CHAPTER 4

CLASSIFICATION ACCURACY ENHANCEMENT THROUGH META-LEARNERS

4.1 INTRODUCTION

As the size of the data increases, most of the conventional feature ranking models are difficult to find the essential key features due to high computational time and memory. Feature selection and ranking techniques contribute greatly to the data classification problem. In this chapter, a hybrid feature selection based ensemble classification learning technique to improve the feature selection and ranking problem in the mixed data types is developed. Most classification models of the conventional ensemble are processed with limited feature space and small data. With the size of the feature area increasing, conventional ensemble classification picks a preset number of classification features. A novel feature selection based classification model is used in this chapter to improve the accuracy and efficiency of high-dimensional data. The main goal of the proposed ensemble learner is to identify the data for complex pattern analyzes in high dimensions and imbalance. In this proposed model, the classifier features are classified by using feature selection based classification problem. In the proposed ensemble model, the output of high-dimensions medical data is analyzed using different base classifiers such as naive bays, feed forward neural networks, and an enhanced decision tree.

4.2 TRADITIONAL DATA ANOMALY BASED CLASSIFICATION

The purpose of data mining is not to reveal individual information, but to generalize it across populations. Data mining works by examining the privacy concerns of individual data subjects. The real issue, isn’t how to extract data but how to protect data. To obtain appropriate anomaly care, a patient must submit very sensitive information to a healthcare provider. In order to obtain good treatment, the hospitals must preserve their privacy in patient records, as patients give a personal data to the hospital. The direct divulgation of anomaly data inevitably puts individuals' privacy at risk. The idea of data conservation (PPDM) was introduced in order to address this problem [68] in order to relieve the tension between data mining and data protection. The concept of [69] is to shake individual data values as a first introduction to PPDM.

Since the management tasks of other companies are outsourced, it is necessary to share data. The process of recruiting an external company to provide a service provided previously by an employee is known as outsourcing. Outsourcing involves transitioning administrative responsibility for service provision and internal employee patterns to an external company. The health care provider outsources a range of activities and also that better efficiency, cost savings and increasing core business management time are the main benefits of using external resources [70].

[71] established that the selection of features involved specific redundant analyses. Thus, via redundant analysis, a new paradigm for efficient selection of features was developed. This system separated the analysis and redundant analysis applicable. A new feature selection algorithm was also introduced, and the best feature set was extracted against various learning algorithms. [72] investigated the current ranking algorithms for the selection of functions. They found that the rating and classification varied markedly. Hence, an algorithm for selecting features was proposed. Two ranking models, Ranking SVM and Rank Net, were used in this algorithm for extracting the best subset of features. [73] proposed an algorithm for feature selection using Particle Swarm Optimization (PSO) and Support Vector Machines (SVMs). The PSO had been used to pick the best subset of apps. PSO's fitness function was tested using the SVM with the one-versus-rest approach. The algorithm has been tested with different classification problems. [74] identified a method of selecting features, namely consistent-based selection of features. It was a valuable indicator for various methods of selection of the devices. The proposed method thus obtained a better result in accuracy than the wrapper approach and also obtained a higher reduction in functionality. [75] proposed an algorithm for selection of features using the mining law of association and knowledge gain. The Apriori algorithm was used to find the attributes in question. Knowledge Gain was used in the dataset to delete the obsolete and redundant features. The analysis of the algorithm showed that the classification accuracy had not improved. [76] implemented a system of selection of features using rating criteria based on the filter. The proposed technique has been called TBFS (Threshold Dependent Feature Selection). Using the F-measure, the value of each attribute was normalized between 0 and 1 and the independent attribute was individually paired with the class attribute. This technique was useful in identifying the smaller subset of features, and showed an increase in the accuracy of classification.

