MEDICAL COST PREDICTION

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

GOUNI SUSHMA	**21UK1A0505**
CHALLA SAI CHARAN	**21UK1A0557**
BODDIREDDY PAPIREDDY	**21UK1A0556**
MORA DIVYA	**21UK1A0538**

Under the guidance of

Mr. P.VAMSHI KRISHNA

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) –

506005

DEPARTMENT OFCOMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION

INDUSTRY ORIENTED MINI PROJECT

This is to certify that the Industry Oriented Mini Project entitled "MEDICAL COST PREDICTION" is being submitted by GOUNI SUSHMA (21UK1A0505), CHALLA SAI CHARAN (21UK1A0557), BODDIREDDY PAPIREDDY (21UK1A0556), MORA DIVYA (21UK1A0538) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2024-2025.

Project Guide	HOD
Mr.P.Vamshikrishna	**Dr.Naveen Kumar Rangaraju**
(Assistant Professor)	(Head of the Department)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. SYED MUSTHAK AHAMED,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.NAVEEN KUMAR RANGARAJU** Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **P. VAMSHI KRISHNA,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this Industry Oriented Mini Project.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the project.

GOUNI SUSHMA	**(21UK1A0505)**
CHALLA SAI CHARAN	**(21UK1A0557)**
BODDIREDDY PAPIREDDY	**(2IUK1A0556)**
MORA DIVYA	**(21UK1A0538)**

ABSTRACT

Medical costs are one of the most common reoccurring expenses in a person's life. It is general known that a person's lifestyle and numerous physical factors determine thediseases or disorders they may get, and that these conditions determine medical expenses. According to several research, there are several significant reasons that leads to greater expenditures. smoking, age, and BMI are all factors in personal medical care.The goal of this project is to examine and identify a link between personal medical costs and other characteristics. Then, by generating Machine learning models like linear Regression, Random Forest and comparing them using ANOVA, we use the significant traits as predictors to forecast medical expenditures. In our research, we discovered that smoking, age, and a higher BMI all have a significant connection with higher medical expenditures, showing that they arekey contributors to the charges, and that the regression can predict the charges with more than 75% accuracy. According to the World Health Organization, personal medical and healthcare spending is growing faster than the global economy This rise in spending has been related to a variety of factors, the most prominent of which are smoking, ageing, and higher BMI. Using insurance data from diverse persons with variables such as smoking, age, number of children, area, and BMI, we hope to uncovera link between medical expenditures and other parameters. The core objective of this project is to empower individuals with the knowledge to make informed healthcare decisions by offering transparency in medical expenses. Additionally, healthcare providers can leverage these predictions to enhance financial planning and resource allocation. Our approach involves data preprocessing, feature selection, and the implementation of various machine learning algorithms such as linear regression, and ensemble methods to identify the most effective predictive model. The expected outcome of this project is a user-friendly tool that delivers reliable cost estimates, thereby bridging the gap between patients and financial clarity in healthcare. This initiative ultimately aspires to contribute to a more efficient and compassionate healthcare system, where financial considerations are less of a burden on patients, leading to better healthcare outcomes and overall satisfaction.

TABLE OF CONTENTS:-

1.INTRODUCTION

1.1OVERVIEW

The prediction of medical costs is a critical task for healthcare providers, insurance companies, and policy-makers due to its potential to influence financial planning, resource allocation, and strategic decision-making. Accurate cost predictions can significantly improve the efficiency and effectiveness of healthcare services by enabling providers to better anticipate and manage patient care expenses. For insurance companies, precise cost forecasting is essential for setting appropriate premiums and managing risk, ensuring that both the insurer and the insured are protected. Policy-makers can use these predictions to develop informed healthcare policies that address cost drivers and improve overall system sustainability. Unpredictable medical expenses can lead to financial strain, causing stress and anxiety for patients who are already dealing with health issues. In this context, the ability to accurately predict medical costs becomes crucial. In recent years, healthcare systems worldwide have faced challenges related to cost management and resource allocation. Patients often encounter unexpected medical bills, making it difficult to manage their finances and plan for future healthcare needs. Providers, on the other hand, struggle with optimizing their services and resources due to fluctuating costs.This project focuses on developing a predictive model for medical cost estimation using machine learning techniques, aiming to provide transparency and assist in financial planning for healthcare expenses.

By leveraging advanced data analytics and machine learning techniques, we can create robust predictive models that analyze a wide range of patient attributes, such as demographics, medical history, lifestyle factors, and clinical data. These models help in identifying patterns and trends that contribute to medical expenses, providing valuable insights for all stakeholders.

This document outlines the comprehensive methodology employed in this project, detailing the steps taken from data collection and preprocessing to model development, evaluation, and deployment. The results demonstrate the effectiveness of machine learning in predicting healthcare costs, highlighting its potential to transform healthcare financial management and policy formulation.

1.2 PURPOSE

The primary purpose of this medical cost prediction project is to address the uncertainty and financial burden associated with healthcare expenses. By developing an advanced machine learning model, the project aims to provide accurate and personalized medical cost estimates. This initiative serves multiple key purposes:

1.<u>Empowering Patients</u>: Enable patients to plan and manage their finances better by providing transparent and accurate predictions of medical expenses. Help patients make informed healthcare decisions by giving them a clearer understanding of potential costs associated with different medical treatments and procedure. Alleviate the stress and anxiety caused by unexpected medical bills, thus enhancing the overall patient experience.

2. <u>Assisting Healthcare Providers</u>: Aid healthcare providers in optimizing resource allocation by predicting costs more accurately, leading to more efficient management of services and facilities. Enhance the quality of healthcare services by allowing providers to better anticipate and plan for financial needs. Improve patient satisfaction through greater financial transparency and better communication about potential costs.

3. **<u>Advancing Healthcare Systems:</u>** Utilize comprehensive data analysis to generate insights that can inform policy-making and improve healthcare systems at a broader level. Contribute to more effective cost management strategies within the healthcare sector, potentially leading to more sustainable and affordable healthcare solutions. Promote innovation by integrating advanced machine learning techniques into healthcare cost prediction, setting a precedent for future technological advancements in the field.

4.<u>Educational and Research Objective</u>: Provide an opportunity for students and researchers to apply machine learning techniques to real-world problems, enhancing their skills and knowledge. Contribute to the academic community by presenting findings and methodologies that can be used as a reference for future research projects.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

The rising costs of healthcare and the unpredictability of medical expenses present significant challenges for both patients and healthcare providers. Several key problems underline the need for a medical cost prediction project:

1. **Unpredictable Medical Expenses:**

   Financial Strain on Patients they often face unexpected and substantial medical bills that can lead to financial hardship. The lack of transparency in healthcare costs makes it difficult for individuals to plan and manage their finances effectively.The fear of unknown medical expenses can cause significant stress and anxiety for patients, impacting their overall well-being and health outcomes.

2. **Lack of Cost Transparency:**

   Healthcare costs can vary widely depending on numerous factors such as the provider, location, and specific medical procedures. This inconsistency makes it challenging for patients to anticipate and understand their potential financial obligations.The complexity of medical billing and pricing structures often leaves patients confused and uninformed about the true cost of their care.

3. **Resource Allocation Challenges for Providers:**

   Healthcare providers struggle with efficiently managing resources due to the unpredictability of patient needs and associated costs. This can lead to either overutilization or underutilization of healthcare. Providers face difficulties in financial planning and budgeting without accurate predictions of patient costs, affecting the overall efficiency and quality of care.

2.2 PROPOSED SOLLUTION

The proposed solution for addressing the challenges of unpredictable medical expenses and lack of cost transparency involves developing a robust predictive model using advanced machine learning techniques. This project will begin by collecting and integrating comprehensive datasets encompassing patient demographics, medical histories, treatment specifics, and regional healthcare costs from reliable sources. Through meticulous data preprocessing, including cleaning and feature engineering, the aim is to enhance data quality and relevance for accurate analysis. Machine learning algorithms such as linear regression, decision trees, and ensemble methods will be explored to construct and refine the predictive model.

1. Data Collection and Integration: Gather comprehensive datasets from reliable sources, including patient demographics, medical history, treatment details, and regional healthcare costs. claims. Integrating diverse data sources such as electronic health records (EHRs), insurance claims, and regional healthcare cost databases to create a unified and comprehensive dataset.

2.Data Preprocessing and Feature Engineering: Clean the data to handle missing values and inconsistencies. Select and engineer relevant features that significantly impact medical costs. Identifying and selecting relevant features that significantly impact medical costs, such as age, gender, pre-existing conditions, types of treatments, and regional cost variations.

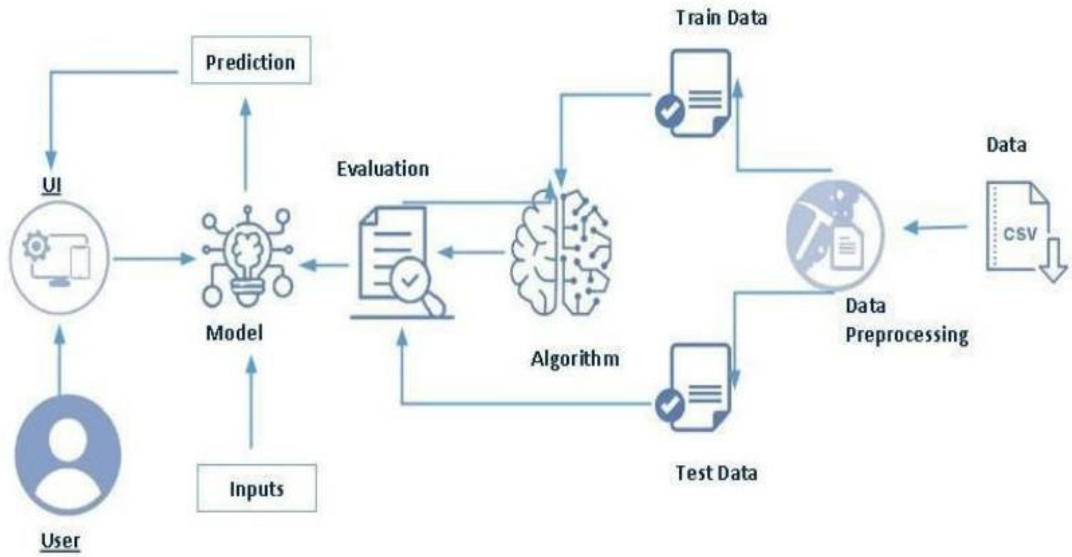
3. **Machine Learning Model Development**: Experiment with various algorithms (e.g., linear regression, random forests) to identify the most effective model. Train, validate, and fine-tune the model to optimize performance. Training the selected models on the preprocessed dataset and validating their performance using cross-validation techniques to ensure accuracy and generalizability.

4. Model Evaluation: Assess the model using performance metrics like Mean Absolute Error (MAE) and R-squared (R^2) value. Compare different models to select the best one. Comparing the performance of different models to select the best one for medical cost prediction.

5. User Interface and Deployment: Develop a user-friendly web or mobile application for patients and providers to input data and receive cost predictions. Ensure real-time predictions and implement robust security measures to protect sensitive data. It allows patients and healthcare providers to input relevant information and receive personalized cost estimates in real-time.

3.THEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM



3.2 . SOFTWARE DESIGNING

The following software and tools were used to develop the Medical Cost Prediction System:

Development Environment:

Google Colab: Google Colab will serve as the development and execution environment for your predictive modeling, data preprocessing, and model training tasks. It provides a cloud-based Jupyter Notebook environment with access to Python libraries and hardware acceleration.

Dataset (CSV File): The dataset in CSV format is essential for training and testing your predictive model. It should include historical air quality data, weather information, pollutant levels, and other relevant features.

Feature Selection/Drop: Feature selection or dropping unnecessary features from the dataset can be done using Scikit-learn or custom Python code to enhance the model's efficiency.

