Jurnal Komputer dan Informatika
Vol 20 No 1, April 2025: hlm 90 - 100

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

# KLASIFIKASI BINTANG KATAI DAN RAKSASA DENGAN METODE K-NN DAN DECISION TREE

**Felix Ferdinand,**
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
*E-mail: [1]felix.535220161@stu.untar.ac.id*

## ABSTRAK

Bintang adalah objek langit yang memancarkan cahaya saat malam hari. Bintang memiliki beberapa sifat seperti, tingkat kecerahannya dan suhunya berdasarkan warna. Sifat-sifat tersebut bisa digunakan untuk klasifikasi dan yang biasanya digunakan adalah klasifikasi Harvard dan klasifikasi Morgan-Keenan. Kedua klasifikasi tersebut dapat digunakan untuk mengklasifikasi besarnya suatu bintang sebagai bintang katai dan bintang raksasa. Dengan banyaknya bintang yang harus diklasifikasikan, klasifikasi manual dapat memakan banyak waktu. Maka, pada penelitian ini digunakan algoritma K-NN dan *Decision Tree* untuk mengetahui keakuratan klasifikasi bintang katai dan bintang raksasa. Dataset yang digunakan berasal dari katalog Gaia DR3 dan Hipparcos and Tycho. Sebanyak 25182 baris atau *record* digunakan. Komposisi eksperimen yang dilakukan pada penelitian ini adalah komposisi 80% data latih dan 20% data uji (80/20) dan komposisi 60% data latih dan 20% data uji (60/20). Eksperimen dilakukan 10 kali dengan mengacak dataset untuk data latih dan data uji. Hasil nilai rata-rata akurasi yang didapatkan dengan *Decision Tree* sebesar 81.5% untuk komposisi 80/20 dan komposisi 60/40. Sedangkan, dengan K-NN didapatkan nilai rata-rata akurasi sebesar 80.9% untuk komposisi 80/20 dan sebesar 81.1% untuk komposisi 60/40.

**Kata kunci**—Bintang, Gaia DR3, K-NN, *Decision Tree*

## *ABSTRACT*

*Stars are celestial objects that emit light at night. Stars have several properties such as brightness and temperature based on color. These properties can be used for classification and the most commonly used are the Harvard classification and the Morgan-Keenan classification. Both classifications can be used to classify the size of a star as a dwarf or giant star. With so many stars to classify, manual classification can take a lot of time. Therefore, in this study, the K-NN and Decision Tree algorithms are used to determine the accuracy of the classification of dwarf and giant stars. The dataset used in this study comes from the Gaia DR3 and Hipparcos and Tycho catalogues. A total of 25182 rows or records were used. The composition of experiments performed in this study is 80% training data and 20% test data (80/20) and 60% training data and 20% test data (60/20). Experiments were conducted 10 times by randomizing datasets for training data and test data. The average accuracy value obtained with Decision Tree is 81.5% for 80/20 composition and 60/40 composition. Meanwhile, with K-NN, the average accuracy value is 80.9% for the 80/20 composition and 81.1% for the 60/40 composition.*

***Keywords—****Stars, Gaia DR3, K-NN, Decision Tree*

## 1. PENDAHULUAN

Bintang adalah objek yang ada di langit. Objek langit tersebut memancarkan cahayanya sehingga terlihat pada malam hari. Namun, terkadang bintang tidak dapat dilihat secara langsung dengan kasatmata. Penyebab sulitnya melihat bintang pada malam hari adalah polusi cahaya. Polusi cahaya ini biasanya berasal dari cahaya lampu kota saat malam hari dan terangnya cahaya ini membuat bintang yang ada di langit sulit dilihat [1]. Sementara itu, pada pagi hari, bintang tidak dapat dilihat dengan jelas karena cahaya matahari yang terang menutupi cahaya bintang. Melihat atau mengamati bintang biasanya dilakukan jauh dari perkotaan atau di tempat yang tidak banyak cahaya [2].

Bintang memiliki banyak sifat atau ciri-ciri. Salah satunya adalah tingkat kecerahannya. Tingkat kecerahan bintang bervariasi, ada yang terang dan ada pula yang redup. Ukuran skala untuk tingkat kecerahan bintang disebut magnitudo tampak atau *apparent magnitude*. Skala magnitudo ini tidak hanya digunakan pada bintang, tetapi juga pada objek langit lainnya seperti planet. Semakin kecil skala magnitudo, maka semakin terang dan semakin besar skala magnitudo, maka semakin redup [3]. Tingkat kecerahan magnitudo tampak berasal dari bumi atau pengamat ke bintang yang diamati dan untuk perhitungan magnitudo tampak diperoleh dengan teknik fotometri, teleskop dan kamera [4]. Jika magnitudo tampak dari bumi ke bintang maka, magnitudo absolut atau *absolute magnitude* dari bumi ke bintang yang diasumsikan berjarak 10 parsec (32.6 tahun cahaya) atau sekitar 300 triliun km [5]. Sebagai contoh, jarak antara matahari dan bumi adalah 1 *astronomical unit* (149.6 juta km) [6]. Matahari yang sebelumnya 149.6 juta km dari bumi diasumsikan atau dipindahkan menjadi 10 parsec. Magnitudo absolut menjadi skala standar yang baik untuk mengukur tingkat kecerahan bintang. Ini karena dengan magnitudo absolut, bintang yang mempunyai magnitudo tampak yang berbeda-beda dapat disatukan dan dapat dibandingkan dengan bintang lainnya menggunakan magnitudo absolut.

