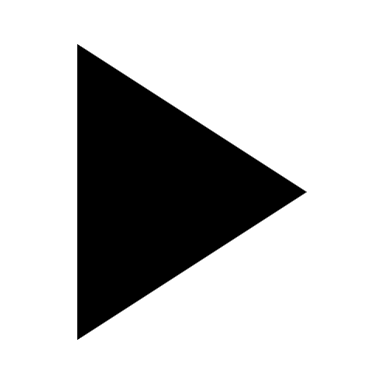


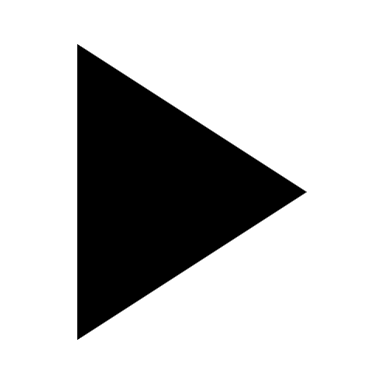
**Heart Disease Risk Prediction**



**Project Overview**

**This project aimed to develop a predictive model for assessing heart disease risk based on various clinical and lifestyle factors. The dataset used was sourced from Kaggle, comprising patient symptoms and risk indicators such as chest pain, high blood pressure, cholesterol levels, smoking status, obesity, and family history**.

**Project Process**



**Project Process stages**

**♦ Data Collection and Preprocessing**

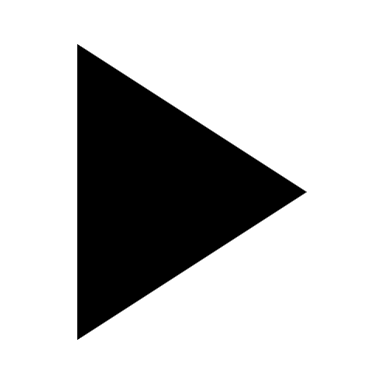
**The dataset underwent an extensive preprocessing phase including checking for missing values, ensuring numerical values, scaling features, and splitting the data into training and testing sets. Key features were extracted and their importance assessed using Random Forest, ensuring that the model focused on the most predictive variables.**

**♦ Model Development**

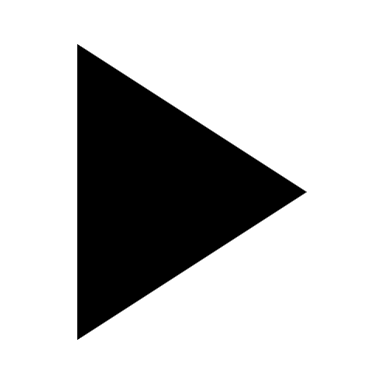
**Multiple classification algorithms were evaluated, including Logistic Regression, Random Forest, and Gradient Boosting. Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were used to benchmark performance. Hyperparameter tuning using Grid Search further enhanced model effectiveness, especially for Random Forest and Gradient Boosting, leading to improved generalization and reduced overfitting.**

**♦ Model Deployment**

**The final model was designed with a focus on interpretability and clinical relevance. Feature importance visualization helped highlight the key drivers of heart disease, aiding both model trust and usability in a healthcare context.**

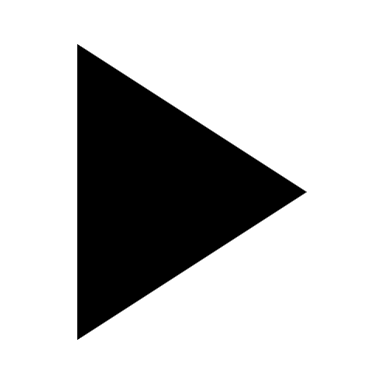
** Challenges**

* **Finding a simple, binary-labeled dataset suitable for heart disease prediction was more difficult than expected. Most available datasets were multiclass or required extensive cleaning. This challenged our goal of maintaining simplicity and interpretability.**
* **Feature selection required domain understanding to avoid discarding clinically relevant variables.**
* **Overfitting in early models necessitated regularization and proper cross-validation strategies.**

