GOVERNMENT COLLEGE OF ENGINEERING
(Formerly IRTT)
ERODE-638 316

BONAFIDE CERTIFICATE

Certified that this project titled **"EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING"** is the bonafide work of **"KAVIN KUMAR R (731119205014), MONISH V S (731119205024), POOVARASU T (731119205034), TIRUMAL T N (731119205045)"** who carried out the project work under my supervision.

SIGNATURE OF HOD

Dr.P.KALYANI,M.E.,Ph.D.,
HEAD OF THE DEPARTMENT
DEPARTMENT OF IT,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE – 638316

SIGNATURE OF SPOC

Dr.G.GOWRISON, M.E.,Ph.D.,
ASSISTANT PROFESSOR(SR)
DEPARTMENT OF ECE,
GOVERNMENT COLLEGE OF
ENGINEERING,ERODE - 638316

SIGNATURE OF FACULTY MENTOR

Dr.M.N.SUDHA,M.E.,Ph.D.,
ASSISTANT PROFESSOR,
DEPARTMENT OF IT,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE – 638316

SIGNATURE OF FACULTY EVALUATOR

Dr.P.THANGAVEL,M.E.,Ph.D.,
ASSISTANT PROFESSOR(SR),
DEPARTMENT OF IT,
GOVERNMENT COLLEGE OF
ENGINEERING, ERODE - 638316

PROJECT REPORT (PNT2022TMID44392)

1.INTRODUCTION

1.1 Project Overview

Kidney Disease are the main reason for a huge number of death in the world over the last few decades and has emerged as the most lifethreatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of Kidney related diseases.

1.2 Purpose

Chronic kidney disease (CKD) is a perennial condition where the kidneys deteriorate and stop functioning gradually. This disease has become one of the major public health concerns worldwide. It is insidious, often recognizable only by laboratory abnormalities until its latest stages. The main motive of this work is to ascertain the existence of chronic kidney disease by imposing various classification algorithms on the patient medical record. This research work is primarily concentrated on finding the best suitable classification algorithm which can be used for the diagnosis of CKD based on the classification report and performance factors.

2.LITERATURE SURVEY

2.1 Existing Problem

Chronic Kidney Disease prediction is one of the most important issues in healthcare analytics. The most intresting and challenging tasks in day-to-day life is prediction in medical field.

2.2 References

The efficiency of ML techniques depends on the selection and use of the appropriate features. Hence,this research analysis several feature optimization approaches along with a max voting ensemble model to establish a highly accurate CKD diagnosis system by using an appropriate feature set. The ensemble model of this research is structured with five existing classifiers. Three types of feature optimization namely feature importance, feature reduction, and feature selection where for each approach two most proficient techniques are analyzed with the mentioned ensemble model. The most outstanding result of 99.5% accuracy by using 10- fold cross-validation. Muhammad MinoarHossain et al.,

The examination of dataset by directed supervised machine learning technique (SMLT) to catch a few data resembles, variable recognizable proof, uni-variate investigation, bi-variate and multi-variate investigation,missing worth medicines and break down the information approval, information cleaning/getting ready, and information perception will be done on the whole given dataset. To classifying information from need, and the outcome shows that the adequacy of the proposed AI calculation method can be contrasted and best exactness with accuracy, callback Diddi Priyanka et at.,

They take a few attributes to measure our analysis about chronic kidney disease and this attribute is one of the major occurrences of chronic kidney disease. Therefore 8 machine learning classifier are used to measure analysis using weka tools namely: Logistic Regression (LG), Naive Bayes (NB), Multilayer Perceptron (MLP), Stochastic Gradient Descent (SGD), Adaptive Boosting (Adaboost), Bagging, Decision Tree (DT), Random Forest (RF) classifier are used. They feature extraction of all attributes using principal component analysis (PCA). We gain the highest accuracy from the Random Forest (RF) and it is 99 % and ROC (receiver operating characteristic) curve value is also highest from other algorithms.Minhaz Uddin et al.,

Saurabh Pal(2022) The purpose of the proposed study is to develop and validate a predictive model for the prediction of chronic kidney disease. Machine learning algorithms are often used in medicine to predict and classify diseases. Medical records are often skewed. They have used chronic kidney disease. Dataset from UCI Machine learning repository with 25 features and applied three machine learning classifiers Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM) for analysis and then used bagging ensemble method to improve the results of the developed model. The clusters of the chronic kidney disease dataset were used to train the machine learning classifiers. the Kidney Disease Collection is summarized by category and non- linear features. We get the best result in the case of decision tree with accuracy of 95.92%. Finally, after applying the bagging ensemble method we get the highest accuracy of 97.23%.

The feature selection techniques, namely, Relief-F and chi-squared feature selection method, were applied to select the important features. Six machine learning classification algorithms were used in this research: decision tree (DT), logistic regression (LR), Naive Bayes (NB), Random Forest (RF), support vector machine (SVM), and Gradient-Boosted Trees (GBT

Classifier) as ensemble learning algorithms. Four methods of evaluation, namely, accuracy, precision, recall, and F1-measure, were applied to validate the results. Manal A Abdel-Fattah et al.,

2.3 Problem Statement Definition

Chronic kidney disease (CKD) means your kidneys are damaged and can't filter blood the way they should. The main risk factors for developing kidney disease are diabetes, high blood pressure, heart disease, and a family history of kidney failure.

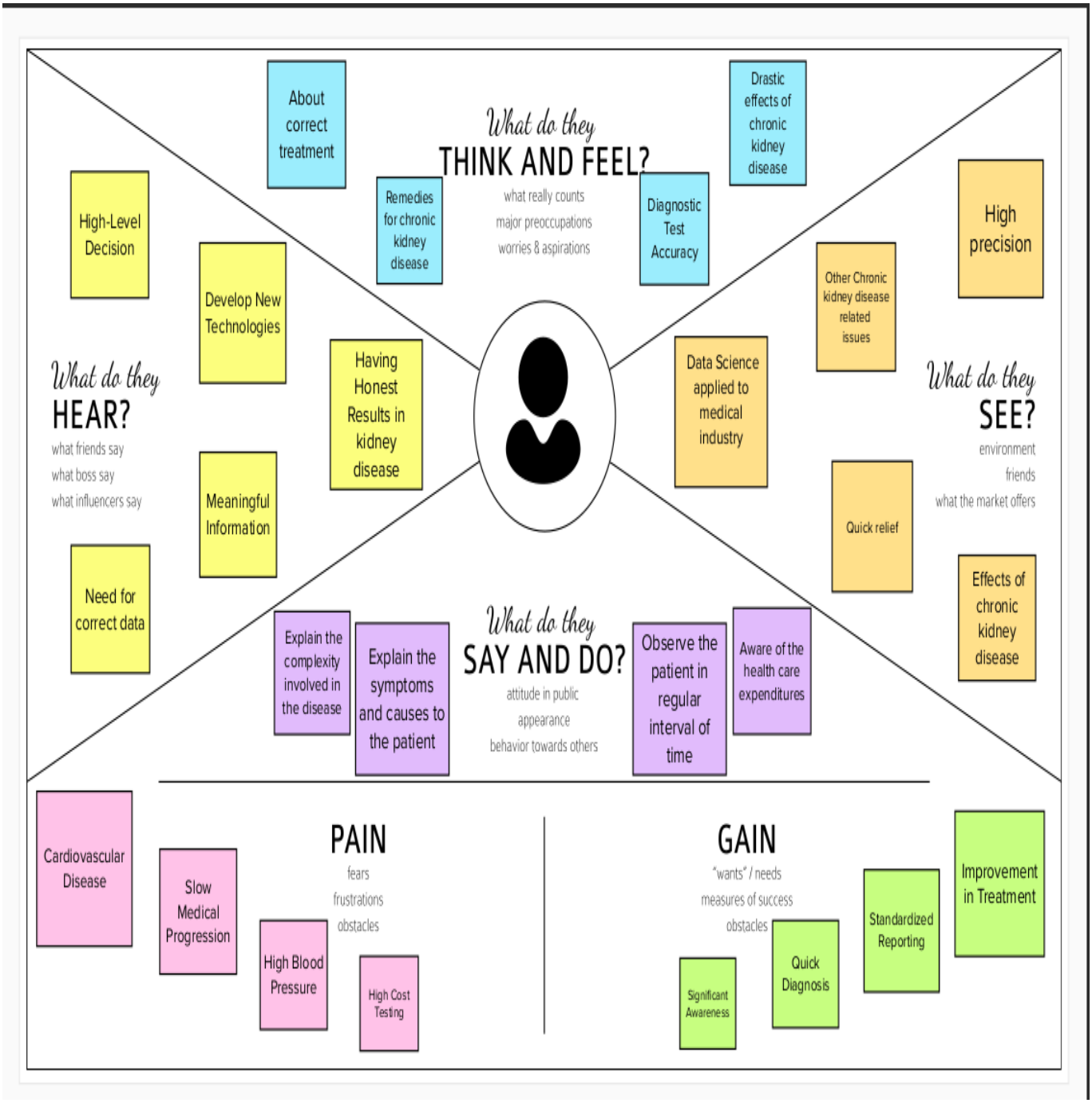
10% of the population worldwide is affected by chronic kidney disease (CKD), and millions die each year because they do not have access to affordable treatment.