Selain tingkat kecerahan bintang, sifat lain dari bintang adalah gelombang yang dipancarkannya. Bintang memancarkan spektrum elektromagnetik di mana terdapat gelombang seperti X-ray dan cahaya tampak (warna) [7]. Warna bintang biasanya berupa biru, merah, kuning, dan jingga. Bintang juga dianggap seperti benda hitam atau *black body* [8]. Benda hitam adalah objek yang hanya menyerap radiasi spektrum elektromagnetik dan memancarkan radiasinya [9]. Radiasi yang dipancarkan oleh bintang akan memancarkan cahaya dan warna. Warna tersebut berasal dari kurva radiasi gelombang warna yang dapat dilihat. Berdasarkan Hukum Wien, kurva radiasi tertinggi dengan panjang gelombang memiliki korelasi dengan suhu [10]. Pada bintang, jika dilakukan pengamatan pada kurva radiasi tertinggi dengan panjang gelombang, dapat terlihat bahwa warna biru yang memiliki suhu tertinggi dan semakin rendah suhu, semakin terlihat warna merah [11]. Namun, pada suhu yang jauh lebih rendah dari bintang seperti suhu tubuh manusia. Radiasi yang dipancarkan oleh manusia adalah suhu tubuh yang tidak terlihat secara langsung dengan kasatmata dan hanya terlihat melalui *infrared* karena kurva radiasi tertinggi di luar gelombang warna [12]. Pengamatan ini dapat dilihat pada simulasi warna benda hitam [13]. Teknik untuk menentukan warna bintang adalah indeks warna di mana filter warna biru dan filter warna merah digunakan untuk melihat tingkat warna tersebut dalam magnitudo [14].

Banyaknya ciri-ciri bintang dapat digunakan untuk klasifikasi bintang. Klasifikasi bintang yang sering digunakan adalah klasifikasi Harvard yang menunjukkan tipe kelas bintang (O, B, A, F, G, K, M) berdasarkan rentang suhu bintang di mana kelas O adalah bintang yang memiliki warna biru dan memiliki suhu paling tinggi, sedangkan kelas M adalah bintang yang memiliki warna merah dan memiliki suhu paling rendah [15]. Klasifikasi Harvard kemudian dibagi menjadi subtipe untuk setiap tipe kelasnya di mana subtipe mengklasifikasikan suhu dari 0 (paling dingin) hingga 9 (paling panas) [16]. Klasifikasi tambahan lainnya adalah klasifikasi Morgan-Keenan yang digunakan untuk mengklasifikasi bintang yang raksasa dan bintang yang katai (kecil) [17]. Kelas yang digunakan adalah angka romawi dari I hingga VII di mana I hingga IV adalah bintang raksasa dan V hingga VII adalah bintang katai [18]. Pengamatan klasifikasi yang dapat digunakan adalah dengan diagram Hertzsprung-Russell (diagram H-R). Diagram ini menunjukkan grafik bintang berdasarkan magnitudo absolut dan magnitudo filter warna atau suhu [19].

Klasifikasi bintang dapat dilakukan secara manual dengan Harvard, Morgan-Keenan, dan diagram H-R. Namun, dataset bintang yang besar dan memiliki banyak parameter atau variabel dapat memakan banyak waktu. Dalam penelitian ini dilakukan klasifikasi dengan menggunakan algoritma machine learning. Dengan penggunaan algoritma machine learning, klasifikasi dapat dilakukan cepat. Penggunaan algoritma machine learning perlu dilakukan evaluasi. Evaluasi dilakukan untuk mengetahui keakuratan model dalam klasifikasi bintang katai dan bintang raksasa terhadap dataset yang digunakan.

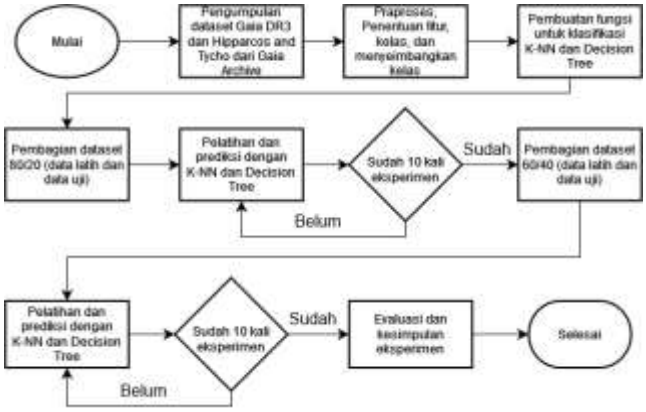
ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

## 2. METODE PENELITIAN

### 2.1 Metode dan Aplikasi

Metode penelitian yang digunakan pada penelitian ini adalah algoritma K-NN dan Decision Tree. Kemudian, untuk program atau aplikasi yang digunakan adalah JupyterLab (Python). Kedua algoritma digunakan dengan bantuan library sklearn. Library lainnya adalah pandas, matplotlib, dan seaborn. Pandas digunakan untuk membuat *data frame* dan membaca dataset. Matplotlib dan seaborn digunakan untuk membuat *confusion matrix*. Data yang digunakan pada penelitian ini adalah dataset katalog Gaia DR3 dan dataset tambahan katalog Hipparcos and Tycho. Eksperimen untuk kedua algoritma dilakukan dengan 80% data latih dan 20% data uji dan juga dilakukan dengan 60% data latih dan 40% data uji. Masing-masing eksperimen tersebut dijalankan 10 kali dengan mengacak dataset yang menjadi data latih dan data uji.

