RSIP Career Basic ML 022

Chronic kidney disease prediction using Watson Auto AI

Made BY: Amrit Kumar

May-June 2020

1. INTRODUCTION

1.1 overview

Our kidneys are vital organs that clean our blood and affect many different bodily functions. Various medical conditions can cause the kidneys to stop working properly for a short time – for example, if blood isn't pumped around the body properly or if an enlarged prostate or kidney stones make urine build up in the kidneys.

But if the kidneys still aren't working at full capacity more than three months later or are permanently damaged, then the problem is considered to be chronic kidney disease. The most common causes in adults are diabetes and high blood pressure.

Older people are especially prone to developing chronic kidney disease. It often goes unnoticed for a long time because poorly functioning kidneys don't cause any problems at first.

Treatment mostly aims to stop the kidney disease from getting worse, or at least delay the process. It is also important to treat complications such as anemia, acidosis or changes in bone metabolism. If the kidneys fail completely, dialysis or a kidney transplant can prolong the person's life.

In the early stages, the kidneys can still clean enough blood although they aren't functioning properly. This state can last for years, and sometimes the kidneys even get better again.

But they may get worse instead, resulting in kidney failure (renal insufficiency). Over time, the kidneys may fail completely in some people. This is referred to as "end-stage kidney disease," and it is life-threatening.

1.2 purpose

A global health problem which is steadily growing is Chronic kidney disease (CKD). It is a chronic condition associated with increased morbidity and mortality, a high risk of many other diseases including cardiovascular disease, and high health care costs. Over two million people worldwide receive dialysis or

kidney transplant treatment to stay alive, yet this number may represent only 10% of people who need treatment to live.

The majority of the 2 million people who receive treatment for kidney failure are in only five relatively wealthy countries, which represent 12% of the global population.

By comparison, only 20% of the world's population is treated in about 100 developing countries, and they represent almost half the global population. Annually, more than one million people in 112 lower-income countries die from untreated kidney failure, due to the huge financial burden of dialysis or kidney transplantation treatment.

Thus, there is significant importance in the early detection, controlling, and managing of the disease. It is necessary to predict the progression of CKD with reasonable accuracy because of its dynamic and covert nature in the early stages, and patient heterogeneity. CKD is often described by severity stages.

Clinical decisions are influenced by the stage, whether a patient is progressing, and the rate of progression. Also, defining the disease stage is quite crucial as it gives several indications that support the determination of required intervention and treatments.

Machine learning algorithms have been used to predict and classify in the healthcare field. Yu et al. have used the Support Vector Machine Algorithm to classify and predict diabetes and pre-diabetes patients, and the results show that SVM is useful to classify patients with common diseases. Similarly, Magnin et al. have classified Alzheimer's disease by using a Support Vector Machine (SVM) to analyze whole-brain anatomical magnetic resonance imaging (MRI) for a set of patients, and the results shows that SVM is a promising approach for Alzheimer's disease early detection. Dessai et al. have done heart disease prediction using the Probabilistic Neural Network Algorithm, Decision tree Algorithm, and Naïve Bayes Algorithm, and PRNN provides the best results compared with other algorithms for heart disease prediction. Cao et al. have done prediction of HBV-induced liver cirrhosis using the Multilayered Perceptron (MLP) Algorithm and the results shows that the MLP classifier gives satisfactory prediction outputs for liver disease, mostly in HBV-related liver cirrhosis patients.

2. LITERATURE SURVEY

2.1 Existing Problem

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated it in the early stages. Usually, people are not aware that medical tests, we take for different purposes could contain valuable information concerning kidney diseases. Consequently, attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease.
The information says that it helps us to measure the severity of the problem, the predicted survival of the patient after the illness, the pattern of the disease and work for curing the disease.

2.2 Proposed Solution

In this proposed system we are able to identify the patients with disease. Once any person gets kidney disease, they may suffer from the disease which may decrease their working capability as well as living quality.

Our aim is to predict patients with chronic kidney failure (ckd) disease and patients who do not (notckd) suffer from the disease. So for that we are building a Machine Learning model to predict the compressive strength of concrete using IBM Watson AutoAI Machine Learning Service.

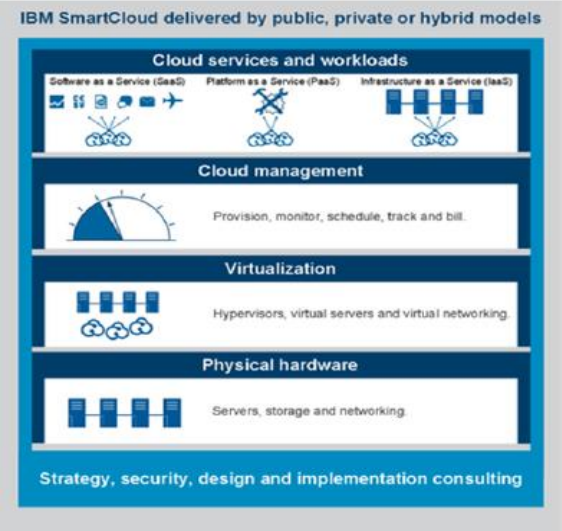
The model is deployed on IBM cloud to get scoring end point which can be used as API in mobile app or web app building. We are developing a web application which is built using node red service. We make use of the scoring end point to give user input values to the deployed model. The model prediction is then showcased on User Interface.