People are at risk for kidney disease if people have diabetes, high blood pressure, heart disease, or a family history of kidney failure. If you have risk factors, get tested for kidney disease and protect your kidneys by making healthy food choices, being more active, aiming for a healthy weight, and managing health conditions that cause kidney damage.

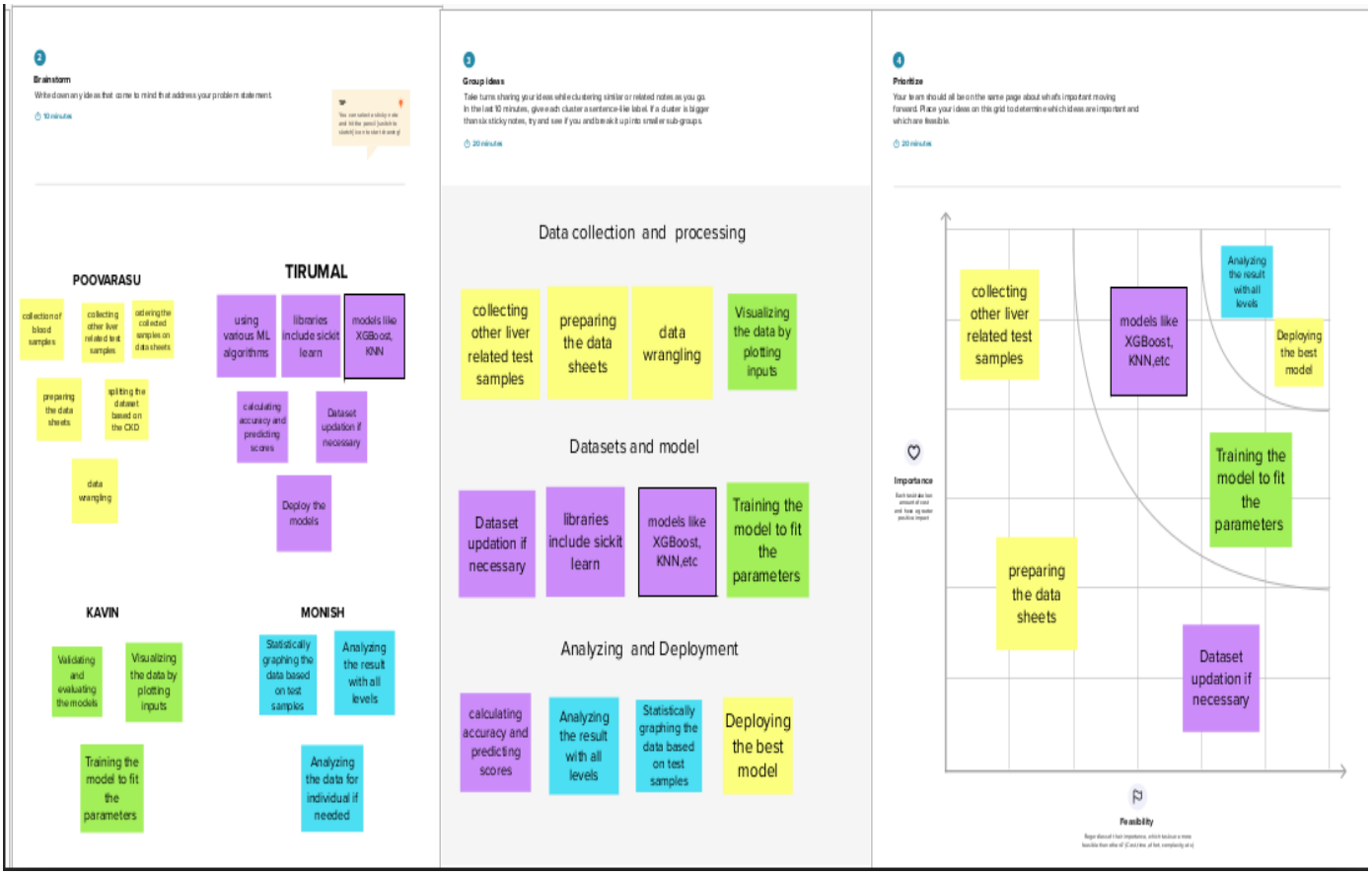
Chronic kidney Disease can be cured, if treated in the early stages. The main aim of this project is to predict whether the patient have chronic kidney disease or not, in more accurate and faster way based on certain diagnostic measurements like Blood Pressure (Bp), Albumin(Al).

