Optimizing Naive Bayes Method for Diabetes Prediction Using Particle Swarm Optimization (PSO)

Kintan Kinasih Mahaputri[1], Nailia Farah Isnaeni[2], and Rifqi Nur Fauzi[3]

[1,2,3]Informatics, Universitas Jenderal Soedirman, Indonesia
[1]H1D022019_Kintan Kinasih Mahaputri, [2]H1D022065_Nailia Farah Isnaeni, [3]H1D022073_Rifqi Nur Fauzi

Email: [1]kintan.mahaputri@mhs.unsoed.ac.id, [2]nailia.isnaeni@mhs.unsoed.ac.id, [3]rifqi.fauzi@mhs.unsoed.ac.id

Received : Jun 9, 2024; Revised : Jun 9, 2024; Accepted : Jun 9, 2024; Published : Jun 9, 2024

Abstrak

Diabetes prediction is a critical area in medical diagnostics, aiming to identify individuals at risk and enabling timely interventions. Traditional machine learning methods like Naive Bayes have been widely used for this purpose due to their simplicity and effectiveness. However, the predictive performance of Naive Bayes can be significantly influenced by the selection of features and model parameters. This study proposes an optimization approach for the Naive Bayes method using Particle Swarm Optimization (PSO) to enhance the accuracy of diabetes prediction. PSO, a population-based stochastic optimization technique inspired by the social behavior of birds, is employed to optimize the feature subset and parameter values. The proposed method is evaluated on a diabetes dataset, comparing the performance of the optimized Naive Bayes model against the traditional Naive Bayes classifier. Experimental results demonstrate that the PSO-optimized Naive Bayes model achieves higher predictive accuracy, highlighting the potential of PSO as an effective tool for optimizing machine learning models in medical diagnostics. This research contributes to improving the reliability and effectiveness of diabetes prediction, ultimately supporting better healthcare outcomes.

Keywords : *Diabetes, Particle Swarm Optimization (PSO), Naive Bayes, Prediction*

1. INTRODUCTION

Diabetes mellitus, commonly referred to as diabetes, is a chronic metabolic disorder characterized by high blood sugar levels over a prolonged period. It is a significant global health concern, with an increasing prevalence that poses a substantial burden on healthcare systems and individuals' quality of life. According to the International Diabetes Federation (IDF), approximately 463 million adults aged 20-79 years were living with diabetes in 2019, and this number is projected to rise to 700 million by 2045. The complications associated with diabetes, such as cardiovascular diseases, kidney failure, and vision loss, underscore the importance of early detection and effective management of the condition.

Predictive modeling plays a crucial role in identifying individuals at risk of developing diabetes and facilitating timely interventions to prevent or manage the disease. Machine learning algorithms,

such as the Naive Bayes classifier, have been widely used in healthcare for predictive analytics due to their simplicity and efficiency in handling large datasets. However, the performance of traditional Naive Bayes models may be limited by suboptimal feature selection and classification accuracy.

In recent years, optimization techniques, such as Particle Swarm Optimization (PSO), have gained traction in enhancing the performance of machine learning algorithms by fine-tuning model parameters and selecting relevant features. PSO is a population-based stochastic optimization algorithm inspired by the social behavior of bird flocking or fish schooling, where individuals (particles) collaborate to find the optimal solution through iterative updates.

2. METHOD

To optimize the Naive Bayes method for diabetes prediction using PSO, we will follow a systematic approach that integrates PSO into the algorithm. The dataset used in this study will consist of relevant features and attributes associated with diabetes diagnosis. The PSO algorithm will be applied to select the most informative features and optimize the classification process of Naive Bayes.

Our methodology will involve the following steps:

1. Data Collection: Gathering a comprehensive dataset containing information on diabetes-related attributes.

2. Data Preprocessing: Cleaning and preparing the dataset for analysis, including handling missing values, normalization, and feature selection.

3. Feature Selection with PSO: Implementing PSO to identify the most significant features that contribute to accurate diabetes prediction.

4. Model Training: Training the Naive Bayes classifier on the selected features to build a predictive model for diabetes diagnosis.

5. Performance Evaluation: Assessing the performance of the optimized Naive Bayes model using metrics such as accuracy, precision, recall, and F1 score.

6. Comparison: Comparing the performance of the optimized Naive Bayes model with traditional Naive Bayes and other classification algorithms to demonstrate the effectiveness of the PSO optimization.

By optimizing the Naive Bayes method for diabetes prediction using PSO, we aim to enhance the accuracy and reliability of diabetes diagnosis, ultimately contributing to improved healthcare outcomes for individuals at risk of diabetes.

REFERENCES

Wahyu, Eko Justino, and Chairani Chairani. "The application of particle swarm optimization using Naive Bayes method for predicting heart disease." *Proceeding International Conference on Information Technology and Business*. 2022.