3. THEORITICAL ANALYSIS

IBM cloud computing is a set of cloud computing services for business offered by the information technology company IBM. IBM Cloud includes infrastructure as a service (IaaS), software as a service (SaaS) and platform as a service (PaaS) offered through public, private and hybrid cloud delivery models, in addition to the components that make up those clouds.

3.1 Block diagram

IBM offers three hardware platforms for cloud computing. These platforms offer built-in support for virtualization. For virtualization IBM offers IBM Websphere application infrastructure that supports programming models and open standards for virtualization. The management layer of the IBM cloud framework includes IBM Tivoli middleware.
Management tools provide capabilities to regulate images with automated provisioning and de-provisioning, monitor operations and meter usage while tracking
costs and allocating billing. The last layer of the framework provides integrated workload tools. Workloads for cloud computing are services or instances of code that can be executed to meet specific business needs. IBM offers tools for cloud based collaboration, development and test, application development, analytics, business-to-business integration, and security.
IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. With its open, flexible multi-cloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI, such as:
● Automate AI lifecycle management with AutoAI
● Visually prepare and build models with IBM SPSS Modeler
● Build models using images with IBM Watson Visual Recognition and texts with IBM Watson Natural Language Classifier
● Deploy and run models through one-click integration with IBM Watson Machine Learning



3.2 Hardware / Software designing

While building models with Watson AutoAI on IBM Cloud platform it automatically selects the best suited Machine Learning Algorithm for the provided dataset and the selection criteria. Watson AutoAI does so by applying pipelining to the datasets, it puts the provided datasets in the multiple Machine Learning Algorithms and selects the best suited algorithm for the provided datasets based on minimum RMSE(RootMean Squared Errors) values of the algorithms. Above working is shown by the following screen shots of AutoAI experiment