3.IDEATION & PROPOSED SOLUTION

3.1 EmpathyMap Canvas



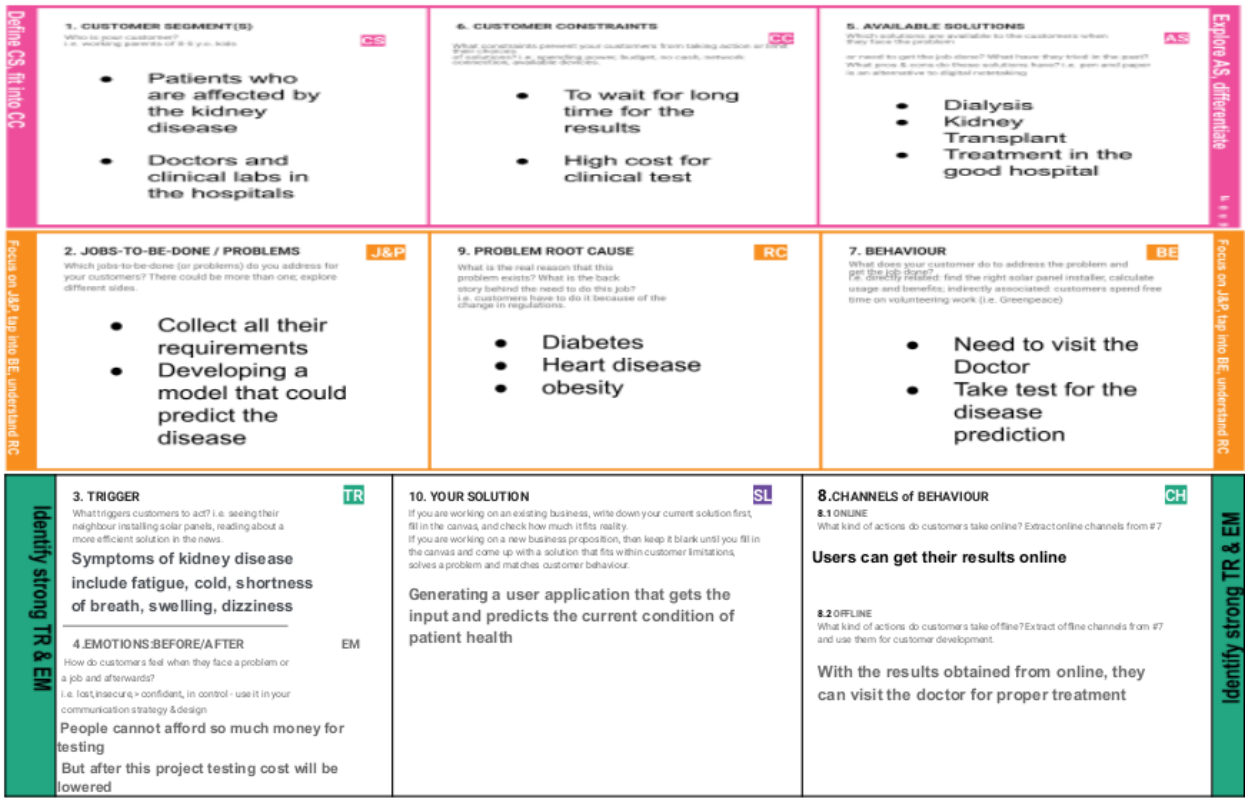
3.2 Ideation & Brainstorming



3.3 PROPOSED SOLUTION

S.NO	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	Chronic kidney disease (CKD) means your kidneys are damaged and can't filter blood the way they should. The main risk factors for developing kidney disease are diabetes, high blood pressure, heart disease, and a family history of kidney failure. To detect CKD Earlier using machine learning.
2.	Idea / Solution description	By taking the past medical history of patients from the dataset as inputs to predict CKD using machine learning concepts.
3.	Novelty / Uniqueness	By analysing different predictions models of machine learning and finding the best of it based on its accuracy.
4.	Social Impact / Customer Satisfaction	Helps in early diagnosis Rapid Recovery Reduction in CKD testing cost
5.	Business model	Commissions for every new features added to the model. One time investment for hospitals
6	Scalability of the Solution	The ability of the model to grow to meet a company's business needs by adding new features**.**

3.4 PROBLEM SOLUTION FIT



4.REQUIREMENT ANALYSIS

4.1 Functional Requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration of their details like bp.**	User should enter the valid datas and his/her original information to predict the result
FR-2	**User Approval**	Confirmation from user to get approve to process the customer data in model.
FR-3	**External Interface**	Flask elements with which a component used to interact and respond
FR-4	**Personal Sharing**	Allows customer to share information to doctor and personal consultant

4.2 Non-functional Requirements

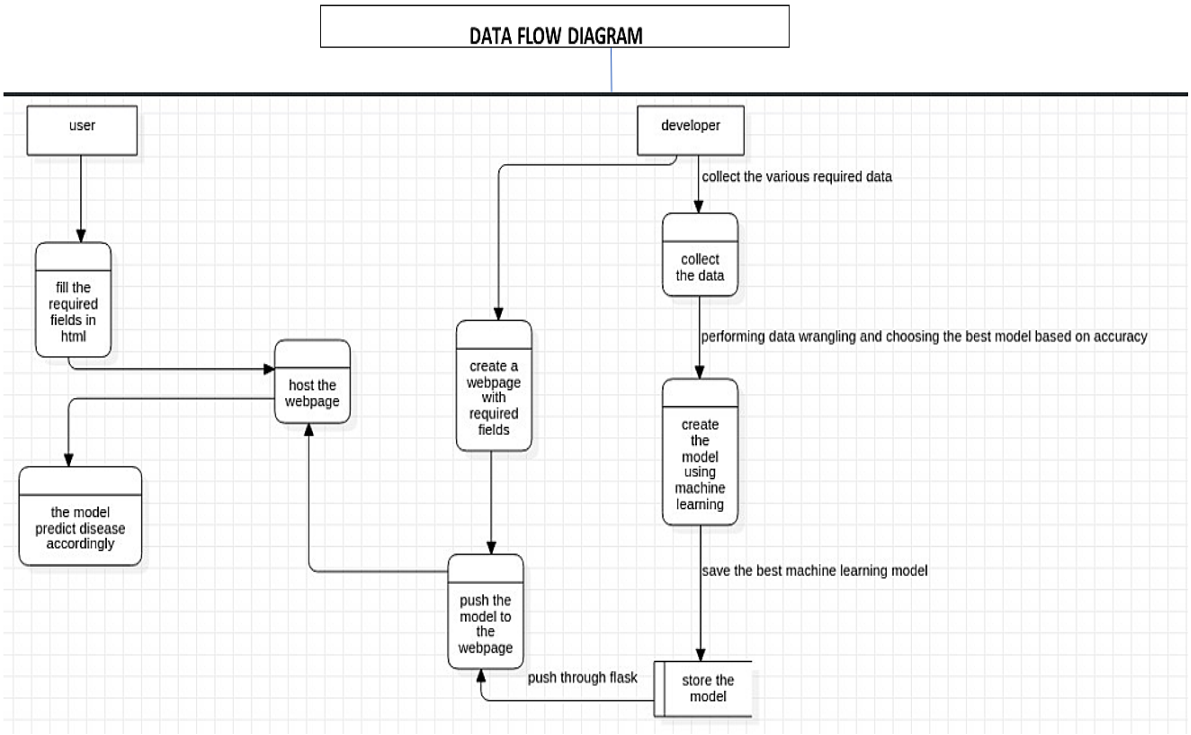
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Provides a interface with are easy to use
NFR-2	**Security**	Protection against possible Difficulties and to avoid Threads
NFR-3	**Reliability**	Provides a consistent usage
NFR-4	**Performance**	Provides effective accomplishment to the tasks assingned
NFR-5	**Availability**	Provides accessibility and Flexibility in a period of time