Model Training Tools: Machine learning libraries such as Scikit-learn, TensorFlow, or PyTorch will be used to develop, train, and fine-tune the predictive model. Regression or classification models can be considered, depending on the nature of the prediction task.

Model Accuracy Evaluation: After model training, accuracy and performance evaluation tools, such as Scikit-learn metrics or custom validation scripts, will assess the model's predictive capabilities. You'll measure the model's ability to predict categories based on historical data.

UI Based on Flask Environment: Flask, a Python web framework, will be used to develop the user interface (UI) for the system. The Flask application will provide a user-friendly platform for users to input location data or view charge predictions, health information, and recommended precautions.

Google Colab will be the central hub for model development and training, while Flask will facilitate user interaction and data presentation. The dataset, along with data preprocessing, will ensure the quality of the training data, and feature selection will optimize the model. Finally, model accuracy evaluation will confirm the system's predictive capabilities, allowing users to rely on the charge predictions and associated health information.

4.EXPERIMENTAL INVESTIGATIONS

In this Project ,we have used insurance Dataset .The dataset consists of patient records with the following attributes:

Age: Age of the patient.

Sex: Gender of the patient.

BMI: Body Mass Index of the patient.

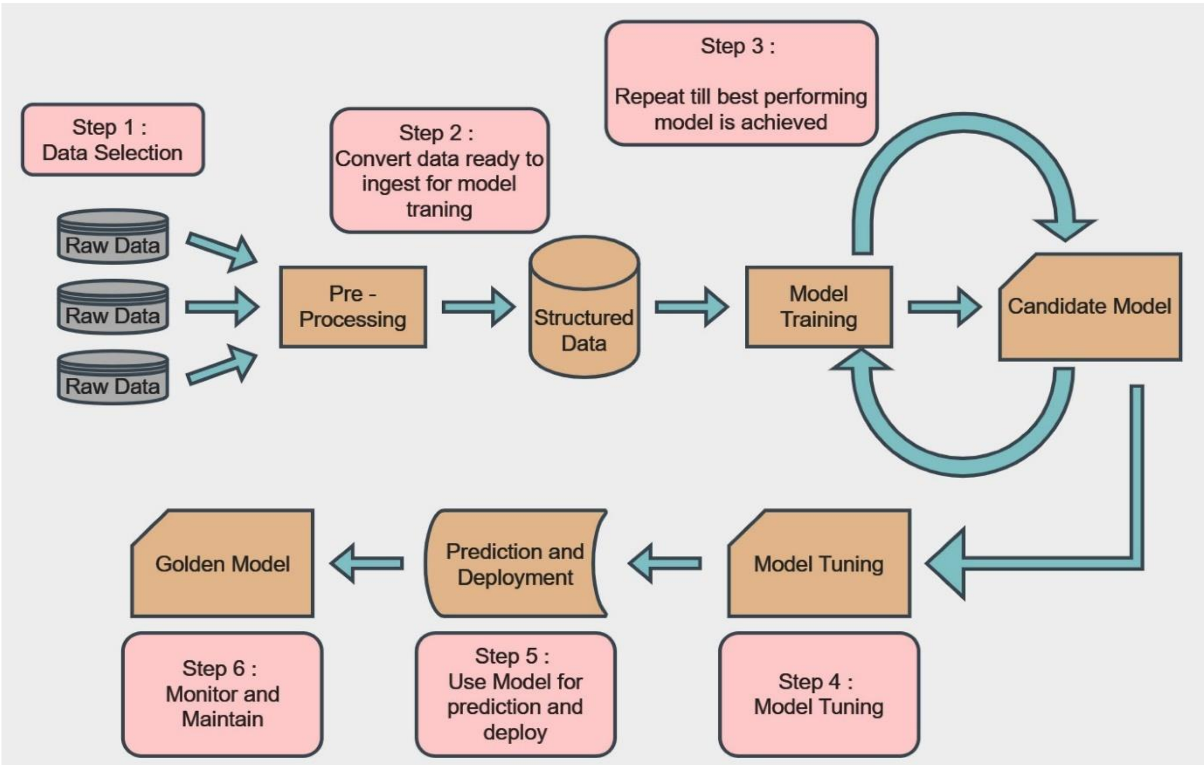
Children: Number of children/dependents.

Smoker: Smoking status of the patient.

Region: Geographical region of the patient.

Charges: Medical expenses incurred.

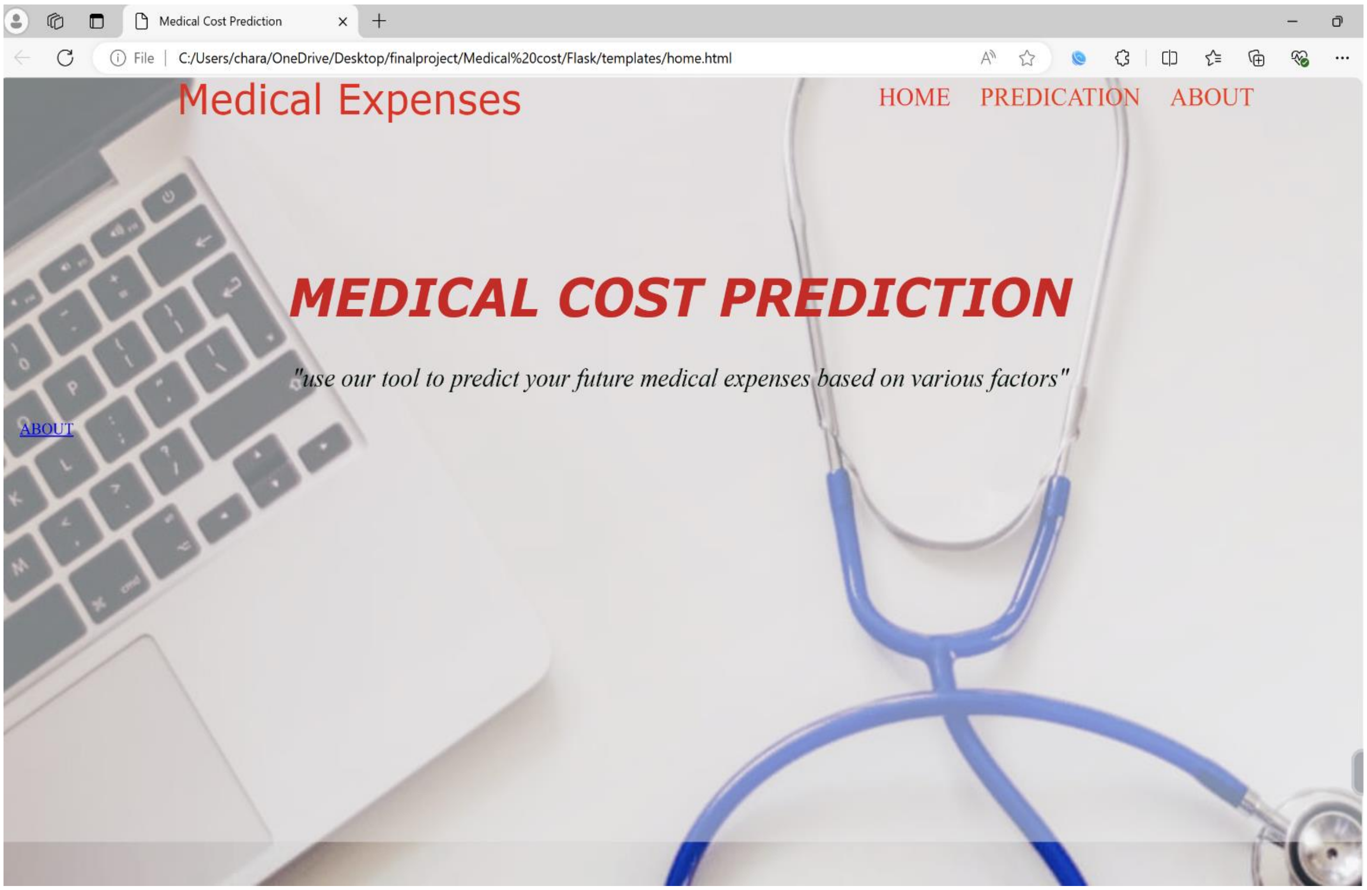
5.FLOWCHART



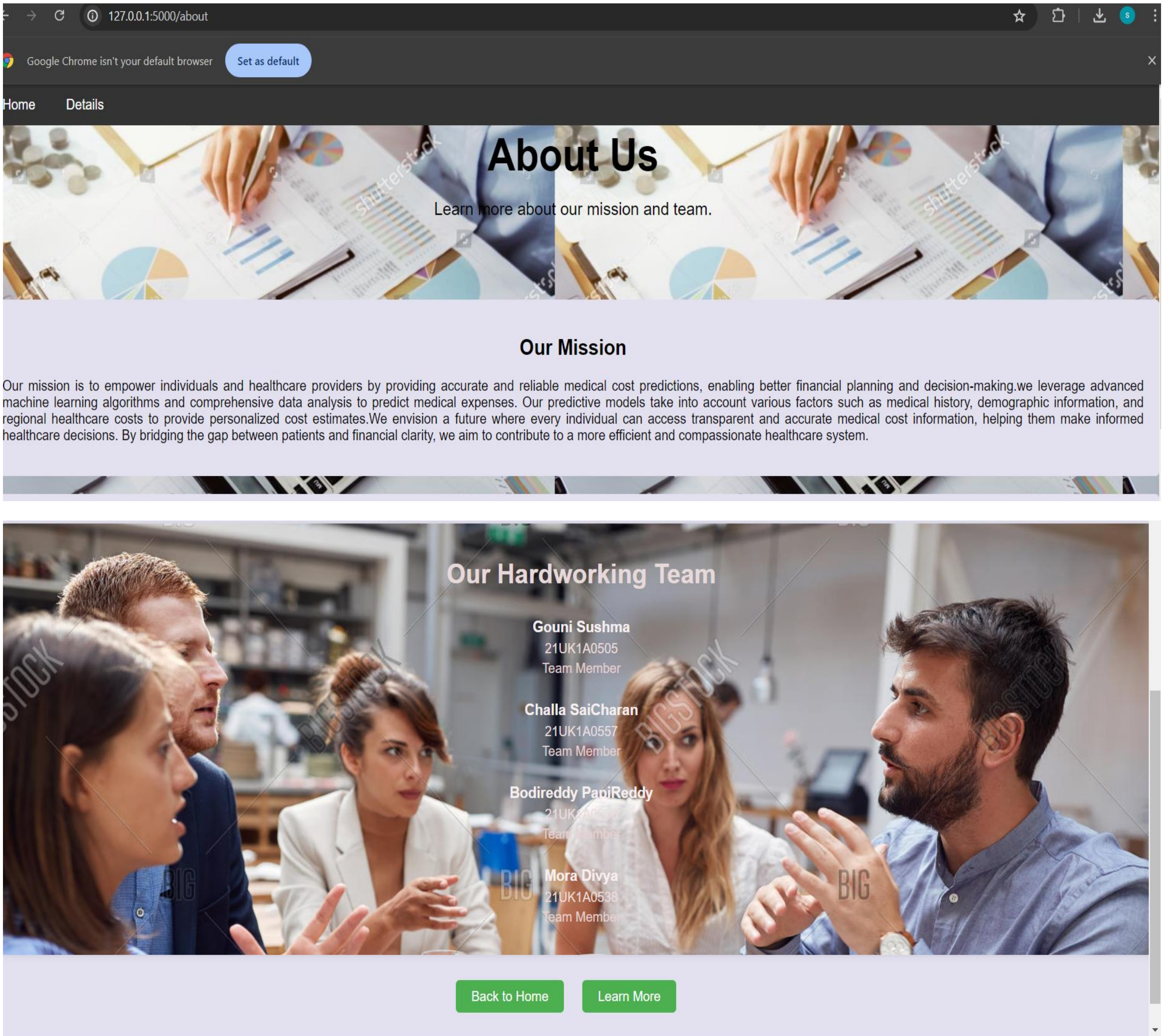
Step 1 :
Data Selection
Raw Data
Raw Data
Raw Data
Step 2 :
Convert data ready to ingest for model traning
Pre - Processing
Structured Data
Step 3 :
Repeat till best performing model is achieved
Model Training
Candidate Model
Model Tuning
Step 4 :
Model Tuning
Prediction and Deployment
Step 5 :
Use Model for prediction and deploy
Golden Model
Step 6 :
Monitor and Maintain

