

Klasifikasi Bintang RR Lyrae / Cepheid / Mira dari *The All-Sky Automated Survey for Supernovae* Menggunakan Algoritma Naïve Bayes

Ayu Lestari^{*1}, Nur Hasanah¹, Eko Juarlin¹

¹Program Studi Fisika, Fakultas MIPA, Universitas Hasanuddin

E-mail: ^{*}¹ayulestarimemang@gmail.com

ABSTRAK

Penelitian ini menggunakan 58.423 data dari *the All-Sky Automated Survey for Supernovae* (ASAS-SN) untuk melakukan klasifikasi bintang RR Lyrae, Cepheid, dan Mira menggunakan pendekatan *machine learning*. Terdapat sembilan kolom yang dijadikan atribut dalam pembuatan model *machine learning*, yaitu: *raj2000*, *dej2000*, *l*, *b*, *mean_vmag*, *amplitude*, *period*, *lksl_statistic*, dan *parallax* dengan kolom *variable_type* digunakan sebagai target label. Dengan memanfaatkan *training dataset* (data latih) dan *testing dataset* (data uji), algoritma Naïve Bayes yang digunakan pada penelitian ini menghasilkan akurasi sebesar 98.6%. Sedangkan berdasarkan hasil evaluasi menggunakan *confusion matrix*, diperoleh presisi dari bintang RR Lyrae, Cepheid, dan Mira masing-masing sebesar 99%, 87%, dan 99%. *Recall* dari ketiga objek masing-masing adalah 99%, 88%, dan 99%, sedangkan nilai *f1-score* masing-masing sebesar 98%, 90%, dan 100%. Kesimpulan dari penelitian ini adalah algoritma Naïve Bayes dapat digunakan dalam klasifikasi objek astronomi dengan tingkat akurasi yang baik.

Kata kunci : *Cepheid, Mira, Naïve Bayes, RR Lyrae.*

ABSTRACT

We used 58.423 data from the All-Sky Automated Survey for Supernovae (ASAS-SN) to classify RR Lyrae, Cepheid, and Mira stars using a machine learning approximation. Nine columns were used as attributes in making machine learning models, namely: *raj2000*, *dej2000*, *l*, *b*, *mean_vmag*, *amplitude*, *period*, *lksl_statistic*, and *parallax* with the *variable_type* column used as the target label. The method used in this research was Naïve Bayes algorithm, which utilized training dataset and testing dataset. This model produced an accuracy of 98.6%. Based on the evaluation results using a confusion matrix, the precision of the stars RR Lyrae, Cepheid, and Mira were 99%, 87%, and 99%, respectively. The recall of the three objects were 98%, 90%, and 100%, respectively, while the *f1-score* values are 99%, 88%, and 99%, respectively. This research concluded that the Naïve Bayes algorithm can be used in the classification of astronomical objects with a good level of accuracy.

Keywords: *Cepheid, Mira, Naïve Bayes, RR Lyrae.*

PENDAHULUAN

Bintang variabel merupakan jenis bintang yang kecerahannya berubah seiring waktu. Bintang ini digunakan untuk mempelajari populasi bintang dan struktur galaksi. Salah satu contoh bintang variabel adalah bintang berdenyut (*pulsating stars*) yang mengembang dan berkontraksi secara berirama. Periode denyutannya berkisar dari detik (katai putih) hingga tahun (raksasa super) (Percy, 2014). Bintang variabel berdenyut sendiri terbagi menjadi beberapa jenis bintang, antara lain RR Lyrae, Cepheid dan Mira.

Dalam beberapa dekade terakhir, teleskop telah berkembang sangat pesat. Peningkatan data astronomi terjadi setiap tahun secara eksponensial, sehingga menjadi mustahil untuk semua data diproses dan dianalisis secara manual. Oleh karena itu, diperlukan sebuah metode dalam analisis dan eksplorasi data yang efektif dan akurat. Solusi yang paling mungkin adalah melibatkan algoritma *machine learning*.