[77] developed a subset selection algorithm for high-dimensional data based on the Clustering function. To separate the features into clusters, the graph-theoretic approach was used. The features which were closely related to the target class were chosen as the best subsets of features. They viewed every cluster as a single trait in this. Consequently, the dimensionality was reduced significantly. The algorithm was compared with various existing algorithms and the prediction accuracy and classification performance showed a minimum improvement. [78] used knowledge theory to examine the discriminatory selection feature algorithm which did not recognize the discriminating and continuous features of the dataset. They also proposed an algorithm for selection of features from the analysis. An Entropy breakpoint definition has been implemented in this algorithm. This algorithm has been tested using various real-world datasets. The experiment result stated that the algorithm had a high computational complexity with very low precision prediction. In Feature Selection, [79] proposed an algorithm for solving the optimization problem. Using Greedy Search method and Greedy Search Loss from ranking method, this algorithm was used to find the best features. [80] introduced a new algorithm called Relief Disc. It functions on Discretization basis. Discretization partitions contains adjacent intervals in finite range. Rather than using random sampling to pick the instance, they suggested taking instance from each interval that reduces computational complexity and retains the consistency of features. There was also no need for user feedback to parameter the sample size. Experimental tests have shown that the current algorithm works better compared to the existing Relief algorithm. Relief-Disc achieved better than Relief according to them. The stochastic search approach is implemented in order to address the drawbacks of the conventional approaches, where some randomness is added in the search process and feature selection process is less sensitive to the specific dataset. The significance of the features is assessed in this process, their consistencies are calculated and a pre-determined search algorithm is used along with a classifier. The consistency of a subset of features is calculated in terms of accuracy of the classification. The function subset contributing the highest accuracy of classification with a minimum number of features is considered to be the best. Different forms of Evolutionary Algorithms (EAs) are used to pick features. However, several algorithms for decision tree learning are degraded in their output performance because of inappropriate, unreliable or indefinite data are presented. Sometimes interdependent relationship among data is taken into consideration while some algorithms limitedly handle the attributes which have discrete

values. In order to determine the missing value of microarray [81], they built a new bicluster based Bayesian Principal Component Analysis. A common microarray missing value estimation method is the Bayesian main component analysis. There is no noticeable overall performance of that system. In this work we have implemented a bi-cluster-based technique to manipulate the local matrix structure. Every bi-cluster gene is identified in the above-mentioned technique along with experimental conditions. In order to get optimum parameters, an automatic training mechanism is implemented. Additional research may be performed in future to substantially reduce the normalized root-mean-square error. Each gene expression can be calculated simultaneously with the theory of microarray. Some conditions are kept and there is a total of 10 of these conditions. Clustering and bi-clustering are here to be listed as two essential tools for the study of the data on gene expression. The data above was obtained from different experiments with microarrays. In a single group all genes with related behavior are included. All the biological knowledge needed is successfully gathered in this process. A specific gene can play many roles, so non-exclusive mechanisms for grouping are required. Gene shaving is used to detect various gene sets with identical expression. On the other hand, genome biology is considered to be the most popular form of clustering with only stable clusters. The value of variance is very high in all these clusters. There are therefore opportunities to overlap between clusters. In this work, an intelligent framework for analysis of microarray data has been implemented. The above-mentioned method refers to the full clustering and biclustering process three numbers of non-exclusive techniques. The main purpose of this work is to detect consistent clusters of genes with large variances between samples [82].

A fuzzy, weighted grouping technique is applied to solve the above-mentioned problems. In each process, this method is congruent and distributed. This technique can be applied in various applications for data clustering. It can be used to obtain the functional relations of different genes in clusters of gene expression. By incorporating ordered weighted averaging and spearman coefficients techniques, the above presented method assesses the similarity measure. Here, density reachable genes are combined to form subclusters. In this work. Finally, through the combination of these subclusters, the finished cluster results were regenerated. An effective voting system that is responsible for deciding the best weights is then introduced. In addition, valid clusters between each individual dataset may also be defined. In a distributed processing environment, the above algorithm is implemented.