6.RESULT

HOME PAGE

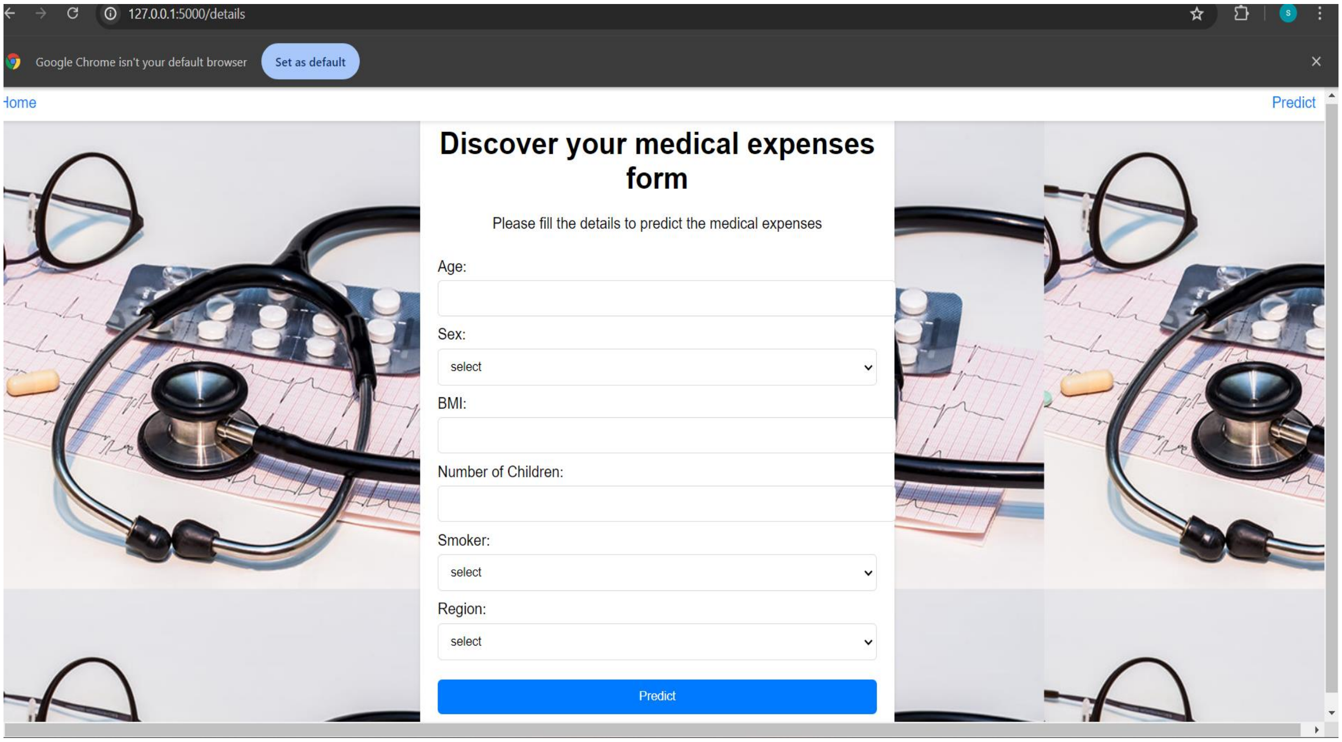


ABOUT PAGE

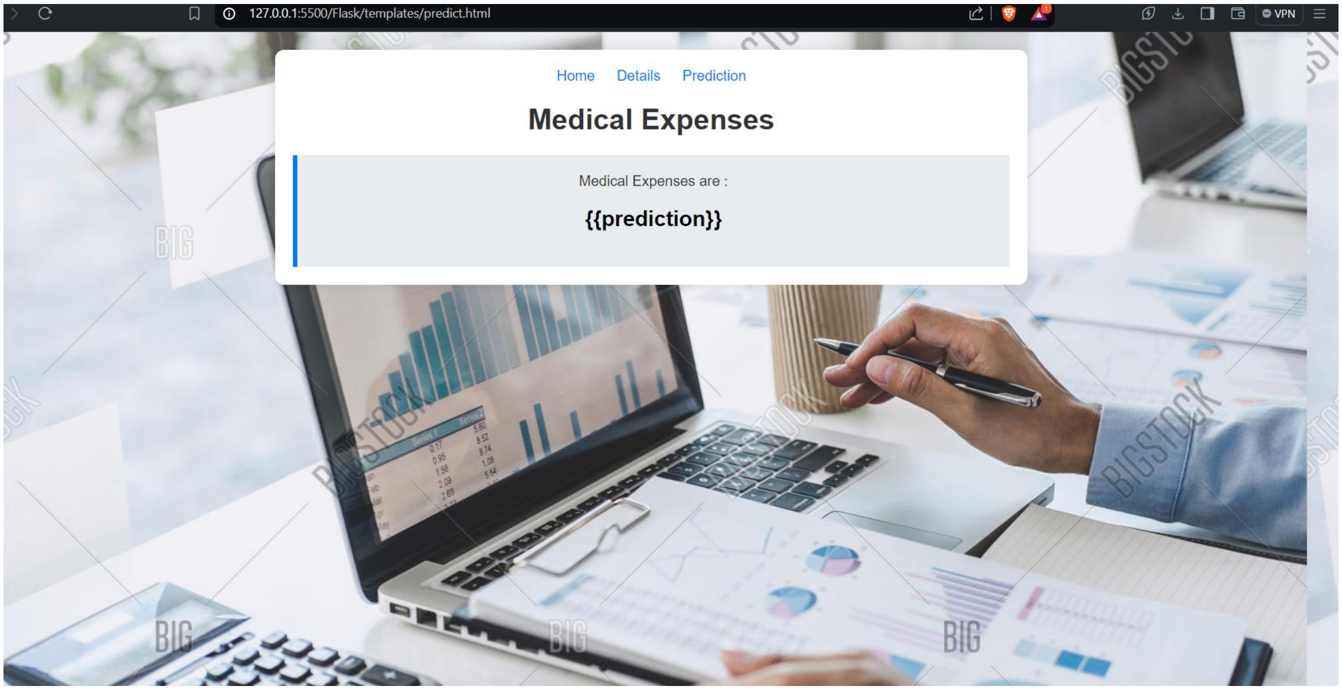


127.0.0.1:5000/about
Google Chrome isn't your default browser
Set as default
Home
Details
About Us
Learn more about our mission and team.
Our Mission
Our mission is to empower individuals and healthcare providers by providing accurate and reliable medical cost predictions, enabling better financial planning and decision-making.we leverage advanced machine learning algorithms and comprehensive data analysis to predict medical expenses. Our predictive models take into account various factors such as medical history, demographic information, and regional healthcare costs to provide personalized cost estimates.We envision a future where every individual can access transparent and accurate medical cost information, helping them make informed healthcare decisions. By bridging the gap between patients and financial clarity, we aim to contribute to a more efficient and compassionate healthcare system.
Our Hardworking Team
Gouni Sushma
21UK1A0505
Team Member
Challa SaiCharan
21UK1A0557
Team Member
Bodireddy PaniReddy
Mora Divya
Back to Home
Learn More

DETAILS PAGE



PREDICT PAGE



7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

1.**Financial Planning**: Predicting medical costs allows individuals and healthcare providers to plan finances more effectively, reducing the uncertainty and stress associated with unexpected medical bills.

2.**Improved Patient Satisfaction**: Transparent cost estimates empower patients to make informed decisions about their healthcare, enhancing overall satisfaction with healthcare services.

3.Resource Optimization: Healthcare providers can better allocate resources and manage budgets based on anticipated costs, improving operational efficiency and patient care delivery.

4.Data-Driven Insights: Analysis of medical cost trends and patterns provides valuable insights for policymakers and administrators to improve healthcare policies and resource allocation strategies.

5. Personalized Healthcare: Tailored cost predictions based on individual patient characteristics enable personalized care planning and financial counseling, potentially improving health outcomes.

DISADVANTAGES:

1. Data Complexity: Integrating and analyzing diverse datasets (e.g., demographics, medical histories) can be complex and require substantial computational resources and expertise.

2. Privacy Concerns: Handling sensitive patient data raises privacy and security concerns, requiring robust measures to protect confidentiality and comply with regulations (e.g., HIPAA).

3. Interpretation Challenges: Complex machine learning models (e.g., random forest, gradient boosting) used for prediction can be challenging to interpret compared to simpler models like linear regression, potentially limiting understanding and trust among stakeholders.

4. Implementation and Adoption: Successfully deploying and integrating predictive models into existing healthcare systems may face resistance due to organizational inertia and cultural factors.

8.APPLICATIONS

1. Patient Financial Planning: Patients can use cost prediction tools to estimate expenses for upcoming medical procedures or treatments, aiding in financial planning and budgeting. It helps patients anticipate and prepare for out-of-pocket expenses, reducing financial stress and enabling informed decision-making about healthcare options.

2. Healthcare Provider Resource Allocation: Healthcare providers can use predictive models to forecast expected costs associated with patient care, optimizing resource allocation and budget management. Enhances operational efficiency by ensuring appropriate staffing, supply chain management, and facility utilization based on anticipated financial needs.

3.Insurance and Payer Strategies: Insurance companies and healthcare payers utilize cost prediction models to estimate future claims and premiums, informing pricing strategies and risk management.

Facilitates more accurate pricing and premium adjustments, improving financial sustainability and profitability for insurance providers.

4.Policy Making and Healthcare Economics: Policymakers and healthcare economists utilize cost prediction insights to inform healthcare policy decisions, resource allocation, and reimbursement strategies. Supports evidence-based policymaking, enhances healthcare system efficiency, and promotes equitable access to affordable healthcare services.

5. Clinical Trial Budgeting and Feasibility: Pharmaceutical companies and research institutions use cost prediction models to estimate expenses associated with clinical trials, aiding in budget planning and feasibility assessments. It Enables more accurate financial planning for research initiatives, optimizing trial design and recruitment strategies while minimizing cost overruns.

9. CONCLUSION

In conclusion, the medical cost prediction project represents a pivotal advancement in healthcare management, aiming to address the longstanding challenges of financial unpredictability and lack of transparency in medical expenses. By harnessing advanced machine learning techniques and integrating diverse datasets encompassing patient demographics, medical histories, and regional cost variations, this project aims to provide precise and personalized predictions of medical expenditures. By leveraging advanced machine learning algorithms and comprehensive datasets, this project endeavors to provide accurate, personalized cost estimates for patients and healthcare providers alike. The implementation of predictive models such as linear regression, random forest, and gradient boosting offers promising avenues for improving financial planning, resource allocation, and decision-making within healthcare systems. Moreover, the project's potential extends beyond immediate cost predictions to influencing policy development, enhancing patient satisfaction, and fostering a more efficient and equitable healthcare environment. These models not only enable patients to better plan and manage their healthcare finances but also empower healthcare providers to optimize resource allocation and improve operational efficiency. Beyond immediate financial planning benefits, the project holds promise in shaping healthcare policy, enhancing decision-making processes, and fostering a more transparent and patient-centric healthcare system. As the project continues to evolve, ongoing advancements in data analytics, real-time data integration, and user interface design will further refine its capability to deliver accurate cost estimates and drive positive outcomes in healthcare delivery and patient care management.As the field continues to evolve with technological innovations and data-driven insights, the ongoing refinement and application of these models hold the promise of transforming healthcare delivery into one that is more transparent, responsive, and patient-centered.