Penerapan *machine learning* dalam bidang astronomi telah banyak dilakukan, seperti klasifikasi galaksi, kuasar dan bintang (Clarke *et al.*, 2020), morfologi galaksi (Martin

et al., 2019), bahkan penyusunan katalog ASAS-SN bintang variabel (Jayasinghe *et al.*, 2020) yang memanfaatkan *machine learning* dalam proses klasifikasi jenis bintang variabel yang berhasil diamati. Klasifikasi dilakukan menggunakan metode *random forest* dengan nilai *f1-score* sebesar 98.5%. Pemanfaatan *machine learning* dengan metode klasifikasi yang berbeda juga diterapkan oleh Čokina *et al.*, (2021). Mereka menggunakan metode *deep learning* dengan algoritma *convolutional neural network* dalam klasifikasi bintang *eclipsing binary* dalam mengkategorikan kurva cahayanya menjadi dua kelas, yaitu *detached binary stars* dan *over-contact binary stars*. Metode ini berhasil mencapai tingkat akurasi sebesar 98%. Terdapat beberapa metode klasifikasi lainnya yang telah dilakukan pada penelitian sebelumnya, yaitu *K-Nearest Neighbor* (KNN), *Decision Tree* (DT), dan *Support Vector Machine* (SVM) (Xiao-Qing dan Jin-Meng, 2020).

Berdasarkan hal yang telah dijelaskan diatas, tujuan dari pekerjaan ini adalah melakukan klasifikasi bintang RR Lyrae, Cepheid, dan Mira menggunakan salah satu

metode *machine learning* yang lain, yaitu metode Naïve Bayes yang menerapkan prinsip probabilistik dalam prosesnya.

METODE PENELITIAN

Data bintang RR Lyrae, Cepheid dan Mira yang digunakan dalam penelitian ini diekstraksi dari database bintang variabel *The All-Sky Automated Survey for Supernovae* (ASAS-SN) (Jayasinghe, 2021). Data terdiri dari 59.072 pengamatan ruang yang terbagi menjadi 82 kolom. Secara garis besar kolom-kolom tersebut memuat keterangan dalam pengamatan, seperti penamaan objek, sistem koordinat, magnitudo fotometri dalam pita V optik sebesar 500 hingga 600 nm, besaran amplitudo dan periode objek, serta sejumlah keterangan dalam proses klasifikasi dalam katalog ASAS-SN.

Data yang telah dikumpulkan harus melalui tahap pengolahan awal data, karena tidak semua data dapat digunakan. Pada dataset ini, pembersihan data dilakukan dengan menghilangkan kolom yang tidak akan mempengaruhi proses klasifikasi dan data yang tidak lengkap (*missing value*). Data yang diambil adalah 58.423. Terdapat sembilan kolom yang akan dijadikan sebagai atribut dalam pembuatan model *machine learning* yang terlihat pada Tabel 1. dengan menjadikan kolom “variable_type” yang berisi tiga jenis bintang variabel, yaitu RR Lyrae, Cepheid dan Mira sebagai target label.

Tabel 1. Atribut dalam dataset ASAS-SN

Atribut	Keterangan
raj2000 (<i>right ascension</i>)	Mengukur timur dan barat pada bola langit, sedangkan J menunjukkan masa Julian dengan mengacu pada 1 Januari 2000
dej2000 (<i>declination</i>)	Mengukur utara dan selatan pada bola langit
l (<i>galactic longitude</i>)	Mengukur jarak sudut suatu objek ke arah timur sepanjang ekuator galaksi dari pusat galaksi
b (<i>galactic latitude</i>)	Mengukur sudut suatu objek ke utara ekuator galaksi (atau bidang tengah) seperti yang dilihat dari Bumi
mean_vmag (<i>mean visible spectrum</i>)	Magnitudo fotometri dalam pita V optik antara 500 dan 600 nm (VizieR)
amplitude	Jarak maksimal dari pusat gerak
period	Waktu untuk satu siklus gerak lengkap
lksl statistic	Ukuran hamburan dalam kurva

(<i>Lafler-Kinman String Length</i>)	cahaya. Kurva cahaya dengan statistik LKSL yang lebih kecil akan memiliki pencar yang lebih rendah (Jayasinghe, 2021)
parallax	Mengukur pergerakan semu bintang dengan latar belakang bintang yang lebih jauh.