4.3 GRAPH-BASED MEDICAL DATA CLASSIFICATION

A graph based medical data classification and feature extraction, extracts required phrase or instances for the large medical data set [83]. The graph-based ranking approach is applied to compute the phrase and sentence scores using similar navigation paths or clustered paths. Using graph model, medical data are modeled as vertices on the graph and they are ordered using the algorithm. When using sparse non-negative matrix factorization instances with similar features, clustering is applied. The most frequent input examples are used to improve the feature extraction. A graph is used to represent medical data of a patient. The medical data is represented as a graph with each line in the file labelled as a vertex. Just like in a graph, there are similarities between the vertices. This model recognizes several limitations; firstly, classification efficiency and association models are influenced by parameter settings, discretion and feature choices. Secondly, due to sparsity and NULL values, this model could not find associations. Third, due to incorrect labeling, the wrong positive rate is increasing.

The new method on discretization of the attributes of the decision tree learning has been suggested by [84]. Each attribute is assumed to be a nominal or categorical attribute as the key problem of conventional decision tree models. To solve this problem, each attribute during the tree construction process is fed a dynamic discretization model on the continuous label. C4.5 used the discretizing approach in preprocessing phases with the elimination of the noise, in conventional decision tree models such as CART. However, the key constraint of this model is that the data must be of a continuous form and not a mixed type. They proposed to construct ensemble trees using the univariate decision tree model as an assembly framework for random project random discretization method. Several models were suggested for the design of ensemble decision-making trees. Randomization is used to create decision tree classifiers based on boosting and bagging. Uniform decision tree models such as ID3, C4.5 are limited to generating a specific classification attribute or node. This maximizes the working time of the decision tree during the induction process to break the instance space. Multivariate decision trees are therefore used in numerical feature sets to maximize the efficiency and runtime of decision tree construction.

The construction in machine learning models of precise ensemble classificators has been an active research subject. Boosting, bagging and stacking are the basic

models of master learning to learn high-dimensional datasets for ensemble classification. Due to high dimensionality and initialization parameters, most conventional ensemble classificators have high error rates. As a way of growing the real good rate in restricted space training, [85] suggested a boosting training model known as Tradaboost. As dimensionality reduces the true positive rate, the principal limitation of this model is. [61] Suggested an expanded Trbagging process model. Proposed The weak classifiers on the training data are learned in this method.

A PSO-based, high dimensional features spectral filter model of the original training data was implemented [86]. The one is the primary component analysis and the other the highest risk assessment. Two reconstruction approaches are employed. In randomization models, various distribution algorithms have been used. In the majority of these methods, Bayesian analytics use randomization and randomization data to estimate the initial data distribution. Feature selection is a significant step in the process of defect prediction and was studied extensively in the machine learning field. [87] Investigated and applied to UCI repository data sets various models of the selection feature representing a classified list of features.

4.4 RANK BASED FEATURE SELECTION CLASSIFICATION MODEL WITH MULTIVARIATE ANALYSIS

One of the main issues to be considered in a scenario for fraud detection is the unbalanced nature of the dataset. The applications for fraud detection require actual data rather than artificial data to provide accurate results. The data set providers will not be able to provide the data as such because they are bound by the confidential law. Until making it public, data anonymization is performed. It will result in misclassifications using the data. Therefore, prior to use, data should be cleaned. Extraction, transformation and loading are the three general phases. But the cleaning process should be adapted to the data being used in real terms. Initially Data Pre-processing has been carried out. In this process, missing values or inconsistent values are replaced with the computed value. If the attribute is numerical then all the missing values are replaced with the computed Max-Min value. If the attribute is categorical, then all the missing values are replaced with probabilistic ranked value. Then on filtered data Feature selection based classification model has been implemented with multivariate analysis.

The applications for fraud detection require actual data rather than artificial data to provide accurate results. The data acquisition companies have to refuse providing the information under legal obligation. Data recognition is accomplished before providing the monitoring function to the general public. It will result in misclassifications of the data. The three phases included in the proposed framework are feature extraction, feature transformation and anomaly detection.

Feature Generation: Scoring the stimulus features isn't an easy task. We based our results on the previous studies to include features already common in the literature. New features can be constructed by transforming or combining the original attributes. This approach is known as feature construction. It is often done by incorporating expert's background knowledge about the problem domain.