Gambar 1 menunjukkan flowchart penelitian ini. Alur dari penelitian ini dimulai dari pengumpulan dataset Gaia DR3 dan Hipparcos and Tycho dari Gaia Archive. Proproses dataset dilakukan untuk menghapus data yang tidak perlu. Penentuan fitur dan kelas juga dilakukan untuk membuat kelas label dan fitur yang penting. Pelatihan dan prediksi dilakukan sebanyak 10 kali untuk melihat apakah model bekerja secara konsisten atau tidak. Evaluasi dan kesimpulan untuk melihat apakah kedua model tersebut memberikan hasil yang baik atau tidak.




**Gambar 1** Flowchart penelitian

### 2.2 Algoritma K-Nearest Neighbors

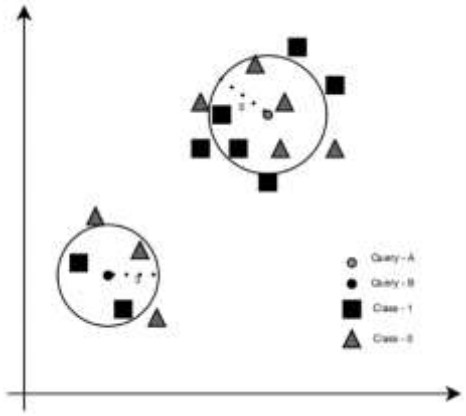
K-Nearest Neighbors atau K-NN adalah algoritma yang bekerja dengan mencari *neighbor* atau tetangga terdekat dan kemiripan dari setiap tetangga. Tetangga terdekat atau data latih yang dekat dengan data uji digunakan nilai k untuk menentukan banyaknya tetangga terdekat yang diambil. Hasil kelas data uji bergantung pada kemiripan dan banyaknya mayoritas kelas dari tetangga terdekat [20]. Ada beberapa rumus untuk untuk menghitung kemiripan tetangga. Salah satunya dengan persamaan Euclidean Distance [21]. Persamaan Euclidean Distance dapat dilihat pada persamaan (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Keterangan untuk persamaan (1) adalah:

- d (x, y) adalah nilai Euclidean Distance.
- n adalah jumlah dimensi atau kolom dalam dataset.
- $x_i$ adalah nilai pada kolom ke-i data latih.
- $y_i$ adalah nilai pada kolom ke-i data uji.

Gambar 2 menunjukkan bagaimana algoritma K-NN bekerja. Terdapat 2 kelas yang ditunjukkan dengan simbol persegi (0) dan segitiga (1). Simbol-simbol yang ada pada gambar tersebut adalah data latih. query A dan query B masing-masing adalah data uji. Data uji query A menggunakan k = 5 dan data uji query B menggunakan k = 3. Query A diklasifikasikan sebagai kelas segitiga karena tetangga yang terdekat diambil 3 data uji yang mayoritas adalah kelas segitiga. Sedangkan, query B diklasifikasikan sebagai kelas persegi yang mayoritas adalah kelas persegi.



**Gambar 2** Penentuan kelas berdasarkan tetangga terdekat [22]

2.3 Algoritma *Decision Tree*

Selain algoritma K-NN, pada penelitian ini juga digunakan algoritma Decision Tree. Algoritma Decision Tree adalah algoritma yang bekerja dengan memecah data dan membuat kondisi fitur data dalam bentuk pohon atau tree. Kemudian, *entropy* digunakan untuk menghitung seberapa acak antar kelas dalam pemecahan dan pemecahan pohon berdasarkan *information gain* tertinggi untuk fitur [23]. *Entropy* dan *information gain* masing-masing memiliki nilai antara 0 hingga 1. Pada pemecahan pohon dipilih *entropy* yang paling rendah dan *information gain* yang paling tinggi [24]. Persamaan *entropy* dan *information gain* dapat dilihat masing-masing pada persamaan (2) dan (3) [25].

$$Entropy(S) = \sum_{i=1}^{n} -p_i \ \log_2 p_i \tag{2}$$

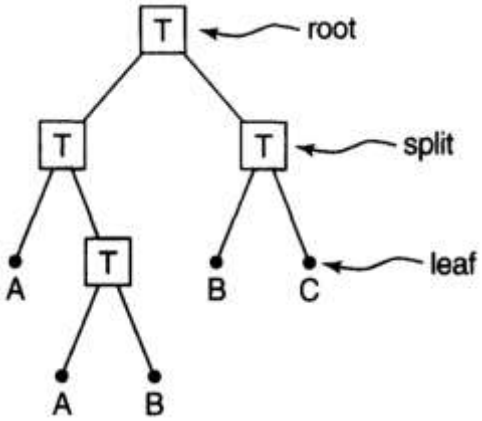
Keterangan untuk persamaan (2) adalah:
- Entropy(S) adalah nilai *entropy* sampel.
- n adalah jumlah atribut S.
- $p_i$ adalah peluang dari kelas ke-i.