Proses selanjutnya adalah melakukan transformasi terhadap nama dalam target label. Tabel 2. menunjukkan hasil transformasi yang dilakukan pada setiap objek dalam dataset.

Tabel 2. Transformasi data jenis objek

Atribut	Transformasi
RRAB, RRC, RRD	RR LYRAE
CWA, CWB, DCEP, DCEPS, RVA	CEPHEID
M	MIRA

Selanjutnya adalah mencari *insights* dari dataset dengan visualisasi data. Metode yang digunakan dalam melihat sebaran data dalam dataset adalah *Kernel Density Estimation* (KDE). Metode ini merupakan suatu pendekatan statistika non-parametrik untuk mengestimasi fungsi distribusi probabilitas dari suatu variabel acak jika diasumsikan bentuk atau model distribusi dari variabel acak tersebut tidak diketahui (Setiawan, 2014). Hasil visualisasi data akan diolah untuk menerapkan metode Naïve Bayes untuk mengklasifikasikan bintang RR Lyrae, Cepheid, dan Mira.

Dataset kemudian dibagi menjadi dua, yaitu data latih (*training dataset*) dan data uji (*testing dataset*). Data latih digunakan untuk membuat sebuah model yang merupakan representasi pengetahuan yang akan digunakan untuk melakukan prediksi kelas data baru yang belum pernah ada. Sedangkan data uji digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Perbandingan pembagian data yang dilakukan sebesar 70:30. Data latih harus lebih besar untuk mengekstrak informasi yang cukup.

Selanjutnya diterapkan metode Naïve Bayes. Metode ini adalah pengklasifikasi probabilistik berdasarkan teorema Bayes. Diasumsikan bahwa kehadiran suatu atribut tidak tergantung pada kehadiran atribut lain yang diberikan suatu kelas (Sun *et al.*, 2018). Alur dari metode Naïve Bayes dijelaskan pada Pers. (2.1) (Bustami, 2014).

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X : Data dengan kelas yang belum

	diketahui
H	: Hipotesis data X merupakan suatu kelas spesifik
$P(H X)$	: Probabilitas hipotesis H berdasarkan kondisi X (<i>posterior probability</i>)
$P(H)$	: Probabilitas hipotesis H (<i>prior probability</i>)
$P(X H)$	: Probabilitas hipotesis X berdasarkan kondisi H
$P(X)$	: Probabilitas hipotesis X

Pertama, dihitung nilai probabilitas setiap kelasnya (*prior probability*). Selanjutnya, klasifikasi akan dilakukan menggunakan metode Gaussian Naïve Bayes. Hal ini dikarenakan data ASAS-SN bersifat kontinu. Artinya, setiap atribut memiliki nilai yang berbeda dalam kumpulan data dan tidak dapat merepresentasikan kemunculannya. Metode ini menggunakan fungsi distribusi Gaussian (distribusi normal) untuk memperkirakan parameter pada probabilitas *likelihood* pada metode Naïve Bayes. Adapun fungsi distribusi Gaussian secara matematis dapat dituliskan:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.2)$$

dimana, μ = nilai rata-rata (*mean*) dan σ = standar deviasi. Adapun persamaan untuk mencari nilai rata-rata adalah sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (2.3)$$

dimana, μ = nilai rata-rata (*mean*), x_i = nilai x ke- i dan n = jumlah sampel. Selanjutnya dihitung nilai standar deviasi untuk setiap atribut berdasarkan kelasnya. Nilai standar deviasi dapat dihitung menggunakan Pers. (2.4).

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (2.4)$$

dimana, σ = standar deviasi, μ = nilai rata-rata (*mean*), x_i = nilai x ke- i dan n = jumlah sampel. Kemudian, *likelihood* semua kelas untuk mendapatkan probabilitas *class-assignment*.

$$P_{\text{assignment}}(c_i) = \frac{\text{likelihood}(c_i)}{\sum_{c_j \in C} \text{likelihood}(c_j)} \quad (2.5)$$

Dalam menghitung nilai probabilitas akhir dari suatu kejadian (*posterior probability*), dipilih kelas dengan probabilitas tertinggi (Putra, 2020).

