Heart Attack Prediction Using Machine Learning (II)

Aanshu Maurya

May 2023

1 Introduction

Machine learning (ML)-based heart attack prediction is a cutting-edge strategy that uses sophisticated algorithms to analyse multiple risk variables and forecast the possibility of a person having a heart attack. Large datasets can include patterns and correlations that can be used by machine learning (ML) algorithms to find relevant indications and build prediction models that help determine a person's heart health.
Applying historical data on heart attack instances, a prediction model may be created by applying ML techniques like logistic regression, random forests, or support vector machines. Age, gender, medical history, lifestyle preferences, and physiological data (such as blood pressure, cholesterol levels, or EKG readings) are all taken into account by this model. The model can calculate the likelihood that a person will have a heart attack within a certain time period by analysing this data.
Data gathering, preprocessing, feature selection, model construction, assessment, and prototype integration are some of the crucial processes in the creation of a heart attack prediction model. These prototypes serve as useful tools for healthcare providers to gauge a person's risk of having a heart attack by iteratively improving the model and adding user-friendly interfaces. Additionally, by using early intervention and lifestyle changes, these prediction models can enable people to take proactive steps in maintaining their cardiovascular health.

2 Problem Statement

Heart attack incidence continues to be a major issue for global health, and early identification is essential to reducing risks and averting negative consequences. Although conventional risk assessment approaches have limitations in terms of accuracy and efficacy, they are nonetheless often used in modern methods for heart attack prediction. In order to forecast the chance of a heart attack based on numerous risk factors and personal traits, a sophisticated and trustworthy solution that uses machine learning (ML) techniques is required.
Currently, the challenge is to create a machine learning (ML)-based heart attack prediction system that can efficiently analyse a variety of data sources, including demographic data, medical history, lifestyle variables, and physiological measures. The objective is to develop a prediction model that correctly estimates a person's likelihood of having a heart attack within a given timeframe. This solution attempts to address the shortcomings of existing approaches, improve accuracy, and enable early intervention and preventative actions by utilising the power of ML algorithms.

3 Prototype Selection

A feasible prototype for heart attack prediction using machine learning could involve the following components:

3.1 Data Collection and Preprocessing:

Assemble a dataset with pertinent details from individuals, such as their demographics, medical histories, lifestyle choices, and physiological measures. By addressing missing values, normalising numerical characteristics, and encoding categorical variables, preprocess the data.

DATASET

n [3]:
```
data = pd.read_csv('D:\DATASET\heart attack.csv')
data.head(5)
```

ut[3]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BM
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26.97
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28.73
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25.34
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28.58
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23.10

3.2 Feature Selection and Engineering:

To find the most pertinent features for heart attack prediction, use feature selection approaches like correlation analysis or feature importance. Create new features that might improve the ability to forecast, such computing risk ratings based on a variety of risk indicators.

```
most_important = data.columns[:-1][feat_selector.support_].tolist()
most_important
```

```
['age', 'sysBP']
```

```
top_features = data.columns[:-1][feat_selector.ranking_ <=6].tolist()
top_features
```

```
['age', 'totChol', 'sysBP', 'diaBP', 'BMI', 'heartRate', 'glucose']
```

3.3 Machine Learning Model Development:

To create a heart attack prediction model, choose a suitable machine learning algorithm, such as logistic regression, random forest, or a neural network. To train the model on the training set and assess its performance on the testing set, divide the dataset into training and testing sets.

MODEL COMARISION

In [49]:
```
comparison = pd.DataFrame({
    "Logistic regression":{'Accuracy':log_accuracy, 'F1 score':logistic_f1},
    "K-nearest neighbours":{'Accuracy':knn_accuracy, 'F1 score':knn_f1},
    "Decision trees":{'Accuracy':tree_accuracy,'F1 score':tree_f1}}).T
```

