

Perbandingan Metode *K-Nearest Neighbor* Dan *Decision Tree* Pada Sistem Rekomendasi Laptop

Muhammad Izdihar Alwin

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Firza Prima Aditiawan

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Muhammad Muharrom Al Haromainy

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: 19081010047@student.upnjatim.ac.id

Abstract. Nowadays, many people need a laptop to serve their daily needs such as work, study or entertainment. However, not everyone knows the indicators in the selection of laptop products according to taste and criteria for laptop needs with the variety of laptop specifications available on the market. To overcome these problems, a laptop recommendation system has been developed. This system will help consumers choose a laptop that suits their needs by recommending various laptop options based on the preferences and features that users enter. In this research, *K-Nearest Neighbor* and *Decision Tree* algorithms are used to see the performance in providing recommendations.

Keywords: *K-Nearest Neighbor*, *Decision Tree*, Laptop

Abstrak. Saat ini, banyak orang membutuhkan laptop untuk melayani kebutuhan sehari-hari seperti bekerja, belajar atau hiburan. Namun tidak semua orang mengetahui indikator dalam pemilihan produk laptop sesuai selera dan kriteria kebutuhan laptop dengan beragamnya spesifikasi laptop yang tersedia di pasaran. Untuk mengatasi masalah tersebut, telah dikembangkan sistem rekomendasi laptop. Sistem ini akan membantu konsumen memilih laptop yang sesuai dengan kebutuhan mereka dengan merekomendasikan berbagai pilihan laptop berdasarkan preferensi dan fitur yang dimasukkan pengguna. Pada penelitian ini, menggunakan algoritma *K-Nearest Neighbor* dan *Decision Tree* untuk melihat performa dalam memberikan rekomendasi.

Kata kunci: *K-Nearest Neighbor*, *Decision Tree*, Laptop

LATAR BELAKANG

Dengan semakin berkembangnya teknologi maka kebutuhan mengenai teknologi pada saat ini semakin meningkat. Perkembangan teknologi menyebabkan penggunaan laptop menjadi suatu kebutuhan yang mendasar, sehingga laptop harus dimiliki agar dapat menyelesaikan pekerjaan dan menggantikan PC. Laptop adalah suatu perangkat computer portable yang cocok digunakan ketika bepergian atau untuk pengguna mobile serta menjadi kebutuhan yang dapat membantu pekerjaan sehari-hari. Adapun komponen yang terdapat pada laptop yaitu ukurannya yang diperkecil, dibuat lebih ringan serta hemat daya. laptop saat ini memiliki berbagai macam merk dan spesifikasi yang dapat membuat orang kesulitan dalam memilih laptop yang sesuai kebutuhan. Dengan adanya permasalahan tersebut maka

masyarakat yang membutuhkan laptop akan menjadi lebih terbantu ketika terdapat suatu sistem yang dapat memberikan rekomendasi laptop yang sesuai dengan kebutuhan (Ikhsani Suwandy Putri & Siti Fathonah, 2023).

Sistem rekomendasi dapat mengatasi masalah pertumbuhan data dengan memberikan rekomendasi kepada pengguna dengan mempelajari informasi dari data masa lalu dan memberikan saran. 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 (Syah, 2020).

KAJIAN TEORITIS

Laptop

Laptop adalah komputer yang mempunyai ukuran kecil dan ringan serta ramping sehingga mudah dibawa, dapat bergerak serta dipindahkan, memiliki berat 1-3 kg dan 1-6 kg. Laptop berasal dari lap dan top, lap artinya pangkuan top artinya atas sehingga dapat diartikan laptop merupakan suatu alat yang dapat disimpan diatas pangkuan dengan ukuran yang kecil (Ikhsani Suwandy Putri & Siti Fathonah, 2023).

Sistem Rekomendasi

Sistem rekomendasi adalah perangkat lunak yang memberikan rekomendasi produk atau item yang menarik bagi pengguna, dengan desain yang dipersonalisasi. Sistem ini berguna untuk memberikan rekomendasi berdasarkan preferensi masa lalu, riwayat pembelian, dan informasi demografis, dan sangat bermanfaat bagi pengguna yang belum berpengalaman dalam memilih alternatif yang relevan (Novianti Indah Putri et al., 2021).

Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (induction-based learning) adalah proses pembentukan definisi definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep konsep yang akan dipelajari. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD (Abdul Koda et al., 2022).

