Data Analytics on Chronic Kidney Disease Data Set Using Association Rule Mining

J. Sarada, Research Scholar, Dept. of Computer Science, SPMVV,Tirupati[1]

Dr. N. V. Muthu Lakshmi, Assistant Professor, Dept. of Computer Science, SPMVV,Tirupati[2]

Abstract: Data analytics is the science of drawing insights from unrefined data/information from various sources. In order to increase the overall efficiency of a business or system, data analytic techniques help to make trends and metrics which will be useful to optimize the processes. Scientists and researchers to verify or disprove scientific models, theories and hypotheses, Data analytics technologies and techniques are widely used especially in commercial & healthcare industries to enable organizations to make knowledge based decisions efficiently. Among many industries, health care is getting prominence as they serve people directly with health and policy issues. In this paper, Chronic Kidney Disease (CKD) data is considered to predict the factors which contribute to the diseases related to kidney commonly called as chronic kidney disease and this prediction can be done with data mining techniques. CKD is one of the fast growing diseases which can cause kidney failure or can lead to death. Research is going on this disease to predict the causes of CKD and symptoms of CKD. Association rule mining technique is applied on chronic kidney disease using well known algorithm apriori using WEKA tool & R to predict the rules. Based on the interest of experts, interested association rules are identified which will be helpful to medical practitioners, hospital management, experts, health analysts, pharmacists, pharmaceutical companies, policy makers, insurance companies and others. The interested rules are useful to predict CKD and also used to take precautions to prevent CKD.CKD can be prevented by reducing or controlling levels of some factors like diabetes, hypertension, sodium, potassium levels etc. Also, this paper computes execution time and compares algorithm with tools. The main objective of this research is to reduce the CKD cases by proposing different conditions which can lead to CKD and also gives different conditions to prevent from CKD.

Keywords: Data Analytics, Association Rule Mining, Health Care Industry, Chronic Kidney Disease

I. INTRODUCTION

Data analytics is the science of drawing insights from unrefined data/information from various sources and whose results are used to increase the overall efficiency of a business or system. To verify or disprove scientific models, theories and hypotheses, scientists and researchers use data analytic techniques which helps them to analyze trends and satisfaction. They also use for optimize the processes or services and also to go for new products and services. Data analytics technologies and techniques are widely available to use especially in service based industries it enables organizations to make right decisions based on the determined knowledge. In general, data analytics or data examination, otherwise called investigation of data is a procedure of reviewing, purifying, changing, and displaying Data with the objective of finding helpful data, recommending conclusions, and supporting basic leadership. Data examination includes making inquiries about what happened, what is going on, and what will happen. Investigation is for the most part the Data crunching, question-noting stage paving the way to the basic leadership stage in the general Intelligence process.

Data analytics includes numerous stages including building up Data at index, setting up the Data for preparing, applying models, recognizing key discoveries and making reports. The objective of Data Analytics is to discover significant bits of knowledge or fining hidden information to improve the services or processes to satisfy the user needs.

Data analytics are widely applied in business, politics, educational institutions, insurance companies, military, healthcare industries and others successfully where intelligence based decisions are crucial to take from large volume of data. These decisions will help the experts to optimize the performance affectively. Here accurate data for analysis is more importance so collecting data from various source in different formats should be efficient even it is a tedious task but data must be examined before apply. Subjective and quantitative are two data examination techniques and each strategy has their own methods and objectives. However, perceptions are types of subjective research, while analyzes and overviews are quantitative.

Next section gives brief outline of various types of data analytics that are being used frequently in real world.

1.1 Different Types of Data Analytics

Data Analytics is used in many industries to improve organizational decision-making and it is in the sciences to verify and invalidate existing models. The most common data analytics are text analytics, picture analytics, video analytics, predictive analytics, Content analytics, pattern analytics and are shown in the below diagram and each is presented here.

- ➢ **Text Analytics**: This analytics can apply in statistical, linguistic and structural techniques to extract and classify information from textual sources, or a species of unstructured data. Many of these foundational text processing and indexing techniques have been deployed in text-based enterprise search.
- ➢ **Picture Analytics** is the preparation of picture to dissect the data. A picture is equal to a huge number of words worth. So the picture investigation assumes a critical part in examining or to foresee fundamental data. Picture analytics including visual mapping, model based analysis and user interactions.
- ➢ **Video Analytics** is one the explanatory strategy utilized for information examination. The information is in various structures like content, pictures and recordings. Prescient examination is the procedure to foresee future esteems from the current esteems or present

esteem. Present examination is utilized as a part of investigation to foresee patterns to conduct designs and to extricate data from information.

