Prototype Development and Business/Financial Modelling for Heart Disease Predictor App

Abburi Dhanusha, Dhanamaneni Varenya, Krishnendu Jana (Team Lead), Rishav Karmahapatra

Contents

Abburi Dhanusha's GitHub

https://github.com/DhanushaAbburi/HeartAttackAnalysis

Dhanamaneni Varenya's GitHub

https://github.com/varenyaa/heartdiseaseprediction_AIprototype/blob/main/Heartdisease_prediction.ipynb

Krishnendu Jana's GitHub

https://github.com/KrishnenduJana/ML-Model-Deployment

Rishav Karmahapatra's GitHub

https://github.com/Rishav19962/heart-attack-predictor/blob/main/Heart%20disease.ipynb

Abstract

Heart disease is a critical health concern affecting millions of individuals worldwide. Accurate prediction of heart disease is crucial for early diagnosis and effective treatment. This abstract presents a comprehensive dataset containing various patient factors that can aid in the prediction of heart disease. The dataset encompasses demographic variables such as age and sex, as well as clinical attributes including chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels coloured by fluoroscopy, and thalassemia.

Introduction

Heart disease is a prevalent and serious medical condition that affects millions of individuals worldwide. It encompasses a range of cardiovascular disorders, including coronary artery disease, heart failure, and arrhythmias. Timely diagnosis and accurate prediction of heart disease are crucial for effective treatment and prevention of adverse events.

Heart disease is a widespread health issue that poses a significant burden on healthcare systems globally. Accurate prediction of heart disease plays a vital role in early intervention and effective management of patients. In this report, we analyse a comprehensive dataset specifically designed for heart disease prediction to gain insights into the factors associated with the condition.

The dataset used in this analysis encompasses various patient-related factors that have been identified as potential indicators of heart disease. These factors include demographic information such as age and sex, as well as clinical measurements and indicators, covering a wide range of aspects related to cardiovascular health. The dataset provides a rich resource for investigating the relationships between these factors and the presence or absence of heart disease.

The primary objective of this analysis is to explore the dataset and uncover patterns and associations that can inform heart disease prediction. By conducting a thorough examination

of the dataset, we aim to identify the most relevant factors and understand their impact on the likelihood of heart disease occurrence. Such insights can help healthcare professionals in risk assessment, treatment planning, and preventive measures.

To accomplish our goal, we employ various analytical techniques, including statistical analysis and data visualization, to gain a comprehensive understanding of the dataset. Additionally, we use machine learning algorithms, such as logistic regression, to develop predictive models based on the available patient factors. By applying these models, we aim to evaluate the predictive power of the dataset and assess its potential in aiding heart disease diagnosis and prognosis.

The analysis presented in this report contributes to the field of heart disease research by providing valuable insights into the factors associated with the condition. The dataset serves as a valuable resource for researchers, clinicians, and policymakers in understanding the complexities of heart disease and developing effective strategies for prevention and treatment.

The subsequent sections of this report detail the methodology employed, the findings derived from the dataset analysis, and a discussion of the implications and potential applications of the results. By thoroughly exploring the dataset, we aim to contribute to the existing body of knowledge in heart disease research and support evidence-based decision-making in cardiovascular healthcare.

Problem Statement

The most advanced heart disease predictors currently available do not take into account risk factors and indicators unique to the Prototype Development, Business & Financial Modelling industries, which may influence those who work in these fields. This information gap makes it very difficult to recognize and evaluate the particular cardiac dangers posed by the professionals in the field. As a result, people working in these areas might not be aware of the dangers they face from heart disease and might not get the proper preventative care or prompt medical attention.

Prototype Selection

Here we have chosen one prototype idea that helps health sector with satisfying following criteria-

- The good or service could be created in the near future.
- The good or service must be useful or able to endure in the long run.
- Direct monetization should be possible for the good or service.

Prototype Development

Here we have developed a prototype for this project. Here we have deployed a model in a website which can predict a person have a heart disease of not.

Customer and Business Need Assessment

To develop an effective heart disease prediction solution, it is essential to conduct a customer and business need assessment to understand the specific requirements and challenges faced by patient. Here are some key aspects to consider during the assessment:

Customer Need Assessment

- Customers, including patients and healthcare professionals, require a heart disease prediction model that can provide accurate risk assessments. They want a reliable tool that can accurately predict the likelihood of developing heart disease based on individual factors such as medical history, lifestyle, and genetic predisposition.
- Customers understand the importance of early detection in managing heart disease. They seek a prediction model that can identify individuals at high risk of developing heart disease before symptoms appear. Early detection allows for timely interventions and lifestyle modifications that can reduce complications and improve outcomes.

