

# **KLASIFIKASI DATASET MENGGUNAKAN METODE NAÏVE BAYES**

*Diajukan Sebagai Salah Satu Syarat untuk Menyelesaikan  
Pendidikan Program Strata – 1 pada  
Jurusan Teknik Informatika*



**Oleh :**

**JABES PUTRA PAULUS PASARIBU**

**09021181320066**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2021**

## LEMBAR PENGESAHAN SKRIPSI

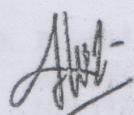
### KLASIFIKASI DATASET MENGGUNAKAN METODE *NAÏVE BAYES*

OLEH:

JABES PUTRA PAULUS PASARIBU

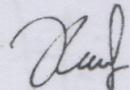
09021181320066

Pembimbing I,



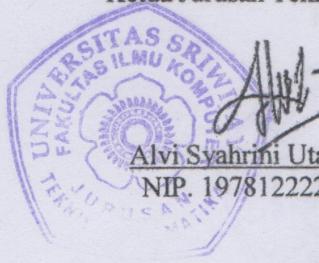
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

Indralaya, 15 Januari 2021  
Pembimbing II,



Kanda Januar Miraswan, M.T.  
NIP. 199001092019031012

Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

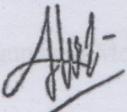
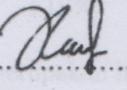
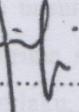
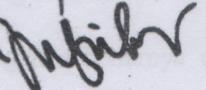
## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari rabu, tanggal 30 Desember 2020 telah dilaksanakan Ujian Sidang Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

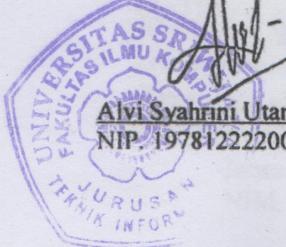
Nama : Jubes Putra Paulus Pasaribu

NIM : 09021181320066

Judul : Klasifikasi Dataset Menggunakan Metode *Naive Bayes*

1. Pembimbing I  
Alvi Syahrini Utami, M.Kom. .....  
NIP. 197812222006042003 
2. Pembimbing II  
Kanda Januar Miraswan, M.T. .....  
NIP. 199001092019031012 
3. Penguji I  
Rifkie Primartha, M.T. .....  
NIP. 197706012009121004 
4. Penguji II  
Nabila Rizky Oktadini, M.T. .....  
NIP. 199110102018032001 

Mengetahui,  
Ketua Jurusan Teknik Informatika

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Jubes Putra Paulus Pasaribu

NIM : 09021181320066

Program Studi : Teknik Informatika

Judul Skripsi : Klasifikasi Dataset Menggunakan Metode *Naive Bayes*

Hasil Pengecekan *Software iThenticate/ Turnitine* : 19%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/ plagiat. Apabila ditemukan unsur penjiplakan/ plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, Januari 2021



Jubes Putra Paulus Pasaribu  
NIM. 09021181320066

## MOTTO DAN PERSEMBAHAN

MOTTO :

❖ “*Serahkanlah segala kekuatiranmu kepada-Nya, sebab Ia yang memelihara kamu*”

*1 Petrus 5 : 7*

❖ “*Takut akan Tuhan adalah permulaan pengetahuan, tetapi orang bodoh menghina hikmat dan didikan*”

*Amsal 1 : 7*

❖ “*Not knowing oneself, that's the worst.*”

❖ “*Knowing is not enough, we must apply*”

❖ “*Empty your cup, In-order to fill it again*”

*By : Chou*

Kupersembahkan Karya Tulis ini kepada :

- Tuhan Yesus Kristus
- Kedua Orang Tuaku (Mama Lasmaria Sihotang dan Bapak Alm. Lamsihar Pasaribu)
- Adik-adik Tercinta (Lambok, Santa dan Putri)
- Almamater Saya
- Sahabat yang selalu memotivasi
- “Diri sendiri”

# **Dataset Classification Using Naive Bayes Method**

Jabes Putra Paulus Pasaribu (09021181320066)