Feature Selection: Selecting the most suitable set of attributes that represent a problem, from large set of attributes is also a challenging task. Some attributes might be irrelevant, redundant, or containing useful information only when combined together. We must select the best possible features before feeding them into the algorithm since this influence the quality of the prediction model as well as the computing resources (such as calculation time, memory usage etc.).

Feature selection is important for the purpose of customer features using various feature combinations. Classification models can be evaluated with regards to classification error rate. So, the best performing feature combination is the feature combination to select from all features.

In feature selection, there are two fundamental search procedures, the forward and backward selection. Forward selection starts from scratch and adds new variables one-by-one while evaluating the optimal search path. The backward selection does the opposite: the search starts from a model based on all variables and eliminates one-by one. The results of both approaches can differ due to non-independent variables and different stopping points when a certain quality threshold value is reached. In other wrapper application fields also other search techniques such as evolutionary search and simulated annealing are used.

Therefore, prior to use, data should be cleaned. Extraction, transformation and loading are the three general phases. But the cleaning process should be adapted to the data being used in real terms.

4.4.1. Data Pre-Processing Algorithm

Database D,

For each data record in D

Do

For each feature F in the record

Do

If(F!=NULL)

Then

Continue;

Else

F_type=check_type(F);

If(F_type==numerical)

Then

Miss_Value= $\frac{Max(F)*\sigma_F^2 - Min(F)*\mu^2{}_F}{N(N-1)*[\text{Max(F)} - \text{Min(F)}]}$

Value(F)=Miss_Value;

End if

If(type==Categorical)

Then

Freq[]=frequency(F);// each category of class attribute.

Probability of each instance value per class.

Prob[]= $\sum_{i=1}^{m} \Pr ob(x_j / C_i)$;

i=1,2,3…m classes

j=1,2…n instances

rank=Max{freq[]}/Max{Prob[]};

Fill the value with the max ranked class value.

End if

Done

Done

In this algorithm, missing values or inconsistent values are replaced with the computed value. If the attribute is numerical then all the missing values are replaced with the computed Max-Min value. If the attribute is categorical, then all the missing values are replaced with probabilistic ranked value.

4.4.2. Fraud Detection Attribute Selection Algorithm

Input: Filtered data FDB

Output: Ranked feature attributes

Procedure:

For each filtered feature ff in FDB

Do

Compute entropy E(ff);

Compute mutual information between attributes.

MI(ff)=Max{MI{ff,F-ff}};

Partition the feature ff into m classes as $\mathrm{p}_1, \mathrm{p}_2,,, p_m$

Find the similarity between instances of two distinct partitions as

$$\mathrm{Sim}(\mathrm{p}_i, p_j) = \frac{2 * \sum_{i,j} | x_i - x_j |^2}{N_i(N_j - 1)};$$

Where N_i is the number of instances in ith partition and N_j is the number of instances in j thpartition.

Rank of the attribute is defined as

R(ff)=E(ff)+M.I(ff)+Max{ $\mathrm{Sim}(\mathrm{p}_i, p_j)$ }

Done

Input k as user defined threshold

For each r in R(ff) do

If(r>k)

Then

Select as fraud feature attribute.

Else

Continue;

done

In this algorithm, rank based fraud detection attributes are selected using a novel approach. In this model, entropy and mutual information measures are computed to each attribute with the remaining attributes. Also, similarity measure is computed to all the data partitions for intra cluster variations. The rank of an attribute is computed using the entropy, mutual information and similarity measure. Feature attributes are selected using the user defined threshold.