K-Nearest Neighbor

Algoritma KNN adalah salah satu algoritma yang sering digunakan untuk melakukan klasifikasi. Algoritma termasuk dalam algoritma lazy learning yang mudah untuk diimplementasikan. Dalam penggunaan algoritma KNN data dibagi menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan algoritma untuk melakukan dasar prediksi, sedangkan data uji terdiri dari nilai yang diprediksi oleh algoritma (Aisha & Kusumawati, 2022). K-Nearest Neighbor merupakan sebuah algoritma yang digunakan untuk melakukan klasifikasi data dengan mempertimbangkan jarak terdekat antara objek data. Penentuan nilai K yang optimal untuk algoritma ini didasarkan pada data yang telah dikumpulkan. Jika nilai K yang dipilih terlalu besar, maka akan mengurangi efek noise pada klasifikasi, namun dapat membuat batas antara setiap klasifikasi menjadi lebih samar. Pengukuran jarak yang digunakan adalah Euclidean Distance. Jarak Euclidean Distance didefinisikan pada persamaan berikut (Homepage et al., 2021).

$$d_i = \sqrt{\sum_{i=1}^p (x_1 - x_2)^2} \quad (1)$$

Keterangan:

- d = Jarak
- i = Variabel data
- p = Dimensi data
- x_1 = Sampel data
- x_2 = Data uji

Decision Tree

Metode Decision Tree dianggap sebagai algoritma yang banyak digunakan untuk mengkategorikan data atau objek, dengan hasil klasifikasi diperoleh dengan jelas dan mampu mengatasi missing value, menangani data kontinu, menangani pruning data, dan menggunakan nilai gain ratio sebagai kunci utama untuk memecahkan masalah yang ada. Permasalahan yang memiliki output bernilai diskrit dapat dipecahkan menggunakan Decision Tree (Oktafiani & Rianto, 2023). Decision Tree merupakan teknik klasifikasi yang digunakan untuk mengekstrak hubungan yang relevan dalam data. Decision Tree adalah program yang membuat pohon keputusan berdasarkan pada set data input berlabel. Decision Tree membagi data training dengan bantuan perolehan informasi. Atribut yang memiliki frekuensi tinggi dipertimbangkan untuk memisahkan data berdasarkan informasi yang tersedia dalam dataset (Maulidah et al., 2020). Untuk mengetahui nilai entropy maka digunakan persamaan berikut

$$\text{Entropy}(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

Keterangan:

- S = Himpunan kasus
- A = Atribut
- n = Jumlah partisi S
- p_i = Proporsi dari S_i terhadap S

Setelah didapatkan nilai entropy, selanjutnya adalah menghitung nilai gain dengan menggunakan persamaan berikut

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (3)$$

Keterangan:

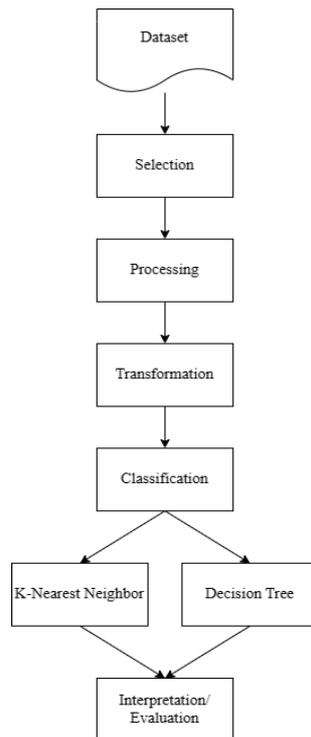
- S = Himpunan kasus
- A = Atribut
- n = Jumlah partisi atribut
- $A|S_i|$ = Jumlah kasus pada partisi ke-i
- $|S|$ = Jumlah kasus dalam S

Confusion Matrix

Confusion matrix merupakan cara untuk melakukan sebuah evaluasi terhadap klasifikasi. Matrix ini berisikan data target prediksi yang dibandingkan dengan data target aktual, Accuracy atau rasio prediksi benar memberikan gambaran terkait seberapa akurat sebuah model Machine Learning memprediksi nilai dengan benar dengan keseluruhan data (Wayan et al., 2023).

METODE PENELITIAN

Pada penelitian ini, peneliti mengikuti alur analisis data mining. Data mining dapat diartikan proses pencarian pola data yang tidak diketahui atau tidak diduga sebelumnya. Ide dasarnya adalah menggali suatu sumber yang berharga dari tempat yang sama sekali tidak diperkirakan sebelumnya. Data mining juga dikenal dengan istilah knowledge discovery in databases atau KDD. Proses knowledge discovery terdiri dari pembersihan data (data cleaning), integrasi data (data integration), pemilihan data (data selection), transformasi data (data transformation), data mining, evaluasi pola (pattern evaluation) dan penyajian pengetahuan (knowledge presentation) (Anggara et al., 2022). Tahapan proses KDD dalam Data Mining dapat dilihat pada gambar berikut.