10.FUTURE SCOPE

1.Enhanced Predictive Models: Advanced Machine Learning Techniques leads to Continued development of advanced machine learning algorithms such as deep learning, reinforcement learning, and ensemble methods to improve prediction accuracy and robustness. Incorporating real-time data streams from wearable devices, electronic health records (EHRs), and IoT sensors to provide dynamic and personalized cost estimates.

2.Personalized Healthcare Solutions: Incorporating genomic data, biomarkers, and personalized treatment plans into cost prediction models to tailor estimates based on individual patient characteristics and health risks. Incorporating social determinants of health (e.g., socioeconomic status, lifestyle factors) to provide holistic cost predictions that consider broader influences on healthcare utilization.

3.Blockchain and Data Security: Utilizing blockchain technology to enhance data security, transparency, and interoperability in sharing sensitive healthcare data across stakeholders while ensuring patient privacy and regulatory compliance.

4.Policy and Healthcare System Impact: Conducting longitudinal studies to evaluate the impact of healthcare policy changes (e.g., reimbursement reforms, new care delivery models) on medical costs and adapting prediction models accordingly. Continuously optimizing predictive models to support healthcare system efficiency initiatives, including reducing administrative costs, improving care coordination, and enhancing patient outcomes.

5.Global Health Applications: Applying cost prediction models across different healthcare systems globally to compare cost structures, identify best practices, and support international health policy development. Addressing healthcare cost challenges in emerging markets by developing scalable and adaptable prediction models that account for local healthcare infrastructure and economic condition

11.BIBILOGRAPHY

[1] “ Obermeyer, Z., & Emanuel, E. J. (2016). *Predicting the Future — Big Data, Machine Learning, and Clinical Medicine*. New England Journal of Medicine, 375, 1216-1219.

[2] J. Cubanski, C. Swoop and T. Neuman,“How Much Is Enough? Out-of-Pocket Spending Among Medicare Beneficiaries: A Chartbook,” The Henry J. Kaiser Family Foundation, Menlo Park, CA, 2014. [Online]. Available:http://files.kff.org/attachment/how-much-isenough-out-of-pocket-spending-among-medicare-beneficiaries-a-chartbook-

[3] J. Cubanski and T. Neuman,“The Facts on Medicare Spending and Financing,” The HenryJ Kaiser Family Foundation, Menlo Park, CA, 2017. [Online]. Available: http://files.kff.org/attachment/Issue-Brief-The-Facts-on-Medicare-Spending-and-Financing. [Accessed Nov. 2, 2017].

[4] The Henry J. Kaiser Family Foundation. (2017). Total Number of Medicare Beneficiaries.[Online]. Available: https://www.kff.org/ medicare/state-indicator/total-medicarebeneficiaries/. [Accessed Nov. 2, 2017].

[5] Scikit-learn.org. (2018). About us — scikit-learn 0.19.1 documentation. [online] Available at: http://scikit-learn.org/stable/about.html#people [Accessed 23 Apr. 2018].

12.APPENDIX

<u>Model building</u>

1)Dataset

2)Google colab and VS code Application Building

1. HTML file (Home file, About file, Details file, Predict file)
1. Inline CSS file
2. Models in pickle format

SOURCE CODE:

<u>HOME.HTML</u>

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title>Medical Cost Prediction</title>
    <meta name="viewport"content="width=device-width, initial-scale=1.0">
  </head>
  <body>
    <style>
      body {
        box-sizing: border-box;
        margin: 0;
        padding: 0;
        background-image:url(../static/assets/img/details.jpg);
        object-fit: contain;
        background-size: 100% 800px;
        background-repeat: no-repeat;


    }
    #lg{
```

```
    background: linear-gradient(to bottom right, rgba(233, 225, 225, 0.5), rgba(227, 221, 221, 0.5),
rgba(238, 241, 243, 0.5));
    width: 100%;
    height: 750px;
    margin: -20px 0;
  }
  #name {
      font-size: 40px;
      margin: 0 12%;
      padding: 15px;
      color: rgb(238, 18, 18);
     font-family: Verdana, Geneva, Tahoma, sans-serif;
    }
  #nav ul{
      list-style-type: none;
      display: flex;
      justify-content: flex-end;
      display: flex;
      justify-content: flex-end;
      color: rgb(238, 18, 18);
      gap: 10px;
      margin:-55px 100px;
    }
  #nav li a {
    text-decoration: none;
    font-size: 25px;
    color: rgb(249, 47, 11);
    padding: 10px;
  }
  #nav li a.active{
     color: rgb(251, 11, 23);
  }
  @media screen and (min-width:0) and(max-width: 768px) {
    #nav ul{
```

```
      width: auto;
      height: auto;
      overflow: hidden;
     position: relative;
     margin-top:25px;
     display: block;
     color: rgb(174, 4, 4);
}
#nav li{
   float:left;
}
#nav li a {
padding:10px;
}
}
#title{
  font-size: 50px;
  font-family: Verdana, Geneva,Tahoma, sans-serif;
  font-style: oblique;
  text-align: center;
  margin-top: 200px;
  color: rgb(214, 11, 25);

}
#quo{
   text-align: center;
   margin-top: 20px;
   font-size:25px;
   font-style: italic;
   color: rgb(0, 16, 6);
}
#predbut{
    display: flex;
    justify-content: center;
```

```
        align-items: center;
        height: 20vh;
    }
    #box{
         width: 250px;
         height: 40px;
         border: 2px solid rgb(22, 2, 109);
         border-radius: 5px;
         padding: 15px 5px 8px 12px;
         backdrop-filter: blur(5px);
    }
    #box a{
        text-decoration: none;
        font-size: 25px;
        color: rgb(9, 9, 8);
    }
    </style>
   <div id="lg">
    <header id="name">Medical Expenses</header>
   <nav id="nav">
    <ul>
      <li><a href="#"class="active">HOME</a></li>
       <li><a href="/details">PREDICATION</a></li>
       <li><a href="/about">ABOUT</a></li>
  </ul>
  </nav>
          <h1 id="title">MEDICAL COST PREDICTION</h1>
          <p id="quo">"use our tool to predict your future medical expenses based on various
factors"</p>
          <li><a href="/about">ABOUT</a></li>
 </div>
</body>
</html>
```

ABOUT.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>About Us - Medical Cost Prediction</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            margin: 0;
            padding: 0;
            background-color:#ddd;
            background-image: url(../static/assets/img/i2.jpg);
        }
        .navbar {
            overflow: hidden;
            background-color: #333;
            position: fixed;
            top: 0;
            width: 100%;
            z-index: 1000;
        }
        .navbar a {
            float: left;
            display: block;
            color: white;
            text-align: center;
            padding: 14px 20px;
            text-decoration: none;
        }
        .navbar a:hover {
```

```
    background-color: #ddd;
    color: black;
}
.header {
    text-align: center;
    padding: 50px;
}
.header h1 {
    font-size: 50px;
    margin: 0;
}
.header p {
    font-size: 20px;
}
.content {
    text-align: center;
    padding: 20px;
}
.content a {
    display: inline-block;
    margin: 10px;
    padding: 10px 20px;
    background-color: #4CAF50;
    color: white;
    text-decoration: none;
    border-radius: 5px;
}
.content a:hover {
    background-color: #45a049;
}
.about-section {
    background-color: #e3e0f0;
    margin:20px 0;
    padding: 20px;
```

```
  border-radius:5px;
 }
.about-section h2  {
   text-align: center;
}
.about-section p {
  text-align: justify;
}
.container {
  background-color: #fff;
  width: 100%;
  box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
  padding: 20px;
  background-image: url(../static/assets/img/meet.jpg);
  border-radius: 8px;
  text-align: center;
}
.container h1 {
  color: #e8dada;
  margin-bottom: 20px;
}
.team-member {
  padding: 10px;
  color: #bc1515;


}
.team-member:last-child {
  border-bottom: none;
}
.team-member p {
  margin: 5px 0;
}
.team-member .name {
  font-size: 18px;
```

```
            font-weight: bold;
            color: #f8eff1;
            text-align: center;
        }
        .team-member .id {
            color: #ecdddd;
            text-align: center;
        }
        .team-member .role {
            color: #e7d1d1;
            text-align: center;
        }
    </style>
</head>
<body>
    <div class="navbar">
        <a href="/home">Home</a>
        <a href="/details">Details</a>
    </div>
    <div >
        <div class="header">
            <h1>About Us</h1>
            <p>Learn more about our mission and team.</p>
        </div>
        <div class="about-section">
            <h2>Our Mission</h2>
            <p>Our mission is to empower individuals and healthcare providers by providing accurate and reliable medical cost predictions, enabling better financial planning and decision-making.we leverage advanced machine learning algorithms and comprehensive data analysis to predict medical expenses. Our predictive models take into account various factors such as medical history, demographic information, and regional healthcare costs to provide personalized cost estimates.We envision a future where every individual can access transparent and accurate medical cost information, helping them make informed healthcare decisions. By bridging the gap between patients and financial clarity, we aim to contribute to a more efficient and compassionate healthcare
```

```html
system.</p>
        </div>
        <div class="about-section">
            <div class="container">
                <h1>Our Hardworking Team</h1>
                <div class="team-member">
                    <p class="name">Gouni Sushma</p>
                    <p class="id">21UK1A0505</p>
                    <p class="role">Team Member</p>
                </div>
                <div class="team-member">
                    <p class="name">Challa SaiCharan</p>
                    <p class="id">21UK1A0557</p>
                    <p class="role">Team Member</p>
                </div>
                <div class="team-member">
                    <p class="name">Bodireddy PapiReddy</p>
                    <p class="id">21UK1A0556</p>
                    <p class="role">Team Member</p>
                </div>
                <div class="team-member">
                    <p class="name">Mora Divya</p>
                    <p class="id">21UK1A0538</p>
                    <p class="role">Team Member</p>
                </div>
            </div>
        <div>
        <div class="content">
            <a href="/home">Back to Home</a>
            <a href="/details">Learn More</a>
        </div>
    </div>
</body>
</html>
```

DETAILS.HTML

```
 <!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Medical Expenses</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            background-color: #f8f8f8;
            margin: 0;
            padding: 0;
            background-image:url(../static/assets/img/departments-1.jpg);
            display: flex;
            flex-direction: column;
            align-items: center;
            justify-content: center;
            height: 100vh;
        }
        header {
            position: absolute;
            top: 0;
            width: 100%;
            background-color: #fff;
            padding: 10px;
            box-shadow: 0 2px 4px rgba(0,0,0,0.1);
            display: flex;
            justify-content: space-between;
            align-items: center;
        }
        header a {
            text-decoration: none;
```

```
    color: #007bff;
    margin: 0 10px;
}
.container {
    background-color: #fff;
    padding: 20px;
    border-radius: 10px;
    box-shadow: 0 2px 4px rgba(0,0,0,0.1);
    max-width: 500px;
    width: 100%;
    text-align: center;
}
h1 {
    margin-bottom: 20px;
}
form {
    display: flex;
    flex-direction: column;
    align-items: flex-start;
}
label {
    margin-top: 10px;
}
input, select {
    width: 100%;
    padding: 10px;
    margin-top: 5px;
    border: 1px solid #ddd;
    border-radius: 5px;
}
button {
    margin-top: 20px;
    padding: 10px;
    width: 100%;
```

```
            background-color: #007bff;
            color: #fff;
            border: none;
            border-radius: 5px;
            cursor: pointer;
        }
        button:hover {
            background-color: #0056b3;
        }
    </style>
</head>
<body>
    <header>
        <a href="/home">Home</a>
        <a href="/predict">Predict</a>
    </header>
    <div class="container">
        <h1>Discover your medical expenses form</h1>
        <p>Please fill the details to predict the medical expenses</p>
             <form action="/predict" method="post">
            <label for="age">Age:</label>
            <input type="number" id="age" name="age" min="0" max="100" required>

            <label for="sex">Sex:</label>
            <select id="sex" name="sex" required>
                <option value="select">select</option>
                <option value="male">Male</option>
                <option value="female">Female</option>
            </select>

            <label for="bmi">BMI:</label>
            <input type="number" id="bmi" name="bmi" step="0.1" required>

            <label for="children">Number of Children:</label>
```

```html
            <input type="number" id="children" name="children" min="0" required>

            <label for="smoker">Smoker:</label>
            <select id="smoker" name="smoker" required>
                <option value="select">select</option>
                <option value="yes">Yes</option>
                <option value="no">No</option>
            </select>
            <label for="region">Region:</label>
            <select id="region" name="region" required>
                <option value="select">select</option>
                <option value="northwest">Northwest</option>
                <option value="northeast">Northeast</option>
                <option value="southeast">Southeast</option>
                <option value="southwest">Southwest</option>
            </select>
                <button type="submit">Predict</button>
        </form>
    </div>
</body>
</html>
```