- ➢ **Predictive Analytics** uses many techniques in data mining, machine learning, modeling and statistics to analyze a current data to make predictions about future. It is reusable and is created by training an algorithm using historical data and saving a model for reuse purpose. The Predictive Analytics software solutions have built in algorithms such as time series, decision tree, outlier, regressions and neural networks. Basically this model analyzes the past performance for future predictions.
- ➢ **Content Analytics** is the way toward breaking down the substance of information investigation like tweet message on twitter, remark content or post on Face book and video content in you tube.
- ➢ **Pattern Analytics** is the investigation procedure directed by web based business sites to discover the client enthusiasm on the item and to know current premiums of clients.

1.2 Applications of Data Analytics

Data Analytics purpose is to help the people in making decisions in order to improve the performance. The range of technologies and techniques for analyzing vast volumes of data is expanding at a rapid pace. Statistical methods, machine learning methods and data mining methods are used for data analysis. In business, to gain deep insights into customer behaviour, system performance, and new revenue opportunities, data analytics strategies will benefit greatly from being on top of the latest data analytics trends. Some of data analytics applications which are used successfully [1] are presented here.

- In banking sector, data analytics applications are used to analyses business forecasting and competitive analysis.
- Data analytics in Social media, data is analyzed to predict interesting values and to find relationships between data. The Social media data analysis is performed for sentiment analysis, content analysis, context analysis, influence analysis and Network analysis.
- In Education domain, data analytics applications are used to exchange new analytical methods and it helps to address new problems raised in research.
- Data analytics applications in entertainment and communication domain, gathers views, comments, opinions or feedbacks about the announcements or show which helps to provide more services in entertainment.
- In Consumer Product Domain, data analytics applications are used to encourage new innovations and improve product design or service.
- In Political Field, data analytics utilizes the collected information to identify the general population talk and to connect with various individuals.
- Data analytics plays an important role in Healthcare which helps in decision makin**g.** Healthcare comprises hospitals, medical devices, human resources, out sourcing, telemedicine, health insurance, drugs and medical equipments. The Indian healthcare sector is growing at a brisk pace due to its strengthening coverage, services and increasing expenditure by public as well private patients. A healthcare service provider, which includes hospitals, nursing homes and diagnostics centers, and pharmaceuticals, constitutes 65 per cent of the overall market for an income generation.

2. DATA ANALYTICS ON HEALTH CARE INDUSTRY

One of the most promising areas is healthcare industry where data analytics can be applied and data analytics on health care can be termed as healthcare data analytics or healthcare analytics. This analytics can provide more services to the people efficiently and also immediately attention is provided to avoid bad situations to the patients. It can have the potential to avoid preventable diseases, reduce costs of treatment, predict outbreaks of epidemics, and improve the quality of life in general. The term disease broadly refers to any condition that impairs the normal functioning of the body. Now days, human lifespan is increasing along world population by getting well treatments, well technologies, advance prevention methods, acquiring knowledge related to health issues from the world wide web. But, even then more number of diseases is getting by the people because of communicable diseases, congenital disease, genetic disease, with various causes of pollution, pesticide foods, water contamination, existing & new viruses, bacteria, and unexpected illness. Healthcare industry in India [2] has become one of India's largest sectors both in terms of revenue and employment. Therefore today's world, Health professionals has to face new challenges in treatment and facing ambiguity diseases with various existing and new methods to satisfy the patient needs and services. But it is possible to predict patterns for Health professionals easily by having suitable and large data from past as well as present easily as technology has developed to address various issues like capability of storing large variety of data with rapid speed, availability of suitable software's. Hence, health care data analyst takes large data from different sources and then applies pre-processing techniques to get useful data for analysis. Data mining techniques can be applied to get interested patterns from these refined data easily and quickly as because tools are easily available to use.