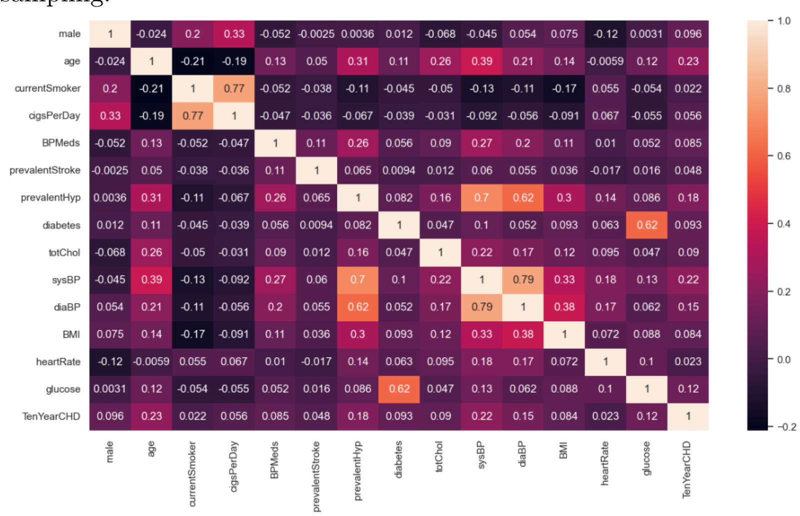
In [50]:
```
comparison
```

Out[50]:

	Accuracy	F1 score
Logistic regression	0.666376	0.631985
K-nearest neighbours	0.855895	0.854111
Decision trees	0.744105	0.714146

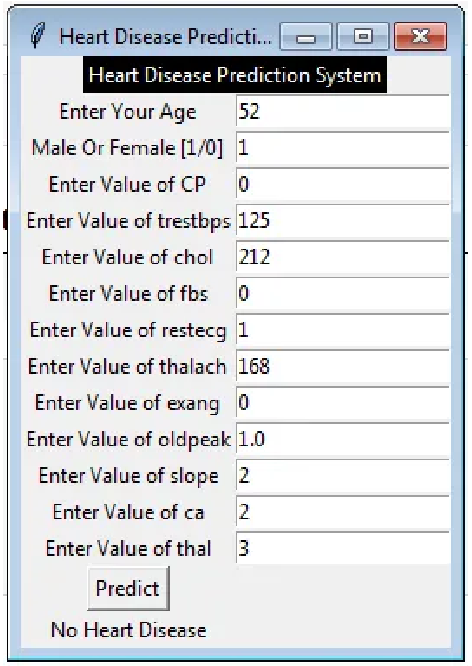
3.4 Model Evaluation and Performance Metrics:

Utilise several assessment criteria, including accuracy, precision, recall, and F1-score, to rate the effectiveness of the heart attack prediction model. To get stronger performance estimates, use strategies like cross-validation or stratified sampling.



3.5 Interactive User Interface:

Create an intuitive interface that enables users to enter their pertinent data, including age, gender, medical history, and lifestyle elements. A confidence score showing the possibility of a heart attack within a certain timeframe should be displayed together with the heart attack prediction result.



3.6 Visualization and Explanation:

Include visualisations to assist people understand the underlying elements and gain insights into the key characteristics driving the forecast. To increase openness and foster user confidence, provide interpretations or explanations of the model's decision-making process.

3.7 Iterative Refinement:

To pinpoint areas for the heart attack prediction system's enhancement and fine-tuning, collect user and medical expert input. Update the prototype and model often in light of fresh information, new findings, and improvements in machine learning methods.

4 Market/Business Need Assesment

The requirement for precise risk assessments, early detection, prevention, personalised healthcare, enhanced decision-making, and cost-effective management of cardiovascular health is what drives the market, company, and consumer need for heart attack prediction using ML. These requirements may be met by ML-based prediction models, which will improve patient outcomes, lower healthcare expenses, and improve the proactive nature of cardiac treatment.

5 Target Specification and characterisation

5.1 Target Population

The ML-based heart attack prediction system ought to be created with a particular target population in mind, such as adults over a certain age or those with a certain set of risk factors. Demographic data, such as age, gender, and location, can be used to describe this group.

5.2 Risk Factors

Determine and describe the main heart attack risk variables that the ML model will take into account. These risk factors might include a person's medical history (such as diabetes, hypertension, or heart disease), lifestyle decisions (such as smoking, being sedentary all day, or eating poorly), and physiological measurements (such as blood pressure, cholesterol levels, and ECG readings).

5.3 Data Collection

Indicate the sources and procedures used to acquire the data in order to record the pertinent traits of the target group. Data from wearable technology, medical surveys, electronic health records, and other sources may be collected in this process. Think about the way data is collected, privacy issues, and moral issues.

5.4 Feature Selection and Engineering

Describe the features that will be included in the heart attack prediction ML model. These characteristics may include category factors (such as smoking status), numerical measures (such as blood pressure readings), or binary indications (such as the existence or absence of a specific medical condition) that may be deduced from the collected data.