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

Keterangan untuk persamaan (3) adalah:
- Gain (S, A) adalah nilai *information gain* fitur pada sampel.
- Entropy(S) adalah nilai *entropy* sampel
- n adalah jumlah subset dari atribut A.
- $S_i$ adalah jumlah sampel subset ke-i.
- S adalah jumlah sampel.
- Entropy ($S_i$) adalah nilai entropy sampel ke-i

Sesuai dengan gambar 3, decision tree bermula dari root node yang kemudian akan dilakukan pemecahan atau pembagian hingga leaf node. Data uji akan melewati kondisi-kondisi dari decision tree dari root node hingga mencapai leaf node yang akan menjadi kelas untuk data uji.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)




**Gambar 3** Decision Tree dengan pemecahan dan kelasnya (A, B, C) [26]

2.4 Evaluasi Klasifikasi

Evaluasi yang biasa digunakan untuk menilai model machine learning adalah *confusion matrix*. Hasil dari matriks tersebut menggambarkan bagaimana keakuratan model dalam memprediksi kelas dengan kelas yang sebenarnya. Pada *confusion matrix* terdapat *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). TP adalah hasil prediksi positif dan kelas sebenarnya juga positif. FP adalah hasil prediksi positif di mana kelas sebenarnya negatif. TN adalah hasil prediksi yang negatif dan kelas sebenarnya juga negatif. FN adalah hasil prediksi negatif di mana kelas sebenarnya positif. Tabel 1 menunjukkan *confusion matrix* jika terdapat 2 target kelas.

**Tabel 1** *Confusion Matrix* untuk klasifikasi biner

| | | Kelas prediksi | |
|---|---|---|---|
| | | **0** | **1** |
| Kelas sebenarnya | **0** | TN | FP |
| | **1** | FN | TP |

Perhitungan lainnya adalah precision, recall, f1-score, dan akurasi. Precision adalah persentase kelas sebenarnya yang positif dengan semua prediksi positif. Recall adalah persentase hasil prediksi positif yang benar dengan kelas sebenarnya yang positif. F1-score adalah persentase keseimbangan antara precision dan recall. Akurasi adalah persentase hasil prediksi yang benar dengan kelas sebenarnya. Persamaan untuk menghitung precision, recall, f1-score, dan akurasi masing-masing di persamaan (4), (5), (6), dan (7) [27].

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Keterangan untuk persamaan (4), (5), (6), dan (7) adalah:

- TP adalah nilai *true positive*.
- FP adalah nilai *false positive*.
- TN adalah nilai *true negative*.
- FN adalah nilai *false negative*.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Data

Data untuk penelitian ini berasal dari Gaia DR3 dan Hipparcos and Tycho. Dataset ini adalah data yang berasal dari teleskop luar angkasa yang dioperasi oleh Badan Antariksa Eropa (European Space Agency) [28], [29]. Katalog data tersebut dapat didapatkan melalui Gaia Archive dengan membuat *query* ADQL (Astronomical Data Query Language). Terdapat beberapa parameter atau kolom pada dataset. Data tersebut dapat dilihat dari *query results* yang bisa diunduh. Data yang dicari dengan *query* diunduh dengan pilihan ekstensi .csv. Gambar 4 menunjukkan tampilan halaman web untuk mencari katalog dataset. *Query* untuk dataset dapat dilihat pada lampiran.


**Gambar 4** Halaman web Gaia Archive

### 3.2 Praproses dataset, penentuan kelas, dan penentuan fitur.

Dataset yang sudah diunduh dilakukan praproses. Dataset .csv diimpor atau dibacakan ke JupyterLab dengan bantuan library pandas. Dataset berisi 60000 baris dan 11 kolom. Tidak ada data yang kosong karena *query* yang digunakan sudah mencegah munculnya data yang kosong. Beberapa kolom perlu dihapuskan karena kolom tersebut adalah ID dari record data. Keterangan dari kolom tersebut dapat diamati di tabel 2.

**Tabel 2** Detail kolom atau fitur dari dataset

| Nama Kolom | Keterangan |
|---|---|
| source_id | ID unik data Gaia DR3 |
| hip | ID unik data Hipparcos and Tycho |
| parallax | Nilai parallax (milli-arcseconds) |
| phot_g_mean_mag | Nilai magnitudo tampak (mag) |
| mg_gspphot | Nilai magnitudo absolut (mag) |
| bp_rp | Nilai magnitudo filter warna (mag) |
| bp_g | |
| g_rp | |
| distance_gspphot | Nilai jarak ke bintang (pc atau parsec) |
| spectraltype_esphs | Tipe spektral bintang data Gaia DR3 (Klasifikasi Harvard) |
| sptype | Tipe spektral bintang data Hipparcos and Tycho (Klasifikasi Harvard dan Morgan-Keenan) |