Business Need assessment

- Businesses developing heart disease prediction models must prioritize accuracy and reliability. Healthcare providers and organizations require models that have been validated with high-quality data and demonstrate consistent performance in predicting heart disease risk. The model should be robust enough to handle diverse patient populations and provide reliable predictions.
- The prediction model should have a user-friendly interface for both patients and healthcare professionals. Clear and intuitive presentation of results, along with appropriate explanations and visualizations, can facilitate better understanding and interpretation of the predictions. A user-friendly interface will encourage adoption and make it easier for healthcare providers to integrate the model into their workflow.

Benchmarking Alternative Product

Benchmarking alternative products of a heart disease prediction model involves comparing and evaluating similar existing solutions in the market. Here are some potential alternative products that can be considered for benchmarking:

Framingham Risk Score (FRS)

The Framingham Risk Score is a widely used cardiovascular disease prediction model that estimates the 10-year risk of developing heart disease. It utilizes factors such as age, gender, blood pressure, cholesterol levels, smoking status, and diabetes to calculate the risk. Benchmarking against FRS can provide insights into its performance and identify areas of improvement.

QRISK

QRISK is another cardiovascular disease prediction algorithm commonly used in the United Kingdom. It takes into account additional risk factors such as ethnicity, body mass index (BMI), family history of heart disease, and socioeconomic status. Comparing a heart disease prediction model against QRISK can help assess its performance and determine if additional risk factors enhance predictive accuracy.

Commercially Available Predictive Models

Several commercial solutions offer heart disease prediction models that utilize different algorithms and datasets. Benchmarking against these commercial products can help evaluate their accuracy, usability, and integration capabilities, providing insights into market standards and customer expectations.

Business Model

A business model for a heart disease prediction model can be structured in various ways depending on the specific goals, target audience, and revenue generation strategies. Here is an example of a business model for a heart disease prediction model:

Target Customer

- Healthcare Providers: Target hospitals, clinics, and medical practices to integrate the heart disease prediction model into their workflow for risk assessment and personalized care.
- Health Insurance Companies: Collaborate with insurance companies to leverage the prediction model in risk stratification and offering personalized insurance plans for individuals based on their risk profiles.
- Research Institutions: Provide access to the prediction model for academic and research purposes, enabling further advancements in heart disease prediction and management.

Key Activities

- Model Development: Continuously refine and improve the heart disease prediction model based on research, updated data, and feedback from healthcare professionals.
- Data Acquisition and Analysis: Collect and analyse comprehensive datasets of patient information, medical records, lifestyle factors, and genetic data to enhance the accuracy and predictive capabilities of the model.
- Integration and Deployment: Collaborate with healthcare providers to seamlessly integrate the prediction model into their existing systems and workflows, ensuring ease of use and compatibility.

Revenue Streams

- Licensing and Subscription: Offer licensing agreements or subscription plans to healthcare providers, insurance companies, and research institutions for access to the heart disease prediction model.
- Software-as-a-Service (SaaS): Another approach is to offer the heart disease prediction system as a cloud-based Software-as-a-Service (SaaS) platform. Users can access the platform through a subscription-based model, paying a monthly or annual fee to utilize the prediction system. This model can cater to healthcare professionals, medical researchers, or institutions seeking a convenient and scalable solution for heart disease prediction
- Consultation Services: Provide expert consultation services to healthcare organizations for implementing and optimizing the use of the prediction model within their practice.
- Data Analytics and Insights: Offer insights and analytics services based on aggregated and anonymized data to pharmaceutical companies, research institutions, or public health organizations for research and development purposes.

Partnerships Business

- Healthcare Institutions: Collaborate with hospitals, clinics, and medical practices to integrate the prediction model into their systems and leverage their expertise for validation and real-world implementation.

- Research Institutions: Establish partnerships with academic and research institutions to exchange knowledge, access data for model training, and collaborate on research projects.
- Data Providers: Form partnerships with data providers, such as electronic health record systems or wearable device manufacturers, to access relevant patient data for model development and validation.

Product Prototype Description

The heart disease predictor cloud-based web application prototype aims to demonstrate the core functionality of risk assessment, personalized recommendations, and educational resources. The prototype provides a foundation for further development, integration with additional features, and refinement based on user feedback and market requirements. Here is a description of the key components and features of the product prototype:

- User Registration and Authentication: Users can create accounts and securely log into the application using their credentials. The registration process collects essential information such as age, gender, and contact details to personalize the risk assessment.
- Dashboard: Upon logging in, users are presented with a dashboard that provides an overview of their heart disease risk status. The dashboard may include visual representations of risk scores, personalized recommendations, and educational resources.
- Risk Calculation and Prediction: The application utilizes a predictive algorithm based on machine learning techniques to analyse the user's inputs and calculate their risk of heart disease. The algorithm takes into account established risk factors and generates a personalized risk score or probability of developing heart disease within a specific timeframe.
- Risk Visualization and Interpretation: The risk assessment results are presented to the user in an intuitive and visually appealing manner. Visualizations such as charts or graphs may be used to display the risk level and provide a clear understanding of the factors contributing to the calculated risk.
- Personalized Recommendations: Based on the risk assessment results, the application provides personalized recommendations for lifestyle modifications, preventive measures, and further medical evaluation if necessary. These recommendations may include suggestions for diet, exercise, medication, and regular check-ups.
- Scalability and Cloud Deployment: The cloud-based web application is designed to be scalable, allowing for efficient handling of increased user traffic and data storage requirements. Cloud deployment ensures accessibility and availability of the application across various devices and locations.

