

## Sistem Rekomendasi Pada Tokopedia Menggunakan Algoritma K-Nearest Neighbor

Rubangi<sup>1</sup>, Rianto<sup>2</sup>

<sup>1,2</sup> Magister Teknologi Informasi, Program Pascasarjana, Universitas Teknologi Yogyakarta

<sup>1</sup>e-mail: rubangi815@gmail.com

<sup>2</sup>e-mail: rianto@staff.uty.ac.id

Diterima	Direvisi	Disetujui
05-12-2021	30-12-2021	14-01-2022

**Abstrak** - Tokopedia merupakan salah satu perusahaan perdagangan elektronik yang memiliki data pertumbuhan yang sangat pesat dengan seiring berjalannya waktu. Adanya data terus menerus bertambah besar sehingga dapat terjadinya masalah bagi user. User sering mengalami kendala dalam promosi produk yang sering dikunjungi pembeli. Selain itu pembeli sering mengalami kesulitan untuk menemukan kebutuhan produk terbaik yang sesuai kebutuhan pembeli. Dengan adanya masalah tersebut yang terjadi maka dapat diatasi dengan adanya sistem rekomendasi produk tertentu untuk ditawarkan kepada pembeli. Sistem rekomendasi penelitian ini mengimplementasikan dengan Algoritma K-Nearest Neighbor pada Rating Product Reviews Tokopedia. Algoritma K-Nearest Neighbor yang digunakan untuk menentukan top-n rekomendasi produk tertentu untuk ditawarkan kepada pembeli. Hasil dari penelitian yang dilakukan pada data 2040 rating produk dengan menggunakan algoritma K-Nearest Neighbors yaitu nilai Accuracy sebesar 73.53%, Precision sebesar 73.64%, Recall sebesar 99.62%.

Kata Kunci : Sistem Rekomendasi, K-Nearest Neighbors, Tokopedia

*Abstract* - Tokopedia is one of the electronic trading companies that has very rapid growth data over time. The existence of data continues to grow up so that there can be problems for the user. Users often experience problems in promoting products that are frequently visited by buyers. In addition, buyers often have difficulty finding the best product needs that match the needs of buyers. These problems that occur, can be overcome by having a specific product recommendation system to be offered to buyers. This research recommendation system implements the K-Nearest Neighbor Algorithm on Tokopedia's Product Reviews Rating. The K-Nearest Neighbor algorithm is used to determine top-n recommendations for certain products to be offered to buyers. The results of research conducted on 2040 product rating data using the K-Nearest Neighbors algorithm are the Accuracy value of 73.53%, Precision of 73.64%, Recall of 99.62%.

Keywords: Recommendation System, K-Nearest Neighbors, Tokopedia

### PENDAHULUAN

Semakin berkembangnya teknologi informasi maka suatu produk penjualan di internet akan semakin bertambah dan berkembang. Pembeli sering mengalami kesulitan pada saat mencari produk yang dibutuhkan sesuai keinginan pembeli karena banyaknya produk pada Tokopedia. Pengguna juga dapat terjadi munculnya kesulitan adanya data transaksi, data produk dan data pembeli yang sangat banyak sehingga dapat timbulnya kesulitan untuk menawarkan dan mempromosikan penjualan produk yang tepat kepada pembeli. Pengguna pada Tokopedia mempunyai jumlah data sangat banyak

sehingga akan dapat pertambahan data setiap saat yang sangat banyak. Adanya pertambahan data tersebut yang sangat besar maka sering terjadi masalah pada sistem Tokopedia. Data penjualan online di Tokopedia akan menghasilkan informasi yang dapat dibutuhkan sesuai dengan kebutuhan pengguna.