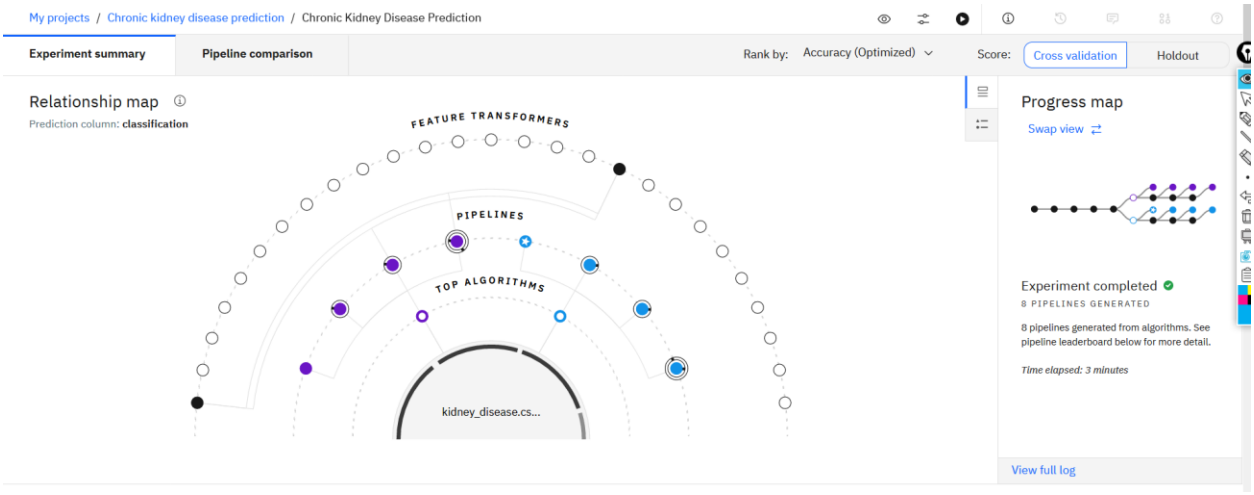


Fig:-Result of Running Pipelines

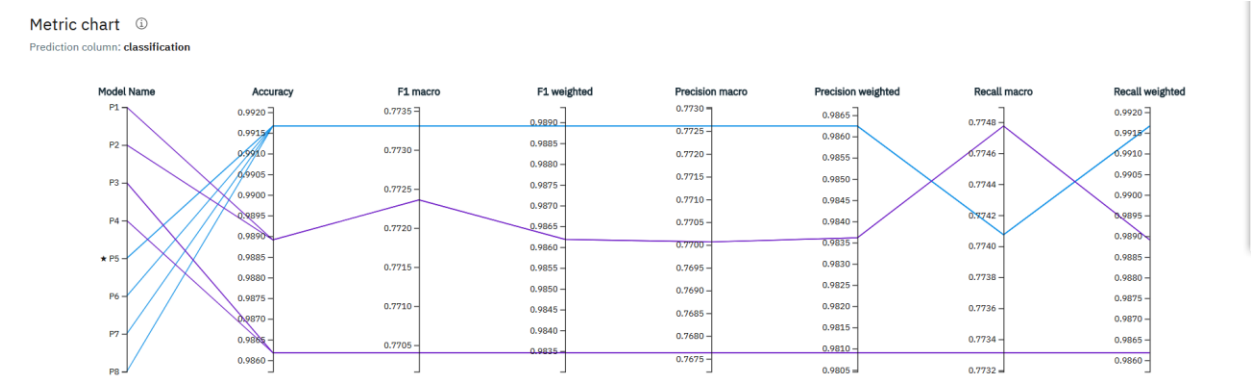


Fig:-Pipeline Comparison

Pipeline leaderboard

Rank	Name	Algorithm	Accuracy (Opti...	F1 macro	F1 micro	F1 weigh...	Precision...	Precision...	Precision...	Recall m...	Recall mi...
★ 1	Pipeline 5	Random Forest Classifier	0.992	0.773	0.992	0.989	0.773	0.992	0.986	0.774	0.992
2	Pipeline 6	Random Forest Classifier	0.992	0.773	0.992	0.989	0.773	0.992	0.986	0.774	0.992
3	Pipeline 7	Random Forest Classifier	0.992	0.773	0.992	0.989	0.773	0.992	0.986	0.774	0.992
4	Pipeline 8	Random Forest Classifier	0.992	0.773	0.992	0.989	0.773	0.992	0.986	0.774	0.992
5	Pipeline 1	Extra Trees Classifier	0.989	0.772	0.989	0.986	0.770	0.989	0.984	0.775	0.989
6	Pipeline 2	Extra Trees Classifier	0.989	0.772	0.989	0.986	0.770	0.989	0.984	0.775	0.989
7	Pipeline 3	Extra Trees Classifier	0.986	0.770	0.986	0.983	0.768	0.986	0.981	0.773	0.986
8	Pipeline 4	Extra Trees Classifier	0.986	0.770	0.986	0.983	0.768	0.986	0.981	0.773	0.986

Fig:-Pipeline Leader board

Watson AutoAI provide Artificial Intelligence platform for implementing different algorithms. After successful implementation of algorithm , IBM Cloud also provide different services for hosting your algorithm through web apps. One such platform is provided through the Node Red App. It is built upon Nodejs for creating web apps.

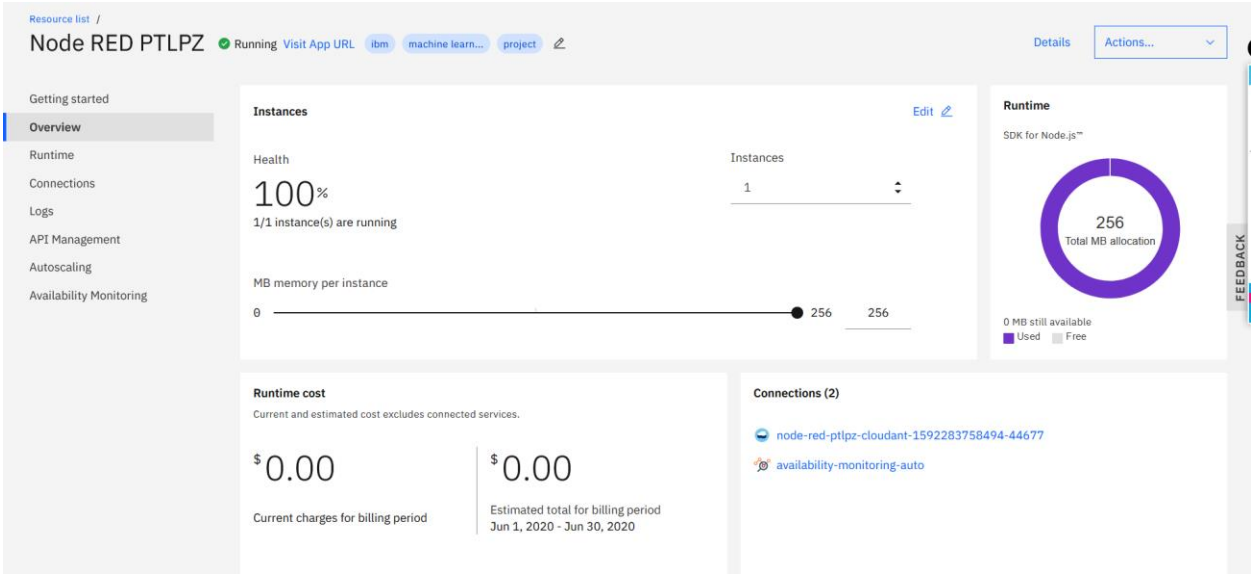


Fig:- Node-Red Service

4. EXPERIMENTAL INVESTIGATION

While working on the implementation of your Machine Learning or Artificial Intelligence Project , IBM Cloud also provides very essential feature to test your implementation before you can deploy it , wherever you wanted to be. In context with the Chronic Kidney Disease Prediction after creating and saving of the best fit algorithm based on minimum RMSE(Root Mean Squared Error), we test our Watson AutoAI model :

