HEART DISEASE ANALYSIS

Project Report

Team ID: LTVIP2026TMIDS89151

1. INTRODUCTION

1.1 Project Overview

The Heart Disease Analysis project aims to develop a comprehensive platform for examining patient health records and estimating the probability of heart disease. By combining data analytics, interactive visualizations, and predictive modeling techniques, the system supports healthcare professionals in efficiently evaluating risk factors and making well-informed clinical decisions.

The solution consolidates patient demographic details, diagnostic test results, and lifestyle attributes to compute a structured Risk Score. These insights are then presented through dynamic dashboards, enabling clear interpretation and actionable decision-making in healthcare settings.

1.2 Purpose

The main objective of this project is to enhance early diagnosis and preventive management for individuals vulnerable to heart disease. By utilizing data analytics and predictive modeling techniques, the project intends to:

- Assist healthcare providers in rapidly identifying patients with elevated risk levels.
- Present patient information through clear visual dashboards for better understanding and analysis.
- Minimize delayed diagnosis by emphasizing critical risk factors and warning signs.
- Function as a clinical decision support system for hospitals, clinics, and healthcare monitoring organizations.

2. IDEATION PHASE

2.1 Problem Statement

Heart disease is one of the leading causes of mortality worldwide. Early detection and intervention significantly improve patient outcomes. However, most hospitals face challenges such as **manual data analysis, delayed diagnosis, and scattered patient records**. The project addresses these challenges by providing a **centralized analytical platform** that processes patient data and predicts heart disease risk accurately.

2.2 Empathy Map Canvas

- **Think & Feel:** Patients often experience stress or uncertainty regarding their heart health, while doctors require dependable analytical tools to support precise clinical decisions.
- **Hear:** Guidance from healthcare professionals, diagnostic reports, lab findings, and recommendations from colleagues or medical peers.
- **See:** Numerous patient records, diagnostic charts, lab result summaries, and detailed medical histories.
- **Say & Do:** Patients describe symptoms and concerns; doctors manually review reports and interpret data for diagnosis.
- **Pain Points:** Manual analysis consumes time, increases the possibility of human error, and may delay early detection of heart disease.
- **Gain Points:** Faster risk evaluation, intuitive visual dashboards, and dependable predictive results that support proactive treatment.

2.3 Brainstorming

- Apply predictive analytics techniques to estimate heart disease risk accurately.
- Develop interactive dashboards for effective data visualization and exploration.
- Integrate dynamic filters for parameters such as age, gender, blood pressure, cholesterol, and other clinical factors.
- Generate an automated Risk Score calculation for each patient profile.
- Enable story-based reporting and summary reports for hospital management and administrative decision-making.

3. REQUIREMENT ANALYSIS

3.1 Customer Journey Map

1. Patient data is collected during routine check-ups.
2. Data is uploaded to the Heart Disease Analysis system.
3. The system preprocesses and cleans the data.
4. Interactive dashboards allow doctors to filter and visualize patient data.
5. Risk Score is calculated and displayed for each patient.
6. Reports are generated for high-risk patients to aid in preventive measures.

3.2 Solution Requirement

- **Functional Requirements:** Data upload, preprocessing, Risk Score calculation, dashboard visualization, filtering, reporting.
- **Non-Functional Requirements:** Accuracy of predictions, fast response time (<3 seconds), user-friendly interface, secure data handling.

3.3 Data Flow Diagram

- **Input:** Patient demographic and medical data.
- **Processing:** Data cleaning, normalization, risk score calculation, visualization.
- **Output:** Dashboards, reports, and notifications for high-risk patients.

3.4 Technology Stack

- **Frontend/Visualization:** Tableau, Power BI
- **Backend/Processing:** Python (pandas, NumPy), Jupyter Notebook
- **Database:** MySQL / CSV datasets
- **Libraries:** scikit-learn for predictive modeling, matplotlib/seaborn for graphs

4. PROJECT DESIGN

4.1 Problem Solution Fit

The proposed system effectively tackles major challenges in early heart disease identification by consolidating patient data into a unified platform, automating risk evaluation processes, and delivering meaningful visual insights. This approach enhances clinical decision-making and supports timely, evidence-based healthcare interventions.