Pada tahap evaluasi performa model menggunakan *confusion matrix*. *Confusion matrix* berisi tiga bagian informasi: presisi, *recall*, *f1-score*. Presisi merupakan proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya (Rosandy, 2016).

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2.6)$$

Recall adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar (Rosandy, 2016). *Recall* yang rendah untuk kelas individu akan menunjukkan bahwa kelas tersebut sering salah diklasifikasikan sebagai kelas lain (Čokina *et al.*, 2021).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.7)$$

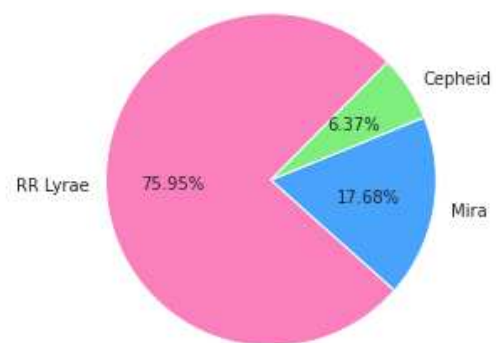
F1-score adalah rata-rata harmonik presisi dan *recall* dan digunakan sebagai metrik kinerja secara keseluruhan.

$$F_1 - \text{score} = \frac{2TP}{2TP+FP+FN} \quad (2.8)$$

HASIL DAN PEMBAHASAN

Visualisasi Data

Berdasarkan hasil dari visualisasi data, didapatkan informasi bahwa objek bintang variabel terbanyak dari dataset adalah RR Lyrae sebesar 75.95%, kemudian Mira dan Cepheid yang masing-masing 17.68% dan 6.37%. Persentase ini juga menunjukkan nilai *prior probability* dari masing-masing kategori. *Prior probability* merupakan nilai probabilitas kemunculan setiap kelas pada dataset. Visualisasi data dapat dilihat pada Gambar 1.

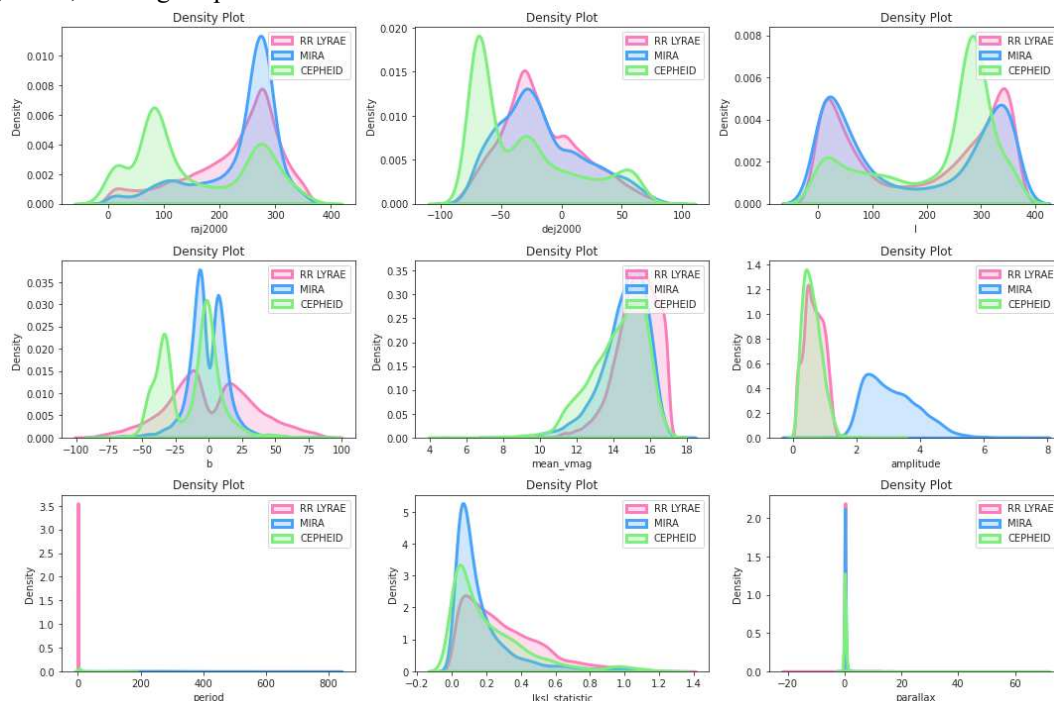


Gambar 1. Persentase RR Lyrae, Cepheid dan Mira dalam dataset

Selanjutnya dicari visualisasi distribusi densitas setiap atribut dalam dataset yang dapat ditunjukkan pada Gambar 2. Sebagai contoh untuk grafik atribut asensio rekta (kiri atas)