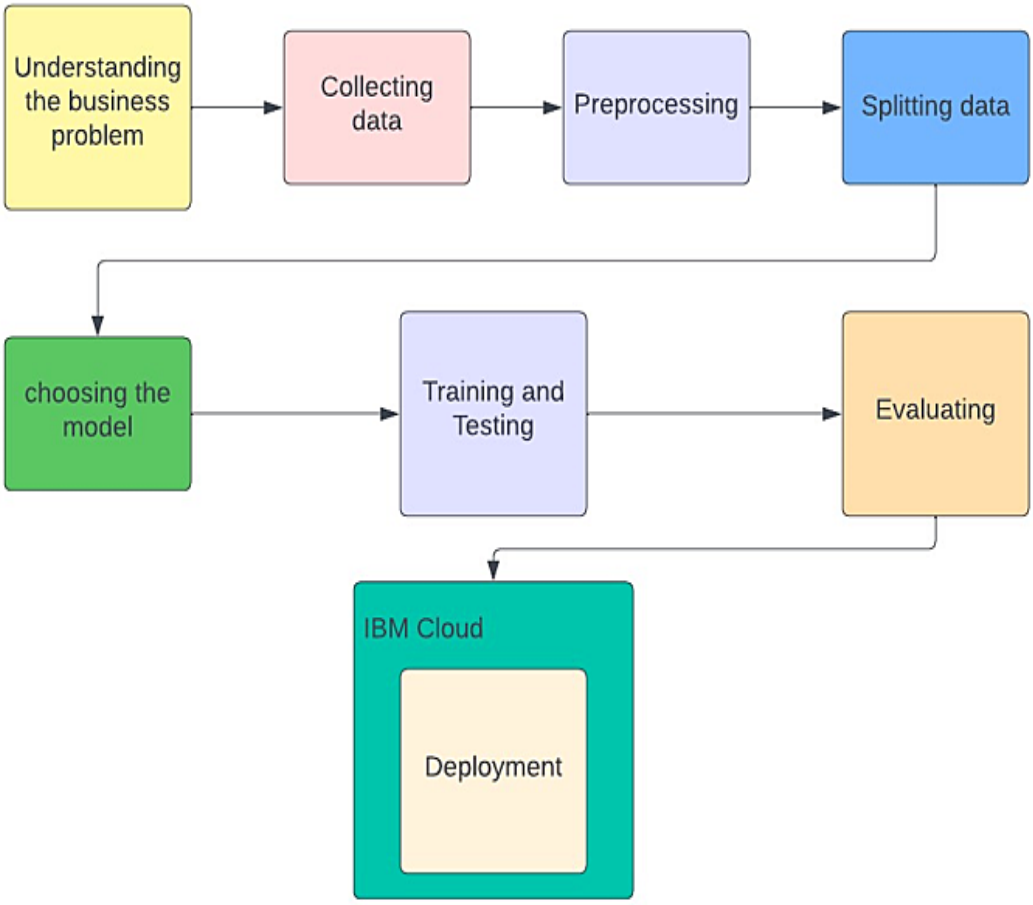
5.PROJECT DESIGN

5.1 Data Flow Diagram & User Stories

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 Solution and Technical Architecture



6.PROJECT PLANNING AND DESIGNING

6.1 Sprint Planning and Estimation

Sprint 1,2,3 and 4:

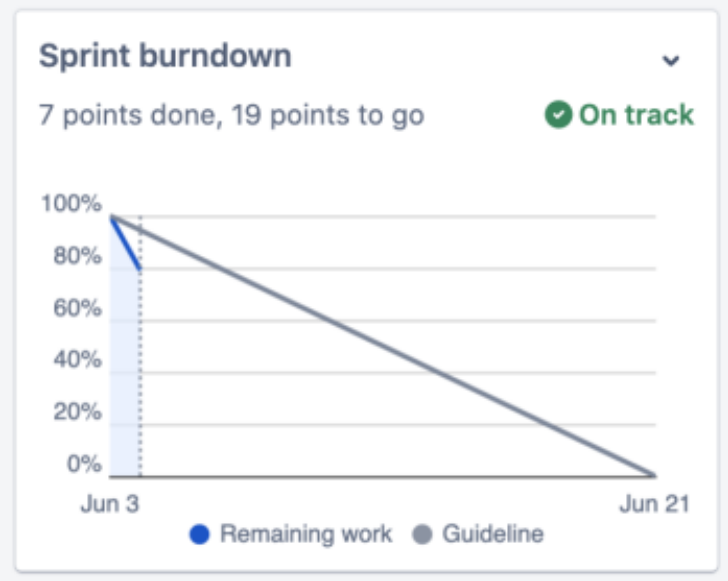
• As a user, I can enter the requirements like bp,wbc,rbc,count,etc. As a user, I will agree to process these information in model which is in cloud.

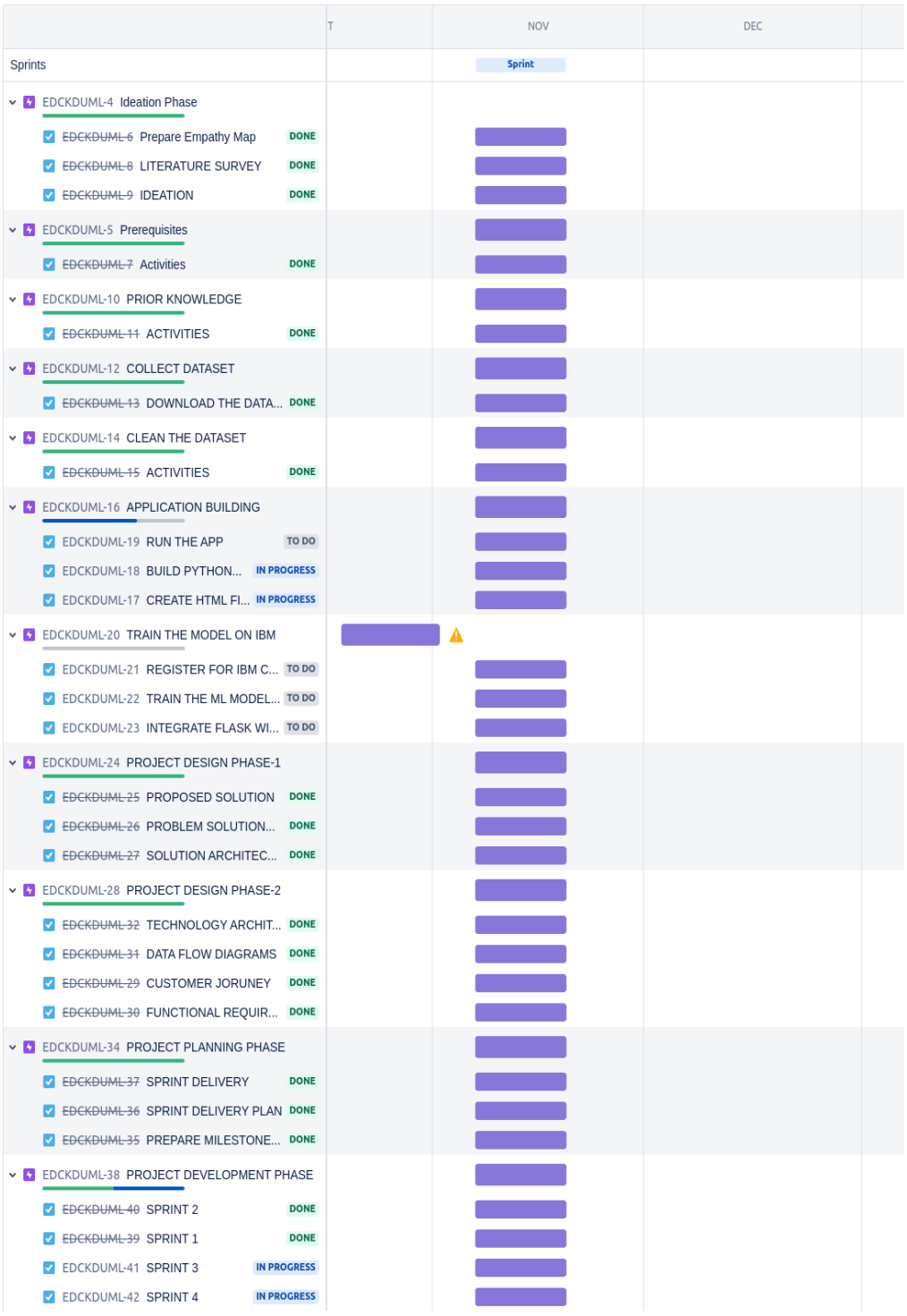
• As a developer i must create the machine learning model in ibm cloud and to connect the through API.

• As a Developer I must create html page for the user to enter their respective requirement values and flask integration.

