ABSTRACT

Novia Putri Bahirah, 10120869

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI CUSTOMER CHURN PADA PLATFORM E-COMMERCE

Thesis. Department of Information Systems, Faculty of Computer Science and Information Technology, Gunadarma University, 2024.
Keywords: Churn, CRISP-DM, Data Mining, E-Commerce, Feature Importance

(xxi + 188 + L-24)

A company with a high churn rate can have a negative impact on the business it manages. This research aims to create a machine learning model to predict customer churn in e-commerce and identify features in the dataset that have a significant effect on the probability of potential churn. The research method used is CRISP-DM, including Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation. Three classification algorithms were applied, namely Gradient Boosting Classifier, XGBoost, and CatBoost Classifier. The evaluation was conducted at various test time periods and XGBoost was selected as the best model because it has an AUC value in the range of 0.87 to 0.90 indicating that this model can distinguish between positive (churn) and negative (non-churn) classes with good performance, has the highest recall metric and the least number of false negatives compared to other algorithms indicating that XGBoost is better at identifying most customers who actually churn in e-commerce, and the metric values tend to be stable indicating that the model is able to handle patterns or changes in trends in e-commerce datasets with good performance at various test time periods. The results show that the features that have the most significant effect on the probability of potential customer churn based on the results of feature importance from XGBoost, namely frequency, recency, and monetary.

Bibliography (2000-2025)