4.2 Proposed Solution

Data Upload Module: Enables users to import patient data through CSV or Excel file formats with validation checks to ensure correct structure and format.

• **Data Preprocessing Module:** Cleans and prepares the dataset by managing missing values, detecting outliers, and applying normalization or scaling techniques.

• **Visualization Dashboard Module:** Offers interactive filters, dynamic charts, KPIs, and automatically computed Risk Scores for detailed patient analysis.

• **Report Generation Module:** Produces structured reports in PDF format or Tableau Story dashboards to support hospital management and administrative review.

4.3 Solution Architecture

• **Data Layer**: Responsible for storing and managing patient information, including demographics, clinical measurements, and medical history in a structured database.

• **Processing Layer:** Python-based scripts and predictive models compute the Risk Score using predefined formulas and machine learning algorithms to assess heart disease probability.

• **Presentation Layer:** Interactive Tableau dashboards visualize patient data through charts, graphs, KPIs, and dynamic filters for easy analysis by healthcare professionals.

• **Reporting Layer:** Generates automated reports and summary insights, particularly highlighting high-risk patients for timely medical intervention and decision-making.

5. PROJECT PLANNING & SCHEDULING

5.1 Project Planning

The **planning phase** is critical to ensure the timely completion of the Heart Disease Analysis project. A detailed schedule was created to allocate tasks, track progress, and manage resources effectively. The project was divided into multiple phases, each with defined objectives, deliverables, and timelines:

Phase 1 – Requirement Gathering (2 Weeks):

o Conducted discussions with doctors and healthcare staff to identify system expectations and clinical needs.

o Defined core functional requirements such as patient data upload, automated Risk Score computation, and report generation.

o Specified non-functional requirements including performance efficiency, data security, and user-friendly interface design.

o **Outcome:** Finalized and approved Requirement Specification Document.

Phase 2 – Data Acquisition & Cleaning (1 Week):

o Gathered patient datasets from public healthcare sources and sample hospital databases.

o Detected and handled missing values, anomalies, and inconsistent data entries.

o Applied preprocessing methods such as data normalization, scaling, and encoding of categorical variables.

o **Outcome:** Structured and cleaned dataset prepared for modeling and visualization.

Phase 3 – Dashboard & Predictive Model Implementation (2 Weeks):

o Developed interactive dashboards in Tableau to visualize patient metrics and calculated Risk Scores.

o Built predictive models in Python using algorithms like Random Forest and Logistic Regression for heart disease prediction.

o Integrated filters, calculated fields, and interactive components to support dynamic data exploration by healthcare professionals.

o **Outcome:** Fully operational dashboards linked with processed datasets and predictive outputs.

Phase 4 – System Testing & Performance Validation (1 Week):

o Performed functional testing to verify proper operation of dashboards, filters, and scoring calculations.

o Carried out performance assessments to measure response time under large datasets and simultaneous user access.

o **Outcome:** System validated and approved for User Acceptance Testing (UAT).

Phase 5 – Reporting & Documentation (1 Week):

o Generated detailed reports and Tableau Story visualizations highlighting high-risk patient insights.

o Prepared comprehensive documentation including technical reports, user guides, and testing summaries.
o **Outcome:** Complete project package finalized for submission and stakeholder review.

Project Management Tools Used:

Gantt charts and Trello boards were utilized to organize tasks, monitor dependencies, and track milestone completion. Weekly review meetings helped maintain progress and address potential challenges proactively.