2.1 Data Mining Techniques on Health Care Industry

Data mining is the process of discovering hidden patterns from large data and this is very important step in data analysis. Data mining tasks can be classified into two which are called predictive and descriptive. Predictive model determines patterns on the existing and/or historical data to make predictions. Descriptive model discovers patterns or relationships in data and it serves as a way to explore the properties of the examined data. Classification, regression analysis, time series analysis, prediction and so on are predictive models of data mining whereas association Rules, clustering, sequence discovery are descriptive models. In this paper, descriptive model such as association rule mining technique is considered for generating useful and strong associations between attributes in health care data which is chronic kidney disease dataset to predict the attributes which may contribute more to get kidney diseases which may be chronic.

One of the challenging issue in health care is how medical data as data is to be collected from many sources governed by different states, hospitals, and administrative departments. Integration of these data sources would require developing a new infrastructure where all data providers collaborate with each other.

The three important data mining techniques which are successfully used in various healthcare domains are discussed here.

The authors in [3] have done work on extracting patterns & detecting trends using Neural Networks for Conventional Pathology Data. Using integrated decision tree model with neural network classification methods was applied out in [4] to categorize skin disease efficiently. Integration of Clustering and Classification techniques were applied for biomedical databases [5] for diagnosing Functional Magnetic Resonance Images. In paper [6], prediction and diagnose of heart disease by using different data mining algorithms such as J48, REPTREE, Naïve Bayes, Bayes Net, Simple CART. The author analyzes the performance of these algorithms through evaluation criteria such as Kappa Statistics, Mean Absolute Error, Root Mean Squared, Relative Absolute Error and Root Relative Squared Error. Accuracy of J48, REPTREE, Naïve Bayes, Bayes Net and CART are 99.0741%, 99.0741%, 97.222%, 98.1481% and 99.0741% respectively. In paper [7], data mining applications are applied on Healthcare Sector. The authors mainly focused on different data mining techniques and various tools and its effects on medical field and predicting diseases. This paper compares and combines the different data mining algorithm also develop a diseases predication algorithm to predict the cancer. It reduces the cost increases the accuracy of separating desirable information from the huge data and helps the doctor in making decision on respective patient health condition.

Abhishek et.al [8] have used two neural network techniques, Back Propagation Algorithm (BPA), Radial Basis Function (RBF) and one non-linear classifier Support Vector Machine (SVM) and compared in accordance with their efficiency and accuracy. They used WEKA 3.6.5 tool for implementation to find the best technique among the above three algorithms for Kidney Stone Diagnosis. The main purpose of their thesis work was to propose the best tool for medical diagnosis, like kidney stone identification, to reduce the diagnosis time and improve the efficiency and accuracy. From the experimental results they concluded, the back propagation (BPA) significantly improved the conventional classification technique for use in medical field. In paper [9] Tommaso Di Noia et al., developed a software tool for predicting the End Stage of Kidney Disease(ESKD) Artificial Neural Network classifier is used to check the probability for the end stage. This research explains the ten networks in a time of thirty eight years at University of Bari. This tool can be used as an android mobile app as well as an online web application and it is essential for the usefulness of clinical approvals.

Data mining plays an important role in diagnosing diseases in advance which help the patients to recover soon and also help the experts or doctors to treat the patients well to recover fast with lower cost with fewer side effects. Further, research work also being carried out in various disease of human body which are Lymphoma Disease, Lung Cancer, Liver diseases, Skin Disease, Chest Disease, Diabetes, Breast Cancer, and Parkinson Disease. Chronci Kidney Desease (CKD) is a major public health problem to study and if not diagnosed or not treated well, patient may die or may face cardiac problems or brain functioning not done or other problems which may deterioration the life span of a kidney patient. If detected at advanced stage, more time, expensive treatment sometimes go for kidney transplantation to maintain health. So this makes to study this major problem faced by many people even children, adults with different problems related to kidney which may be facing this disease that to chronic.

In this paper association rule mining is considered and is applied over chronic kidney disease to predict the important attribute/attributes and their values with combination which leads to CKD more frequently.

