بسم الله الرحمن الرحيم

Sudan University of Science and Technology

College of Graduate Studies

College of Computer Science and Information Technology

Supplementary Research for the master Degree titled:

Effect of Features Selection Method for Kidney Disease Classification Using Naïve Bayes and J48 Classifiers

تأثير طريقة إختيار الخصائص في تصنيف بيانات مرضى الكُلى بإستخدام خوارزميتي Naïve Bayes و J48

Prepared by:

Tageldin Musa Bakheet Mohammed Noor

Supervision :

Dr. Ali Ahmed Alfaki

May 2018

الاستـــهلال

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

قال تعالى :

(يَا بُنَيَّ إِنَّهَا إِن تَكُ مِثْقَالَ حَبَّةٍ مِّنْ خَرْدَلٍ فَتَكُن فِي صَخْرَةٍ أَوْ فِي السَّمَاوَاتِ أَوْ فِي الْأَرْضِ يَأْتِ بِهَا اللَّهُ ۚ إِنَّ اللَّهَ لَطِيفٌ خَبِيرٌ (١٦) يَا بُنَيَّ أَقِمِ الصَّلَاةَ وَأْمُرْ بِالْمَعْرُوفِ وَانْهَ عَنِ الْمُنكَرِ وَاصْبِرْ عَلَىٰ مَا أَصَابَكَ ۖ إِنَّ ذَٰلِكَ مِنْ عَزْمِ الْأُمُورِ(17)

صدق الله العظيم

سورة لقمان الايات (١٦-١٧)

الإهـــداء

Dedication

To who reached the letter and led the honesty and advised the nation.

To the Prophet of mercy and the light of the worlds"

“Our master Muhammad peace be upon him”

To whom God has endowed with honor and glory

Teach me bestowal without waiting. For who I carry his name with all pride .

I ask God to extend in your age.

And your words will remain stars I will guide today and tomorrow and forever

“Dear father”

To my angel in life. to the meaning of love and to the meaning of compassion and Dedication

To the smile of life and the secret of existence to whom was her invitation the secret of my success

“Dear mother”

To my brothers and sisters

To those who love brotherhood and distinguished themselves by giving and to the springs of pure truth to those with whom I was happy with them

To whom I knew how to find them and taught me not to waste them.

“Brothers and my friends”

Acknowledgments

In the name of Allah the most Merciful Praise be to Allah, Lord of the Worlds, and prayers and peace be upon the of our master Muhammad and his family and companions, and thanks to God who helped me complete this research in its final form.

"He who does not thank people does not thank God"

Thanks all the thanks to

Dr. Ali Ahmed Alfaki

For the effort and help me in the research and ask God to make it in the balance of his deeds.

I would like to thank all who have contributed to Provide a helping hand to accomplish this search,

I especially thank **my teachers** at the Faculty of Computer Science and Information Technology - Sudan University of Science and Technology.

As well as thanks to the College of Graduate Studies

And

College of Computer Science and Information Technology

Sudan University of Science and Technology

Abstract

There are many fields today that use database systems to store data. Over time this has lead to the creation huge amount of data. The importance of these huge data is that are knowledge and relationships between them. When analyzing these data we get knowledge and facts that help decision makers to make the right decisions which are called **data mining**.The medical field is one of the most important fields to apply data mining techniques. The purpose of this study is to know the effect of features selection algorithms in increasing the accuracy of the classifier (model). We have applied this study on the Chronic Kidney Disease (CKD) dataset which contains 25 features used to diagnose (CKD) to know the effect of features selection algorithms in the increase classifier accuracy. In this study we have used naïve bayes and J48 classifiers with the wrapper features selection evaluator to select best features that have high effect in classifier accuracy and eliminate unimportant features through experiments we have noticed when used naïve bayes a classifier with the wrapper features selection evaluator the degree of accuracy increased from 95% to 99.5%. But when used the j48 classifier with wrapper features selection evaluator the degree of accuracy not significant.

المستخلص:

هناك العديد من المجالات اليوم التي تستخدم أنظمة قواعد البيانات لتخزين البيانات. ومع مرور الزمن أدى ذلك إلى إنشاء كمية هائلة من البيانات .وتتمثل اهمية هذه البيانات الضخمة في انها تكوّن معارف وعلاقات فيما بينها.وعند تحليل هذه البيانات نصل إلى معارف وحقائق تساعد متخذي القرار في اتخاذ القرارات الصحيحة وهذا ما يسمى **بتنقيب البيانات** . إن المجال الطبي هو واحد من أهم المجالات لتطبيق تقنيات تنقيب البيانات.الغرض من هذه الدراسة هو معرفة تأثير خوارزميات اختيار الميزات في زيادة دقة المصنف (النموذج) لقد قمنا بتطبيق هذه الدراسة على مجموعة بيانات أمراض الكلى المزمنة (**CKD**) التي تحتوي على (٢٥) خاصية تستخدم لتشخيص مرضى الكلى (CKD). وذلك لمعرفة تأثير خوارزميات اختيار الخصائص في زيادة وتحسين دقة المصنف. في هذه الدراسة لقد قمنا باستخدام خوارزميتي (naive bayes) و (j48) مع استخدام واحدة من خوارزميات اختيار الخصائص وهي خوارزمية ال(wrapper) وذلك لاختيار الخصائص التي لها تأثير عالي في زيادة دقة المصنف وإزالة الخصائص الاقل أهمية. ومن خلال التجارب لاحظت أنه عند استخدام خوارزمية (naive bayes) مع (wrapper) أدى ذلك لتحسين دقة المصنف من ٩٥% إلى ٩٩,٥%. اما عند استخدم خوارزمية (J48) مع (wrapper) لم تتغير دقة المصنف.

Table of Contents

CHAPTER (I)

CHAPTER (II)

CHAPTER (III)

CHAPTER (IV)

CHAPTER (V)

List of Figures:

CHAPTER (I)

1.1 Introduction

Chronic kidney disease (renal failure) has spread significantly in recent times, threatening the lives of many men, women, young people and children and affecting their health. Renal insufficiency is a term in medicine called in cases of kidney failure in the performance of its functions. There are two types of kidney failure, namely acute renal insufficiency and chronic renal insufficiency. Renal failure in general is a failure in the work of the kidney and its functions, leading to a general imbalance in the human body.There are many reasons that result in renal function failure. The most common causes are diabetes, high blood pressure, kidney inflammation (kidney glomerulonephritis), polycystic kidney disease which is a genetic disease that can cause kidneys to be formed and lead to their failure. Unknown reasons about 20% of dialysis patients never know what the real cause of kidney failure is these patients are often used the therapy for the first time after the Progress of renal failure, and at this stage it is difficult to determine the cause of the disease.

In present days, computers have brought significant improvements to technology that leads to the creation of huge volumes of data. Moreover, the advancement of the healthcare database management systems creates a huge number of medical databases. Creating knowledge and management of large amounts of heterogeneous data has become a major field of research, namely data mining [1].

Simply stated, data mining refers to extracting or “mining” knowledge from large amounts of data. Now-a-days, Data mining technique is combined with machine learning to extract hidden patterns as well as for analysis purposes [2].Data mining is defined as a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database.

Medical data mining has great potential like exploring the hidden patterns which can be utilized for clinical diagnosis of any disease dataset. There are two strategies to perform data mining namely supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used. Classification is a supervised learning used to discover hidden patterns from existing medical data [3]. Classification is very critical for therapy of patients also supervised learning techniques that are used in both medical and clinical research are Classification.

In this study we have focused on the usage of classification techniques in the field of medical science and bioinformatics. The main objective of the classification technique is to predict the target class accurately for each case in the data.

Feature selection is a technique to reduce the dimensionality. The main use of this method is to extract small subsets of relevant features from the original dataset based on evaluation criterion. Generally it is used to improve accuracy. **Subset selection** process is used in machine learning to select relevant subsets which contains the least number of dimensions. There are two approaches in feature selection known as forward and backward selection. The main idea is to choose a subset of input variables by eliminating features with little or no predictive information. These methods can be divided into three broad classes. One is Filter methods and another one is Wrapper method and the third one is embedded method [4].

1.2 Problem Statement:

The problem is that the diagnosis of kidney disease depends on a number of tests and a comprehensive medical examination of the patients. There are a large number of tests are taken to diagnose kidney failure, some of these tests are not important in the process of classification. Sometimes some of these tests may take a long time this may lead in some cases to the death of the patient. The problem also represent in using large number of features in classification leads to decrease classification model accuracy and performance. So this study focuses on the choice of the best features using wrapper evaluator to eliminate unimportant features to improve classification model accuracy and increased performance.

1.3 Significance:

The importance of this study is that the presence of Model (system depends on the computer) to classification Cases of patients helps physicians make the right decision for the diagnosis of the disease because the model has the ability to predict the results of some tests by deduced from relations Tests with each other And accurately in the fastest time possible Which helps in the process of Diagnose and determine treatment in time this will help in maintaining life People.

