

**SKRIPSI**

**KLASIFIKASI BINTANG RR LYRAE / CEPHEID / MIRA  
MENGUNAKAN METODE NAÏVE BAYES**

**Disusun dan diajukan oleh**

**AYU LESTARI**

**H021 17 1313**



**DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2022**

**KLASIFIKASI BINTANG RR LYRAE / CEPHEID / MIRA  
MENGUNAKAN METODE NAÏVE BAYES**

**SKRIPSI**

*Diajukan sebagai Salah Satu Syarat  
Memperoleh Gelar Sarjana Sains  
pada Program Studi Fisika Departemen Fisika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin*



**AYU LESTARI**

**H021 17 1313**

**DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2022**

**LEMBAR PENGESAHAN**

**KLASIFIKASI BINTANG RR LYRAE / CEPHEID / MIRA  
MENGUNAKAN METODE NAÏVE BAYES**

**Disusun dan diajukan oleh:**

**AYU LESTARI  
H021 17 1313**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 27 Januari 2022 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pertama,

  
Nur Hasanah, S.Si., M.Si.  
NIP. 19831122 200912 2 001

  
Eko Juarlin, S.Si., M.Si.  
NIP. 19811106 200812 1 002

Ketua Program Studi,

  
Prof. Dr. Arifin, M.T.  
NIP. 19670520 199403 1 002

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Ayu Lestari  
NIM : H021171313  
Program Studi: Fisika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

### **Klasifikasi Bintang RR Lyrae / Cepheid / Mira Menggunakan Metode Naïve Bayes**

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau seluruh skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 27 Januari 2022

Yang Menyatakan,



Ayu Lestari

## ABSTRAK

Kami menggunakan 10.000 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: asensio rekta, deklinasi, garis bujur galaksi, garis lintang galaksi, rata-rata spektrum tampak, amplitudo, periode, statistik *Lafler-Kinman String Length* dan paralaks dengan kolom jenis bintang digunakan sebagai target label. Metode yang digunakan dalam penelitian ini adalah algoritma Naïve Bayes. Metode ini memanfaatkan *training dataset* (data latih) dan *testing dataset* (data uji). Terdapat tiga prosedur dalam pembagian data latih dan data uji dengan persentase data yang berbeda, yaitu sebesar 10%, 20% dan 30% untuk data latih. Setiap kasus dengan proporsi data latih yang kecil berhasil melakukan klasifikasi terhadap data uji yang lebih besar dengan rata-rata nilai akurasi untuk kasus I, kasus II dan kasus III masing-masing sebesar 98.8%, 98.7% dan 98.8%. Kesimpulan dari penelitian ini adalah algoritma Naïve Bayes dapat digunakan dalam klasifikasi objek astronomi dengan tingkat akurasi yang baik.

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

## **ABSTRACT**

We used 10.000 data from the All-Sky Automated Survey for Supernovae (ASAS-SN) to classify RR Lyrae, Cepheid and Mira stars using a machine learning approach. There are nine columns that are used as attributes in making machine learning models, namely: right ascension, declination, galactic longitude, galactic latitude, mean visible spectrum, amplitude, period, the Lafler-Kinman String Length statistic and parallax with the variable type column used as the target label. The method used in this research is the Naïve Bayes algorithm. This method utilizes the training dataset and testing dataset. There are three procedures in the distribution of data with different percentages, namely 10%, 20% and 30% for the training dataset. Each case with a small proportion of the training dataset succeeded in classifying the larger testing dataset with an average accuracy value for case I, case II and case III are 98.8%, 98.7% and 98.8%, respectively. The conclusion of this research is that the Naïve Bayes algorithm can be used in the classification of astronomical objects with a good level of accuracy.

**Keywords:** ASAS-SN, Cepheid, Mira, Naïve Bayes, RR Lyrae.

## KATA PENGANTAR

Puji dan syukur penulis haturkan ke Hadirat Tuhan Yang Maha Esa atas segala limpahan rahmat, nikmat, petunjuk dan karunia-Nya terutama kesehatan dan kemudahan, sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul “**Klasifikasi Bintang RR Lyrae / Cepheid / Mira Menggunakan Metode Naïve Bayes**” yang merupakan salah satu syarat untuk mendapatkan gelar Sarjana Sains di Departemen Fisika Program Studi Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Shalawat serta salam juga semoga senantiasa selalu tercurahkan kepada Nabi Muhammad SAW.

