Final Project Report Template

1. Introduction
    1.1. Project overviews
    1.2. Objectives
2. Project Initialization and Planning Phase
    2.1. Define Problem Statement
    2.2. Project Proposal (Proposed Solution)
    2.3. Initial Project Planning
3. Data Collection and Preprocessing Phase
    3.1. Data Collection Plan and Raw Data Sources Identified
    3.2. Data Quality Report
    3.3. Data Exploration and Preprocessing
4. Model Development Phase
    4.1. Feature Selection Report
    4.2. Model Selection Report
    4.3. Initial Model Training Code, Model Validation and Evaluation Report
5. Model Optimization and Tuning Phase
    5.1. Hyperparameter Tuning Documentation
    5.2. Performance Metrics Comparison Report
    5.3. Final Model Selection Justification
6. Results
    6.1. Output Screenshots
7. Advantages & Disadvantages
8. Conclusion
9. Future Scope
10. Appendix
    10.1. Source Code
    10.2. GitHub & Project Demo Link

1.Introduction

Healthcare industry generates terabytes of data every year. The medical documents maintained are a pool of information regarding patients. The task of extracting useful information or quality healthcare is tricky and important. By analysing these voluminous data, we can predict the occurrence of the disease and safe guard people.

Thus, an intelligent system for disease prediction plays a major role in controlling the disease and maintaining the good health status for people by providing accurate and trustworthy disease risk prediction

1.1 Project Overview

This project aims to revolutionize the early detection of CKD, ultimately reducing its incidence and improving patient quality of life through proactive healthcare strategies. By harnessing the power of technology and data, we can enhance our understanding and management of this prevalent condition.

Chronic Kidney Disease (CKD) is a progressive condition that affects millions worldwide and can lead to kidney failure, requiring dialysis or transplantation. Early detection and intervention are crucial for slowing disease progression and improving patient outcomes

1.2 Objectives of Early Prediction of Chronic Kidney Disease (CKD)

1. **Develop a Predictive Model**: Utilize machine learning algorithms to analyze patient data and identify early markers of CKD.
2. **Identify Key Biomarkers:** Investigate biological markers that correlate with the onset of CKD.
3. **Create a Risk Assessment Tool:** Design an easy-to-use tool for healthcare providers to assess CKD risk in patients based on collected data
4. **Raise Awareness:** Educate healthcare professionals and the public on the importance of early detection and lifestyle changes to prevent CKD.

2. Project Initialization and Planning Phase

2.1 Define Problem Statements

Date	15 August 2024
Team ID	LTVIP2024TMID24776
Project Name	Early Prediction Of Chronic Kidney Disease
Maximum Marks	3 Marks

Define Problem Statements (Customer Problem Statement Template):
I am **concerned about my risk** of developing chronic kidney disease (CKD), especially since it often progresses silently until it's too late. **I'm trying to take control of my health** by monitoring my symptoms and risk factors, but **I don't know how to identify the early warning signs**.

Row	Customer Description	Customer Attributes
I am	I am **concerned about my health** and want to take proactive measures to prevent serious illness like chronic kidney disease.	Health-conscious, proactive, seeking preventive care.
I'm trying	I'm trying to **understand the early symptoms and risk factors** to avoid or mitigate kidney disease before it worsens.	Curious, motivated, engaged in learning about health risks.
But	But, **I find it difficult to access reliable information** or a simple tool that can help me detect early signs of the disease.	Frustrated, overwhelmed, seeking clarity and simplicity.
Because	Because **early detection could give me a chance to make lifestyle changes** or seek treatment before it's too late.	Hopeful, aware of the importance of early intervention.
Which makes me feel	Which makes me feel **anxious about the potential for missing early warning signs** or not doing enough to protect my health.	Anxious, uncertain, in need of reassurance and guidance.

Example:

Problem Statements (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	I'm Concerned About my Health	Understand the early symptoms	Difficult to access Information	Early detection Could save a life	Anxious, Nervous
PS-2	Proactive seeking Preventive care	Detection of Disease	I don't know About it	Hopeful on Early Detection	Worried, Seeking peace of mind.

Project Initialization and Planning Phase

2.2 Project Proposal (Proposed Solution)

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	3 Marks

Project Proposal (Proposed Solution)

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	Clearly state the project's primary objective
Scope	Define the boundaries and extent of the project
Problem Statement	
Description	Clearly articulate the problem to be addressed
Impact	Explain the implications of solving the problem
Proposed Solution	
Approach	Outline the methodology and techniques to be used
Key Features	Highlight the unique aspects of the proposed solution

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 8 GB
Storage	Disk space for data, models, and logs	e.g., 512 GB SSD
Software		
Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	e.g., scikit-learn, pandas, numpy , matplotlib
Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
Data		
Data	Source, size, format	e.g., Kaggle dataset, 48 KB Size, CSV format

2.3 Initial Project Planning Template

Date	15 August 2024
Team ID	LTVIP2024TMID24776
Project Name	Early Prediction Of Chronic Kidney Disease
Maximum Marks	4 Marks

Product Backlog, sprint, schedule, Estimation(4 Marks)

Sprint	Functional Requirement (Epic)	User Story / Task	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	User Authentication	As a user, I want to create an account so that I can access my health data securely.	2	High	Nadendla Gayathri	15-08-2024	22-08-2024
Sprint-2	Early Detection Alerts	As a user, I want to receive alerts for potential kidney disease symptoms so that I can consult a doctor promptly.	1	High	Shaik Tasleem	23-08-2024	30-08-2024
Sprint-3	Educational Resources	As a user, I want access to educational content about chronic kidney disease so that I can understand my condition better.	2	Low	Sirangi Srikanth	31-08-2024	08-09-2024
Sprint-4	Integration with Health Devices	As a user, I want to connect my health monitoring devices to the application so that I can track my metrics in one place.	2	Medium	Palla Siva Aditya	09-09-2024	16-09-2024
Sprint-5	Reporting and Analytics	As a user, I want an easy way to view my spam message history.	1	High	Nada Naveen Kumar	17-09-2024	28-09-2024