3. ASSOCIATION RULE MINING ON CKD DATA SET

Association rules are another type of learning portrayal. It has two steps to get strong rules one is generating frequent item sets and another step is to generate association rules. Generation of all the possible combinations of rules that can be derived from a generated frequent item set based on user specified minimum confidence is the first step and then in second stage prune the rules which are not interested to the user and in this way the ultimate rule set contains only the best/strong of them. Apriori algorithm [10] designed apriori algorithm which is most popular in finding frequent item sets. WEKA tool is used to generate association rules over the CKD data set.
.

Different data mining methodologies and machine learning algorithms are connected for forecast of incessant illnesses. Here we are worried about Chronic kidney sickness (CKD), otherwise called perpetual renal ailment, is a strange capacity of kidney or a dynamic disappointment of renal capacity over a time of months or years. Frequently, interminable kidney malady is analyzed because of screening of individuals known to be in danger of kidney issues, for example, those with hypertension or diabetes and those with a blood relative with CKD. It is separated from intense kidney infection in that the lessening in kidney work must be available for more than 3 months. This work dominatingly centered on, forecast of ceaseless kidney infection. Interminable Kidney illness is anticipated utilizing grouping strategies of information mining [11]. Endless kidney malady expectation is a standout amongst the most focal issues in restorative basic leadership since it is one of the main sources of death. Along these lines, robotized device for early expectation of this infection will be helpful to cure.

Once kidney has problem that to chronic where medicines are not sufficient to cure then medical practitioners suggest kidney transplantation or the two forms of dialysis are which are peritoneal and hemodialysis. Peritoneal dialysis uses the body's own peritoneal membrane beneath the outer layers of the abdominal wall to filter the blood. This dialysis can be further classified into two such as available continuous cycling peritoneal dialysis (CCPD) and continuous ambulatory peritoneal dialysis (CAPD). Blood in the entire body is cleansed outside the body through a machine and this treatments take several hours at a time and usually need to be done three or more times a week and this process is called hemodialysis,. But hemodialysis can be done in a dialysis center or it can be done at home and this method is used more frequently. If patient is undergoing dialysis properly he may have chances of getting cardiac arrest.

The following section presents how significant attributes are identified with association rule mining technique using apriori algorithm.

3.1 Proposed Methodology

The Methodology for Finding Significant attributes from using association rule mining technique which leads to CKD are shown in the figure 1 and are explained with seven steps, given below.
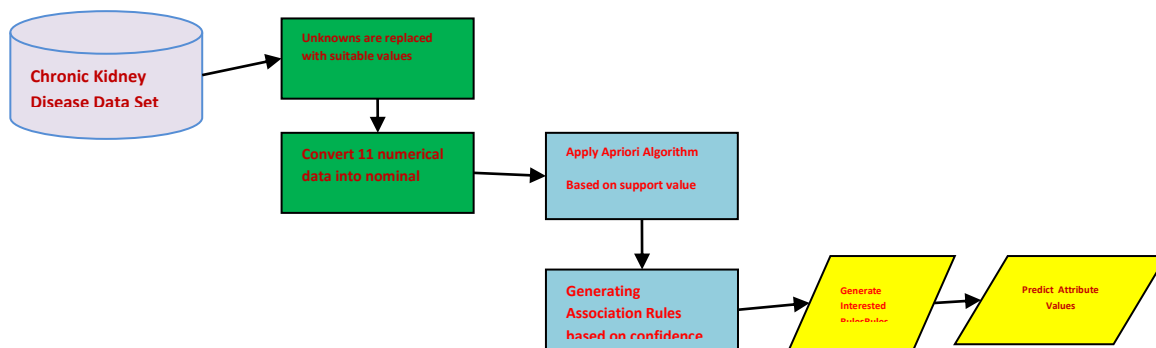
Figure 1. Methodology to Predict Significant Attributes from CKD Data Set

Step1: The first task is Collecting data from UCI Machine learning repository [12] called Chronic Kidney Disease (CKD) which is clinical data. This data set consisting of 400 instances and 250 attributes which are in ARFF file format which has been converted into CSV file format. Clinical Attributes and attribute values are analyzed to understand the significance in identifying CKD. Among 25 attributes, 11 attributes are numerical and 14 attributes are nominal and some of the attribute are specified here.