Perjalanan penyelesaian skripsi ini adalah satu masa yang sangat berharga bagi penulis. Skripsi bukan hanya tentang menyelesaikannya dalam bentuk tulisan, prosesnya bukan hanya tentang membaca berbagai macam referensi, seminar proposal, seminar hasil, revisi, mengolah data, dsb. Namun lebih dari itu, skripsi adalah sebuah perjalanan yang sarat akan makna hidup, menemukan fitrah, makna belajar yang semua itu pula mengajarkan penulis tentang menjadi diri sendiri, tangguh, rendah hati dan tidak mudah putus asa dalam mencapai tujuan. Dalam perjalanan penyelesaian skripsi ini, penulis menyadari betul bahwa ada banyak pihak yang memberikan bantuan dan dukungan, baik secara langsung maupun secara tidak langsung. Oleh karena itu penulis ingin memberikan sebuah bentuk penghargaan dan ucapan terima kasih yang tak terhingga kepada:

1. Kedua orang tua tercinta, Ayahanda **Kamaruddin Beta** dan Ibunda **Hj. Patiha** yang tiada henti memberikan doa dan dukungan serta semangat untuk bisa menyelesaikan skripsi ini. Terima kasih atas didikan yang penuh cinta dan kasih sayang sehingga penulis bisa menjadi individu seperti saat ini. Semoga Allah senantiasa memberi kesehatan dan meridhai penulis dalam membahagiakan dan membanggakan beliau.
2. Kepada kakak-kakak terkasih, **Kak Erni, Kak Adnan dan Kak Anita** yang telah memberi dukungan secara moral maupun materil kepada penulis. Serta keponakan-keponakan penulis, **Nako dan Naka** yang telah memberi keramaian dan kelucuan dalam rumah.

3. Bapak **Prof. Dr. Arifin, M.T.**, selaku ketua Departemen Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin, yang telah memberi masukan dan motivasi kepada penulis untuk segera menyelesaikan tugas akhir. Terima kasih atas segala ilmu dan nasehat yang Bapak berikan kepada penulis selama menjadi mahasiswa Prodi Fisika Unhas.
4. Bapak **Prof. Dr. Tasrief Surungan, M.Sc.**, selaku Dosen Penasehat Akademik yang telah mendampingi, membimbing, dan memberikan dukungan kepada penulis. Terima kasih atas segala ilmu dan semangat yang diberikan selama penulis menjadi mahasiswa bimbingan PA sehingga penulis bisa menjadi pribadi yang semakin baik.
5. Ibu **Nur Hasanah, S.Si., M.Si.**, selaku Dosen Pembimbing Utama yang telah meluangkan waktu dan segenap energinya untuk mendukung penyelesaian tugas akhir ini. Terima kasih telah bersedia mengantar penulis untuk menjadi Sarjana Sains dengan segala sikap dan kepribadian penulis dengan tidak memberikan simpati kepada penulis dan terus membimbing penulis bukan hanya sebatas perbaikan skripsi, tetapi lebih dari itu adalah tentang bagaimana penulis bisa menjadi diri sendiri dan menjadi lebih baik. Terima kasih atas segala ilmu yang Ibu berikan secara keseluruhan kepada penulis selama menjadi mahasiswa Prodi Fisika Unhas.
6. Bapak **Eko Juarlin, S.Si., M.Si.**, selaku Dosen Pembimbing Pertama atas segala waktu dan energinya juga untuk mendukung penyelesaian tugas akhir ini. Terima kasih selalu mendampingi dan mengajarkan kepada penulis banyak hal. Terima kasih atas segala ilmu yang Bapak berikan secara keseluruhan kepada penulis selama menjadi mahasiswa Prodi Fisika Unhas.
7. Bapak **Prof. Dr. rer-nat Wira Bahari Nurdin, M.Eng. Sc.** dan Ibu **Dr. Sri Dewi Astuti, S.Si., M.Si.**, selaku Tim Penguji yang telah banyak meluangkan waktu dan energi untuk memberikan ilmu, saran dan diskusi dalam perbaikan skripsi ini.
8. Kak **Bryan Lawrencius Sondakh**, selaku mentor penulis selama melakukan kegiatan magang di PT. Aplikasi Karya Anak Bangsa (GOJEK Makassar). Terima kasih Kak Bryan telah memberikan penulis kesempatan untuk menjadi

bagian dari *Gojek East Indonesia Analytics Team*. Berdiskusi dengan Kak Bryan bukan hanya sekadar pekerjaan, namun Kak Bryan juga turut memberikan *insights* dalam penyelesaian tugas akhir ini. Penulis merasakan kurva perkembangan secara eksponensial selama dibimbing oleh Kak Bryan. Sebuah pengalaman yang sangat berharga dan akan selalu penulis jadikan bekal untuk perjalanan karir penulis ke depan.

