

ABSTRAK

E-commerce selalu membutuhkan *customer* dan tanpa *customer*, *e-commerce* tidak akan bisa bertahan. Dalam *e-commerce*, *customer* akan membeli produk karena pengalaman pengguna yang baik, pengiriman cepat dan handal, personalisasi produk dan rekomendasi, promosi dan diskon, kemudahan pembayaran, program loyalitas dan *reward*, *review* produk dari *customer* sebelumnya serta rekomendasi berdasarkan preferensi pembelian sebelumnya. Namun, pada kenyataannya, *e-commerce* menghadapi beberapa masalah seperti *review* produk yang buruk, kategori produk yang terbatas, kurangnya personalisasi, dan tingkah laku *customer* yang tiba-tiba meninggalkan *e-commerce* yang menyebabkan *customer* akan beralih dan tidak menggunakan *e-commerce* tersebut. Salah satu metode yang populer dalam mengatasi masalah ini adalah *logistic regression*, sebuah teknik *machine learning* yang terbukti efektif dalam berbagai prediksi biner. Penelitian ini bertujuan untuk mengimplementasikan metode *logistic regression* dalam memprediksi *churn customer* dan tingkat *churn rate* dengan meningkatkan akurasi dan menyusun rekomendasi penanganan *churn customer* berdasarkan hasil visualisasi data yang sudah dimodelkan.

Dalam penelitian ini, digunakan dataset *Fashion Campus*, di mana *preprocessing* dilakukan melalui *data cleaning*, *data integration*, *data transformation*, *data reduction*, *exploratory data analysis*, dan *feature engineering*. Variabel yang digunakan dibagi menjadi dua yaitu dependen sebagai hasil prediksi yaitu '*segmentation_churn*' dan variabel independen sebagai variabel input yang memprediksi '*basket_size*', '*total_order*', '*total_sales*', '*monetary_segementation*', dan '*first_segmentation*'. Data dibagi menjadi dua yaitu 80% untuk *data training* dan 20% untuk *data testing*. Evaluasi model dilakukan menggunakan *confusion matrix* untuk mengetahui nilai dari akurasi, *precision*, *recall*, dan *f-1 score*. Selain itu, validasi model juga diperlukan dengan menggunakan metode ROC AUC untuk mengukur seberapa baik model dalam memberikan prediksi klasifikasi *churn*.

Hasil penelitian ini menunjukkan bahwa penerapan *hyperparameter tuning* pada model *logistic regression* memberikan dampak dalam peningkatan performa model prediksi walaupun tidak signifikan. Pemilihan parameter yang tepat, seperti '*C*', '*class_weight*', '*dual*', '*fit_intercept*', '*intercept_scaling*', '*l1_ratio*', '*max_iter*', '*multi_class*', '*n_jobs*', '*penalty*', '*random_state*', '*solver*', '*tol*', '*verbose*', dan '*warm_start*', berkontribusi pada peningkatan performa model prediksi *churn*. Peningkatan nilai akurasi dari 80,21% menjadi 81,07% sebesar 0,86%, nilai *precision* yang meningkat dari 77,74 % menjadi 79,08% sebesar 1,34%, nilai *recall* yang meningkat dari 80,21% menjadi 81,07% sebesar 0,86%, dan *f-1 score* yang meningkat dari 77,16% menjadi 78,35% sebesar 1,19% menjadi informasi penting bahwa optimasi parameter yang optimal akan menghasilkan prediksi *churn* yang akurat dan dapat diandalkan dalam rekomendasi untuk mengurangi tingkat *churn* dan menjaga retensi pelanggan.

Kata Kunci: *E-commerce*, Rekomendasi, *Churn*, *Logistic Regression*, *Hyperparameter Tuning*.

ABSTRACT

E-commerce always needs customers and without customers, e-commerce will not survive. In e-commerce, customers will buy products because of good user experience, fast and reliable delivery, product personalization and recommendations, promotions and discounts, ease of payment, loyalty and reward programs, product reviews from previous customers and recommendations based on previous purchase preferences. However, in reality, e-commerce faces several problems such as poor product reviews, limited product categories, lack of personalization, and the behavior of customers who suddenly leave e-commerce which causes customers to switch and not use the e-commerce. One of the popular methods in overcoming this problem is logistic regression, a machine learning technique that is proven effective in various binary predictions. This research aims to implement the logistic regression method in predicting customer churn and churn rate by increasing accuracy and developing recommendations for handling customer churn based on the results of data visualization that has been modeled.

In this research, the Fashion Campus dataset is used, where preprocessing is done through data cleaning, data integration, data transformation, data reduction, exploratory data analysis, and feature engineering. The variables used are divided into two, namely the dependent as the prediction result 'segmentation_churn' and the independent variables as input variables that predict 'basket_size', 'total_orders', 'total_sales', 'monetary_segmentation', and 'first_segmentation'. The data is divided into two, 80% for training data and 20% for testing data. Model evaluation is done using confusion matrix to determine the value of accuracy, precision, recall, and f-1 score. In addition, model validation is also required using the ROC AUC method to measure how well the model provides churn classification predictions.

The results of this study indicate that the application of hyperparameter tuning to the logistic regression model has an impact on improving the performance of the prediction model, although not significantly. The selection of appropriate parameters, such as 'C', 'class_weight', 'dual', 'fit_intercept', 'intercept_scaling', 'l1_ratio', 'max_iter', 'multi_class', 'n_jobs', 'penalty', 'random_state', 'solver', 'tol', 'verbose', and 'warm_start', contributed to the improvement of the churn prediction model performance. The increase in accuracy value from 80,21% to 81,07% by 0,86%, precision value that increased from 77.74% to 79.08% by 1.34%, recall value that increased from 80.21% to 81.07% by 0.86%, and f-1 score that increased from 77.16% to 78.35% by 1.19% are important information that optimal parameter optimization will produce accurate and reliable churn prediction in recommendations to reduce churn rate and maintain customer retention.

Keywords: *E-commerce, Recommendation, Churn, Logistic Regression, Hyperparameter Tuning.*