Kolom source_id, hip, spectraltype_esphs, bp_g, g_rp, dan distance_gspphot dihapus untuk mengurangi dimensi dari dataset. Penentuan dan pembuatan kelas dengan membuat fungsi Python. Fungsi Python ini membuat kolom baru dengan nama ‘class’ dan menerima kolom sptype dengan kondisi karakter ‘I’, ‘II’, ‘III’, ‘IV’, ‘V’, ‘VI’, dan ‘VII’. Seperti yang dijelaskan sebelumnya, angka romawi I, II, III, IV adalah bintang raksasa. Sedangkan angka romawi V, VI, VII adalah bintang katai. Fungsi tersebut mengembalikan angka 0 untuk bintang katai, 1 untuk bintang raksasa, dan 2 untuk yang lainnya. Setelah menjalankan fungsi tersebut, baris dengan kelas yang bernilai 2 dihapus.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Penghapusan baris dengan kelas yang bernilai 2 mengakibatkan dataset menjadi 27436 baris. Kelas bintang katai sebanyak 14530 dan kelas bintang raksasa sebanyak 12906. Kelas dalam dataset dilakukan *resample* atau diseimbangkan dengan library sklearn. Sebanyak 1624 kelas bintang katai dihapuskan agar jumlahnya sama dengan kelas bintang raksasa. Penyeimbangan ini dilakukan agar algoritma mendapatkan pembelajaran kelas yang setara. Total dataset setelah penyeimbangan sebanyak 25812.

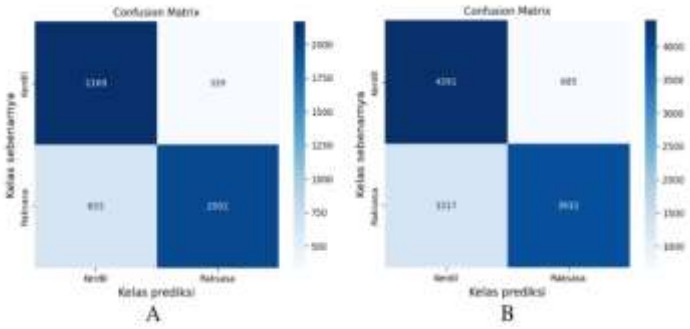
Kemudian, kolom sptype dihapus karena sudah ada kolom 'class' sebagai kelas dari dataset. Penentuan fitur dan kelas dari dataset ini adalah menginisialisasi variabel X sebagai fitur dengan kolom parallax, phot_g_mean_mag, mg_gspphot, dan bp_rp. Variabel y sebagai kelas dengan kolom 'class' di mana bintang katai bernilai 0 dan bintang raksasa bernilai 1. Selanjutnya, pembuatan algoritma klasifikasi K-NN dan Decision Tree. Tabel 3 menunjukkan tabel dataset 5 baris pertama setelah praproses, penentuan kelas, penentuan fitur, dan penyeimbangan kelas.

**Tabel 3** Potongan dataset 5 baris pertama

| parallax | phot_g_mean_mag | mg_gspphot | bp_rp | class |
|---|---|---|---|---|
| 5.737754 | 7.475046 | 0.9976 | 1.252862 | 0 |
| 1.990087 | 8.631828 | -0.0964 | 0.065264 | 0 |
| 6.701261 | 6.589071 | 0.5653 | -0.041393 | 0 |
| 8.112837 | 6.462860 | 0.5355 | 1.152445 | 0 |
| 4.310796 | 7.068493 | 0.1503 | 1.525983 | 0 |

3.3 *Algoritma K-NN*

Eksperimen dengan algoritma K-NN menggunakan k = 7. Eksperimen yang dilakukan sebanyak 10 kali. Setiap eksperimen dilakukan pengacakan Komposisi data latih dan data uji. Komposisi data latih dan data uji tersebut digunakan 80/20 dan 60/40. Pada komposisi 80/20 dengan eksperimen pertama, hasil persentase untuk akurasi, precision, recall, dan f1-score yang diperoleh adalah 81%. *Confusion matrix* eksperimen pertama juga menunjukkan hasil yang cukup baik. Namun, banyak prediksi untuk bintang raksasa mengklasifikasikannya sebagai bintang kerdil. Kesalahan ini lebih banyak dari pada memprediksi kelas kerdil di mana sebanyak 339 salah prediksi. Kemudian, pada eksperimen pertama dengan komposisi dataset 60/40, hasil persentase sama seperti eksperimen pertama dengan komposisi dataset 80/20. Hasil eksperimen pertama persentase akurasi, precision, recall, dan f1-score yang diperoleh adalah 81%. *Confusion matrix* pada eksperimen ini juga menunjukkan hasil yang mirip dengan eksperimen komposisi 80/20 di mana semua nilai TP, FP, FN, TN naik sekitar 2 kali lipat. Naiknya nilai tersebut karena lebih banyak data uji yang digunakan dari pada komposisi 80/20. Kedua *confusion matrix* tersebut ditunjukkan pada gambar 5.