Also the Significance of this study represent in the using features selection evaluator to reduce dataset that leading to improve classification models accuracy and performance because this evaluators built base on statistic approaches to calculated correlation between this features and removed unimportant features.

1.4 Hypothesis:

1. Features Selection algorithms leads to improve accuracy model
2. Using all the features in dataset that lead to decrease the accuracy and performance.
3. The use of decision tree (j48) in the form designs ability to find the best results in terms of speed.
4. Model helps in the diagnosis of kidney patients quickly.

1.5 Objectives:

1. To apply naïve bayes (NBC) and (J48) classifiers based on all (CKD) dataset features (using original dataset) to classification data.
2. To apply the wrapper features selection algorithm to select best features and eliminate unimportant features.
3. To apply naïve bayes (NBC) and (J48) classifiers methods based on selected features (using reduced dataset).
4. To Compare accuracy before using wrapper (using original dataset)) with accuracy after using wrapper (using reduced dataset).

1.6 Top reasons to use feature selection because:

- It enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the right subset is chosen.
- It reduces over fitting.

1.7 **Scope**:

This search covers the Offline not online classification and considers The Chronic-Kidney-Disease dataset.

This study applied by using data groups mentioned above for kidney patients from UCI machine learning repository, temporal boundaries 2017 – 2018.

1.8 Thesis organization

This research consists five chapters. Chapter (I) represents an overview of the research. Chapter (II) Literature review is divided into four sections. First one describes data mining and data classification, section two represent naïve bayes and j48 classifier, and section three describes the features selection algorithms, and section four consist related work. Chapter (III) the proposed model is divided into three sections. Section one describes proposed model and section two describes material and tools used and section three represent dataset used and data pre-processing. Chapter (IV) explains experiments and results, and Chapter (V) conclusion recommendations.

CHAPTER (II)

Literature review

2.1 Introduction

This chapter divided into four sections. First one describes data mining and data classification, section two describes the features selection algorithms , and section three represent wrapper method, naïve bayes and j48 classifier, , and section four consist related work.

2.2 DATA MINING AND DATA CALSSIFICATION MODELS

2.2.1 Data mining definition:

Data mining refers to extracting or "mining" knowledge from large amounts of data [2]. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long. "Knowledge mining," a shorter term may not reflect the emphasis on mining from large amounts of data. Others view data mining as simply an essential step in the process of knowledge discovery.

2.2.2 Data mining steps:

Knowledge discovery as a process is depicted in *Figure 2.1* below and consists of an iterative sequence of the following steps

- Data cleaning (to remove noise and inconsistent data)
- Data integration (where multiple data sources may be combined)
- Data selection (where data relevant to the analysis task are retrieved from the database)

- Data transformation(where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
- Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
- Pattern evaluation (to identify the truly interesting patterns representing knowledge Based on some interestingness measures)
- Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)

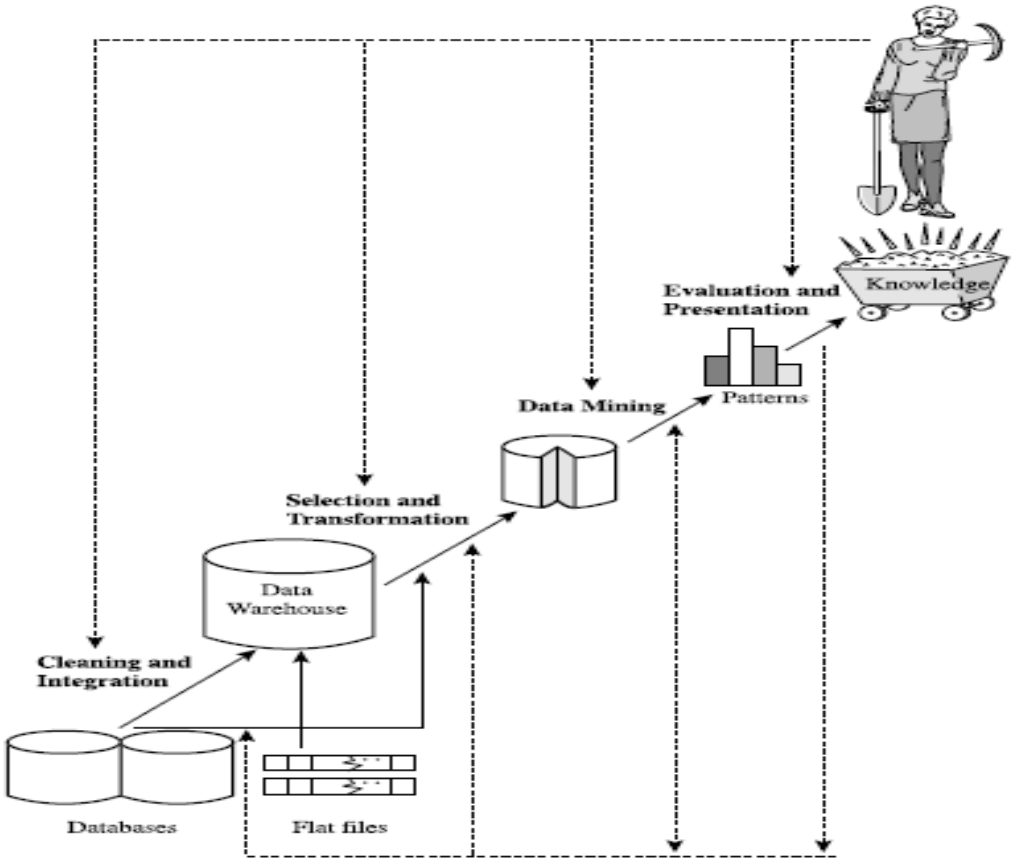

Figure 2.1 Data mining as a step in the process of knowledge discovery [2]

:

2.2.3 Data Classification:

Classification is the problem of identifying to which of a set of categories (sub populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Many real-world problems can be modeled as classification problems such as assigning a given email into "spam" or" non spam" classes, automatically assigning the categories (e.g., "Sports" and "Entertainment") of coming news, and assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.).

There are two strategies to perform data mining, namely supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used. Classification is a supervised learning used to discover hidden patterns from existing medical data. Classification is very critical for therapy of patients. Classification is an important data mining task and the main purpose of classification is to propose a classification function or classification model (called classifier).The classification models can align the data in the database or dataset to a specific class. Classification construction methods include: Naïve Bayes, , Support Vector Machine, Multi-Layer Perceptron , Logistic Regression, Decision Tree, Back Propagation Neural Network Random Forest.

In Figure2.2 demonstrated a general process of data classification usually consists of two phases the training phase and the prediction phase. In the training phase, data is analyzed into a set of features based on the feature (e.g "A", "B", "AB" or "O", for blood type), ordinal (e.g. "large", "medium" or "small"), integer-valued (e.g. the number of occurrences of a part word in an email) or real-valued (e.g. a measurement of blood pressure). Some algorithms work only in terms of discrete data such as ID3 and require that real-valued or integer-valued data be discretized into groups (e.g. less than 5, between 5 and 10, or greater than 10)

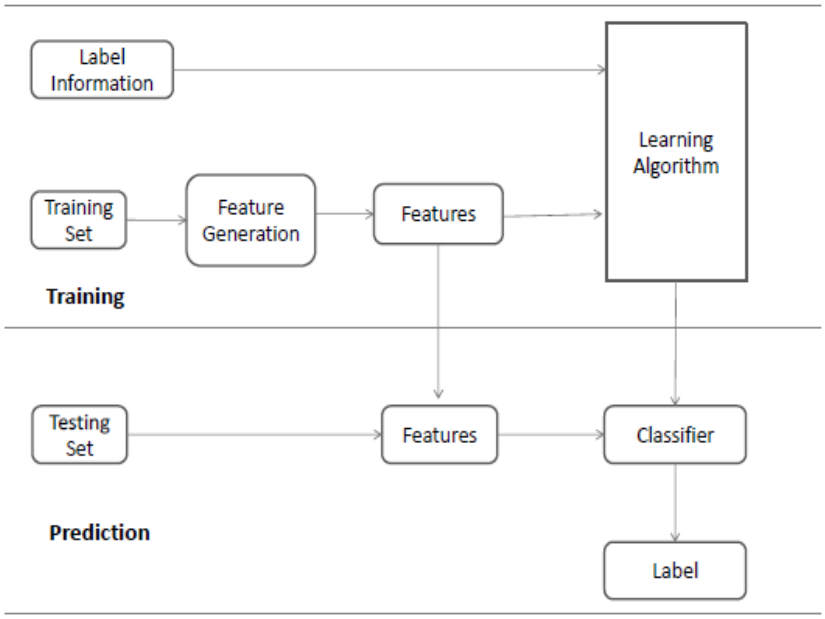

FIGURE 2.2 A General Process of Data Classification [4]

.