Product Details

The heart disease prediction web application works by utilizing various data sources, algorithms, and user inputs to generate personalized risk calculation for patients. Here is an overview of how the product works:

- Model Training: Machine learning models are trained using historical data and validated medical research to learn patterns and predictive relationships between the input variables and the likelihood of developing heart disease.
- Risk Assessment: The trained models analyse the user's input data, applying the learned patterns and algorithms to calculate the individual's risk of heart disease. This may involve generating

a risk score or probability indicating the likelihood of developing heart disease within a specified time frame.

- User Interface: The heart disease predictor web application offers a user-friendly interface where users can input their data, view risk assessment results, and access personalized recommendations. The interface is designed to be intuitive, visually appealing, and easy to navigate.
- Continuous Improvement: The application continually evolves and improves based on feedback, updates in medical research, and advancements in machine learning. Regular updates ensure that the prediction model remains accurate and up-to-date.

Data Source

The heart disease dataset used in this analysis consists of several columns representing various factors that can contribute to the prediction of heart disease. Each column provides specific information about a patient's characteristics and clinical measurements. Here is a description of the columns in the dataset:

- Age: The age of the patient in years.
- Sex: The sex of the patient (0: female, 1: male).
- CP (Chest Pain Type): The type of chest pain experienced by the patient, categorized into four types (0-3).
- Trestbps: The resting blood pressure of the patient (in mm Hg).
- Chol: The cholesterol level of the patient (in mg/dl).
- FBS (Fasting Blood Sugar): The fasting blood sugar level of the patient (>120 mg/dl: 1, <=120 mg/dl: 0).
- Restecg (Resting Electrocardiographic Results): The results of resting electrocardiography, categorized into three levels (0-2).
- Thalach: The maximum heart rate achieved by the patient during exercise.
- Exang (Exercise Induced Angina): The presence (1) or absence (0) of exercise-induced angina.
- Oldpeak: The ST depression induced by exercise relative to rest.
- Slope: The slope of the peak exercise ST segment, categorized into three levels (0-2).
- CA (Number of Major Vessels): The number of major vessels colored by fluoroscopy (0-3).
- Thal (Thalassemia): A blood disorder categorized into three levels (0-2).
- Target: The presence (1) or absence (0) of heart disease.

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

The provided code utilizes the logistic regression algorithm to predict heart disease based on the given dataset. Logistic regression is a widely used machine learning technique for binary classification tasks. The code first loads the dataset and splits it into features (X) and the target variable (y). Then, it performs a train-test split to divide the data into training and testing sets. The logistic regression model is trained on the training set and evaluated on the testing set using the accuracy score metric.

The goal of the code is to train the logistic regression model to predict the presence or absence of heart disease based on the provided patient factors. By evaluating the model's accuracy, it provides insights into the performance of the logistic regression algorithm for heart disease prediction using the given dataset.

Overall, the heart disease dataset and code aim to analyse and predict the occurrence of heart disease based on patient characteristics and clinical measurements, contributing to the understanding and management of cardiovascular health.