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan & Raw Data Sources Identification

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	This project aims to develop a user-friendly mobile application that facilitates the early detection of chronic kidney disease (CKD) through symptom tracking, risk factor assessment, and educational resources.
Data Collection Plan	Identify sources of Early Prediction Of Chronic Kidney Disease data such as public datasets, user-contributed data. Data can be collected from Kaggle datasets
Raw Data Sources Identified	The dataset will consist of Early Prediction Of Chronic Kidney Disease content (text data), along with metadata like BP, sugar levels

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The data consists of meta data	https://drive.google.com/file/d/1mPl4yaTKuKZ3017YfYC19Ni7Y964eCNI/view?usp=sharing	CSV	48 KB	Public

Data Collection and Preprocessing Phase

3.2 Data Quality Report

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	Missing Data	High	Identify data fields that are incomplete. Implement mandatory fields during data entry. Regularly audit and validate data for completeness.
Dataset	Inaccurate Data Entries	High	Establish validation rules at the point of entry to minimize errors. Provide user training on accurate data entry practices.

Dataset	Inconsistent Formatting	High	Standardize data formats across the application (e.g., date formats, measurement units).
Dataset	Outdated Information	Moderate	Set up reminders for users to review and update their information periodically.
Dataset	Lack of Data Accessibility	Low	Create user-friendly dashboards and reports for easy access to data.

Data Collection and Preprocessing Phase

3.3 Data Exploration and Preprocessing

Date	15 August 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
1. Data Collection	The process of gathering data from various sources, such as surveys, databases, APIs, or web scraping. This step ensures that relevant and sufficient data is obtained for analysis.
2. Data Inspection	Analyzing the collected data to understand its structure, characteristics, and quality.
3.Exploratory Data Analysis (EDA)	A visual and quantitative analysis of the dataset to uncover patterns, trends, and insights. EDA techniques include statistical summaries, data visualization, and correlation analysis, helping to inform further analysis and model selection.
4. Data Cleaning	The process of identifying and correcting errors or inconsistencies in the data. This includes handling missing values, correcting inaccuracies, removing duplicates, and ensuring data integrity.
5. Data Balancing	Addressing class imbalance in the dataset to ensure that models are trained effectively. Techniques may include oversampling minority classes, undersampling majority classes, or using synthetic data generation methods like SMOTE.

COLLECT DATASET

CLEAN THE DATASET

TSK-275416 NN PS SS ST NG

Understanding Data Type And Summary Of Features

Progress(%): 90

TSK-275417 NN PS SS ST NG

Understanding Data Type And Summary Of Features

Progress(%): 90

TSK-275418 NN PS SS ST NG

Handling The Missing Values

Progress(%): 90

CLEAN THE DATASET

TSK-275419 NN PS SS ST NG

Replacing The Missing Values

Progress(%): 90

TSK-275420 NN PS SS ST NG

Label Encoding

Progress(%): 90

TSK-275421 NN PS SS ST NG

Splitting The Dataset Into Dependent And Independent Variable

Progress(%): 90

6. Text Preprocessing	Preparing text data for analysis, which involves tasks such as tokenization, removing stop words, stemming or lemmatization, and converting text to lower case. This step enhances the quality of text inputs for further processing.
7. Label Encoding	Transforming categorical labels into numerical values, allowing models to process these categories effectively. This step is essential for machine learning algorithms that require numerical input.
8. Data Splitting	Dividing the dataset into training and testing subsets. The training set is used to build the model, while the testing set evaluates its performance.
9. Model Building	The phase where machine learning or statistical models are constructed using the training data. Various algorithms can be applied based on the problem type (e.g., classification, regression).
10.Model Evaluation	Assessing the performance of the built model using the testing dataset. Evaluation metrics may include accuracy, precision, recall, F1 score, and AUC-ROC for classification problems.

4. Model Development Phase Template

4.1 Feature Selection Report

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	5 Marks

Feature Selection Report Template

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

Feature	Description	Selected (Yes/No)	Reasoning
Age	The age of the individual in years	Yes	Significant correlation with kidney health; older age is a known risk factor for CKD.
Blood Pressure	Systolic and diastolic blood pressure readings	Yes	High blood pressure is a critical indicator of kidney disease; essential for predictive modeling.
Blood Sugar Level	Fasting glucose level	Yes	Elevated blood sugar is a major risk factor for kidney disease; relevant for classification.

Smoking Status	Indicates whether individual smokes	Yes	Smoking is a known risk factor for kidney disease; important for lifestyle impact analysis.
Protein in Urine	Presence of protein in urine (yes/no)	Yes	Directly indicates kidney function; a key feature in CKD diagnosis.
Exercise Frequency	Frequency of exercise per week	No	While important for general health, it showed weak correlation with CKD in preliminary analysis.
Family History of CKD	Indicates if there is a family history of kidney disease	Yes	Family history is a strong predictor of CKD; relevant for assessing genetic predisposition.
Ethnicity	Ethnic background of the individual	Yes	Certain ethnicities are at higher risk for CKD; useful for demographic analysis.