Adanya masalah tersebut yang terjadi maka dapat diatasinya dengan adanya sistem rekomendasi produk tertentu untuk dipromosikan dan ditawarkan serta memberikan rekomendasi kepada pembeli Tokopedia. Dalam kemudahan pada Sistem Rekomendasi Tokopedia sehingga diharapkan dapat



membantu pengguna dalam mengerjakan pembelian produk pada Tokopedia secara online. Sistem Rekomendasi ialah sebuah model aplikasi dari hasil observasi keadaan keinginan pelanggan sehingga membutuhkan sebuah model yang tepat sesuai keinginan untuk mempermudah untuk mengambil keputusan saat menentukan produk tertentu (Kadyanan, 2017). Pada Sistem Rekomendasi yang ditawarkan akan memberikan item-item berupa produk yang tersedia untuk para pembeli serta membantu pengguna untuk dapat mengidentifikasi beberapa kebutuhan produk sesuai keinginan pembeli. Sistem rekomendasi yang dikembangkan penelitian menggunakan algoritma K-Nearest Neighbor untuk mengklasifikasi rating product reviews Tokopedia. Algoritma K-Nearest Neighbor merupakan sebuah metode algoritma yang hasil dari sampel uji baru untuk diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN yang bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan sampel latih (Hidayanti, 2020). Penggunaan metode algoritma K-Nearest Neighbor ini digunakan merupakan suatu model paling simpel untuk pendukung keputusan dapat mengklasifikasikan berdasarkan data yang jarak terdekat. Selain itu metode K-Nearest Neighbor juga rumus yang paling sederhana untuk dipakai dalam pencarian jarak. Pada pengujian hasil performa dari penerapan algoritma K-NN ini untuk sistem rekomendasi dapat mengukur keberhasilan algoritma dalam memberikan rekomendasi yang tepat kepada pengguna.

Dari beberapa hasil penelitian sebelumnya (Prasetya, 2017) yang berjudul sistem rekomendasi E-Commerce dengan menggunakan metode algoritma K-Nearest Neighbor, pada penelitian ini membahas mengenai pebisnis online E-commerce dengan salah satu metode klasifikasi yang dapat untuk pengambilan keputusan para pembeli yang timbul adanya kesulitan untuk mencari produk tertentu di internet. Dari berbagai metode seperti Collaborative Filtering, Content based, dan Hybrid yang digunakan untuk klasifikasi. Pada sistem rekomendasi dengan menggunakan data rating pembeli untuk menentukan top-n pada rating produk setiap pembelian produk. Hasil penelitian dari sistem rekomendasi pada E-Commerce ini metode Content Based digunakan mengungguli lebih baik dari pada metode lainnya karena data yang digunakan sparse, yaitu kondisi dimana jumlah rating yang diberikan relatif sedikit terhadap banyaknya produk yang tersedia pada E-commerce tersebut.

Penelitian (Fitrianti et al., 2019) terhadap algoritma K-Nearest neighbor untuk sentimen review Restoran teks bahasa Indonesia. Dengan penerapan algoritma K-Nearest Neighbor untuk analisis sentimen membantu restoran memiliki kecenderungan sentimen positif. Penelitian ini data digunakan 205 dataset kemudian dipre-processing menjadi 199 dataset review pelanggan restoran

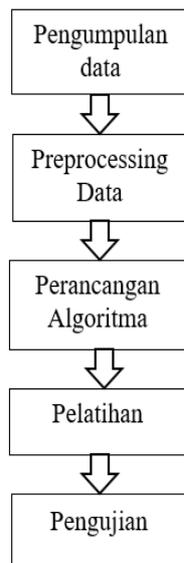
dikarenakan data tersebut terdapat kata-kata yang tidak sesuai atau noise. Hasil pengujian dari analisis sentimen review restoran dengan bahasa Indonesia dengan algoritma K-Nearest neighbor diperoleh hasil akurasi confusion matrix sebesar 96,61%.

Penelitian lainnya (Aisha Alfani W. P. R., Fahrur Rozi, 2021) menggunakan metode algoritma K-Nearest Neighbor untuk prediksi penjualan produk. Data penjualan produk dipakai adalah Unilever, Nestle, dan P&G. Data dilakukan pengujian menggunakan metode algoritma K-NN dengan 60 data latih dan 20 data uji untuk prediksi penjualan produk berguna mengetahui produk terlaris yang terjual sehingga mempermudah dalam penyediaan stok barang yang akan datang. Algoritma K-Nearest Neighbor digunakan untuk klasifikasi data penjualan produk dengan hasil didapatkan nilai akurasi tinggi klasifikasi penjualan produk yaitu sebesar 86,66% dan klasifikasi akurasi rendah penjualan produk sebesar 40%.