Department of Informatics, Faculty of Computer Science

Sriwijaya University

## **ABSTRACT**

In this study, a system was built to do dataset classification using *Naive Bayes* method. The study aims to look at the classification performance of *Naive Bayes* whether it is influenced by dataset characteristics. The study used three data sets with a fairly different number of attributes, instantaneous numbers and data types. *Split validation* 70% and *K-fold Cross validation* with  $k = 10$  used as evaluation method. The results of this study showed that data type, number of attributes and instant number influenced classification results. Accuracy results tend to get better when more and more dataset attributes. Soybean with 35 attributes and 683 instant numbers as the largest dataset, is classified quite well and produces an accuracy of 92.79%. The study also showed that testing with k-fold cross validation value  $k=10$  resulted in a better classification than testing with a split validation ratio of 70%.

Keywords : data mining, classification, *Naive Bayes*, dataset

## **Klasifikasi Dataset Menggunakan Metode Naive Bayes**

Jabes Putra Paulus Pasaribu (09021181320066)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

### **ABSTRAK**

Pada penelitian ini dibangun sebuah sistem untuk melakukan klasifikasi dataset dengan menggunakan metode *Naive Bayes*. Penelitian bertujuan untuk melihat kinerja klasifikasi *Naive Bayes* apakah dipengaruhi karakteristik dataset. Penelitian ini menggunakan tiga set data dengan jumlah atribut, jumlah instan dan tipe data yang cukup berbeda. *Split validation 70%* dan *K-fold Cross validation* dengan nilai  $k=10$  digunakan sebagai metode evaluasi. Hasil penelitian ini menunjukkan bahwa tipe data, jumlah atribut dan jumlah instan mempengaruhi hasil klasifikasi. Hasil akurasi cenderung semakin baik bila semakin banyak jumlah atribut dataset. Soybean bertipe nominal dengan jumlah atribut terbanyak yaitu 35 dan jumlah instan 683 diklasifikasikan dengan cukup baik dan menghasilkan akurasi sebesar 92,79%. Penelitian ini juga menunjukkan bahwa pengujian dengan *k-fold cross validation* nilai  $k=10$  menghasilkan klasifikasi yang lebih baik daripada pengujian dengan *split validation* rasio 70%.

Kata kunci – data mining, klasifikasi, dataset, *Naive Bayes*

## KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yesus Kristus atas rahmat dan karunia-Nya sehingga Penulis mampu menyelesaikan Tugas Akhir ini yang berjudul “Klasifikasi Dataset Menggunakan Metode *Naive Bayes*”. Tugas akhir ini disusun untuk memenuhi salah satu syarat lulus tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis pada kesempatan ini mengucapkan terimakasih kepada semua yang telah mendukung, membimbing, membantu dan memberi arahan dalam penulisan tugas akhir ini diantaranya adalah :

1. Tuhan Yesus Kristus atas berkat, rahmat dan penyertaan-Nya.
2. Diri sendiri, Jabel yang sudah berhasil menghadapi beberapa cobaan selama kuliah dan terlebih dalam penyelesaian Tugas Akhir ini.
3. Mama yang selalu mendukung, memberi semangat, tetap percaya dan selalu menyebutkan namaku dalam setiap doanya, Alm. Bapak yang sudah tenang di surga, juga adik-adik tercinta yang tetap mendukung dan sabar atas kelebihan dan kekuranganku, yang sudah memberi dukungan moral dan materil juga setiap doa yang selalu dipanjatkan untuk kesuksesan saya. Keluarga besar serta kerabat yang juga selalu memberi semangat dalam penulisan tugas akhir ini.
4. Bapak Jaidan Jauhari, M.T. selaku dekan Fakultas Ilmu Komputer Universitas Sriwijaya,
5. Ibu Alvi Syahrini Utami, M. Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya yang senantiasa memberi arahan dan motivasi terhadap mahasiswa termasuk Penulis,
6. Ibu Alvi Syahrini Utami, M. Kom. selaku pembimbing pertama dan Bapak Kanda Januar Miraswan, M.T. selaku pembimbing kedua yang telah banyak memberi motivasi, arahan dan ilmu dalam penulisan tugas akhir ini sehingga tugas akhir ini dapat terselesaikan,