9. Kak **Veny Charnita Ginting**, selaku mentor penulis saat mengikuti program belajar kolaborasi IMPLUSE ITB dan *Amazon Web Services (AWS)*. Terima kasih atas waktu dan energi yang Kak Veny berikan untuk membimbing dan memberi masukan kepada penulis bahkan setelah programnya selesai. Terima kasih telah mengantarkan penulis untuk lulus Program Generasi Gigih oleh Yayasan Anak Bangsa Bisa dengan memberi masukan yang sangat membangun pada esai dan resume penulis.
10. Seluruh **Dosen FMIPA Unhas**, khususnya kepada seluruh **Bapak dan Ibu Dosen Pengajar Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin**. Terima kasih telah memberikan ilmu yang bermanfaat dan mendidik penulis selama menjadi mahasiswa Prodi Fisika Unhas.
11. Bapak/Ibu Staf Pegawai FMIPA Unhas, terutama Staf Departemen Fisika, **Pak Syukur, Ibu Evi, Ibu Rana dan Pak Ahmad** yang selalu membantu penulis selama berada di kampus.
12. Teman-teman Laboratorium Teori dan Komputasi 2017, **Eka Yesiana, Destri Pratiwi, Aron Wiliyam, Safrullah, Agung Prawiranegara, Muhammad Sabran dan Betuel**. Terima kasih atas segala kebersamaan yang telah terbangun selama kurang lebih 4 tahun, terima kasih atas kolaborasi yang semakin terbentuk. Semoga kesuksesan selalu bersama dengan kita semua.
13. Teman-teman yang senantiasa berkolaborasi selama proses penyelesaian skripsi ini. Terima kasih kepada **Yesi, Zahra, Aron dan Fahri** atas segala bantuan dan dukungan yang diberikan kepada penulis.

14. Kepada sahabat seperjuangan penulis selama dibangku perkuliahan, **Evita, Hajar, Suci, Rachel, Fadil dan Trisna**. Terima kasih untuk selalu menemani dalam suka duka selama menjadi mahasiswa Prodi Fisika Unhas.
15. Kepada sahabat-sahabat Kosong High Level-ku, **Suci, Wiwi, Titi dan Ayu Rahayu**, terima kasih atas persahabatan yang telah terjalin sejak bangku SMP. Saya sangat bersyukur menjadi salah satu bagian dari kalian.
16. Teman-teman seperjuangan **Fisika Angkatan 2017** yang tidak sempat disebutkan satu – persatu namanya, banyak cerita telah dilalui bersama. Semoga menjadi kenangan indah untuk kita semua. Semoga kita semua sukses dimasa mendatang dan silaturahmi diantara kita tetap terjalin.
17. Teruntuk **diriku sendiri**, terima kasih telah mau berjuang dan memberikan versi terbaik dirimu.
18. Seluruh pihak yang turut terlibat dalam perjalanan penyelesaian skripsi ini yang penulis tidak sebutkan satu – persatu. Terima kasih atas segala dukungan dan bantuannya sehingga memudahkan penulis menyelesaikan skripsi ini.

Penulis berharap agar penelitian ini dapat memberikan banyak manfaat terutama dalam pengembangan keilmuan Fisika. Penulis menyadari bahwa penelitian ini masih jauh dari kesempurnaan, untuk itu penulis akan selalu terbuka terhadap umpan balik guna untuk melakukan penelitian yang lebih baik ke depannya. Akhir kata, semoga kita terus mendapat limpahan karunia dari Tuhan Yang Maha Pengasih berupa ilmu yang selalu dapat digunakan untuk mengabdikan kepada masyarakat.