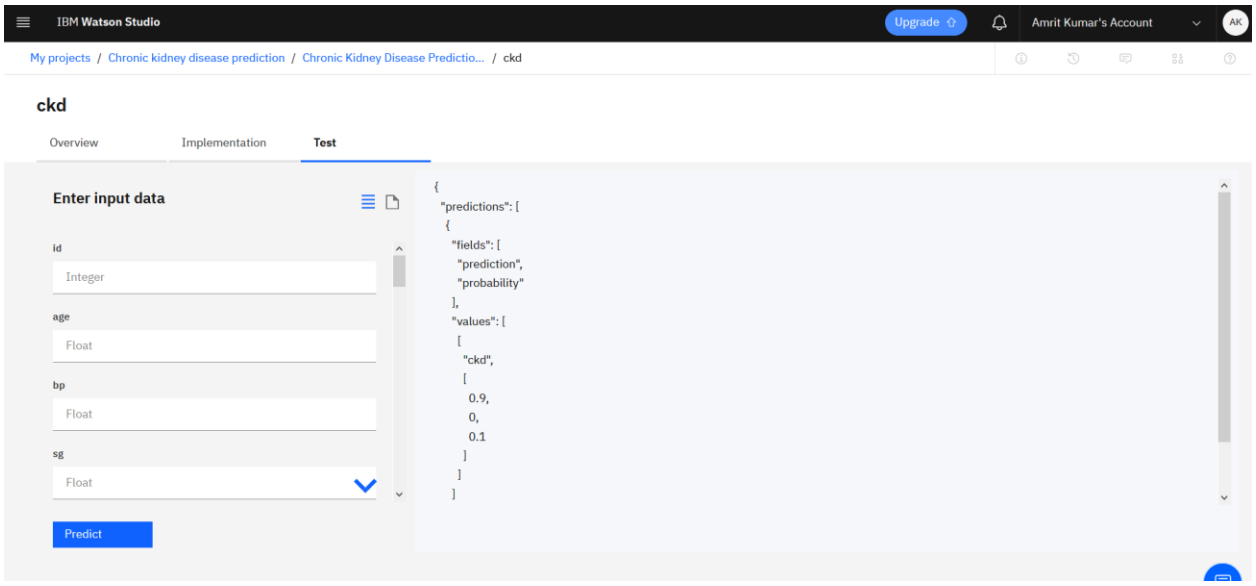
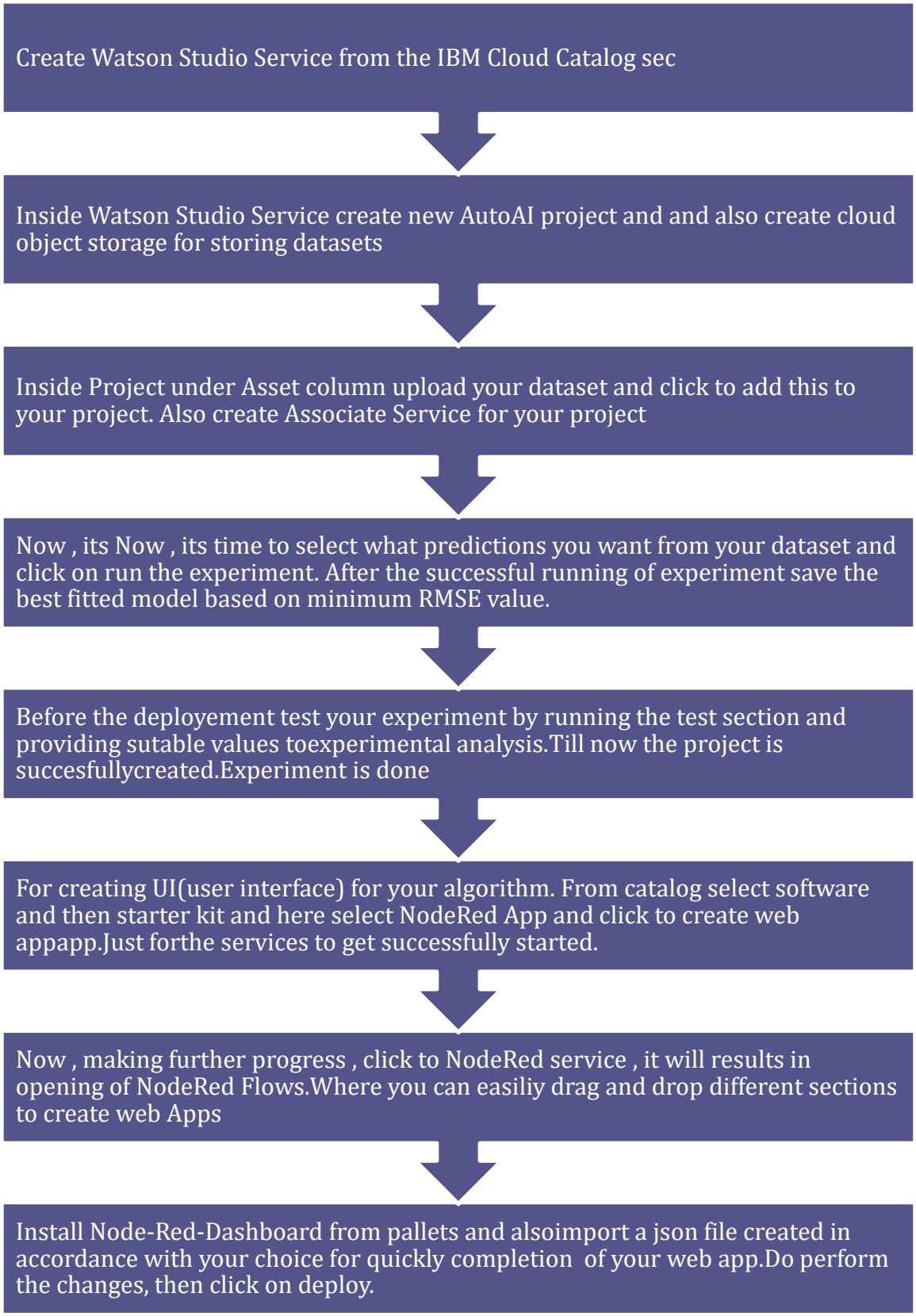


