NAME

Mahrael Raafat Fawzi

Tasnim Reda Shehata

Abdelqader Ismail Abdelqader

ID

2305305

2305243

2305156

Heart Disease Risk Prediction System

1. Introduction

Cardiovascular diseases are a leading cause of mortality worldwide. Early detection and risk assessment play a crucial role in preventing severe outcomes. This project focuses on developing a heart disease risk prediction system using a combination of **Machine Learning (ML)** and an **Expert System**. The primary goal is to evaluate the efficiency of automated decision-making models and rule-based expert knowledge in predicting heart disease risks.

2. Methodology

This system integrates two core approaches:

- **Machine Learning Model**: A **Decision Tree Classifier** is used to classify individuals based on their health attributes.
- **Expert System**: Implemented using the **Experta** library, the system applies rule-based reasoning to classify heart disease risk levels.

2.1 Libraries Used

The following libraries were used in the implementation:

- **pandas**: For data manipulation and preprocessing.
- **numpy**: For numerical computations.
- **scikit-learn**: To build and evaluate the Decision Tree Classifier.
- **matplotlib & seaborn**: For data visualization.
- **experta**: To create the rule-based expert system.
- **streamlit**: To develop an interactive web-based user interface.

2.2 Data Preprocessing and Feature Selection

- **Dataset**: The system utilizes a heart disease dataset consisting of various health indicators, such as age, cholesterol level, blood pressure, and other cardiovascular-related parameters.
- **Preprocessing Steps**:
  - Handling missing values.
  - Encoding categorical features.
  - Standardizing numerical attributes.
- **Feature Selection**: `SelectKBest` method with ANOVA F-value was applied to select the most relevant features for model training.

2.3 Machine Learning Model (Decision Tree Classifier)

- A **Decision Tree Classifier** was chosen for its interpretability and efficiency.
- Hyperparameter tuning was conducted using `GridSearchCV` to optimize parameters such as `max_depth`, `min_samples_split`, and `criterion`.
- Model evaluation metrics include **Accuracy, Precision, Recall, F1-score, and Confusion Matrix**.

2.4 Expert System Implementation

- The expert system was built using the **Experta** library, defining rules based on medical thresholds (e.g., high cholesterol, hypertension levels, and BMI values).
- The rule engine classifies individuals into different risk categories: **Low, Medium, or High Risk**.

2.5 Model Comparison and Visualization

- The performance of the ML model and the expert system was compared using standard evaluation metrics.
- Results were visualized using **matplotlib** and **seaborn** to illustrate differences in classification accuracy and risk assessment.

3. Results and Discussion

- The **Decision Tree Classifier** achieved an accuracy of **X%**, indicating its potential for heart disease risk prediction.
- The **Expert System** provided interpretable results aligned with medical domain knowledge but lacked adaptability to complex data patterns.
- The **comparison** showed that while ML models are data-driven and adaptive, expert systems offer transparency in decision-making.
- Future improvements could involve integrating more advanced ML models (e.g., Random Forest, XGBoost) and refining expert rules based on medical research.

4. Deployment Considerations

- The system is designed to be interactive and user-friendly using **Streamlit** and **Voila**.
- A web-based interface allows users to input health attributes and receive risk predictions.
- A `requirements.txt` file ensures easy installation of dependencies, and containerization with **Docker** can facilitate deployment.
- Codebase is structured into logical folders with appropriate documentation.
- A **GitHub repository** is maintained, including a `README.md` file with setup instructions and usage examples.

5. Conclusion

This project demonstrates a **hybrid approach** to heart disease risk prediction by leveraging both **Machine Learning** and **Expert Systems**. While ML models exhibit high predictive accuracy, expert systems provide **interpretable and medically grounded** insights. Combining both approaches enhances reliability and transparency in clinical decision-making. Future work can focus on integrating real-time health monitoring data and refining expert rules for greater precision.

6. References

- Relevant medical studies and datasets used.
- Documentation for **Scikit-learn**, **Experta**, and **Streamlit**.
- Other sources supporting the feature selection and classification approach.