Gambar 1. Alur analisis data mining

1) *Selection*

Selection merupakan proses memilih data atau menyeleksi data yang relevan untuk digunakan. Sering kali tidak semua data yang ada pada dataset digunakan, hanya beberapa yang sesuai untuk keperluan analisis. Penelitian ini menggunakan dataset 2022 March Laptop data yang dipublish oleh Santosh Kumar pada platform Kaggle. Dataset ini berisi berbagai kriteria laptop seperti nama brand, model, prosesor, ram, storage, os, harga dan sebagainya.

2) *Pre-Processing*

Pre-processing adalah proses membuang atau menghapus data-data yang tidak lengkap, error, dan tidak konsisten pada dataset. Pada dataset dengan jumlah data 896 dilakukan pre-processing sehingga pada penelitian ini dataset yang digunakan berjumlah 21 data.

3) *Transformation*

Transformation merupakan tahap untuk menghasilkan data yang siap digunakan dengan mengubah data dari bentuk asalnya sesuai kebutuhan untuk mempermudah proses penggalian data dalam menemukan suatu pengetahuan baru.

4) *Classification*

Classification adalah proses penemuan model atau fungsi untuk menentukan dan mengelompokkan tipe data yang bertujuan agar dapat memprediksi kelas dari objek yang

label kelasnya tidak diketahui. Pada tahap ini juga dilakukan pemilihan algoritma dan implementasi dari algoritma yang dipilih. Pada penelitian ini menggunakan *K-Nearest Neighbor* dan *Decision Tree*.

5) *Interpretation*

Interpretation adalah proses memahami berbagai jenis data penelitian yang sudah diolah dan menafsirkan atau menyimpulkan hasil dari tahap sebelumnya. Pada tahap ini ditampilkan hasil dari implementasi algoritma *K-Nearest Neighbor* dan *Decision Tree*

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan bahasa pemrograman Python untuk melakukan perbandingan algoritma K-Nearest Neighbor dan Decision Tree

1. K-Nearest Neighbor

Pada tahap ini dilakukan classification dengan algoritma K-Nearest Neighbor. Dataset yang digunakan terdapat pada tabel 2. Langkah pertama adalah menghitung jarak euclidean dari data ke-1 sampai dengan data ke-20 yang digunakan sebagai data latih sedangkan data ke-21 digunakan sebagai data uji. Langkah kedua adalah menentukan peringkat dari jarak euclidean

Tabel 1. Hasil *Euclidean Distance*

no	jarak euclidean	peringkat	K = 1	K = 3	K = 5
19	8,659	1	1	1	1
2	9,057	2		1	1
8	10,953	3		1	1
15	12,206	4			1
13	14,516	5			1

Pada tabel 1, setelah menentukan nilai k, dapat dilihat bahwa pada data ke-19, 2, 8, 15, dan 13 bernilai 1 (recommend) sehingga data ke-21 juga bernilai 1. Langkah selanjutnya adalah melihat performa algoritma menggunakan *confusion matrix*. Hasil dari *confusion matrix* dapat dilihat pada gambar berikut

Confusion Matrix KNN:

```
[[15 14]
 [10 19]]
```

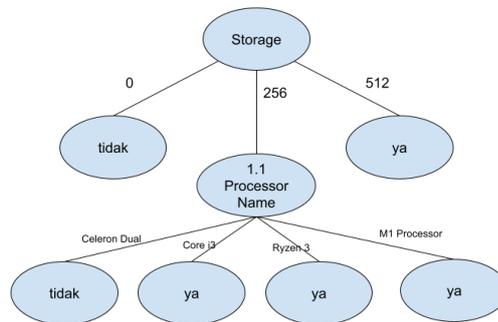
True Positives (TP): 19
 True Negatives (TN): 15
 False Positives (FP): 14
 False Negatives (FN): 10

Akurasi: 0.5862068965517241
 Presisi: 0.5757575757575758
 Recall: 0.6551724137931034
 F1-score: 0.6129032258064515

Gambar 1. Confusion matrix KNN

2. Decision Tree

Pada tahap ini dilakukan classification dengan algoritma Decision Tree. Dataset yang digunakan terdapat pada tabel 1. Langkah pertama adalah mengelompokkan semua atribut dan menghitung jumlah dari tiap atribut kemudian menghitung entropy dan gain dari tiap atribut yang telah dikelompokkan. Setelah menentukan nilai entropy dan gain tiap atribut, langkah berikutnya adalah menentukan nilai max gain untuk dijadikan sebagai node pertama atau akar pertama. Nilai max gain terdapat pada atribut storage dengan nilai gain 0,57617. Langkah selanjutnya membuat pohon keputusan dengan atribut storage sebagai node pertama.