➤ **Blood Pressure(BP)**- Over time, high blood pressure can damage blood vessels throughout body. This can reduce the blood supply to important organs like the kidneys. High blood pressure also damages the tiny filtering units in kidneys. As a result, the kidneys may stop removing wastes and extra fluid from blood. The extra fluid in blood vessels may build up and raise blood pressure more. It is a numerical data.

➤ **Specific Gravity or Urine Specific Gravity (USG)** is a comparison of the density of urine to that of water. It is an estimate of urine osmolality. Urine that is too concentrated could signal that kidneys are not functioning properly. It is a nominal.

➤ **Diabetes Mellitus (DM)-** Diabetes mellitus, usually called diabetes, is a disease in which body does not make enough insulin or cannot use normal amounts of insulin properly. Insulin is a hormone that regulates the amount of sugar in your blood. A high blood sugar level can cause problems in many parts of your body. It is a nominal

➤ **Serum Creatinine(SC):** Creatinine is a chemical waste product in the blood that passes through the kidneys to be filtered and eliminated in urine. Serum creatinine level can determine the stages of CKD. It is a nominal

➤ **Sodium:** Sodium is one of the body's three major electrolytes. Electrolytes control the fluids going in and out of the body's tissues and cells. Salt is a major source of electrolytes. Sodium contributes to regulate blood pressure and blood volume, Helping transmit impulses for nerve function and muscle contraction and regulating the acid-base balance of blood and body fluids. Too much sodium can be harmful for people. It is numerical data.

Step2: So many instances are having unknowns by '?'. Unknowns can be replaced by mean or mode or median value or frequent value. Depends on the significance unknown values in attributes are replaced by frequent or mean value.

For example attribute name bp that is blood pressure value having '?' is replaced with 'high blood pressure value' because hypertension attribute value shows that patient has BP. Another attribute named sc Serum creatinine value "?" is replaced by most frequent value.

Step 3: Here 11 attributes are numerical so they have been converted into nominal by binning method in preprocessing technique.The attributes converted into nominal are shown in table 1.

Table 1. Description of 11 Attributes Converting From Numerical to Nominal

Slno	Name of the Attribute	Description	Possible values		
1.	Age	Age In Years	Young	Adult	Senior Citizen
			2-30	30-60	60 above
2	Bp	Blood Pressure In mm/Hg	Low	Normal	High
			50	50-140	140 above
3	Bgr	Blood Glucose Random, mgs/dl	Low	Normal	High
			1,2	3	4,5
4	Bu	Blood Urea, mgs/dl	Low	Normal	High
			7	7 to 20	20 above
5	Sc	Serum Creatinine, mgs/dl	Low	Normal	High
			0.5	0.5-13.5	13.5 above
6	Sod	Sodium, mEq/L	Low	Normal	High
			120	120-150	150 above
7	Pot	Potassium, mEq/L	Low	Normal	High
			2.5	2.5-7.6	7.6above

8	Hemo	Hemoglobin, gms	Low	Normal	High
			6.1	6.1-17.8	17.8 above
9	Pcv	Packed Cell Volume, g/dL	Low	Normal	High
			40	40.7-50.3%	50.3 above
10	Wbcc	White Blood Cell Count, cells/cumm	Low	Normal	High
			2200	2200-26400	26400 above
11	Rbcc	Red Blood Cell Count, millions/cmm	Low	Normal	High
			2.1	2.1-6.7	6.7 above

Step 4: Apply Apriori algorithm based on Minimum support and confidence specified by the user. Minimum support is 40 %.

Step 5. Association rules are generated from frequent item sets which are obtained in fourth step based Minimum confidence value is 60% and some of the strong rules are given here.

Strong Association Rules:

➢ **Rule : sod=high and dm=yes ==> class=ckd with 75% confidence.**

This rule states that if the patient has sodium level is high that is above 150 and also has diabeties then that patient may get chronic kidney diseases (ckd).

➢ **Rule: rbc=abnormal and sc=high sod=high ==> class=ckd**

This rule specifies states that the patient having less number of red blood cells count and also sodium level in urine is high, then the confidence of this association rule is 80% then. So this patient may get chance of chronic kidney diseases.

➢ **Rule: htn=yes and cad=no ==> class=ckd (not Interested rule)**

This rule states, if the patient is suffering with hypertension and even through no coronary artery disease but may get chronic kidney diseases(ckd) with confidence 60%.

