

Comparison of RAM Classification with Decision Tree Algorithms and KNN

Komparasi Klasifikasi RAM dengan Algoritma Decision Tree dan KNN

Solehudin Al Ayyubi, Indira Setia Amalia, Amalia Anjani Arifiyanti

{sholehudinyubi@gmail.com, indirasetiaa@gmail.com, amalia_anjani.fik@upnjatim.ac.id}

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Abstract. *A laptop is a rather small personal computer consisting of a keyboard, screen, microprocessor and usually a rechargeable battery. In the current era of technology, when buying a laptop, we need to have thorough and detailed considerations regarding our needs in using a laptop, especially laptop RAM. By comparing Decision Tree and KNearest Neighbors algorithm, a more accurate algorithm was found to make a prediction in buying a laptop with suitable variable consideration. The result shows that Decision Tree algorithm is more accurate to be used in predicting the suitable laptop RAM. Decision Tree accuracy is 68%, this result is higher than KNN accuracy which is only 66%.*

Keywords - *Classification; Data Mining; Decision Tree; KNN; Laptop RAM*

Abstrak. *Laptop adalah komputer pribadi yang agak kecil terdiri dari papan tombol, layar tampilan, mikroprosesor dan adanya baterai yang biasanya dapat diisi ulang. Di era teknologi saat ini, Dalam membeli laptop, perlu pertimbangan yang menyeluruh dan rinci terkait kebutuhan kita dalam menggunakan laptop terutama RAM laptop. Dengan membandingkan antara algoritma Decision Tree dan K-Nearest Neighbor ditemukan algoritma yang lebih akurat untuk melakukan prediksi dalam membeli laptop dengan pertimbangan variabel yang sesuai. Hasil penelitian menunjukkan bahwa algoritma Decision Tree lebih akurat untuk digunakan dalam memprediksi RAM laptop yang sesuai dengan hasil akurasi Decision Tree sebesar 68% sedangkan KNN hanya sebesar 66%.*

Kata Kunci – *Klasifikasi; Data Mining; Decision Tree; KNN; RAM Laptop*

I. PENDAHULUAN

Penggunaan komputer saat ini tidak lepas dengan kehidupan masyarakat. Seiring dengan perkembangan teknologi perangkat keras, komputer yang dahulu berukuran besar lambat laun semakin *compact*. Laptop merupakan salah satu bentuk perkembangan teknologi komputer yang mana dapat dibawa kemanapun karena ringan dan ukurannya cukup kecil. Oleh karena fleksibilitasnya, penggunaan laptop saat ini telah banyak menggantikan personal komputer [1]. Spesifikasi perangkat keras laptop saat ini sangatlah beragam. Setiap pengguna yang membutuhkan laptop dapat mencari spesifikasi laptop menyesuaikan dengan tujuan penggunaannya [2][3].

Pemilihan laptop merupakan hal yang penting karena berhubungan dengan performa laptop dalam mendukung tujuan penggunaan laptop. Kriteria harga, *processor*, dan RAM (*Random Access Memory*) merupakan tiga hal yang digunakan sebagai variabel dalam penentuan pembelian laptop [4][5]. Dari segi perangkat keras, *processor* dan RAM dapat menjadi penentu kecepatan pemrosesan pada komputer. Oleh karenanya, semakin tinggi versi processornya maka semakin cepat pemrosesan kegiatan pada laptop. Hal ini juga berlaku pada RAM, semakin besar kapasitas RAM pada laptop maka semakin cepat pemrosesan laptop.

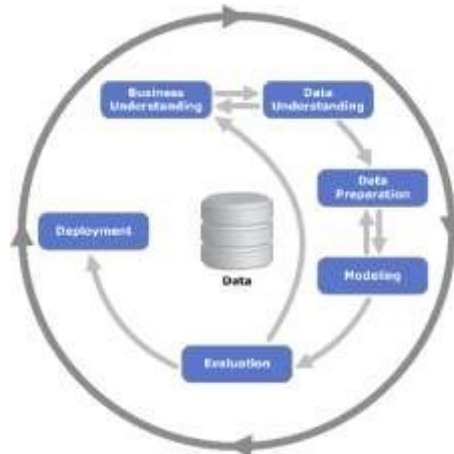
Klasifikasi yang merupakan salah satu *task data mining* dapat digunakan untuk memprediksi spesifikasi laptop yang salah satunya adalah RAM. Algoritma yang digunakan pada klasifikasi mengacu pada mesin pembelajar berjenis *supervised learning*. Algoritma mesin pembelajar ini diantaranya adalah Decision Tree dan KNN. Keduanya merupakan algoritma mesin pembelajar yang banyak digunakan pada kegiatan klasifikasi [6][7]. Dalam penelitian ini akan dilakukan komparasi dua algoritma klasifikasi yaitu *Decision Tree* dan KNN dalam perancangan model klasifikasi untuk menentukan RAM yang cocok pada laptop.

