Performance Comparison of Data Mining Techniques for Predicting Heart Disease Survivability

Author: Lakshmi, K. R., Krishna, M. V., & Kumar, S. P. (2013).

Performance comparison of data mining techniques for predicting of heart disease survivability. International Journal of Scientific and Research Publications

Abstract

Heart disease is one of the leading causes of death globally, making early prediction and diagnosis essential for improving patient outcomes. This study evaluates and compares the performance of various data mining techniques, including Decision Trees, Naïve Bayes, Neural Networks, and Support Vector Machines (SVM), for predicting heart disease survivability. Using a dataset comprising patient demographics, clinical test results, and medical history, the performance of these algorithms was assessed through metrics such as accuracy, precision, recall, and F1-score. The findings reveal the strengths and weaknesses of each technique, emphasizing the potential of data mining in supporting healthcare professionals. The study underscores the importance of optimizing predictive models to enhance the accuracy and reliability of heart disease diagnostics.

Introduction

Heart disease remains a major health challenge, contributing to significant mortality and morbidity rates worldwide. Early prediction of heart disease survivability is critical for timely medical intervention, which can significantly reduce the risk of severe outcomes. With the advent of data mining techniques, healthcare professionals can utilize computational models to analyze vast clinical datasets, uncover patterns, and predict disease outcomes. This paper focuses on comparing the performance of popular data mining algorithms, such as Decision Trees, Naïve Bayes, Neural Networks, and Support Vector Machines (SVM), in predicting heart disease survivability. The study aims to provide insights into the effectiveness of these techniques and their potential applications in healthcare decision-making.

Methodology

1. **Data Collection**:
   The study utilized a clinical dataset containing information on patient demographics, medical history, and test results. Key attributes included age, gender, blood pressure, cholesterol levels, and other relevant health indicators.
2. **Preprocessing**:
   - **Data Cleaning**: Missing or inconsistent data points were addressed through imputation techniques.
   - **Feature Selection**: Attributes most relevant to heart disease survivability were identified using statistical and correlation analysis.
   - **Normalization**: Data was standardized to ensure compatibility across algorithms.
3. **Algorithms Evaluated**:
   - **Decision Trees**: A tree-based structure that splits data based on key attributes to classify outcomes.
   - **Naïve Bayes**: A probabilistic classifier based on Bayes' theorem, suitable for handling categorical data.
   - **Neural Networks**: A computational model inspired by biological neural systems, effective for complex, nonlinear data.
   - **Support Vector Machines (SVM)**: A classification technique that identifies the optimal hyperplane for separating data points.
4. **Performance Metrics**:
   The models were evaluated using the following metrics:
   - **Accuracy**: Overall correctness of predictions.
   - **Precision**: Proportion of true positive predictions.
   - **Recall**: Ability to identify actual positive cases.

  - **F1-score**: Harmonic mean of precision and recall, balancing false positives and false negatives.

5. **Implementation**:
   The algorithms were implemented using standard machine learning tools and frameworks. A comparative analysis was conducted to highlight their strengths and limitations.

Conclusion

The study demonstrates that data mining techniques offer valuable tools for predicting heart disease survivability, with each algorithm exhibiting unique advantages and limitations. While Decision Trees and Naïve Bayes provide interpretability and simplicity, Neural Networks and SVMs offer superior accuracy for complex datasets. The findings highlight the importance of selecting appropriate algorithms and optimizing their parameters to improve prediction reliability. Future work should focus on integrating real-time patient data and exploring ensemble approaches to enhance model performance further. This research underscores the transformative potential of data mining in enabling data-driven healthcare solutions.