****

**Insights**

* **Lifestyle factors like smoking, obesity, and chronic stress, along with clinical indicators like high cholesterol and blood pressure, were consistently among the top predictors.**
* **Random Forest and Gradient Boosting provided strong predictive performance with robust handling of feature interactions.**

**FAQ**

**Q: What is the main objective of this project?**

**Answer:** The primary goal is to build a machine learning model that can predict whether a patient is at risk of heart disease based on clinical and lifestyle-related features. This supports early diagnosis and preventative healthcare.

**Q: What dataset did you use?**

**Answer:** We used a publicly available binary classification dataset from Kaggle that includes features such as age, cholesterol, blood pressure, smoking status, family history, and other key indicators relevant to heart disease.

**Q: Why did you choose a binary classification problem?**

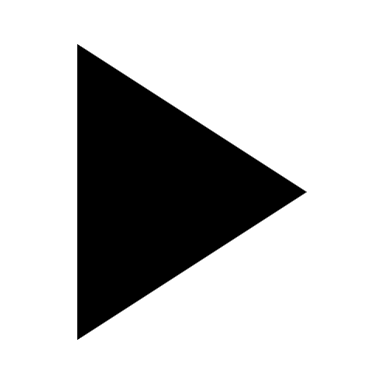
**Answer:** We wanted to simplify the modeling process and make the results easily interpretable for healthcare professionals. Binary classification (at-risk or not-at-risk) aligns well with real-world medical screening scenarios.

**Q: How did you evaluate model performance?**

**Answer:** We used accuracy, precision, recall, F1-score, and ROC-AUC to assess model performance. Given the importance of correctly identifying at-risk patients, we focused heavily on recall and F1-score.

**Q: What are the next steps for this project?**

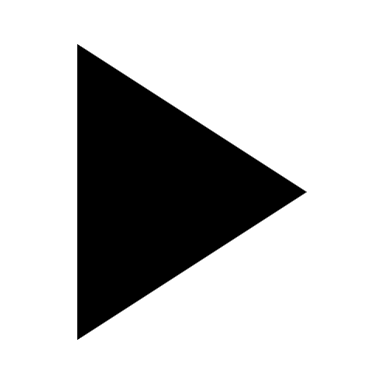
**Answer:** Future work includes testing the model with larger and more diverse datasets, integrating feedback from medical professionals, and developing a user-friendly web interface for clinical environments.

****

**Recommendations for Clinical Integration**

**To translate this model into practical use for healthcare professionals, the following recommendations are proposed:**

1. **Integration into EHR Systems:** Embed the model within electronic health record platforms to provide real-time risk assessments during patient visits.
2. **Clinical Decision Support:** Use the model to flag high-risk patients for early intervention, lifestyle counseling, or diagnostic testing.
3. **Patient Engagement:** Provide risk feedback to patients in a visual and interpretable format to promote health literacy and preventive action.
4. **Continuous Monitoring and Updates:** Retrain the model periodically with new patient data to maintain accuracy and reflect changes in population health trends.

****

**Conclusion**

**The project successfully demonstrated the potential of machine learning in predicting heart disease risk using a combination of clinical symptoms and lifestyle factors. With proper integration and oversight, the model can support proactive care, early diagnosis, and improved patient outcomes in real-world healthcare settings.**

**Team members**

**Abdelrahman Elsayed Mohamed**

**Abdallah Mohamed Abdallah**

**Mazen Ehab Gamal**

**Omar Ayman Mohamed**

**1**

**2**

**Mazen Mostafa Abo-El Yazeed**

**3**

**4**

**5**

**Seif El-Eslam Mohamed Yahia**

**6**

**5**

**Undersupervision of**

**Eng / Mahmoud Khorshid**