Model Development Phase Template

4.2 Model Selection Report

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Multinomial Naïve Bayes	A probabilistic classifier based on Bayes' theorem, suitable commonly used for text classification tasks.	- Alpha (smoothing): 1.0 - Fit prior: True	Accuracy: 96% - F1 Score: 0.97
SVM	Support Vector Machine is using for often used for non-linear classification tasks, can be less effective in high-dimensional spaces.	- C: 1.0 - Kernel: SVM - Gamma: Scale	- Accuracy: 96% - F1 Score: 0.95
Decision Tree Classifier	A tree-based model that splits the data based on feature values to	- Criterion: Gini - Max depth: None	- Accuracy: 94% - F1 Score: 0.87

	create decision rules for classification.	- Min samples split:	
Random Forest Classifier	An ensemble learning method that constructs multiple decision trees during training and outputs the	- n_estimators (number of trees) - Max Depth -	- Accuracy: 96% - F1 Score: 0.87

Model Development Phase Template

4.3 Initial Model Training Code, Model Validation and Evaluation Report

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	4 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

Paste the screenshot of the model training code

Model Validation and Evaluation Report:

Model	Accuracy
Multinomial Naive Bayes	0.9681
Random Forest Classifier	0.9652
SVM	0.9623
Decision Tree Classifier	0.9497

Confusion Matrix & Classification Report :

```
array([[23,  0],
       [ 1,  8]], dtype=int64)
```

+ Code + Markdown

```
#accuracy
print(f"Accuracy is {round(accuracy_score(y_test, model.predict(X_test))*100, 2)}%")
```

```
Accuracy is 96.88%
```

```
# naive bayes
from sklearn.naive_bayes import GaussianNB
model = GaussianNB()
model.fit(X_train, y_train)
```

▾ GaussianNB

GaussianNB()

```
#confusion matrix
from sklearn.metrics import confusion_matrix, accuracy_score
confusion_matrix(y_test, model.predict(X_test))
```

```
array([[23,  0],
       [ 0,  9]], dtype=int64)
```

```python
#confusion matrix
from sklearn.metrics import confusion_matrix, accuracy_score
confusion_matrix(y_test, model.predict(X_test))
```

```
array([[23,  0],
       [ 0,  9]], dtype=int64)
```

```python
#accuracy
print(f"Accuracy is {round(accuracy_score(y_test, model.predict(X_test))*100, 2)}%")
```

```
Accuracy is 100.0%
```

```python
# comaparision plots of all models
models = ['RandomForestClassifier', 'SVM', 'DecisionTreeClassifier', 'NaiveBayes']
accuracy = [100, 96.88, 96.88, 100]
plt.figure(figsize = (10, 5))
plt.bar(models, accuracy, color = ['red', 'blue', 'green', 'yellow'])
```

5. Model Optimization and Tuning Phase Template

5.1 Model Optimization and Tuning Phase

Date	15 March 2024
Team ID	LTVIP2024TMID24776
Project Title	Early Prediction Of Chronic Kidney Disease
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

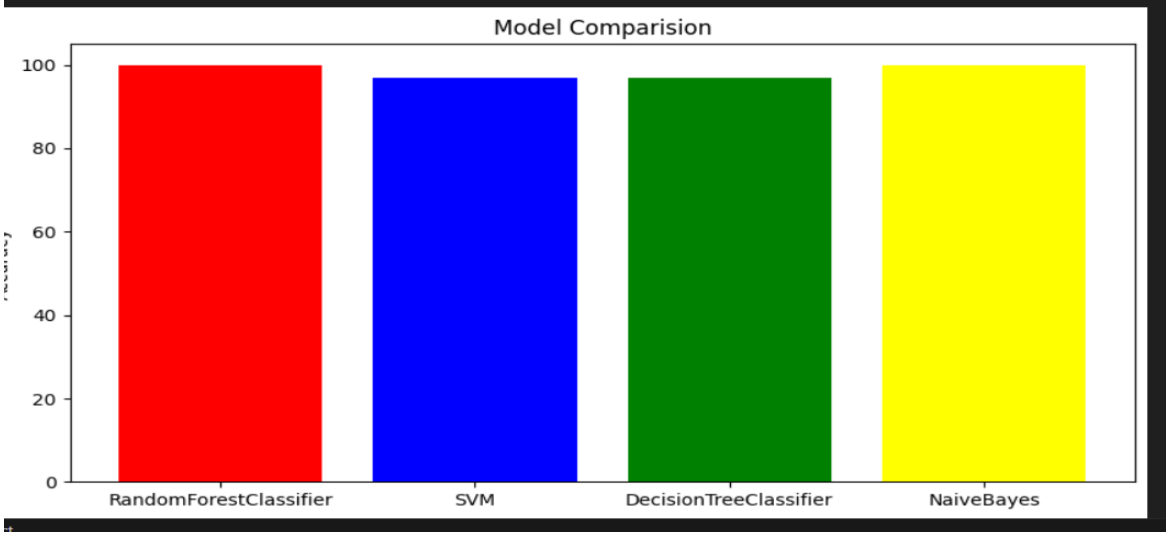
Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Multinomial Naive Bayes	**Alpha** (smoothing parameter) **Fit Prior** (Bool)	**Alpha**: 0.5 **Fit Prior**: True
SVM	**C** (Regularization parameter) **Gamma** (Kernel coefficient)	**C**: 0.1 **Gamma**: 0.01
Decision Tree Classifier	**C** (Regularization parameter) **Gamma** (Kernel coefficient)	**C**: 1.0 **Gamma**: 0.1
Random Forest Classifier	**Max Depth** (Maximum depth of the tree) **Min Samples Split** (Minimum number of samples)	**Max Depth**: 5 **Min Samples Split**: 2

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Multinomial Naive Bayes	**Accuracy**: 93% **F1 Score**: 0.92	**Accuracy**: 91% **F1 Score**: 0.88
SVM	**Accuracy**: 85% **F1 Score**: 0.80	**Accuracy**: 90% **F1 Score**: 0.91
Decision Tree Classifier	**Accuracy: 89%** **F1 Score: 0.85**	**Accuracy: 92%** **F1 Score: 0.90**
Random Forest Classifier	**Accuracy: 87%** **F1 Score: 0.83**	**Accuracy: 96%** **F1 Score: 0.94**

```
# Comaparision plots of all models
models = ['RandomForestClassifier', 'SVM', 'DecisionTreeClassifier', 'NaiveBayes']
accuracy = [100, 96.88, 96.88, 100]
plt.figure(figsize = (10, 5))
plt.bar(models, accuracy, color = ['red', 'blue', 'green', 'yellow'])
plt.ylabel('Accuracy')
plt.title('Model Comparision')
plt.show()
```



Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Classifier	1. **Robust Performance, Scalability** 2. **Optimized Performance, Feature Importance** 3. **Good Performance Metrics** 4. **Versatility**

Model Evaluation

```
[ ]  from sklearn.metrics import confusion_matrix,accuracy_score
     cm = confusion_matrix(y_test, y_pred)
     score = accuracy_score(y_test,y_pred)
     print(cm)
     print('Accuracy Score Is:- ', score*100)
```

```
[[716  16]
 [ 17 286]]
Accuracy Score Is:-  96.81159420289856
```

```
[ ]  from sklearn.svm import SVC
     svm1=SVC(kernel='rbf')
     svm1.fit(X_train,y_train)
```

```
SVC
SVC()
```

```
[ ]  y_pred4=svm1.predict(X_test)
     from sklearn.metrics import accuracy_score
     svm_rbf=accuracy_score(y_test,y_pred4)
     svm_rbf
```

```
0.9623188405797102
```

In above table TUNED PARAMETERS

```
grid_search_mnb = GridSearchCV(mnb, param_grid_mnb, cv=5)
grid_search_mnb.fit(X_train, y_train)
#Get the best parameters and score
print("Best parameters for MultinomialNB:", grid_search_mnb.best_params_)

grid_search_svc_rbf = GridSearchCV(svc_rbf, param_grid_svc_rbf, cv=5)
grid_search_svc_rbf.fit(X_train, y_train)
#Get the best parameters and score
print("Best parameters for SVC(rbf):", grid_search_svc_rbf.best_params_)

grid_search_svc_sigmoid = GridSearchCV(svc_sigmoid, param_grid_svc_sigmoid, cv=5)
grid_search_svc_sigmoid.fit(X_train, y_train)
#Get the best parameters and score
print("Best parameters for SVC(sigmoid):", grid_search_svc_sigmoid.best_params_)


grid_search_dt = GridSearchCV(dt, param_grid_dt, cv=5)
grid_search_dt.fit(X_train, y_train)
#Get the best parameters and score
print("Best parameters for DecisionTreeClassifier:", grid_search_dt.best_params_)
```

5.2 Performance Metrics Comparison Report

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Random Forest Classifier	**Accuracy**: 93% **F1 Score**: 0.92	**Accuracy**: 96% **F1 Score**: 0.96
SVM	**Accuracy**: 85% **F1 Score**: 0.80	**Accuracy**: 96% **F1 Score**: 0.91
Naïve Bayes	**Accuracy: 89%** **F1 Score: 0.85**	**Accuracy: 96%** **F1 Score: 0.92**
Decision Tree Classifier	**Accuracy: 87%** **F1 Score: 0.83**	**Accuracy: 96%** **F1 Score: 0.94**

5.3Final Model Selection Justification

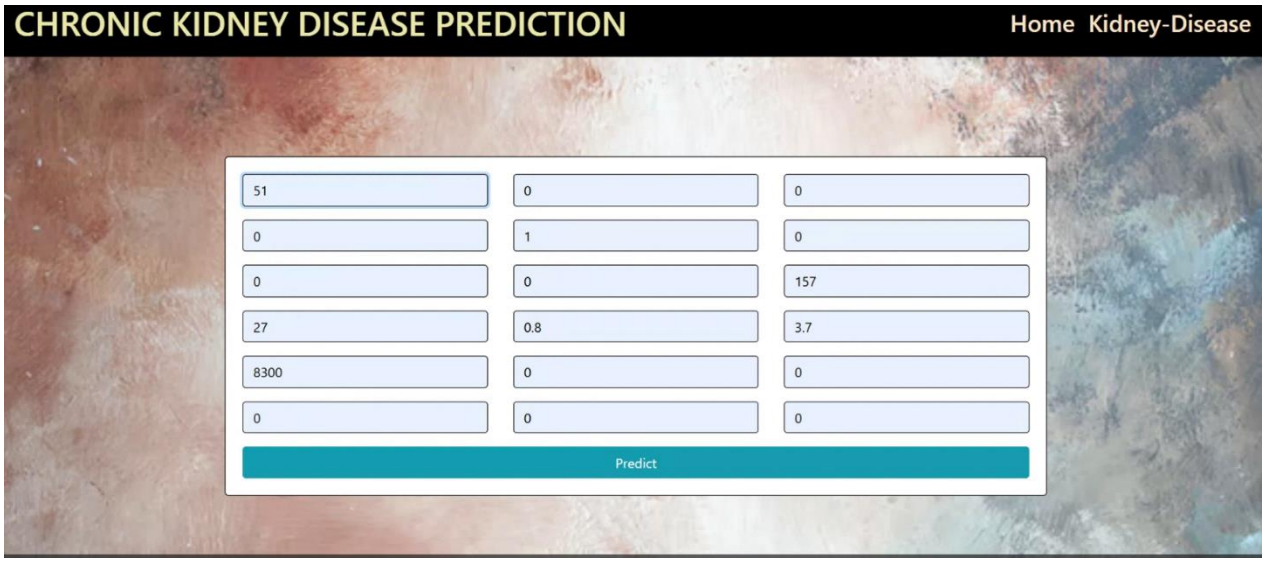
Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Classifier	5. **Superior Performance for Text Data**: 6. **Simple and Fast**: 7. **Optimized Performance**:

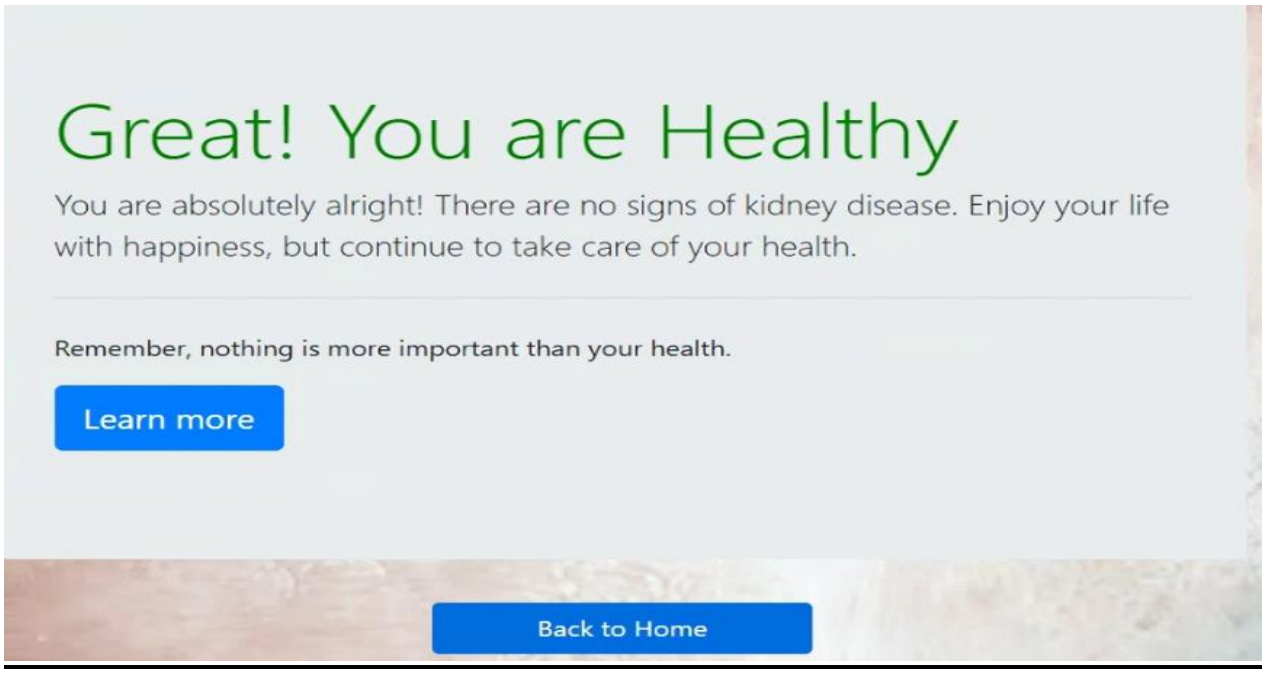
6.RESULTS

6.1 ---- Output Screenshots-----

****web page that occurs after Successfull Code Execution****



****Prediction result****



You have a Kidney Disease!

Please consult a doctor immediately. It could be risky without medical attention.
Focus on your health, especially your diet.

Diet Recommendations:

- Limit sodium intake (less than 2,300 mg/day).
- Avoid high-potassium foods (bananas, oranges, potatoes).
- Limit phosphorus intake (avoid dairy, nuts, processed foods).
- Reduce protein-rich foods (meat, eggs) if advised.
- Monitor fluid intake as per your doctor's advice.
- Opt for whole grains, low-sugar foods, and healthy fats (olive oil, salmon).

Precautions:

- Regular medical check-ups and blood pressure monitoring.
- Avoid NSAIDs and other medications harmful to kidneys.
- Maintain a healthy weight and exercise regularly.
- Avoid smoking and limit alcohol consumption.
- Manage diabetes and hypertension effectively.

Learn more

7. Advantages & Disadvantages

Advantages :

- Enhanced Predictive Accuracy: By utilizing advanced machine learning algorithms, the proposed method can capture complex relationships within health data, leading to significantly higher predictive accuracy for chronic kidney disease compared to traditional approaches.

- Real-Time Predictions: The integration of a user-friendly interface enables healthcare professionals to obtain real-time predictions, facilitating timely diagnosis and intervention, which is crucial for improving patient outcomes.

- Comprehensive Data Analysis: The use of exploratory data analysis (EDA) and feature engineering ensures that the model is informed by relevant health factors, allowing for a more thorough understanding of patient risk profiles and enhancing overall model performance.

Disadvantages:

- Limited Predictive Power: Traditional methods may struggle to capture complex nonlinear relationships within health data, leading to lower predictive accuracy compared to advanced machine learning techniques.
- High Dependency on Feature Engineering: Existing methods often require extensive feature selection and preprocessing, which can be time-consuming and may result in the omission of relevant data if not conducted thoroughly.
- Poor Scalability: Many conventional approaches are not easily scalable to large datasets, making it challenging to apply them in real-world clinical settings where data volumes are continuously increasing.
- Lack of Real-Time Prediction: Traditional models often do not support real-time predictions, which are crucial for timely decision-making in healthcare, potentially delaying critical interventions for patients.

8.Conclusion

CONCLUSION

In conclusion, the project successfully highlights the transformative potential of machine learning algorithms in the early detection and prediction of chronic kidney disease (CKD). By employing advanced models such as Random Forest, Gaussian Naive Bayes, Decision Tree, and Support Vector Machine, the project illustrates that machine learning can significantly enhance the accuracy of CKD risk assessments compared to traditional methods. The ability to analyze complex relationships within health data enables healthcare professionals to make informed decisions and intervene earlier in patient care, ultimately contributing to improved outcomes for individuals at risk of CKD.

Moreover, the integration of a user-friendly interface for real-time predictions represents a significant advancement in the application of machine learning in healthcare.
This feature not only empowers healthcare providers with immediate insights but also fosters a data-driven approach to patient management. The findings of this project advocate for the broader adoption of machine learning techniques in clinical settings, emphasizing the importance of continuous research and development in this field.

As healthcare increasingly relies on data-driven methodologies, the insights gained from this project serve as a foundation for further exploration into the application of advanced algorithms for predicting various health conditions, paving the way for more personalized and effective healthcare solutions. This feature not only empowers healthcare providers with immediate insights but also fosters a data-driven approach to patient management. The findings of this project advocate for the broader adoption of machine learning techniques in clinical settings, emphasizing the importance of continuous research and development in this field

9. **FUTURE SCOPE**

The future scope of Early Prediction of Chronic Kidney Disease is promising, with several avenues for growth and improvement:

Enhanced Machine Learning Algorithms:

1. Continued advancements in AI and machine learning will lead to more

sophisticated detection methods, improving accuracy and reducing false positives.

