

CARDIO DISEASE PREDICTOR PRESENTATION

TEAM MEMBERS

HABIBA OSAMA :

ROMANY ADEL

GEORGE JOSEPH

FADY THABET

BASSANT MAGDY

INTRODUCTION

- This project applies Machine Learning techniques to health data.
- The goal is to build a predictive model that determines whether a person is likely to have cardiovascular disease or not.
- The model relies on several health factors, including:
 - Age
 - Weight & Body Mass Index (BMI)
 - Blood Pressure (systolic / diastolic)
 - Cholesterol and Glucose levels
 - Lifestyle-related factors

Dataset Overview

- Features: Age, Gender, Height, Weight, Blood Pressure (ap_hi, ap_lo), Cholesterol, Glucose, Lifestyle habits.
- Target variable: cardio (0 = No, 1 = Yes).
- Rows: ~70k (after cleaning → reduced).

DATA PREPROCESSING

- Removed unnecessary columns (Unnamed: 0, id).
- Converted age from days → years.
- Calculated $BMI = \text{weight} / (\text{height}^2)$.
- Filtered invalid/outlier values (BMI between 10–50, $ap_lo < ap_hi$, $ap_hi \leq 300$).
- Handled cholesterol inconsistencies by dropping mismatched cases.

EXPLORATORY DATA ANALYSIS

- Histograms: BMI, blood pressure distributions.
- Target distribution: Cardiovascular disease (imbalanced at first).
- Gender distribution: Males vs females.
- Heatmap: Correlation between features.

Handling Imbalance

- Used SMOTE (Synthetic Minority Oversampling Technique).
- Balanced the dataset ($y=0$ vs $y=1$).

MACHINE LEARNING MODELS

- Logistic Regression > acc= 0.947
- Decision Tree> acc=0.93
- Naive Bayes> acc= 0.91
- KNN> acc= 0.943
- Random Forest> acc= 0.95

DEEP LEARNING Model

- Neural Network Architecture:
 - Input layer (features).
 - Hidden Layer 1: 128 neurons + ReLU + Dropout.
 - Hidden Layer 2: 64 neurons + ReLU + Dropout.
 - Output Layer: Sigmoid (binary classification).
- Optimizer: Adam
- Loss: Binary Crossentropy
- Accuracy: 95% on test set.

REFERENCES

S

- Kaggle
dataset:<https://www.kaggle.com/datasets/akshatshaw7/cardiovascular-disease-dataset>
- Notebooklink:https://colab.research.google.com/drive/19wclJiFRfQV6wKJwiKShLEfDc9NxnW_3?usp=sharing

TEAM MEMBERS ROLES

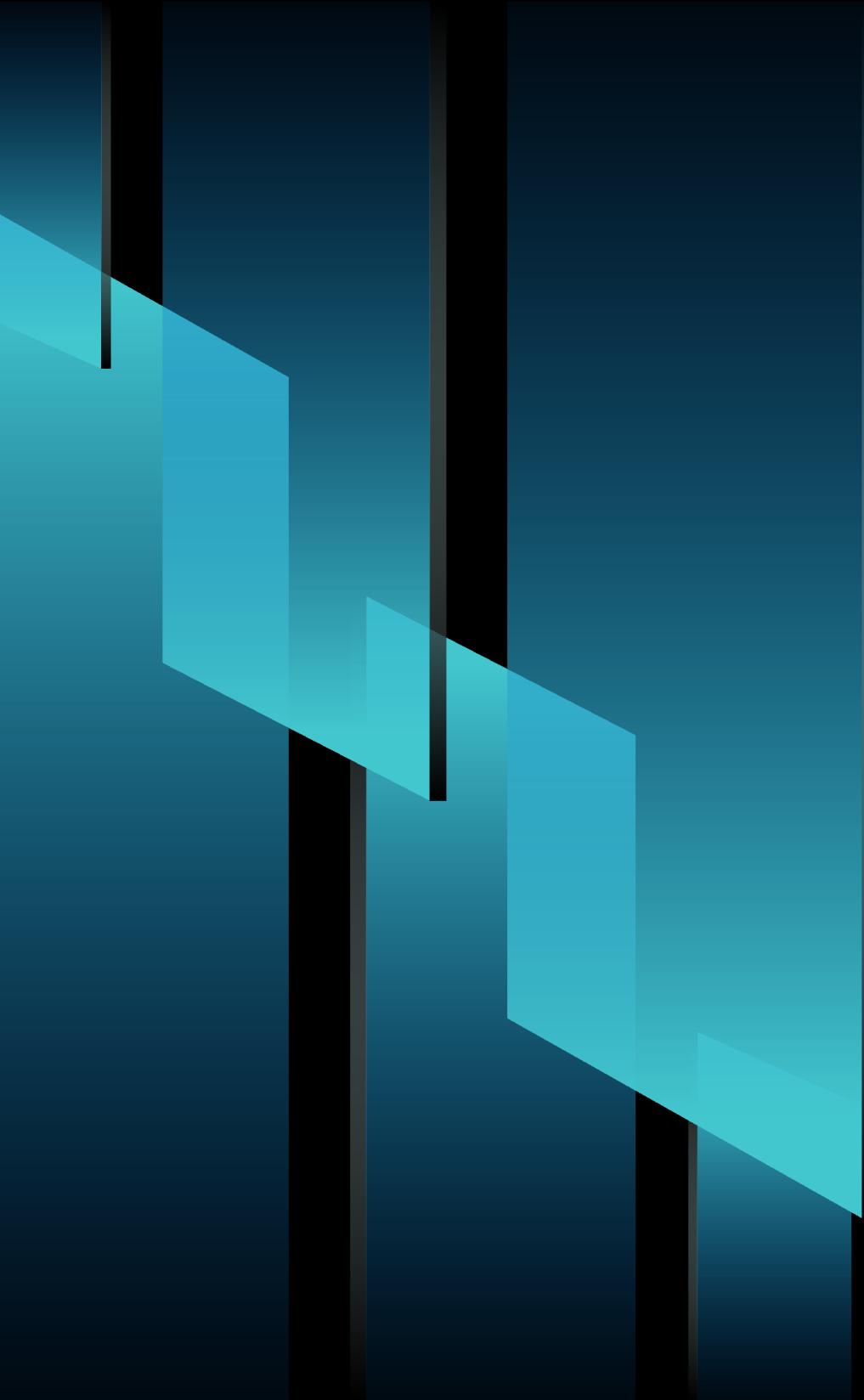
Bassant Magdy : EDA

Romany Adel : Data Cleaning & preprocessing

Habibia Osama : Data Visualization

Fady Thabet : AI models ,Neural Network & Dashboard

George Joseph : JUI



THANK YOU