4.5 EXPERIMENTAL RESULTS

In this section, experimental findings will be simulated on various datasets using the proposed model in line with existing approaches. In this the data collection [14] used in the experimental study is summarized. As test data for performance measurement in the experimental results, in this 10 % of the training data are used. Proposed ensemble methods that improve true positive performance and accuracy on entire high-dimensional datasets. For the construction of decision patterns, the proposed model uses the full training data set, therefore the accuracy of predicting every cross validation is more reliable than the conventional ensemble classification patterns. It is evident from the experimental results that feature selection optimization improves in the simple classifications along with a truly positive and negative classification scale. In Network (KDD) dataset the classification rate increased to 98.2%, which is better accuracy than the previous models (94%), In credit dataset the classification rate increased to 97.3%, which is better accuracy than the previous models (82%) In Diabetes dataset the classification rate increased to (97%), which is better accuracy than the previous models (74.5%)

Results on Medical Datasets

In this section, we have implemented our proposed cancer micro-array model and compared the results with conventional decision tree modeling. Table summarizes the Microarray data set used for experimental evaluation. 10% of the training data are used as test data for performance evaluation in the experimental results.

Ensemble methods allow more accurate prediction of true positives of high dimensional datasets. The proposed model uses the entire training data for construction of decision patterns, therefore its prediction accuracy tends to be more accurate than the traditional ensemble models.

Table 4.1: Datasets and its properties

Dataset Name	Features	Type
lung-Michigan	7000	Numeric
DLBCL-Stanford	4000	Numeric
Leukemia	6817	Numeric
lungCancer_train	12000	Numeric
Lymphoma	4026	Numeric

Ovarian Cancer dataset:

Ovarian cancer dataset is one of the high dimensional continuous attributes dataset. In this data a large number of feature space with limited number of instances are taken from the website(http://csse.szu.edu.cn/staff/zhuzx/Datasets.html). This dataset contains 15154 attributes and 253 instances for classification process.

Table 4.2: Performance analysis of classification accuracy on ovarian dataset on 10 fold cross validation.

10 % Test Data Ovarian Dataset			
Model	**True Positive Rate**	**Error Rate**	**Run time(ms)**
PSO + Naïve Bayes	0.87	0.2492	4526
PSO + Random Forest	0.8923	0.2293	4297
PSO + Neural Network	0.9055	0.1987	3975
Statistical+Decision tree (Proposed Model-1)	0.9645	0.012	3183
Proposed-Ensemble (Proposed Model-2)	0.987	0.023	3046

Table 4.2 Describes the performance of the improved feature selection algorithm with ensemble classification to the traditional models on microarray cancer das. From the table, it is clearly observed that the proposed model has high accuracy and less error rate compare to traditional models in terms of error rate, accuracy and runtime.

Table 4.3: Comparative study of present model to the traditional models on four microarray datasets by using accuracy measure

Datasets	**PSO + CART**	**PSO + SVM**	**PSO + FFNN**	**Statistical + Ensemble**	**FS + Ensemble**
Lung Cancer	0.7144	0.8353	0.8525	0.9535	96.03
Lung Michigan	0.8143	0.8353	0.845	0.9635	0.9903
Ovarian	0.8353	0.8744	0.8574	0.97	0.984
Lymphoma	0.8467	0.8835	0.8934	0.982	0.9845
Error Rate	0.3144	0.3042	0.294	0.174	0.201
Runtime (ms)	6983	6194	6364	3068	3465

Table 4.4: Comparative study of present model to traditional models by using average recall and precision rate on the training microarray datasets

Model	Recall	Precision
PSO + Naïve Bayes	0.845	0.864
SVM + Random Forest	0.885	0.9054
PSO + FFNN	0.9143	0.9133
IPSO + Ensemble	0.9735	0.9735
FS + Ensemble	0.9824	0.9819

Table 4.4, illustrates the performance of the present model to the existing approaches by using average recall and precision rates. These average rates are simulated on the training microarray datasets.

Pima indians diabetes mellitus dataset is taken from the UCI machine learning website. There are 768 instances of this data set. Each person is identified by 8 attributes in the following data set. The number values are all attributes. This data set was created by UCI Machine Learning Databases Repository. The data were gathered from a larger data set of mellite and digestive and kidney disease national institutes.