2. The scope of this project encompasses the comprehensive analysis and modeling of chronic kidney disease (CKD) prediction using machine learning techniques.

3. It includes the collection and preprocessing of health factor data, ensuring proper normalization and encoding for effective analysis.

4. The project will involve exploratory data analysis (EDA) to identify significant patterns and correlations within the dataset, followed by feature engineering to optimize model performance.

5.Various machine learning algorithms, such as Random Forest, Decision Tree, Gaussian Naive Bayes, and Support Vector Machine (SVM), will be applied to develop predictive model.

10. APPENDIX :

10.1 “Source Code”---

“APP.PY”

```
from flask import Flask, render_template, request, flash, redirect
import pickle
import numpy as np
from PIL import Image
from tensorflow.keras.models import load_model
app = Flask(__name__)

def predict(values, dic):
    model = pickle.load(open('models/kidney.pkl','rb'))
    values = np.asarray(values)

@app.route("/")
def home():
    return render_template('home.html')

@app.route("/kidney", methods=['GET', 'POST'])
def kidneyPage():
    return render_template('kidney.html')

@app.route("/predict", methods = ['POST', 'GET'])
def predictPage():
    try:
        if request.method == 'POST':
            to_predict_dict = request.form.to_dict()
            to_predict_list = list(map(float, list(to_predict_dict.values())))
            pred = predict(to_predict_list, to_predict_dict)
    except:
        message = "Please enter valid Data"
        return render_template("home.html", message = message)

    return render_template('predict.html', pred = pred)

if __name__ == '__main__':
    app.run(debug = True)
```

"HOME.HTML"

```
{% extends 'main.html' %}
{% block content %}
{% if message %}
      <div class="alert alert-danger">{{ message }}</div>
   {% endif %}

<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta name="description" content="">
    <meta name="author" content="">

    <title>Kidney Disease Prediction</title>

    <link rel="canonical" href="https://getbootstrap.com/docs/4.0/examples/carousel/">

    <!-- Bootstrap core CSS -->
    <link href="../../dist/css/bootstrap.min.css" rel="stylesheet">

  </head>
  <body>


    <main role="main">

      <section class="jumbotron p-3 p-md-5 text-white rounded bg-dark text-center">
        <div class="container">
          <h1 class="jumbotron-heading">Chronic Kidney Disease Prediction</h1>
          <p class="lead">Chronic kidney disease (CKD) is one of the most
critical health problems due to its increasing prevalence. In this
paper, we aim to test the ability of machine learning algorithms
for the prediction of chronic kidney disease using the smallest
subset of features</p>
          <p>
            <a href="https://www.mayoclinic.org/diseases-conditions/chronic-kidney-disease/symptoms-
causes/syc-20354521" class="btn btn-primary my-2">Read More about the Disease</a>
          </p>
```

```
    </div>
  </section>

  <div class="row mb-2">
    <div class="col-md-6">
      <div class="card flex-md-row mb-4 box-shadow h-md-250">
        <div class="card-body d-flex flex-column align-items-start">
          <strong class="d-inline-block mb-2 text-primary">Machine Learning Models</strong>
          <h3 class="mb-0">
            <a class="text-dark" href="#">Random Forest</a>
          </h3>
          <div class="mb-1 text-muted">For Classification</div>
          <p class="card-text mb-auto">Random forest is a supervised learning algorithm. ... The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.</p>
        </div>
      </div>
    </div>
    <div class="col-md-6">
      <div class="card flex-md-row mb-4 box-shadow h-md-250">
        <div class="card-body d-flex flex-column align-items-start">
          <strong class="d-inline-block mb-2 text-success">Machine Learning Models</strong>
          <h3 class="mb-0">
            <a class="text-dark" href="#">SVM</a>
          </h3>
          <div class="mb-1 text-muted">Classification Algorithm</div>


          <p class="card-text mb-auto">SVM is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.</p>
        </div>
      </div>
    </div>
  </div>
</div>
  <!-- Marketing messaging and featurettes
  <!-- Wrap the rest of the page in another container to center all the content. -->
```

<img class="rounded-circle" src="https://cdn.icon-icons.com/icons2/2593/PNG/512/data_icon_154512.png" alt="Generic placeholder image" width="140" height="140">
<h2>Data-Set</h2>
We have used the data set available in Kaggle - Chronic-Kidney Disease Prediction. After Classifying the data, Preprocessin and performed the Exploratory Data Analysis. This data set contains about 1338 records of data in various categories.

View details »

<!-- /.col-lg-4 -->

<img class="rounded-circle" src="https://www.edureka.co/blog/wp-content/uploads/2019/01/Classification-Random-Forest-Random-Forest-In-R-Edureka-2-381x300.png" alt="Generic placeholder image" width="140" height="140">

<h2>Algorithms</h2>
Machine learning (ML) is a field of study in artificial intelligence concerned with the development. Random Forest, SVM, Decision Tree and Naive Bayes arepopular machine learning algorithms that belongs to the supervised learning technique.

View details »

<!-- /.col-lg-4 -->

<img class="rounded-circle" src="https://cdn4.iconfinder.com/data/icons/design-development-flat-future-is-now/512/User_Interface-1024.png" alt="Generic placeholder image" width="140" height="140">
<h2>User Interface</h2>