Exploratory Data Analysis

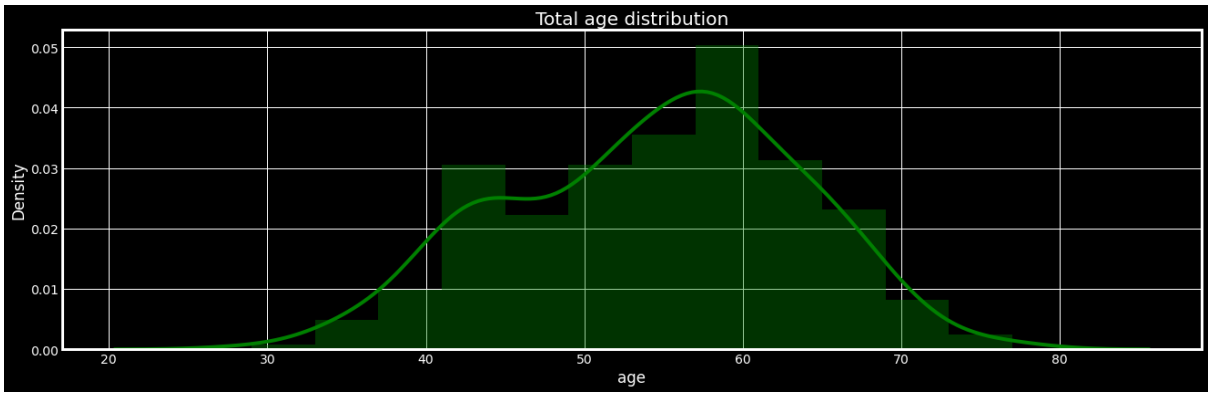


Figure 1

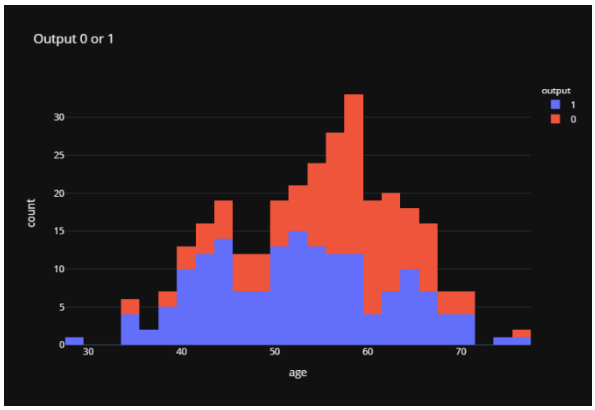


Figure 2

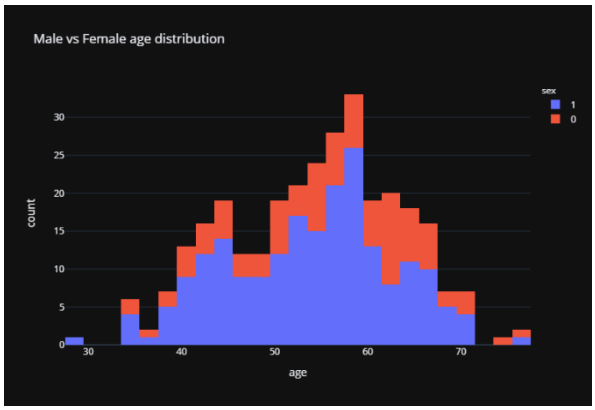


Figure 3

- Figure 1: The first plot shows the total distribution of age for all the individuals in the survey. Most of them were between 50-60 age group and very few on either side.
- Figure 2: The second plot shows the distribution of age separately for people with higher risk (output = 1) and for peoplw with lower risk (output = 0)
- Figure 3: the third plot shows the distribution of age separately for women (sex = 1) and men (sex = 0)

Machine Learning Model Selection

XGBoost Classifier: XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that can be used as a classifier for heart disease prediction. It is an ensemble learning method that

combines the predictions of multiple weak decision trees to make accurate predictions. Here's an overview of how the XGBoost classifier can be applied to heart disease prediction:

- Data Preprocessing: Clean and preprocess the data to handle missing values, outliers, and categorical variables. Data normalization or scaling may also be applied to ensure all features are on a similar scale.
- Training and Testing Data Split: Divide the dataset into a training set and a testing set. The training set is used to train the XGBoost classifier, while the testing set is used to evaluate its performance.
- XGBoost Model Training: Apply the XGBoost algorithm to the training set. The algorithm works by iteratively building weak decision trees and combining their predictions to minimize a specific loss function. It uses a gradient boosting framework to optimize the model's performance.
- Hyperparameter Tuning: Fine-tune the XGBoost classifier by adjusting its hyperparameters. Some important hyperparameters include the number of trees (n_estimators), maximum tree depth (max_depth), learning rate (eta), and regularization parameters. Hyperparameter tuning techniques such as cross-validation or grid search can be used to find the optimal set of hyperparameters.
- Prediction and Evaluation: Once the XGBoost classifier is trained, use the testing set to evaluate its performance. Predict the presence or absence of heart disease for the test samples based on their feature values and compare the predictions with the actual labels. Utilize evaluation metrics such as accuracy, precision, recall, and F1-score to assess the model's performance.
- Feature Importance Analysis: XGBoost provides a feature importance score, which indicates the relative importance of each feature in predicting heart disease. Analyze the feature importance scores to gain insights into the most influential features for heart disease prediction.
- Prediction and Deployment: With the trained XGBoost classifier, you can predict heart disease for new, unseen samples. The model can be deployed in a production environment, such as a web application or an API, to provide real-time predictions for individuals.