Pada penelitian ini mengimplementasi pada sistem rekomendasi pada Tokopedia menggunakan algoritma K-Nearest Neighbors. Data yang digunakan berjumlah 2040 data rating produk hasil penilaian produk yang dibeli pelanggan. Hasil performa algoritma K-Nearest Neighbors dapat diukur berdasarkan nilai akurasi yang diperoleh.

## METODOLOGI PENELITIAN

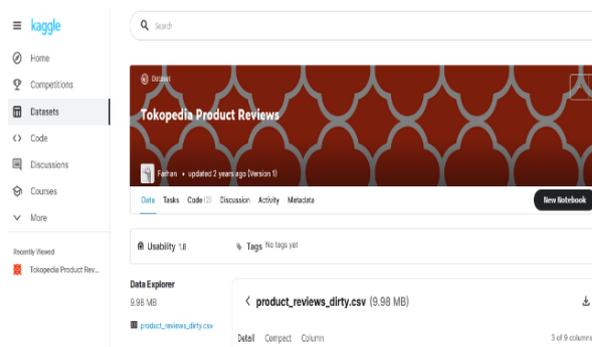
Penelitian ini dilakukan dimulai mengumpulkan data. Setelah data dikumpulkan maka akan dilakukan beberapa proses *pre-processing* data. Kemudian proses mengelola data dengan tool Rapidminer menggunakan algoritma K-Nearest Neighbors untuk mendapatkan hasil klasifikasi nilai akurasi. Pada metode penelitian ini yaitu: (1) pengumpulan data; (2) preprocessing data; (3) perancangan algoritma; (4) pelatihan; (5) pengujian. Tahapan-tahapan pada metode penelitian dapat dilihat pada gambar 1 dibawah ini.



Sumber: Tahapan langkah penelitian (2021)  
Gambar 1. Tahapan Penelitian

### 1. Pengumpulan data

Penelitian ini pengumpulan data digunakan yaitu dari Tokopedia produk Reviews. Dataset yang digunakan penelitian diambil dari Kaggle.com. Data yang digunakan berjumlah sebesar 2040 data rating review produk. Sumber dataset yang digunakan dilihat pada gambar 2.



Sumber data: Kaggle.com (2021)  
Gambar 2 Sumber Data

### 2. Preprocessing data

Langkah awal sebelum menggunakan data dengan metode atau teknik data mining maka harus dilakukan beberapa langkah yaitu preprocessing terhadap data. Tahap preprocessing yaitu untuk mencari data yang memiliki missing value yang berarti data yang didapat belum normal dan tahapan ini membuat label/*Class*, merubah atribut string menjadi numerik (Maulidah et al., 2020). Pada tahapan preprocessing data dengan cara seleksi beberapa atribut dari dataset kemudian atribut akan diberikan target role dari atribut tersebut. Setelah tahap preprocessing maka selanjutnya proses

klasifikasi dengan metode K-NN dengan nilai parameter ditentukan  $k=3$ ,  $k=5$ ,  $k=7$ ,  $k=9$ ,  $k=11$ . Hasil didapatkan untuk mengetahui performa klasifikasi, pengukuran, yaitu dengan akurasi, presisi dan *recall* (Deolika et al., 2019).

### 3. Perancangan Algoritma

salah satu algoritma yang sering digunakan untuk melakukan klasifikasi data adalah Algoritma K-Nearest Neighbors. Hal ini tujuan dari algoritma K-Nearest Neighbors adalah melakukan mengklasifikasikan obyek berdasarkan atribut dan training sample (Ernawati & Wati, 2018). Algoritma K-Nearest Neighbors merupakan salah satu metode klasifikasi terhadap objek berdasarkan data jaraknya paling dekat dengan objek tersebut (Kafil, 2019). pada algoritma K-Nearest Neighbor melakukan klasifikasi mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan. Algoritma *K-Nearest Neighbor* ini yang digunakan penelitian data dibagi 2 bagian yaitu pelatihan data dan uji data untuk berdasarkan untuk prediksi. (Syah, 2020) ada beberapa langkah dari algoritma K-Nearest Neighbor yaitu:

- Menentukan parameter  $k$  (jumlah tetangga terdekat).
- Perhitungan  $k$ -data *training* yang jaraknya paling dekat dengan data *testing*.
- Urutkan parameter jarak berdasarkan nilai terkecil sejumlah  $k$ .
- Penetapan kelompok data *testing* berdasarkan label mayoritas  $k$ .