After representing data through these extracted features, the learning algorithm will utilize the label information as well as the data itself to learn a map function f (or generation models such as the vector space model for text data [4]. These features may either be categorical (e. a classifier) from features to labels as,

$$f(\text{features}) \rightarrow \text{labels} \quad . (0.1)$$

In the prediction phase, data is represented by the feature set extracted in the training process, and then the map function (or the classifier) learned from the training phase will perform on the feature represented data to predict the labels. Note that the feature set used in the training phase should be the same as that in the prediction phase. There are many classification methods mentioned in the previous paragraph.

2.3 FEATURE SELECTION ALGORITHMS:

A "feature" or "attribute" or "variable" refers to an aspect of the data. Usually before collecting data, features are specified or chosen. Features can be discrete, continuous, or nominal. Generally, features are characterized as:

1- *Relevant*: These are features which have an influence on the output and their role cannot be assumed by the rest.

2- *Irrelevant*: Irrelevant features are defined as those features not having any influence on the output, and whose values are generated at random for each example.

3- *Redundant*: A redundancy exists whenever a feature can take the role of another perhaps the simplest way to model redundancy.

Feature selection is a widely employed technique for reducing dimensionality among practitioners. It aims to choose a small subset of the relevant features from the original ones according to certain relevance evaluation criterion, which usually leads to better learning performance (e.g., higher learning accuracy for classification), lower computational cost, and better model interpretability [5][6].

According to whether the training set is labeled or not, feature selection algorithms can be categorized into supervised [7][8], unsupervised [9] and semi-supervised feature selection [9] [10].

Supervised feature selection methods can further be broadly categorized into filter models, wrapper models and embedded models in this study focus on this models

2.3.1 Filter Models:

The approach of filter model separates feature selection from classifier learning so that the bias of a learning algorithm does not interact with the bias of a feature selection algorithm. It relies on measures of the general characteristics of the training data such as distance, consistency, dependency, information, and correlation. Relief [11] Fisher score and Information Gain based methods [12] are among the most representative algorithms of the filter model.

A typical filter algorithm consists of two steps. In the first step, it ranks features based on certain criteria. Feature evaluation could be either univariate or multivariate. In the univariate scheme, each feature is ranked independently of the feature space, while the multivariate scheme evaluates features in a batch way. Therefore, the multivariate scheme is naturally capable of handling redundant features. In the second step, the features with highest rankings are chosen to induce classification models.

- Univariate algorithm:

In the univariate scheme, each feature is ranked independently of the feature space, univariate algorithm includes Information gain and Gain ratio the advantages of this algorithms fast and Independent of the classifier in this section will discuss the information gain and gain algorithm.

A- Information Gain

Information gain is one of the most popular feature selection methods. It is used to measure the dependence between features and labels and calculates the information gain between the *i-th* feature f_i and the class labels *Cas*

$$IG(f_i, C) = H(f_i) - H(f_i|C), \quad (2)$$

where $H(f_i)$ is the entropy of f_i and $H(f_i|C)$ is the entropy of f_i after observing C:

$$H(f_i) = -\sum_j p(x_j) log_2(p(x_j)),$$
$$H(f_i|C) = -\sum_k p(c_k) \sum_j p(x_j|c_k) log_2(p(x_j|c_k)) \quad (3)$$

In information gain, a feature is relevant if it has a high information gain

B- Gain ratio (GR)

Is a modification of the information gain that reduces its bias. When choosing an attribute, Gain ratio considers the number and size of branches. It takes the intrinsic information of a split into account. Intrinsic information is entropy of distribution of instances into branches (i.e. how much info do we need to tell which branch an

instance belongs to). Attribute value decreases as the increase in the intrinsic information.

$$\text{Gain Ratio(Attribute)} = \frac{\text{Gain(Attribute)}}{\text{Intrinsic} - \text{Info(Attribute)}} \qquad (4)$$

- Multivariate algorithm:

Multivariate scheme evaluates features in a batch way. Therefore, the multivariate scheme is naturally capable of handling redundant features, the advantages of multivariate Models feature dependencies, Independent of the classifier, better computational complexity

Than wrapper methods this method includes Correlation-based feature selection (CFS), and Fast Correlation-based feature selection (FCBF).

A- CORRELATION-BASED FEATURE SELECTION(CFS):

CFS searches feature subsets according to the degree of redundancy among the features. The evaluator aims to find the subsets of features that are individually highly correlated with the class but have low inter-correlation. The subset evaluators use a numeric measure, such as conditional entropy, to guide the search iteratively and add features that have the highest correlation with the class. The downside of univariate filters for e.g. information gain is, it does not account for interactions between features, which is overcome by multivariate

Filters for e.g. CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Correlation coefficients are used to estimate correlation between subset of attributes and class, as well as inter-correlations between the features. Relevance of a group of features grows with the correlation between features and classes, and decreases with growing inter-correlation. CFS is used to determine the best feature subset and is usually combined with search strategies such as forward selection, backward elimination, bi-directional search, best-first search and genetic search. Equation for CFS is given.

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}}$$

Where r_{zc} is the correlation between the summed feature subsets and the class variable, k is the number of subset features, r_{zi} is the average of the correlations between the subset features an the class variable, and r_{ii} is the average inter-correlation between subset features.

B- FAST CORRELATION BASED FS (FCBF):

FCBF (Yu and Liu, ICML 2003) uses also the symmetrical uncertainty measure. But the search algorithm is very different. It is based on the "predominance" idea. The correlation between an attribute X* and the target Y is predominant if and only

$$\text{If } \rho_{y,x^*} \geq \delta \text{et } \forall X(X \neq X^*),\ \rho_{x,x^*} < \rho_{y,x^*}$$

Concretely, a predictor is interesting if its correlation with the target attribute is significant (delta is the parameter which allows to assess this one) there is no other predictor which is more strongly correlated to it.

2.3.2 Embedded Models:

Embedded Models embedding feature selection with classifier construction , have the advantages of (1) wrapper models - they include the interaction with the classification model and (2) filter models - they are far less computationally intensive than wrapper methods [Kudo and sklansky, 2000]

There are three types of embedded methods. The first are pruning methods that first Utilizing all features to train a model and then attempt to eliminate some features by setting the corresponding coefficients to 0, while maintaining model performance such as recursive Feature elimination using support vector machine (SVM) [Goyon and banhill , 2002]. The second are models with a build-in mechanism for feature selection such as ID3 [J.R.Quinlan , 1986] and C4.5 [J.R.Quinlan , 1993]. The third are regularization models with objective functions that minimize fitting errors and in the mean time force the coefficients to be small or to be exact zero. Features with

coefficients that are close to 0 are then eliminated [Ma and huang, 2008]. Due to good performance, regularization models attract increasing attention.

2.3.3 WRAPPER METHODS:

Wrapper methods use the predictor as a black box and the predictor performance as the objective function to evaluate the variable subset. Since evaluating 2N subsets becomes a NP-hard problem, sub optimal subsets are found by employing search algorithms which find a subset heuristically. A number of search algorithms can be used to find a subset of variables which maximizes the objective function which is the classification performance. The Branch and Bound method [Cirish and Ferat, A survey 2014] used tree structure to evaluate different subsets for the given feature selection number. But the search would grow exponentially for higher number of features. Exhaustive search methods can become computationally intensive for larger datasets.

Therefore simplified algorithms such as sequential search or evolutionary algorithms such as Genetic Algorithm (GA) [Cirish and Ferat, A survey 2014] or Particle Swarm Optimization (PSO) [Cirish and Ferat ,A survey 2014] which yield local optimum results are employed which can produce good results and are computationally feasible.

We broadly classify the Wrapper methods into Sequential Selection Algorithms and Heuristic Search Algorithms. The sequential selection algorithms start with an empty set (full set) and add features (remove features) until the maximum objective function is obtained. To speed up the selection, a criteria is chosen which incrementally increases the objective Function until the maximum is reached with the minimum number of features. The heuristic search algorithms evaluate different subsets to optimize the objective function. Different subsets are generated either by searching around in a search space or by generating solutions to the optimization problem. First we will look at sequential selection algorithms followed by the heuristic search algorithms.

1. Sequential selection algorithms

These algorithms are called sequential due to the iterative nature of the algorithms. The Sequential Feature Selection (SFS) algorithm [Pudil and Kittler, 1994] [Reunanen ,2003]starts with an empty set and adds one feature for the first step which gives the highest value for the objective function. From the second step onwards the remaining features are added individually to the current subset and the new subset is evaluated. The individual feature is permanently included in the subset if it gives the maximum classification accuracy. The process is repeated until the required number of features are added. This is a naive SFS algorithm since the dependency between the features is not accounted. A Sequential Backward Selection (SBS) algorithm can also be constructed which is similar to SFS but the algorithm starts from the complete set of variables and removes one feature at a time whose removal gives the lowest decrease in predictor performance.