Diabetes dataset patterns:

[preg=3]: 75 ==> [insu=0]: 30
[age=22]: 72 ==> [insu=0]: 29
[preg=5]: 57 ==> [insu=0, class=tnegative]: 23
[pres=70]: 57 ==> [class=t_positive]: 23
[age=21, class=tnegative]: 58 ==> [insu=0]: 25
[preg=5]: 57 ==> [skin=0, insu=0]: 25
[preg=5]: 57 ==> [skin=0]: 25
[age=21]: 63 ==> [insu=0]: 28
[pres=70]: 57 ==> [insu=0]: 26
[class=tnegative]: 500 ==> [insu=0]: 236
[preg=0, class=tnegative]: 73 ==> [insu=0]: 35
[preg=0]: 111 ==> [insu=0]: 54
[class=t_positive]: 268 ==> [insu=0]: 138
[preg=4, class=tnegative]: 45 ==> [insu=0]: 24
[preg=7]: 45 ==> [class=t_positive]: 25
[preg=7]: 45 ==> [insu=0]: 25
[pres=80]: 40 ==> [insu=0]: 23
[preg=4, insu=0]: 41 ==> [class=tnegative]: 24
[insu=0, class=tnegative]: 236 ==> [skin=0]: 139
[preg=2, insu=0]: 39 ==> [skin=0]: 23

[preg=0, insu=0]: 54 ==> [skin=0]: 32
[pres=70]: 57 ==> [class=tnegative]: 34
[preg=4]: 68 ==> [insu=0]: 41
[insu=0]: 374 ==> [skin=0]: 227
[skin=0]: 227 ==> [insu=0, class=tnegative]: 139
[skin=0, insu=0]: 227 ==> [class=tnegative]: 139
[skin=0]: 227 ==> [class=tnegative]: 139
[preg=6]: 50 ==> [insu=0]: 31
[pres=78]: 45 ==> [class=tnegative]: 28
[insu=0]: 374 ==> [class=tnegative]: 236
[preg=8]: 38 ==> [insu=0]: 24
[preg=5]: 57 ==> [class=tnegative]: 36
[preg=5]: 57 ==> [insu=0]: 36
[pres=72]: 44 ==> [class=tnegative]: 28
[insu=0, class=t_positive]: 138 ==> [skin=0]: 88
[preg=5, insu=0]: 36 ==> [class=tnegative]: 23
[preg=5, class=tnegative]: 36 ==> [insu=0]: 23
[preg=3]: 75 ==> [class=tnegative]: 48
[preg=0, insu=0]: 54 ==> [class=tnegative]: 35
[preg=0]: 111 ==> [class=tnegative]: 73
[preg=4, insu=0]: 41 ==> [skin=0]: 27
[preg=4]: 68 ==> [class=tnegative]: 45
[pres=74]: 52 ==> [class=tnegative]: 35
[pres=80]: 40 ==> [class=tnegative]: 27
[preg=6]: 50 ==> [class=tnegative]: 34
[preg=5, insu=0]: 36 ==> [skin=0]: 25
[pres=64]: 43 ==> [class=tnegative]: 30
[pres=62]: 34 ==> [class=tnegative]: 24
[age=25]: 48 ==> [class=tnegative]: 34
[age=28]: 35 ==> [class=tnegative]: 25
[pres=68]: 45 ==> [class=tnegative]: 33
[age=27]: 32 ==> [class=tnegative]: 24
[age=26]: 33 ==> [class=tnegative]: 25
[preg=1, insu=0]: 41 ==> [class=tnegative]: 32
[preg=1]: 135 ==> [class=tnegative]: 106

Table 4.5: Number of instances, attribute with computed ranked attributes of diabetes data for feature selection and ranking

Number of instances	Number of features	Ranked Attributes
500	10	6
700	14	8
900	16	11
1200	18	15
1500	21	14

