

## **Project Proposal**

### **1. Summary of the Problem**

Heart attacks are a leading cause of mortality in the United States. Because many heart attacks occur unexpectedly, it is often a life-and-death scenario. If we could create an early detection model that could identify individuals at high risk early, this could prevent fatal outcomes. However, healthcare providers often struggle with assessing heart attack risk quickly and accurately based on available patient vitals, especially in cases where a patient is just establishing care with a new practice. Since then, we haven't even had past visits on file to reference. Current processes can be time-consuming and reliant on each individual healthcare professional's experience, leading to an uneven patient experience and low-quality consistency.

### **2. Benefits to the Customer and Support for Decision-Making**

The proposed data product will enable healthcare providers to input vital signs such as blood pressure, cholesterol levels, and heart rate to receive a heart attack risk assessment quickly. This predictive tool will enhance clinical decision-making by providing data-driven insights, allowing physicians to prioritize high-risk patients and take preventative actions sooner. The system would streamline diagnostics and improve patient care outcomes, ultimately leading to better long-term management of heart disease.

### **3. Data Product Outline**

The heart attack risk prediction model is built upon a machine learning algorithm that is trained on a large dataset of patient vitals and heart attack outcomes. It consists of:

- **Input:** Patient vitals (age, sex, cholesterol, blood pressure, etc.)
- **Output:** A calculated risk score indicating the probability of a heart attack.
- **User Interface:** A web-based application or embedded module for use by clinicians in hospitals and healthcare settings.

The system will integrate seamlessly into current clinical workflows, ensuring ease of use and accessibility.

### **4. Description of Data Used to Construct the Data Product**

The data product will be constructed using a dataset of anonymized patient records, which includes information on:

- **Age, sex, cholesterol, blood pressure, and other relevant vitals**
- **Historical heart attack outcomes**

This data will allow us to train and validate machine learning models that accurately predict heart attack risk based on a patient's current vitals.

### **5. Objectives and Hypotheses of the Project**

**Objective:** To develop a machine learning model that accurately predicts the likelihood of a heart attack based on a patient's vitals.

**Hypotheses:**

- Certain combinations of vitals, such as elevated cholesterol and high blood pressure, significantly increase the risk of heart attack.
- The model can outperform traditional risk assessment tools used by clinicians by offering real-time predictions.

The overarching goal is to improve heart attack risk assessment accuracy and speed, enabling better prevention efforts.

## **6. Outline of Project Methodology**

The methodology will involve the following steps:

1. **Data Collection:** Gather patient data on heart attack risk factors and outcomes.
2. **Data Preprocessing:** Clean and prepare the dataset by handling missing values, standardizing features, and splitting it into training and test sets.
3. **Model Development:** Test multiple machine learning models, including Logistic Regression, Random Forest, and Gradient Boosting, to identify the most effective approach.
4. **Model Validation:** Evaluate model performance using metrics such as accuracy, precision, and recall.
5. **Deployment:** Integrate the trained model into a user-friendly application for healthcare providers.
6. **Monitoring and Maintenance:** Continuously monitor the system's performance, updating the model as new data becomes available.

## **7. Funding Requirements**

The project will require funding for the following:

- **Data acquisition and storage**
- **Machine learning model development**
- **Application development (UI, backend integration)**
- **Testing and validation in clinical settings**
- **Training for healthcare staff**

**Estimated Budget: \$200,000 - \$300,000**, depending on the scale and integration requirements.

## **8. Impact of Solution on Stakeholders**

The primary stakeholders include:

- **Healthcare providers:** Improved diagnostic tools and enhanced decision-making capabilities.
- **Patients:** Quicker identification of high-risk individuals will lead to faster interventions and potentially save lives.
- **Healthcare systems:** Reduced costs associated with emergency heart attack treatments by enabling preventative measures.

## 9. Ethical and Legal Considerations

Working with patient health data requires strict adherence to **HIPAA** and **GDPR** regulations. All patient data used for this project will be anonymized to protect individual identities. Additionally, clear communication regarding the limitations of the predictive model will be provided to healthcare professionals to prevent over-reliance on the system.

**Precautions:** The system will be designed to ensure that only authorized personnel have access, and all data processing will occur in secure environments. Regular audits will be performed to ensure compliance with all applicable data privacy laws.

## 10. Expertise Relevant to the Proposed Solution

As a team with a strong background in data science, machine learning, and healthcare data management, we have successfully completed similar predictive modeling projects in the past. Our team includes specialists in:

- **Data analytics and machine learning** with experience in healthcare.
- **Software engineering** for building and integrating user-friendly applications.
- **Compliance experts** to ensure ethical and legal adherence when handling sensitive patient data.