**Gambar 5** *Confusion matrix* eksperimen pertama K-NN 80/20 (A) dan 60/40 (B)

Berdasarkan eksperimen komposisi K-NN yang dilakukan (tabel 4), akurasi terbaik ada di komposisi 60/40 dengan 81.1%. Namun, nilai akurasi tersebut tidak signifikan dari komposisi 80/20

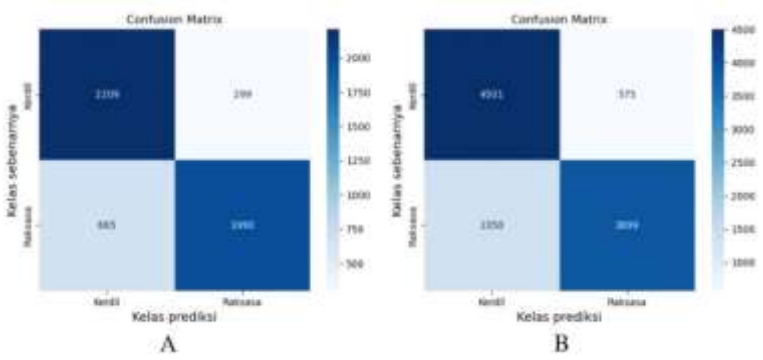
dengan 80.9%. Hal yang sama juga pada evaluasi lainnya. Persentase yang didapatkan tidak signifikan atau jauh. Dapat disimpulkan bahwa kedua komposisi tidak mempengaruhi akurasi, precision, recall, dan f1-score.

**Tabel 4** Nilai rata-rata evaluasi K-NN dalam 10 eksperimen

| Data latih | Data uji | Nilai rata-rata | | | |
|---|---|---|---|---|---|
| | | Akurasi | Precision | Recall | F1-Score |
| 80% | 20% | 80.9% | 81.5% | 81% | 80.9% |
| 60% | 40% | 81.1% | 81.4% | 81.1% | 81.1% |

3.4 Algoritma Decision Tree

Eksperimen dengan algoritma Decision Tree dilakukan dengan menggunakan parameter criterion *entropy* untuk sklearn dan untuk max_depth dengan maksimal 5 tingkat kedalaman pohon. Hasil eksperimen pertama komposisi 80/20 dengan persentase untuk akurasi, precision, recall, dan f1-score masing-masing adalah 81%, 82%, 82%, dan 81%. *Confusion matrix* yang ditampilkan pada eksperimen sangat mirip dengan *confusion matrix* K-NN. Dapat diamati bahwa model memiliki banyak kesalahan ketika memprediksi kelas bintang raksasa sebenarnya dan memberikan label bintang kerdil dari pada saat memprediksi kelas bintang kerdil sebenarnya. Hasil eksperimen pertama berikutnya adalah komposisi 60/40. Hasil persentase untuk akurasi, precision, recall, dan f1-score masing-masing adalah 81%, 82%, 81%, dan 81%. Hasil tersebut hanya berbeda sedikit dan tidak signifikan dengan komposisi 80/20. *Confusion matrix* ini juga mirip seperti *confusion matrix* komposisi 80/20, hanya saja dengan 2 kali lipat nilai-nilai *confusion matrix* karena lebih banyak data uji yang digunakan. Kedua *confusion matrix* decision tree ini ditunjukkan pada gambar 6.




A B

**Gambar 6** *Confusion matrix* eksperimen pertama Decision Tree 80/20 (A) dan 60/40 (B)

Berdasarkan eksperimen komposisi dataset yang dilakukan dengan decision tree (tabel 5), kedua nilai akurasi sama, yaitu 81.5%. Nilai evaluasi lainnya seperti precision, recall, dan f1-score hanya berbeda 1% untuk kedua komposisi. Dapat disimpulkan bahwa komposisi 80/20 dan 60/20 tidak mempengaruhi nilai evaluasi algoritma untuk klasifikasi bintang katai dan bintang raksasa. Perbedaan persentase tersebut tidak signifikan.

**Tabel 5** Nilai rata-rata evaluasi Decision Tree dalam 10 eksperimen

| Data latih | Data uji | Nilai rata-rata | | | |
|---|---|---|---|---|---|
| | | Akurasi | Precision | Recall | F1-Score |
| 80% | 20% | 81.5% | 81.9% | 81.6% | 81.5% |
| 60% | 40% | 81.5% | 82% | 81.5% | 81.4% |

3.5 Perbandingan kinerja *K-NN dan Decision Tree*

Kinerja algoritma K-NN dan Decision Tree dengan komposisi 80/20 menunjukkan bahwa akurasi yang unggul adalah akurasi Decision Tree dengan 81.5% dari pada akurasi K-NN dengan 80.9% (tabel 6). Walaupun tidak signifikan, perbedaan ini menunjukkan bahwa Decision Tree lebih akurat dalam klasifikasi bintang katai dan bintang raksasa. Nilai rata-rata evaluasi lainnya

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

menunjukkan keunggulan dari pada algoritma K-NN. Seperti keseimbangan f1-score dari decision tree yang lebih unggul dari K-NN. Dapat dikatakan bahwa decision tree menjadi algoritma yang unggul saat klasifikasi bintang katai dan bintang raksasa dengan komposisi 80/20.