terlihat bahwa bintang Mira dan RR Lyrae paling banyak berada pada posisi asensio rekta sekitar 280° . Sedangkan untuk bintang Cepheid memiliki dua puncak, yaitu pada posisi asensio rekta sekitar 80° dan 280° . Terdapat pula grafik mengenai amplitudo bintang yang diukur berdasarkan tingkat kecerahan bintang, sehingga dinyatakan dalam magnitudo. Amplitudo bintang RR Lyrae berkisar antara 0 hingga 1.8 magnitudo, bintang Cepheid berkisar antara 0

hingga 3.8 magnitudo, sedangkan bintang Mira memiliki amplitudo 1.8 hingga 8 magnitudo.



Gambar 2. Distribusi densitas untuk setiap atribut

Perhitungan Naïve Bayes

Berikut adalah contoh dari data latih dan data uji dari masing-masing jenis bintang yang digunakan dalam pembuatan model *machine learning*. Pada Tabel 3. menunjukkan sampel dari data latih sedangkan Tabel 4. menunjukkan sampel dari data uji. Akan dilakukan perhitungan Naïve Bayes terhadap sampel data uji.

Tabel 3. Contoh data latih

Atribut	Target label		
	RR LYRAE	CEPHEID	MIRA
raj2000	0.015	1.212	0.684
dej2000	35.363	50.235	48.714
l	111.261	115.347	114.708
b	-26.342	-11.949	-13.378
mean vmag	12.97	11.44	13.61
amplitude	0.78	0.36	4.13
period	0.707	9.213	396.38
lksl statistic	0.01	1.1	0.14
parallax	0.236	0.18	0.326

Tabel 4. Contoh data uji

Atribut	Nilai	Target label
raj2000	0.017	?
dej2000	18.407	
l	106.229	
b	-42.793	
mean vmag	13.81	
amplitude	0.31	
period	0.485	
lksl statistic	0.09	
parallax	0.114	

Perhitungan *likelihood* didapatkan menggunakan distribusi Gaussian (normal). Setelah data dibagi berdasarkan jenis bintangnya kemudian dilakukan perhitungan nilai rata-rata (*mean*) dan standar deviasi untuk setiap atributnya. Nilai rata-rata dan standar deviasi untuk setiap kasus dapat dilihat pada Tabel 5.

Tabel 5. Nilai rata-rata dan standar deviasi

Jenis Bintang	Atribut	Rata-Rata	Standar Deviasi
---------------	---------	-----------	-----------------

RR LYRAE	raj2000	225.801	82.958
	dej2000	-16.492	33.831
	l	190.06	134.694
	b	1.217	32.720
	mean vmag	15.248	1.119
	amplitude	0.662	0.300
	period	0.506	0.135
	lksl statistic	0.301	0.232
	parallax	0.17	0.408
CEPHEID	raj2000	155.528	103.63
	dej2000	-31.036	42.651
	l	217.093	110.264
	b	-13.716	21.167
	mean vmag	14.223	1.496
	amplitude	0.605	0.307
	period	11.233	18.754
	lksl statistic	0.22	0.236
	parallax	0.188	1.444
MIRA	raj2000	236.688	74.557
	dej2000	-16.9	35.718
	l	175.512	138.775
	b	-1.165	15.575
	mean vmag	14.6	1.27
	amplitude	3.109	0.848
	period	267.458	85.273
	lksl statistic	0.174	0.188
	parallax	0.241	0.342

Kedua kuantitas ini selanjutnya digunakan untuk menghitung nilai distribusi Gaussian merujuk pada Pers. (2.2). Perhitungan dilakukan dengan memasukkan data uji berdasarkan atribut yang sama dengan atribut nilai rata-rata dan standar deviasinya untuk setiap kelasnya. Sehingga setiap atribut memiliki nilai distribusinya tersendiri. Nilai distribusi Gaussian untuk masing-masing data uji dapat dilihat pada Tabel 6.