6. FUNCTIONAL AND PERFORMANCE TESTING

6.1 Performance Testing

Performance testing was conducted to ensure the system works efficiently under real-world scenarios. The following aspects were tested:

- **Data Upload Testing:**
  - o Confirmed that the system successfully processed 500+ patient records without failures or noticeable lag.
  - o Monitored CPU performance and memory consumption to ensure dashboards stayed responsive.

- **Filter Validation Testing:**
  - o Applied combinations of filters such as age, gender, blood pressure, and cholesterol levels.
  - o Ensured that charts, visual elements, and risk score tables refreshed instantly.
  - o Verified that filtering operations maintained calculation accuracy and visual consistency.

- **Computation Testing:**
  - o Evaluated the Risk Score algorithm across numerous records to validate precision.
  - o Tested edge scenarios, including null values and extreme data inputs, to avoid incorrect outputs.

- **Performance Testing:**
  - o Assessed dashboard response time as the dataset size increased.
  - o Confirmed that all visual components loaded within 3 seconds to maintain a seamless user experience.

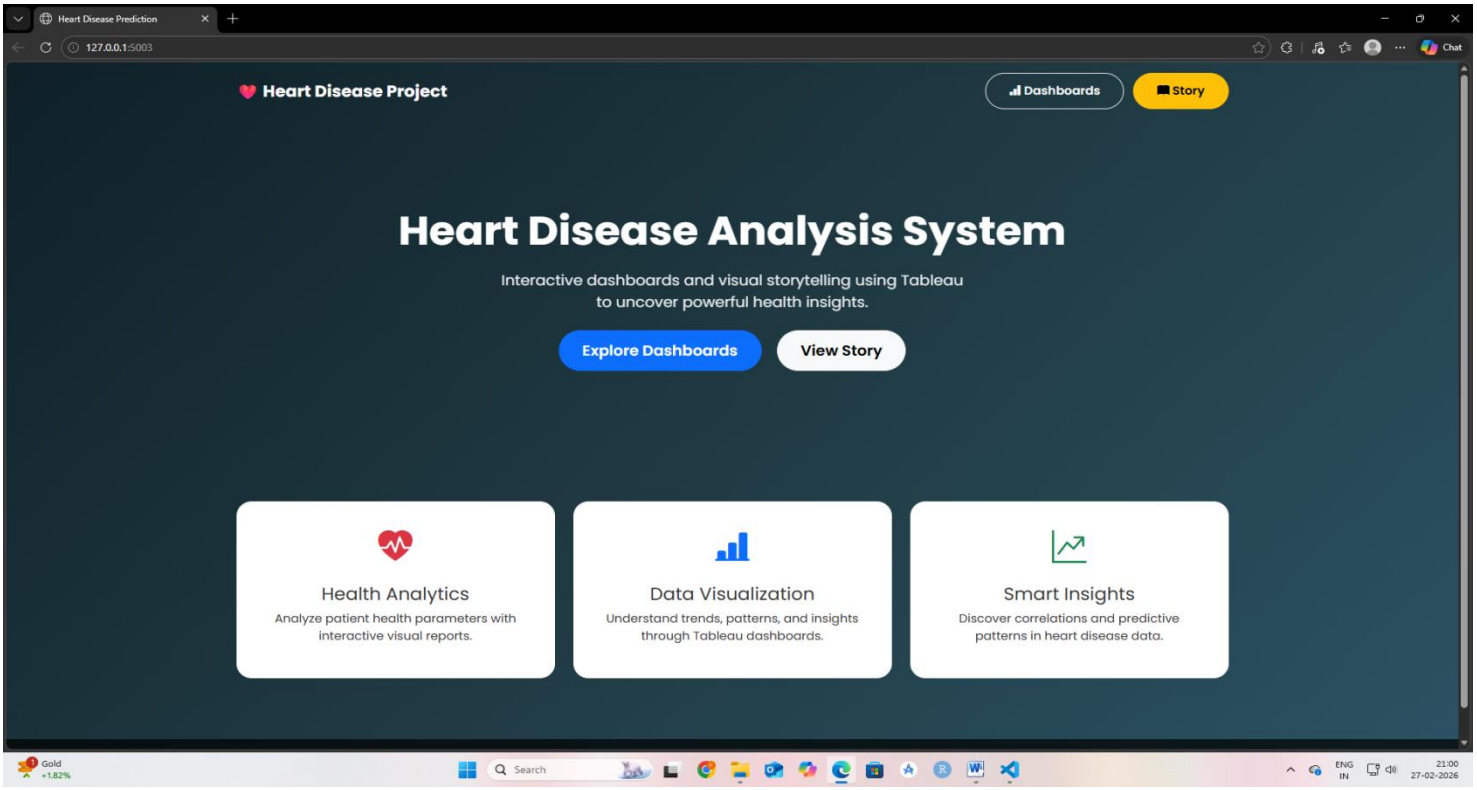
- **Load & Stability Testing:**
  - o Simulated concurrent access by multiple users on the dashboards.
  - o Verified system stability under heavy workload without crashes or slowdowns.

