

ABSTRAK

Nama : Muhammad Thufail
NIM : 41519010157
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Perbandingan Algoritma K-Nearest Neighbor, C4.5 dan Regresi Linier untuk Klasifikasi Prediksi Produk-Produk Terlaris Di Toko Cahaya Batik
Pemimbing : Dr. Afiyati, S.Si, MT

Batik, sebagai warisan budaya dunia UNESCO, memiliki beragam variasi kerajinan di seluruh Indonesia. Menjaga dan melestarikan batik memiliki pentingnya bagi warga Indonesia. Salah satu contoh upaya promosi adalah dengan mendirikan Toko Cahaya Batik di Pasar Tanah Abang, Jakarta, yang menawarkan berbagai produk batik seperti daster, hijab, gamis, setelan baju, dan mukena. Dalam penelitian ini, dilakukan perbandingan klasifikasi dan prediksi penjualan historis menggunakan algoritma KNN, C4.5, dan Regresi Linier. Dataset yang digunakan diperoleh dari data penjualan historis Toko Cahaya Batik dalam rentang waktu Februari 2022 hingga Februari 2023. Dataset tersebut kemudian melewati tahap preprocessing, termasuk pembersihan data dan penyesuaian format data untuk pemrosesan menggunakan model yang sesuai. Selain itu, data juga diberi label "Laris" dan "Tidak Laris" berdasarkan Nama Barang dan Jumlah Pembelian (Qty). Pengujian dilakukan pada model K-Nearest Neighbor dan C4.5 Decision Tree dengan menggunakan *Confusion Matrix*. Hasilnya menunjukkan bahwa algoritma K-Nearest Neighbor mencapai akurasi klasifikasi sebesar 86%, sementara C4.5 Decision Tree mencapai akurasi 71%. Untuk prediksi menggunakan model Regresi Linier, evaluasi dilakukan dengan menggunakan MAE, MSE, dan RMSE. Hasil pengujian menunjukkan nilai MAE sebesar 491.0422, MSE sebesar 390208.4036, dan RMSE sebesar 624.6666. Dari prediksi regresi linier menggunakan persamaan regresi, diharapkan terjadi penjualan sekitar 3545.0153 total produk pada bulan Desember 2023.

Kata Kunci: Batik, Klasifikasi, Prediksi, K-Nearest Neighbor, C4.5, Decision Tree, Regresi Linier, Confusion Matrix

ABSTRACT

Name : Muhammad Thufail
NIM : 41519010157
Study Program : Informatic Engineering
Title Thesis : Comparison of K-Nearest Neighbor, C4.5, and Linear Regression Algorithms for Classification Prediction of Bestselling Products at Cahaya Batik Store.
Counsellor : Dr. Afiyati, S.Si, MT

As a UNESCO World Cultural Heritage, Batik showcases various craft variations throughout Indonesia. Preserving and conserving batik holds great importance for the Indonesian people. One example of promotional efforts is the establishment of Toko Cahaya Batik in Tanah Abang Market, Jakarta, which offers a wide range of batik products such as dresses, hijabs, robes, clothing sets, and prayer garments. This research focuses on comparing the classification and prediction of historical sales using the KNN, C4.5, and Linear Regression algorithms. The dataset used is obtained from the historical sales data of Toko Cahaya Batik, covering the period from February 2022 to February 2023. The dataset goes through preprocessing stages, including data cleaning and formatting adjustments to ensure compatibility with the chosen models for data processing. Additionally, the data is labeled as "Popular" and "Not Popular" based on the Product Name and Quantity Purchased. The testing is performed on the K-Nearest Neighbor and C4.5 Decision Tree models using the Confusion Matrix. The results show that the K-Nearest Neighbor algorithm achieves a classification accuracy of 86%, while the C4.5 Decision Tree algorithm achieves 71% accuracy. For the Linear Regression model prediction, the evaluation is done using MAE, MSE, and RMSE. The testing results yield MAE of 491.0422, MSE of 390208.4036, and RMSE of 624.6666. Based on the linear regression prediction equation, approximately 3545.0153 total products are expected to be sold in December 2023.

Keywords: Batik, Classification, Prediction, K-Nearest Neighbor, C4.5, Decision Tree, Linear Regression, Confusion Matrix.