IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Performance Evaluation Using Classifier Algorithm On Endometrial Cancer Data

A. Hency Juliet[1*], R. Padmajavalli[2]

[1]Research & Development Centre, Bharathiar University, Coimbatore and Assistant Professor, Department of Computer Application, Mar Gregorios College, Chennai, India
[2]Research & Development Centre, Bharathiar University, Coimbatore and Associate Professor, Department of Computer Application, Bhaktavatsalam Memorial College for Women, Chennai, India

[*]*Corresponding Author:* hencyjuliet@gmail.com

Available online at: www.ijcseonline.org

Accepted: 20/Sept/2018, Published: 31/Oct/2018

Abstract— Many factors affecting the success of data mining techniques, the pureness of data are one of the factors. The inclusion of irrelevant and noisy data in the pattern analyzing phase, can results poor predicting performance. To discover information from the endometrial carcinoma data set, the pre-processing technique such as cleaning, transforming and modelling are applied. Diverse kinds of pre-processing techniques were functions in the data set in order to work with the full pledged data set. ***Methodology Used:*** The data mining tool WEKA is used for feature selection. Using various classifiers, evaluators and search methods six features are selected out of eighteen, attributes. The performance evaluation was done using RStudio. The accuracy of the classifiers model Random Forest and Naïve Bayes are checked for the minimized and full data sets. The hybrid model was formed by combing both the models to improve the performance of the classifier model. ***Findings:*** The Hybrid model was adopted for the performance evaluation by combining naïve Bayes and random forest classifier and the accuracy of the new model is 93.55%.

Keywords — WEKA; Endometrial; R; carcinoma; classifiers; Naïve Bayes; Random Forest.

I. INTRODUCTION

Data mining is the modus operandi of pulling out the data from the enormous dataset [10]. Data preprocessing is the evolution of cleaning, altering and modeling the data with the target of discovering new patterns [2]. The preprocessing techniques include Data cleaning, integration, reduction, and transformation. Attribute selection is the process of identifying and selecting the relevant and unique information as possible.

A. Data cleaning:

It is referred to as removing noisy and correcting inconsistencies in the data set. The scheme in data cleaning is filling the missing-data, smoothing the noisy-value, identifying and removing-outliers and resolving consistencies [18]. The data, available in the real world, may be incomplete, noisy and inconsistent and having incorrect attributes values. The data collection mechanisms used may be defective. While entering the data human or computer error may happen. Erroneous data may result from discrepancies in naming regulations. Dirty data can cause confusion for the mining procedure [18].

B. Data Integration

Merging data from many sources into one source is referred to as data integration. It may comprise several databases, data cubes, or flat files. Careful integration can decline and skip redundancies and inconsistencies in the final dataset [4]. Entity identification, redundancy, correlation analysis, and tuple duplication are some issues in data integration.

C. Data Transformation

Transforming data into an appropriate form for mining is referred to as data transformation. So that mining process may be more efficient and the patterns found may be easier to understand [5].

D. Feature Selection and Extraction:

Finding the Best Attributes:

The effectiveness of data mining can reduce due to too much information. Some of the columns of data attributes may not contribute meaningful information to the model [13]. Inappropriate features may include noise to the data so that the model precision may change. Wide data increases processing faces for data mining algorithms. The calculation rate of algorithmic processing will be higher if the dimensionality is higher [18]. Reducing in dimension is a popular preprocessing step in data mining to minimize noise [14]. Two approaches in dimension reduction

- Feature selection — choosing the most appropriate features for processing.
- Feature extraction — joining attributes into a new reduced set of an attribute.

Feature Selection:

Attribute importance function is used for feature selection in data mining. It is an organized function that rank features based on their implication in predicting a target [6]. The features of build data ranked based on the predictive implications [9]. Features selection uses ranking and measure of importance, to select the most appropriate features[21].

Feature Extraction

Feature extraction is a feature reduction method. Feature selection is nothing like, ranks the present attributes according to their prognostic implication, whereas feature extraction actually converts the attributes [11]. Much lesser and more affluent set of features will be the outcome of the feature extraction process. Data compression, data decomposition, and projection, and pattern recognition are some purpose of feature extraction. It also helps to enhance the pace and helpfulness of supervised learning [12].

The endometrial cancer patient' datasheet consists of eighteen features, in order to select the attributes relevant to the prediction and to produce the better result the feature minimization techniques were applied to the dataset. The performance of the classifier was measured using the R language. The classifier Random forest and Naïve Bayes are used in this study. In order to improve the performance of the classifier, the hybrid model was formed. The variable postmenopausal status was taken as the target variable. Totally 1040 datasets are used in this study.

