

Performa Algoritma *User K-Nearest Neighbors* pada Sistem Rekomendasi di Tokopedia

Rama Dian Syah

Fakultas Teknologi Informasi, Universitas Gunadarma, Jl. Margonda Raya 100, Depok, Jawa Barat, Indonesia, 16424
e-mail: rama_ds@staff.gunadarma.ac.id

Submitted Date: July 24th, 2020
Revised Date: September 22nd, 2020

Reviewed Date: September 22nd, 2020
Accepted Date: September 30th, 2020

Abstract

The biggest marketplace in Indonesia such as Tokopedia has data on e-commerce activities that always increase with time. Large data growth in Marketplace can cause problems for users. Buyers who have difficulty in finding the best product that suits their needs and sellers who have difficulty in promoting products that are often visited by buyers can be overcome. The recommendation system can overcome these problems by providing specific product recommendations to be promoted and offered to buyers. This research implements the Recommendation System using the Item Rating Prediction Method by applying the User K-Nearest Neighbors Algorithm. The Recommendation System provides recommendations based on ratings on products given by the buyer. Algorithm performance in Recommendation System is measured by the parameters of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE). The performance values obtained are RMSE = 0.713, MAE = 0.488 and NMAE = 0.122. Performance values below 1 proves that the User K-Nearest Neighbors Algorithm is suitable as a rating prediction model on recommendation system.

Keywords: Recommendation System; User K-Nearest Neighbors; Tokopedia

Abstrak

Pasar online terbesar di Indonesia seperti Tokopedia memiliki data aktifitas perdagangan secara elektronik yang selalu bertambah seiring waktu. Pertumbuhan data yang besar pada pasar online dapat menimbulkan masalah bagi pengguna. Pembeli yang mengalami kesulitan pada pencarian produk terbaik yang sesuai dengan kebutuhannya dan penjual yang kesulitan pada promosi produk yang sering dikunjungi pembeli dapat diatasi. Sistem rekomendasi dapat mengatasi masalah tersebut dengan memberikan rekomendasi produk tertentu untuk dipromosikan dan ditawarkan kepada pembeli. Penelitian ini mengimplementasikan Sistem Rekomendasi menggunakan Metode *Item Rating Prediction* dengan menerapkan Algoritma *User K-Nearest Neighbors*. Sistem Rekomendasi memberikan rekomendasi berdasarkan penilaian pada produk yang diberikan oleh pembeli. Performa algoritma pada Sistem Rekomendasi diukur dengan parameter *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE) dan *Normalized Mean Absolute Error* (NMAE). Nilai performa yang didapat yaitu RMSE = 0.713, MAE = 0.488 dan NMAE = 0.122. Nilai performa dibawah 1 membuktikan bahwa algoritma *User K-Nearest Neighbors* cocok sebagai model prediksi rating pada sistem rekomendasi.

Kata Kunci: Sistem Rekomendasi; *User K-Nearest Neighbors*; Tokopedia

1. Pendahuluan

Data merupakan elemen penting dalam perkembangan teknologi saat ini. Data dapat menghasilkan informasi yang dibutuhkan untuk perkembangan suatu sistem agar sesuai dengan kebutuhan pengguna. Pertambahan data yang

sangat besar terjadi pada sistem online *Marketplace*.

Kemudahan transaksi yang ditawarkan oleh *Marketplace* sangat membantu pengguna dalam melakukan aktifitas jual beli secara online. *Marketplace* terbesar di Indonesia seperti

Tokopedia memiliki jumlah pengguna yang sangat banyak sehingga menimbulkan pertambahan data yang sangat cepat.

Masalah yang dapat ditimbulkan seperti pembeli yang mengalami kesulitan pada pencarian produk yang sesuai dengan kebutuhannya atau penjual yang ingin mempromosikan produknya yang sering dikunjungi pembeli dapat diatasi.

Sistem rekomendasi dapat mengatasi masalah pertumbuhan data dengan memberikan rekomendasi kepada pengguna *Marketplace*. Sistem rekomendasi merupakan sistem yang digunakan untuk mempelajari informasi dari data masa lalu dan memberikan saran (Choudhary & Tulasi, 2019). Sistem Rekomendasi merupakan salah satu implementasi dari *Machine Learning*.

Sistem Rekomendasi memiliki beberapa metode dan algoritma yang dapat diterapkan. Pengujian performa dari penerapan algoritma pada Sistem Rekomendasi dapat mengukur keberhasilan algoritma dalam memberikan rekomendasi kepada pengguna.

Beberapa penelitian telah dilakukan terhadap Sistem Rekomendasi pada *E-Commerce*. Kai Wang melakukan penelitian dengan mengusulkan Sistem Rekomendasi pada *E-Commerce* menggunakan algoritma RNN (K.