• As a user, I must verify the predicting results in webpage. As a Developer i must push the webpage in IBM Cloud in order to use by the customer and publish the release version .

6.2 Reports from JIRA




T
NOV
DEC
Sprints
Sprint
EDCKDUML-4 Ideation Phase
~~EDCKDUML-6~~ Prepare Empathy Map DONE
~~EDCKDUML-8~~ LITERATURE SURVEY DONE
~~EDCKDUML-9~~ IDEATION DONE
EDCKDUML-5 Prerequisites
~~EDCKDUML-7~~ Activities DONE
EDCKDUML-10 PRIOR KNOWLEDGE
~~EDCKDUML-11~~ ACTIVITIES DONE
EDCKDUML-12 COLLECT DATASET
~~EDCKDUML-13~~ DOWNLOAD THE DATA... DONE
EDCKDUML-14 CLEAN THE DATASET
~~EDCKDUML-15~~ ACTIVITIES DONE
EDCKDUML-16 APPLICATION BUILDING
EDCKDUML-19 RUN THE APP TO DO
EDCKDUML-18 BUILD PYTHON... IN PROGRESS
EDCKDUML-17 CREATE HTML FI... IN PROGRESS
EDCKDUML-20 TRAIN THE MODEL ON IBM
EDCKDUML-21 REGISTER FOR IBM C... TO DO
EDCKDUML-22 TRAIN THE ML MODEL... TO DO
EDCKDUML-23 INTEGRATE FLASK WI... TO DO
EDCKDUML-24 PROJECT DESIGN PHASE-1
~~EDCKDUML-25~~ PROPOSED SOLUTION DONE
~~EDCKDUML-26~~ PROBLEM SOLUTION... DONE
~~EDCKDUML-27~~ SOLUTION ARCHITEC... DONE
EDCKDUML-28 PROJECT DESIGN PHASE-2
~~EDCKDUML-32~~ TECHNOLOGY ARCHIT... DONE
~~EDCKDUML-31~~ DATA FLOW DIAGRAMS DONE
~~EDCKDUML-29~~ CUSTOMER JORUNEY DONE
~~EDCKDUML-30~~ FUNCTIONAL REQUIR... DONE
EDCKDUML-34 PROJECT PLANNING PHASE
~~EDCKDUML-37~~ SPRINT DELIVERY DONE
~~EDCKDUML-36~~ SPRINT DELIVERY PLAN DONE
~~EDCKDUML-35~~ PREPARE MILESTONE... DONE
EDCKDUML-38 PROJECT DEVELOPMENT PHASE
~~EDCKDUML-40~~ SPRINT 2 DONE
~~EDCKDUML-39~~ SPRINT 1 DONE
EDCKDUML-41 SPRINT 3 IN PROGRESS
EDCKDUML-42 SPRINT 4 IN PROGRESS

7.CODING & SOLUTIONING

7.1 Feature 1

In general, CKD regular testing takes longer time for the results so that diagnosis and treatment get late.In our project the model predicts the disease quickly.

7.2 Feature 2

As the model is deployed in IBM Cloud, wide range of people and hospitals get benefited.

8.TESTING

8.1 Test Cases

Test Scenario	Expected Result
Verify the UI elements	UI works fine
Verify the button elements working	Home page loads
Verify the button elements working	Predict page loads
Verify the UI elements	Shows the description about the project and the details of team members
Verify the predict button working	Predict page loads
Verify the UI elements	Show the Input boxes for the user to enter their inputs
verify the working of prediction page	Predict page works fine
verify the UI elements	Result page works fine

8.2 User Acceptance Testing

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

This report showsthe number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	5	0	0	0	5
Duplicate	0	0	7	0	7
External	0	5	0	0	5
Fixed	2	0	0	5	7
Not Reproduced	0	7	0	0	7
Skipped	5	0	0	0	5
Won't Fix	6	0	0	8	14
Totals	17	12	7	13	50