2. Heuristic search algorithms

Genetic Algorithm (GA) can be used to find the subset of features [15] wherein the chromosome bits represent if the feature is included or not. The global maximum for the objective function can be found which gives the best suboptimal subset. Here again the objective function is the predictor performance. The GA parameters and operators can be modified within the general idea of an evolutionary algorithm to suit the data or the application to obtain the best performance or the best search result. A modified version of the GA called the CHCGA can be used for feature selection [15].

2.4 WRAPPER AND CLASSIFIER USED

2.4.1 Wrapper

Wrapper is one of the features algorithms used to reduce the number of features before applying classifier to enhance the accuracy of model.

In wrapper methods, we try to use a subset of features and train a model using them based on the inferences that we draw from the previous model decide to add or remove features from subset. The problem is essentially reduced to a search problem. These methods are usually computationally very expensive. Figure3.3 below showed the general framework of wrapper method.

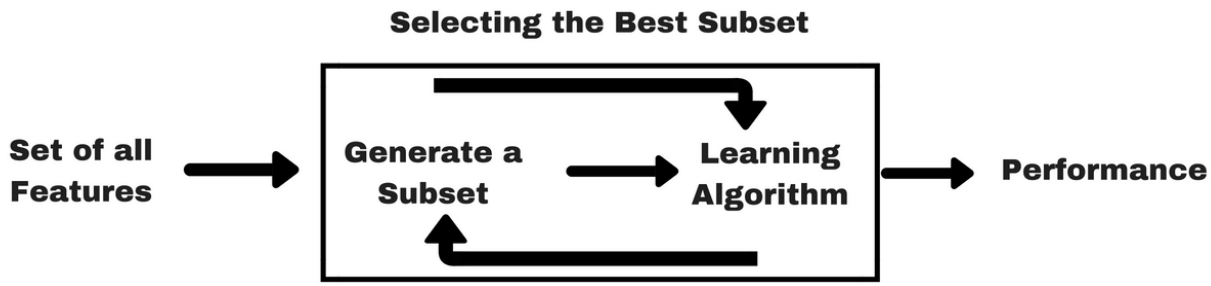

Figure 3.3 General Framework of Wrapper method

Some common examples of wrapper methods are forward feature selection, backward feature elimination, recursive feature elimination, etc.

- Forward Selection: Forward selection is an iterative method in which we start with having no feature in the model. In each iteration, we keep adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model.
- Backward Elimination: In backward elimination, we start with all features and remove the least significant feature at each iteration which improves the performance of the model. We repeat this until no improvement is observed on removal of features.

- Recursive Feature elimination: It is a greedy optimization algorithm which aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination.

Given a predefined classifier, a typical wrapper model will perform the following steps:

Step 1: searching a subset of features

Step 2: evaluating the selected subset of features by the performance of the classifier,

Step 3: repeating Step 1 and Step 2 until the desired quality is reached.

A general framework for wrapper methods of feature selection for classification is shown in Figure 3.4 and it contains three major components:

- Feature selection search - how to search the subset of features from all possible feature subsets.
- Feature evaluation - how to evaluate the performance of the chosen classifier, and
- Induction Algorithm.

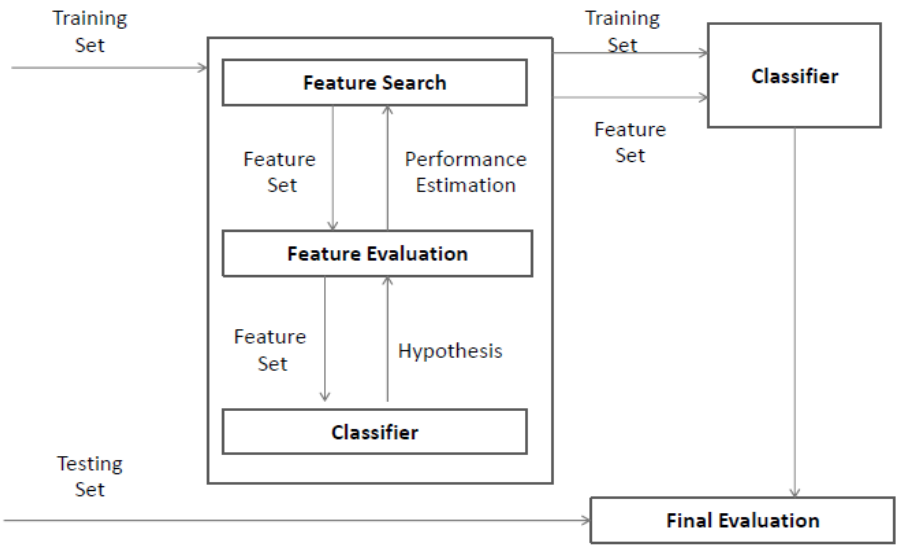

Figure 3.4: Framework for Wrapper Methods steps[4]

In wrapper models, the predefined classifier works as a black box. The feature search Component will produce a set of features and the feature evaluation component will use the classifier to estimate the performance, which will be returned back to the feature search component for the next iteration of feature subset selection. The feature set with the highest estimated value will be chosen as the final set to learn the classifier. The resulting classifier is then evaluated on an independent testing set that is not used in during the training process.

2.4.2 Naïve bayes classifier

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". This restricted individuality assumption infrequently clutches true in real world applications, hence the characterization as Naive yet the algorithm inclines to perform well and learn rapidly in various supervised classification problems . An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. The following paragraph represents and explains the Bayes theorem:

Let H be some hypothesis that the data tuple X belongs to a specified class C, X be a data tuple.

P(H/X) - is the posterior probability of H conditioned on X.

P(H) - is the prior probability of H.

P(X) - is prior probability of X.

$$P(H/X) = \frac{P(H/X)P(H)}{P(X)} \qquad (1)$$

Example: Predicting a class label using naïve Bayesian classification. We wish to predict the class label of a tuple using naïve Bayesian classification, given the same training data in the table2.1 below. The data tuples are described by the attributes *age, income, student, and credit rating*. The class label attribute, buys computer, has two distinct values (*namely, fyes, nog*). Let C1 correspond to the class buys computer = yes and C2 correspond to buys computer = no. The tuple we wish to classify is x = (age = youth, income = medium, student = yes, credit rating = fair)

Table 2.1 Class-labeled training tuples from the AllElectronics customer database.

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

We need to maximize P(XjCi)P(Ci), for i = 1, 2. P(Ci), the prior probability of each Class, can be computed based on the training tuples:

P(*buys computer* = *yes*) = 9/14 = 0.643
P(*buys computer* = *no*) = 5/14 = 0.357
To compute *PX*j*Ci*), for *i* = 1, 2, we compute the following conditional probabilities:
P(*age* = *youth* j *buys computer* = *yes*) = 2/9 = 0.222

P(*age* = *youth* j *buys computer* = *no*) = 3/5 = 0.600

P(*income = medium* j *buys computer = yes*) = 4/9 = 0.444

P(*income = medium* j *buys computer = no*) = 2/5 = 0.400

P(*student = yes* j *buys computer = yes*) = 6/9 = 0.667

P(*student = yes* j *buys computer = no*) = 1/5 = 0.200

P(*credit rating = fair* j *buys computer = yes*) = 6/9 = 0.667

P(*credit rating = fair* j *buys computer = no*) = 2/5 = 0.400

Using the above probabilities, we obtain

P(Xjbuys computer = yes) = P(age = youth / buys computer = yes) ×P(income = medium / buys computer = yes) × P(student = yes / buys computer = yes) ×P(credit rating = fair / buys computer = yes)

= 0.222×0.444×0.667×0.667 = 0.044.

Similarly,

P(Xjbuys computer = no) = 0:600×0:400×0:200×0:400 = 0.019.

To find the class, Ci, that maximizes P(XjCi)P(Ci), we compute

P(Xjbuys computer = yes)P(buys computer = yes) = 0.044×0.643 = 0.028

P(Xjbuys computer = no)P(buys computer = no) = 0.019×0:357 = 0.007

Therefore, the naïve Bayesian classifier predicts buys computer = yes for tuple X.

2.4.3 J48 classifier

J48 classifier is a simple C4.5 decision tree for classification. It is supervised method of classification. It creates a small binary tree. It is univariant decision tree. It is an extension of ID3 algorithm. In this classifier Divide and Conquer approach is used to classify the data. It divides the data into range based on the attribute value for that value that are found in training sample.

Algorithm: Generate decision tree. Generate a decision tree from the training tuples of data partition D.

