

ABSTRAK

Analisis data penjualan memainkan peran penting dalam mendukung pengambilan keputusan bisnis, terutama untuk mengoptimalkan manajemen stok dan meningkatkan efisiensi operasional. permasalahan utama yang dihadapi oleh Vapestore XYZ di Karawang adalah kesulitan dalam memprediksi jumlah penjualan produk secara akurat, sehingga sering terjadi ketidakseimbangan antara persediaan dan permintaan pasar. Hal ini dapat menyebabkan kerugian akibat kelebihan stok maupun kekurangan barang. Saat ini, estimasi kebutuhan stok masih mengandalkan intuisi dan pengalaman pribadi, tanpa dukungan analisis data yang objektif. Penelitian ini bertujuan untuk membangun model prediksi penjualan dengan mengombinasikan metode *K-Means* untuk klasterisasi produk dan Regresi Linear untuk prediksi kuantitas penjualan. Data penjualan diambil langsung dari aplikasi POS toko, kemudian melalui tahapan pembersihan, pelabelan, dan klasterisasi menjadi tiga kelompok, yaitu “Kurang Laku”, “Laku”, dan “Sangat Laku”. Prediksi penjualan dilakukan menggunakan Regresi Linear dengan memanfaatkan hasil klasterisasi dan variabel waktu sebagai input. Evaluasi performa model dilakukan menggunakan metrik *error*, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*. Berdasarkan hasil pengujian, model Regresi Linear yang dikembangkan memperoleh *MAE* sebesar 3,20, *MSE* sebesar 52,34, dan *RMSE* sebesar 7,23. Nilai-nilai *error* ini menunjukkan bahwa model mampu memberikan estimasi penjualan yang cukup mendekati data aktual, sehingga dapat diandalkan dalam perencanaan stok. Visualisasi hasil prediksi dalam bentuk tabel dan heatmap memudahkan identifikasi tren penjualan dan perbandingan performa antar produk. Temuan penelitian ini membuktikan bahwa kombinasi metode *K-Means* dan Regresi Linear efektif digunakan untuk mendukung pengambilan keputusan stok dan strategi pemasaran pada toko ritel vape. Pengembangan lebih lanjut disarankan dengan memperkaya dataset dan mengeksplorasi metode prediksi lain guna meningkatkan performa model.

Kata Kunci: Prediksi penjualan; *K-Means*; Regresi linear; *Data mining*; Evaluasi model

ABSTRACT

Sales data analysis plays an important role in supporting business decision making, especially to optimise stock management and improve operational efficiency. The main problem faced by Vapestore XYZ in Karawang is the difficulty in accurately predicting the number of product sales, so there is often an imbalance between inventory and market demand. This can cause losses due to overstocks or shortages of goods. Currently, the estimation of stock requirements still relies on intuition and personal experience, without the support of objective data analysis. This research aims to build a sales prediction model by combining the K-Means method for product clustering and Linear Regression for sales quantity prediction. Sales data is taken directly from the store POS application, then goes through the stages of cleaning, labelling, and clustering into three groups, namely 'Less Sold', "Sold", and 'Very Sold'. Sales prediction is performed using Linear Regression by utilising the clustering results and time variables as inputs. Model performance evaluation is performed using error metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Based on the test results, the developed Linear Regression model obtained MAE of 3.20, MSE of 52.34, and RMSE of 7.23. These error values indicate that the model is able to provide sales estimates that are close enough to the actual data to be reliable in stock planning. Visualisation of the prediction results in the form of tables and heatmaps makes it easy to identify sales trends and compare performance between products. The findings of this study prove that the combination of K-Means and Linear Regression methods is effectively used to support stock decision making and marketing strategies in vape retail stores. Further development is recommended by enriching the dataset and exploring other prediction methods to improve model performance.

Keyword: Sales prediction; K-Means; Linear Regression; Data mining; Model evaluation.