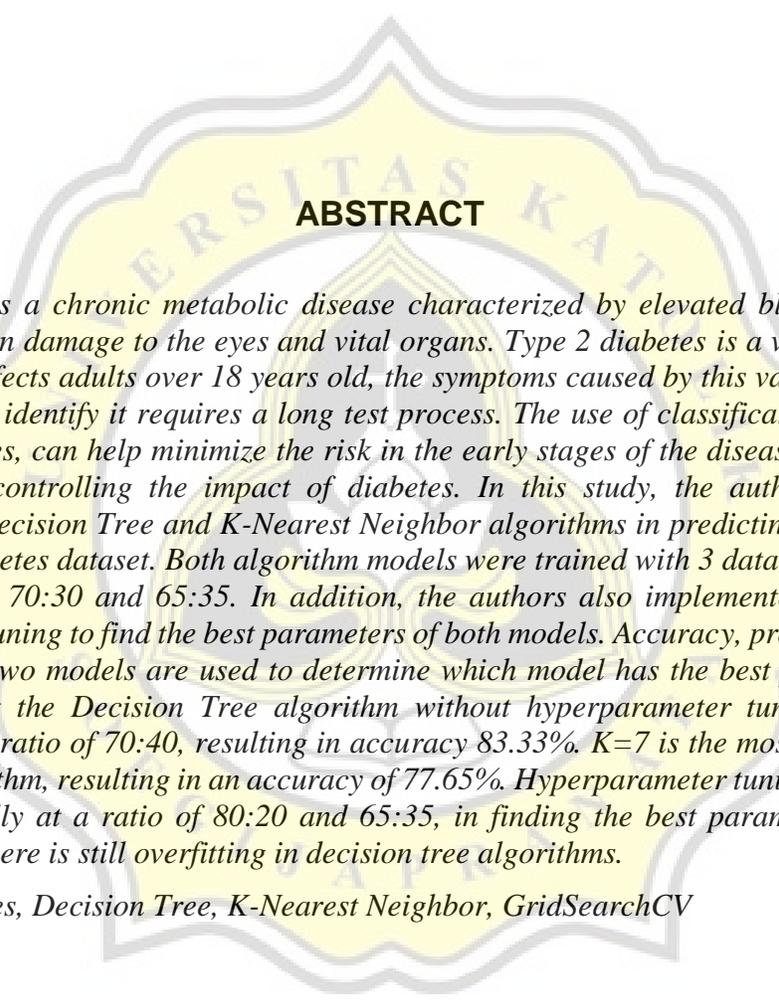**PROJECT REPORT**

COMPARISON OF DECISION TREE ALGORITHM AND K-NEAREST NEIGHBOR (KNN) ALGORITHM PERFORMANCE IN DIABETES CASE STUDY

SILVANO PRATAMA JUBILATE DEO
19.K1.0016

Faculty of Computer Science
Soegijapranata Catholic University
2023

ABSTRACT

Diabetes is a chronic metabolic disease characterized by elevated blood sugar levels, which can result in damage to the eyes and vital organs. Type 2 diabetes is a variant of diabetes that most often affects adults over 18 years old, the symptoms caused by this variant are not very noticeable and to identify it requires a long test process. The use of classification algorithms in predicting diabetes, can help minimize the risk in the early stages of the disease and help health practitioners in controlling the impact of diabetes. In this study, the authors compare the performance of Decision Tree and K-Nearest Neighbor algorithms in predicting diabetes, on the Pima Indian Diabetes dataset. Both algorithm models were trained with 3 dataset sharing ratios, which are 80:20, 70:30 and 65:35. In addition, the authors also implemented GridSearchCV hyperparameter tuning to find the best parameters of both models. Accuracy, precision, recall and F-1 score of the two models are used to determine which model has the best performance. The results show that the Decision Tree algorithm without hyperparameter tuning has the best performance at a ratio of 70:40, resulting in accuracy 83.33%. K=7 is the most optimal K value in the KNN algorithm, resulting in an accuracy of 77.65%. Hyperparameter tuning GridSearchCV can work optimally at a ratio of 80:20 and 65:35, in finding the best parameters in decision algorithms. But there is still overfitting in decision tree algorithms.

Keyword: Diabetes, Decision Tree, K-Nearest Neighbor, GridSearchCV

6