PROJECT REPORT

THE EFFECT OF CHI-SQUARE FEATURE SELECTION ON THE NAIVE BAYES ALGORITHM IN ANALYZING THE SENTIMENT OF GOJEK APPLICATION REVIEWS ON GOOGLE PLAY STORE

RAFAEL HANDIKA DWINANTA
21.K1.0028

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT

This study analyzes customer sentiment in reviewing the Gojek application to find out whether Chi-Square feature selection can improve the performance of the sentiment analysis model. This study uses 12,000 Gojek review data, starting with labeling positive, negative, or neutral based on user ratings of the reviews. Naive Bayes with and without Chi-Square feature selection is used in testing related to accuracy, precision, recall, and F1 score. The best performance is obtained by using alpha 0.5 combined with the best 2000 Chi-Square features, which produces 86.96% accuracy, 87.84% precision, 86.96% recall, and 85.29% F1 score on imbalanced data. SMOTE is also used to handle the low number of neutral reviews, but it produces lower accuracy. In conclusion, Chi-Square feature selection in the Naive Bayes algorithm can improve model accuracy on imbalanced and balanced datasets.

Keyword: sentiment analysis, Naive Bayes, Chi-Square, feature selection, Gojek reviews