**Heart Failure Prediction**

**CBIO313: Data mining and Machine learning**

**Dr. Mohamed El-Sayeh**

**May 2024**

**Made by:**

Jomana Hesham Kamal

**ID:**

211002130

**Introduction**

Heart failure is a critical medical condition that occurs when the heart is unable to pump sufficient blood to meet the body's needs. Early prediction of heart failure can significantly improve patient outcomes by enabling timely intervention and treatment. In this study, we evaluate the performance of several machine learning classifiers for predicting heart failure. The classifiers assessed include K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine (SVM). The performance of these models is compared using various metrics such as precision, recall, F1-score, and accuracy.

**Steps**

1. **Data Collection**: The dataset used for this study contains medical records of patients, including features relevant to heart failure prediction.
2. **Data Preprocessing**: The data underwent preprocessing steps such as handling missing values, feature scaling, and encoding categorical variables to ensure compatibility with the machine learning algorithms.
3. **Model Training**: The dataset was split into training and testing sets. Each classifier was trained on the training set using default hyperparameters.
4. **Model Evaluation**: The performance of each model was evaluated on the testing set. Metrics such as precision, recall, F1-score, and accuracy were calculated to compare the models' effectiveness.

**Results**

K-Nearest Neighbors (KNN)

* Precision: 0.970
* Recall: 0.980
* F1-score: 0.975
* Accuracy: 0.985

| Class | Precision | Recall | F1-score | Support |
| --- | --- | --- | --- | --- |
| 0 | 0.99 | 0.99 | 0.99 | 698 |
| 1 | 0.97 | 0.98 | 0.98 | 302 |

Random Forest

* Precision: 0.980
* Recall: 0.990
* F1-score: 0.985
* Accuracy: 0.991

| Class | Precision | Recall | F1-score | Support |
| --- | --- | --- | --- | --- |
| 0 | 1.00 | 0.99 | 0.99 | 698 |
| 1 | 0.98 | 0.99 | 0.99 | 302 |

Decision Tree

* Precision: 0.974
* Recall: 0.983
* F1-score: 0.979
* Accuracy: 0.987

| Class | Precision | Recall | F1-score | Support |
| --- | --- | --- | --- | --- |
| 0 | 0.99 | 0.99 | 0.99 | 698 |
| 1 | 0.97 | 0.98 | 0.98 | 302 |

Logistic Regression

* Precision: 0.763
* Recall: 0.735
* F1-score: 0.749
* Accuracy: 0.851

| Class | Precision | Recall | F1-score | Support |
| --- | --- | --- | --- | --- |
| 0 | 0.89 | 0.90 | 0.89 | 698 |
| 1 | 0.76 | 0.74 | 0.75 | 302 |

Support Vector Machine (SVM)

* Precision: 0.921
* Recall: 0.891
* F1-score: 0.906
* Accuracy: 0.944

| Class | Precision | Recall | F1-score | Support |
| --- | --- | --- | --- | --- |
| 0 | 0.95 | 0.97 | 0.96 | 698 |
| 1 | 0.92 | 0.89 | 0.91 | 302 |

**Conclusion**

The comparative analysis of various machine learning classifiers for heart failure prediction revealed the following insights:

* The Random Forest classifier demonstrated the highest overall performance with an accuracy of 0.991 and a balanced precision, recall, and F1-score across both classes.
* K-Nearest Neighbors (KNN) and Decision Tree classifiers also performed well, with accuracies of 0.985 and 0.987, respectively.
* The Support Vector Machine (SVM) achieved a respectable accuracy of 0.944 but fell short of the top-performing models.
* Logistic Regression had the lowest performance among the tested classifiers, with an accuracy of 0.851, indicating it might not be the best choice for this specific task.

In conclusion, the Random Forest classifier emerged as the most effective model for predicting heart failure, offering the best balance between precision, recall, and overall accuracy.

**Gethub Repo Link**

https://github.com/jomanaa03/Heart\_Failure\_Prediction