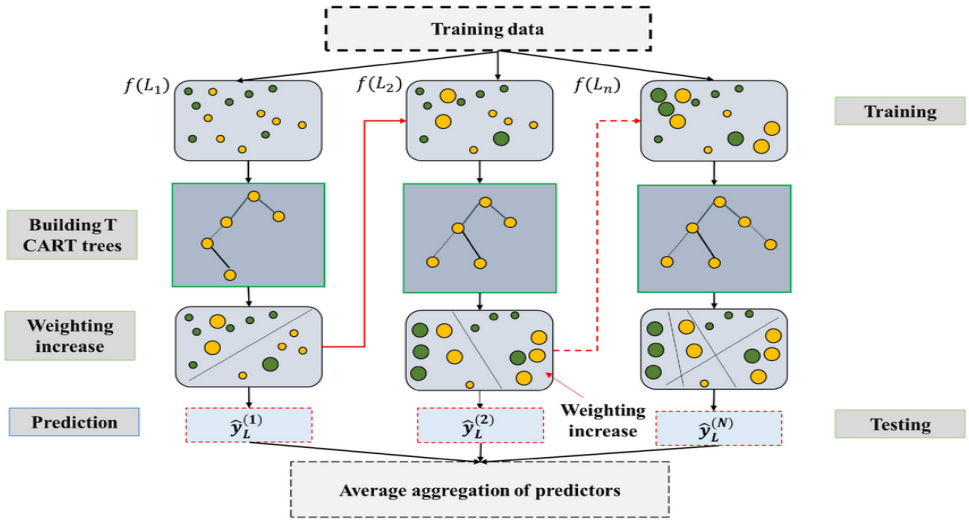


Figure : Graphical-scheme-of-XGBoost-model

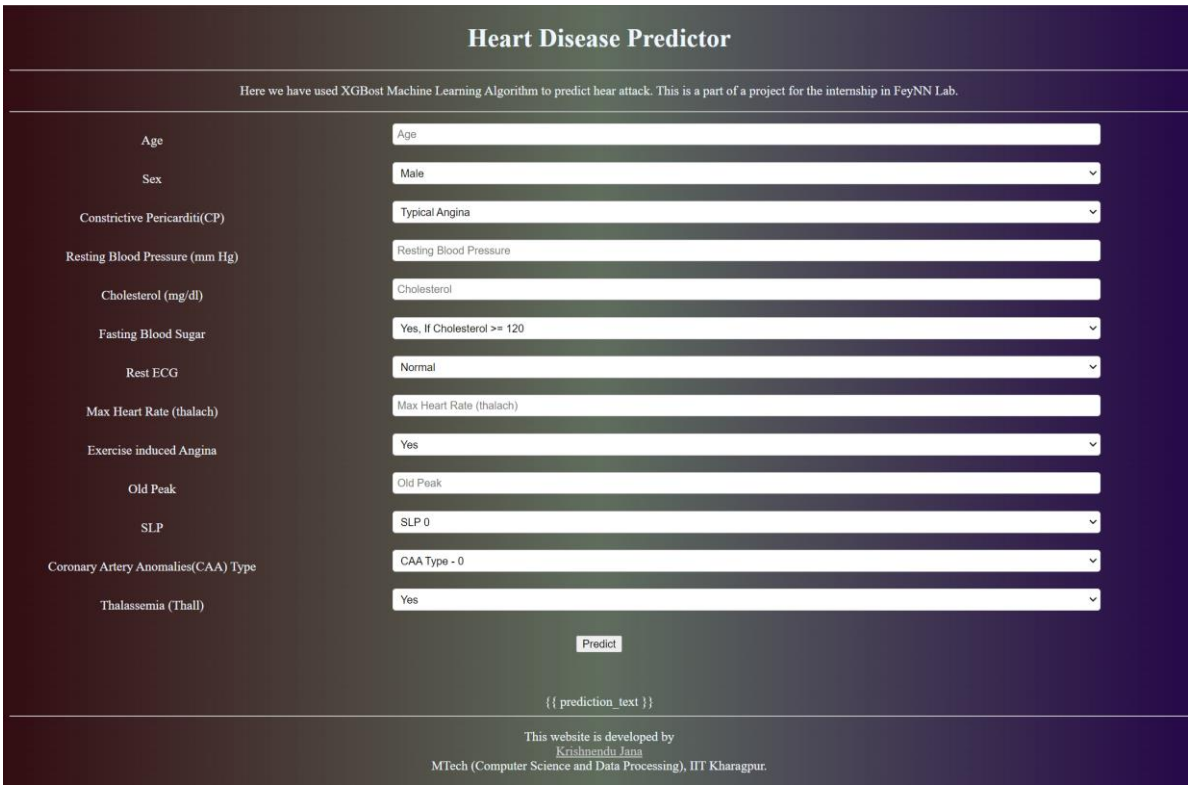
Accuracy of the model:

Precision: 0.968 Recall: 0.938 F1-Score: 0.952 Accuracy: 95.082 % Mean Square Error: 0.049