**Tabel 6** Nilai rata-rata evaluasi dengan komposisi 80/20

| **Algoritma** | **Nilai rata-rata** | | | |
|---|---|---|---|---|
| | Akurasi | Precision | Recall | F1-Score |
| K-NN | 80.9% | 81.5% | 81% | 80.9% |
| Decision Tree | 81.5% | 81.9% | 81.6% | 81.5% |

Kinerja algoritma Decision Tree masih lebih unggul dari pada K-NN dalam komposisi 60/20. Nilai rata-rata akurasi Decision Tree yang didapatkan dari 10 eksperimen adalah 81.5%. Nilai rata-rata evaluasi lainnya juga menunjukkan keunggulan. Pada komposisi 60/40, Decision Tree lebih unggul dari pada K-NN saat klasifikasi bintang katai dan bintang raksasa dengan dataset Gaia DR3 dan Hipparcos and Tycho.

**Tabel 7** Nilai rata-rata evaluasi dengan komposisi 60/40

| **Algoritma** | **Nilai rata-rata** | | | |
|---|---|---|---|---|
| | Akurasi | Precision | Recall | F1-Score |
| K-NN | 81.1% | 81.4% | 81.1% | 81.1% |
| Decision Tree | 81.5% | 82% | 81.5% | 81.4% |

Jika diamati pada hasil nilai rata-rata pada kedua komposisi. Hasil kinerja algoritma K-NN dan Decision Tree pada kedua komposisi mirip. Dengan ukuran dataset sekitar 25 ribu baris atau record, komposisi 80/20 dan 60/20 tidak memberikan perbedaan yang signifikan. Namun, Decision Tree masih unggul di kedua komposisi tersebut. Ini berarti algoritma Decision Tree lebih baik digunakan ketika ukuran dataset besar.

## 4. KESIMPULAN

Melihat banyaknya bintang ataupun objek langit dan banyaknya bintang yang ditemukan melalui teleskop. Keperluan untuk klasifikasi digunakan untuk mengelompokkan jenis-jenis bintang. Namun, banyaknya bintang akan membuat klasifikasi dengan manual memakan waktu. Penggunaan metode machine learning dapat membantu klasifikasi bintang sesuai dengan jenisnya yaitu, katai dan raksasa. Dengan ukuran dataset sebesar 25812 baris atau record. Hasil dari penelitian ini adalah dengan menggunakan algoritma Decision Tree menghasilkan nilai rata-rata akurasi dalam komposisi 80% data latih dan 20% data uji (80/20) sebesar 81.5%. Nilai rata-rata akurasi tersebut juga sama pada komposisi 60% data latih dan 40% data uji (60/40). Kedua hasil ini lebih unggul tetapi tidak signifikan dari pada menggunakan algoritma K-NN yang menghasilkan nilai rata-rata akurasi untuk komposisi 80/20 sebesar 80.9% dan 81.1% untuk komposisi 60/40. Saran untuk pengembangan penelitian selanjutnya adalah menggunakan metode machine learning lainnya seperti, Neural Network. Kemudian, saran lainnya adalah menggunakan dataset yang lebih kecil dan mengklasifikasi berdasarkan kelas dari klasifikasi Harvard.

## UCAPAN TERIMA KASIH

Penelitian ini menggunakan data dari misi Gaia milik Badan Antariksa Eropa (ESA) (https://www.cosmos.esa.int/gaia), yang diolah oleh Konsorsium Pengolahan dan Analisis Data Gaia (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Pendanaan untuk DPAC disediakan oleh lembaga-lembaga nasional, khususnya lembaga-lembaga yang berpartisipasi dalam Perjanjian Multilateral Gaia.

# DAFTAR PUSTAKA

[1] A. A. Bustari, R. B. Pratama dan C. J. Al-Aryachiyah, "Analisis Tingkat Polusi Cahaya Berdasarkan Light Pollution Map Dan Implikasinya Terhadap Pengamatan Astronomi Indonesia," dalam *Kalijaga Innovation and Research Technologies Conference 7*, Sleman, 2022.

[2] A. M. Rahmadhan, A. Marlina dan U. Mustaqimah, "KONSEP PENANGGULANGAN POLUSI CAHAYA PADA OBSERVATORIUM BINTANG DI KABUPATEN KARANGANYAR," *Senthong*, vol. 6, no. 3, pp. 1045-1054, 2023.

[3] P. Protte dan S. M. Hoffmann, "Accuracy of magnitudes in pre-telescopic star catalogs," *Astronomische Nachrichten*, pp. 827-840, 2020.

[4] Sutrisno dan T. J. R. Fadilla, "Identification and Determining The Pseudo Magnitude Of Meissa Star Based on Observation Data From The Eastern Sky In The Mentigen Hills of The Pseudo Magnitude of Meissa Star Based on Observation Data From The Eastern Sky," dalam *The 10th National Physics Seminar (SNF 2021)*, Jakarta, 2021.

[5] S. A. A. Albakri, M. N. A. Hussien dan H. Herdan, "Measurement of the distance to the central stars of Nebulae by using Expansion methods with Alladin Sky Atlas," dalam *1st International Conference in Physical Science and Advance Materials*, Istanbul, 2020.

[6] M. S. Devetaković, Đ. D. Đorđević, G. D. Đukanović, A. D. Krstić-Furundžić, B. S. Sudimac dan A. Scognamiglio, "Design of Solar Systems for Buildings and Use of BIM Tools: Overview of Relevant Geometric Aspects," *FME Transactions*, vol. 47, no. 2, pp. 387-397, 2019.