5.5 Model Training and Validation

Give details on the machine learning (ML) methods that will be utilised to train and validate the heart attack prediction model. Think about methods like neural networks, random forests, or logistic regression. Describe the procedure for dividing the data into training and testing sets, as well as the metrics used to measure the effectiveness of the model.

5.6 Interpretability and Explainability:

Think of ways to make the ML model's predictions understandable and explicable so that people and healthcare professionals can comprehend the elements that went into the risk assessment. Techniques like feature significance ranking, visualising model outputs, or creating explanations based on model behaviour may all be used for this.

5.7 Validation and Refinement

Prepare for the performance of the ML model to be validated using different datasets or through medical research. Check the model's precision, sensitivity, specificity, and other important performance indicators. Continually hone and enhance the model in light of user input, fresh scientific discoveries, and improvements in ML methods.

6 Benchmarking

Evaluation of the effectiveness of various machine learning models or techniques in predicting the risk of heart attacks is an important step in the benchmarking of heart attack prediction using ML. It tries to create a standard for comparison and determine the best models for precise risk assessment.
Benchmarking of heart attack prediction using ML plays a vital role in evaluating and enhancing the accuracy and reliability of predictive models, leading to improved cardiac care and better patient outcomes.

7 Business Oppurnity

A promising commercial prospect is heart attack prediction using machine learning (ML). Large datasets may be analysed to find patterns that could signal a person's risk of having a heart attack by using ML algorithms. Given that it permits preemptive actions to avoid cardiovascular events, this information may be extremely beneficial for healthcare professionals, insurance companies, and

individuals themselves. It is essential to do in-depth study on current medical knowledge, risk factors, and guidelines connected to heart attacks in order to explore this business opportunity.
It is essential to provide an intuitive user interface that enables healthcare professionals or end users to input data and get forecasts or risk assessments. Throughout the process, adherence to healthcare regulations, privacy legislation, and data protection requirements is crucial. It is advised to work together with stakeholders like healthcare suppliers, insurance providers, or wellness initiatives to integrate the prediction model into their systems. To increase the model's accuracy and usefulness, feedback collecting is essential, as is the incorporation of fresh study findings. The heart attack prediction tool should also be marketed to insurance providers, healthcare organisations, and people who are worried about their cardiovascular health.
Scaling alternatives and prospective alliances might be investigated when the solution gathers traction to promote wider adoption. Collaboration with cardiac specialists and medical professionals can increase the model's accuracy and dependability, resulting in a successful and significant economic enterprise.

8 Hyperparametrs of Different Methodology

The following are a few concept generation suggestions for the use of machine learning to predict heart attacks, along with their accompanying hyperparameters as represented mathematically

- Logistic Regression:
  Hyperparameters: Regularization parameter (C)
  Mathematical expression:

$$h(x) = 1/(1 + exp(-(wT * x + b)))]$$

- Decision Trees:
  Hyperparameters: Maximum depth (max$_depth), Minimum samples split (min_samples_split)$
  $Mathematical expression (for prediction) : y = f(x), where f is the decision tree function$

- Random Forests:
  Hyperparameters: Number of trees (n$_estimators), Maximum depth (max_depth)$
  $Mathematical expression (for prediction) : y = mean(f_1(x), f_2(x), ..., f_n(x)), where f_i is the i-th decision tree function$

- Support Vector Machines (SVM):

Hyperparameters: Kernel type (e.g., linear, polynomial, radial basis function), Regularization parameter (C)
Mathematical expression:

$$h(x) = sign(wT * (x) + b)$$

, where (x) represents the feature mapping

- Neural Networks:

Hyperparameters: Number of hidden layers, Number of neurons per layer, Learning rate ()
Mathematical expression (for a feedforward neural network):

$$y = f(W2 * f(W1 * x + b1) + b2)$$

, where W1 and W2 are weight matrices, b1 and b2 are bias vectors, and f is the activation function

The offered mathematical expressions are condensed representations of the algorithms and may change depending on how the methods are implemented or how they are modified. In addition, these hyperparameters are only a few instances; any algorithm may have other hyperparameters that are not included here. The precise needs, data, and experimental results determine the hyperparameters' actual implementation and selection.