➢ **Rule: sc=high and sod=high and htn=yes ==> class=ckd**

This rule states, if the patient having serum creatinine level in blood is high, sodium level in urine is high, and also sufferes with hypertension then this association rule says that 85% confidence that he/she may get chance chronic kidney diseases (CKD).

➢ **Rule: sod=high and wbcc=low ==> class=ckd**

This rule specifies, if the person has sodium level is high and white blood cells count is very less then this association rule says that 65% confidence that he/she may get chronic kidney diseases (ckd).

➢ **Rule :sod=high and dm=yes pe=yes ==> class=ckd**

This rule states, if the patient has high sodium level also has diabetes and also having Pedal Edema then this association rule says that 92% confidence that he/she may get chronic kidney diseases (ckd).

➢ **Rule: age =seniorcitizen and sod1=high and pe=yes ==> class=ckd**

This rule states, that the person age is above 60 years, and high sodium levels in urine and also pedal edema then the association rule says that 96% confidence that he/she may get chronic kidney diseases (ckd).

➢ **Rule: bgr=high and pot=high ==> class=ckd**

This rule specifies, if the patient has high sugar levels, high potassium levels then the association rule says that 67% confidence that he/she may get chance of chronic kidney diseases (ckd).

➢ **Rule: sc=high and sod=high pot=high ==> class=ckd**

This rule tells, if the person is having serum creatinine level in blood is high, sodium level in urine is high, and also high potassium then this association rule says that 95% confidence that he/she may get chance of chronic kidney diseases (ckd).

Step 6: Obtain only interested rules based on high confidence rules and attributes significance. Few pruned rules are given below.

☒ **Rule: htn=yes and cad=no ==> class=ckd**

This rule states, if the patient is suffering with hypertension and even through no coronary artery disease but may get chronic kidney diseases(ckd) with confidence 60%. Based on the rule it is correct but based on attributes significance it may not be true. So this rule is pruned.

☒ **Rule: bacteria=present and rbcc=less ==> class=ckd with confidenc 61%.**
This rule is true based on data but not interested to analysis so this has been proved.

Step 7: Identifying Predictive attributes which are used to represent chances of getting CKD.

These rules give the different combinations of attributes which can cause CKD with the effect of other attributes. Significant attributes and their values are given below.

Table 2: Predictive Attributes determined from Association Rule Mining

S.No.	Attribute	Value	S.No.	Attribute	Value
1	Hypertension	Yes, Above 140 mm/Hg	8	Sodium level	High, Above 150 mEq/L
2	Specific Gravity	1.015	9	Hemoglobin	Low, <=6.1 Gms
3	Diabetes mellitus	Yes, mgs/dl	10	Potassium	High, 7.6 above mEq/L
4	Serum Creatinine	High, Above 13.5 mgs/dl	11	Pedal Adima	Yes
5	Red Blood Cells Count	Low, <=2.1 Millions/Cmm	12	Age	Senior Citizen Above 60 years
6	White Blood Cells Count	Low, 2200 cells/cumm	13	Appetite	Poor
7	Blood Urea Nitrozen mgs/dl	High,	14	Albumin	High, Above 5

From this analysis some observation can be made which are as follows.
Usually High Sodium level increases Blood Pressure but also increases CKD. High potassium increases advanced CKD. Less count of Red blood cells, Hemoglobin, packed cell volume and CKD leads to anemia. There is a chance that 70% of the people suffering from CKD might have hypertension or High blood pressure with chronic artery disease. If potassium and sodium levels are high that leads to advanced CKD. Bacteria, virus (HIV), Hepatitis and parasites (malaria),pus cells in urine indicates infection in the kidney that may be acute or chronic. Above 60 years of people having serum creatinine, blood sugar, specific gravity are high 90 % chances to get diseases with kidney functioning. However the children and young age people also suffering from kidney disease either acute or chronic but because of high medication they have taken for some other problems or diabetics or hypertension or high albumin urea so people are suffering from this problem irrespective of age. Early prediction is more than late diagnosis & curing. Children with chronic kidney failure may not have any symptoms until about 80% of their kidney function is lost. Then, they may feel tired, have nausea or vomiting, having difficulty in concentrating. Fluid build-up appears as swelling in the skin, high blood pressure and fluid congestion in the lungs. Blood and urine tests should be carried out regularly to examine patients to detect the disease early.
.