### 4. Pelatihan

Pada penelitian ini data pelatihan pada dataset dilakukan untuk melatih algoritma K-Nearest Neighbors. Pelatihan dataset digunakan dari data Tokopedia product reviews yang sudah diolah dan dipreprocessing dengan menggunakan algoritma K-Nearest Neighbors sebelumnya. Data yang akan pakai untuk pelatihan adalah data dibagi menjadi dua yaitu, sebagai *training* dataset dengan persentase 70% dan *testing* dataset 30%. Pada tahapan pelatihan data yang digunakan sebanyak 1428 dari dataset atau 70% dari dataset.

### 5. Pengujian

Pengujian dalam penelitian ini menggunakan algoritma K-Nearest Neighbors. Pengujian dataset dilakukan untuk mengetahui performa dari algoritma K-NN yaitu *accuracy*, *precision*, dan *recall* dari metode algoritma yang dipakai. Tahap pengujian ini menjelaskan bagaimana menggunakan algoritma K-Nearest Neighbors dalam menentukan nilai akurasi. Performa dari metode K-NN yang didapatkan akan

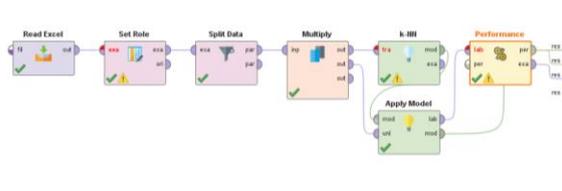
membuktikan keakuratan model algoritma. Pada pengujian dataset sebanyak 30% atau 612 dataset.

## HASIL DAN PEMBAHASAN

Pada penelitian ini dataa dipakai atau digunakan untuk pengklasifikasian sistem rekomendasi pada Tokopedia sebanyak 2040 data pengguna. Tool Aplikasi yang dipakai mengelola data dan membuat model yang telah dirancang sebelumnya yaitu dengan RapidMiner versi 9.0.

### 1. Model model K-Nearest Neighbors

Dalam penelitian ini untuk sistem rekomendasi pada Tokopedia dengan rancangan model K-Nearest Neighbors dilakukan menggunakan tool RapidMiner dengan melakukan beberapa operator yaitu Read Excel, set role, split data, multiply, algoritma K-NN, apply model dan performance. model yang dirancang terdapat pada gambar 3.



Sumber: Model K-Nearest Neighbors (2021)

Gambar 3. Model Algoritma K-Nearest Neighbors

Pada gambar 3 diatas beberapa tahapan yang dilakukan untuk klassifikasi rating product reviews dengan metode K-Nearest Neighbor sebagai berikut:

- Operator Read Excel merupakan dataset rating product reviews Tokopedia yang diinputkan untuk dianalisis.
- Operator Set Role digunakan untuk dataset rating product reviews mengubah atribut beberapa atribut yang akan diubah menjadi atribut keterangan menjadi label.
- Operator Slipt Data dipakai untuk membagi data menjadi dengan ratio 0.7 dan 0.3.
- Operator Multiply dipakai untuk membagi data yang sudah diberi atribut ke K-NN dan Applay Model.
- Pelatihan dataset rating product reviews menggunakan operator K-NN.
- Untuk penerapan model menggunakan operator Apply Model.
- Pengujian dataset rating product reviews dengan model menggunakan operator performance untuk mengetahui *accuracy*, *precision* dan *recall* dari model K-NN.

### 2. Performa Algoritma K-Nearest Neighbors

Berdasarkan hasil nilai klasifikasi rating product reviews Tokopedia model algoritma K-Nearest

Neighbors yang dirancang dengan pengujian data berjumlah 2040 data. Pengujian data sebagai *training* dataset berjumlah 70 % dan 30 % *testing* dataset. Hasil dari klasifikasi data testing diperoleh tingkat keakuratan pengujian metode K-Nearest Neighbors yaitu dengan nilai parameter K=3, K=7, K=9, K=11 seperti terlihat tabel 1 dibawah ini.