Prototype Deployment

To develop a prototype using Flask for the heart disease predictor in a webpage, we have follow these steps:

1. Set up Flask: Install Flask using pip, create a new Flask project, and set up the necessary file structure.
2. Create the HTML templates: Design and create the necessary HTML templates for the webpage. Consider creating templates for the home page, prediction form, and result display.
3. Implement routes and views: Define the routes in your Flask application and create corresponding view functions to handle requests. For example, you may have a route for the home page, prediction form submission, and result display.
4. Build the prediction model: Implement the heart disease prediction algorithm using machine learning techniques. Train the model using an appropriate dataset and extract relevant features for prediction.
5. Integrate the prediction model: In the appropriate Flask view function, load the trained prediction model and use it to predict the heart disease risk based on user input.
6. Render templates and display results: Pass the prediction results to the appropriate template and render it to display the prediction outcome to the user.
7. Implement user input handling: Retrieve and validate user input from the prediction form, ensuring that the required fields are filled and the data is in the correct format.
8. Apply styling and improve user experience: Enhance the visual appearance of the webpage using CSS and JavaScript as desired. Ensure that the user interface is intuitive and user-friendly.
9. Test and debug: Perform thorough testing of the prototype to identify and fix any bugs or issues. Validate that the heart disease predictor functions as intended and provides accurate predictions.
10. Deployment: Once the prototype is tested and ready, deploy it to a web server or hosting platform so that it can be accessed by users.



Financial Modelling

Financial Modeling for Heart Disease Predictor on Prototype Development, Business & Financial Modeling:

Revenue Streams

- Licensing Fees: Generate revenue by licensing the heart disease predictor to companies in the Prototype Development, Business & Financial Modeling industry.
- Subscription Model: Offer a subscription-based revenue model, providing access to advanced features, updates, and ongoing support.

Cost Structure

- Research and Development: Allocate funds for continuous research and development to enhance the accuracy and effectiveness of the heart disease predictor.
- Technology Infrastructure: Invest in robust server infrastructure, cloud services, and data storage systems to handle user data securely and efficiently.
- Employee Expenses: Cover salaries, benefits, and training costs for the team involved in developing, maintaining, and supporting the heart disease predictor.
- Marketing and Promotion: Allocate funds for marketing campaigns, digital advertising, and other promotional activities to reach potential customers in the target industry.
- Legal and Compliance: Budget for legal services and compliance measures to ensure adherence to data protection regulations and industry standards.
- Operational Costs: Account for general operational expenses, such as office rent, utilities, and administrative costs.

Key Performance Indicators (KPIs)

- User Acquisition: Track the number of new users or companies subscribing to the heart disease predictor platform.
- Revenue Growth: Monitor the growth in revenue generated from licensing fees and subscriptions.
- Customer Retention: Measure the percentage of customers who renew their subscriptions or continue to use the heart disease predictor over time.
- Accuracy Metrics: Evaluate the accuracy and performance of the heart disease predictor through metrics like sensitivity, specificity, and predictive value.
- Cost-to-Income Ratio: Monitor the ratio of operating costs to revenue to ensure a sustainable and profitable business model.
- Customer Satisfaction: Conduct surveys or gather feedback to assess user satisfaction and identify areas for improvement.

Financial Projections

Develop financial projections based on assumptions regarding market size, pricing strategies, user adoption rate, and customer retention. Consider factors such as the expected growth rate of the target

industry and the competitive landscape. Create revenue forecasts and cost estimates to determine the expected profitability and return on investment over a specific period.

Sensitivity Analysis

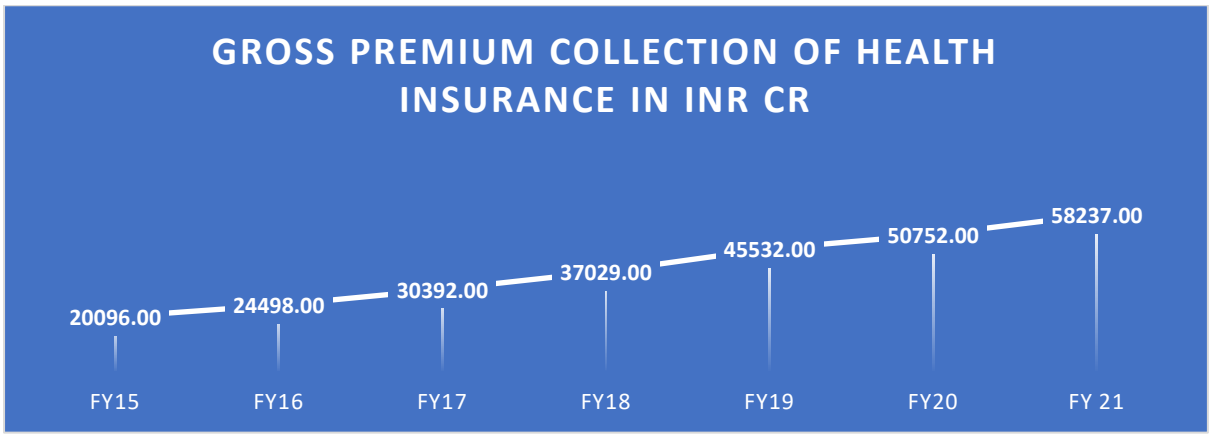
Perform sensitivity analysis to assess the impact of key variables on financial projections. Identify critical factors such as changes in subscription pricing, market saturation, or customer acquisition costs. Evaluate different scenarios and their potential effects on revenue, expenses, and profitability.

