Report on Research Papers

-2310080078

Mythri

PAPER1 Title:

An integrated data mining algorithms and meta-heuristic technique to predict the readmission risk of diabetic patients.

Research Objective:

1.Integrate data mining algorithms and meta-heuristic techniques to develop a more accurate classifier;

2. Predict the early readmission risk of diabetic patients;

3. Employ a Chi-square analysis to identify/confirm and rank the significant factors affecting the early readmission risk;

4. Apply several classification algorithms such as random forest, neural network, and support vector machine to introduce the best-fitted model; and

5. Adjust the hyper-parameters of the support vector machine using a Genetic Algorithm to increase the prediction performance.

Dataset:

The research dataset was obtained from the UC Irvine Machine Learning Repository, including 101765 instances with 50 features representing patient and hospital outcomes, collected from 130 US hospitals.

Models Used:

This research applies various classification algorithms, including SVM, RF, and NN, using Rapid Miner software. Moreover, hyper- parameter tuning for the SVM is done by integrating a Genetic algo rithm using Python software, which identifies the most appropriate “c” and “gamma” parameters.

Limitations:

While SMOTE performs well on many datasets, it has the drawback of generating noisy samples and needs modification for “Nominal” and “Continuous” features

Future Work:

Furthermore, similar databases could be modelled with the designed models in this research, particularly with GA-SVM, and compare the results. GA approach might be used to further enhance the accuracy of other classification algorithms. Other meta-heuristic algo rhythms, such as particle swarm optimization (PSO) and Grey Wolf Optimizer (GWO), can also be used for this purpose. Future researchers might develop a hybrid feature selection technique, for example, based on PSO and a Chi-square analysis, to enhance the prediction accuracy.

Method:

- Dataset description
- Data preprocessing
    - Data cleansing

    - Sampling
    - Missing values management
- Feature selection/confirmation/ranking
  via Chi-square analysis
- Normalization
- Modelling
    - Support vector machine
    - Genetic algorithm
    - Neural network
    - Random forest
- Validation

Importance:

Reduces hospital readmission costs and enhances patient care.

Can be applied to similar healthcare datasets for predictive analysis

Limitations

- The accuracy of GA-SVM still lags behind Random Forest.
- Alternative classification techniques may further enhance prediction performance

Results

- RF (Accuracy: 74.04%), GA-SVM (73.52%), SVM (72.40%), NN (70.44%).
- GA-SVM improves SVM accuracy by 1.12%

PAPER2 Title:

Application of Data Mining Technology in Exam Score Analysis

Research Objective

- Analyze the relationship between exam scores and influencing factors.
- Develop a model for performance prediction based on data mining techniques.

Dataset

- Collected from a university educational administration system.
- Includes student scores, number of questions attempted, and demographic data.

Models Used

- Improved K-means and Apriori algorithms.
- Decision tree (C4.5 and ID3) for performance prediction.

Limitations

- The model does not consider psychological or environmental factors.
- The dataset might be biased due to a lack of diverse student groups.

Future Work

- Implementing deep learning models to refine predictions.
- Expanding the dataset to include broader demographics.

Method

- Data preprocessing: Feature extraction and discretization.

- Training classifiers using K-means and Apriori for association rule mining.

Importance

- Helps students and teachers understand performance trends.
- Can be integrated into educational systems for personalized recommendations.

Limitations

- External factors affecting student performance are not included.
- Limited generalizability due to dataset constraints.

Results

- C4.5 outperformed ID3 in exam performance prediction.
- Improved K-means and Apriori provided valuable insights into study patterns.

PAPER3 Title:

A Novel Study

Research Objective

- Investigate the application of machine learning techniques for classification tasks.
- Develop an optimized model with improved accuracy and efficiency.

Dataset

- The dataset details were not explicitly provided, but preprocessing steps were emphasized.

Models

- Implemented various machine learning classifiers.
- Used optimization techniques to enhance model performance.

Limitations

- Computational constraints affecting model scalability.
- Need for more diverse datasets for generalization.

Future Work

- Explore deep learning methods for improved performance.
- Apply the model to different domains for validation.

Method

- Data preprocessing: Cleaning, transformation, and feature extraction.
- Model training and evaluation using standard machine learning techniques.

Importance

- Enhances classification accuracy in complex datasets.
- Can be adapted for various real-world applications.

Limitation

- Model performance is limited by dataset quality and size.

Results

- Improved accuracy compared to baseline models.
- Model performance validated through multiple evaluation metrics.