PROJECT REPORT

COMPARISON OF K-NEAREST NEIGHBORS (KNN) AND NAÏVE BAYES ALGORITHMS FOR FRAUD DETECTION OF CREDIT CARD TRANSACTIONS

Jefry Roberto Hasudungan Simanjuntak

19K10057

Faculty of Computer Science

Soegijapranata Catholic University

2023

ABSTRACT

This study aims to evaluate the performance of the KNN and Naïve Bayes algorithms in Detection of credit card transactions. Two datasets, Data 1 and Data 2, were utilized, each employing different outlier handling techniques. The experimental results highlight the performance of the KNN and Naïve Bayes algorithms for each dataset. For Data 1, missing values were removed, while for Data 2, outlier values were eliminated. The following results were obtained for Data 1 (missing value removal) are KNN algorithm in Data1 KNN Oversampling (20:80) achieved an accuracy of 98.98. Naïve Bayes algorithm in Data1 Naivebayes (30:70) achieved an accuracy of 95.16%.And for Data 2 (outlier removal) are KNN algorithm in Data2 KNN Oversampling (20:80) achieved an accuracy of 99.79%. Naïve Bayes algorithm in Data2Biasa Multinomial Naïve Bayes (30:70) achieved an accuracy of 99.31%. The findings indicate that the KNN algorithm demonstrates robustness in handling datasets with numerous outliers. It effectively addresses data imbalance through oversampling and undersampling techniques. On the other hand, Naïve Bayes exhibits limitations in handling datasets with a high number of outliers and performs better when the dataset contains fewer outliers. This research provides insights into the effectiveness of the KNN and Naïve Bayes algorithms for outlier handling in data classification, contributing to the selection of appropriate algorithms for handling specific types of datasets.

Keywords: KNN algorithm, Naïve Bayes algorithm, outlier handling, data classification, data preprocessing.