Makassar, 27 Januari 2022



Ayu Lestari

## DAFTAR ISI

|   |      |
|---|------|
| <b>HALAMAN SAMPUL</b> .....                           | i    |
| <b>HALAMAN JUDUL</b> .....                            | ii   |
| <b>LEMBAR PENGESAHAN</b> .....                        | ii   |
| <b>PERNYATAAN KEASLIAN</b> .....                      | ii   |
| <b>ABSTRAK</b> .....                                  | v    |
| <b>ABSTRACT</b> .....                                 | vi   |
| <b>KATA PENGANTAR</b> .....                           | vii  |
| <b>DAFTAR ISI</b> .....                               | xi   |
| <b>DAFTAR GAMBAR</b> .....                            | xiii |
| <b>DAFTAR TABEL</b> .....                             | xiv  |
| <b>DAFTAR LAMPIRAN</b> .....                          | xv   |
| <b>BAB I PENDAHULUAN</b> .....                        | 1    |
| I.1 Latar Belakang .....                              | 1    |
| I.2 Rumusan Masalah .....                             | 2    |
| I.3 Tujuan.....                                       | 2    |
| <b>BAB II TINJAUAN PUSTAKA</b> .....                  | 3    |
| II.1 Bintang Variabel ( <i>Variable Stars</i> ).....  | 3    |
| II.2 Metode Naïve Bayes .....                         | 6    |
| <b>BAB III METODE PENELITIAN</b> .....                | 9    |
| III.1 Studi Literatur .....                           | 9    |
| III.2 Pengumpulan Data .....                          | 9    |
| III.3 Pembersihan Data ( <i>Data Cleansing</i> )..... | 9    |
| III.4 Transformasi Data .....                         | 10   |
| III.5 Visualisasi Data.....                           | 11   |
| III.6 Metode Naïve Bayes .....                        | 11   |
| III.7 Evaluasi Performa Model.....                    | 12   |
| III.8 Bagan Alir .....                                | 14   |
| <b>BAB IV HASIL DAN PEMBAHASAN</b> .....              | 16   |
| IV.1 Visualisasi Data.....                            | 16   |

|              |  |           |
|--------------|--|-----------|
| IV.2         | Klasifikasi Bintang RR Lyrae / Cepheid / Mira..... | 22        |
| IV.3         | Evaluasi Performa Model.....                       | 29        |
| <b>BAB V</b> | <b>KESIMPULAN DAN SARAN</b> .....                  | <b>32</b> |
| V.1          | Kesimpulan.....                                    | 32        |
| V.2          | Saran.....   | 32        |
|              | <b>DAFTAR PUSTAKA</b> .....                        | <b>33</b> |
|              | <b>LAMPIRAN</b> .....                              | <b>38</b> |

## DAFTAR GAMBAR

|                    |  |    |
|--------------------|--|----|
| <b>Gambar 2.1</b>  | (a) RR Lyrae, (b) kurva cahaya RR Lyrae .....  | 4  |
| <b>Gambar 2.2</b>  | (a) RS Puppis, (b) kurva cahaya Cepheid.....   | 5  |
| <b>Gambar 2.3</b>  | (a) Chi Cyg, (b) kurva cahaya Mira.....  | 6  |
| <b>Gambar 3.1</b>  | Diagram alir metode Naïve Bayes.....   | 14 |
| <b>Gambar 3.2</b>  | Diagram alir penelitian .....  | 15 |
| <b>Gambar 4.1</b>  | Persentase RR Lyrae, Cepheid dan Mira dalam dataset.....   | 16 |
| <b>Gambar 4.2</b>  | Distribusi densitas untuk sistem koordinat ekuator,<br>(a) <i>right ascension</i> (asensio rekta), (b) <i>declination</i> (deklinasi).....         | 17 |
| <b>Gambar 4.3</b>  | Distribusi densitas untuk sistem koordinat galaksi,<br>(a) <i>galactic longitude</i> ( <i>l</i> ), (b) <i>galactic latitude</i> ( <i>b</i> ) ..... | 18 |
| <b>Gambar 4.4</b>  | Distribusi densitas untuk <i>mean visible magnitude</i> .....  | 19 |
| <b>Gambar 4.5</b>  | Distribusi densitas untuk amplitudo .....  | 19 |
| <b>Gambar 4.6</b>  | Distribusi densitas untuk periode .....  | 20 |
| <b>Gambar 4.7</b>  | Distribusi densitas untuk ukuran hamburan cahaya .....   | 20 |
| <b>Gambar 4.8</b>  | Distribusi densitas untuk paralaks .....   | 21 |
| <b>Gambar 4.9</b>  | Nilai presisi, <i>recall</i> dan <i>f1-score</i> setiap jenis bintang<br>berdasarkan proporsi pembagian data dalam dataset .....                   | 30 |
| <b>Gambar 4.10</b> | <i>Confusion matrix</i> .....  | 31 |