II. METODE

Metodologi yang digunakan dalam penelitian ini mengacu pada model *Cross-Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6 fase yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* yang dapat dilihat pada gambar 1 [8].

A. Pengumpulan data

Data yang digunakan adalah data yang diambil dari situs AtapData. Pada situs tersebut, data yang dipilih adalah data laptop dengan berbagai merek dan spesifikasinya. *Dataset* tersebut memiliki beberapa atribut yaitu *company*, *product*, *typename*, *inches*, *screen resolution*, *CPU*, *RAM*, *memory*, *gpu*, *opsys*, *weight*, *price_euros*. Data laptop yang dikumpulkan berbentuk file dalam format csv.



Gambar 1. Tahapan CRISP-DM [9].

B. Data mining

Data mining merupakan proses ekstraksi informasi yang sifatnya implisit atau tidak diketahui secara gamblang melalui penggalian pola pada data. *Data mining* digunakan untuk menggali informasi ini agar dapat dimanfaatkan. Proses ini dinyatakan sebagai eksplorasi dan analisis data dengan kuantitas data yang berjumlah banyak dan berdimensi tinggi. Proses ini dilakukan secara otomatis dengan bantuan mesin pembelajaran.

C. Klasifikasi

Klasifikasi merupakan salah satu task data mining yang digunakan untuk memprediksi kategori/*class* dari suatu data. Klasifikasi mayoritas menggunakan mesin pembelajaran bersifat *supervised learning*. Oleh karenanya, diperlukan pembagian data yang seagamainya digunakan untuk membuat model klasifikasi dan sisanya digunakan untuk pengujian model klasifikasi.

D. Decision tree

Algoritma pembelajar *Decision Tree* menggunakan konsep pohon keputusan. Algoritma ini menghasilkan aturan-aturan yang merupakan representasi dari pohon keputusan. Hal ini digunakan untuk memprediksi *class* berdasarkan aturan-aturan yang telah dibangun. Pohon keputusan digunakan untuk melakukan eksplorasi data dengan cara menemukan hubungan tersembunyi antara jumlah variabel input dengan sebuah variabel target.

E. KNN (*k-nearest neighbors*)

KNN merupakan algoritma klasifikasi yang menggunakan konsep *neighbors*. Semakin dekat suatu obyek dengan obyek pada model klasifikasi pada maka obyek tersebut akan dinyatakan memiliki kategori/*class* yang sama. Oleh karenanya, prediksi kelas oleh ditentukan dengan jarak *neighbors* terdekat.

F. Confusion matrix

Matrix ini populer digunakan untuk evaluasi performa model klasifikasi [10][11]. Matrix ini dapat digunakan pada binary class maupun multiple class. Confusion matrix berupa tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

III. HASIL DAN PEMBAHASAN

A. Business understanding/pemahaman bisnis

Pemilihan RAM untuk laptop perlu diprediksi secara otomatis. Prediksi ini dapat dilakukan dengan menggunakan klasifikasi. Untuk itu, diperlukan data mengenai spesifikasi perangkat keras laptop untuk perancangan model klasifikasinya. Algoritma mesin pembelajaran yang digunakan pada penelitian ini digunakan algoritma yang populer

digunakan yaitu KNN dan *Decision Tree*. Dilakukan komparasi hasil performa model klasifikasi dari kedua algoritma tersebut, sehingga dapat dipilih model klasifikasi yang memiliki performa paling optimal.