Fig:-Test Run of Project

In the above test run it is quite clear that on the basis of inputted data this model predicts the probability of a Patient that if a person are closer or not to having the chronic kidney disease.

5. FLOWCHART


Create Watson Studio Service from the IBM Cloud Catalog sec
Inside Watson Studio Service create new AutoAI project and and also create cloud object storage for storing datasets
Inside Project under Asset column upload your dataset and click to add this to your project. Also create Associate Service for your project
Now , its Now , its time to select what predictions you want from your dataset and click on run the experiment. After the successful running of experiment save the best fitted model based on minimum RMSE value.
Before the deployement test your experiment by running the test section and providing sutable values toexperimental analysis.Till now the project is succesfullycreated.Experiment is done
For creating UI(user interface) for your algorithm. From catalog select software and then starter kit and here select NodeRed App and click to create web appapp.Just forthe services to get successfully started.
Now , making further progress , click to NodeRed service , it will results in opening of NodeRed Flows.Where you can easiliy drag and drop different sections to create web Apps
Install Node-Red-Dashboard from pallets and alsoimport a json file created in accordance with your choice for quickly completion of your web app.Do perform the changes, then click on deploy.

6. RESULT

By using the IBM Cloud's Watson AutoAI , we are able to model our given dataset into a good predictable Machine Learning Algorithm and by the help of NodeRed service we are able to turn our Machine Learning Algorithm into a beautiful Web App:

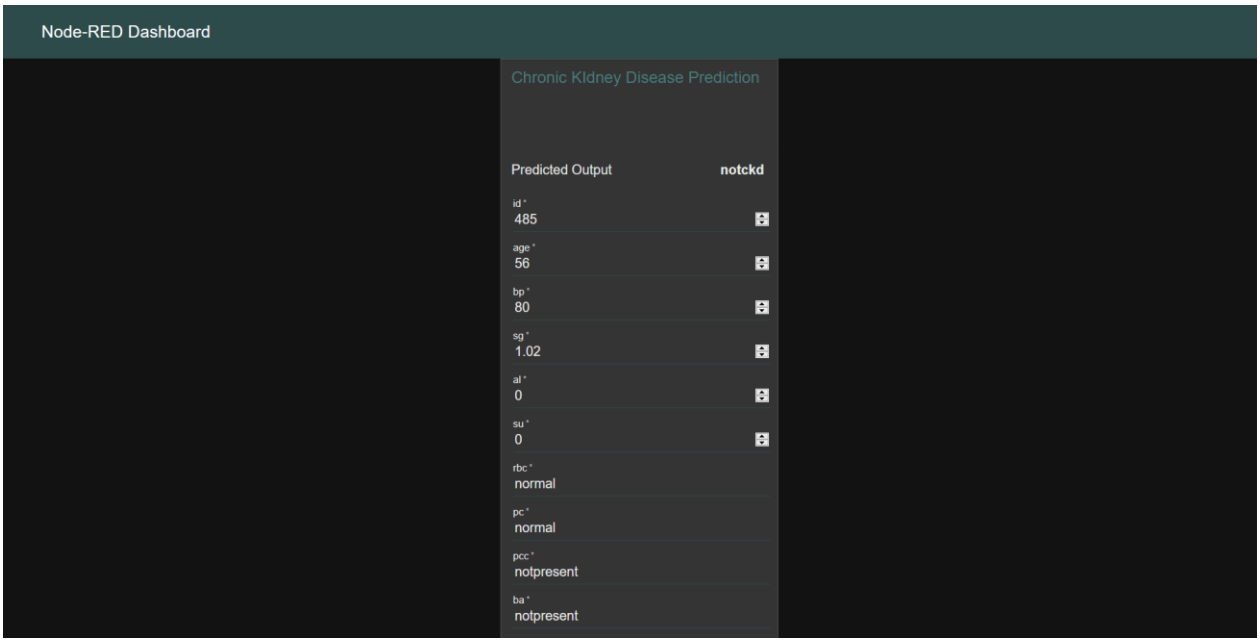


Fig:-Screenshot of Web App UI 1

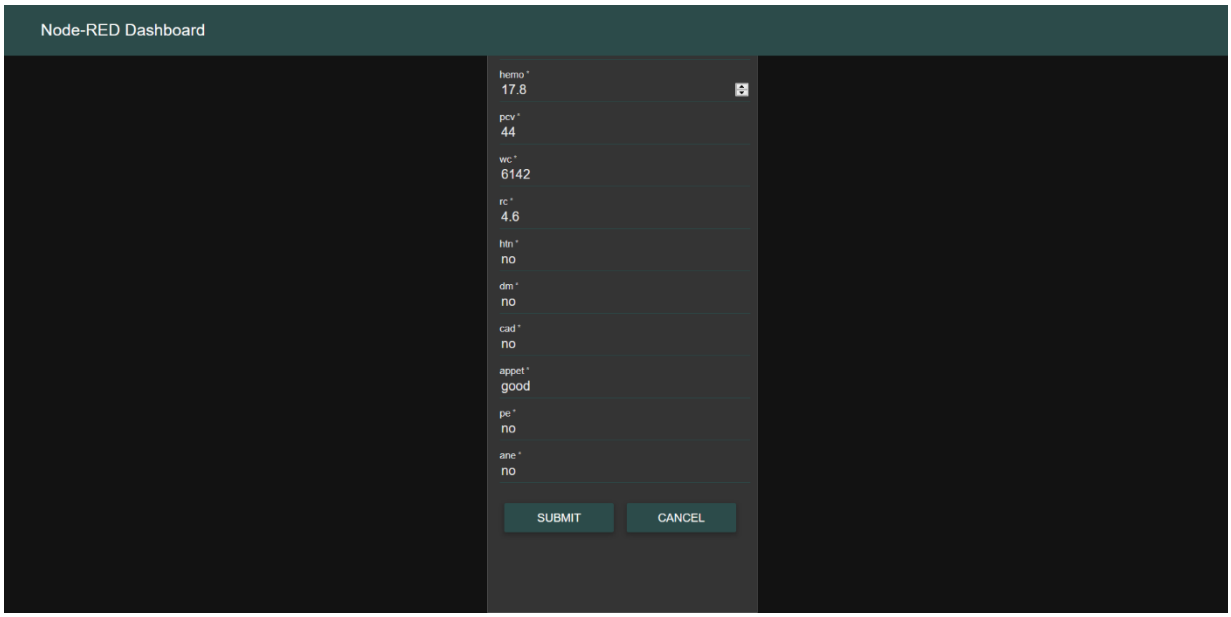


Fig:-Screenshot of Web App UI 2

7.ADVANTAGES AND DISADVANTAGES

Advantages:

- The model improves continuously as the new data is provided model accuracy and efficiency is increased.
- It can predict if a person is suffering from chronic kidney disease from taking the data from normal tests.
- chronic kidney disease are difficult to track but with model we can predict if a person is suffering from one or not easily and accurately.
- Machine learning model do not require human intervention every time hence it can learn on its own and make the model better and provide better results.
- It can reduce death count due kidney diseases by identifying it in patient's and treating it before it gets out of hand.
- It can easily identify trends and pattern which we can use to investigate the cause of the problem.
- Kidney diseases are curable but identifying them is difficult thus with this the problem can avoided.
- The data from this model can be used to identify any other diseases.
- The model can be used to diagnose more number of patient in less time period.
- It can be used to increase the efficiency of the diagnosis takes place.
- The model can be directly used in clinical practice to diagnose patients.

Disadvantages:

- The strength of the data is not higher because of the size of the data set and the missing attribute values.
- To build a machine learning model targeting chronic kidney disease with overall accuracy of 99.99%, will need millions of records with zero missing values.
- Training of model and testing it can take some time if the dataset is increased to get higher accuracy training time is going to increase significantly.

8. APPLICATIONS

- Predicting kidney and renal failure in clinics
- Predicting trends between different disease and kidney failure for insurance companies.
- Hospitals can use it before giving tests for kidney disease.
- Testing centres an use it to check whether patient in question should take the test.
- Hospital management can use it to predict the chances of person renal and kidney failure and weather an RRT is required or not.

9. CONCLUSION

The main purpose of the application is to build a Machine Learning model to predict Chronic Kidney disease using IBM Watson AutoAI Machine Learning Service. The model is deployed on IBM cloud to get scoring end point which can be used as API in mobile app or web app building. We are developing a web application which is built using node red service. We make use of the scoring end point to give user input values to the deployed model. The model prediction is then showcased on User Interface.

10. FUTURE SCOPE

he main advantage of the existing work is in the use of comorbidities for prediction of RRT. A large heterogeneous population should be used to create and evaluate the model's performance before it can be applied to clinical practice. It must be understood that by using ML algorithms, our study provides a screening approach for predicting the chances of upcoming RRT based on the clinical data, therefore this should neither be considered as clinical guideline nor a diagnostic / therapeutic tool for CKD patients. On the other hand, the results at this point are more interesting from the point of view of policy-makers, such as hospital managers or health officials, or insurance companies. Using predictive models on a general population with the data available can allow for better planning and allocation of resources. Future scope lies in coming up with a prediction model that would factor in the more clinical data (use of specific drugs / associated comorbidities / dietary interventions / degree of blood pressure control / degree of blood sugar control) in predicting the outcomes and providing a possible chance for us to tailor the therapeutic interventions accordingly.

We need to realise that the ML algorithms our study provides need to be considered as a possible screening tool to predict the time frame of progression of CKD patient before he/she would need RRT.