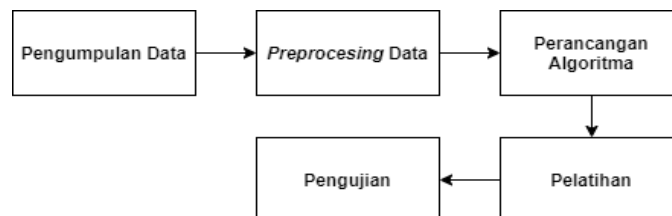
Wang et al., 2019). Hasil penelitiannya yaitu performa dari algoritma diukur berdasarkan parameter *Mean Absolute Error* (MAE).

Penelitian yang lainnya dilakukan oleh Arie Satia Dharma tentang Sistem Rekomendasi menggunakan algoritma KNN berdasarkan personalisasi pengguna (Dharma & Samosir, 2019). Hasil penelitiannya adalah performa dari algoritma KNN yang diukur berdasarkan parameter *Root Mean Square Error* (RMSE).

Pada penelitian ini dilakukan pengujian performa Algoritma *User K-Nearest Neighbors* yang diterapkan pada Sistem Rekomendasi di Tokopedia. Data yang digunakan berjumlah 40608 pengguna Tokopedia yang memberikan rating pada produk. Pengujian performa Algoritma diukur berdasarkan parameter *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE) dan *Normalized Mean Absolute Error* (NMAE).

2. Metode Penelitian

Penelitian ini dilakukan dengan 5 tahapan penelitian yaitu: (1) Pengumpulan Data; (2) *Preprocessing* Data; (3) Perancangan Algoritma; (4) Pelatihan; (5) Pengujian. Diagram alur tahapan penelitian terlihat pada Gambar 1.

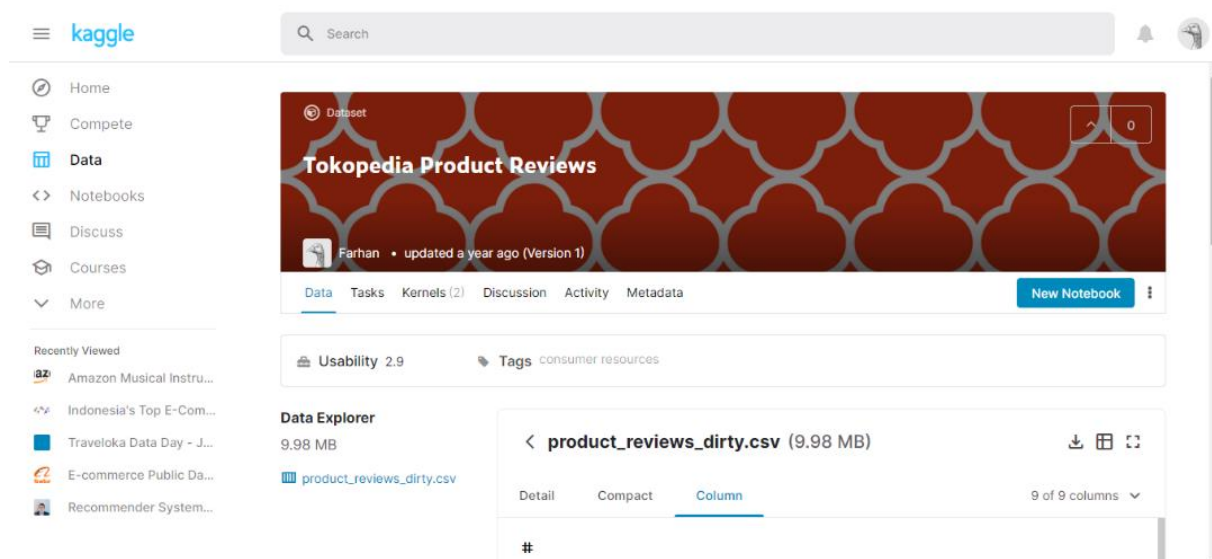


Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pengumpulan data dilakukan dengan berbagai teknik untuk bahan penelitian (Roh et al., 2019). Data yang digunakan pada penelitian ini berasal dari *Tokopedia Product Reviews*. Data

diambil dari *website* Kaggle.com. Data yang digunakan berjumlah 40608 *reviews* pengguna Tokopedia. Data berukuran 9.98 MB. Sumber data yang digunakan terlihat pada Gambar 2.



