pos
D3	Diameter of the tree at various pos
D4	Diameter of the tree at various pos
VOL1	Volume of the Tree
VOL2	Volume of the Tree
WetWt	Weight of the wetness in the tree
DRYWT	Dry weight of the tree
MOIST	Moisture content in the wood
Decay	Decompositon level of the tree
WDENSITY	Wood density of the tree with respect to vol1
Den2	Wood density of the tree with respect to Vol2

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 05 July 2023 **Revised:** 25 August 2023 **Accepted:** 15 September 2023*

Knot Vol	Volume of the wood at knot
Sample Date	Date on which the data is collected
Comments	About any special features of the tree

IV. RESULTS AND DISCUSSION

1) Pearson Coefficient

Pearson coefficient technique applied on the dataset and found the optimal feature subset based on coefficient values. The attributes are selected based on the criteria whose coefficient values are greater than >50. Under this criteria 6 attributes were chosen from the total 21 attributes. The machine learning algorithm is trained with these 6 attributes and measured the performance of the model. As a result the model gives the accuracy of 87% in predicting the wood density.

Table 2. Attribute Selection using R2 Value

Attributes	Pearson value Co-Efficient	Selected Attributes (>0.5)	R^2 value
Lognum	0.199	Species	0.869
Species	-0.527	DryWT	
Time	0.120	DECAY	
Year	-0.208		
Subtype	-0.384		
Radpos	0.165		
D1	0.367		
D2	-0.408		
D3	-0.393		
D4	-0.002		
VOL1	0.001		
VOL2	0.002		
WetWt	0.425		
DRYWT	0.76		
MOIST	-0.389		
Decay	-0.817		
Den2	0.340		
Knot Vol	0.223		
Sample Date	0.208		
Comments	0.001		

2) Correlation Matrix

This technique is applied on the dataset and picked up the feature subset and trained the machine learning model. Using this technique 6 attributes were taken and the performance of the model is measured. The machine learning model trained with these 6 attributes produces 90% in prediction accuracy.

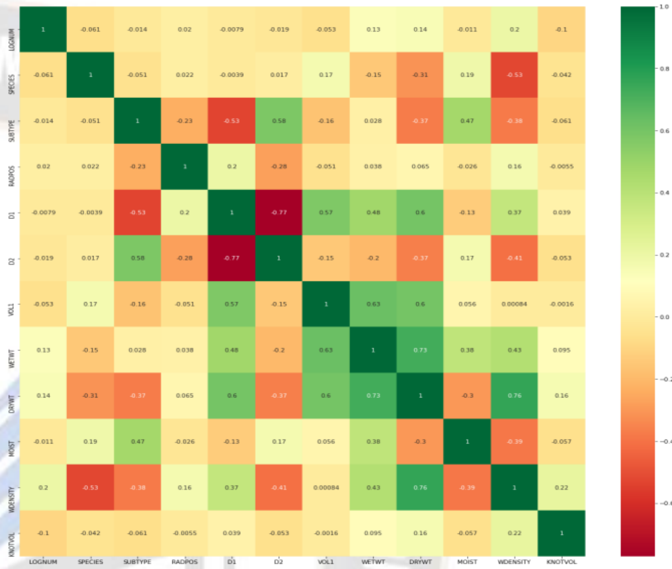

Figure 8 Feature Selection using Correlation Matrix

3) Variable Importance

This technique finds the importance variables from the full set of features available in the dataset. Here we chosen the attributes having the value >5.Under this criteria 5 attributes are chosen and trained the model which gives 86% in prediction accuracy.

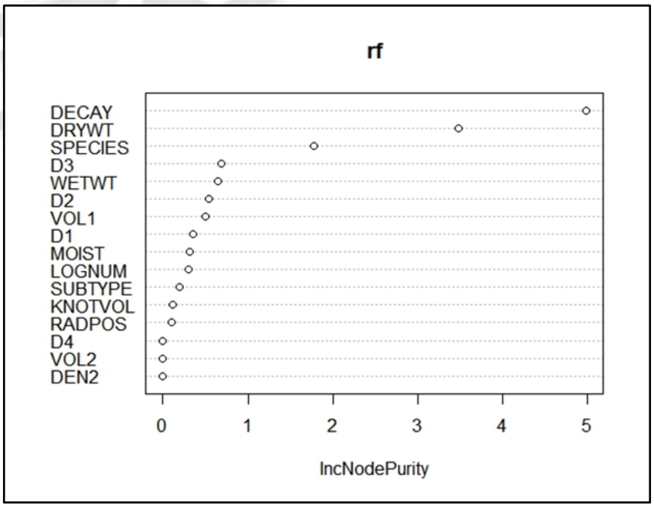

Figure 9. Feature Selection using Variable Importance

4) Wrapper Method

In this technique we implemented forward selection method and detected the subset features which contribute maximum to the output prediction accuracy R^2.

Table 3. Attribute Selection using Wrapper Method

Wrapper Method	R^2 value
dat1[,c(2,11,14,16)]	0.908
dat1[,c(7,11,14,16)]	0.9135
dat1[,c(11,14,16)]	0.90
dat1[,c(2,7,11,14,16)]	0.916
dat1[,c(1,2,5:7,11,14,16)]	0.9185
dat1[,c(1,2,5:9,10,11,14,16)]	0.9226
dat1[,c(1,2,5:9,10,11:12,14,16)]	0.9226
dat1[,c(1,2,5:9,10,11:14,16)]	0.9229
dat1[,c(1,2,5:9,10,11:16)]	0.9239
dat1[,c(2,7,11,13,14,16)]	0.937

From the table we can select the optimal subset feature which gives maximum result as 94% accuracy in predicting the output.