7. Bapak Rifkie Primartha, M.T. dan Ibu Nabila Rizky Oktadini, M.T. selaku penguji Tugas Akhir yang telah memberikan nasihat dan saran yang membangun,
8. Bapak Ali Ibrahim, M.T. selaku dosen pembimbing akademik yang selalu memberikan saran dan arahan selama masa perkuliahan,
9. Seluruh Dosen Jurusan Teknik Informatika yang telah memberikan bekal ilmu selama masa perkuliahan,
10. Sahabat KK Lima, Jou, Ady, Wasti, Santi dan seluruh Angkatan 2013 PDO Sion sebagai tempat berbagi keluh kesah juga atas dukungan dan motivasi,
11. Teman-teman bedeng Nixon, Friski, Friska, Tiffany, Vero, warga PDO Sion, seluruh keluarga Naimarata dan Sirajaoloan UNSRI,
12. Teman seperjuangan Teknik Informatika Angkatan 2013,
13. Sahabat terbaik Yuli Arnita Sihaloho,
14. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.

Penulis menyadari masih banyak terdapat kekurangan dalam penyusunan Tugas Akhir ini yang disebabkan oleh keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran membangun sangat diharapkan untuk menyempurnakan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Indralaya, 15 Januari 2021

Jabes Putra Paulus Pasaribu

## DAFTAR ISI

<b>COVER .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>LEMBAR TANDA LULUS.....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>iv</b>
<b>MOTTO DAN PERSEMBAHAN.....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>DAFTAR GAMBAR.....</b>	<b>xv</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN .....</b>	<b>I-1</b>
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang Masalah .....	I-1
1.3    Perumusan Masalah.....	I-3
1.4    Tujuan Penelitian.....	I-3
1.5    Manfaat Penelitian.....	I-3
1.6    Batasan Masalah.....	I-4
1.7    Sistematika Penulisan.....	I-4
1.8    Kesimpulan.....	I-5
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>II-1</b>
2.1    Pendahuluan .....	II-1
2.2    Penelitian Terdahulu.....	II-1
2.3    Seleksi dan Pengolahan Data .....	II-2
2.4    Praproses Data .....	II-3
2.5    Data Mining.....	II-4
2.6    Klasifikasi.....	II-5
2.6.1    Konsep Pembuatan Model dalam Klasifikasi .....	II-6
2.6.2    Proses Klasifikasi Data .....	II-6

2.7	<i>Naive Bayes</i> .....	II-7
2.8	Data yang Digunakan .....	II-9
2.8.1	Data Iris.....	II-9
2.8.2	Data Soybean (large).....	II-11
2.8.3	Data Ionosphere .....	II-12
2.9	Kesimpulan.....	II-12
 BAB III METODOLOGI PENELITIAN .....		III-1
3.1	Pendahuluan .....	III-1
3.2	Unit Penelitian .....	III-1
3.3	Metode Pengumpulan Data .....	III-1
3.4	Tahapan Penelitian .....	III-2
3.4.1	Menentukan Ruang Lingkup Penelitian.....	III-2
3.4.2	Menentukan Dasar Teori yang Berkaitan dengan Penelitian.....	III-3
3.4.3	Menetapkan Kriteria Pengujian.....	III-3
1.	<i>Cross Validation</i> .....	III-3
2.	<i>Split Validation</i> .....	III-4
3.4.3.1	Pengelompokan data menggunakan metode <i>Naïve Bayes</i> .....	III-4
3.4.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-5
<u>III-3.4.5</u>	Melakukan Pengujian Penelitian.....	III-6
3.4.6	Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-7
3.5	Metode Pengembangan Perangkat Lunak .....	III-9
3.5.1	Fase Insepsi .....	III-9
3.5.2	Fase Elaborasi .....	III-9
3.5.3	Fase Konstruksi.....	III-10
3.5.4	Fase Transisi .....	III-10
3.5.5	Penjadwalan Penelitian .....	III-11
 BAB IV PENGEMBANGAN PERANGKAT LUNAK .....		IV-1
4.1	Pendahuluan .....	IV-1

