**JMJ BCA COLLEGE**

CHIPGI, SIRSI-581402

**KARNATAKA UNIVERSITY, DHARWAD**



**PROJECT REPORT**

**ON**

**“Chronic Kidney Disease Prediction using Machine Learning”**

Submitted in partial fulfilment of the requirement for the award of degree

**BACHELOR OF COMPUTER APPLICATION (NEP)**

**Submitted By**

**SAMPRITA NAGAPATI HEGDE, U02KH21S0037**

**Under the Guidance of**

Mr. Xavier Fernandes &

Zetacoding Innovative Solutions, Bengaluru

**Samprita Nagapati Hegde** declare that the undersigned “**Chronic Kidney Disease Prediction using Machine Learning**” report has been prepared under the guidance of Xavier Fernandes. This is being submitted to Karnataka University, Dharwad in partial fulfilment of University regulation for the award of the degree of BCA. We further declare that this report is based in the discussion made with the “**Chronic Kidney Disease Prediction using Machine Learning**”.

Place:JMJ, Chipgi, Sirsi

**SAMPRITA NAGAPATI HEGDE, U02KH21S0037**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER** | **DESCRIPTION** | **PAGE NO.** |
| **CHAPTER 1** | **INTRODUCTION** |  |
|  | Chronic Kidney Disease Overview | 1 |
|  | 1.2 Blood Pressure Overview for Kidney | 2 |
|  | 1.3 Stages of CKD | 3 |
|  | 1.4 Chronic Kidney Disease Tests & Diagnosis | 5 |
|  | 1.5 Managing Chronic Kidney Disease | 8 |
| **CHAPTER 2** | **SYSTEM ANALYSIS** |  |
|  | 2.1 Aim and Objectives | 9 |
|  | 2.2 Scope of the internship | 9 |
|  | 2.3 Existing System | 10 |
|  | 2.4 Proposed System | 10 |
| **CHAPTER 3** | **SYSTEM REQUIREMENT SPECIFICATIONS** |  |
|  | 3.1 Functional Requirements | 12 |
|  | 3.1.1 Jupyter Notebook (Anaconda Navigator) | 13 |
|  | 3.1.2 Python IDLE | 14 |
|  | 3.1.3 Google Colab | 16 |
|  | 3.2 Non-functional Requirements: | 18 |
|  | 3.3 User Input (Dataset) | 18 |
| **CHAPTER 4** | **SYSTEM DESIGN** |  |
|  | 4.1 System Architecture | 20 |
|  | 4.2 Sequence Diagram | 24 |
| **CHAPTER 5** | **SYSTEM IMPLEMENTATION** |  |
|  | 5.1 CKD Implementation | 26 |
|  | 5.2 Dataset Feature Distributions | 30 |
|  | 5.3 Machine Learning Algorithms | 33 |
|  | 5.4 Experimental Results | 40 |
|  | **CONCLUSION** | 42 |
|  | **REFERENCES** | 43 |
|  | **SNAPSHOTS** | 44-50 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **DESCRIPTION** | **PAGE NO.** |
| **1** | Figure 1.1 Healthy Kidney Vs Diseased Kidney | 1 |
| 2 | Figure 1.3: Stages of CKD | 4 |
| 3 | Figure 1.4: GFR results show whether your kidneys are filtering at a normal level. | 6 |
| 4 | Figure 1.5: Urine Test for Albumin. | 7 |
| 5 | Figure 2.1 Chronic Kidney Disease Proposed Diagram | 11 |
| 6 | Figure 3.1: Jupyter Notebook Dashboards | 13 |
| 7 | Figure 3.2: Notebook support for plotting | 14 |
| 8 | Figure 3.3: Python IDLE Download Page | 15 |
| 9 | Figure 3.4: Python IDLE prompt to write and execute code | 15 |
| 10 | Figure 3.5: Welcome page of Google Colab | 16 |
| 11 | Figure 3.6: Upload the Notebook File | 17 |
| 12 | Figure 3.7: Start the Application Page | 17 |
| 13 | Figure 3.8: Start the Application Page | 19 |
| 14 | Figure 4.1: System Architecture diagram | 20 |
| 15 | Figure 4.2: Sequence Diagram | 25 |
| 16 | Figure 5.1: Describing about dataset | 27 |
| 17 | Figure 5.2: Description of CKD Dataset | 28 |
| 18 | Figure 5.3: Dataset | 28 |
| 19 | Figure 5.5: Visualization of the Dataset | 30 |
| 20 | Figure 5.6: Correlation between Features of Dataset | 32 |
| 21 | Figure 5.7: Confusion matrix | 40 |
| 22 | Figure 5.8: Comparison study of models in bar plot | 41 |

**ABSTRACT**

The Project on which I worked during my academic period is “**Chronic Kidney Disease Prediction using Machine Learning**. Machine Learning is used across many ranges around the world. The healthcare industry is no exclusion. Machine Learning can play an essential role in predicting presence/absence of locomotors disorders, CKD diseases and more. Such information, if predicted well in advance, can provide important intuitions to doctors who can then adapt their diagnosis and dealing per patient basis. Lifestyle diseases are common among the population today not only in India but also in almost every country. Lifestyle diseases are caused because of the habits that we have on a day to day basis. The way one lives his life is the major cause of it. It includes heart disease, hypertension, chronic kidney disease and liver disease, etc., which all may hear of.

CKD` disease has been a major life threatening problem in a human being. Due to several reasons such as eating habits, lack of exercise, becoming overweight, and smoking and unhealthy lifestyle habits cause’s ckd diseases. This motivates us to get a clear idea of the risk factors in our life regarding the cause of ckd disease. The proposed method suggests the levels of risk factors according the patients data, so that one can take care of their health properly in order to prevent the CKD disease.

In this project work, proposed methodology is to segregate CKD disease patients based on their risk factors. Firstly applying the feature selection, this is to reduce the records of data and to train the model faster. The four classifiers are SVM, KNN, random forest and ADA Boost Models are applied. After applying the classifiers a proposed model will segregate the CKD disease patients based on their risk factors according to age of the patients. This will be helping for the doctors to analyze the risk factors of the patients. We work on predicting CKD Diseases in people using Machine Learning algorithms. Applied some the data analysis concept to analyse the data and perform the prediction whether patient has CKD or No CKD.