|  |  |  |
| --- | --- | --- |
| **Tech Saksham**  Capstone Project Report  **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FUNDAMENTALS** |  |  |

**“HEART DISEASE PREDICTION”**

**“ANNA UNIVERSITY REGIONAL CAMPUS TIRUNELVELI”**

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| au950021135016 | EDHISHA SP |

|  |  |
| --- | --- |
|  |  |
|  | RAMER BOSE  Sr. AI Master Trainer |
|  |  |

**ABSTRACT**

Heart disease is a term covering any disorder of the heart.

Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases has increased significantly over the past few decades in India it has become the leading cause of death in India. A study shows that from 1990 to 2016 the death rate due to heart disease increased around 34 percent from 155.7 to 209.1 deaths per one lakh population in India.

Thus preventing Heart disease has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives. This is where Machine Learning comes into play. Machine Learning helps in predicting Heart diseases, and the predictions made are quite accurate.

1. Problem statement
2. Data collection
3. Existing solution
4. Proposed solution with used models
5. Result

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 1 |
| 2 | Chapter 2: Services and Tools Required | 4 |
| 3 | Chapter 3: Project Architecture | 6 |
| 4 | Chapter 4: Project Outcome | 8 |
| 5 | Conclusion | 9 |
| 6 | Future Scope | 10 |
| 7 | References | 11 |
| 8 | Code | 12 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Problem Statement**

The problem statement of heart disease typically revolves around the high prevalence, significant morbidity, and mortality rates associated with various cardiac conditions. It encompasses understanding the causes, risk factors, diagnosis, treatment, and prevention strategies related to heart diseases The problem statement of heart disease prediction typically involves developing a predictive model that can accurately identify the likelihood of an individual having heart disease based on certain input features such as medical history, lifestyle factors, and possibly genetic information. The goal is to create a reliable tool that healthcare professionals can use to assess a patient's risk of heart disease and make informed decisions regarding prevention, diagnosis, and treatment.

* 1. **Proposed Solution**

Aproper solution for heart disease prediction using modern technology involves a combination of advanced machine learning algorithms, big data analytics, and wearable sensor technologies. Gather comprehensive health data from various sources, including electronic health records, medical imaging, genetic information, lifestyle factors (such as diet and exercise), and wearable devices (like smartwatches or fitness trackers. Identify relevant features from the collected data that are strongly correlated with heart disease risk. This may include factors like blood pressure, cholesterol levels, family history, smoking status, physical activity, and more. Choose appropriate machine learning models for classification tasks, such as logistic regression, decision trees, random forests, support vector machines (SVM), or neural networks. Train the selected models on the preprocessed data. Use techniques like cross-validation to ensure robustness and avoid overfitting. Fine-tune the parameters of the models to optimize their performance using techniques like grid search or random search. Evaluate the trained models using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score. Once the best-performing model is selected, deploy it in a production environment using frameworks like Flask or Django for creating APIs. Continuously monitor the performance of the deployed model and update it periodically with new data to ensure its effectiveness over time.

* 1. **Feature**

**Real-Time Analysis**: The dashboard will provide real-time analysis of customer data.

**Customer Segmentation**: Identify relevant features from the collected data that are strongly correlated with heart disease risk. This may include factors like blood pressure, cholesterol levels, family history, smoking status, physical activity, and more

**Predictive Analysis**: It will use historical data to predict future customer behavior

* 1. **Advantages**

Using machine learning to classify cardiovascular disease occurrence can help diagnosticians. This research develops a model that can correctly predict cardiovascular diseases to reduce the fatality caused by cardiovascular diseases. Heart disease prediction offers several significant advantages in healthcare. By leveraging advanced data analytics and predictive modeling techniques, healthcare providers can identify individuals at higher risk of developing heart disease before symptoms manifest, enabling early intervention and preventive measures

* 1. **Scope**

The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction. The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, eg, relationships between medical factors related to heart disease and patterns, to be established

* 1. **Future Work**

This project predicts people with cardiovascular disease by extracting the patient's medical history that leads to fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, etc. The model can be used to provide an enhanced, more accurate framework that would lead to a better human disease prediction model

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

**2.1 Services Used**

1. Electronic Health Records (EHR): EHR systems store patient health information, including medical history, lab results, diagnostic tests, and treatment plans. Analyzing this data can provide valuable insights for predicting heart disease risk.