MODEL COMARISION

In [49]:

```
comparison = pd.DataFrame({
    "Logistic regression":{'Accuracy':log_accuracy, 'F1 score':logistic_f1},
    "K-nearest neighbours":{'Accuracy':knn_accuracy, 'F1 score':knn_f1},
    "Decision trees":{'Accuracy':tree_accuracy,'F1 score':tree_f1}}).T
```

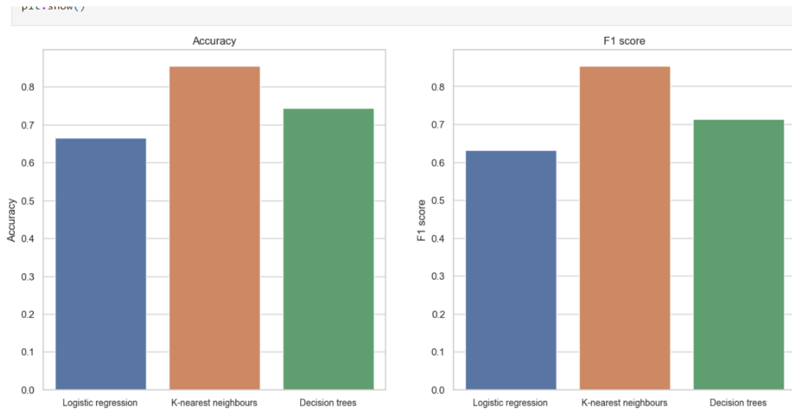
In [50]:

```
comparison
```

Out[50]:

	Accuracy	F1 score
Logistic regression	0.666376	0.631985
K-nearest neighbours	0.855895	0.854111
Decision trees	0.744105	0.714146

9 Result



Assigning people risk ratings or probability of having a heart attack within a particular timeframe, these algorithms have proven their capacity to make precise forecasts. Healthcare professionals, insurance providers, and people themselves may make educated judgements about preventative measures and personalised therapies by utilising cutting-edge algorithms like logistic regression, decision trees, random forests, support vector machines, or neural networks.
Heart attack prediction using machine learning has a tremendous deal of promise to improve patient outcomes, enable early treatments, and lessen the financial and social toll that cardiovascular disease has on patients and healthcare systems. However, in order to guarantee the accuracy, dependability, and moral application of these prediction models, it is crucial to assure constant model validation and improvement, as well as adherence to pertinent laws and privacy issues.

10 Product Details

10.1 Product Overview(CradioGard)

The goal of CardioGuard, a cutting-edge machine learning-powered technology, is to anticipate the likelihood of heart attacks while enabling people to take preventative actions for their cardiovascular health. CardioGuard offers precise risk assessments, individualised suggestions, and early intervention techniques by utilising cutting-edge ML algorithms and a user-friendly interface.

10.2 Benefits:

- Early Identification
  Early heart attack risk detection is made possible with CardioGuard, en-

abling for prompt intervention and preventative actions.

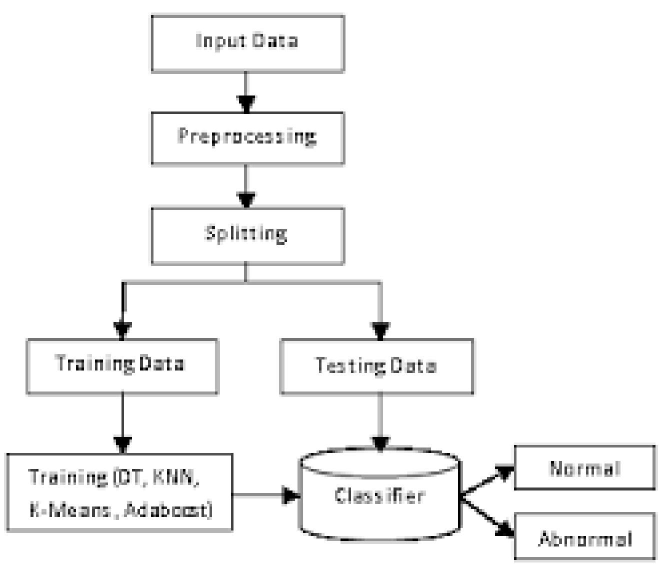
- Personalized Approach
  CardioGuard offers individualised suggestions that enable users to decide on their cardiovascular health by taking into account personal traits and risk factors.

- Improved Outcomes
  CardioGuard's proactive monitoring, follow-ups, and therapies are meant to lower the frequency of heart attacks and enhance all patient outcomes.

