

ABSTRAK

Pertumbuhan bisnis e-commerce di Indonesia, melalui berbagai situs online seperti Tokopedia, Shopee, Beli-beli memiliki kecenderungan untuk meningkat setiap tahunnya. Pada perkembangannya muncul fenomena dimana ada lapak yang mampu menjual lebih sukses dibandingkan lapak lainnya. Sehingga penelitian ini berfokus untuk mengembangkan model prediksi penjualan produk jenis tertentu dan mengambil interpretasi dari model yang telah dibuat. Hasil Interpretasi tersebut dapat digunakan untuk merancang strategi penjualan untuk meningkatkan kesuksesan lapak.

Model prediksi ini dibangun dengan pendekatan klasifikasi metode Pohon Keputusan. Kategori akan dibagi menjadi dua, yaitu berhasil terjual, dan tidak terjual. Kemudian model prediksi dikembangkan berdasarkan variabel-variabel informasi yang terpublikasi pada situs ecommerce. Tahap pertama adalah mengembangkan 3 model prediksi menggunakan parameter yang sama untuk mengetahui apakah kesimpulan yang didapat akan berbeda terhadap masing-masing produk. Kemudian Tahap kedua, mengembangkan model prediksi masing-masing data menggunakan *k-fold cross-validation* dengan tujuan mendapatkan hasil yang *robust*. Dan tahap ketiga mengambil kesimpulan berdasarkan masing-masing pohon keputusan yang telah dikembangkan. Pengembangan model ini menggunakan 1739 data untuk 3d printer, 612 data untuk printer laser, dan 2722 data untuk desktop.

Model awal 3d printer memberikan *F1-score* sebesar 86.9%, model awal printer laser memberikan *F1-score* sebesar 65.99%, dan model awal desktop memberikan *F1-score* sebesar 66.09%. Ketiga model awal tersebut memiliki kesamaan yaitu variabel 'dilihat' memiliki pengaruh yang besar dalam penentuan kesuksesan penjualan, akan tetapi hanya data 3d printer yang mampu memberikan hasil yang serupa apabila variabel 'dilihat' tidak digunakan dalam pengembangan model. Hasil penelitian menunjukkan bahwa variabel kesuksesan penjualan ketiga data ini memiliki hasil yang berbeda-beda dan model prediksi masing-masing mampu memprediksi dengan nilai *F1-score* 92.02% untuk 3d printer, nilai *F1-score* 42,55% untuk printer laser, dan nilai *F1-score* 65.99% untuk desktop. Model prediksi yang memiliki nilai *F1-score* lebih dari 50% dapat digunakan sebagai pedoman dalam perencanaan penjualan produk-produk tersebut.

Kata Kunci : Data Mining, Machine Learning, Pohon Keputusan, Klasifikasi, Kesuksesan Penjualan, Ecommerce

ABSTRACT

Ecommerce business growth in Indonesia based on site such as Tokopedia, Shopee, Beli-beli have tendency to increase each year. Ecommerce rapid development arise a fenomena where there are shops that are capable of selling more successfully than other shops. This research purpose is focusing on building a model that is capable of predicting whether a product has a chance that it will be sold. From those models, we will make an interpretation that can be use to make salse strategy to increasing the success of a shop.

The prediction model is built using Decision Tree method. The success category is divided into two, sold, and not sold. Prediction model is developed using variable information that appears in ecommerce sites. First step in research is to built 3 prediction models using the same parameter, which will be used to find out whether model prediction will produce different or the same result. Second step is developing the model using k-fold cross validation to obtain robust result. Third step is to take a conclusion based on model prediction in second step. This prediction model is using 1739 data point for 3d printer, 612 data for printer laser, and 2722 data for desktop.

Early model predictions give f1-score 86.9% for 3d printer, 65.99% f1-score for printer laser, and 66.09% for desktop. Those three models have similarity where variabel ‘dilihat’ had more influence to the success of product sold, however only 3d printer data that have similar result when model is built without variabel ‘dilihat’. This research result shows those three data have different prediction models where data 3d printer f1-score is 92.02%, printer laser f1-score is 42.55%, and desktop f1-score is 65.99%. Those prediction models that have f1-score more than 50% can be used as a sales strategy for each product.

Keywords : Data Mining, Machine Learning, Decision Tree, Classification, Sales Success, Ecommerce.