PREDICT.HTML

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Medical Expenses</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            margin: 0;
            padding: 0;
```

```
    background-image:url(../static/assets/img/img2.jpg);
    background-color: #f4f4f4;
}
.container {
    max-width: 800px;
    margin: 20px auto;
    padding: 20px;
    text-align: center;
    border-radius: 10px;
    background-color: #fff;
    box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
}
h1 {
    color: #333;
}
nav {
    margin-bottom: 20px;
}
nav a {
    margin: 0 10px;
    text-decoration: none;
    color: #007bff;
}
nav a:hover {
    text-decoration: underline;
    text-align: center;
}
.prediction {
    padding: 20px;
    background-color: #e9ecef;
    border-left: 5px solid #007bff;
    text-align: center;
}
.prediction p {
```

```
            margin: 0;
            color: #333;
        }
    </style>
</head>
<body>
    <div class="container">
        <nav>
            <a href="/home">Home</a>
            <a href="/details">Details</a>
            <a href="/predict">Prediction</a>
        </nav>
        <h1>Medical Expenses</h1>
        <div class="prediction">
            <p>Hence, based on calculation, the medical expenses are: </p>
            <h2 id="prediction">{{prediction}}</h2>
        </div>
    </div>
</body>
</html>
```

APP.PY

```
from flask import Flask, render_template, request
import numpy as np
import pandas as pd
import pickle

app = Flask(__name__)
# Load the model once when the app starts
with open('rf.pkl', 'rb') as f:
    model = pickle.load(f)
@app.route('/',methods=['GET'])
```

```
def home():
    return render_template('home.html')

@app.route('/about',methods=['GET'])
def about():
    return render_template('about.html')

@app.route('/details',methods=['GET'])
def details():
    return render_template('details.html')

@app.route("/predict", methods=['POST'])
def predict():
        if request.method == 'POST':
        # Extract features from form datails
        age = int(request.form['age'])
        sex = request.form['sex']
        bmi = float(request.form['bmi'])
        children = int(request.form['children'])
        smoker = request.form['smoker']
        region = request.form['region']

        # Encode categorical variables
        sex_encoded = 1 if sex == 'male' else 0
        smoker_encoded = 1 if smoker == 'yes' else 0
        region_encoded = {
            'northeast': 0,
            'northwest': 1,
            'southeast': 2,
            'southwest': 3
        }.get(region, -1)

        # Ensure region is valid
        if region_encoded == -1:
```

```
        return render_template('predict.html', pred='Invalid region specified!')

    # Create feature array
    features = np.array([[age, sex_encoded, bmi, children, smoker_encoded, region_encoded]])

    # Predict using the model
    prediction = model.predict(features)[0]

    if prediction < 0:
        return render_template('predict.html', prediction='Error calculating Amount!')
    else:
        return render_template('predict.html', prediction='Expected amount is
{0:.3f}'.format(prediction))

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

CODE SNIPPETS

MODEL BUILDING

Importing The Libraries

```
[ ] import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn  as sns
```

Read The Dataset

```
[ ] df=pd.read_csv("/content/insurance .csv")
```

```
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
[ ] df.tail()
```

	age	sex	bmi	children	smoker	region	charges
1333	50	male	30.97	3	no	northwest	10600.5483
1334	18	female	31.92	0	no	northeast	2205.9808
1335	18	female	36.85	0	no	southeast	1629.8335
1336	21	female	25.80	0	no	southwest	2007.9450
1337	61	female	29.07	0	yes	northwest	29141.3603

```
df.columns
```

```
Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'], dtype='object')
```

```
df.shape
```

```
(1338, 7)
```

```
df.dtypes
```

```
age           int64
sex          object
bmi         float64
children      int64
smoker       object
region       object
charges     float64
dtype: object
```

```
df.nunique()
```

```
age           47
sex            2
bmi          548
children       6
smoker         2
region         4
charges     1337
dtype: int64
```

```
df["charges"].max()
```

```
63770.42801
```

Stastical Analysis

```
numeric_columns = ['age', 'bmi', 'children', 'charges']
```

```
mean = df[numeric_columns].mean()
mean
```

```
age            39.207025
bmi            30.663397
children        1.094918
charges     13270.422265
dtype: float64
```

```
median = df[numeric_columns].median()
median
```

```
age           39.000
bmi           30.400
children       1.000
charges     9382.033
dtype: float64
```

```
df.mode()
```

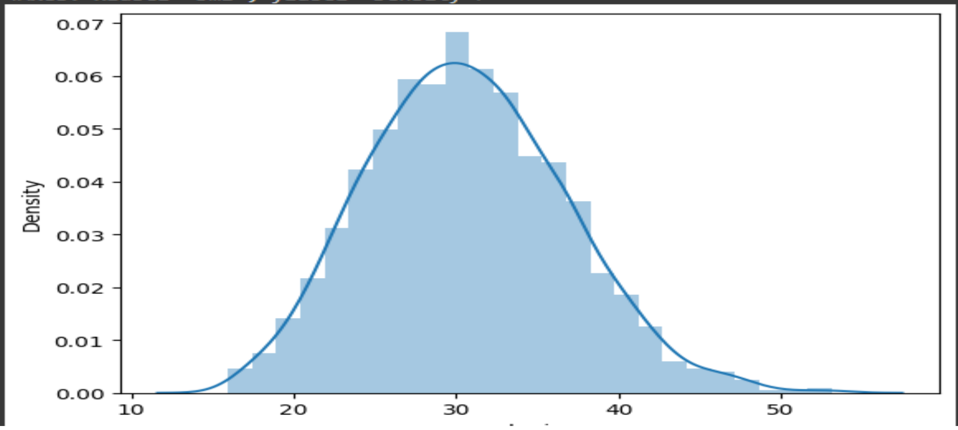
	age	sex	bmi	children	smoker	region	charges
0	18	male	32.3	0	no	southeast	1639.5631

Data Visualization

Univariate Analysis

```
sns.distplot(df['bmi'])
```

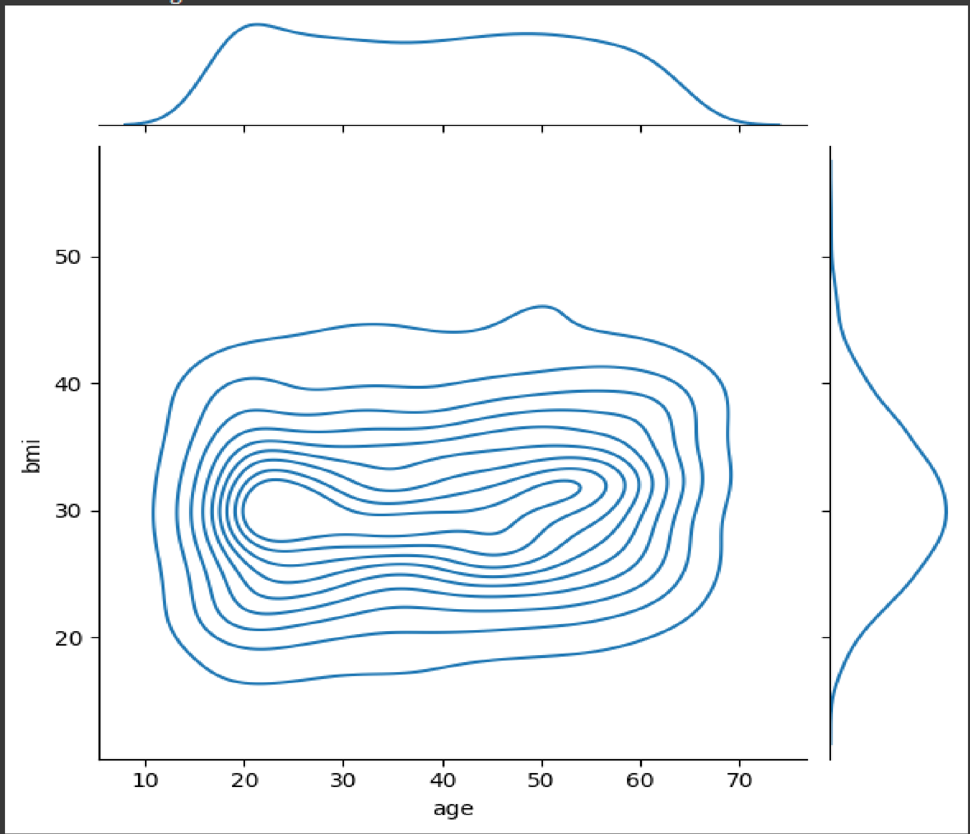
```
<ipython-input-18-a1d1d0e2654e>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['bmi'])
<Axes: xlabel='bmi', ylabel='Density'>
```



Bivariate Analysis

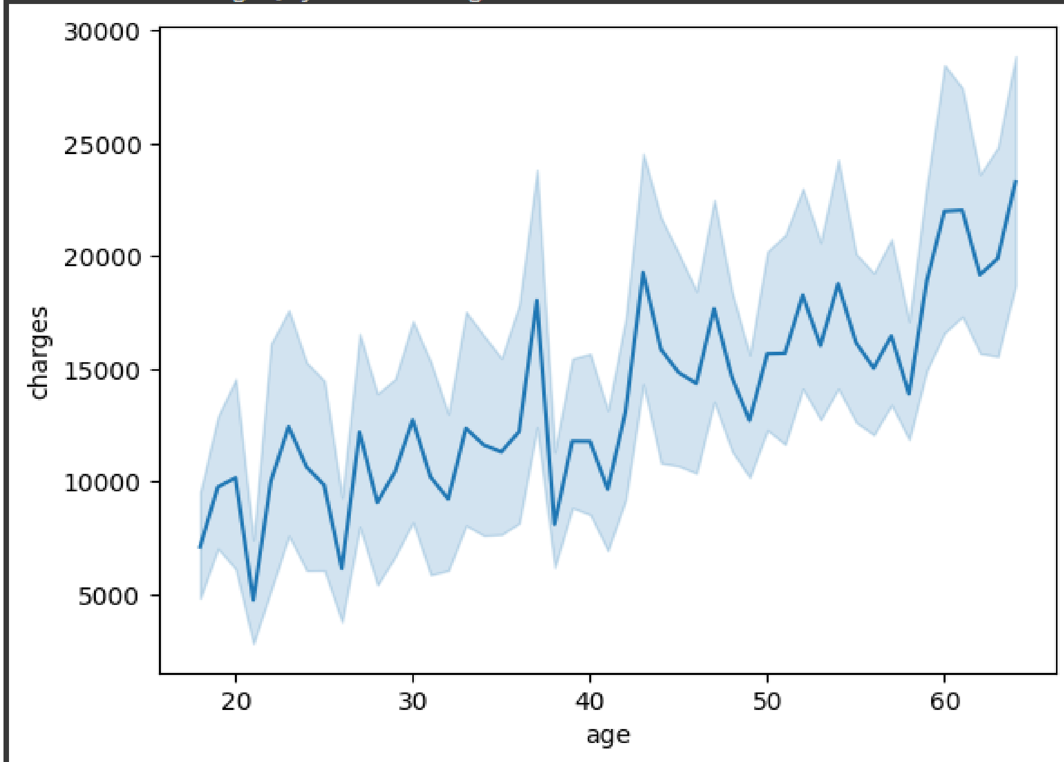
```
sns.jointplot(x=df['age'], y=df['bmi'], kind="kde")
```

```
<seaborn.axisgrid.JointGrid at 0x7a570b1e1f30>
```