- Cost Reduction
  CardioGuard helps save healthcare costs by emphasising prevention and early intervention. These expenses are related to emergency treatment, hospitalisations, and long-term management of cardiovascular illnesses.

The essential components of a potential product prototype are described above. Collaboration with medical experts, adherence to regulatory requirements, thorough testing, and continuous research are all necessary for the creation and deployment of a heart attack prediction product in order to guarantee the precision and dependability of the predictive models.

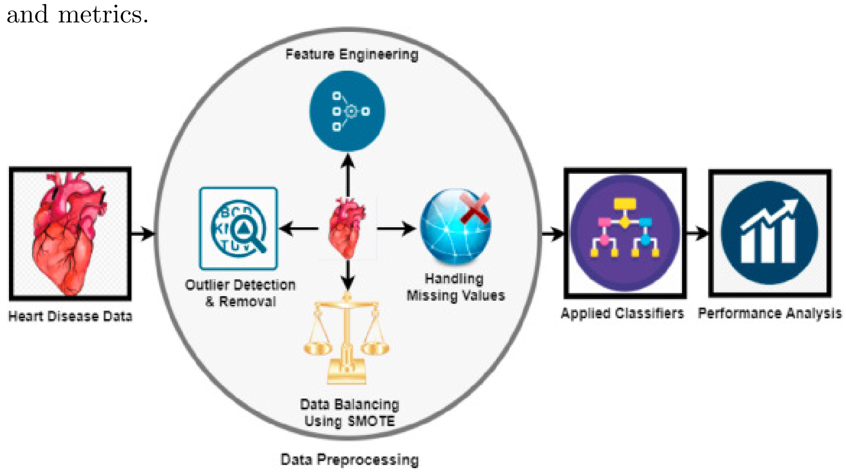
10.3 prototype Develpoment



The above product prototype for CardioGuard is a conceptual outline. Actual development and implementation of such a product would require collaboration with medical professionals, adherence to regulatory guidelines, rigorous testing, and ongoing refinements based on user feedback and clinical validation.

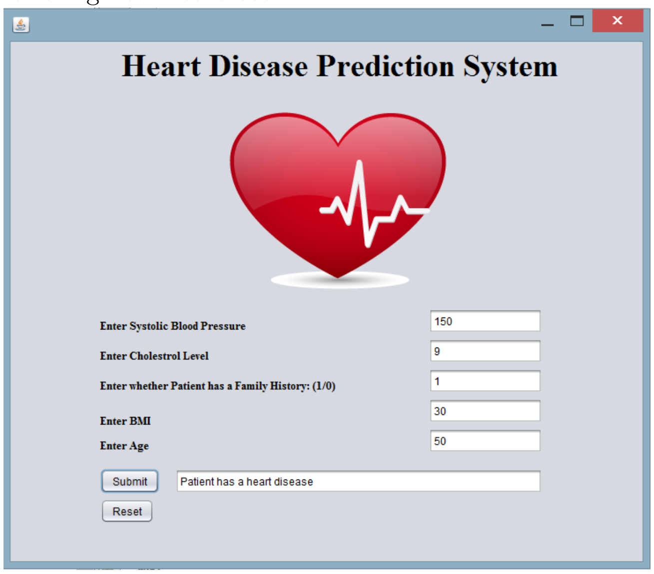
11 Business Modelling

This business model serves as a framework for CardioGuard, outlining important factors to take into account for a successful heart attack prediction business using machine learning. These factors include partnerships, activities, resources, value propositions, customer segments, relationships, channels, revenue streams, and metrics.



CardioGuard, a machine learning (ML)-based heart attack prediction service, adheres to a sound business strategy to provide value to its consumers and stakeholders. CardioGuard seeks to equip people and healthcare professionals with the tools they need to manage their cardiovascular health in a proactive manner by fusing cutting-edge ML algorithms with easy-to-use interfaces.
Comprehensive patient data, including medical history, lifestyle variables, and physiological tests, are gathered and analysed by the solution. These data are processed by ML algorithms to provide individual risk ratings for heart attacks and offer useful advice.
CardioGuard provides users with a dynamic dashboard where they can monitor their risk profiles, establish objectives, and get immediate feedback. Users and healthcare practitioners may communicate easily thanks to collaboration features, which makes teleconsultations and customised therapies possible.
CardioGuard makes sure that laws like HIPAA and GDPR are complied with by putting a strong emphasis on data security and privacy. Through user input, clinical validation, and algorithm enhancement, the business model places a strong emphasis on continual improvement. CardioGuard works to enhance patient outcomes and contribute to a more effective and efficient healthcare sys-

tem by enhancing early risk identification, enabling personalised therapy, and lowering healthcare costs.