II. RELATED WORK

This segment examines the allied work on datasets using data mining techniques. P.Ravisankar et.al [15] uses data mining techniques such as Multilayer Feed Forward Neural Network, Support Vector Machines, Genetic Programming, Group Method of Data Handling, Logistic Regression, and Probabilistic Neural Network to fraud. Each of these techniques is tested on a dataset and compared with and without feature selection. GP and PNN outperformed with feature selection and with marginally equal accuracies. Chia-Ming Wang [3] et al, formulated, the feature selection problem as a multi-objective optimization problem, and new criteria were proposed to fulfill the goal. Foremost, data were pre-processed with missing value replacement scheme, re-sampling procedure, data type transformation procedure, and min-max normalization procedure[20]. After that, a wide variety of classifiers and feature selection methods were conducted and evaluated. Huan Liu et al [9] introduced the concepts and algorithms of feature selection. Illustrative examples to show how existing feature selection algorithms can be integrated into a Meta algorithm that can take advantage of individual algorithms and provided guidelines in the selection of feature selection algorithms are presented. Mark A. Hall et al [12] presents a benchmark comparison of several attribute selection methods for supervised classification. All the methods produced an attribute ranking, a useful device for isolating the individual merit of an attribute. Attribute selection is achieved by cross-validating the attribute ranking with respect to a classification learner to find the best attribute [13].

III. METHODOLOGY

In this study, the endometrial cancer data was taken from the InterNational cancer Institute, Neyyoor, Tamil Nadu. The data was retrieved from the patients' data sheet, attributes such as patient's age at diagnosis, symptoms, menopause age, diet, area, family history, occupational factor, lesion, history of trauma, spouse relationship, parous, religion, menopausal status, etc... are taken. The attributes and its range codes are described in table 1. The raw data had missing values, incomplete and inconsistency. Then pre-processing techniques such as data cleaning, transformation, reduction and feature selection applied to the dataset. The process flow of pre-processing technique is illustrated in the figure 3. Using the pre-processing technique the data was cleaned and transformed to the format necessary for the data mining process. Initially, the data is in the format of text category, and then it was converted into a quantitative attribute. For example, the field menopause status takes the value yes (value - 1) if the patient has post menopause, otherwise no. but it was replaced by 1 and 0.

The presence of the missing value, noise, outlier or duplication of data in the database referred to as incomplete or inconsistency, these can be detected and removed. Removal of the attribute was done using subset function. Most of the attributes are the categorical format, using factor the categorical format are converted into quantitative data. It is the most critical step in the data mining process which deals with the preparation and transformation of the initial data set [19]. Attribute selection techniques are coined by the term Filter and Wrapper, to depict the character of the metric used to assess the meaning of attribute [16]. Wrappers assess attribute using accuracy estimates provided by the real target learning algorithm. Filters, on the other hand, use the common quality of the data to assess the attributes and drive autonomously on any machine learning algorithm [12]. So that selected attributes alone taken for the data analysis purpose out of eighteen attributes, six attributes are selected for the process.

Body mass index (BMI) is a measure of body fat based on height and weight that applies to adult men and women.

BMI = (Weight in Kilograms / (Height in

Centimeters) 2) X10000 for example, a person who weighs 99.79 Kilograms and is 190.50 centimeters tall has a BMI of 27.5. From the BMI Obesity was calculated, obesity = BMI of 30 or greater. So the attribute height and weight were replaced by BMI first then it was replaced by Obesity. Table-1 illustrates the BMI calculation. Table 2 specifies the data description of features.

Table 1 ICIEC – Patient Datasheet - Feature Description

Attribute Name	Range
AgeWhenCancerDiagnosed	Age Below 50 - 1 Age 51 - 60 - 2 Age Above 60 - 3
GeneralCondition	Good - 1 Fair - 2 Bad - 3
FirstSignofSymptom	Date of Onset
TotalDurationOfSymptom	Year, Months, Weeks
DelayBeforeConsulting _Months	Year , Months Weeks
HistoryOfTrauma	Yes -1, No - 0
OccupationalFactor	Yes -1, No - 0
MenopausalAge	Below 50 - 1, Above 50 - 2
MenopauseStatus	PreMeno - 0 PostMeno - 1
PreviousTreatment	None - 0 Surgery - 1 Radiotheraphy - 2 OtherTreatment – 3
Lesion	New - 0 Healed - 1 Residual- 2 Recurrent - 3 Massive - 4
FamilyHistory	Positive - 1 Negative – 0
SpouseRelationship	Patternal side -1 Maternal side - 2 Both - 3
Parous	Nulliparous - 0 Otherwise – 1
Diet	Veg - 0 NonVeg - 1
Religion	Hindu - 1 Christian - 2 Muslim - 3
Area	Urban - 1 Rural - 2
Obesity	Yes – 1, No – 0

Table-2 Sample BMI Calculation

Patient Height	Patient Weight	BMI	Obesity
162	76	28.959	No
158	98	39.25653	Yes
162	86	32.76939	Yes
168	59	20.9042	No
152	60	25.96953	No
165	64	23.50781	No
160	48	18.75	No
172	82	27.71769	No
170	90	31.14187	Yes

Figure 1 represents the distribution of Age at the time of diagnosis. The patients whose age are greater than or equal to 50 are affected a lot.