B. Data understanding/ pemahaman data dan data preparation/ penyiapan data

Total dataset model laptop yang dikumpulkan sebanyak 1303 laptop dengan 13 atribut. Data ini perlu dipilih menyesuaikan dengan kebutuhan. Pemilihan data dilakukan pada tahap pemahaman data dan penyiapan data. Pada tahap ini, dilakukan proses identifikasi dan penghapusan data yang memiliki *outlier/noise*, data yang tidak konsisten, dan data yang tidak lengkap (*missing values*). Dari total data yang berhasil dikumpulkan, 442 diantaranya memiliki missing value sehingga baris data ini dihapus. Total data yang digunakan pada penelitian ini setelah penghapusan *missing value* sebanyak 861 baris data. Lalu, dari 13 atribut hanya dipilih 4 atribut dengan 3 atribut prediktor dan 1 atribut label/*class*.

Setelah proses pemilihan dan penghapusan data dan atribut, tahap selanjutnya adalah transformasi data. Beberapa atribut yang bertipe numerik dikonversi menjadi nominal sehingga dapat diproses lebih lanjut. Atribut label/*class* didefinisikan menjadi *binary class* yaitu kelas 0 dan kelas 1. Kelas 0 merepresentasikan ketidaksesuaian, sedangkan kelas 1 merepresentasikan kesesuaian RAM.

C. Pembuatan model dan evaluasi model

Dilakukan pembuatan model klasifikasi dengan dua skenario yaitu: 1) pemodelan dengan metode *decision tree*, dan 2) pemodelan dengan metode KNN. Untuk pembuatan model ini, *dataset* dibagi menjadi dua jenis yaitu data *training* dan data *testing*. Pembagian data pada penelitian ini menggunakan metode *hold-out* dengan proporsi 60%:40%, dimana 60% total data digunakan sebagai data *training* yaitu sebanyak 516 dan 40% sisanya digunakan sebagai data *testing* yaitu sebesar 345 data. Hasil model klasifikasi dievaluasi menggunakan metode *confusion matrix*.

Pada skenario pertama yaitu pembuatan model dengan metode *Decision Tree* dihasilkan hasil evaluasi performa seperti pada Tabel 1. Diketahui pada kelas 0, ketepatan prediksi sebanyak 212 data dan pada kelas 1, ketepatan prediksi sebanyak 24 data. Kesalahan prediksi terjadi pada dua kelas tersebut. Pada kelas 0 yang diprediksi salah yaitu diprediksi ke kelas 1 sebanyak 29 data dan 80 data kelas 1 diprediksi ke kelas 0. Ketepatan prediksi pada kelas 0 sebesar 88%, sedangkan tingkat prediksi benar pada kelas 0 sebesar 23,1%. Dari hasil ini dapat dinyatakan bahwa model klasifikasi pada skenario ini memiliki kecenderungan prediksi ke kelas 0.

Tabel 1. *Confusion Matrix* Metode *Decision Tree*

Prediksi \ Aktual	Kelas 0	Kelas 1
Kelas 0	212	29
Kelas 1	80	24

Pada skenario kedua yaitu pembuatan model dengan metode KNN dihasilkan hasil evaluasi performa seperti pada Tabel 2. Ketepatan prediksi pada kelas 0 sebanyak 197 data, sedangkan pada kelas 1, ketepatan prediksi sebanyak 30 data. Kesalahan prediksi terjadi pada dua kelas tersebut. Pada kelas 0 yang diprediksi salah yaitu diprediksi ke kelas 1 sebanyak 44 data dan 74 data kelas 1 diprediksi ke kelas 0. Ketepatan prediksi pada kelas 0 sebesar 81,7%, sedangkan tingkat prediksi benar pada kelas 0 sebesar 28,9%. Dari hasil ini dinyatakan bahwa model klasifikasi pada skenario kedua ini memiliki kecenderungan prediksi ke kelas 0.

Tabel 2. *Confusion Matrix* Metode KNN

Prediksi \ Aktual	Kelas 0	Kelas 1
Kelas 0	197	44
Kelas 1	74	30

Berdasarkan *Confusion Matrix* pada tabel 1 dan 2, dihasilkan nilai untuk akurasi, *precision*, *recall*, F1, dan ROC AUC dari metode *Decision Tree* dan metode KNN. Pada Tabel 3 terlihat bahwa nilai akurasi dari *Decision Tree* memiliki persentase lebih tinggi daripada nilai akurasi dari KNN, dimana nilai akurasi *Decision Tree* sebesar 68% dan nilai akurasi KNN sebesar 66%. Sedangkan persentase nilai ROC AUC dari *Decision Tree* lebih tinggi 1% daripada persentase nilai dari KNN.