11.BIBLIOGRAPHY

- Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients
  https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0233976#sec023

- Prediction of Chronic Kidney Disease Using Machine Learning Algorithm
  https://ijarcce.com/wpcontent/uploads/2018/11/IJARCCE.2018.71021.pdf

APPENDIX

Source Code:

[{"id":"b80a685a.9c8be8","type":"tab","label":"Flow 1","disabled":false,"info":""},{"id":"1902b374.8fd0bd","type":"ui_base","z":0,"theme":{"name":"theme-dark","lightTheme":{"default":"#0094CE","baseColor":"#0094CE","baseFont":"Arial Black,Arial Black,Gadget,sans-serif","edited":false,"reset":false},"darkTheme":{"default":"#097479","baseColor":"#2b4a49","baseFont":"Century Gothic,CenturyGothic,AppleGothic,sans-serif","edited":false,"reset":false},"customTheme":{"name":"Untitled Theme 1","default":"#4B7930","baseColor":"#4B7930","baseFont":"-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-Sans,Ubuntu,Cantarell,Helvetica Neue,sans-serif"},"themeState":{"base-color":{"default":"#097479","value":"#2b4a49","edited":true},"page-titlebar-backgroundColor":{"value":"#2b4a49","edited":false},"page-backgroundColor":{"value":"#111111","edited":false},"page-sidebar-backgroundColor":{"value":"#000000","edited":false},"group-textColor":{"value":"#477a79","edited":false},"group-borderColor":{"value":"#555555","edited":false},"group-backgroundColor":{"value":"#333333","edited":false},"widget-textColor":{"value":"#eeeeee","edited":false},"widget-backgroundColor":{"value":"#2b4a49","edited":false},"widget-borderColor":{"value":"#333333","edited":false},"base-font":{"value":"Century Gothic,CenturyGothic,AppleGothic,sans-serif"}},"angularTheme":{"primary":"teal","accents":"blue","warn":"red","background":"blue-grey","palette":"dark"}},"site":{"name":"Node-RED Dashboard","hideToolbar":"false","allowSwipe":"false","lockMenu":"true","allowTempTheme":"true","dateFormat":"DD/MM/YYYY","sizes":{"sx":50,"sy":50,"gx":0,"gy":0,"cx":6,"cy":6,"px":0,"py":0}}},{"id":"c4035464.585408","type":"ui_group","z":"","name":"Chronic KIdney Disease Prediction","tab":"614ba06b.dd718","order":1,"disp":true,"width":"6","collapse":false},{"id":"614ba06b.dd718","type":"ui_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false},{"id":"184ce0c5.a5bb9f","type":"ui_group","z":"","name":"Default","tab":"35784888.2d2568","order":1,"disp":true,"width":"6","collapse":false},{"id":"35784888.2d2568","type":"ui_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false},{"id":"34e83295.63a1be","type":"ui_spacer","name":"spacer","group":"c4035464.585408","order":1,"width":1,"height":1},{"id":"35a6ad46.0828b2","type":"function","z":"b80a685a.9c8be8","name":"PreToken","func":"global.set(\"id\",msg.payload.id)\nglobal.set(\"age

\",msg.payload.age)\nglobal.set(\"bp\",msg.payload.bp)\nglobal.set(\"sg\",msg.payload.sg)\nglobal.set(\"al\",msg.payload.al)\nglobal.set(\"su\",msg.payload.su)\nglobal.set(\"rbc\",msg.payload.rbc)\nglobal.set(\"pc\",msg.payload.pc)\nglobal.set(\"pcc\",msg.payload.pcc)\nglobal.set(\"ba\",msg.payload.ba)\nglobal.set(\"bgr\",msg.payload.bgr)\nglobal.set(\"bu\",msg.payload.bu)\nglobal.set(\"sc\",msg.payload.sc)\nglobal.set(\"sod\",msg.payload.sod)\nglobal.set(\"pot\",msg.payload.pot)\nglobal.set(\"hemo\",msg.payload.hemo)\nglobal.set(\"pcv\",msg.payload.pcv)\nglobal.set(\"wc\",msg.payload.wc)\nglobal.set(\"rc\",msg.payload.rc)\nglobal.set(\"htn\",msg.payload.htn)\nglobal.set(\"dm\",msg.payload.dm)\nglobal.set(\"cad\",msg.payload.cad)\nglobal.set(\"appet\",msg.payload.appet)\nglobal.set(\"pe\",msg.payload.pe)\nglobal.set(\"ane\",msg.payload.ane)\nvar apikey=\"qbdxHbpHFaR73uvH4SHL76WAiFeT-M6GU4YMrvlUvTKm\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;","outputs":1,"noerr":0,"x":320,"y":460,"wires":[["d37618c8.600498"]]},{"id":"d37618c8.600498","type":"http request","z":"b80a685a.9c8be8","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":510,"y":460,"wires":[["7c9ffc71.cf9814"]]},{"id":"451e92f.9ad476c","type":"debug","z":"b80a685a.9c8be8","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":1130,"y":360,"wires":[]},{"id":"7c9ffc71.cf9814","type":"function","z":"b80a685a.9c8be8","name":"Pre Prediction","func":"var id = global.get('id')\nvar age = global.get('age')\nvar bp = global.get('bp')\nvar sg = global.get('sg')\nvar al = global.get('al')\nvar su = global.get('su')\nvar rbc = global.get('rbc')\nvar pc = global.get('pc')\nvar pcc = global.get('pcc')\nvar ba = global.get('ba')\nvar bgr = global.get('bgr')\nvar bu = global.get('bu')\nvar sc = global.get('sc')\nvar sod = global.get('sod')\nvar pot = global.get('pot')\nvar hemo = global.get('hemo')\nvar pcv = global.get('pcv')\nvar wc = global.get('wc')\nvar rc = global.get('rc')\nvar htn= global.get('htn')\nvar dm = global.get('dm')\nvar cad = global.get('cad')\nvar appet = global.get('appet')\nvar pe = global.get('pe')\nvar ane = global.get('ane')\nvar token=msg.payload.access_token\nvar instance_id=\"9773e91c-9115-422b-bbea-8b3c8755822f\"\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance_id}\nmsg.payload={\"input_data\": [{\"fields\": [\"id\",\"age\",\"bp\",\"sg\",\"al\",\"su\",\"rbc\",\"pc\",\"pcc\",\"ba\",\"bgr\",\"bu\",\"sc\",\"sod\",\"pot\",\"hemo\",\"pcv\",\"wc\",\"rc\",\"htn\",\"dm\",\"