Tabel 1. Pengujian Metode K-Nearest Neighbors

No	K	Accuracy	Precision	Recall
1	3	66.18%	73.56%	85.66%
2	5	71.64%	73.11%	96.94%
3	7	72.69%	74.38%	96.85%
4	9	73.53%	73.64%	99.62%
5	11	73.32%	74.82%	97.13%

Sumber: Pengujian K-NN (2021)

Berdasarkan tabel 1. Hasil pengujian pada masing-masing parameter K1 s/d K11 diperoleh nilai *accuracy* tertinggi yaitu pada parameter k=9 nilai *accuracy* 73,53 %. Nilai *accuracy* terendah pada pengujian parameter k=3 yaitu sebesar 66,18%. *Precision* nilai tertinggi dihasilkan pada pengujian menggunakan parameter k=11 sebesar 74,82%. *Precision* nilai terendah pada pengujian parameter k=5 sebesar 73,11%. Nilai *recall* tertinggi dihasilkan pada pengujian menggunakan parameter k=9 sebesar 99,62 %. Nilai *recall* terendah pada pengujian menggunakan parameter k=3 sebesar 85,66%. Sehingga didapatkan klasifikasi rating produk review dengan hasil performa algoritma K-Nearest Neighbors yang dapat dilihat tabel 2.

Tabel 2. Performa K-Nearest Neighbors

No	Jenis performa	Nilai
1	Accuracy	73.53%
2	Precision	73.64%
3	Recall	99.62%

Sumber: Hasil Performa K-NN (2021)

Berdasarkan Tabel 2. Hasil *accuracy* pengujian performa algoritma K-Nearest Neighbors pada klasifikasi rating produk review Tokopedia diperoleh hasil nilai tertinggi yaitu parameter K=9 dengan nilai *accuracy* sebesar 73,53%, *precision* sebesar 73.64%, *recall* sebesar 99.62%.

## KESIMPULAN

Pada penelitian ini disajikan sistem rekomendasi pada Tokopedia menggunakan algoritma K-Nearest Neighbors. Hasil pengujian akurasi performa dari model algoritma K-Nearest Neighbors yang dirancang sebelumnya dengan hasil *accuracy* sebesar 73,53%, *precision* sebesar 73.64% dan *recall* sebesar 99.62%. Sistem rekomendasi pada Tokopedia menggunakan metode algoritma K-Nearest Neighbors untuk mengklasifikasikan rating product reviews Tokopedia yang dapat diterapkan dan implementasi pada pembelian produk untuk

direkomendasi produk tertentu kepada pembeli, sehingga pada penelitian selanjutnya dapat dikembangkan lebih lanjut dengan menggabungkan beberapa metode algoritma untuk sistem rekomendasi sehingga menghasilkan rekomendasi penjualan produk lebih baik.

## REFERENSI

- Aisha Alfani W. P. R., Fahrur Rozi, F. S. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(1), 155–160.
- Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informatika*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, VI(1), 64–69. <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626>.
- Fitrianti, R. P., Kurniawati, A., & Agusten, D. (2019). Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informatika (SNATi) 2019*, 27–32.
- Hidayanti, W. P. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape ( Rokok El ektrik ) pada “ Lombok Vape On ”. 3(2).
- Kadyanan, I. G. A. G. A. (2017). Perancangan Sistem Rekomendasi dalam Industri Kuliner di Bali. *Jurnal Ilmiah*, X(1), 1–6. <https://docplayer.info/96219900-Perancangan-sistem-rekomendasi-dalam-industri-kuliner-di-bali.html>
- Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 3(2), 59–66. <https://doi.org/10.36040/jati.v3i2.860>
- Maulidah, M., Gata, W., Aulianita, R., & Agustyaningrum, C. I. (2020). Algoritma Klasifikasi Decision Tree untuk Rekomendasi Buku Berdasarkan Kategori Buku. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 89–96.
- Prasetya, C. S. D. (2017). Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 4(3), 194. <https://doi.org/10.25126/jtiik.201743392>
- Syah, R. D. (2020). Performa Algoritma User K-Nearest Neighbors pada Sistem Rekomendasi di Tokopedia. *Jurnal Informatika Universitas Pamulang*, 5(3), 302. <https://doi.org/10.32493/informatika.v5i3.6312>