8. CONCLUTION AND FUTURE WORK
Kidneys remove wastes and extra fluid from the body and also remove acid that is produced by the cells of a body and maintains a healthy balance of water, minerals salts, such as sodium, calcium, phosphorus, and potassium in blood. Kidneys also make hormones to control blood pressure also to make red blood cells and also keeps bones strong and healthy. Without this balance, nerves, muscles and other tissues in the body may not work normally.

So Kidney function must be good but so many people are facing this disease frequently. Once the patient has the problem in kidney which is chronic he/she may be suggested to go for dialysis or kidney transplantation or medication but the cost spent for this is very high which is not affordable by a common people. As we know that detection is better than curing so more research is going on this topic to improve the quality of life for people with chronic illnesses.

In this paper, Chronic Kidney Data (CKD) set is considered which is a clinical observations from 400 patients to do analysis to predict the attribute/attributes with particular values which leads to CKD. These attribute values may also be considered as symptoms for suffering from kidney disease. These predictive attributes can be determined by applying data mining technique such as association

rule mining which determines strong rules based on user specified confidence using apriori algorithm. But all the rules are not interested to the expert so only interested rules are collected by knowing interest factors from experts/analysts. In this way, attribute/attributes and their specified values are predicted.

The main objective of this research is to reduce the CKD cases by proposing different conditions which can lead to CKD and also gives different conditions to prevent from advanced CKD by detecting CKD stages at early. In future works, the research can be further extended to get more accurate results to alert people to avoid bad situations before occurring by enhancing the analysis considering other factors such as family history, food habits, medicines used for other diseases, drinking water composition, toxic materials in water, genetic disorder, living style and other bacterial & viral infections on CKD.

9. REFERENCES

[1]. Ajit Kumar Roy, Kumar, J Biom, "Impact of Big Data Analytics on Healthcare and Society", Biostat 2016, volume7, Issue: 3, pages:7 http://dx.doi.org/10.4172/2155-6180.1000300

[2]. https://www.ibef.org/industry/healthcare-india.aspx]

[3]. Doron Shalvi & Nicholas DeClaris, "An Unsupervised Neural Network Approach To Medical Data Mining Techniques", IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227) 1998, ISBN 0-7803-4859-1,Volume 3, Page 1688 – 2561.

[4]. L.Chang & C.H.Chen, "Applying Decision Tree and Neural Network To Increase Quality Of Dermatologic Diagnosis", Article in Expert Systems with Applications- Elsevier, Volume: 36(2), pp. 4035-4041, March 2009.

[5]. Michael Barnathan & Jingjing Zhang & Vasileios, "A Web-Accessible Framework for the Automated Storage and Texture Analysis of Biomedical Images", 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro， P1-3. May, 2008, Print ISSBN: 978-1-4244-2002-5.

[6].Shashi Chhikara1, Purushottam Sharma,"Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal For Research In Applied Science And Engineering Technology (IJRASET), *Vol. 2 Issue 5, May 2014, ISSN: 2321-9653.*
[7]. M. Durairaj, V.Ranjani, "Data Mining Application In Health care Sector: A Study", International Journal of scientific and Technology Research, Vol. 2, Issue 10, 2013.
[8]. Abhishek, Gour Sundar Mitra Thakur, Dolly Gupta, (2012) "Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis", International Journal of Computer Science and Information Technologies, Vol. 3 (3), pp 3900-3904 .

[9]. Tommaso Di Noia, Vito Claudio Ostuni, Francesco Pesce,Giulio Binetti, David Naso, Francesco Paolo Schena, EugenioDi Sciascio, An end stage kidney disease predictor based on an artificial neural networks ensemble, Elsevier Publication,Expert Systems with Applications 40, Page No 4438-4445,2013.

[10] Agrawal R., Imielinski T., and Swami A., Mining association rules between set of items in large databases. In Proceedings of ACM SIGMOD, pages 207-216, May 1993.

 [11] Han Kamber, Data preprocessing techniques for data mining concepts and Techniques, Second Edition, Elsevier.

[12].https://archive.ics.uci.edu/ml/machine-learning-databases/00336/ .