Input:

- *Data partition, D, which is a set of training tuples and their associated class labels;*
- *attribute list, the set of candidate attributes;*
- *Attribute selection method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split point or splitting subset.*

Output: A decision tree.
Method:

(1) create a node N;
(2) if tuples in D are all of the same class, C then
(3) return N as a leaf node labeled with the class C;
(4) if attribute list is empty then
(5) return N as a leaf node labeled with the majority class in D; // majority voting
(6) apply Attribute _selection_ method(D, attribute_ list) to find the "best" splitting criterion;
(7) label node N with splitting criterion;
(8) if splitting attribute is discrete-valued and multiway splits allowed then // not restricted to binary trees
(9) attribute _list← attribute _list −splitting _attribute; // remove splitting attribute
(10) for each outcome j of splitting criterion // partition the tuples and grow subtrees for each partition
(11) let Dj be the set of data tuples in D satisfying outcome j; // a partition
(12) if Dj is empty then
(13) attach a leaf labeled with the majority class in D to node N;
(14) else attach the node returned by Generate decision tree(Dj, attribute list) to node N;
endfor
(15) return N;

2.5 RELATED WORK:

This section consists of the reviews of various technical and review articles on data mining techniques applied to predict Chronic Kidney Disease. Many researchers have used different data mining techniques for future prediction.

[Dr. S. Vijayarani and Mr. S. Dhayanand , 2015 In this research work classification process is used to classify four types of kidney diseases. Comparison of Support Vector Machine (SVM) and Naïve Bayes classification algorithms is done based on the performance factors classification accuracy and execution time. From the results, it can be concluded that the SVM achieves increased classification

performance, yields results that are accurate, hence it is considered as best classifier when compared with Naïve Bayes classifier algorithm. Perhaps, Naïve Bayes classifier classifies the data with minimum execution time.

[Lambodar Jena 2015] Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease in this study researcher used the same dataset was used and applied different classifier (j48 (99%)and , naïve Bayes(95%), Multilayer perception(99.75%) , SVM (62%), Conjuctive Rule(94%) Decision Table (99%)) Multilayer perception algorithm gives more classification accuracy i.e. 99.75% comparing to all other classifiers. However, it is interesting to note that all algorithms have classification accuracy more than 90% except SVM which performs very poor. Hence it is concluded that Multilayer Perceptron performs well in case of chronic-kidney-disease dataset .the open issue of this study if you can using the feature selection algorithms then that lead to increasing accuracy of the modes and improve performance by decrease the time execution with building the modes.

[Naganna Chetty , Kunwar Singh Vaisla Sithu D Naganna 2015]. The researcher in this study built classification models with different classification algorithms, Wrappersubset attribute evaluator and best first search method to predict and classify the CKD and non CKD patients. They predicted good accuracy with best first search. Sequential Minimal Optimization(97.75%) followed by IBK(95.75%) and then followed by Naïve Bayes(95%) on original dataset and the same classifiers predicted accuracy i.e., IBK (Implements the K-nearest neighbor) (100%) followed by Naïve Bayes(99%) and it is followed by SMO(Self organizing Map) (98.25%) on reduced dataset. These classifiers have differently classified correctly and incorrectly instances differently on original reduced datasets. However IBK classifier performs better than other two.

[Dr. Uma N Dulhare Mohammad Ayesha,2016]. Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier in this study the researcher built the model by using Naïve Bayes Classifier and used four attribute Evaluator to reducing the attributes. The Attribute Evaluator used (WrapperSubsetEval attribute

evaluator with SMO (Self organizing Map) classifier and best first search, WrapperSubsetEval attribute evaluator with IBK (Implements the K-nearest neighbor) classifier and best first search WrapperSubsetEval attribute evaluator with Naïve bayes classifier and best first search and OneR attribute evaluator with Naïve Bayes). The conclusion of this study proposed Naive Bayes with OneR "One Rule" The number of attributes in dataset is also reduced by 80% using OneR algorithm and improved the accuracy by 12.5 % as compared to the existing system.. Our proposed system extract the action rules for the respective chronic renal disease stages so that the necessary treatments can be taken according to the action rules stated to avoid advancing of CKD to the next stage.

[Ramya and Dr. N. Radha ,2016] have developed a system to predict the kidney function failure by applying four classification techniques on test data from patient medical report. They have 1000 records with 15 attributes. They also compared these four techniques like Back propagation, Neural network, Radial Basis Function and Random Forest. Their results show that RBF (Radial Basis Function) has better accuracy for predicting the chronic kidney disease.

2.6 Summary

In this chapter described the concepts of data mining and data classification and also explained the feature selection methods such as filter model and embedded model and last one wrapper model, also described the naïve bayes and j48 classifiers, and also discussed the related work. In the next chapter will describe proposal model.

CHAPTER (III)

PROPOSED MODLE

3.1 Introduction

This chapter divided into three sections, first one describes proposed model and section two describes material and tools used and section three represent dataset used and data pre-processing.

3.2 PROPOSED MODEL

In this section represent the proposed methodology used in this study. The methodology consists of three steps:

Step1: applying classification model by using original dataset.

Step2: Applying feature selection or (data prepossessing) to reduce the data.

Step3: applying classification model by using reduced datasets.

In Figure 3.1 below showed framework for the proposed methodology for mining pattern.

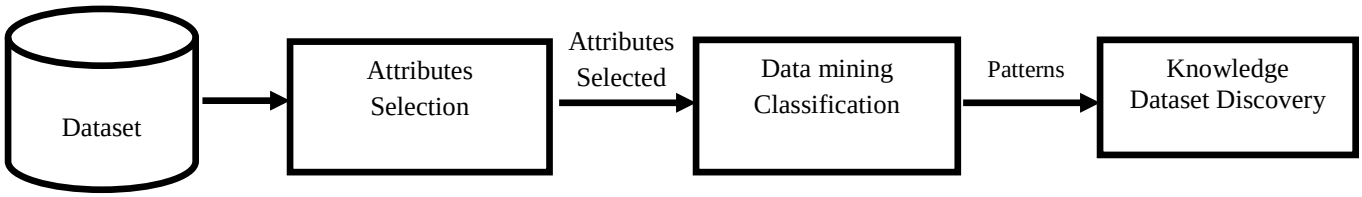

Figure 3.1 Proposed Frameworks for Mining Patterns [3]

This framework consist of dataset and the method that used to select best attributes after that using data mining classification to extract pattern that lead to Knowledge Discovery.

Figure3.2 below shows a methodology for attributes selection. Attribute selection is a process of reducing the dimension of a dataset by eliminating the

attributes of less importance. In this figure we have used the wrapper subset evaluator with best first search method to select features for Naïve Bayes and j48 classifier.

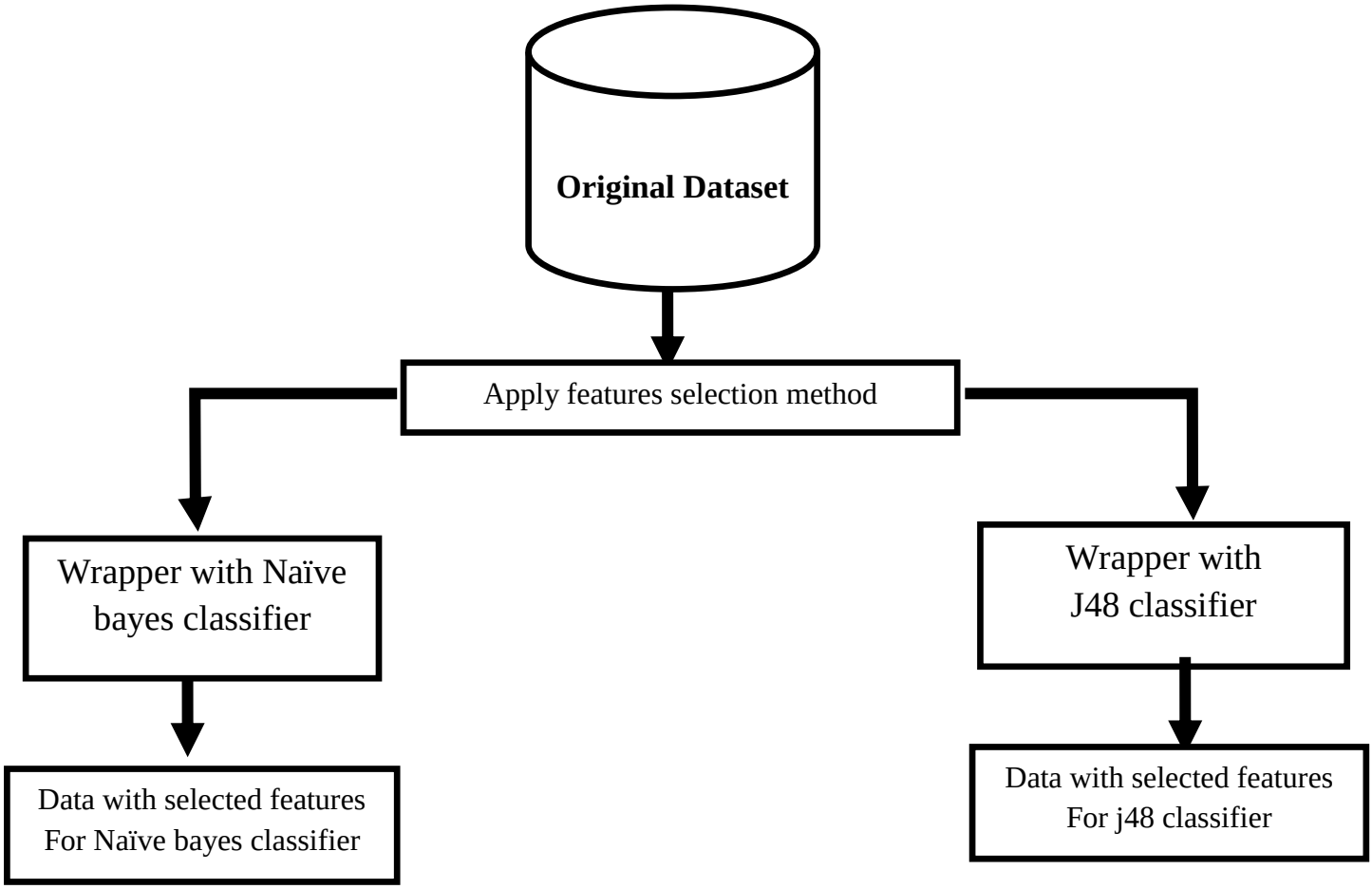

Figure 3.2 Frameworks for Attributes Selection

3.3 Material and Tools

In this study used the Weka (Waikato Environment for Knowledge Analysis) program to experiments and tests and in the following paragraph a brief description of this program.