4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan .....	IV-2
a.	Fitur Membuka File .....	IV-3
b.	Fitur Klasifikasi <i>Naive Bayes</i> .....	IV-3
4.2.3	Analisis dan Desain.....	IV-4
4.2.3.1	Analisis Perangkat Lunak.....	IV-4
a.	Analisis Kebutuhan Perangkat Lunak .....	IV-4
b.	Analisis Data .....	IV-5
c.	Analisis Metode <i>Naive Bayes</i> .....	IV-5
4.2.3.2	Desain Perangkat Lunak .....	IV-6
1.	Model Use Case.....	IV-6
2.	Diagram Aktivitas .....	IV-11
4.3	Fase Elaborasi.....	IV-13
4.3.1	Pemodelan Bisnis .....	IV-13
4.3.1.1	Perancangan Data.....	IV-14
4.3.1.2	Perancangan Antarmuka .....	IV-14
4.3.2	Kebutuhan Sistem .....	IV-15
4.3.3	Diagram Sequence .....	IV-15
4.4	Fase Konstruksi .....	IV-21
4.4.1	Kebutuhan Sistem .....	IV-21
4.4.2	Diagram Kelas.....	IV-21
4.4.3	Implementasi.....	IV-23
4.4.3.1	Implementasi Kelas .....	IV-23
4.4.3.2	Implementasi Antarmuka .....	IV-24
4.5	Fase Transisi.....	IV-25
4.5.1	Pemodelan Bisnis .....	IV-25
4.5.2	Kebutuhan Sistem .....	IV-26
4.5.3	Rencana Pengujian .....	IV-26
4.5.3.1	Rencana Pengujian Use Case Membuka File.....	IV-26
4.5.3.2	Rencana Pengujian Use Case Klasifikasi <i>Naive Bayes</i> .....	IV-27

4.5.3.3 Rencana Pengujian Use Case Menentukan Atribut .....	IV-27
4.5.3.4 Rencana Pengujian Use Case Melakukan Validasi.....	IV-28
4.5.4 Implementasi.....	IV-28
4.5.4.1 Pengujian Use Case Membuka File .....	IV-28
4.5.4.2 Pengujian Use Case Klasifikasi <i>Naive Bayes</i> .....	IV-29
4.5.4.3 Pengujian Use Case Menentukan Jenis Atribut .....	IV-31
4.5.4.4 Pengujian Use Case Melakukan Validasi .....	IV-32
 BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan .....	V-1
5.2 Data Hasil Penelitian .....	V-1
5.2.1 Pembagian Data .....	V-1
a. <i>Split Validation</i> .....	V-1
b. <i>K-fold Cross Validation</i> .....	V-2
5.2.2 Hasil Pengujian Metode Naive Bayes .....	V-2
a. Split Validation.....	V-3
b. Cross Validation .....	V-5
5.3 Analisis Hasil Penelitian .....	V-7
 BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran .....	VI-2
 <b>DAFTAR PUSTAKA .....</b>	<b>xvii</b>
 <b>LAMPIRAN.....</b>	<b>1</b>