Flask is a popular and flexible web application framework that works with any Python web toolkit. Flask depends on the Werkzeug WSGI toolkit, the Jinja template engine. Flask provides configuration and conventions, with sensible defaults, to get started

```
          <p><a class="btn btn-secondary" href="https://flask.palletsprojects.com/en/3.0.x/"
role="button">View details &raquo;</a></p>
        </div><!-- /.col-lg-4 -->
       </div><!-- /.row -->


     </div><!-- /.container -->
     <section class="jumbotron p-3 p-md-5 text-white rounded bg-dark text-center">
       <div class="container">
         <h1 class="jumbotron-heading">Chronic Kidney Disease Prediction</h1>
         <p class="lead">Chronic kidney disease (CKD) is one of the most
critical health problems due to its increasing prevalence. In this
paper, we aim to test the ability of machine learning algorithms
for the prediction of chronic kidney disease using the smallest
subset of features</p>
         <p>
           <a href="{{ url_for('kidneyPage') }}" class="btn btn-primary my-2">Check out the Project</a>
         </p>
       </div>
     </section>


   </main>
   <!-- Bootstrap core JavaScript
   <!-- Placed at the end of the document so the pages load faster -->
   <script src="https://code.jquery.com/jquery-3.2.1.slim.min.js" integrity="sha384-
KJ3o2DKtIkvYIK3UENzmM7KCkRr/rE9/Qpg6aAZGJwFDMVNA/GpGFF93hXpG5KkN"
crossorigin="anonymous"></script>
   <script>window.jQuery || document.write('<script src="../../assets/js/vendor/jquery-
slim.min.js"><\/script>')</script>

   <script src="../../assets/js/vendor/popper.min.js"></script>

   <script src="../../dist/js/bootstrap.min.js"></script>

   <!-- Just to make our placeholder images work. Don't actually copy the next line! -->
  <script src="../../assets/js/vendor/holder.min.js"></script>
  </body>
</html>
{% endblock %}
```

KIDNEY.HTML

```
{% extends 'main.html' %}
{% block content %}

<div class="row"  style="margin-bottom: 125px;">
   <div class="col-md-2"></div>
   <div class="col-md-8">
      <center><h1>Kidney Disease Predictor</h1></center>

      <div class="card card-body" style="border: 1px solid black;">

         <form class="form-horizontal" action="{{ url_for('predictPage') }}" method="POST">
            <div class="row">
               <div class="col-md-4">
                  <div class="form-group">

                     <input style="border: 1px solid black;" class="form-control" type="text" name="age"
placeholder="age">
                  </div>
               </div>
               <div class="col-md-4">
                  <div class="form-group">
                     <input style="border: 1px solid black;" class="form-control" type="text" name="bp"
placeholder="bp">
                  </div>
               </div>
               <div class="col-md-4">
                  <div class="form-group">
                     <input style="border: 1px solid black;" class="form-control" type="text" name="al"
placeholder="al">
                  </div>
               </div>
            </div>
            <div class="row">
               <div class="col-md-4">
                  <div class="form-group">
                     <input style="border: 1px solid black;" class="form-control" type="text" name="su"
placeholder="su">
                  </div>
               </div>
```

```
                <div class="col-md-4">

                    <div class="form-group">

                        <input style="border: 1px solid black;" class="form-control" type="text" name="rbc"
placeholder="rbc">
                    </div>
                </div>

                <div class="col-md-4">
                    <div class="form-group">
                        <input style="border: 1px solid black;" class="form-control" type="text" name="pc"
placeholder="pc">
                    </div>
                </div>
            </div>
            <div class="row">
                <div class="col-md-4">
                    <div class="form-group">
                        <input style="border: 1px solid black;" class="form-control" type="text" name="pcc"
placeholder="pcc">
                    </div>
                </div>
                <div class="col-md-4">
                    <div class="form-group">

                        <input style="border: 1px solid black;" class="form-control" type="text" name="ba"
placeholder="ba">

                    </div>
                </div>
                <div class="col-md-4">
                    <div class="form-group">
                        <input style="border: 1px solid black;" class="form-control" type="text" name="bgr"
placeholder="bgr">
                    </div>
                </div>
            </div>
            <div class="row">
                <div class="col-md-4">
                    <div class="form-group">
```

```
                    <input style="border: 1px solid black;" class="form-control" type="text" name="bu"
placeholder="bu">
                  </div>
                </div>
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="sc"
placeholder="sc">
                  </div>
                </div>
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="pot"
placeholder="pot">
                  </div>
                </div>
              </div>
              <div class="row">
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="wc"
placeholder="wc">
                  </div>
                </div>
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="htn"
placeholder="htn">
                  </div>
                </div>
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="dm"
placeholder="dm">
                  </div>
                </div>
              </div>
              <div class="row">
                <div class="col-md-4">
                  <div class="form-group">
                    <input style="border: 1px solid black;" class="form-control" type="text" name="cad"
```

```
placeholder="cad">
                    </div>
                </div>
                <div class="col-md-4">
                    <div class="form-group">
                        <input style="border: 1px solid black;" class="form-control" type="text" name="pe"
placeholder="pe">
                    </div>
                </div>
                <div class="col-md-4">
                    <div class="form-group">
                        <input style="border: 1px solid black;" class="form-control" type="text" name="ane"
placeholder="ane">

                    </div>
                </div>
            </div>
            <input type="submit" class="btn btn-info btn-block" value="Predict">

        </form>
      </div>
   </div>
   <div class="col-md-2"></div>
</div> <br>

<h1 style="text-align: center"> Sample-Inputs in the Data Set</h1> <br>

<table class="table">
  <thead>
    <tr>
      <th scope="col">age</th>
      <th scope="col">bp</th>
      <th scope="col">al</th>
      <th scope="col">su</th>
      <th scope="col">rbc</th>
      <th scope="col">pc</th>
      <th scope="col">pcc</th>
      <th scope="col">ba</th>
      <th scope="col">bgr</th>
      <th scope="col">bu</th>
      <th scope="col">sc</th>
```

```
    <th scope="col">pot</th>
    <th scope="col">wc</th>
    <th scope="col">htn</th>
    <th scope="col">dm</th>
    <th scope="col">cad</th>
    <th scope="col">pe</th>
    <th scope="col">ane</th>
    <th scope="col">Disease</th>

  </tr>
</thead>
<tbody>
  <tr>

    <td>24</td>
    <td>100</td>
    <td>2</td>
    <td>0</td>
    <td>1</td>
    <td>0</td>
    <td>1</td>
    <td>0</td>
    <td>136</td>
    <td>60</td>
    <td>1.9</td>
    <td>3.7</td>
    <td>9600</td
    <td>1</td>
    <td>1</td>
    <td>0</td>
    <td>0</td>
    <td>1</td>
    <td>Present</td>
  </tr>
  <tr>
    <td>68</td>
    <td>80</td>
    <td>3</td>
    <td>0</td>
    <td>0</td>
    <td>1</td>
```

```
      <td>0</td>
      <td>0</td>
      <td>157</td>
      <td>162</td>

      <td>9.6</td>

      <td>4.9</td>
      <td>11000</td>
      <td>0</td>
      <td>1</td>
      <td>0</td>
      <td>0</td>
      <td>1</td>
      <td>Present</td>
    </tr>
    <tr>
      <td>51</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>1</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>121</td>
      <td>27</td>
      <td>0.8</td>
      <td>3.7</td>
      <td>8300</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>0</td>
      <td>Healthy</td>
    </tr>
  </tbody>
</table>

{% endblock %}
```