WEKA is a tool for data preparation and research developed at the University of Waikato in New Zealand. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

3.4 Dataset Used:

Dataset used is taken from UCI machine learning repository. This Dataset special for the Chronic-Kidney-Disease (CKD) it consists of 25 attributes out of which 1 is class attribute, 13 nominal attributes and 11 numerical attributes and 400 instances (250 CKD, 150 notckd) and the table 3.1 below gives more information about this dataset.

TABLE 3.2 DESCRIPTION OF ATTRIBUTES IN THE CHRONIC-KIDNEY-DISEASE DATASET

Sl. No	Attribute	Description	Type	Permissible values
1	Age	Age	numerical	age in years
2	Bp	blood pressure	numerical	in mm/Hg
3	Sg	specific gravity	nominal	(1.005,1.010,1.015,1.020,1.025)
4	Al	albumin	nominal	(0,1,2,3,4,5)
5	Su	sugar	nominal	(0,1,2,3,4,5)
6	Rbc	red blood cells	nominal	normal ,abnormal
7	Pc	pus cell	nominal	Normal ,abnormal
8	Pcc	pus cell clumps	nominal	present, not present
9	Ba	bacteria	nominal	Present , not present
10	Bgr	blood glucose random	numerical	in mgs/dl
11	Bu	blood urea	numerical	in mgs/dl
12	Sc	serum creatinine	numerical	in mgs/dl
13	Sod	sodium	numerical	in mEq/L
14	Pot	potassium	numerical	in mEq/L
15	Hemo	hemoglobin	numerical	in gms
16	Pcv	packed cell volume	numerical	in cells/cumm
17	Wbbc	white blood cell count	numerical	in cells/cumm
18	Rbcc	red blood cell count	numerical	millions/cmm
19	Htn	hypertension	nominal	Yes ,no
20	Dm	diabetes mellitus	nominal	Yes ,no
21	Cad	coronary artery disease	nominal	Yes ,no
22	Appet	appetite	nominal	Good ,poor
23	Pe	pedal edema	nominal	Yes ,no
24	Ane	anaemia	nominal	Yes ,no
25	Class	class	nominal	Ckd , notckd

Table 3.2 Class Distribution: (2 classes)

Class	Number of instances	Total
Ckd	250	400 instances
Notckd	150	

3.5 Data pre-processing

Before running any classification algorithms on the data, the data must first be cleaned and transformed in what is called a pre-processing stage. During this pre-processing stage, several processes take place, including evaluating missing values, eliminating noisy data such as outliers, normalizing, and balancing unbalanced data

Real world data generally contains missing values. One way of dealing with missing values is to omit the entire record which contains the missing value, a method called Case Deletion. However, it is identified that if a data set with 30 variables misses 5% of the values (spread randomly throughout attributes and records), one would have to omit approximately 80% of the records from the data set. Instead of removing the records with missing values, different data imputation algorithms have been studied and compared. Missing numerical attributes are Median or Mid-Range imputed, as its name implies, replaces the missing values in the record with the median value of that attribute taken across the data set. However for nominal attributes Mode Imputation is done, replaced the missing values in the record with the mode value of that attribute taken across the dataset.

The data preprocessing applied on this dataset includes:

- Merge any nominal attribute that multiple values such as(sugar (Su((0,1,2,3,4,5)))In the small range (2 or 3 outcomes).
- Discretize any numerical attribute that multiple distinct values to three categorize.
- Ignore the missing value for each attribute.
- Applied the features selection algorithms to select best features.

3.6 Summary

In this chapter described the proposed model and methodology framework also represented material and tools used (Weka program) and discussed the Chronic-Kidney-Disease (CKD) dataset and last described the data pr-processing. In next chapter will dissection the experiments and results.

CHAPTER (IV)

EXPERIMENTS AND RESULTS

4.1 Introductions

This chapter explains the experiments that we have undertaken for the purpose of study and which focus on the effect of the selection of features in improving the accuracy of the model. in the following paragraphs explain the stages of building the model and discuss the results obtained from these experiments.

4.2 Model build:

In this work we used weka program to apply wrapper features selection algorithm and J48 and naïve bayes classifier to build the model. And comparing the result for this classifier based on classification accuracy.

Classification Accuracy is measure used to comparing the results that taken from the classifiers.In this section we describe the concept of model accuracy and how precision is calculated

- Accuracy is defined in the terms of correctly classified instances divided by the total number of instances present in the dataset.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

Where TP – True positive, FP False positive, TN true Negative FN False Negative.

- TP Rate: It is the ability which is used to find the high true-positive rate. The true-positive rate is also called as sensitivity.

$$\text{TPR} = \frac{TP}{TP+FN}$$

- Precision is given the correlation of number of modules correctly classified to the number of entire modules classified fault-prone. It is quantity of units correctly predicted as faulty.

$$\text{Precision}= \frac{TP}{TP+FP}$$

4.3 The first experiment

In this experiment we have apply naïve bayes (NBC) and (J48) classifiers methods based on all dataset features (using original dataset). (CKD) dataset consists of 13nominal attributes, 11 numerical attributes and 1 class and 400 instances (250 CKD, 150 notckd) for chronic kidney disease.Figure 4.1 shows the distribution of the patient based on the class label (CKD or notckd).

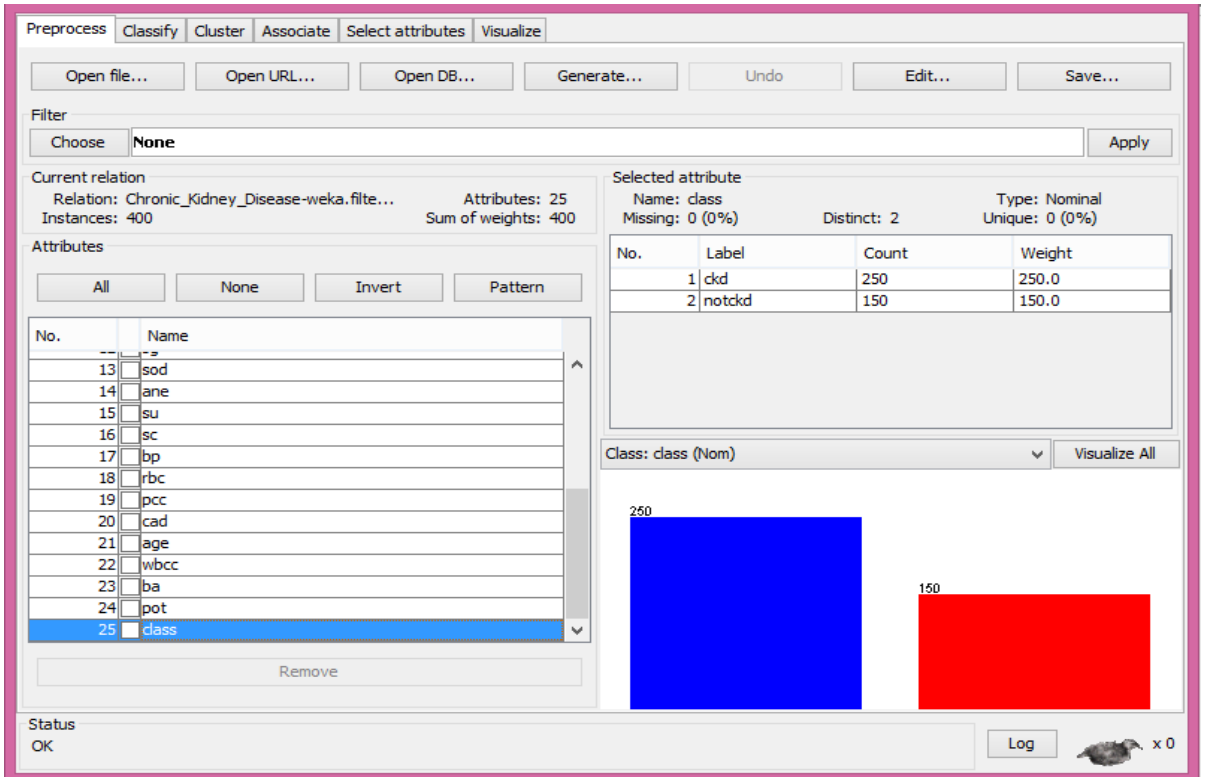

Figure 4.1 the distribution of the patient based on the class label (CKD or notckd).