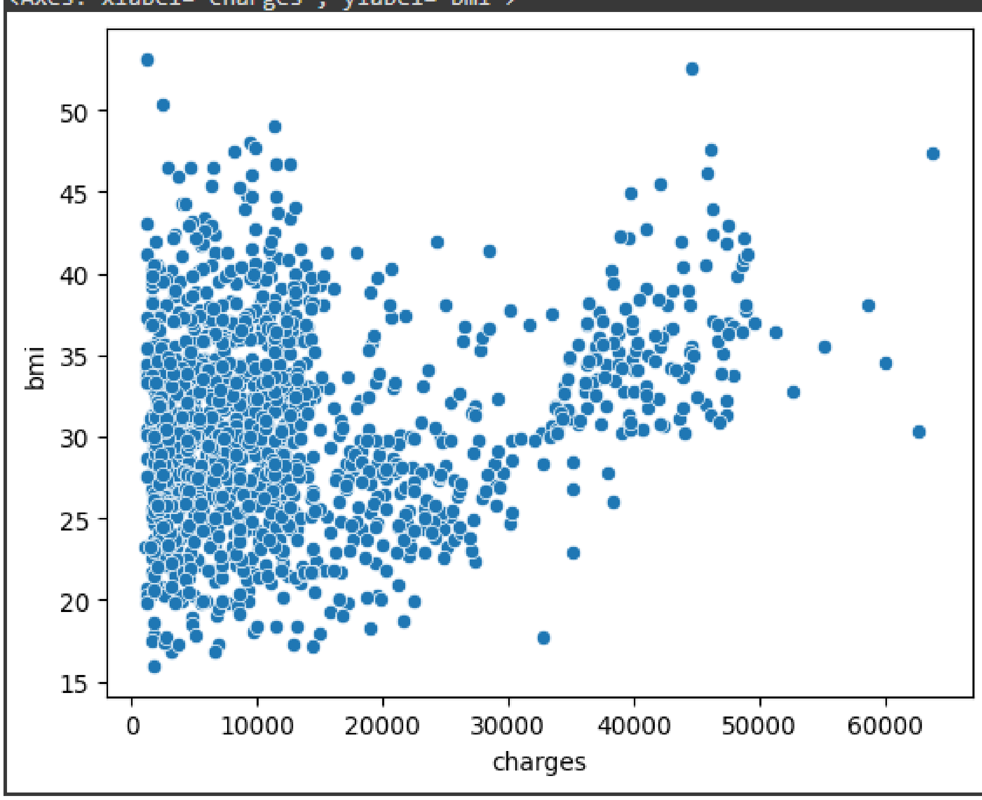
```
sns.lineplot(x=df['age'], y=df['charges'])
```

<Axes: xlabel='age', ylabel='charges'>




```
sns.scatterplot(x='charges', y='bmi', data=df)
```

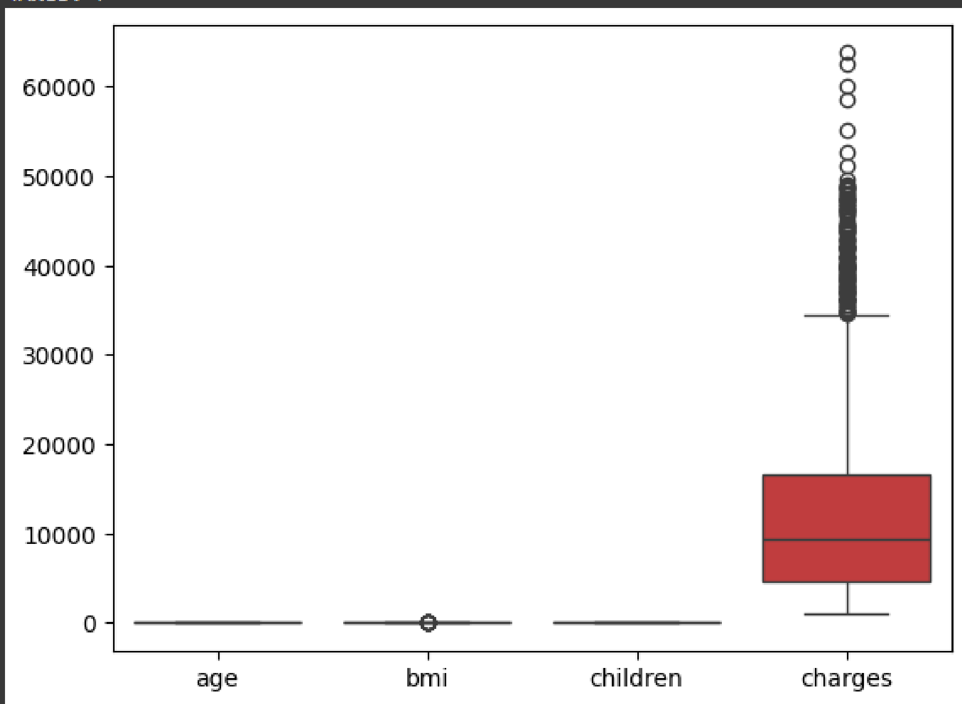
<Axes: xlabel='charges', ylabel='bmi'>



Detecting and Handling Outliers

```
sns.boxplot(df)
```

```
<Axes: >
```




```
IQR = df['bmi'].quantile(0.75)-df['bmi'].quantile(0.25)
IQR
```

```
8.3975
```

```
lowerBound=df['bmi'].quantile(0.25)-(1.5*IQR)
lowerBound
```

```
13.7
```

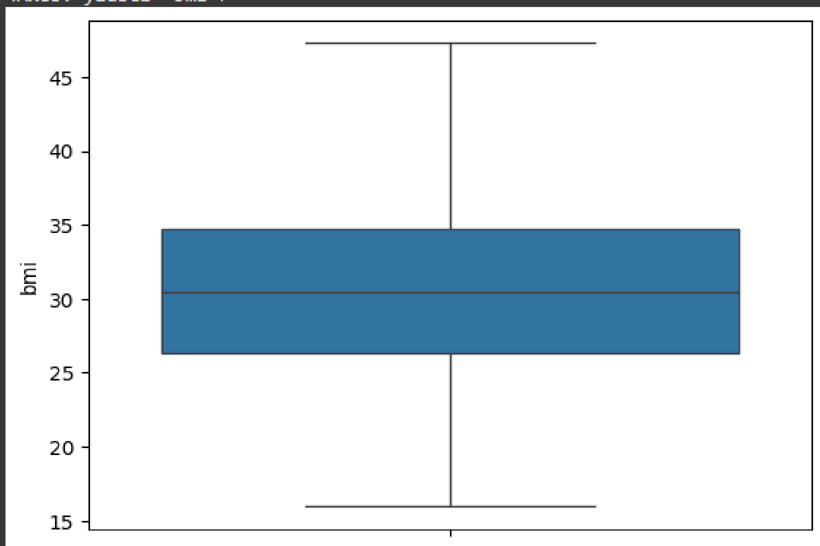
```
upperBound=df['bmi'].quantile(0.75)+(1.5*IQR)
upperBound
```

```
47.290000000000006
```

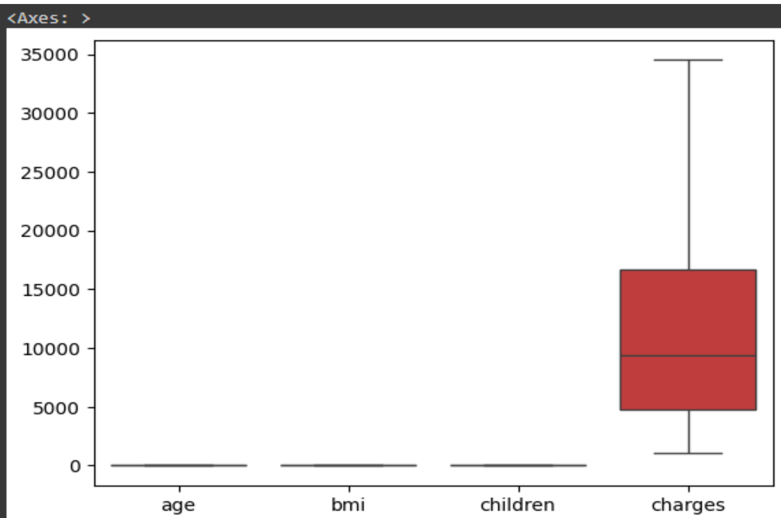
```
df['bmi']=np.where(df['bmi']>upperBound,upperBound,df['bmi'])
df['bmi']=np.where(df['bmi']<lowerBound,lowerBound,df['bmi'])
```

```
sns.boxplot(df['bmi'])
```

```
<Axes: ylabel='bmi'>
```




```
<Axes: >
```



Descriptive Analysis

```
df.describe()
```

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.650034	1.094918	12479.369251
std	14.049960	6.056926	1.205493	10158.056096
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000

Label Encoding

```
[ ] from sklearn.preprocessing import LabelEncoder
```

```
[ ] label_encoder = LabelEncoder()
```

```
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
[ ] df['sex'] = label_encoder.fit_transform(df['sex'])
    df['smoker'] = label_encoder.fit_transform(df['smoker'])
    df['region'] = label_encoder.fit_transform(df['region'])
```

```
[ ] df.head()
```

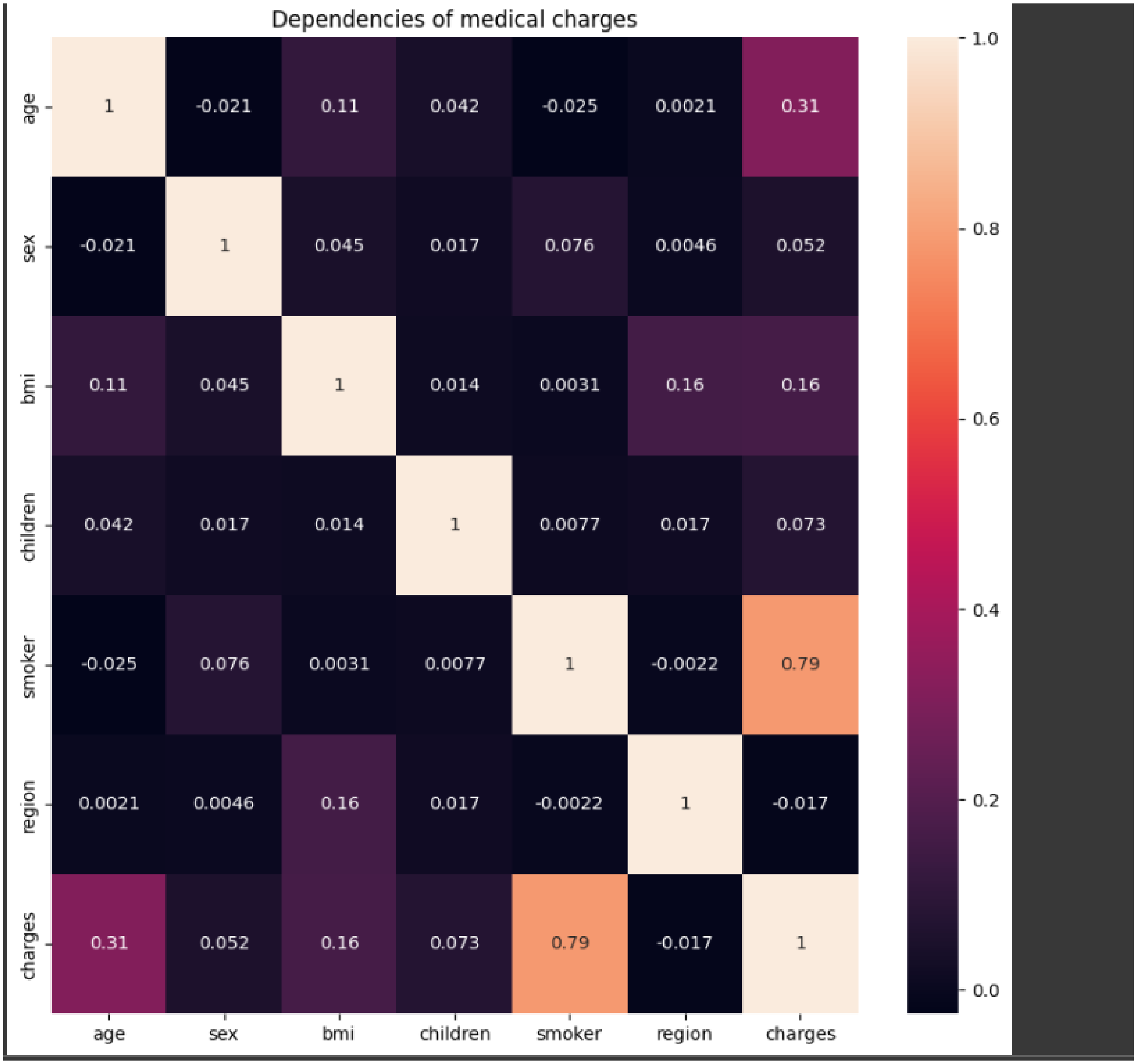
	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520

```
df
```