## DAFTAR TABEL

	Halaman
<b>Tabel III-1. Tabel Pembagian Data Latih dan Data Uji .....</b>	<b>III-4</b>
Tabel III-2. Tabel Hasil Metode .....	III-6
Tabel III-3. Tabel Rancangan <i>Confussion Matrix</i> .....	III-6
Tabel III-4. Presisi, Recall dan Akurasi .....	III-8
Tabel III-5. Penjadwalan Penelitian.....	III-11
<b>Tabel IV-1. Kebutuhan Fungsional.....</b>	<b>IV-4</b>
<b>Tabel IV-2. Kebutuhan Non-Fungsional .....</b>	<b>IV-4</b>
<b>Tabel IV-3. Definisi Aktor <i>Use Case</i>.....</b>	<b>IV-7</b>
<b>Tabel IV-4. Definisi <i>Use Case</i>.....</b>	<b>IV-7</b>
<b>Tabel IV-5. Skenario <i>Use Case</i> Membuka File.....</b>	<b>IV-8</b>
<b>Tabel IV-6. Skenario <i>Use Case</i> Klasifikasi <i>Naive Bayes</i> .....</b>	<b>IV-8</b>
<b>Tabel IV-7. Skenario <i>Use Case</i> Menentukan Jenis Atribut .....</b>	<b>IV-9</b>
<b>Tabel IV-8. Skenario <i>Use Case</i> Melakukan Validasi .....</b>	<b>IV-10</b>
<b>Tabel IV-9. Implementasi Kelas .....</b>	<b>IV-23</b>
<b>Tabel IV-10. Rencana Pengujian <i>Use Case</i> Membuka File.....</b>	<b>IV-26</b>
<b>Tabel IV-11. Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i> .....</b>	<b>IV-27</b>
<b>Tabel IV-12. Rencana Pengujian <i>Use Case</i> Menentukan Atribut .....</b>	<b>IV-27</b>
<b>Tabel IV-13. Rencana Pengujian <i>Use Case</i> Melakukan Validasi .....</b>	<b>IV-28</b>
<b>Tabel IV-14 . Pengujian <i>Use Case</i> Membuka File.....</b>	<b>IV-29</b>
<b>Tabel IV-15 . Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i> .....</b>	<b>IV-29</b>
<b>Tabel IV-16 . Pengujian <i>Use Case</i> Menentukan Jenis Atribut .....</b>	<b>IV-31</b>
<b>Tabel IV-17 . Pengujian <i>Use Case</i> Melakukan Validasi .....</b>	<b>IV-32</b>
<b>Tabel V-1. Pembagian Data Latih dan Data Uji <i>Split Validation</i> .....</b>	<b>V-2</b>
<b>Tabel V-2. Pembagian Data Latih dan Data Uji K-fold Cross Validation..</b>	<b>V-2</b>
<b>Tabel V-2. Pembagian Data Latih dan Data Uji K-fold Cross Validation..</b>	<b>V-2</b>
<b>Tabel V-3. Presisi, Recall dan Akurasi Split Validation Data Iris .....</b>	<b>V-3</b>
<b>Tabel V-4. Presisi, Recall dan Akurasi Split Validation Data Soybean.....</b>	<b>V-3</b>
<b>Tabel V-5. Presisi, Recall dan Akurasi Split Validation Data Ionosphere ..</b>	<b>V-4</b>
<b>Tabel V-6. Presisi, Recall dan Akurasi Cross Validation Data Iris .....</b>	<b>V-5</b>
<b>Tabel V-7. Presisi, Recall dan Akurasi Cross Validation Data Soybean ....</b>	<b>V-6</b>
<b>Tabel V-8. Presisi, Recall dan Akurasi Cross Validation Data Ionosphere</b>	<b>V-7</b>

## DAFTAR GAMBAR

	Halaman
Gambar II-1. Klasifikasi.....	II-5
Gambar II-2. Tiga Jenis Bunga Iris .....	II-10
Gambar II-3. Bagian-bagian Bunga.....	II-10
Gambar III-1. Diagram Tahapan Penelitian.....	III-2
Gambar III-2. Gannt Chart-1.....	III-15
Gambar III-3. Gannt Chart-2.....	III-16
Gambar III-4. Gannt Chart-3.....	III-17
Gambar III-5. Gannt Chart-4.....	III-18
Gambar III-6. Gannt Chart-5.....	III-19
Gambar IV-1. Diagram <i>Use Case</i> .....	IV-6
Gambar IV-2. Diagram aktivitas <i>Use Case</i> membuka File.....	IV-11
Gambar IV-3. Diagram aktivitas <i>Use case</i> Klasifikasi <i>Naive Bayes</i> .....	IV-12
Gambar IV-4. Diagram aktivitas <i>Use case</i> Menentukan Jenis Atribut....	IV-12
Gambar IV-5. Diagram aktivitas <i>Use case</i> Melakukan Validasi.....	IV-13
Gambar IV-6 . Rancangan Antarmuka Klasifikasi <i>Naive Bayes</i> .....	IV-14
Gambar IV-7. <i>Sequence Diagram Use Case</i> Membuka File.....	IV-17
Gambar IV-8. Sequence Diagram <i>Use Case</i> Klasifikasi <i>Naive Bayes</i> .....	IV-18
Gambar IV-9. <i>Sequence Diagram Use Case</i> Menentukan Jenis Atribut ..	IV-19
Gambar IV-10. <i>Sequence Diagram Use Case</i> Melakukan Validasi .....	IV-20
Gambar IV-11. Diagram Kelas Klasifikasi <i>Naive Bayes</i> .....	IV-22
Gambar IV-12. Antarmuka Main .....	IV-24
Gambar IV-13. Hasil Klasifikasi <i>Naive Bayes</i> .....	IV-25