[7] J. McKinney, L. Armus, A. Pope, T. Díaz-Santos, V. Charmandaris, H. Inami, Y. Song dan A. S. Evans, "Regulating Star Formation in Nearby Dusty Galaxies: Low Photoelectric Efficiencies in the Most Compact Systems," *The Astrophysical Journal*, vol. 908, no. 2, pp. 1-11, 2021.

[8] A. Serenelli, R. D. Rohrmann dan M. Fukugita, "Nature of blackbody stars," *Astronomy & Astrophysics (A&A)*, vol. 623, pp. 1-7, 2019.

[9] T. S. Kuhn, Black-Body Theory and the Quantum Discontinuity, 1894-1912, Chicago: University of Chicago Press, 1987.

[10] H. Zhang, Z. Huang, M. Ding, Q. Wang, Y. Feng, Z. Li, S. Wang, L. Yang, S. Chen, W. Shang, J. Zhang, T. Deng, H. Xu dan K. Cui, "A photon-recycling incandescent lighting device," *Science Advances*, vol. 9, no. 15, pp. 1-8, 2023.

[11] M. Surace, E. Zackrisson, D. J. Whalen, T. Hartwig, S. C. O. Glover, T. E. Woods and A. Heger, "On the detection of supermassive primordial stars – II. Blue supergiants," *Monthly Notices of the Royal Astronomical Society*, vol. 488, no. 3, pp. 3995-4003, 2019.

[12] D. R. B. dan M. Badi, "Non-Contact Temperature Measurement Applicable for Covid-19," *Journal of Advancement in Electronics Design*, vol. 5, no. 2, pp. 1-14, 2022.

[13] PhET - University of Colorado Boulder, "Blackbody Spectrum," [Online]. Available: https://phet.colorado.edu/sims/html/blackbody-spectrum/latest/blackbody-spectrum_all.html. [Diakses 1 Mei 2024].

[14] D. Deng, Y. Sun, M. Jian, B. Jiang dan H. Yuan, "Intrinsic Color Indices of Early-type Dwarf Stars," *The Astronomical Journal*, vol. 159, no. 5, pp. 1-11, 2020.

[15] J. Airey dan U. Eriksson, "Unpacking the Hertzsprung-Russell Diagram: A Social Semiotic Analysis of the Disciplinary and Pedagogical Affordances of a Central Resource in Astronomy," *Designs for Learning*, vol. 11, no. 1, pp. 99-107, 2019.

[16] C. Dafonte, A. Rodríguez, M. Manteiga, Á. Gómez dan B. Arcay, " A Blended Artificial Intelligence Approach for Spectral Classification of Stars in Massive Astronomical Surveys," *Entropy*, vol. 22, no. 5, pp. 1-26, 2020.

[17] S. J. Dick, "Astronomy's Three Kingdom System: A Comprehensive Classification System of Celestial Objects," *Knowledge Organization*, vol. 46, no. 6, pp. 460-466, 2019.

[18] G. Banerjee, "Spectroscopy: The Tool to Study the Stars," *Mapana – Journal of Sciences*, vol. 20, no. 4, pp. 9-31, 2021.

[19] D. Huppenkothen, J. Pampin, J. R. A. Davenport dan J. Wenlock, "The Sonified Hertzsprung-Russell Diagram," dalam *28th ICAD 2023 Conference*, Norrköping, 2023.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[20] P. Cunningham dan S. J. Delany, "k-Nearest Neighbour Classifiers - A Tutorial," *ACM Computing Surveys,* vol. 54, no. 6, pp. 1-25, 2021.

[21] N. Ali, D. Neagu dan P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Applied Sciences,* vol. 1, no. 1559, 2019.

[22] S. Uddin, I. Haque, H. Lu, M. A. Moni dan E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports,* vol. 12, no. 6256, pp. 1-12, 2022.

[23] I. D. Mienye, S. Y. dan Z. Wang, "Prediction performance of improved decision tree-based algorithms: a review," dalam *2nd International Conference on Sustainable Materials Processing and Manufacturing (SMPM 2019),* 2019.

[24] I. Setiawati, A. Permana dan A. Hermawan, "IMPLEMENTASI DECISION TREE UNTUK MENDIAGNOSIS PENYAKIT LIVER," *Journal of Information System Management (JOISM),* vol. 1, no. 1, pp. 13-17, 2019.

[25] A. Izyuddin dan S. Wibisono, "APLIKASI PREDIKSI PENJUALAN ACMENGGUNAKAN DECISION TREE DENGAN ALGORITMA C4.5," *Jurnal Manajemen Informatika dan Sistem Informasi (MISI),* vol. 3, no. 2, pp. 146-156, 2020.

[26] M. A. Friedl dan C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment,* vol. 61, no. 3, pp. 399-409, 1997.

[27] I. M. K. Karo dan Hendriyana, "KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN ALGORITMA MACHINE LEARNING DAN Z-SCORE," *Jurnal Teknologi Terpadu,* vol. 8, no. 2, pp. 94-99, 2022.

[28] Gaia Collaboration, T. Prusti, J.H.J. de Bruijne, et al., "The Gaia mission," *A&A 595,* pp. A1, 2016b.

[29] Gaia Collaboration, A. Vallenari, A. G. A. Brown, et al., "Gaia Data Release 3. Summary of the content and survey properties," *A&A 674,* pp. A1, 2023j.