Result: All functional and performance evaluations were completed successfully, demonstrating that the system is stable, efficient, and prepared for User Acceptance Testing (UAT) deployment.

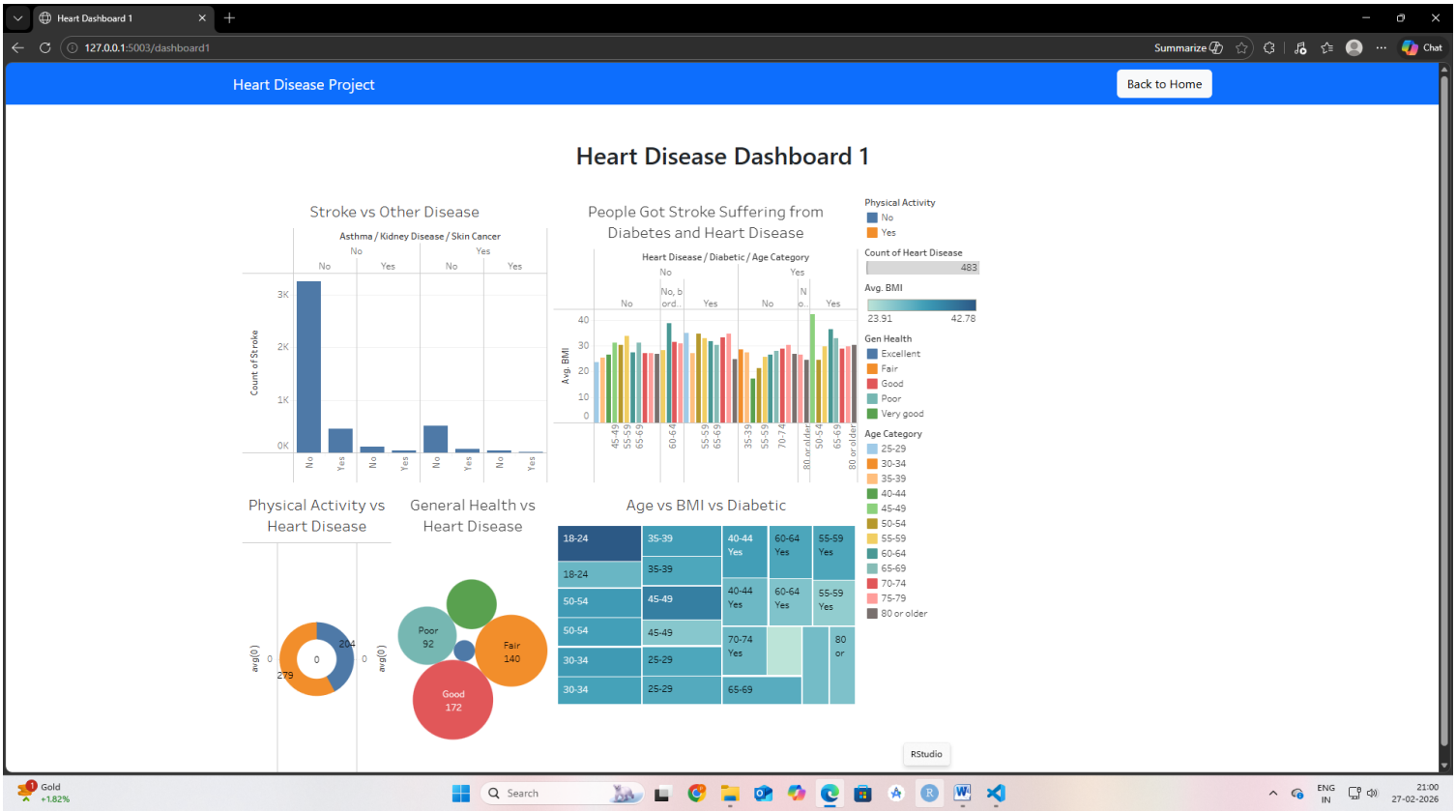
7. RESULTS

7.1 Output Screenshots

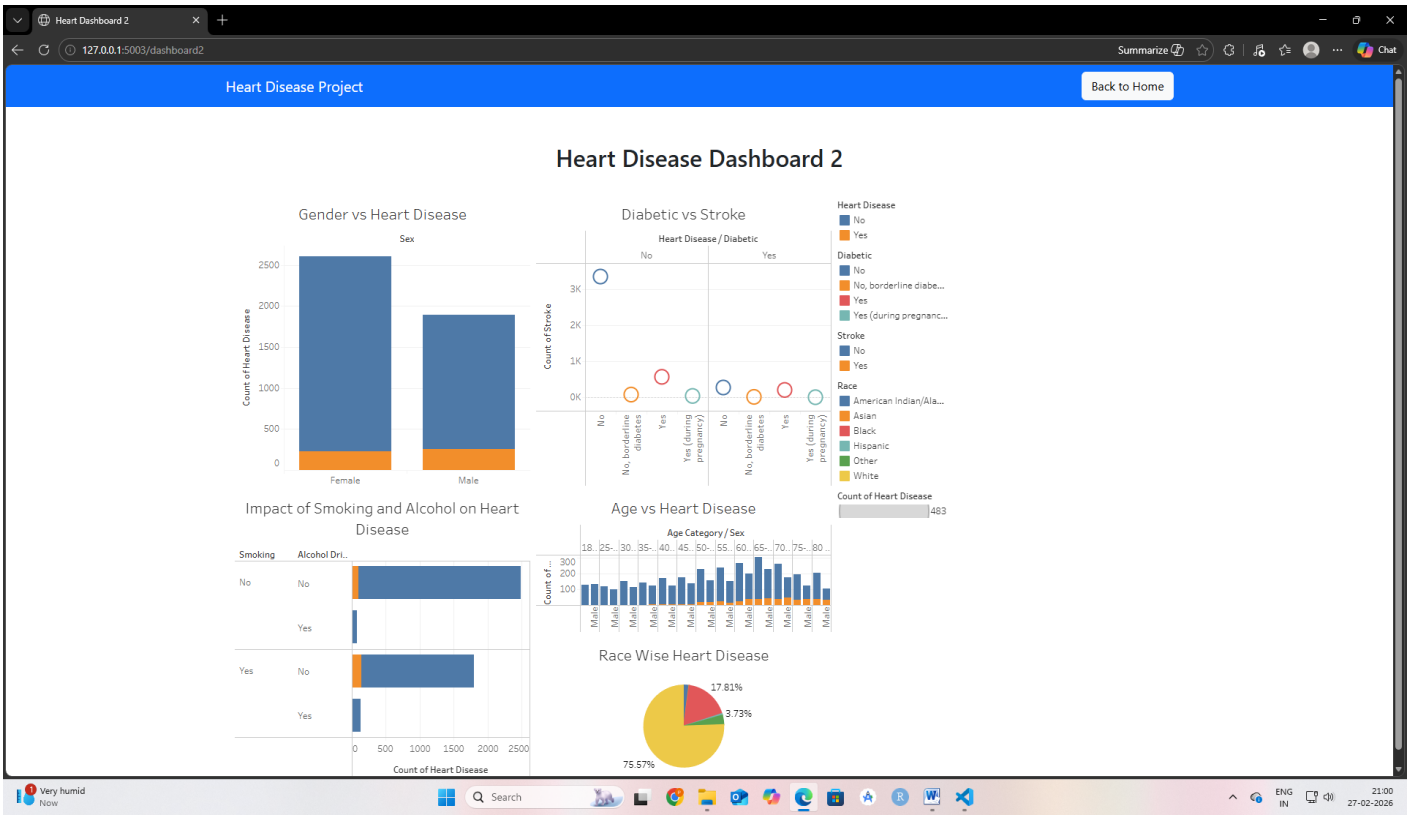
7.1.1 Home



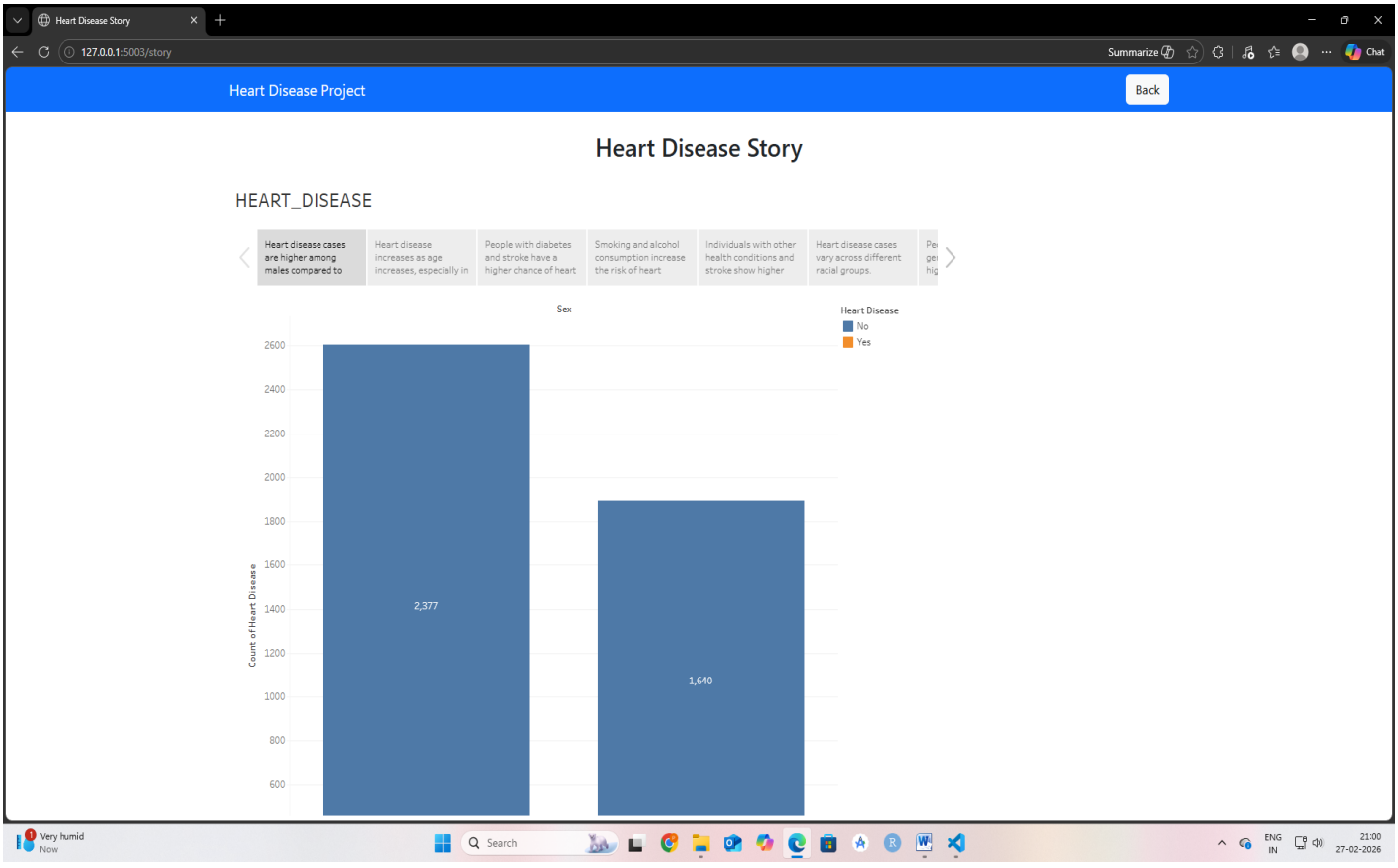
7.1.2 Dashboard



7.1.3 Dashboard



7.1.4 Story



7. ADVANTAGES & DISADVANTAGES

Advantages:

• Centralized storage of patient information enhances operational efficiency and data accessibility.

• Predictive algorithms effectively generate reliable risk assessment scores.

• Dynamic dashboards enable real-time analysis and visualization of patient records.

• Supports early diagnosis and preventive care, reducing potential health complications.

• Analytical reports and visual stories strengthen communication with management and stakeholders.