## DAFTAR TABEL

|                  |  |    |
|------------------|--|----|
| <b>Tabel 3.1</b> | Atribut dalam dataset ASAS-SN .....                    | 10 |
| <b>Tabel 3.2</b> | Transformasi data jenis objek .....                    | 11 |
| <b>Tabel 3.3</b> | Model <i>confusion matrix</i> .....                    | 13 |
| <b>Tabel 4.1</b> | Hasil pembagian data latih dan data uji .....          | 22 |
| <b>Tabel 4.2</b> | Contoh data latih .....                                | 22 |
| <b>Tabel 4.3</b> | Contoh data uji .....                                  | 23 |
| <b>Tabel 4.4</b> | Nilai rata-rata dan standar deviasi setiap kasus ..... | 23 |
| <b>Tabel 4.5</b> | Nilai distribusi Gaussian .....                        | 24 |
| <b>Tabel 4.6</b> | Nilai <i>likelihood</i> data uji 1 .....               | 26 |
| <b>Tabel 4.7</b> | Nilai <i>likelihood</i> data uji 2 .....               | 27 |
| <b>Tabel 4.8</b> | Nilai <i>likelihood</i> data uji 3 .....               | 28 |
| <b>Tabel 4.9</b> | Performa model .....                                   | 29 |

## DAFTAR LAMPIRAN

|  |    |
|--|----|
| <b>Lampiran 1.</b> Plot sebaran data untuk sistem koordinat ekuator .....  | 38 |
| <b>Lampiran 2.</b> Plot sebaran data untuk sistem koordinat galaksi.....   | 39 |
| <b>Lampiran 3.</b> Sampel dataset bintang RR Lyrae, Cepheid dan Mira ..... | 40 |
| <b>Lampiran 4.</b> Sintaks simulasi menggunakan Scilab.....                | 42 |

# BAB I

## PENDAHULUAN

### I.1 Latar Belakang

Bintang variabel merupakan bintang yang kecerahannya berubah seiring waktu [1]. Bintang variabel dapat digunakan untuk mempelajari populasi bintang dan struktur galaksi [2]. Secara garis besar bintang variabel terbagi atas dua, yaitu bintang variabel ekstrinsik dan intrinsik. Bintang variabel ekstrinsik adalah bintang yang keluaran cahayanya berubah karena beberapa proses di luar bintang. Contoh yang paling umum adalah *eclipsing binary*. Sedangkan bintang variabel intrinsik adalah bintang yang perubahan kecerahan disebabkan oleh beberapa perubahan di dalam bintang [1]. Salah satu contoh bintang variabel intrinsik adalah bintang berdenyut (*pulsating stars*). Bintang variabel berdenyut mengembang dan berkontraksi secara berirama. Periode berkisar dari detik (katai putih) hingga tahun (raksasa super) [3]. Beberapa jenis bintang variabel berdenyut adalah RR Lyrae, Cepheid dan Mira.

Dalam beberapa dekade terakhir, teleskop robotik telah berkembang sangat pesat. Sehingga data survei langit telah menghasilkan sumber data pengamatan yang melimpah. Data ini berisi banyak informasi tentang bintang variabel. Banyaknya data astronomi yang dikumpulkan, menjadi mustahil untuk semua data diproses dan dianalisis secara manual. Proses ini dapat memakan waktu yang lama dan kemungkinan terjadinya bias sangat besar. Oleh karena itu, diperlukan sebuah metode dalam analisis dan eksplorasi data yang efektif dan akurat. Solusi yang paling mungkin adalah melibatkan algoritma *machine learning* [4].

Berbagai penerapan *machine learning* sudah dilakukan, seperti klasifikasi galaksi, kuasar dan bintang [5], morfologi galaksi [6], verifikasi gelombang gravitasi [7] dan masih banyak lagi. Termasuk dalam penyusunan katalog ASAS-SN bintang variabel [8] 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 telah

diterapkan oleh M. Cokina, et. al [4]. 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 *Kohonen Self Organising Maps* (SOM) [9], *AdaBoost classifier* [10], *K-Nearest Neighbor* (KNN), *Decision Tree* (DT), dan *Support Vector Machine* (SVM) [11].