MAIN.HTML

```
<!DOCTYPE html>
<html>
<head>
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <meta name="og:title" content="Kidney-Disease Prediction">
  <meta name="og:image" content="static/logo1.png">
  <title>Kidney Disease Predictor</title>
  <link rel="icon" href="{{ url_for('static', filename = 'logo1.png') }}" type="image/icon type">
  <p> Kidney Disease </p>
  <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css"
integrity="sha384-
ggOyR0iXCbMQv3Xipma34MD+dH/1fQ784/j6cY/iJTQUOhcWr7x9JvoRxT2MZw1T"
crossorigin="anonymous">
    <link href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css"
rel="stylesheet"/>
    <link rel="canonical" href="https://getbootstrap.com/docs/4.0/examples/sticky-footer/">
    <script src="https://code.jquery.com/jquery-3.3.1.slim.min.js" integrity="sha384-
q8i/X+965DzO0rT7abK41JStQIAqVgRVzpbzo5smXKp4YfRvH+8abtTE1Pi6jizo"
crossorigin="anonymous"></script>
    <script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.14.7/umd/popper.min.js"
integrity="sha384-
UO2eT0CpHqdSJQ6hJty5KVphtPhzWj9WO1clHTMGa3JDZwrnQq4sF86dIHNDz0W1"
crossorigin="anonymous"></script>
    <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/js/bootstrap.min.js"
integrity="sha384-JjSmVgyd0p3pXB1rRibZUAYoIIy6OrQ6VrjIEaFf/nJGzIxFDsf4x0xIM+B07jRM"
crossorigin="anonymous"></script>
  <style>
    html, body{ height:100%; margin:0; }
header{ height:50px;}
footer{ height:75px; background:black; }
/* Trick */
body{
  display:flex;
  flex-direction:column;
}
footer{
  padding:10px;
  margin-top:auto;
  margin-bottom: auto;
}</style>
```

```
</head>
<body>
  <nav class="navbar navbar-expand-lg navbar-dark fixed-top bg-dark" style="background-color: black !important;">

      <a style="text-decoration: none; color: white" href="{{ url_for('home') }}"><h1> CHRONIC KIDNEY DISEASE PREDICTION</h1></a>

      <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarNav" aria-controls="navbarNav" aria-expanded="false" aria-label="Toggle navigation">
        <span class="navbar-toggler-icon"></span>
      </button>
      <div class="collapse navbar-collapse" id="navbarNav">
        <ul class="navbar-nav ml-auto">
          <li class="nav-item active">
            <a href="{{ url_for('home') }}" class="nav-link"><h3>Home</h3></a>
          </li>
          <li class="nav-item active">
            <a class="nav-link" href="{{ url_for('kidneyPage') }}"><h3>Kidney-Disease</h3></a>
          </li>
        </ul>
      </div>
    </nav>
  <br>
<br>
<br>
<br>
<main>
  <div class="container-fluid" style="margin-bottom: 20px;">
  {% block content %}

  {% endblock %}
    </div>
  </main><footer>
    <center>
      <ul style="list-style-type:none;">
    </ul>
  </center>
  </footer>
</body>
</html>
```

“ PREDICT.HTML ”

```
{% extends 'main.html' %}
{% block content %}
   <div class="row"  style="margin-bottom: 477px;">
      <div class="col-md-3"></div>
      <div class="col-md-6">
         {% if pred == 1 %}
            <div class="jumbotron ">
  <h1 class="display-4">You have a Kidney Disease !</h1>
  <p class="lead">Please Consult the Doctor Immideately. It was too risky without consultation. Make sure of health in your diet.</p>
  <hr class="my-4">
  <p>Proper Doctor Consultation Needed.</p>
  <p class="lead">
   <a class="btn btn-primary btn-lg" href="https://www.who.int/" role="button">Learn more</a>
  </p>
</div>
         {% else %}
            <div class="jumbotron">
  <h1 class="display-4">Great! You are Healthy</h1>
  <p class="lead">You are Absolutely Alright ! There is no Marks for Kidney Disease. Enjot=y you life with full of Happiness.</p>
  <hr class="my-4">
  <p>Be careful at your health. Nothing is important than your health.</p>
  <p class="lead">
   <a class="btn btn-primary btn-lg" href="https://www.who.int/" role="button">Learn more</a>
  </p>
</div>
         {% endif %}
         <div class="row">
            <div class="col-md-4"></div>
            <div class="col-md-4"><a href="{{ url_for('home') }}" class="btn btn-block btn-primary">Back to Home</a></div>
            <div class="col-md-4"></div>
         </div>
      </div>
      <div class="col-md-3"></div>
   </div>
{% endblock %}
```

10.2 GITHUB LINK :

https://github.com/gayathrinadendla/Early-Prediction-of-Chronic-Kidney-Disease/tree/main

PROJECT DEMO LINK :

https://drive.google.com/file/d/1o-_HRDInip2cVhkG9YcZYV7sfwEr4Hlh/view?usp=sharing