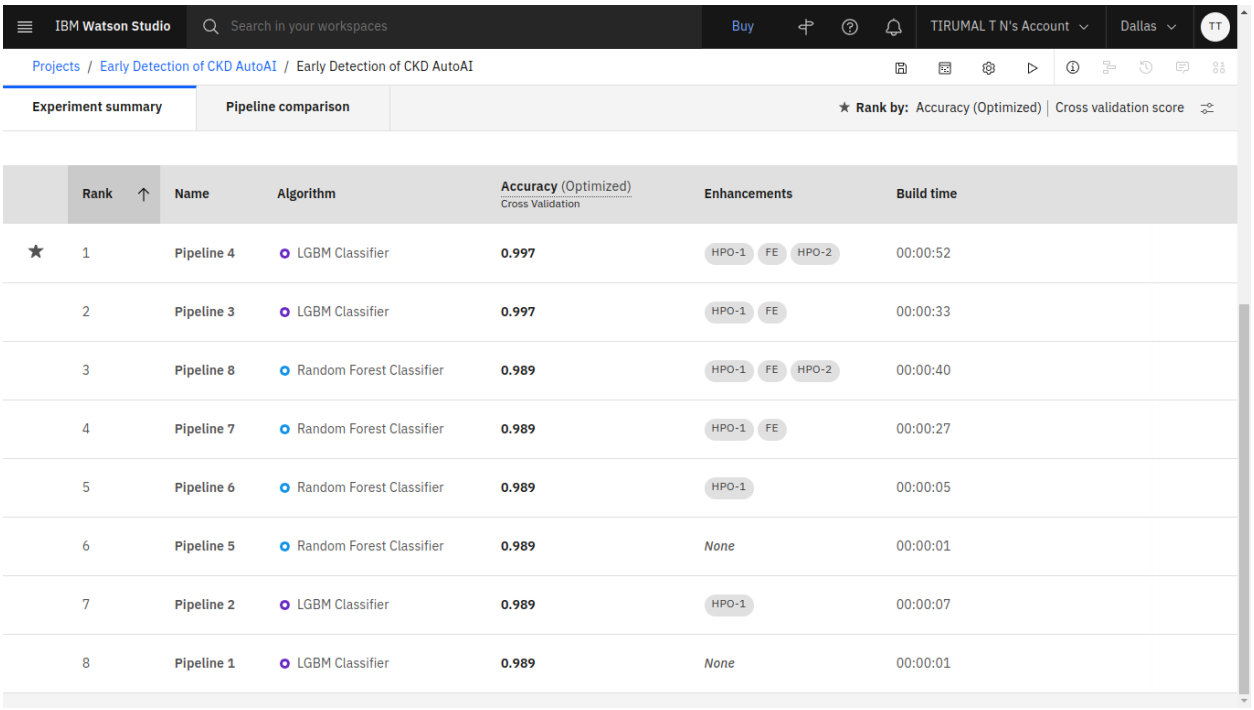
3. Test Case Analysis

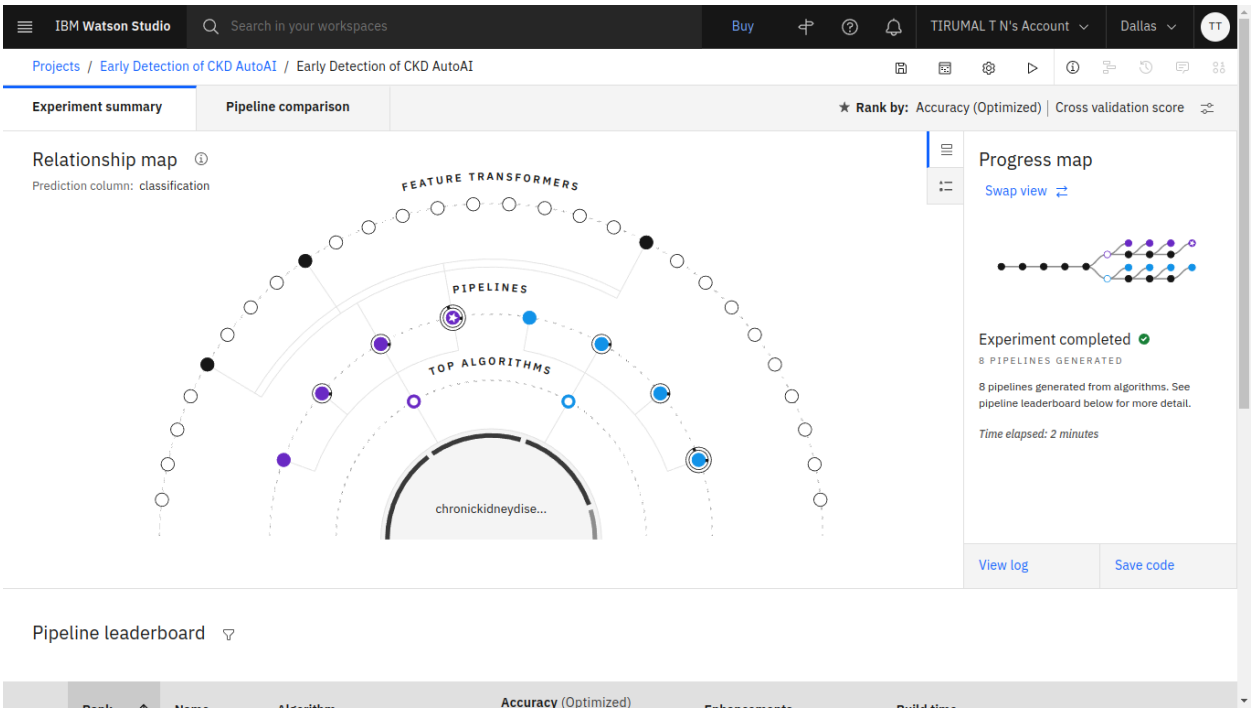
This report shows the number of test cases that have passed, failed,and untested

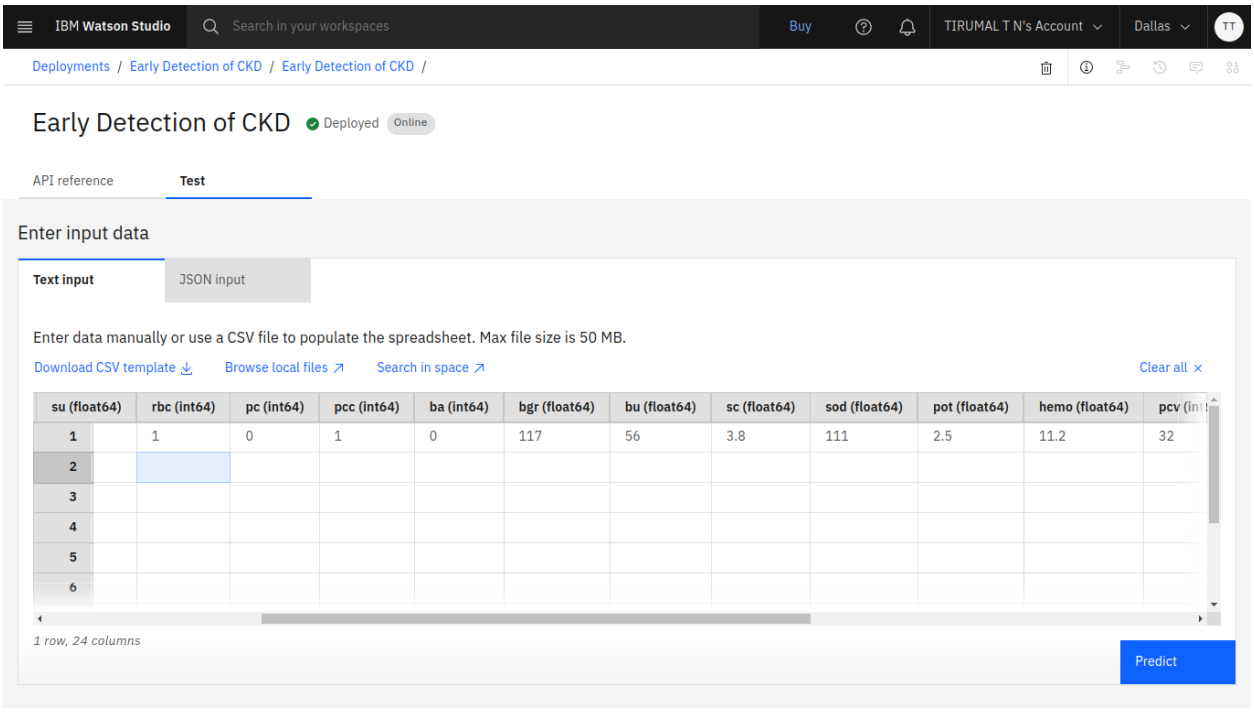
Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

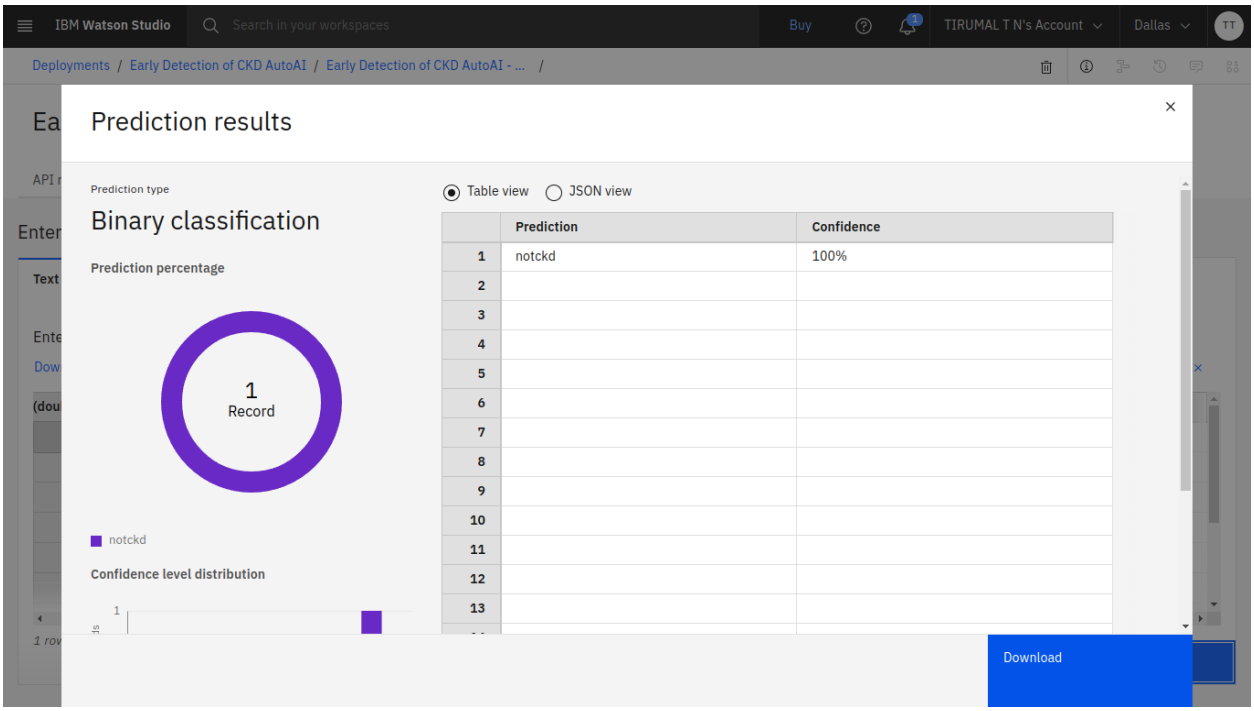
9.RESULTS

9.1 Performance Metrics




IBM Watson Studio
Search in your workspaces
Buy
TIRUMAL T N's Account
Dallas
Projects / Early Detection of CKD AutoAI / Early Detection of CKD AutoAI
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Relationship map
Prediction column: classification
FEATURE TRANSFORMERS
PIPELINES
TOP ALGORITHMS
chronickidneydise...
Progress map
Swap view
Experiment completed
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 2 minutes
View log
Save code
Pipeline leaderboard
Accuracy (Optimized)