Tabel 6. Nilai distribusi Gaussian

Jenis Bintang	Atribut	Rata-Rata
RR LYRAE	raj2000	0.00012
	dej2000	0.00693
	l	0.00244
	b	0.00493
	mean vmag	0.15613
	amplitude	0.66810
	period	2.92003
	lksl statistic	1.13712
	parallax	1.13712
CEPHEID	raj2000	0.00125
	dej2000	0.00478
	l	0.00218
	b	0.00734
	mean vmag	0.25670
	amplitude	0.81897
	period	0.01805

MIRA	lksl statistic	1.45247
	parallax	0.27591
	raj2000	0.00003
	dej2000	0.00685
	l	0.00254
	b	0.00072
	mean vmag	0.25887
	amplitude	0.00203
	period	0.00003
	lksl statistic	1.92044
	parallax	1.92044

Nilai *likelihood* kemudian didapatkan dengan mengalikan nilai probabilitas dan nilai distribusi Gaussian. Perhitungan ini dikelompokkan berdasarkan jenis kelasnya. Untuk *likelihood* RR Lyrae didapatkan dengan mengalikan nilai probabilitas RR Lyrae dan nilai distribusi Gaussian setiap atribut pada kelas RR Lyrae. Hal yang sama juga dilakukan untuk mendapatkan nilai *likelihood* Cepheid dan Mira. Nilai *likelihood* untuk data uji dapat dilihat pada Tabel 7.

Tabel 7. Nilai *likelihood*

Jenis kelas	<i>Likelihood</i>
RR LYRAE	2.96E-12
CEPHEID	2.57E-14
MIRA	1.86E-21

Nilai *likelihood* masing-masing bintang selanjutnya digunakan untuk mencari nilai *P_{assignment}* merujuk pada Pers. (2.5), maka didapatkan:

Tabel 7. Nilai *P_{assignment}*

Jenis kelas	<i>P_{assignment}</i>
RR LYRAE	9.91E-01
CEPHEID	8.61E-03
MIRA	6.23E-10

Sehingga dapat diketahui bahwa data uji masuk ke dalam kelompok RR Lyrae karena *P_{assignment}* RR Lyrae lebih tinggi dibanding yang lain, yaitu sebesar 9.91E-01.

Klasifikasi Bintang RR Lyrae / Cepheid / Mira

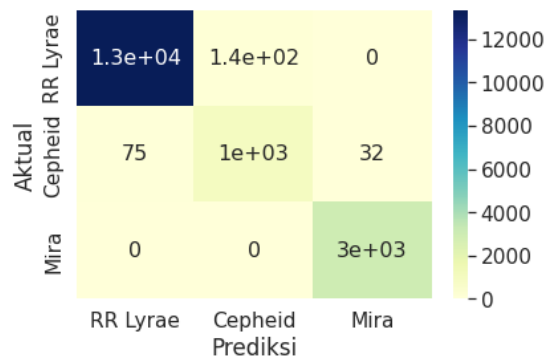
Setelah menerapkan metode Naïve Bayes pada data latih dan data uji kemudian dilakukan evaluasi menggunakan *confusion matrix* seperti yang terlihat pada Tabel 8.

Tabel 8. Hasil evaluasi menggunakan *confusion matrix*

Jenis Bintang	Presisi	Recall	F1-score
RR Lyrae	0.99	0.98	0.99

Cepheid	0.87	0.90	0.88
Mira	0.99	1.00	0.99

Model ini menghasilkan akurasi sebesar 98.6%. *Confusion matrix* terlihat pada Gambar 3.



Gambar 3. *Confusion matrix*

Berdasarkan *confusion matrix*, terdapat salah prediksi pada bintang RR Lyrae dan Cepheid. Bintang RR Lyrae yang terprediksi sebagai RR Lyrae sebanyak 13354, terprediksi sebagai Cepheid sebanyak 145 dan tidak ada yang terprediksi sebagai Mira. Untuk bintang Cepheid yang terprediksi sebagai Cepheid sebanyak 1013, sedangkan yang terprediksi sebagai RR Lyrae dan Mira masing-masing sebanyak 75 dan 32. Sedangkan untuk Mira semuanya terprediksi sebagai Mira.