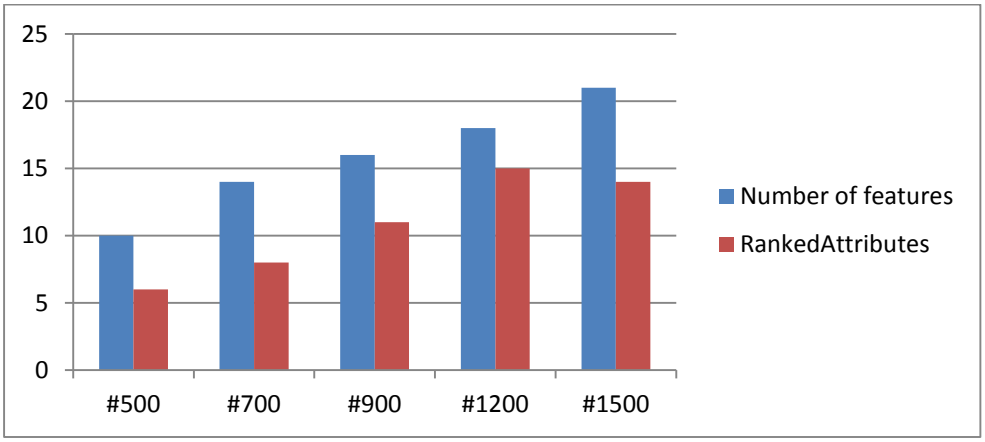

Figure 4.1: Comparison of the number of features and selected ranked attributes on PIMA dataset.

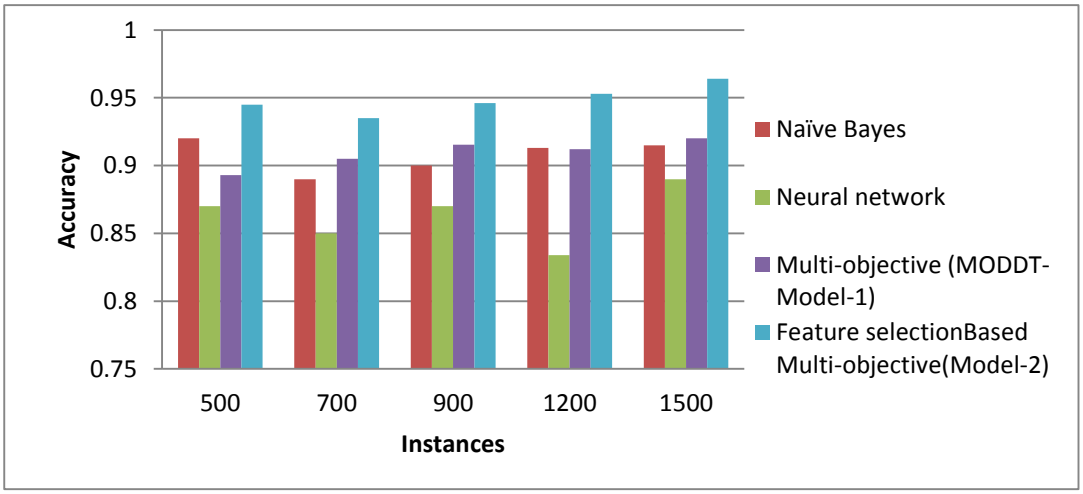

Figure 4.2: Accuracy Comparison of Proposed And Existing Algorithms

A model of selection of credit fraud dataset has been introduced on the basis of this Rank. Anomaly attributes can be found using traditional models since the size and number of instances are high, and are inefficient and ineffective. The hybrid function

selection algorithm is used to extract effective ranking attributes. Experimental findings show that, as opposed to conventional models, the proposed framework effectively recognizes the related attributes in terms of time and dimensions.