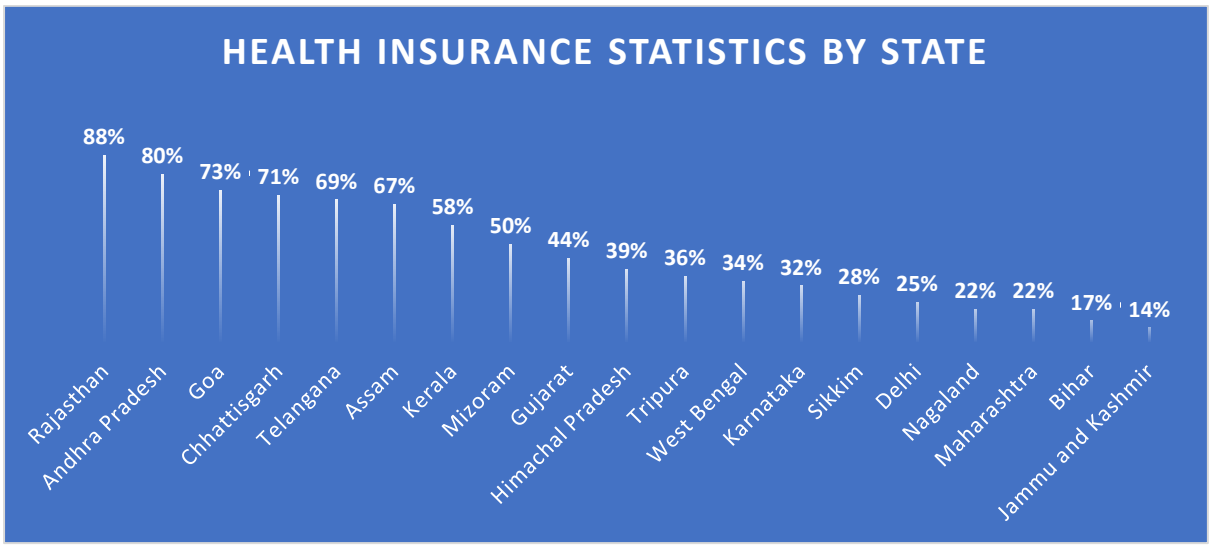
Funding and Investment

If required, create a funding strategy, considering options like venture capital, angel investors, or grants. Develop an investment pitch highlighting the market opportunity, competitive advantages, financial projections, and the potential impact of the heart disease predictor on the Prototype Development, Business & Financial Modeling industry.

Health Insurance for Heart Conditions Patients

Health insurance companies in India have stepped up and introduced dedicated heart or cardiac care insurance plans to help people cope up with the exorbitant heart treatment cost. Buying a critical illness plan or a regular health insurance plan may not be feasible if you have a pre-existing condition. However, in case of heart care insurance you can get coverage even with pre-existing diseases.