Berdasarkan hal yang telah dijelaskan diatas, maka akan dilakukan penelitian mengenai klasifikasi bintang RR Lyrae, Cepheid dan Mira menggunakan salah satu metode *machine learning* yang lain, yaitu metode Naïve Bayes. Beberapa penelitian telah menunjukkan keunggulan algoritma Naïve Bayes terhadap algoritma SVM dan KNN [12-13]. Hal ini membuktikan kemampuan algoritma Naïve Bayes dalam klasifikasi. Penelitian ini diharapkan dapat membuat model *machine learning* dalam mengklasifikasikan ketiga jenis bintang, yaitu RR Lyrae, Cepheid dan Mira.

## **I.2 Rumusan Masalah**

Rumusan masalah dari penelitian ini antara lain:

1. Bagaimana mengolah data ASAS-SN dalam klasifikasi bintang RR Lyrae, Cepheid dan Mira dengan metode Naïve Bayes?
2. Bagaimana tingkat akurasi dari penerapan metode Naïve Bayes?

## **I.3 Tujuan**

Tujuan dilaksanakannya penelitian ini antara lain:

1. Menghasilkan model *machine learning* yang mampu mengklasifikasikan bintang RR Lyrae, Cepheid dan Mira menggunakan metode Naïve Bayes.
2. Menganalisis tingkat akurasi dari metode Naïve Bayes.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **II.1 Bintang Variabel (*Variable Stars*)**

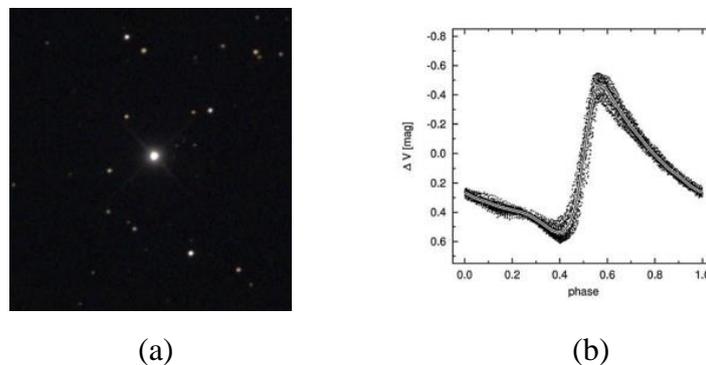
Terdapat miliaran bintang di alam semesta. Mereka adalah sumber cahaya alami selain kilat selama badai Guntur [14]. Bintang didefinisikan sebagai bola gas yang disatukan oleh gravitasinya sendiri. Bintang terdekat dengan Bumi adalah Matahari [15]. Ketika melihat ke langit malam, sebagian besar beranggapan bahwa bintang tidak berubah. Selain berkelap-kelip karena efek atmosfer, bintang tampak tetap dan konstan bagi mata. Namun, bintang mengalami perubahan kecerahan dari waktu ke waktu. Beberapa menunjukkan perilaku periodik, cerah dengan cepat kemudian berkurang kecerahannya secara perlahan hanya untuk terulang kembali. Proses ini merupakan contoh dari bintang variabel. Bintang variabel adalah bintang dengan kecerahan yang bervariasi [16].

Salah satu jenis bintang variabel adalah bintang berdenyut (*pulsating stars*). Bintang berdenyut adalah bintang yang secara fisik mengembang dan mengerut [16]. Proses *pulsating* pada bintang disebabkan oleh ekspansi dan kontraksi periodik dari lapisan permukaan bintang [17]. Dalam [16], dijelaskan bahwa di bagian dalam bintang terdapat medan gravitasi sebagai akibat dari adanya massa bintang yang menyatukan semua komponen penyusun bintang. Terdapat pula tekanan yang membuat bintang mengembang. Pada bintang non-variabel, kedua kuantitas ini seimbang sehingga bintang akan tetap pada ukurannya. Pada bintang variabel yang telah mengalami kontraksi melalui beberapa gangguan, saat suhunya naik maka atom helium di lapisan luar bintang menjadi terionisasi. Helium yang terionisasi menghentikan panas dan cahaya dari dalam bintang agar tidak keluar semudah sebelumnya. Artinya, suhu dan tekanan meningkat lebih cepat di dalam bintang. Ketika tekanan menjadi lebih besar daripada gaya gravitasi, maka bintang akan mengembang. Lalu jika bintang mencapai keadaan setimbang, tekanan akan menurun dan gas mendingin serta ion helium akan bergabung kembali menjadi atom utuh. Dalam keadaan ini, bintang akan jauh lebih transparan dan mendingin lebih cepat. Hal ini menandakan tekanan gas menurun secara tajam sehingga gaya

gravitasi akan jauh lebih kuat dan bintang kembali mengerut. Berbagai jenis variabel berdenyut dibedakan oleh periode *pulsating* dan bentuk kurva cahayanya, beberapa diantaranya adalah RR Lyrae, Cepheid dan Mira.