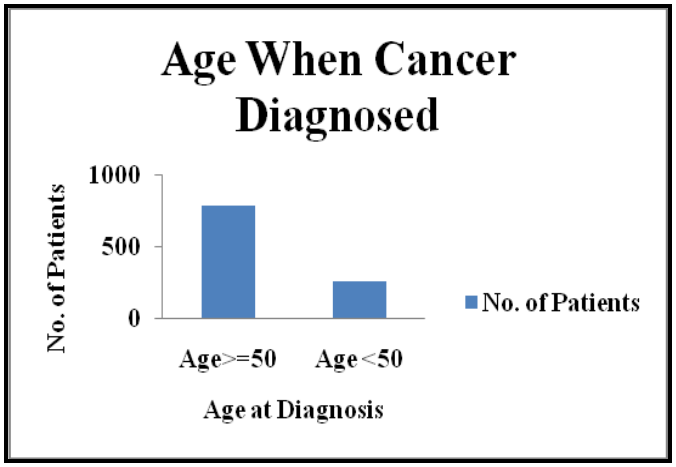


Figure 1 Distribution of Age at diagnosis

Figure 2 represents the distribution of Menopause Status. The patients with post menopausal are more in number than the patients with pre menopausal

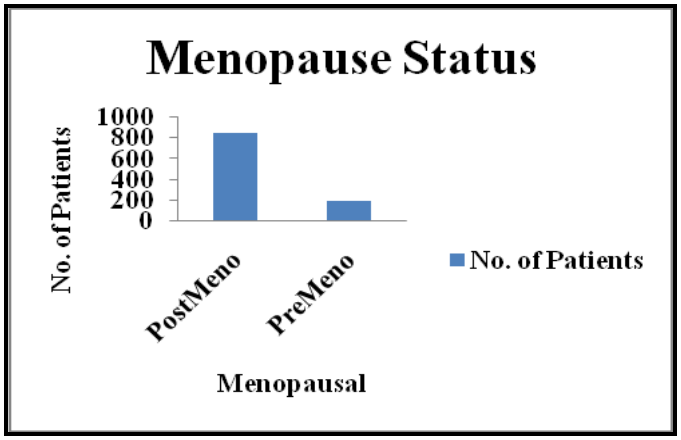


Figure 2 Distribution of Menopausal Status

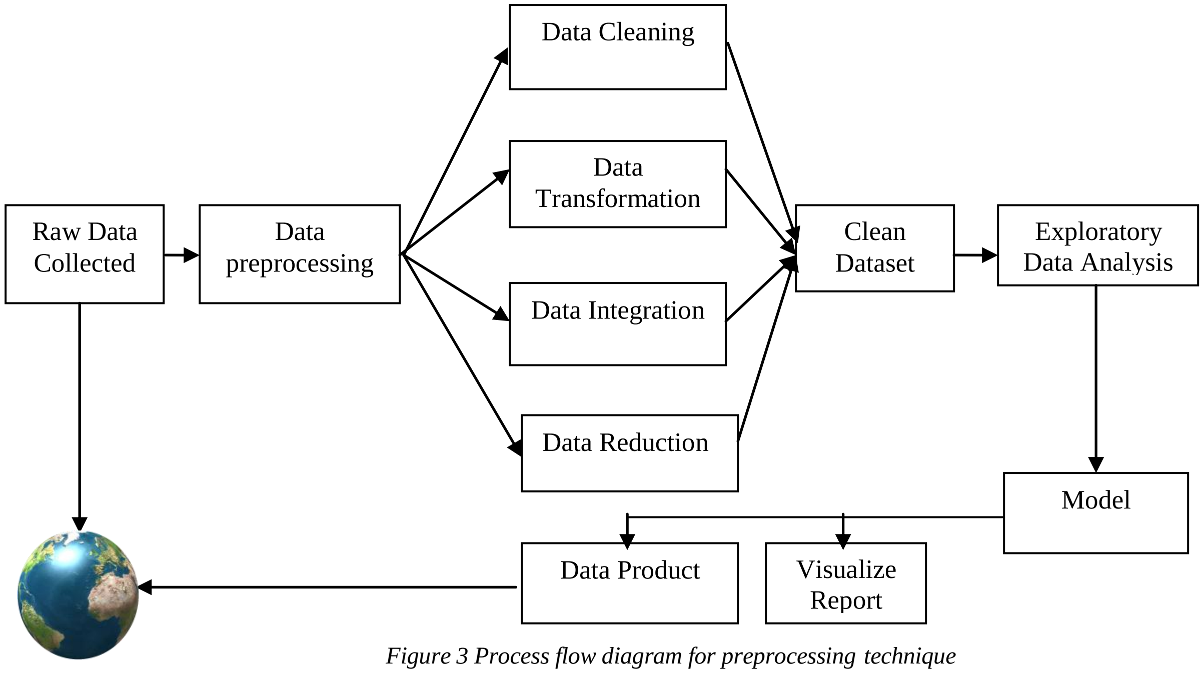


Figure 3 Process flow diagram for preprocessing technique

A. Overview of Endometrial Cancer

The well-known type of adenocarcinoma is referred to as endometriosis cancer. Endometriosis cancers are created by cells in glands that appear a lot like the usual uterine inside layer (endometrial) [1]. The two kinds of endometrial cancer are type1 and type2. Type 1 endometrial cancers are deliberation to be origin by surfeit estrogen. They built up from a distinctive hyperplasia sometimes, an abnormal overgrowth of cells in the endometrium [5]. Type 1 cancers are typically not very destructive and are slow to widen to new tissues. Type 2 endometrial cancers create a small number of endometrial cancers. It doesn't give the impression to be caused by a huge quantity of estrogen [17].

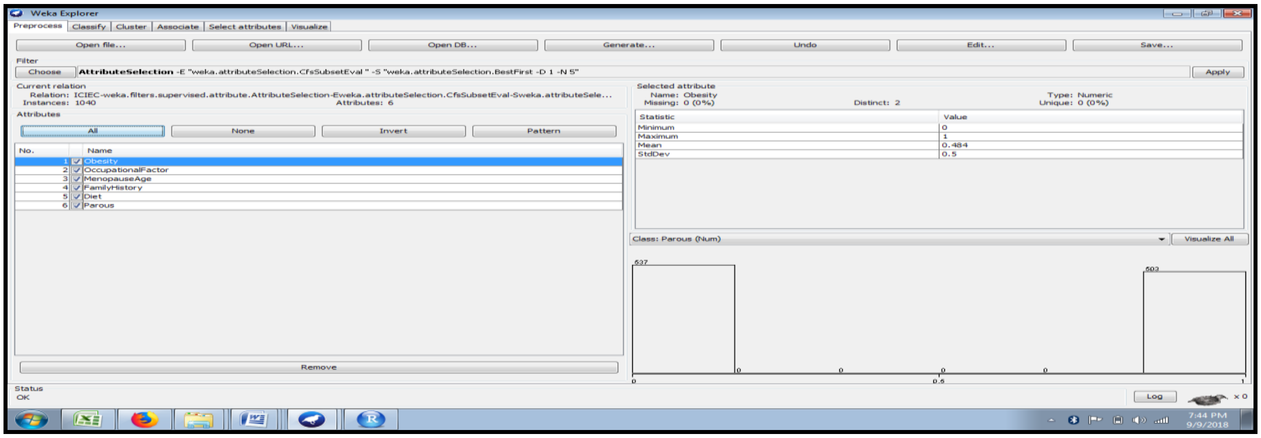
B. Overview of R

R is a programming language and free software environment for statistical computing and graphics. The R language is extensively used amid statisticians and data miners for developing statistical software and data analysis. R is a GNU package. The source code for the R software environment is written primarily in C, FORTRAN, and R [7]. R has a command line interface, there are several graphical front-ends, most notably RStudio [8]. Integrated development environments are available.

IV. Results and Discussion

Using WEKA the feature selection process was carried out. The attribute selection process of the supervised filter of weka was used for this purpose. It extracted six features out of eighteen features. The feature extraction was represented in figure 4. Various evaluators are applied on the endometrial cancer data. The variable postmenopausal status was taken as the target variable. Totally 1040 datasets are used in this study. Then 70% of the data set is taken for train data and 30% of the data are taken as test data. Initially, the model was built using the train data. Then the prediction was conducted on the test data. Then the performance of naïve Bayes and random forest algorithm was evaluated using evaluation parameters, for the minimized features dataset and for a dataset with all the features. The screenshot of Rstudio with accuracy details are represented in figures 5 to 7. The classifier Random forest, Naïve Bayes are performed well on the ICIEC data. The accuracy of the model was evaluated on minimized feature data as well as on the full data. The parameters such as Gini Coefficient, accuracy, and Area under the curve, specificity and sensitivity are calculated to evaluate the performance of the classifier model.