Table 4.1 shows the result of first experiment before reducing dataset by using J48 and naïve bayes used cross validation test applied. The j48 classifier taken 99% and only (4 instances) incorrectly classified and naive bayes taken 95% and only (20 instances) incorrectly classified.

Classifier	Total Of Instances	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy
J48	400	396	4	99%
Naïve bayes	400	380	20	95%

The figure 4.2 shows Graphical representation of classification accuracies for J48 and naïve bayes (NBC) clarifiers.

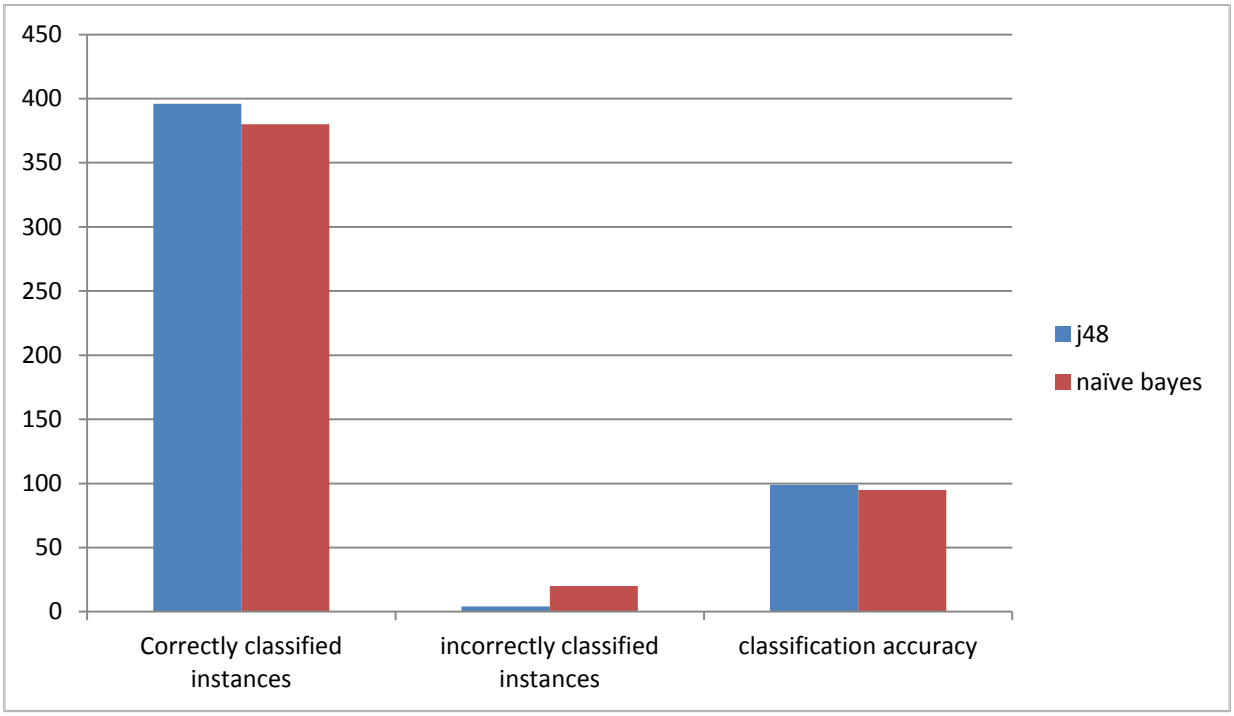

Figure4.2 Result of Classification based on Original Dataset

4.4 Reducing dataset

In this section we have used Wrapper attribute evaluator and best first search method with Naïve bayes (NBC) and J48 classifier to reduction dataset. Table 4.2 below shows the result of attributes reduction by the attribute evaluator. The *Wrapper Subset Evaluator with* Naïve bayes (NBC) selects only (5) attributes (*hemo , al , sc, su , and wbcc*) from (25) total of attributes with 80% attributes reduction. and The *Wrapper Subset Evaluator* withJ48 selects only (11) (*hemo , rbcc ,htn , dm , bgr, appet , pe ,bu , sg ,sod and sc*) attributes from (25) total of attributes with 54% attributes reduction. The table 4.2 shows result of attributes reduction by using Wrapper attribute evaluator and best first search method with Naïve bayes (NBC) and J48 classifiers.

TABLE 4.2 Representation the result of attributes reeducation.

Attribute Evaluator using Best first search method	Initial Attributes	Selected Attributes	Attributes Reduction(%)
Wrapper Subset evaluator with j48 classifier	25	11	54%
Wrapper Subset evaluator with naïve bayes(NBC) classifier	25	5	80%

Figurer 4.3 below shows Graphical representation of data reduction after applied the wrapper features selection evaluator based on the best first search method by using naïve bayes (NBC) and J48 classifiers.

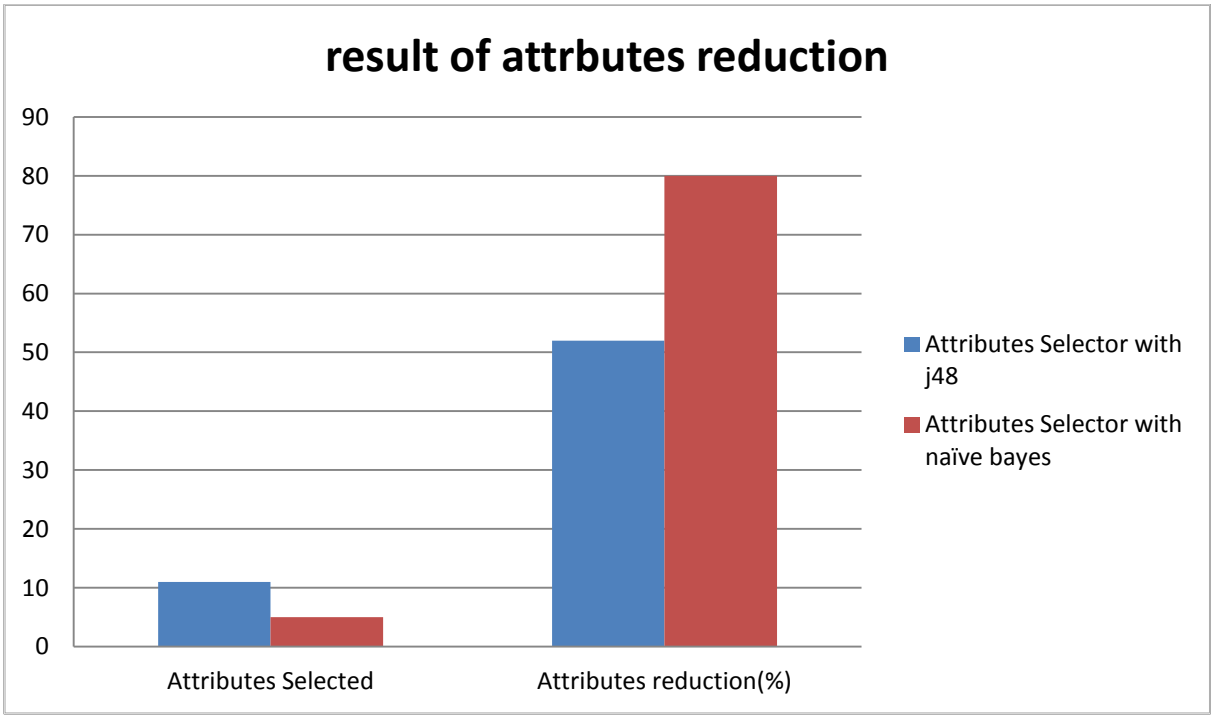

Figure 4.3result of attributes reduction.

4.5 The second experiment

In this experiment we have used (J48) and naïve bayes (NBC) classifiers after reduced dataset by using the wrapper attributes selection evaluator (using reduced dataset) discussed in the previous section and applied preprocessing below:

- Merge any nominal attribute that multiple values such as (sugar (Su ((0,1,2,3,4,5))) in the small range (2 or 3 outcomes).
- Discretize any numerical attribute that multiple distinct value to three categorizes.

Table 4.3 below shows the result of second experiment by using j48 and naïve bayes (NBC) used cross validation test after dataset reduced.