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520
...	...	...	...	...	...	...	...
1333	50	1	30.970	3	0	1	10600.54830
1334	18	0	31.920	0	0	0	2205.98080
1335	18	0	36.850	0	0	2	1629.83350
1336	21	0	25.800	0	0	3	2007.94500
1337	61	0	29.070	0	1	1	29141.36030

1338 rows × 7 columns

```
[ ]  corr = df.corr()

     fig, ax = plt.subplots(figsize=(10, 10))
     sns.heatmap(corr, annot=True, ax=ax)
     plt.title("Dependencies of medical charges")
     plt.show()
```



Splitting of data

```
X=df.drop(['charges'],axis=1)
```

```
y=df['charges']
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
y_train
```

```
560      9193.838500
1285     8534.671800
1142    27117.993780
969      8596.827800
486     12475.351300
            ...
1095     4561.188500
1130     8582.302300
1294    11931.125250
860     34489.350562
1126    10214.636000
Name: charges, Length: 1070, dtype: float64
```

Model Building

Linear Regression

```
from sklearn.linear_model import LinearRegression
```

```
lr= LinearRegression()
lr.fit(X_train,y_train)
```

```
▾ LinearRegression
LinearRegression()
```

```
y_pred1=lr.predict(X_test)
```

```
from sklearn import  metrics
```

```
score1=metrics.r2_score(y_test,y_pred1)
print(score1)
```

```
0.7837015388200166
```

```
s1=metrics.mean_absolute_error( y_test,y_pred1)
print(s1)
```

```
3320.557034987548
```

```
rmse_lr=np.sqrt(metrics.mean_squared_error(y_test,y_pred1))
print("mean_squared_error",rmse_lr)
```

```
mean_squared_error 4845.6792366495965
```

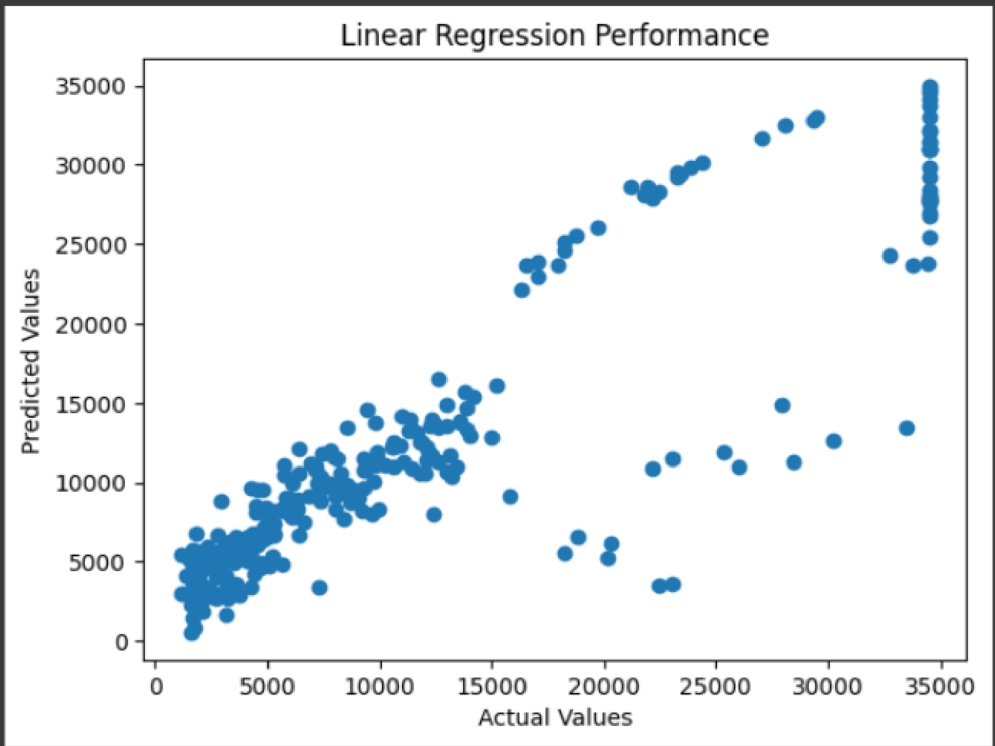
```
[ ] accuracy=lr.score(X_test,y_test)
    print("-------Linear Regression-------")
    print("model accuracy \t\t",accuracy)
    print(f'Accuracy in percentage\t:{accuracy:.1%}')
```

```
-------Linear Regression-------
model accuracy          0.7837015388200166
Accuracy in percentage  :78.4%
```

```
[ ] from sklearn.metrics import mean_squared_error, r2_score # Import appropriate metrics for regression
    from sklearn.model_selection import train_test_split
    print("Regression Metrics:")
    print("Mean absolute Error:", s1)
    print("Root Mean Squared Error:", rmse_lr)
    score1=metrics.r2_score(y_test,y_pred1)
    print("R-squared:",score1)
```

```
Regression Metrics:
Mean absolute Error: 3320.557034987548
Root Mean Squared Error: 4845.6792366495965
R-squared: 0.7837015388200166
```

```
import matplotlib.pyplot as plt

plt.scatter(y_test, y_pred1)
plt.xlabel("Actual Values")
plt.ylabel("Predicted Values")
plt.title("Linear Regression Performance")
plt.show()
```



Support Vector MachineRegressor

```
[ ] from sklearn.svm import SVR
```

```
[ ] svm= SVR()
    svm.fit(X_train,y_train)
```

```
▾ SVR
SVR()
```

```
[ ] y_pred2=svm.predict(X_test)
```

```
[ ] score2=metrics.r2_score(y_test,y_pred2)
    print(score2)
```

```
-0.057306433750309305
```

```
[ ] s2=metrics.mean_absolute_error( y_test,y_pred2)
    print(s2)
```

```
7754.513457705959
```

```
[ ] rmse_svm=np.sqrt(metrics.mean_squared_error(y_test,y_pred2))
    print("root_mean_squared_error",rmse_svm)
```

```
root_mean_squared_error 10713.4262641038
```

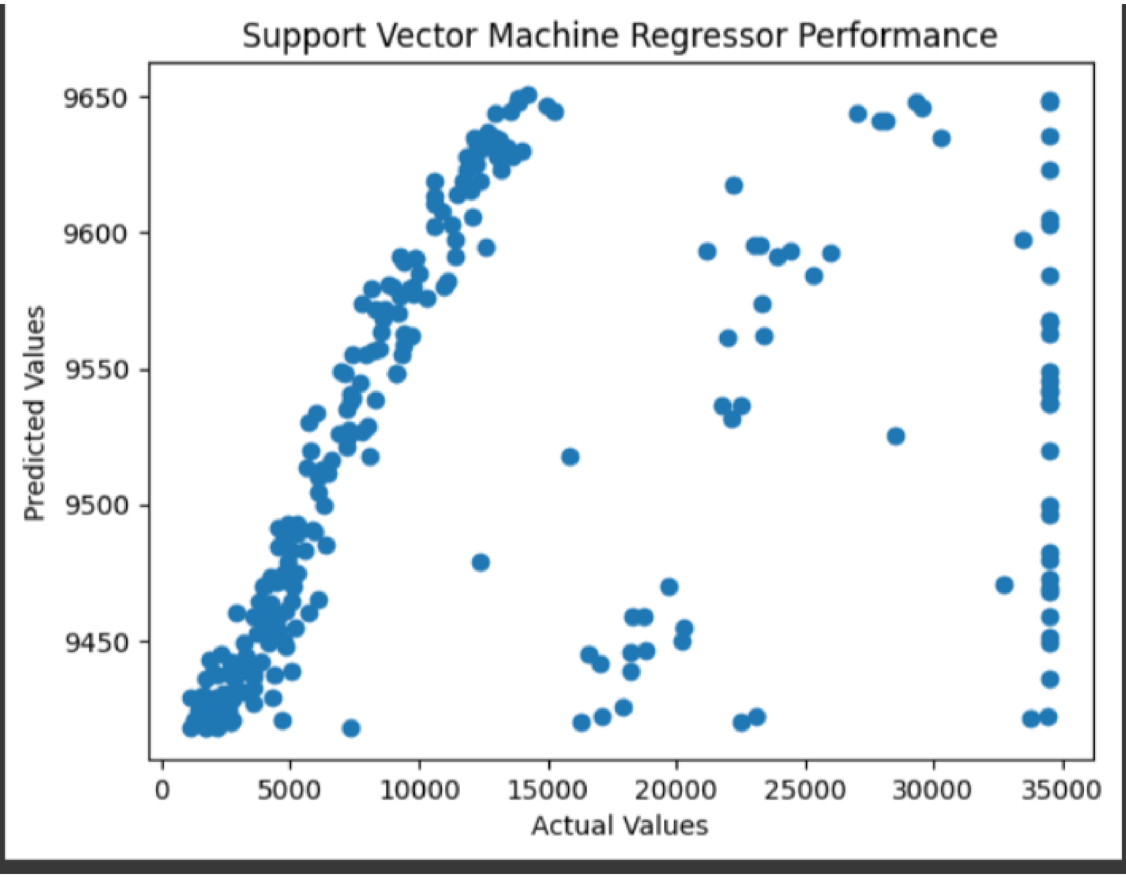
```
accuracy=svm.score(X_test,y_test)
print("-------Support Vector Machine-------")
print("model accuracy \t\t",accuracy)
print(f'Accuracy in percentage\t:{accuracy:.1%}')
```

```
-------Support Vector Machine-------
model accuracy           -0.057306433750309305
Accuracy in percentage  :-5.7%
```

```
[ ] from sklearn.svm import SVR
    from sklearn.metrics import mean_squared_error, r2_score # Import appropriate metrics for regression
    from sklearn.model_selection import train_test_split
    print("Regression Metrics:")
    print("Mean absolute Error:", s2)
    print("Root Mean Squared Error:", rmse_svm)
    score2=metrics.r2_score(y_test,y_pred2)
    print("R-squared:",score2)
```

```
Regression Metrics:
Mean absolute Error: 7754.513457705959
Root Mean Squared Error: 10713.4262641038
R-squared: -0.057306433750309305
```

```
import matplotlib.pyplot as plt

plt.scatter(y_test, y_pred2)
plt.xlabel("Actual Values")
plt.ylabel("Predicted Values")
plt.title("Support Vector Machine Regressor Performance")
plt.show()
```