The Gini coefficient is used to measure the inequality in the distribution. The area under the curve (AUC) is calculated to measure the quality of a classifier. Accuracy is calculated to measure the correctness of the classifier. Specificity relates the classifiers ability to identify negative results and Sensitivity is the proportion of actual positives which are correctly identified as positives. The parameters are calculated for the classifiers model and described in table 3. Obviously, the accuracy with feature selection is better than the accuracy without feature selection. Naïve Bayes and Random Forest are performed well on the minimized dataset.



Filter
AttributeSelection -E "weka.attributeSelection.CfsSubsetEval " -S "weka.attributeSelection.BestFirst -D 1 -N 5"

Current relation
Relation: ICIEC-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelection.CfsSubsetEval-Sweka.attributeSele...
Instances: 1040 Attributes: 6

Attributes

No.		Name
1	☑	Obesity
2	☑	OccupationalFactor
3	☑	MenopauseAge
4	☑	FamilyHistory
5	☑	Diet
6	☑	Parous

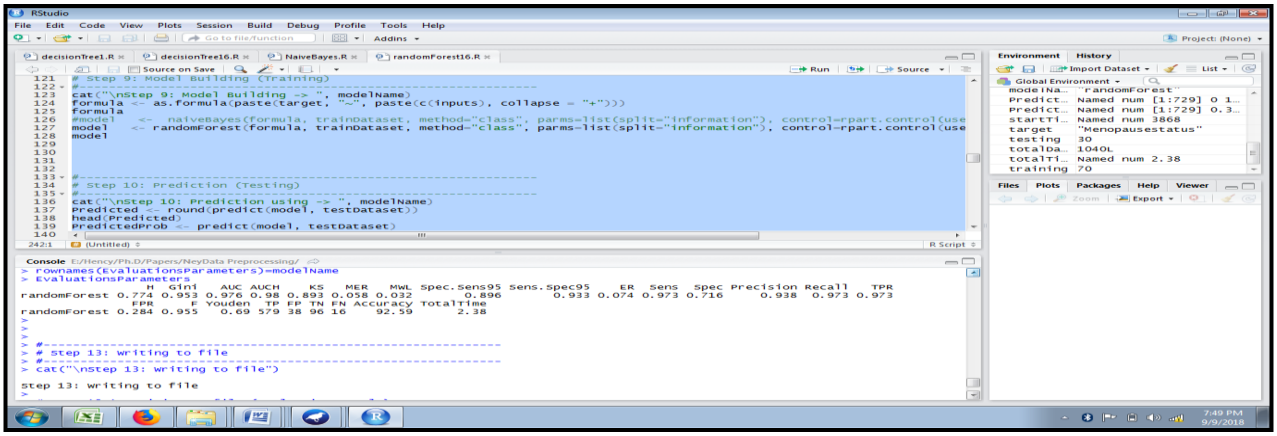
Selected attribute
Name: Obesity Type: Numeric
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

Statistic	Value
Minimum	0
Maximum	1
Mean	0.484
StdDev	0.5

Class: Parous (Num)

Status
OK

Figure 4 Attribute Selection Using WEKA



```
# Step 9: Model Building (Training)
#-----------------------------------------------------------------
cat("\nStep 9: Model Building -> ", modelName)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula
#model    <-  naiveBayes(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.control(use
model    <- randomForest(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.control(use
model




#-----------------------------------------------------------------
# Step 10: Prediction (Testing)
#-----------------------------------------------------------------
cat("\nStep 10: Prediction using -> ", modelName)
Predicted <- round(predict(model, testDataset))
head(Predicted)
PredictedProb <- predict(model, testDataset)
```