• Flexible and scalable architecture capable of supporting large datasets and multiple concurrent users.

Disadvantages:

• Users may require basic training to effectively navigate and interpret dashboard features.

• Model accuracy is highly dependent on the reliability and completeness of input data.

• Integration with existing hospital information systems may demand additional technical resources.

• Predictive models require regular updates and retraining to ensure sustained performance and accuracy.

8. CONCLUSION

The Heart Disease Analysis project effectively combines data analysis, predictive algorithms, and interactive visualizations within a unified platform. The system showcases:

- Reliable performance in generating accurate risk assessment scores.
- Comprehensive dashboards and detailed reports to support healthcare professionals.
- Smooth processing of large volumes of patient data with optimized response time.

In conclusion, the project successfully achieves its goal of enhancing early diagnosis and supporting data-driven medical decisions, highlighting its practical applicability in real-world healthcare environments.

9. FUTURE SCOPE

The Heart Disease Analysis System can be further improved in multiple ways:

• Integration with IoT-based health monitoring devices such as smartwatches and blood pressure monitors for continuous data collection.
• Implementation of advanced machine learning and deep learning algorithms for early risk detection and personalized treatment recommendations.
• Creation of a user-friendly mobile and web application to enable remote access for both patients and healthcare professionals.

• Extension of the system to cover additional chronic conditions like hypertension, obesity, and diabetes for comprehensive health evaluation.
• Connection with hospital management systems and electronic medical records (EMR) to ensure accurate and secure data exchange.
• Regular retraining of predictive models using updated patient datasets to enhance accuracy and long-term system performance.

10.APPENDIX

- **Source Code:**

  [https://github.com/pravalli710/Heart_Disease_Analysis]

- **Dataset Link:** https://drive.google.com/file/d/190Qmq27LeZZ_nWricP3Obl7ys_5otEsp/view
- **Project Demo Link:**
  https://drive.google.com/file/d/1NW-spRtW22-zb4Hj16MDJHi13F5I0m_L/view?usp=drivesdk

[Tableau Public Dashboard – Dashboard1-
https://public.tableau.com/views/heart_diseaseproject/Dashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link
Dashboard2-

https://public.tableau.com/views/heart_diseaseproject/Dashboard2?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link]

[Tableau public Story-

https://public.tableau.com/views/heart_diseaseproject/Story1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link]