RandomForest Regressor

```
[ ] from sklearn.ensemble import RandomForestRegressor
```

```
[ ] rf= RandomForestRegressor()
    rf.fit(X_train,y_train)
```

```
▾ RandomForestRegressor
RandomForestRegressor()
```

```
[ ] y_pred3=rf.predict(X_test)
```

```
score3=metrics.r2_score(y_test,y_pred3)
print(score3)
```

```
0.8302918166174308
```

```
[ ] s3=metrics.mean_absolute_error( y_test,y_pred3)
    print(s3)
```

```
2158.311786770744
```

```
[ ] rmse_rf=np.sqrt(metrics.mean_squared_error(y_test,y_pred3))
    print("root_mean_squared_error",rmse_rf)
```

```
root_mean_squared_error 4292.193966762153
```

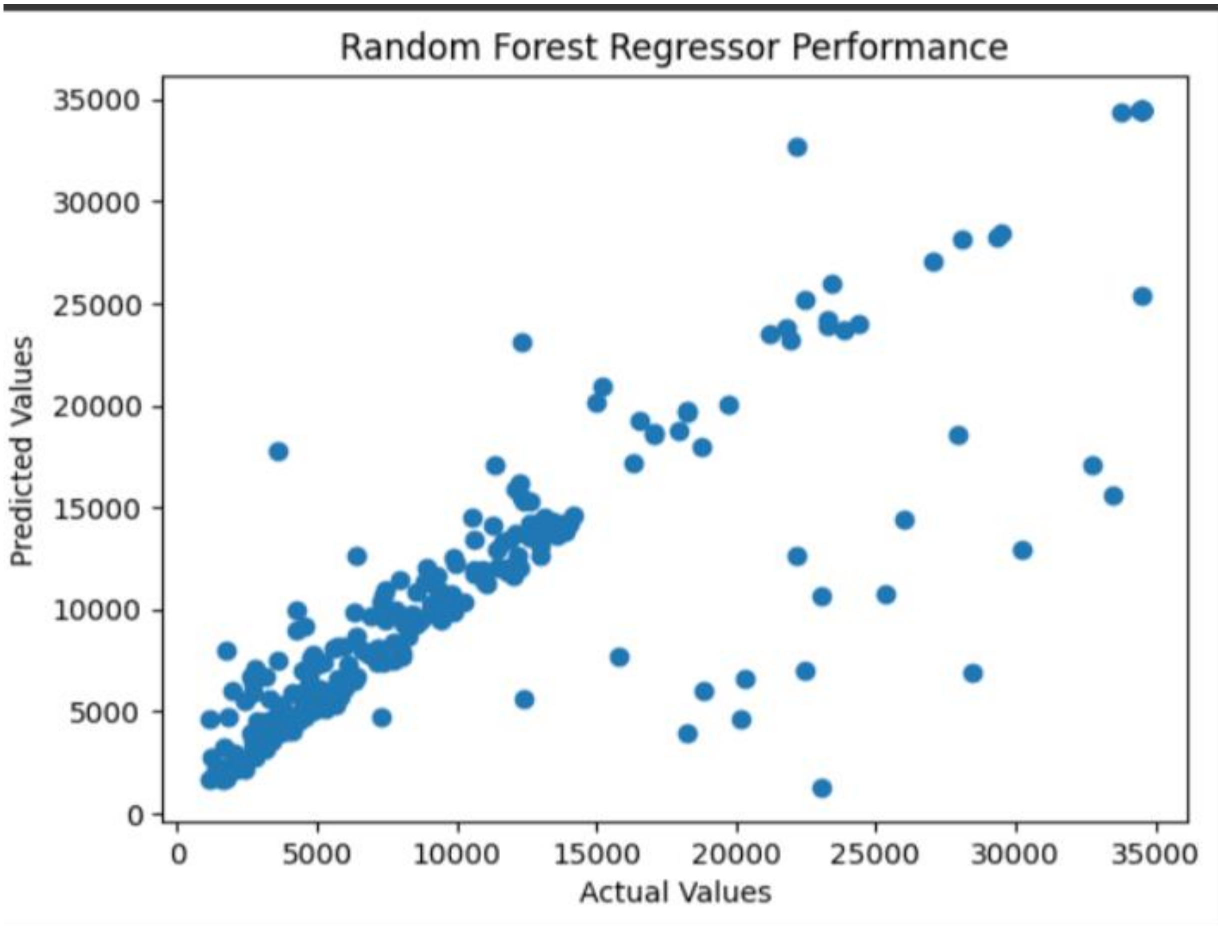
```
[ ] accuracy=rf.score(X_test,y_test)
    print("-------RandomForestRegressor-------")
    print("model accuracy \t\t",accuracy)
    print(f'Accuracy in percentage\t:{accuracy:.1%}')
```

```
-------RandomForestRegressor-------
model accuracy          0.8302918166174308
Accuracy in percentage  :83.0%
```

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score # Import appropriate metrics for regression
from sklearn.model_selection import train_test_split
print("Regression Metrics:")
print("Mean absolute Error:", s3)
print("Root Mean Squared Error:", rmse_rf)
score3=metrics.r2_score(y_test,y_pred3)
print("R-squared:", score3)
```

```
Regression Metrics:
Mean absolute Error: 2158.311786770744
Root Mean Squared Error: 4292.193966762153
R-squared: 0.8302918166174308
```

```
import matplotlib.pyplot as plt

plt.scatter(y_test, y_pred3)
plt.xlabel("Actual Values")
plt.ylabel("Predicted Values")
plt.title("Random Forest Regressor Performance")
plt.show()
```



GradientBoostingRegressor

```
from sklearn.ensemble import GradientBoostingRegressor
```

```
gb= GradientBoostingRegressor()
gb.fit(X_train,y_train)
```

```
▾ GradientBoostingRegressor
GradientBoostingRegressor()
```

```
y_pred4=gb.predict(X_test)
```

```
score4=metrics.r2_score(y_test,y_pred4)
print(score4)
```

```
0.8451154840835637
```

```
s4=metrics.mean_absolute_error( y_test,y_pred4)
print(s4)
```

```
2174.9371457221414
```

```
rmse_gb=np.sqrt(metrics.mean_squared_error(y_test,y_pred4))
print("root_mean_squared_error",rmse_gb)
```

```
root_mean_squared_error 4100.4540432147405
```

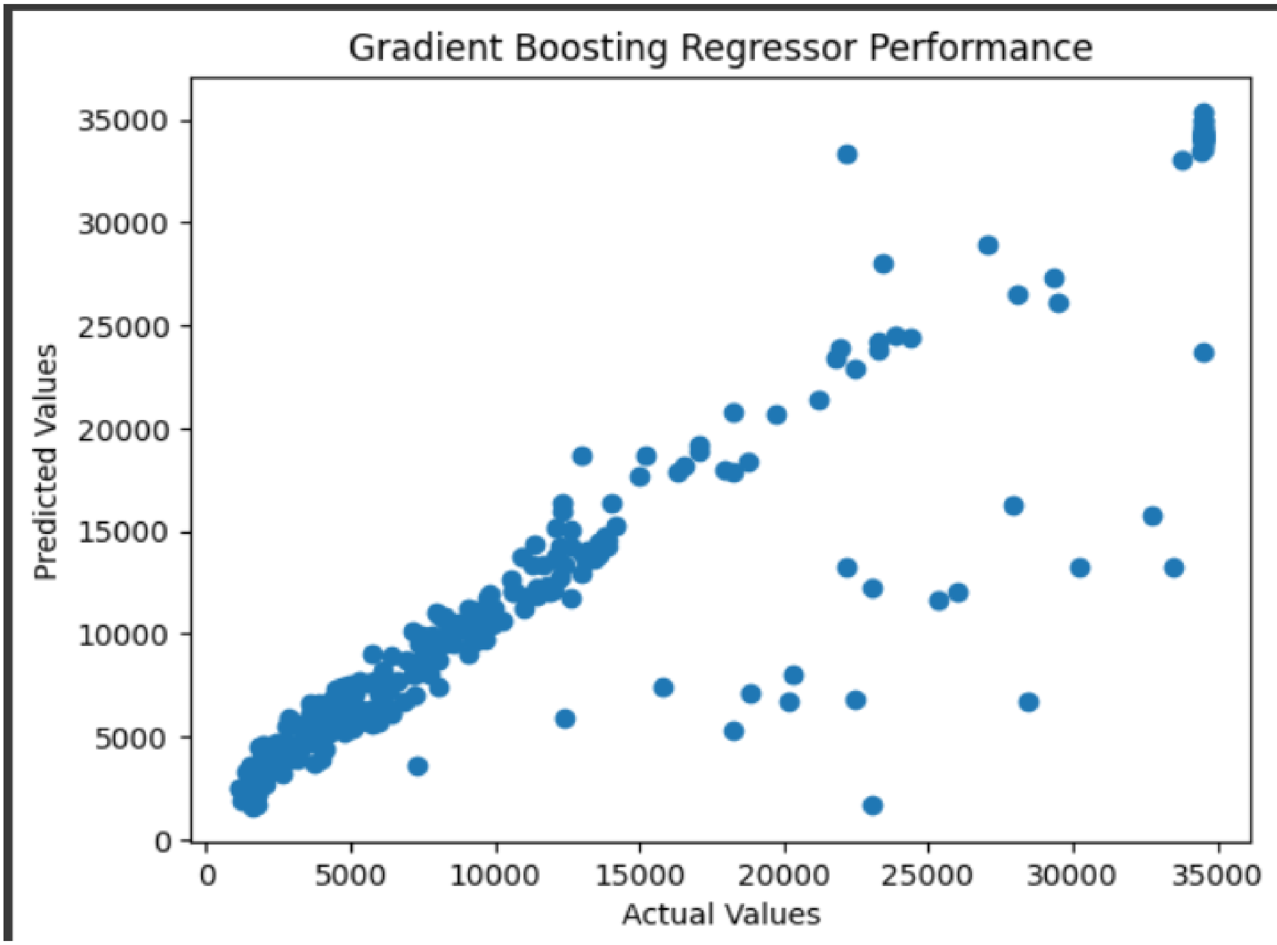
```
accuracy=gb.score(X_test,y_test)
print("-------GradientBoostingRegressor-------")
print("model accuracy \t\t",accuracy)
print(f'Accuracy in percentage\t:{accuracy:.1%}')
```

```
-------GradientBoostingRegressor-------
model accuracy          0.8451154840835637
Accuracy in percentage  :84.5%
```

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error, r2_score # Import appropriate metrics for regression
from sklearn.model_selection import train_test_split
print("Regression Metrics:")
print("Mean absolute Error:", s4)
print("Root Mean Squared Error:", rmse_gb)
score4=metrics.r2_score(y_test,y_pred4)
print("R-squared:", score4)
```

```
Regression Metrics:
Mean absolute Error: 2174.9371457221414
Root Mean Squared Error: 4100.4540432147405
R-squared: 0.8451154840835637
```

```
import matplotlib.pyplot as plt

plt.scatter(y_test, y_pred4)
plt.xlabel("Actual Values")
plt.ylabel("Predicted Values")
plt.title("Gradient Boosting Regressor Performance")
plt.show()
```



Model Selection

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
# Assuming X and y are your feature and target data respectively,
# which are not provided in the problem
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

rf = RandomForestRegressor()
rf.fit(X_train, y_train)
```

```
▾ RandomForestRegressor
RandomForestRegressor()
```

```
[ ] df={'age':19,'sex':0,'bmi':27.9,'children':0,'smoker':1,'region':3}
    df=pd.DataFrame(df,index=[0])
    df
```

	age	sex	bmi	children	smoker	region
0	19	0	27.9	0	1	3

```
[ ] new_pred=rf.predict(df)
    print(new_pred)
```

```
[17141.5437415]
```

Saving the Model

```
[ ] import pickle
    import warnings
```

```
with open("rf.pkl","wb")as f:
  pickle.dump(rf,f)
```