cad\",\"appet\",\"pe\",\"ane\"],\"values\":
[[id,age,bp,sg,al,su,rbc,pc,pcc,ba,bgr,bu,sc,sod,pot,hemo,pcv,wc,rc,htn,dm,cad,appet,pe,ane]]}]}\nreturn
msg","outputs":1,"noerr":0,"x":716.0000076293945,"y":411.9999990463257,"wires":[["4dfa906b.1b78"]]},{"id":"4dfa906b.1b78","type":"http
request","z":"b80a685a.9c8be8","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v4/deployments/c3084a2b-5b41-4068-9bc1-
87fb3e858e44/predictions","tls":"","persist":false,"proxy":"","authType":"","x":809.5000076293945,"y":355.9999990463257,"wires":[["451e92f.9ad476c","8d0dc3b.b6e964"]]},{"id":"8d0dc3b.b6e964","type":"function","z":"b80a685a.9c8be8","name":"","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn
msg;","outputs":1,"noerr":0,"x":669.6000366210938,"y":251.60000133514404,"wires":[["826da6b1.697fc8"]]},{"id":"97e2ad4d.77d13","type":"ui_form","z":"b80a685a.9c8be8","name":"","label":"","group":"c4035464.585408","order":0,"width":0,"height":0,"options":[{"label":"id","value":"id","type":"number","required":true,"rows":null},{"label":"age","value":"age","type":"number","required":true,"rows":null},{"label":"bp","value":"bp","type":"number","required":true,"rows":null},{"label":"sg","value":"sg","type":"number","required":true,"rows":null},{"label":"al","value":"al","type":"number","required":true,"rows":null},{"label":"su","value":"su","type":"number","required":true,"rows":null},{"label":"rbc","value":"rbc","type":"text","required":true,"rows":null},{"label":"pc","value":"pc","type":"text","required":true,"rows":null},{"label":"pcc","value":"pcc","type":"text","required":true,"rows":null},{"label":"ba","value":"ba","type":"text","required":true,"rows":null},{"label":"bgr","value":"bgr","type":"number","required":true,"rows":null},{"label":"bu","value":"bu","type":"number","required":true,"rows":null},{"label":"sc","value":"sc","type":"number","required":true,"rows":null},{"label":"sod","value":"sod","type":"number","required":true,"rows":null},{"label":"pot","value":"pot","type":"number","required":true,"rows":null},{"label":"hemo","value":"hemo","type":"number","required":true,"rows":null},{"label":"pcv","value":"pcv","type":"text","required":true,"rows":null},{"label":"wc","value":"wc","type":"text","required":true,"rows":null},{"label":"rc","value":"rc","type":"text","required":true,"rows":null},{"label":"htn","value":"htn","type":"text","required":true,"rows":null},{"label":"dm","value":"dm","type":"text","required":true,"rows":null},{"label":"cad","value":"cad","type":"text","required":true,"rows":null},{"label":"appet","value":"appet","type":"text","required":true,"rows":null},{"label":"pe","value":"pe","type":"text","required":true,"rows":null},{"label":"ane","value":"ane","type":"text","required":true,"rows":null}],"formValue":{"id":"","age":"","bp":"","sg":"","al":"","su":"","rbc":"","pc":"","pcc":"","ba":"","bgr":"","bu":"","sc":"","sod":"","pot":"","hemo":"","pcv":"","w

c":"","rc":"","htn":"","dm":"","cad":"","appet":"","pe":"","ane":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":110,"y":460,"wires":[["35a6ad46.0828b2"]]},{"id":"826da6b1.697fc8","type":"ui_text","z":"b80a685a.9c8be8","group":"c4035464.585408","order":1,"width":0,"height":0,"name":"","label":"Predicted Output","format":"{{msg.payload}}","layout":"row-spread","x":990,"y":240,"wires":[]}]