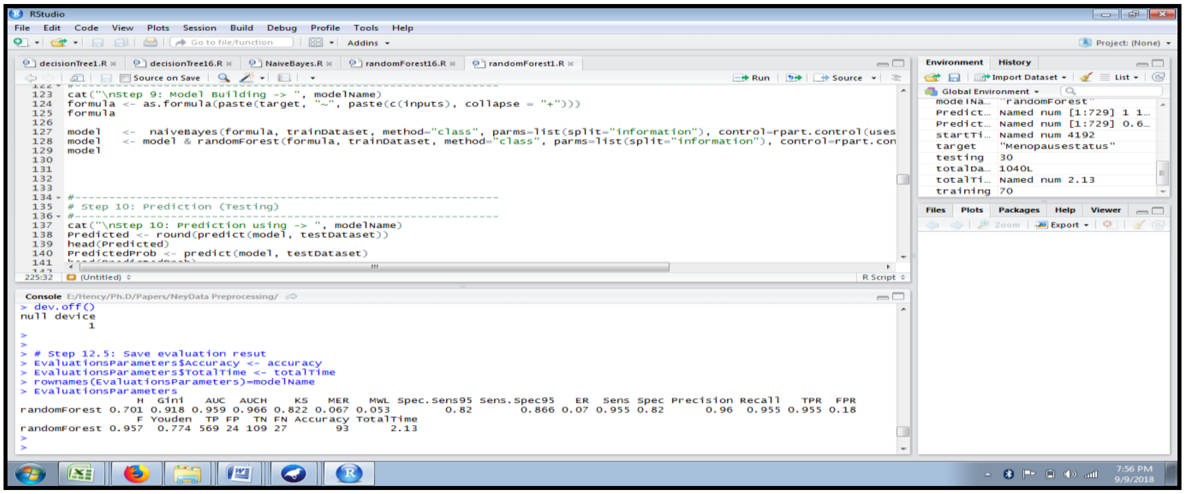
Console E:/Hency/Ph.D/Papers/NeyData Preprocessing/

```
> rownames(EvaluationsParameters)=modelName
> EvaluationsParameters
                 H  Gini   AUC  AUCH    KS   MER   MWL Spec.Sens95 Sens.Spec95    ER  Sens  Spec Precision Recall   TPR
randomForest 0.774 0.953 0.976 0.98 0.893 0.058 0.032       0.896       0.933 0.074 0.973 0.716     0.938  0.973 0.973
               FPR     F Youden  TP FP TN FN Accuracy TotalTime
randomForest 0.284 0.955   0.69 579 38 96 16    92.59      2.38
>
>
>
> #-----------------------------------------------------------------
> # Step 13: Writing to file
> #-----------------------------------------------------------------
> cat("\nStep 13: Writing to file")

Step 13: Writing to file
>
```

```
modelNa…  "randomForest"
Predict…  Named num [1:729] 0 1…
Predict…  Named num [1:729] 0.3…
startTi…  Named num 3868
target    "Menopausestatus"
testing   30
totalDa…  1040L
totalTi…  Named num 2.38
training  70
```

Figure 5 Accuracy of Random Forest Classifier



```
cat("\nStep 9: Model Building -> ", modelName)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula

model    <-  naiveBayes(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.control(uses
model    <- model & randomForest(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.con
model





#-----------------------------------------------------------------
# Step 10: Prediction (Testing)
#-----------------------------------------------------------------
cat("\nStep 10: Prediction using -> ", modelName)
Predicted <- round(predict(model, testDataset))
head(Predicted)
PredictedProb <- predict(model, testDataset)
```

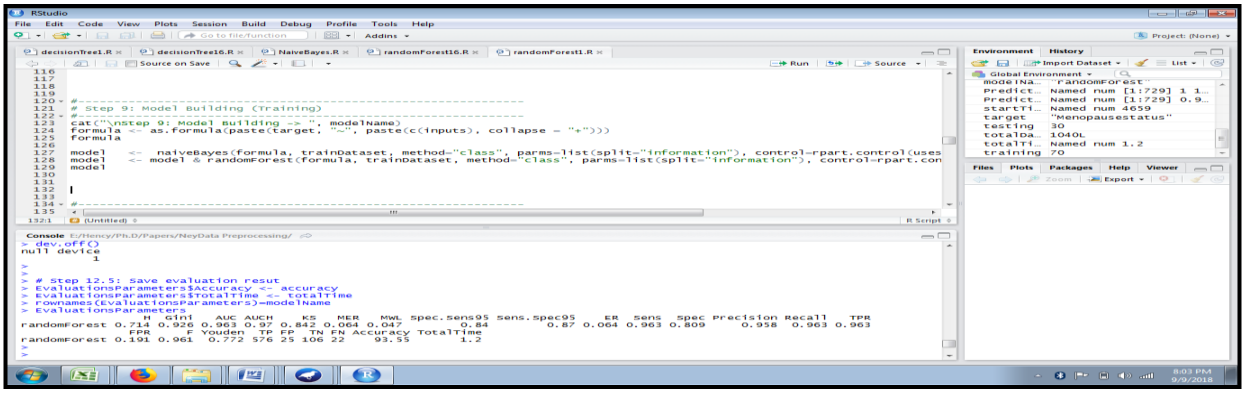
Console E:/Hency/Ph.D/Papers/NeyData Preprocessing/

```
> dev.off()
null device
          1
>
>
> # Step 12.5: Save evaluation resut
> EvaluationsParameters$Accuracy <- accuracy
> EvaluationsParameters$TotalTime <- totalTime
> rownames(EvaluationsParameters)=modelName
> EvaluationsParameters
                 H  Gini   AUC  AUCH    KS   MER   MWL Spec.Sens95 Sens.Spec95   ER  Sens Spec Precision Recall   TPR  FPR
randomForest 0.701 0.918 0.959 0.966 0.822 0.067 0.053        0.82       0.866 0.07 0.955 0.82      0.96  0.955 0.955 0.18
                 F Youden  TP FP  TN FN Accuracy TotalTime
randomForest 0.957  0.774 569 24 109 27       93      2.13
>
>
```

```
modelNa…  "randomForest"
Predict…  Named num [1:729] 1 1…
Predict…  Named num [1:729] 0.6…
startTi…  Named num 4192
target    "Menopausestatus"
testing   30
totalDa…  1040L
totalTi…  Named num 2.13
training  70
```