Table 4.6 Accuracy Comparison on three datasets (KDD, Credit, Diabetes)

Dataset	KDD	Credit	Diabetes
Correlation+knn	94.4	65.9	69.14
Correlation+Naive base	91.5	75.52	74.2
Correlation+Random forest	93.4	71.3	73.95
Correlation+Random tree	92	65.1	71.74
InfoGain + knn	94.4	65.9	69.14
InfoGain +Naive base	91.5	75.52	74.2
InfoGain +Random forest	92.8	70.4	73.69
InfoGain +Random tree	88.5	63.2	69.92
GainRatio + knn	94.4	65.9	69.14
GainRatio +Naive base	91.5	75.52	74.2
GainRatio +Random forest	93.7	71	72.65
GainRatio +Random tree	89	64.8	69.66
ReliefF + knn	94.4	65.9	69.14
ReliefF +Naive base	91.5	75.52	74.2
ReliefF +Random forest	93.5	70	73.69
ReliefF +Random tree	87.5	64.2	69.92
Model-2 Feature Selection +MODD Tree	**98.2**	**97.39**	**97.86**

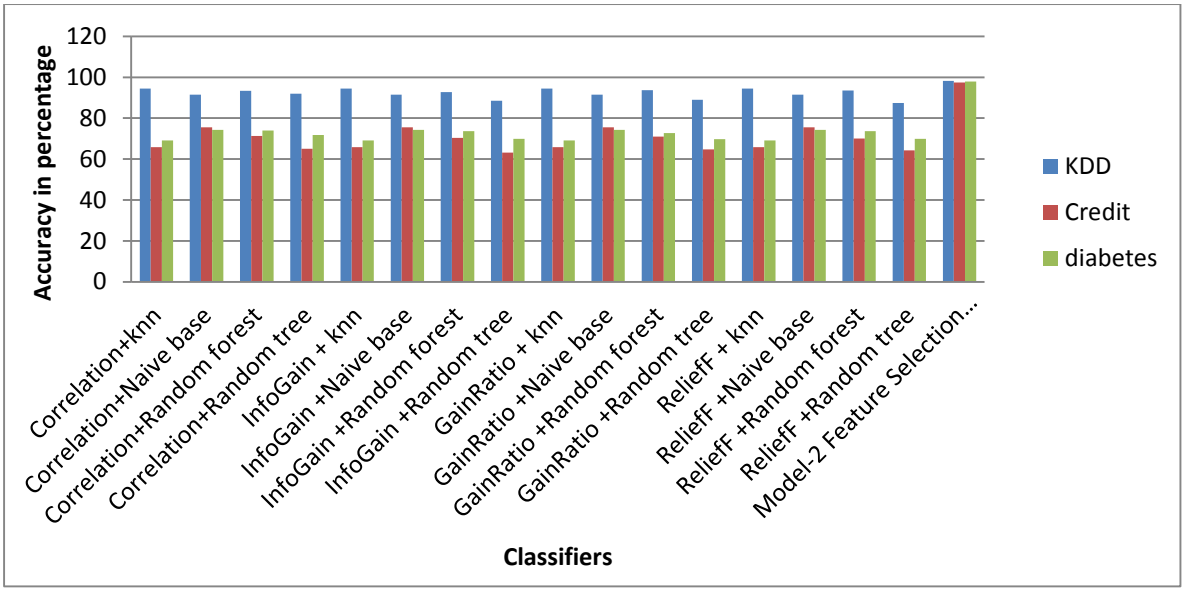

Figure 4.3: Comparative analysis off proposed model to conventional models for accuracy computation on different datasets.

Experimental results show that proposed system efficiently detects the relevant attributes compared to traditional models in terms of time and dimensions are concerned.

4.6 SUMMARY:

In this chapter, hybrid feature selection based classification model is proposed on the mixed type of datasets such as medical, credit card, and KDD datasets. In this work, an advanced feature selection is applied on the mixed type of attributes for attribute ranking and classification problem. An advanced decision tree classifier is used to improve the classification accuracy and error rate on the different types of data types. In Network (KDD) dataset the classification rate increased to 98.2%, which is better accuracy than the previous models (94%), In credit dataset the classification rate increased to 97.3%, which is better accuracy than the previous models (82%) In Diabetes dataset the classification rate increased to 97%, which is better accuracy than the previous models (74.5%). Hybrid Feature Selection classification model gives the better disclosure of diabetes, right kind of of glass, better perceiving of network attacks, cancer decease and credit risks in the credit dataset. Experimental results proved that the filter based multi-objective classification model has better accuracy and feature ranking on different large feature space.