## **DAFTAR LAMPIRAN**

	Halaman
L.1.1. Hasil Koding.....	L-1

## **BAB I PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

### **1.2 Latar Belakang Masalah**

Data mining atau *Knowledge Discovery in Database* (KDD) sering digunakan untuk menjelaskan proses pencarian juga penggalian informasi yang tersembunyi dalam suatu basis data yang besar. Meskipun konsepnya berbeda namun kedua saling berkaitan. Salah satu tahapan dalam keseluruhan proses KDD juga termasuk data mining (Kusrini, 2009). Data mining merupakan analisa terhadap data (biasanya berukuran besar) untuk mencari hubungan yang jelas serta mencari kesimpulan yang belum diketahui sebelumnya dengan cara terkini yang mudah dipahami dan berguna bagi pengguna.

Terdapat beberapa teknik dalam data mining salah satunya adalah klasifikasi. Klasifikasi merupakan teknik yang paling umum dipakai dalam data mining. Klasifikasi adalah sebuah proses untuk mencari model atau fungsi yang menjelaskan dan membedakan kelas atau konsep dari data, dengan tujuan untuk menggunakan model dan melakukan prediksi dari kelas suatu objek dimana tidak diketahui label dari kelas tersebut (Prasetyo, 2019).

Algoritma yang paling sering dipakai dalam teknik klasifikasi adalah algoritma *Naive Bayes*. Klasifikasi *Naive Bayes* mengklasifikasikan data berbasis probabilitas dan statistik yang dikemukakan oleh Thomas Bayes dengan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Taheri et al., 2014). Metode *Naive Bayes* mudah diimplementasikan dan dalam perhitungan sederhana memiliki tingkat akurasi yang tinggi, juga hanya membutuhkan data pelatihan (*training data*) dalam jumlah kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan waktu *training data* yang lebih cepat.

Beberapa penelitian terdahulu mengenai klasifikasi telah dilakukan diantaranya adalah Klasifikasi Tebu Dengan Menggunakan Algoritma *Naive Bayes Classification* pada Dinas Kehutanan dan Perkebunan Pati (Anandita, 2014). Penelitian tersebut bertujuan untuk menerapkan algoritma *Naive Bayes Classification* yaitu perhitungan probabilitas statistik dengan nilai 1 sebagai nilai tertinggi dengan melihat dari pengklasifikasian tebu yang dilakukan berdasarkan kelas produktif atau tidak dengan uji validasi *Confussion Matriks* dengan kinerja sebesar 73,3% yang disebabkan oleh banyaknya data kontinu dengan data diskrit.

Penelitian terkait lainnya mengenai Klasifikasi Peminatan Siswa SMA Menggunakan Metode *Naive Bayes* (Naparin, 2016). Penelitian tersebut menggunakan *k-fold cross validation* untuk menentukan akurasi *Naive Bayes* dalam menentukan peminatan siswa SMA. Dalam penelitian tersebut ditemukan bahwa dari 379 data, *10-fold cross validation* memberikan hasil akurasi terbaik sebesar 99,47%.

Berdasarkan penelitian diatas, penelitian ini akan menganalisis pengaruh karakteristik data terhadap hasil akurasi klasifikasi algoritma *Naive Bayes*.