Table 4.3 Result of Classification based on Reduced Dataset

Classifier	Total Of Instances	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy
J48	400	396	4	99%
Naïve bayes	400	398	2	99.5%

From the result shows in Table 4.3 observe j48 classifier classified 396 instances correctly and 4 instances only incorrectly from (400) instances of kidney disease with classification accuracy 99%

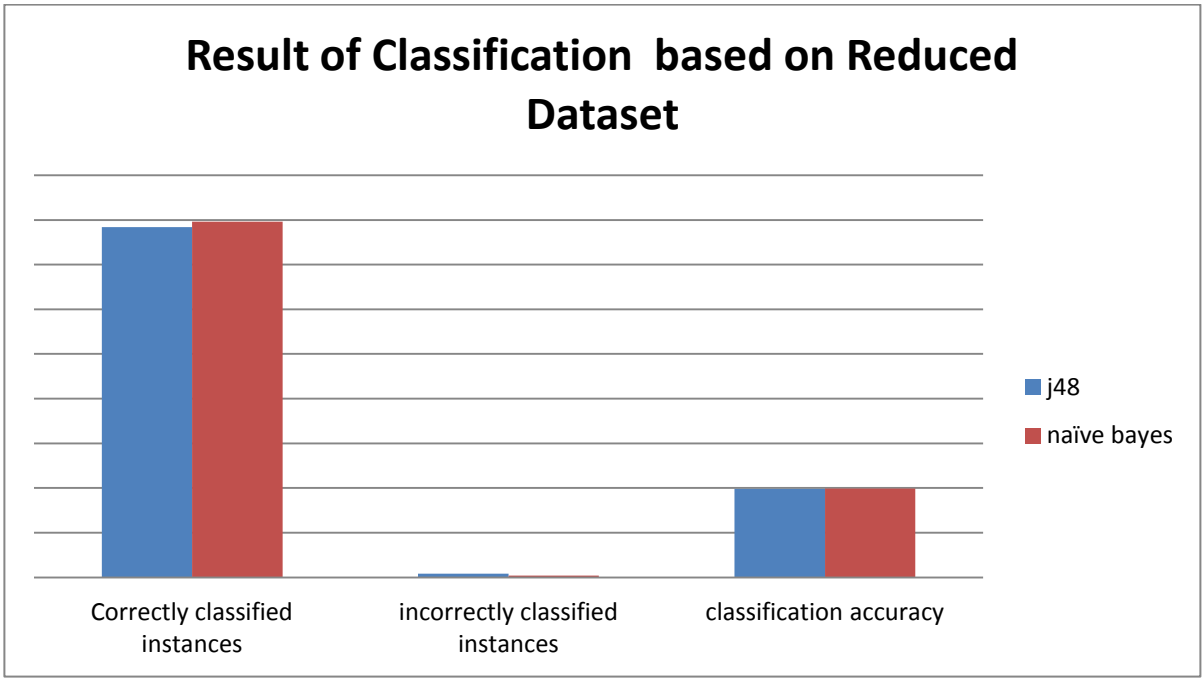

Figure 4.4 Result of Classification on Reduced Dataset

and naïve bayes classifier classified 398 instances correctly and 2 instances incorrectly from (400) instances of kidney disease with classification accuracy 99.5%.and the accuracy of the naïve bayes classifier increased in the rate of 5.5% but result of J48 classifier do not effect. That means the classifier naïve bayes better than J48 clarifier when used the wrapper method evaluator. Graphical representation of classification accuracies showed in Figure 4.4, and also observes the naïve bayes and j48

classifiers have agreed to choose some features because these features have effective in Classification of kidney patients. The (hemo hemoglobin and sc hemoglobin serum creatinine).This confirms that the two very important tests for the diagnosis of kidney disease are creatinine and hemoglobin and also observe the J48 classifier doesn't effect.

4.6 Summary

This chapter divided into three phases, the first one applied model by using original dataset, and the phase tow applied wrapper method to reduce dataset, and in phase three applied model by used the reduced dataset. And also showed short describe for each phases. The next chapter consist conclusion and recommendations.

5.1 CONCLUSION

The model applied to classification Chronic-Kidney-Disease by used the naïve bayes classifier and J48 classifier and used wrapper attribute evaluator to reduce dataset. And study the effect of this evaluator in accuracy model. dataset used is taken from UCI Dataset for CKD. It consists of 25 attributes out of which 1 is class attribute, 13 nominal attributes and 11 numerical attributes and 400 instances (250 CKD, 150 notckd).

The *Wrapper Sub set Eval* with Naïve bayes selects only (5) attributes (*hemo , al , sc, su , and wbcc*) from (25) total of attributes with 80% attributes reduction with classification accuracy 99.5%.and the accuracy of the naïve bayes classifier increased in the rate of 5.5%. and *WrapperSusetEval* with j48 selects only (11) (*hemo , rbcc ,htn , dm , bgr, appet , pe ,bu , sg ,sod and sc*) attributes from (25) total of attributes with 54% attributes reduction with classification accuracy 99% that mean the classifier naïve bayes better than J48 clarifier when used the wrapper method evaluator , and the classification accuracy after the reduced the data set better than classification accuracy before reduced data set or by used original dataset. The precision of J48 classifier was not affected when using wrapper algorithm because the J48 algorithm using internal evaluator to select best features. My contributions the features of (hem (hemoglobin) and Sc (serum creatinine)) the very important features to classification (CKD) chronic kidney diseases because these features selected by tow classifiers, and when used naïve bayes from the best used with wrapper method because this gives high accuracy and performance.

5.2 Recommendations

1- In this focused just on j48 and naïve bayes classifiers also possible used another classifiers.
2- They are many methods to features selection but in this study focused on wrapper method and possible used hybrid features selection methods.
3- To enhance naïve bayes algorithm in the future by built wrapper in naïve bayes algorithm if that possible.

5.3 References

[1] Lambodar Jena , "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", International Journal of Emerging Research in Management &Technology November 2015.

[2] Data mining concept and technique second Edition for "Jiawei Han And Micheline".

[3] Mohammad Ayesha ," Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier " , IEEE International Conference on Computational Intelligence and Computing Research2016.

[4] Jiliang Tang, Salem Alelyani and Huan Liu, "Feature Selection for Classification: A Review"

[5]Yvan Saeys, Inaki Inza and Pedro Larranaga, "A review of feature selection techniques in bioinformatics", Bioinformatics, Volume 23, Issue 19, August 2007, pp 2507-2517.

[6] Zheng, Lijuan, Hongwei Wang, and Song Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews", International Journal of Machine Learning and Cybernetics, 2015.

[7] L. Song, A. Smola, A. Gretton, K. Borgwardt, and J. Bedo. Supervised feature selection via dependence estimation. In International Conference on Machine Learning, 2007.

[8] J. Weston, A. Elisseff, B. Schoelkopf, and M. Tipping. Use of the zero norm with linear models and kernel methods. Journal of Machine Learning Research, 3:1439–1461, 2003.

[9] J.G. Dy and C.E. Brodley. Feature selection for unsupervised learning. The Journal of Machine Learning Research, 5:845–889, 2004.

[10] Z. Xu, R. Jin, J. Ye, M. Lyu, and I. King. Discriminative semi-supervised feature selection via manifold regularization. In IJCAI' 09: Proceedings of the 21th International Joint Conference on Artificial Intelligence, 2009.

[11]H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 1226–1238, 2005.

[12] M. R. Sikonja and I. Kononenko. Theoretical and empirical analysis of Relief and ReliefF. Machine Learning, 53:23–69, 2003.

[13] Girish Chandrashekar and, Ferat Sahin ," A survey on feature selection methods" , Computers and Electrical Engineering 40 (2014) 16–28.

[14] Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. Pattern Recog Lett 1994;15:1119–25.

[15] Reunanen J. Overfitting in making comparisons between variable selection methods. J Mach Learn Res 2003;3:1371–82.

[16] Kudo M, Sklansky J. Comparison of algorithms that select features for pattern classifiers. Pattern Recog 2000;33:327–36.

[17] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. Machine learning, 46(1-3):389–422, 2002.

[18] J. R.Quinlan. Induction of decision trees. Machine learning, 1(1):81–106, 1986.

[19] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

[20]S. Ma and J. Huang. Penalized feature selection and classification in bioinformatics. Briefings in bioinformatics, 9(5):392–403, 2008.

[21] Dr. S. Vijayarani, DATA MINING CLASSIFICATION ALGORITHMS FOR KIDNEY DISEASE PREDICTION, International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August 2015.

[22] Naganna Chetty, Kunwar Singh Vaisla and Sithu D Sudarsan, "Role of attributes selection in classification of Chronic Kidney Disease patients", International Conference on Computing, Communication and Security (ICCCS), 4-5 Dec, 2015, pp 1-6..

[24]Mohammad Ayesha ," Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier " , IEEE International Conference on Computational Intelligence and Computing Research2016.

[25] S.Ramya, and Dr. N.Radha , "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2016.