Figure 6 Accuracy of Hybrid Model without feature minimization

The hybrid model for the performance evaluation was adopted by combing Naïve Bayes and Random Forest and it was represented in Figure 8. The hybrid model was applied to the dataset and the final accuracy was noted. The hybrid model provides the better accuracy than the individual model.



```
#------------------------------------------------------------------
# Step 9: Model Building (Training)
#------------------------------------------------------------------
cat("\nStep 9: Model Building -> ", modelName)
formula <- as.formula(paste(target, "~", paste(c(inputs), collapse = "+")))
formula

model    <-  naiveBayes(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.control(uses
model    <- model & randomForest(formula, trainDataset, method="class", parms=list(split="information"), control=rpart.con
model


|

#------------------------------------------------------------------
```

```
> dev.off()
null device
          1
>
>
> # Step 12.5: Save evaluation resut
> EvaluationsParameters$Accuracy <- accuracy
> EvaluationsParameters$TotalTime <- totalTime
> rownames(EvaluationsParameters)=modelName
> EvaluationsParameters
                 H  Gini   AUC  AUCH    KS   MER   MWL spec.sens95 sens.spec95    ER  Sens  Spec Precision Recall   TPR
randomForest 0.714 0.926 0.963 0.97 0.842 0.064 0.047        0.84        0.87 0.064 0.963 0.809     0.958  0.963 0.963
               FPR     F Youden  TP FP  TN FN Accuracy TotalTime
randomForest 0.191 0.961  0.772 576 25 106 22    93.55       1.2
>
>
```

```
modelNa…  "randomForest"
Predict…  Named num [1:729] 1 1…
Predict…  Named num [1:729] 0.9…
startTi…  Named num 4659
target    "Menopausestatus"
testing   30
totalDa…  1040L
totalTi…  Named num 1.2
training  70
```

Figure 7 Accuracy of Hybrid Model with feature minimization

Table 3 Performance Evaluation of machine learning models

Classifier Model	**Accuracy without Feature Minimization**	**Accuracy With Feature Minimization**
Random Forest	92.59	92.59
Naivebayes	87.93	88.48
Hybrid Model	93.00	**93.55**

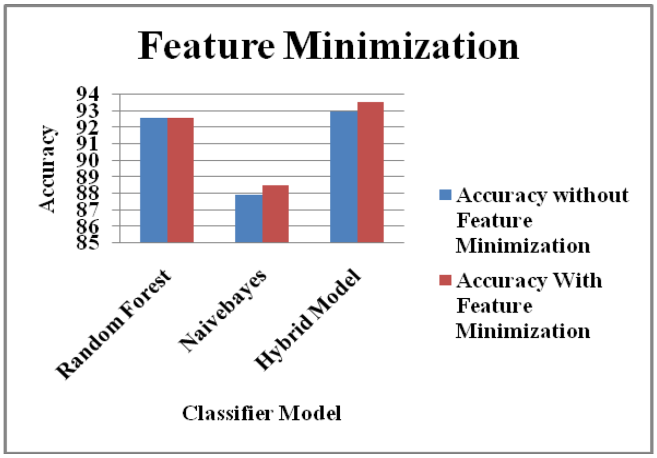


Figure 8 Performance Evaluation of machine learning models

V. Conclusion and Future Scope

To improve the performance of the mining process, pre-processing techniques and feature extractions was applied to the data set. The classifiers accuracy was measured for the dataset with all the features and for the dataset with minimized features. The accuracies of the model with feature minimization are greater than without feature minimization. The Classifier model such as Navie Bayes and Random Forest was performed well on the minimized features. The accuracy of the Random forest is 92.59% for both minimized and full features. When the Random forest was combined with Naïve Bayes, the performance was 93% for without feature minimization and for with feature minimization, the accuracy was 93.55%. The performance of the hybrid model was improved very much. We got 93.55% of accuracy through the hybrid model. Now the minimized data was ready for the further processes.