V. CLASSIFICATION ACCURACY AND COMPARISON STUDY

After obtaining the output from all the feature selection techniques the results are compared. As a result, Wrapper method provides optimal subset feature with 6 attributes and the accuracy 94% highest compared to other techniques. Next to wrapper method, PCA provides highest accuracy of 93%. So we finalize the optimal feature subset from Wrapper technique and fit the machine learning model[12].

Table 4. Classification Accuracy – Comparative table

Techniques	**No of Attributes**	**Prediction Accuracy**
Pearson	6	87%
Correlation	6	90%
Variable Importance	5	86%
PCA	7	93%
Wrapper method	**6**	**94%**

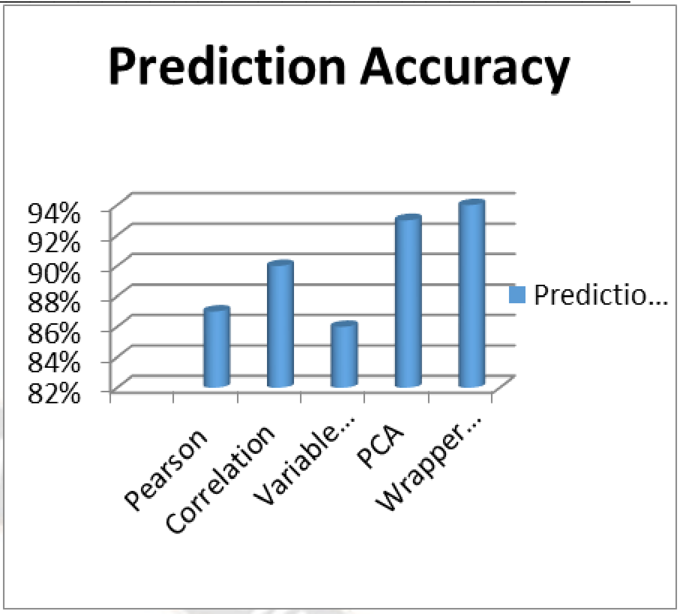

Figure 10. Chart Shows Classification Accuracy using the Wrapper Feature Selection Method

VI. CONCLUSION

The strategy for selecting the best subset of features from a large feature space is presented in this work. The process entails feeding the entire feature set into a feature selection algorithm, pulling features from the feature space in a specific order, comparing the features in various spaces using a feature evaluation criterion, and then repeating the process until the new space has a default number of features. The proposed method combines the advantages of the Filter selection algorithm and the Wrapper selection algorithm, and by using the complementary characteristics of the two algorithms, calculation cost is reduced while algorithm efficiency is improved.. The chosen ideal feature subset is thought to consist of the features in the new feature space. The method conveniently controls the duration of the entire selection process and provides for the choosing of the number of characteristics to be included in the result. It also allows for the saving of the threshold value setting. The paper concludes that the wrapper method can select a better feature subset compared to other methods and improve the classification accuracy of the model.

REFERENCES

[1]. Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M.,Yuan, X., & Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. Applied Soft Computing,74,634-642. https://doi.org/10.1016/j.asoc.2018. 10.036.

[2]. Dulhare, U. N. (2018). Prediction system for heart disease using naive bayes and particle swarm optimization. Biomedical Research, 29(12), 2646-2649. http://dx.doi.org/10.4066/biomedicalresearch. 29-18-620.

[3]. Bhattacharya, A., Goswami, R. T., Mukherjee, K., & Nguyen, N. G. (2019). An ensemble voted feature selection technique for predictive modeling of malwares of android. International Journal of Information System Modeling and Design, 10(2), 46-69. https://doi.org/10.4018/IJISMD.2019040103.

[4]. Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. Egyptian Informatics Journal, 19(3), 179-189. https://doi.org/10.1016/j.eij.2018.03.002.

[5]. L. Ladla and T. Deepa, " Feature Selection Methods And Algorithms", International Journal on Computer Science and Engineering (IJCSE),vol.3(5), pp. 1787-1797, 2011.

[6]. A. G. K. Janecek and G. F. Gansterer et al, "On the Relationship between Feature Selection and Classification Accuracy", In: Proceeding of New Challenges for Feature Selection, pp. 40-105, 2008.

[7]. Z. Zhao, H. Liu, "Searching for Interacting Features", Proceedings of International Joint Conference on Artificial Intelligence, pp. 1156-1161, 2007.

[8]. Weisberg, S. (2005). Applied linear regression. Hoboken: John Wiley & Sons,Incorporated.

[9]. Stańczyk, U. (2015). Feature evaluation by filter, wrapper, and embedded approaches.In Feature selection for data and pattern recognition (pp. 29–44). Berlin, Heidelberg:Springer.

[10]. Jantawan B, Tsai CF.A Comparison of Filter and Wrapper Approaches with Data Mining Techniques for Categorical Variables Selection. International Journal of Innovative Research in Computer and Communication Engineering; 2014.

[11]. Kumari, B., & Swarnkar, T. (2011). Filter versus wrapper feature subset selection in large dimensionality micro array: A review. International Journal of Computer Science and Information Technologies, 2(3), 1048-1053.

[12]. Matharaarachchi, S., Domaratzki, M., & Muthukumarana, S. (2021). Assessing feature selection method performance with class imbalance data. Machine Learning with Applications, 6, 100170.

[13]. Solorio-Fernandez, S., Martínez-Trinidad, J. F., & Carrasco-Ochoa, J. A. (2020). A supervised filter feature selection method for mixed data based on spectral feature selection and information-theory redundancy analysis. Pattern Recognition Letters, 138, 321-328.