IBM Watson Studio
Search in your workspaces
Buy
TIRUMAL T N's Account
Dallas
Deployments / Early Detection of CKD / Early Detection of CKD /
Early Detection of CKD
Deployed
Online
API reference
Test
Enter input data
Text input
JSON input
Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.
Download CSV template
Browse local files
Search in space
Clear all
su (float64) rbc (int64) pc (int64) pcc (int64) ba (int64) bgr (float64) bu (float64) sc (float64) sod (float64) pot (float64) hemo (float64) pcv
1 1 0 1 0 117 56 3.8 111 2.5 11.2 32
2
3
4
5
6
1 row, 24 columns
Predict



10.ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- No need to worry about infrastructure,platform and software.
- Initial setup and maintenance cost is reduced
- All under single umbrella, i.e., everything comes under single framework
- Improves health care services
- Reduces medical supervision procedures
- Ease of maintaining EHR (Electronic Health Care) records
- Proactively preparing for upcoming population health trends
- Acquiring new patients through personalized campaigns
- Reducing costs on appointment no show and readmission penalties

DISADVANTAGES

- Internet Connectivity is mandatory
- More steps to remember during creation of different services
- User interface creation is little bit tougher.

11.CONCLUSION

This study developed an algorithm for predicting CKD at an early stage. The dataset contains input parameters obtained from CKD patients, and the models are trained and validated using the valid parameters. To diagnose CKD, decision tree, random forest, and support vector machine learning models are built.

12.Future Enhancements

SMS/Email module

In the proposed system, admin assigns Id and password for doctors and receptionists and is intimated manually, so we can add SMS/Email module as a future enhancement where doctors and receptionists receive an SMS or Email regarding the Id and password.

Query module

We can add the query module as a future enhancement to the application where doctor, receptionist and admin of the application can interact with each other.

13.APPENDIX

SOURCE CODE:

APP.PY

```
@app.route('/')
def indexPage():
    return render_template('index.html')
@app.route('/home')
def homePage():
    return render_template("home.html")
@app.route('/prediction')
def predict_test():
    return render_template('predict.html')
@app.route('/predict',methods=['post'])
def predict():
    age= request.form["age"]
    bp= request.form["bp"]
    sg= request.form["sg"]
    al= request.form["al"]
    su= request.form["su"]
    rbc= request.form["rbc"]
    pc= request.form["pc"]
    pcc= request.form["pcc"]
    ba= request.form["ba"]
    bgr= request.form["bgr"]
    bu= request.form["bu"]
    sc= request.form["sc"]
    sod= request.form["sod"]
    pot= request.form["pot"]
    hemo= request.form["hemo"]
    pcv= request.form["pcv"]
    wbcc= request.form["wbcc"]
    rbcc= request.form["rbcc"]
    htn= request.form["htn"]
    dm= request.form["dm"]
    cad= request.form["cad"]
    appet= request.form["appet"]
```

```
        pe= request.form["pe"]
        ane= request.form["ane"]
        values = [[int(age), int(bp), float(sg), int(al), int(su), int(rbc), int(pc), int(pcc), int(ba), int(bgr), int(bu), float(sc), int(sod), float(pot), float(hemo), int(pcv), int(wbcc), float(rbcc), int(htn), int(dm), int(cad), int(appet), int(pe), int(ane)]]
        print(values)
        output=model.predict(values)
        print(output)
        if(output[0]==0): return render_template("failure.html")#output="CKD"
        else: return render_template("success.html")#output="No CKD"

```

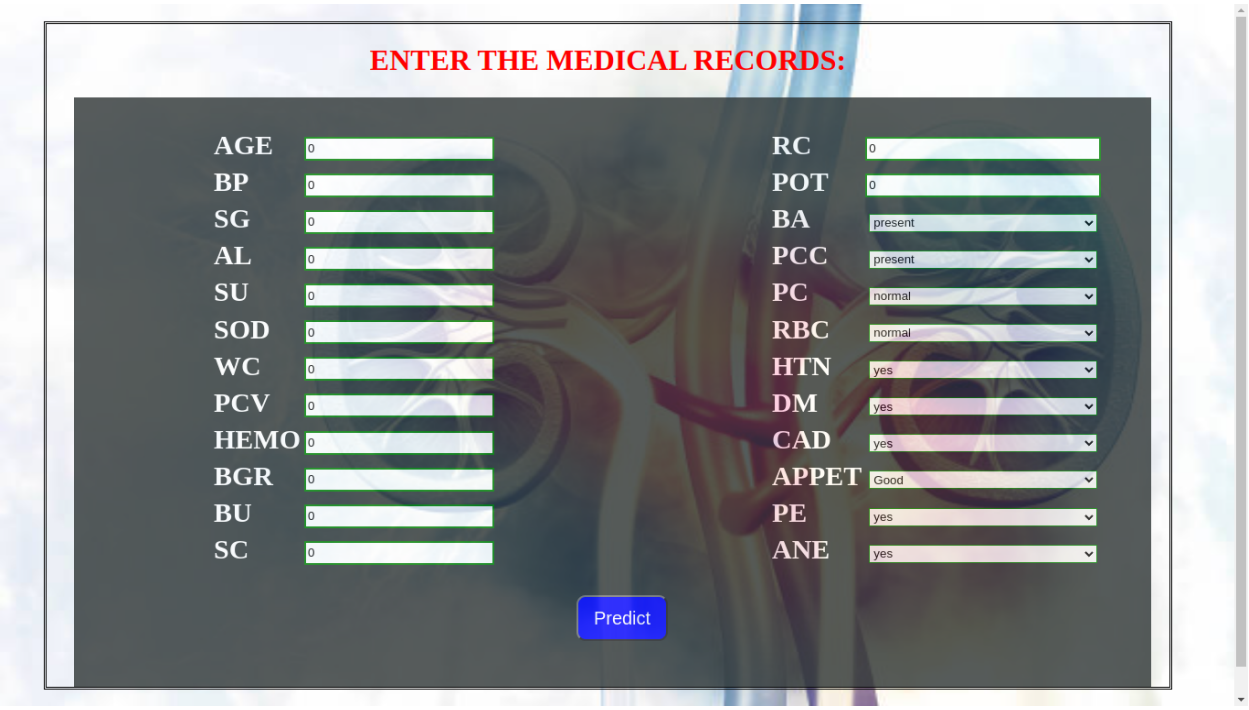
PREDICT.HTML

```
<!DOCTYPE html>
<html>
<head>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='css/predict.css')}}">
  <title>Table test</title>
</head>
<body>
  <form action='/predict' method="post">
    <div class ="table-border">
      <h1 class="alert"><center><b>ENTER THE MEDICAL RECORDS:</b></center></h1>
      <div class="kutty-table">
    <table >
        <tr>
<th class='tab'> AGE </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= age required > </th>
<th class='tab2'> RC </td><th> <input class="tab4" type="number" step="0.01" placeholder="Enter the value" name= rbcc required> </th></tr>
<tr>
<th class='tab'> BP </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= bp required > </th>
<th class='tab2'> POT </td>
<th> <input class="tab4" type="number" step="0.01" placeholder="Enter the value" name= pot required> </th>
</tr>
<tr>
<th class='tab'> SG </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= sg required > </th>
<th class='tab2'>
  <label for="ba">BA</label> </td>
<th>   <select name="ba" id="ba" required >
    <option value="">select the option given below </option>
    <option value=1> present </option>
    <option value=0> not present </option>
  </select></th></tr>
<tr>
<th class='tab'> AL </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= al required > </th>
<th class='tab2'>
  <label for="pcc">PCC</label> </td>
<th>   <select name="pcc" id="pcc" >
    <option value="">select the option given below </option>
    <option value=1> present </option>
    <option value=0> not present </option>
  </select></th></tr><tr>
<th class='tab'> SU </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= su required > </th>
<th class='tab2'>
  <label for="pc">PC</label> </td>
<th>   <select name="pc" id="pc" required >
    <option value="">select the option given below </option>
    <option value=1> normal </option>
    <option value=0> abnormal </option>
  </select></th></tr><tr>
  <th class='tab'> SOD </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= sod required> </th>
<th class='tab2'>
  <label for="rbc">RBC</label> </td>
<th>   <select name="rbc" id="rbc" required >
    <option value="">select the option given below </option>
    <option value=1> normal </option>
    <option value=0> abnormal </option>
  </select></th></tr><tr>
<th class='tab'> WC </td>
<th> <input type="number" step="0.01" placeholder="Enter the value" name= wbcc required> </th>
<th class="tab3">
  <label for="htn">HTN</label> </td>
<th >   <select name="htn" id="htn" required >
    <option value="">select the option given below </option>
    <option value=1> yes </option>
    <option value=0> no </option>
```