### II.1.1 RR Lyrae

Bintang RR Lyrae adalah raksasa putih berdenyut yang lebih tua dengan tingkat logam yang rendah. Mereka umum di gugus bola – kelompok padat bintang tua di lingkaran cahaya galaksi. RR Lyrae mengembang dan berkontraksi dalam siklus perubahan ukuran yang berulang. Perubahan ukuran dapat diamati sebagai perubahan kecerahan semu (magnitudo semu). Denyutan RR Lyrae bersifat periodik. RR Lyrae memiliki  $\sim 0,5$  massa matahari dan memiliki periode pulsasi pendek 0,05 hingga 1,2 hari dan variasi amplitudo 0,3 hingga 2 magnitudo. Bintang RR Lyrae biasanya kelas spektral A [17]. Gambar 2.1 (a) menunjukkan bintang RR Lyrae sedangkan bagian (b) menunjukkan penurunan dan peningkatan berkala dari kecerahan RR Lyrae yang diilustrasikan melalui kurva cahaya.



**Gambar 2.1** (a) RR Lyrae [18], (b) kurva cahaya RR Lyrae [17]

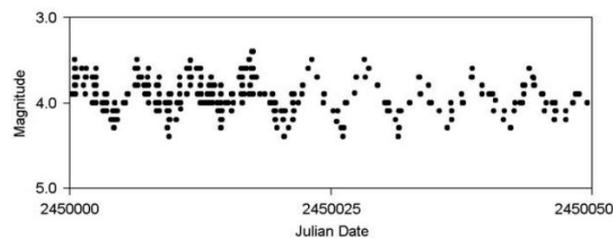
### II.1.2 Cepheid

Saat ini lebih dari 1.000 bintang Cepheid telah dikatalogkan, sementara jumlahnya akan terus bertambah [19]. Cepheid menjadi bintang variabel yang paling banyak dipelajari. Amplitudo besar dan luminositas intrinsiknya membuatnya mudah dideteksi di sebagian besar survei variabilitas fotometrik [20]. Cepheid memiliki siklus perubahan berulang yang periodik – teratur seperti detak jantung, dengan periode 1 hingga 135 hari dengan variasi amplitudo dari 0,1 hingga

2,0 magnitudo [21]. Bintang masif ini ( $\sim 8$  massa matahari) memiliki luminositas tinggi dan kelas spektral F maksimum, dan G hingga K minimum [17]. Gambar objek dan kurva cahaya Cepheid dapat dilihat pada Gambar 2.2.



(a)

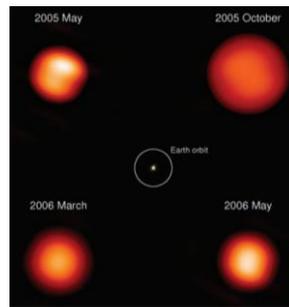


(b)

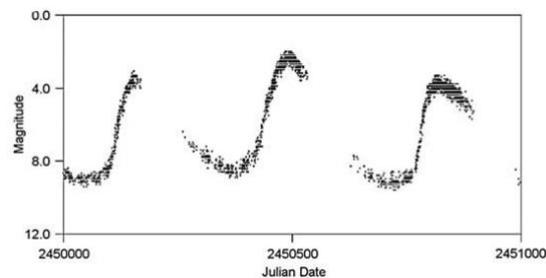
**Gambar 2.2** (a) RS Puppis [22], (b) kurva cahaya Cepheid [17]

### II.1.3 Mira

Bintang Mira adalah raksasa merah berdenyut periodik dengan periode yang relatif stabil, yaitu 100 hingga 1000 hari dengan sebagian besar jatuh antara 150 hingga 450 hari [13]. Mira memiliki variasi amplitudo lebih dari 2,5 hingga 10 magnitudo. Matahari pada akhirnya akan menjadi bintang Mira yang berdenyut. Mira pada diagram H-R menempati wilayah antara bintang berukuran sedang pada deret utama dan cabang raksasa [17]. Salah satu bintang Mira yang berhasil diamati, yaitu *Chi Cyg* dapat dilihat pada Gambar 2.3 (a). Pada bagian (b) menggambarkan kurva cahaya Mira secara umum.