In future we can implement this using ensemble model. If the data is travelled through different models then the models will perfectly learn the data to provide reliable and accurate results.

V. Acknowledgment

The author would like to thank Mr. Julius Gnanamoney, Radiotherapy Clinical Specialist, Kent Oncology Centre, Maidstone Hospital, Hermitage Lane, Maidstone, ME16 9QQ, London, Who supported a lot and provided the detailed information on Endometrial Cancer.

The author also would like to thank Dr. V.G.Sudhakaran, Head Department of Radiation and oncology, and Dr. Arul Pirakasam, International Cancer Institute, Neyyoor, 629802, for the support and guidance.

References

[1] A detailed guide – endometrial cancer. Available in http://www.cancer.org/cancer/endometrialcancer/what-is-endometrial-cancer.

[2] A detailed guide – endometrial cancer. Available in www. cancer. org/ cancer/ endometrial cancer/cancer-risk-factors.

[3] Chia-Ming Wang, Yin-Fu Huang, Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data, Elsevier Journal, Expert Systems with Applications 36 (2009) 5900–5908

[4] E. Friberg & N. Orsini & C. S. Mantzoros & A. Wolk, Diabetes Mellitus And Risk Of Endometrial Cancer: A Meta-Analysis, Springer-Verlag 2007.

[5] F.Angiulli and C. Pizzuti, "Outlier Mining in Large High-Dimensional Data Sets", IEEE Trans. on Knowledge and Data Engineering, vol. 17, no. 2, pp. 203-215, 2005.

[6] Faysal A Saksouk, MD; Chief Editor: Eugene C Lin, MD. Endometrial Carcinoma Imaging.

[7] Wrathematics (27 August 2011). "How Much of R Is Written in R". librestats. Retrieved 2018-08-07.

[8] "7 of the Best Free Graphical User Interfaces for R". linuxlinks.com. Retrieved 9 February 2016.

[9] Huan Liu and Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Transactions on Computers, Pages 191 - 226

[10] J.Han, M.Kamber, J.Pei. Data Mining concepts and Techniques. 3rd Edition, Simon Fraser University, pages-230-240, 2006.

[11] M. Dash. Feature selection via set cover. In Proceedings of IEEE Knowledge and Data Engineering Exchange Workshop, pages 165–171, 1997.

[12] Mark Hall and Geoffrey Holmes, Benchmarking Attribute Selection Techniques for Discrete Class Data Mining, Mark Hall and Geoffrey Holmes, 2002

[13] Mark Hall, correlation based feature selection for district and numeric machine learning, In proc. Of 17th International conference on Machince Learning, ICML2000.

[14] P. Mitra, C. A. Murthy, and S. K. Pal. Unsupervised feature selection using feature similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3):301–312, 2002.

[15] P. Ravisankar, V. Ravi, G. Raghava Rao, I. Bose, Detection of financial statement fraud and feature selection using data mining techniques, Elsevier Journal Decision Support Systems 50 (2011) 491–500

[16] Ramya Rathan, Sridhar R, Balasubraminian S, Association Rule-Spatial Data Mining Approach For Exploration Of Endometrial Cancer Data, International Journal Of Advanced Research In Computer Science And Software Engineering, Volume 3, Issue 10, October 2013.

[17] The Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Endometrial Carcinoma. Nature. May 2, 2013. DOI: 10.1038/nature12113.

[18] Wei-Shinn Ku, A Bayesian Inference-Based Framework for RFID Data Cleansing, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 10, Page - 2177, 2013

[19] K. Pavya, B. Srinivasan, "Enhancing Wrapper Based Algorithms for Selecting Optimal Features from Thyroid Disease Dataset", International Journal of Computer Sciences and Engineering, Vol.6, Issue.3, pp.7-13, 2018.

[20] B. Bakariya, G.S. Thakur, "Effectuation of Web Log Preprocessing and Page Access Frequency using Web Usage Mining", International Journal of Computer Sciences and Engineering, Vol.1, Issue.1, pp.1-5, 2013.

[21] G. Pandey, N. Mishra, "*Optimal Feature Selection in Stream Data Classification Using Improved Ensemble Classifier for High Dimension Data*", International Journal of Computer Sciences and Engineering, Vol.4, Issue.9, pp.12-18, 2016.

Author Profile

Mrs. A.Hency Juliet pursed Bachelor of Science from Manonmaniam Sundaranar University, India in 1995 and Master of Science from Manonmaniam Sundaranar University, India in 1997 She is currently pursuing Ph.D. and currently working as Assistant Professor and Head in Department of Computer Application, MGC, University of Madras since 2005. She has published five research papers in an international journals. Her main research work focuses on Data Mining. She has thirteen years of teaching experience and three years of Research Experience.