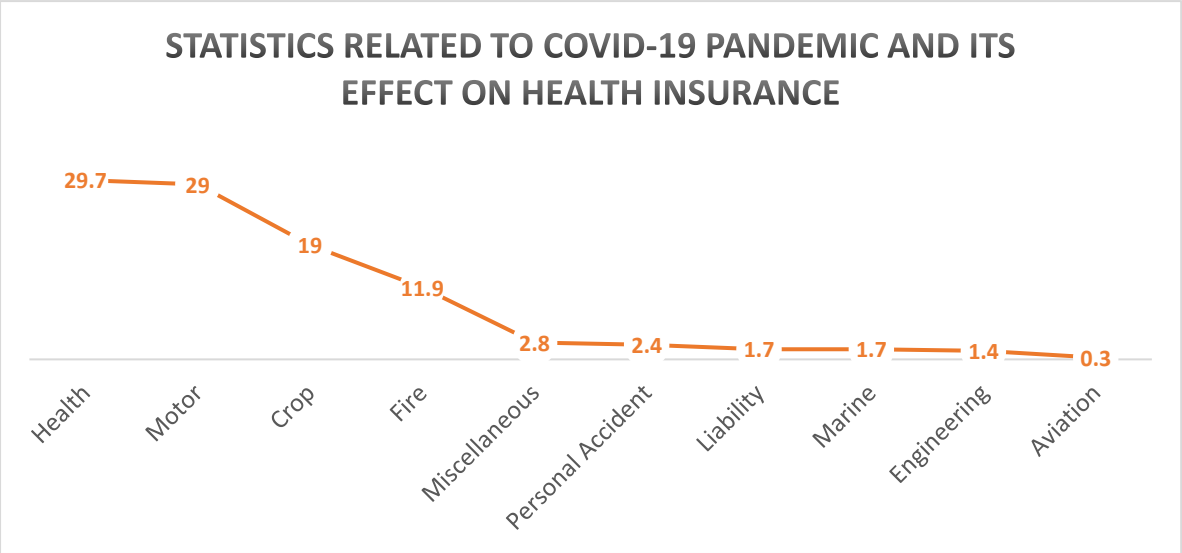




Employer-Sponsored Health Insurance Statistics

Employer-sponsored health insurance is also known as corporate health insurance or group health insurance which is generally provided by the employer of the company or a particular group to its employers.

- As per the Health Trend report shared by world's leading insurance broker and risk advisor, Mercer Marsh, the cost of employer-sponsored medical benefits in India was expected to rise by 15% in 2022.
- The Health Trends report surveyed almost 210 insurers around the world, including 74 in Asia, to identify significant trends that will shape the future of employer-provided medical coverage.
- The same reports show that five countries in Asia including India have higher medical trend rates. India had the highest medical inflation rate of 14%, followed by China (12%), Indonesia (10%), Vietnam (10%), and the Philippines (9%).



From *"Gross Premium Collection of Health Insurance in INR Cr"* chart it shows that the gross premium collection for health insurance has been increasing steadily and linearly from the financial year 2015 to 2021. Additionally, we may infer from Figures 1c and 1d that between the Financial Years 2016 and 2021, there was an increase in the number of persons purchasing health insurance premiums from both group and individual businesses.

Now we see that the growth of the market cap of health insurance is linear which is of the form,

$$y = mx + c \tag{0.1}$$

In the above x represents the financial year (FY15 – FY21, $x = 15$ for FY15, $x = 16$ for FY16,…) and y represent the number of customers of our company. An estimate has been made about the values of the co-efficient a and intercept b by fitting linear regression curve to the data shown in "*Gross Premium Collection of Health Insurance in INR Cr"* we get,

$$m = 6503, c = -\ 78969 \tag{0.2}$$

Now then we get,

$$y = 6503x - 78969 \tag{0.3}$$

Now if C_F be the fixed cost and C_V be the variable cost then profit P will be,

$$P(x) = 2000y - C_V \cdot y - C_F = (2000 - C_V)(6503x - 78969) - C_F \quad (0.4)$$

Future scope

- Integration of Genomic Data: Incorporating genomic data into the heart disease prediction dataset can offer valuable insights into genetic risk factors and personalized medicine approaches. Genetic information, such as single nucleotide polymorphisms (SNPs) or genetic variants associated with heart disease, can be integrated with clinical features to create a comprehensive dataset for more accurate predictions.
- Real-Time Monitoring and Wearable Devices: With the advancement of wearable devices and IoT technologies, future research can explore the integration of real-time physiological data, such as heart rate, blood pressure, or activity levels, into the heart disease prediction dataset. This can enable continuous monitoring and personalized risk assessment, leading to timely interventions and proactive management of heart disease.
- Risk Stratification and Precision Medicine: Heart disease prediction can be further refined to enable risk stratification, which categorizes individuals into different risk groups based on their likelihood of developing specific types or stages of heart disease. This can facilitate the implementation of precision medicine approaches, tailoring preventive strategies and treatments to individual patients' risk profiles.
- Mobile Applications and Digital Health Solutions: Future developments can focus on the integration of heart disease prediction models into mobile applications and digital health solutions. This can empower individuals to monitor their cardiovascular health, receive personalized risk assessments, and access lifestyle recommendations or interventions through user-friendly interfaces.

Conclusion

In conclusion, the heart disease prediction dataset provides valuable insights into the factors and measurements associated with the occurrence of heart disease. Through the utilization of machine learning algorithms and predictive models, this dataset enables accurate and early detection of heart disease, empowering healthcare professionals, researchers, and individuals to make informed decisions regarding prevention, diagnosis, and treatment.

References

[1] Data Set (Kaggle) : *https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset*

[2] "*Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India*" - Ekta Maini, Bondu Venkateswarlu, Baljeet Maini, and Dheeraj Marwahad.

[3] *"Top Health Insurance Statistics Of 2023"* , Forbes.