

Perbandingan Algoritma K-Nearest Neighbor, Random Forest, dan Logistic Regression Terhadap Kategori Review pada Marketplace XYZ

Aria Nanda Herdiawan

ABSTRAK

Dengan perkembangan teknologi di dunia yang sangat cepat, dapat mempermudah dalam melakukan segala hal. Salah satunya adalah berbelanja secara *online*. Salah satu platform yang dapat digunakan untuk berbelanja *online* adalah *marketplace*. Banyak *review* yang diberikan pembeli ketika selesai melakukan transaksi pada *marketplace*. Sehingga dibutuhkan klasifikasi untuk membedakan kategori *review* pembeli, selain itu dapat diketahui perbedaan kinerja dari algoritma klasifikasi *K-nearest neighbor*, *random forest*, dan *logistic regression*. Dengan menggunakan data *review* sebanyak 1000 data yang dibagi menjadi kelas produk 618 data, pengiriman 214 data, dan pelayanan 168 data. Data dilakukan *preprocessing* dengan tahap *case folding*, *cleansing*, normalisasi, tokenisasi, *filtering*, dan *stemming*. Kemudian dihitung bobot kata pada data dengan *tf-idf*. Data akan dibagi menjadi *data train* dan *data test*. Setelah itu dilakukan klasifikasi dengan *K-nearest neighbor*, *random forest*, dan *logistic regression*. Yang menghasilkan akurasi tertinggi dari *K-nearest neighbor* 76%, *random forest* 81%, dan *logistic regression* 80%.

Kata Kunci: *Review, Marketplace, Klasifikasi Teks, K-Nearest Neighbor, Random Forest, Logistic Regression*

ABSTRACT

With the rapid development of technology in the world, it can make doing everything easier. One of them is shopping online. One platform that can be used for online shopping is a marketplace. Many reviews are given by buyers when completing transactions in marketplace. So classification is needed to distinguish categories of buyer reviews, besides that we can know the difference performance of the K-nearest neighbor, random forest, and logistic regression classification algorithms. By using review data of 1000 data which is divided into product classes 618 data, delivery 214 data, and service 168 data. The data was preprocessed using case folding, cleansing, normalization, tokenization, filtering, and stemming stages. Then the word weights in the data are calculated using tf-idf. The data will be divided into train data and test data. After that, classification was carried out with K-nearest neighbor, random forest, and logistic regression. Which produces the highest accuracy of K-nearest neighbor 76%, random forest 81%, and logistic regression 80%.

Key Word: Review, Marketplace, Text Classification, K-Nearest Neighbor, Random Forest, Logistic Regression