Tabel 3. Hasil Pengujian Confusion Matrix

Evaluasi Confusion Matrix	Hasil Decision Tree	Hasil KNN
Akurasi	68%	66%
Precision	59%	57%
Recall	56%	55%
F1	55%	55%
ROC AUC	56%	55%

IV. KESIMPULAN

Dari hasil penelitian yang dilakukan, komparasi pengujian metode *decision tree* dan metode KNN pada *dataset* laptop dengan prediksi untuk menentukan RAM yang sesuai menunjukkan hasil pengujian menggunakan Metode *decision tree* pada memiliki hasil yang lebih akurat daripada menggunakan Metode KNN. Nilai akurasi yang didapat dari metode *decision tree* menggunakan sebesar 68,40%. Sehingga, dapat disimpulkan bahwa pengujian untuk dataset laptop lebih akurat jika menggunakan metode *Decision Tree*. Akan tetapi, model klasifikasi yang dihasilkan baik oleh *decision tree* maupun KNN memiliki kecenderungan prediksi ke kelas 0. Oleh karenanya perlu perbaikan di masa mendatang terhadap model klasifikasi agar prediksi lebih berimbang terhadap kedua kelas. Selain itu, disarankan menggunakan atribut yang lebih banyak agar dapat menghasilkan model klasifikasi yang memiliki performa prediksi yang lebih akurat.

REFERENSI

- [1] H. Nuryanto, *Sejarah Perkembangan Teknologi Informasi dan Komunikasi*. Balai Pustaka (Persero), 2012.
- [2] M. R. Noviansyah, W. Suharso, D. R. Chandranegara, M. S. Azmi, and M. Hermawan, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN LAPTOP PADA E-COMMERCE MENGGUNAKAN METODE WEIGHTED PRODUCT," in *Prosiding SENTRA (Seminar Teknologi dan Rekayasa)*, Nov. 2019, vol. 5, pp. 43–53. doi: 10.22219/SENTRA.V0I5.3025.
- [3] G. P. Sanyoto, R. I. Handayani, and E. Widanengsih, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN LAPTOP UNTUK KEBUTUHAN OPERASIONAL DENGAN METODE AHP (STUDI KASUS: DIREKTORAT PEMBINAAN KURSUS DAN PELATIHAN KEMDIKBUD)," *Jurnal Pilar Nusa Mandiri*, vol. 13, no. 2, pp. 167–174, 2017, Accessed: Jun. 08, 2022. [Online]. Available: <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/233>
- [4] R. Elnatan and H. Tannady2, "Alternatif Pemilihan Laptop Bagi Mahasiswa Di Jakarta Utara Menggunakan Metode Analytic Hierarchy Process," *Jurnal Teknologi*, vol. 10, no. 1, pp. 8–13, Jul. 2020, doi: 10.35134/JITEKIN.V10I1.17.
- [5] J. E. S. Casym and D. N. Oktara, "Aplikasi Analytical Hierarchy Process dalam Mengidentifikasi Preferensi Laptop Bagi Mahasiswa," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, Feb. 2020, pp. 636–640.
- [6] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research*, vol. 9, pp. 381–386, 2020, doi: 10.21275/ART20203995.
- [7] N. Hidayati, J. Suntoro, and G. G. Setiaji, "Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM," *Jurnal Sains dan Informatika*, vol. 7, no. 2, pp. 117–126, Nov. 2021, doi: 10.34128/JSI.V7I2.313.
- [8] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.
- [9] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Computer Science*, vol. 6, pp. 1–43, May 2020, doi: 10.7717/PEERJ-CS.267/SUPP-2. [10] P. Joshi, *Artificial Intelligence with Python*. Packt Publishing, 2017.
- [10] R. Susmaga, "Confusion Matrix Visualization," *Intelligent Information Processing and Web Mining*, pp. 107–116, 2004, doi: 10.1007/978-3-540-39985-8_12.