Gambar 2. Decision Tree node 1.1

Langkah selanjutnya adalah melihat performa algoritma menggunakan *confusion matrix*. Hasil dari *confusion matrix* dapat dilihat pada gambar berikut

Confusion Matrix:

```
[[35, 19], [27, 34]]
```

Akurasi: 0.6

Presisi: 0.6415094339622641

Recall: 0.5573770491803278

F1 Score: 0.5964912280701753

Gambar 3. Confusion matrix Decision Tree

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis perbandingan kedua algoritma, dapat disimpulkan bahwa algoritma K-Nearest Neighbor dan Decision Tree dapat diimplementasi untuk menghasilkan rekomendasi laptop. Namun hasil dari confusion matrix menunjukkan bahwa kedua algoritma memiliki tingkat accuracy, precision, recall, dan f1-score yang tidak terlalu signifikan perbedaannya, sehingga pada kondisi tertentu hasil dari confusion matrix pada kedua algoritma dapat berubah bergantung pada dataset yang digunakan. Adapun saran untuk penelitian berikutnya ialah dapat melakukan komparasi dengan algoritma dan dataset yang berbeda sehingga mendapatkan hasil yang beragam.

DAFTAR REFERENSI

- Abdul Koda, Rahayu, P., Pratama, A., Rafly, A., & Kaslani. (2022). Penentuan Bonus Karyawan Dengan Menggunakan Algoritma K-Nearest Neighbor. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(1), 14–20. <https://doi.org/10.32485/kopertip.v4i1.115>
- Aisha, D., & Kusumawati, R. (2022). IMPLEMENTASI METODE ALGORITMA COLLABORATIVE FILTERING DAN K-NEAREST NEIGHBOR PADA SISTEM REKOMENDASI E-COMMERCE. *JUISIK*, 2(3). <http://journal.sinov.id/index.php/juisik/indexHalamanUTAMAJurnal:https://journal.sinov.id/index.php>
- Anggara, E. D., Widjaja, A., & Suteja, B. R. (2022). Prediksi Kinerja Pegawai sebagai Rekomendasi Kenaikan Golongan dengan Metode Decision Tree dan Regresi Logistik. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1). <https://doi.org/10.28932/jutisi.v8i1.4479>
- Homepage, J., Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- Ikhsani Suwandiy Putri, N., & Siti Fathonah, N. (2023). IMPLEMENTASI SISTEM REKOMENDASI LAPTOP MENGGUNAKAN METODE CONTENT BASED FILTERING DAN K-MEANS BERBASIS MOBILE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1329–1336. <https://doi.org/10.36040/jati.v7i2.6962>
- Maulidah, M., Windu Gata, Rizki Aulianita, & Cucu Ika Agustyaningrum. (2020). ALGORITMA KLASIFIKASI DECISION TREE UNTUK REKOMENDASI BUKU BERDASARKAN KATEGORI BUKU. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 89–96. <https://doi.org/10.51903/e-bisnis.v13i2.251>
- Novianti Indah Putri, Rustiyana, Yudi Herdiana, & Zen Munawar. (2021). Sistem Rekomendasi Hibrid Pemilihan Mobil Berdasarkan Profil Pengguna dan Profil Barang. *TEMATIK*, 8(1), 56–68. <https://doi.org/10.38204/tematik.v8i1.566>

- Oktafiani, R., & Rianto, R. (2023). Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 113–121. <https://doi.org/10.25077/TEKNOSI.v9i2.2023.113-121>
- 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>
- Wayan, I., Argodi, A., Pembangunan, U., Veteran, N. ", Jawa, ", Eva, T., Puspaningrum, Y., Timur, J., Muharrom, M., & Haromainy, A. (2023). IMPLEMENTASI METODE TF-IDF DAN ALGORITMA NAIVE BAYES DALAM APLIKASI DIABETIC BERBASIS ANDROID. *Elektro Dan Ilmu Komputer(TEKNIK)*, 3(2). <https://doi.org/10.55606>