(a)



(b)

**Gambar 2.3** (a) Chi Cyg, (b) kurva cahaya Mira [17]

## II.2 Metode Naïve Bayes

Dalam *machine learning* terdapat beberapa algoritma yang dapat digunakan untuk mengklasifikasikan data ke dalam beberapa kelas seperti Naïve Bayes, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) dan *Random Forest* (RF) [23]. Klasifikasi adalah bidang *machine learning* dalam penambangan data (*data mining*) dan pengenalan pola [24]. Salah satu metode yang sering digunakan dalam klasifikasi adalah 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 [25]. Alur dari metode Naïve Bayes dijelaskan pada Pers. (2.1) sampai (2.7) [26]:

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

Keterangan:

$X$  : Data dengan kelas yang belum diketahui

$H$  : Hipotesis data  $X$  merupakan suatu kelas spesifik

$P(H|X)$  : Probabilitas hipotesis  $H$  berdasarkan kondisi  $X$  (*posterior probability*)

$P(H)$  : Probabilitas hipotesis  $H$  (*prior probability*)

$P(X|H)$  : Probabilitas hipotesis  $X$  berdasarkan kondisi  $H$

$P(X)$  : Probabilitas hipotesis  $X$

Proses klasifikasi memerlukan sejumlah petunjuk atau karakteristik tertentu untuk menentukan jenis kelas pada sampel yang dianalisis. Oleh karena itu, teorema Bayes disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2.2)$$

dimana variabel  $C$  merepresentasikan kelas, sementara variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi.

$$Posterior = \frac{(prior \times likelihood)}{evidence} \quad (2.3)$$

Dalam satu sampel, nilai *evidence* akan selalu tetap untuk setiap kelas, sedangkan nilai *posterior* akan dibandingkan dengan nilai *posterior* kelas lainnya untuk menentukan jenis kelas dari sampel yang akan diklasifikasikan. Penjabaran rumus Bayes dapat dilakukan dengan menjabarkan  $(C|F_1, \dots, F_n)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2, F_3)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2, F_3) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad (2.4)$$

Berdasarkan Pers. (2.4) dapat dilihat bahwa semakin banyak faktor yang mempengaruhi nilai probabilitas maka akan mustahil untuk dilakukan analisis satu persatu. Kemudian diperkenalkan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk ( $F_1, F_2, \dots, F_n$ ) tidak bergantung satu sama lain (*independent*).

$$P(P_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (2.5)$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (2.6)$$

Pers. (2.6) menunjukkan bahwa asumsi independensi naif membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Untuk penjabaran  $P(C|F_1, \dots, F_n)$  dapat dituliskan menjadi:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots \\ &= P(C) \prod_{i=1}^n P(F_i|C) \end{aligned} \quad (2.7)$$

Metode Naïve Bayes terdiri atas tiga jenis, yaitu [27]: Gaussian Naïve Bayes, yaitu mengasumsikan bahwa data setiap atribut diambil dari distribusi Gaussian yang sederhana. Kedua adalah Multinomial Naïve Bayes, yaitu atribut diasumsikan diambil dari distribusi multinomial. Metode ini cocok untuk atribut dengan data yang diskrit. Ketiga yaitu Bernoulli Naïve Bayes yang digunakan jika atribut diasumsikan biner. Terdapat beberapa alasan pemilihan metode Naïve Bayes pada penelitian ini, diantaranya [22]:

- Pengklasifikasi Naïve Bayes mudah diimplementasikan,
- Pengklasifikasi Naïve Bayes tidak membutuhkan banyak data latih,
- Pengklasifikasi Naïve Bayes cepat dan dapat digunakan secara *real-time*. Keuntungan ini mungkin tidak berguna untuk saat ini, tetapi ini akan menjadi penting karena data yang digunakan akan sangat besar dan akan terus bertambah seiring waktu,
- Pengklasifikasi Naïve bayes tidak sensitif terhadap data yang tidak relevan.