KESIMPULAN

Algoritma Naïve Bayes dapat digunakan dalam memprediksi jenis objek astronomi. Metode ini adalah pengklasifikasi probabilistik berdasarkan teorema Bayes. Diasumsikan bahwa kehadiran suatu atribut tidak tergantung pada kehadiran atribut lain yang diberikan suatu kelas. Model ini menghasilkan akurasi sebesar 98.6%. Berdasarkan hasil evaluasi menggunakan *confusion matrix*, presisi dari bintang RR Lyrae, Cepheid dan Mira masing-masing sebesar 99%, 87% dan 99%. *Recall* dari ketiga objek masing-masing adalah 98%, 90% dan 100% sedangkan nilai *f1-score* masing-masing sebesar 99%, 88% dan 99%.

DAFTAR PUSTAKA

- Bustami. (2014). Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika*, 8(1):127-146.
- Clarke, A.O., Scaife, A.M.M., Greenhalgh, R. & Griguta, V. (2020). Identifying Galaxies, Quasars, and Stars with

Machine Learning: A new catalogue of Classifications for 111 Million SDSS Sources Without Spectra. *Astronomy & Astrophysics*.

- Čokina, M., Krešňáková, V.M., Butka, P. & Parimucha, Š. (2021). Automatic Classification of Eclipsing Binary Stars Using Deep Learning Methods. *Astronomy and Computing*, 36.
- Jayasinghe, T., Kochanek, C.S., Stanek, K.Z., Shappee, B.J., Holoién, T.W.-S., Thompson, T.A., Prieto, J.L., Dong, S., Pawlak, M., Pejcha, O., Pojmanski, G., Otero, S., Hurst, N. & Will, D. (2021). The ASAS-SN Catalog of Variable Stars IX: The Spectroscopic Properties of Galactic Variable Stars. *Monthly Notices of the Royal Astronomical Society*, 503(1):200-235.
- Jayasinghe, T., Stanek, K.Z., Kochanek, C.S., Shappee, B.J., Holoién, T.W.-S., Thompson, T.A., Prieto, J.L., Dong, S., Pawlak, M., Pejcha, O., Shields, J.V., Pojmanski, G., Otero, S., Hurst, N., Britt, C.A. & Will, D. (2020). The ASAS-SN Catalogue of Variable Stars – V. Variables in the Southern Hemisphere. *Monthly Notices of the Royal Astronomical Society*, 491(1).
- Martin, G., Kaviraj, S., Hocking, A., Read, S.C. & Geach, J.E. (2019). Galaxy Morphological Classification in Deep-Wide Surveys via Unsupervised Machine Learning. *Monthly Notices of the Royal Astronomical Society*, 491:1408-1426.
- Percy, J.R. (2014). *Variable Stars: Action in the Sky!*. Diktat, Departemen Astronomi dan Astrofisika, Universitas Toronto, Kanada.
- Putra, J.W.G. (2020). *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning Edisi 1.4*.
- Rosandy, T. (2016). Perbandingan Metode Naïve Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS/BMT AL-FADHILA). *Jurnal TIM Darmajaya*, 2(1):52-62.
- Setiawan, E. (2014). *Analisis Penggunaan Kernel Density Estimation pada Metode Loss Distribution Approach untuk Risiko Operasional*. Tesis, Program Magister Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok.

- Sun, N., Sun, B., Lin, J. & Wu, M.Y.C. (2018). Lossless Pruned Naïve Bayes for Big Data Classifications. *Big Data Res.*
- Xiao-Qing, W. & Jin-Meng, Y. (2020). Classification of Star/Galaxy/QSO and Star Spectral Type from LAMOST Data Release 5 with Machine Learning Approaches. *Chinese Journal of Physics*.
<https://doi.org/10.1016/j.cjph.2020.03.008>.
- VizieR. *Catalog*. Diakses 28 September 2021, dari <https://vizier.u-strasbg.fr/viz-bin/VizieR?-source=J/AJ/137/4186>.

