

HEART DISEASE PREDICTION



BY :
SUSAN MURITHI - 158864
SALIMA ALI - 169964

- Goal: Predict heart disease presence using patient data
- Task Type: Binary Classification
- Target: target (1 = Disease, 0 = No Disease)
- Dataset: UCI Heart Disease (Kaggle)

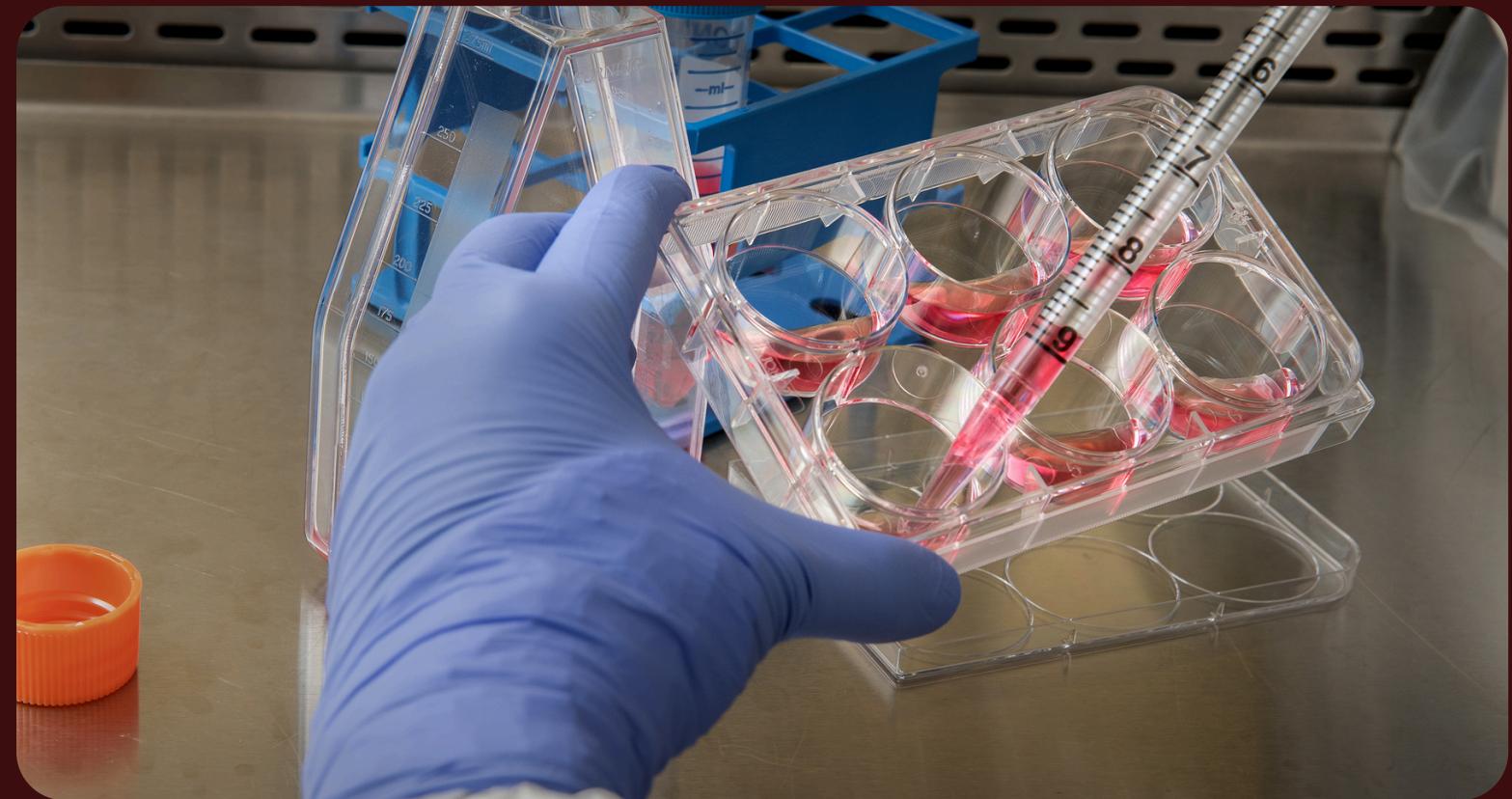


DEFINING THE PROBLEM

DATASET OVERVIEW

- ~304 records, 13 features
- Mix of numerical and categorical features
- No missing values
- Preprocessing:
 - Label encoding
 - Scaling (for k-NN and ANN)
 - 50% training / 50% testing split