Gambar 2. Sumber Data

2.2 Preprocessing Data

Preprocessing Data dilakukan untuk transformasi data mentah menjadi data yang berkualitas serta menghilangkan duplikasi data (Syah, 2020). *Preprocessing* Data dilakukan

dengan seleksi atribut dari beberapa atribut yang ada pada dataset. Atribut yang digunakan akan diberikan *target role* sebagai tugas dari atribut tersebut. Berikut atribut yang akan digunakan terdapat pada Gambar 2 .

attribute name	target role
rating	label
product_id	item identification
shop_id	user identification

Gambar 2. Atribut yang digunakan

Keterangan Gambar 2:

- *Target Role Label* merupakan atribut yang khusus digunakan untuk target pembelajaran operator. Atribut yang menjadi target pembelajaran yaitu atribut rating.
- *Target Role Item Identification* digunakan untuk mengidentifikasi produk pada Tokopedia.
- *Target Role User Identification* digunakan untuk mengidentifikasi pengguna yang memberikan penilaian pada produk di Tokopedia.

2.3 Perancangan Algoritma

Algoritma yang digunakan dalam penelitian ini adalah *User K-Nearest Neighbors*. Algoritma ini termasuk salah satu dari metode

Collaborative Rating Prediction. Algoritma ini membagi data menjadi 2 bagian yaitu data latih dan data uji. Data latih digunakan oleh algoritma sebagai dasar prediksi. Data uji digunakan oleh algoritma sebagai penilaian hasil prediksi (Prasetya, 2017).

Tahapan Algoritma KNN yaitu:

1. Penetapan *k* (jumlah tetangga terdekat).
2. Perhitungan jarak data latih dengan data uji.
3. Pengurutan jarak tersebut berdasarkan nilai yang terkecil sejumlah *k*.
4. Penetapan kelompok data uji berdasarkan label mayoritas *k*.

2.4 Pelatihan

Pelatihan pada dataset dilakukan untuk melatih algoritma (Shafique & Hato, 2015). Pelatihan dataset *Tokopedia Product Reviews* yang sudah diolah pada saat *preprocessing* data dengan algoritma *User K-Nearest Neighbors* yang sudah dirancang. Dataset yang digunakan dibagi menjadi dua data dengan persentase 80% sebagai *training* dataset dan 20% sebagai *testing* dataset.

2.5 Pengujian

Pengujian dilakukan untuk mengetahui performa algoritma terhadap dataset yang telah diproses (Belavagi & Muniyal, 2016). Algoritma yang akan dilakukan pengujian dan evaluasi yaitu algoritma *User K-Nearest Neighbors*. Nilai performa yang didapat akan membuktikan keakurasian model yang telah dirancang. Terdapat 3 Parameter uji yang digunakan dalam pengujian yaitu (Wang & Lu, 2018):

1. *Root Mean Square Error* (RMSE)

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\hat{r}_n - r_n)^2}{N}}$$

2. *Mean Absolute Error* (MAE)

$$MAE = \frac{\sum_{n=1}^N |\hat{r}_n - r_n|}{N}$$

3. *Normalized MAE* (NMAE).

$$NMAE = \frac{\sum_{n=1}^N |\hat{r}_n - r_n|}{\sum_{n=1}^N r_n}$$

Keterangan dari rumus parameter uji:

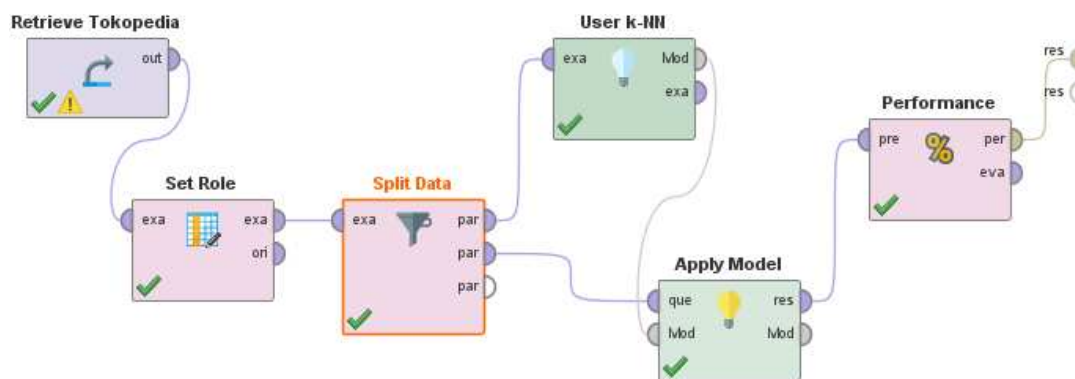
- $\hat{r}_n$ = rating prediksi
- r_n = rating pengguna pada dataset
- N = jumlah pasangan prediksi rating antara pengguna pada dataset dan hasil prediksi

3 Hasil dan Pembahasan

Dataset yang digunakan pada model sistem rekomendasi di Tokopedia berjumlah 40607 data pengguna dengan 3 atribut. *Tool* RapidMiner versi 9 digunakan untuk mengelola data menggunakan model yang dirancang.