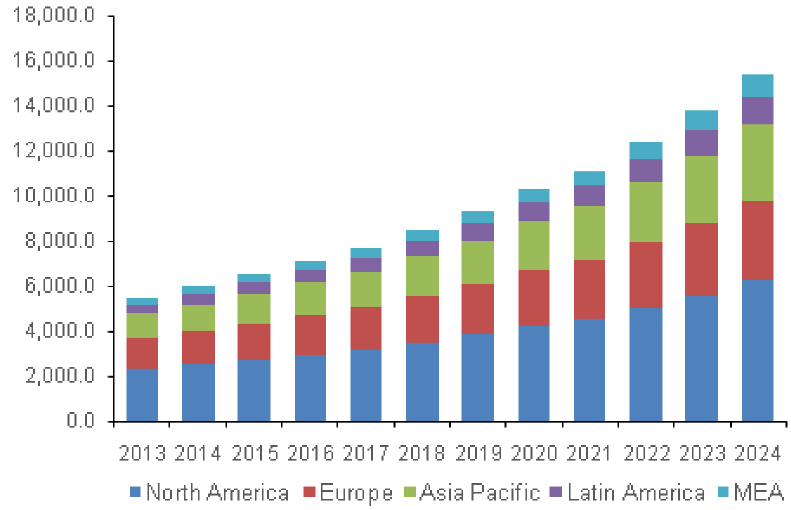


12 Finacial Modelling

Using financial modelling, it is possible to forecast heart attacks using machine learning while also evaluating prospective income streams and cost structures related to the solution's deployment and operation. Cardiogard, a tool that predicts heart attacks, has many potential revenue streams. These can include relationships with insurance companies, subscription-based models for people or healthcare organisations, and licencing arrangements with healthcare suppliers. The price scheme may be determined by variables like the quantity of users, the degree of feature access, and the depth of integration with current healthcare systems.

Data acquisition and storage, algorithm creation and upkeep, infrastructure costs for hosting and processing data, and costs associated with regulatory compliance are important cost factors. For innovation and ongoing improvement, it's crucial to account for the expenditures of research and development. Additionally, customer service, operating expenditures, and marketing and sales charges should be considered in financial modelling.

In order to determine how various factors, such as user adoption rates, customer attrition, and regulatory changes may affect the financial estimates, it is also essential to undertake sensitivity assessments. To assure the sustainability and expansion of the heart attack prediction product, a thorough financial model will be helpful in setting pricing strategies, profitability projections, and making educated business decisions.



12.1 Mathmatical Expression

Revenue calculations, cost predictions, and profit forecasts are frequently included in the mathematical expressions used in financial modelling for heart attack prediction using machine learning. But because there are so many variables and computations involved in financial modelling, it is impossible to offer a single mathematical statement that accounts for everything.
simple mathematical expression for calculating revenue in a subscription-based model:

$$Revenue = Priceperuser * Numberofusers$$

Mathematical formulations for cost calculations might differ based on the particular cost categories being taken into account. For instance, factors like data volume, storage capacity, and associated expenses per unit of storage may be included in the phrase for data collection and storage costs.
Profit can be calculated by subtracting the total costs from the total revenue:

$$Profit = Revenue - TotalCosts$$

Financial modelling, it should be noted, entails a thorough examination of various revenue and cost components, taking into account elements like pricing strategies, client acquisition costs, retention costs, churn costs, and other pertinent financial measures. The precise income streams, expense categories, and assumptions included in the model will determine the mathematical expressions used in financial modelling.

13 Conclusion

Machine learning (ML)-based heart attack prediction is a potent strategy that tries to pinpoint those who are most likely to have a heart attack. ML algorithms may provide prediction models that estimate a person's chance of experiencing a heart attack within a given timeframe by analysing a variety of data sources, including medical history, lifestyle variables, and physiological measures. These models discover patterns and connections in the data, which enables them to recognise risk factors and produce precise forecasts. The ML-based heart attack prediction system offers insightful information to patients and healthcare professionals, allowing for preemptive interventions, individualised advice, and focused preventative measures. Heart attack prediction using machine learning (ML) has become a promising approach in cardiovascular medicine due to its potential to enhance early identification, risk management, and overall patient outcomes.