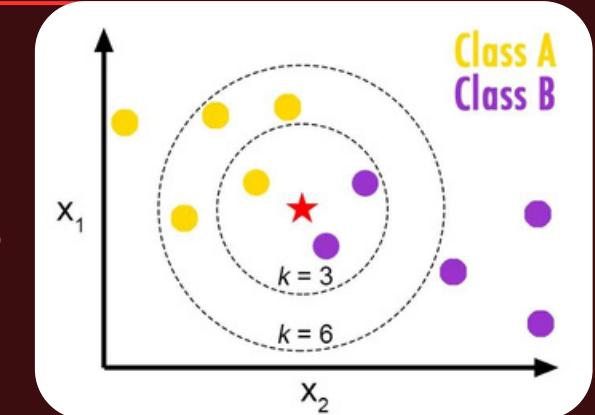
| Feature | Description | Type |
|-----------------|---|-------------|
| age | Age of the patient in years | Numerical |
| sex | Gender of the patient (0 = female, 1 = male) | Categorical |
| cp | Chest pain type (0 = typical angina, 1 = atypical, 2 = non-anginal, 3 = asymptomatic) | Categorical |
| trestbps | Resting blood pressure (in mm Hg) | Numerical |
| chol | Serum cholesterol in mg/dl | Numerical |
| fbs | Fasting blood sugar > 120 mg/dl (1 = true, 0 = false) | Categorical |
| restecg | Resting ECG results (0 = normal, 1 = ST-T wave abnormality, 2 = LV enlargement) | Categorical |
| thalach | Maximum heart rate achieved | Numerical |
| exang | Exercise-induced angina (1 = yes, 0 = no) | Categorical |
| oldpeak | ST depression induced by exercise relative to rest | Numerical |
| slope | Slope of the peak exercise ST segment (0 = upsloping, 1 = flat, 2 = downsloping) | Categorical |
| ca | Number of major vessels colored by fluoroscopy (0–3) | Numerical |
| thal | Thalassemia (1 = normal, 2 = fixed defect, 3 = reversible) | Categorical |



MODEL SUMMARIES

k-NN:

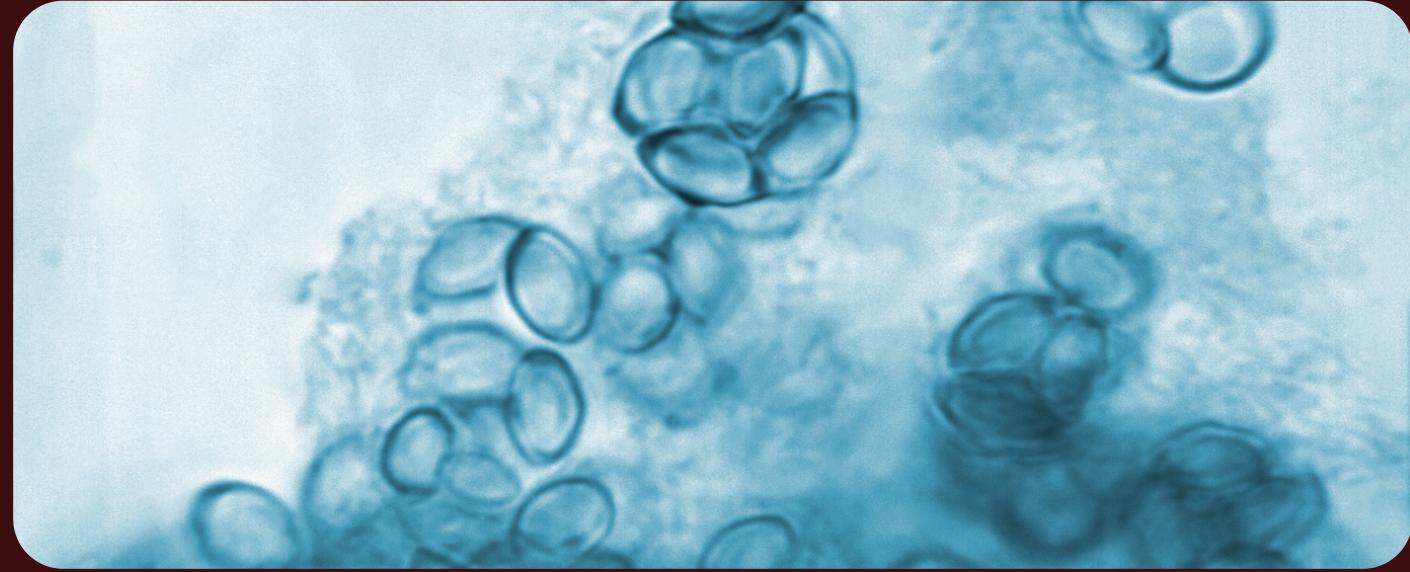
Simple, distance-based, needs scaling



Decision Tree:
Interpretable, prone to overfitting

ANN: Captures complex patterns, needs tuning





MODEL ANALYSIS

| | Model | Accuracy (%) | True Positives (TP) | True Negatives (TN) | Total Test Cases |
|---|---------------|---------------------|----------------------------|----------------------------|-------------------------|
| 0 | KNN | 84.0 | 68 | 59 | 152 |
| 1 | Decision Tree | 74.0 | 58 | 55 | 152 |
| 2 | ANN | 80.0 | 65 | 57 | 152 |

CONCLUSION



Ranking : KNN > ANN > Decision Tree

- KNN is the best performing model for this dataset, suggesting that proximity-based classification worked well.
- ANN is a close second and might outperform with further tuning.
- Decision Trees, while interpretable, underperformed slightly and may require pruning or boosting for better results.

REAL LIFE APPLICABILITY

Doctors and Nurses: The model can help them quickly check if someone might have heart problems during regular hospital visits.

Health Apps: It can be used in mobile apps to help people in remote or rural areas check their heart health without needing to visit a doctor.

Healthy Living Advice: If the model says someone is at risk, they can get advice on how to eat better, exercise, or reduce stress to stay healthy.

Insurance Companies: It can help insurance companies understand a person's health risk before giving them a health cover or deciding the price of a plan.

IMPROVEMENTS

- Model Tuning
- Additional Evaluation Metrics
- Feature Importance



THANK YOU

QUESTIONS