Model Card: Heart Attack Risk Prediction

Model Overview

This model card provides details for the Decision Tree and Logistic Regression models used in predicting heart attack risk based on health-related parameters. These models were trained and evaluated using a dataset containing various clinical and lifestyle factors.

Intended Use

- This model is designed to predict heart attack risk based on key health indicators.
- Intended for educational and exploratory purposes, not for medical diagnosis.
- Can be extended for real-world applications with more extensive clinical validation.

Training Data

- **Dataset Source:** Kaggle
- **Number of Records:** 50,000
- **Number of Features:** 19
- **Features Used:**
    - Age
    - Gender
    - Cholesterol Level
    - Blood Pressure
    - Smoking Status
    - Exercise-Induced Angina
    - Diabetes
    - Family History

Model Details

1. Logistic Regression Model

- Algorithm: Logistic Regression (binary classification)
- Hyperparameters Used:
    - Max-iter: 100
    - C: 10
    - fit_intercept: True

- **Performance:**
    - Accuracy: 50.00%
    - Mean Squared Error (MSE): 0.49
    - Confusion Matrix:

      [[0 1511 0]

      [0 3754 0]

      [0 2235 0]]

2. Decision Tree Model

- Algorithm: Decision Tree Classifier
- Hyperparameters Used:
    - max_depth: 5
    - criterion: Gini
- Performance:
    - Accuracy: 49.89%
    - Mean Squared Error (MSE): 0.55
    - Confusion Matrix:

      [[12 1428 61]

      [66 3570 118]

      [43 2129 63]]

Limitations & Considerations

- The accuracy of both models is low, suggesting that more advanced methods, such as Neural Networks, may be needed for better performance.
- Feature selection & engineering may improve predictive power.
- Data quality and balance significantly impact model reliability.
- Further medical validation is required before real-world deployment.

Future Improvements

- Implementing Neural Networks to capture non-linearity in the data.
- Collecting additional health-related variables.
- Exploring ensemble methods like boosting and stacking.

Ethical Considerations

- Bias & Fairness: Model predictions must be evaluated for fairness across different demographic groups.
- Privacy: Ensure patient data confidentiality when applying in real-world scenarios.
- Interpretability: Decision trees offer transparency, while logistic regression provides probability estimates.

References

- Dataset Source: Kaggle :https://www.kaggle.com/datasets/arifmia/heart-attack-risk-dataset/data