2. Machine Learning Platforms: Platforms like TensorFlow, PyTorch, and sci-kit-learn offer tools and libraries for building and deploying machine learning models. These platforms enable developers to create predictive algorithms using various techniques such as logistic regression, random forests, support vector machines, and deep learning.

3. Cloud Computing: Cloud computing services such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure provide scalable infrastructure for storing and processing large healthcare datasets. Cloud-based solutions facilitate data analysis, model training, and real-time predictions

**2.2 Tools and Software used**



**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 Architecture**

**USER FRONTEND BACKEND**

|  |  |  |
| --- | --- | --- |
|  | **HTML 5** | **NODEJS 14.0**  **Database** |

**The architecture of a heart disease prediction system typically involves several key components:**

1. Data Collection: Gather relevant data sources such as electronic health records, medical imaging, genetic information, lifestyle factors, and wearable device data. This may involve accessing databases, APIs, or integrating with external systems.

2. Preprocessing: Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies. This step may involve data cleaning, normalization, feature engineering, and encoding categorical variables.

3. Feature Selection: Identify the most informative features from the preprocessed data that are strongly correlated with heart disease risk. This step may involve statistical analysis, feature importance ranking, and domain knowledge.

4. Model Development: Build predictive models using machine learning algorithms such as logistic regression, random forests, support vector machines, or deep learning neural networks. Train the models on labeled datasets using techniques like cross-validation to optimize performance

5. Validation and Evaluation: Validate the predictive models using independent test datasets and evaluate their performance metrics such as accuracy, sensitivity, specificity, and area under the ROC curve (AUC). This step ensures the models generalize well to unseen data.

6. Deployment: Deploy the trained models into a production environment where they can be accessed by end-users or integrated with healthcare systems. This may involve deploying as web services, APIs, or embedding within applications.

7. Monitoring and Maintenance: Monitor the performance of the deployed models over time and update them as needed to adapt to changing data distributions or improve predictive accuracy. This may involve ongoing model evaluation, retraining, and version control

**CHAPTER 4**

**PROJECT OUTCOME**

Model Training and Prediction :

We can train our prediction model by analyzing existing data because we already know whether each patient has heart disease. This process is also known as supervision and learning. The trained model is then used to predict if users suffer from heart disease.

Splitting:

First, data is divided into two parts using component splitting. In this experiment, data is split based on a ratio of 80:20 for the training set and the prediction set. The training set data is used in the logistic regression component for model training, while the prediction set data is used in the prediction component.

The following classification models are used - Logistic Regression, Random Forest Classifier, SVM, Naive Bayes Classifier, Decision Tree Classifier, LightGBM, XGBoost

Prediction:

The two inputs of the prediction component are the model and the prediction set. The prediction result shows the predicted data, actual data, and the probability of different results in each group.

Evaluation:

The confusion matrix, also known as the error matrix, is used to evaluate the accuracy of the model.

**CONCLUSION**



**FUTURE SCOPE**

The future scope of heart disease prediction is promising, driven by advancements in technology, data analytics, and personalized medicine. Innovations such as wearable sensors, remote monitoring devices, and genomic sequencing are poised to revolutionize cardiovascular risk assessment by providing real-time physiological data and identifying genetic predispositions. Integration of artificial intelligence and machine learning algorithms will enable more accurate and personalized risk prediction models, capable of analyzing complex datasets and identifying subtle patterns indicative of heart disease. Moreover, the integration of telehealth platforms and digital health ecosystems will facilitate seamless data exchange, enabling proactive interventions and personalized treatment plans tailored to individual patient needs. Collaborative efforts among healthcare stakeholders, researchers, and technology developers will be crucial in harnessing the full potential of heart disease prediction, ultimately leading to improved patient outcomes, reduced healthcare costs, and enhanced population health.

**REFERENCES**

1. Project Github link, Ramar Bose, 2024
2. Project video recorded link (YouTube/github), Ramar Bose, 2024
3. Project PPT & Report GitHub link, Ramar Bose, 2024

**CODE**

https://github.com/Edhisha016/HEART-DISEASE-PREDICTION