3.1 Model Algoritma User K-Nearest Neighbors

Model yang dirancang menggunakan beberapa operator yaitu *Set Role*, *Split Data*, *User k-NN*, *Apply Model* dan *Performance*. Model yang dirancang terdapat pada Gambar 3.



Gambar 3. Model Algoritma *User K-Nearest Neighbors*

Pada Gambar 3 tahapan yang dilakukan yaitu:

1. Dataset yang diinput merupakan dataset tokopedia yang ditampung pada *Retrieve Data*.
2. Atribut dari dataset akan diberikan tugas atribut menggunakan operator *Set Role*.
3. Partisi dataset dengan rasio 0.8 dan 0.2 menggunakan operator *Split Data*.
4. Pelatihan dataset menggunakan operator *User k-NN*.
5. Penerapan model menggunakan operator *Apply Model*

6. Pengujian Model menggunakan operator *Performance*.

3.2 Performa Algoritma User K-Nearest Neighbors

Nilai performa dari model yang dirancang dengan Algoritma *User K-Nearest Neighbors* diukur dengan parameter *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE) dan *Normalized Mean Absolute Error* (NMAE). Performa dari model yang sudah dirancang terdapat pada Tabel 1.

Tabel 1. Nilai Performa Model dengan Algoritma *User k-NN*

Performa Vector	Nilai Performa
RMSE	0.713
MAE	0.488
NMAE	0.122

Parameter MAE dan RMSE merupakan parameter yang digunakan untuk menghitung perbedaan antara nilai prediksi rating dengan nilai rating pengguna sesungguhnya (Wang & Lu, 2018). Parameter NMAE merupakan normalisasi dari nilai MAE. Nilai MAE, RMSE dan nilai NMAE semakin kecil maka hasil prediksi *rating* dari sistem rekomendasi semakin akurat (Muliadi & Lestari, 2019).

4 Kesimpulan

Pada penelitian ini disajikan Sistem Rekomendasi dengan Algoritma *User K-Nearest Neighbors* untuk prediksi peringkat penilaian produk oleh pengguna di Tokopedia. Performa dari model yang dirancang didapatkan nilai RMSE = 0.713, MAE = 0.488 dan NMAE = 0.122. Nilai performa yang didapatkan sangat kecil yaitu dibawah 1 yang membuktikan bahwa Algoritma *User K-Nearest Neighbors* cocok sebagai model untuk prediksi rating pada Sistem Rekomendasi.

Daftar Pustaka

- Belavagi, M. C., & Muniyal, B. (2016). Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection. *Procedia Computer Science*, 89, 117–123. <https://doi.org/10.1016/j.procs.2016.06.016>
- Choudhary, B. T., & Tulasi, B. (2019). Recommender system for personalised travel itinerary. *International Journal of Electrical and Computer Engineering*, 9(5), 4460–4465. <https://doi.org/10.11591/ijece.v9i5.pp4460-4465>
- Dharma, A. S., & Samosir, T. (2019). The User Personalization with KNN for Recommender System. *Sinkron*, 3(2), 45–48. <https://doi.org/10.33395/sinkron.v3i2.10047>
- Muliadi, K. H., & Lestari, C. C. (2019). Rancang Bangun Sistem Rekomendasi Tempat Makan Menggunakan Algoritma Typicality Based Collaborative Filtering. *Techno.Com*, 18(4), 275–287. <https://doi.org/10.33633/tc.v18i4.2515>
- 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>
- Roh, Y., Heo, G., & Whang, S. E. (2019). A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering*, PP(c), 1–1. <https://doi.org/10.1109/tkde.2019.2946162>
- Shafique, M. A., & Hato, E. (2015). Formation of Training and Testing Datasets, for Transportation Mode Identification. *Journal of Traffic and Logistics Engineering*, 3(1), 77–80. <https://doi.org/10.12720/jtle.3.1.77-80>
- Syah, R. D. (2020). Metode Decision Tree untuk Klasifikasi Hasil Seleksi Kompetensi Dasar pada CPNS 2019 di Arsip Nasional Republik Indonesia. *Jurnal Ilmiah Informatika Komputer*, 25(2), 107–114. <https://doi.org/10.35760/ik.2020.v25i2.2750>
- Wang, K., Zhang, T., Xue, T., Lu, Y., & Na, S.-G. (2019). E-Commerce Personalized Recommendation Analysis by Deeply-learned Clustering. *Journal of Visual Communication and Image Representation*, 71. <https://doi.org/10.1016/j.jvcir.2019.102735>
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>