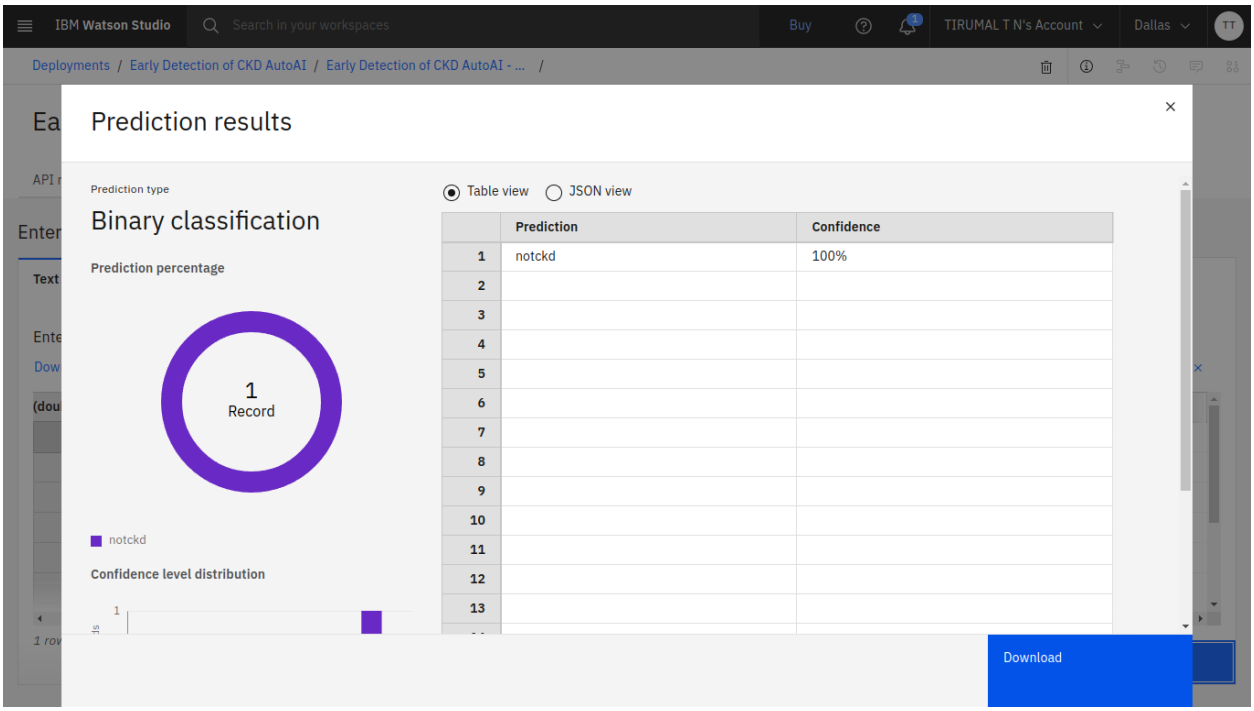
```
    </select>

  </th>
  </tr>
  <tr>
  <th class='tab'> PCV </td>
  <th> <input type="number" step="0.01" placeholder="Enter the value" name= pcv required> </th>
  <th class='tab3'>
    <label for="dm">DM</label> </td>
  <th>

    <select name="dm" id="dm" required >
      <option value="">select the option given below </option>
      <option value=1> yes </option>
      <option value=0> no </option>
    </select>

  </th>
  </tr>
  <tr>
  <th class='tab'> HEMO </td>
  <th> <input type="number" step="0.01" placeholder="Enter the value" name= hemo required> </th>
  <th class='tab3'>
    <label for="cad">CAD</label> </td>
  <th>

    <select name="cad" id="cad" required >
      <option value="">select the option given below </option>
      <option value=1> yes </option>
      <option value=0> no </option>
    </select>

  </th>
  </tr>
  <tr>
  <th class='tab'> BGR </td>
  <th> <input type="number" step="0.01" placeholder="Enter the value" name= bgr required > </th>
  <th class='tab3'>
    <label for="appet">APPET</label> </td>
  <th>

    <select name="appet" id="appet" required >
      <option value="">select the option given below </option>
      <option value=1> Good </option>
      <option value=0> Poor </option>
    </select>

  </th>
  </tr>
  <tr>
  <th class='tab'> BU </td>
  <th> <input type="number" step="0.01" placeholder="Enter the value" name= bu required > </th>
  <th class='tab3'>
    <label for="pe">PE</label> </td>
  <th>

    <select name="pe" id="pe" required >
      <option value="">select the option given below </option>
      <option value=1> yes </option>
      <option value=0> no </option>
    </select>

  </th>
  </tr>
  <tr>
  <th class='tab'> SC </td>
  <th> <input type="number" step="0.01" placeholder="Enter the value" name= sc required > </th>
  <th class='tab3'>
    <label for="ane">ANE</label> </td>
  <th>

    <select name="ane" id="ane" required >
      <option value="">select the option given below </option>
      <option value=1> yes </option>
      <option value=0> no </option>
    </select>
  </th>
  </tr>

      </table>

      <button class="submit-button" type="submit" name="submit">Predict</button>
        </div>
  </div>
```

```
    </form>
</body>
</html>
```

SCREENSHOTS:



SCREENSHOT OF SAMPLE ENTERED VALUES



SCREENSHOT OF OUPUT IN CLOUD AND LOCALHOST

DEMO LINK :

⟹ https://drive.google.com/file/d/14ubo7554wJ_6HI81jds3hXVypTj5o5Wb/view?usp=drivesdk

TEAM ID: PNT2022TMID44392