### **1.3 Perumusan Masalah**

Perumusan masalah dari penelitian ini adalah :

1. Bagaimana mekanisme algoritma *Naive Bayes* dalam klasifikasi data?
2. Adakah pengaruh tipe data dan jumlah atribut pada performa algoritma *Naive Bayes*?

### **1.4 Tujuan Penelitian**

Tujuan dari dilakukannya penelitian ini adalah

1. Memahami algoritma Naive Bayes dalam klasifikasi data
2. Membuat perangkat lunak yang dapat melakukan klasifikasi data dengan metode Naïve Bayes dan mengetahui bagaimana mekanisme performa algoritma *Naive Bayes* terhadap tipe data ataupun jumlah atribut yang berbeda.

### **1.5 Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini adalah mempermudah dalam melakukan klasifikasi data dengan metode Naïve Bayes juga dapat dijadikan rujukan untuk penelitian terkait.

## 1.6 Batasan Masalah

1. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari UCI Repository (<http://archive.ics.uci.edu/ml>) berupa data *soybean*, *iris* dan *ionosphere*.
2. Dataset yang digunakan berekstensi \*.csv.

## 1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti sistematika penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya antara lain :

### **BAB I. PENDAHULUAN**

Pada bab ini membahas tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

### **BAB II. TINJAUAN PUSTAKA**

Pada bab ini membahas tentang dasar-dasar teori yang digunakan dalam penelitian, seperti definisi-definisi sistem informasi, algoritma penelitian dan jenis-jenis pengembangan perangkat lunak.

### **BAB III. METODOLOGI PENELITIAN**

Pada bab ini membahas tentang tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi tentang perencanaan manajemen proyek pada pelaksanaan penelitian.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan dibahas mengenai perancangan, lingkungan implementasi, dan hasil pengujian perangkat lunak klasifikasi *Naive Bayes*.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini

## **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan pengembangan perangkat lunak klasifikasi *Naive Bayes* ini.

### **1.8 Kesimpulan**

Berdasarkan uraian di atas, maka dalam penelitian ini akan dikembangkan sebuah perangkat lunak untuk melakukan klasifikasi dataset menggunakan algoritma *Naive Bayes*.

## DAFTAR PUSTAKA

- Anandita, E. R. (2014). Klasifikasi Tebu Dengan Menggunakan Algoritma Naive Bayes Clasification pada Dinas Kehutanan dan Perkebunan Pati. *Ilmu Komputer, vol.1*, 1–13.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah, 10(2)*, 160–165.  
<https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Han, J., Kamber, M., & Pei, J. (2012). Introduction. In *Data Mining*.  
<https://doi.org/10.1016/b978-0-12-381479-1.00001-0>
- Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, 1(1)*, 21–2014. [www.bluefame.com](http://www.bluefame.com),
- Kusrini. (2009). *Algoritma Data Mining* (T. A. Prabawati (ed.)). Penerbit Andi.  
<https://books.google.co.id/books?id=-Ojclag73O8C&printsec=frontcover#v=onepage&q&f=false>
- Larose, D. T. (2006). Data Mining Methods and Models. In *Data Mining Methods and Models*. <https://doi.org/10.1002/0471756482>
- Naparin, H. (2016). Klasifikasi Peminatan Siswa SMA Menggunakan Metode Naive Bayes. *Systemic: Information System and Informatics Journal, 2(1)*, 25–32. <https://doi.org/10.29080/systemic.v2i1.104>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN.

*Computer Engineering, Science and System Journal*, 4(1), 78.

<https://doi.org/10.24114/cess.v4i1.11458>

Prasetyo, E. (2019). *Data mining mengolah data menjadi informasi menggunakan matlab.*

Taheri, S., Yearwood, J., Mammadov, M., & Seifollahi, S. (2014). *Attribute weighted Naive Bayes classifier using a local optimization*. 0941–0643.

<https://doi.org/10.1007/s00521-012-1329-z>

Turban, E., Aronson, J., & Llang, T. (2003). Decision Support Systems and Intelligent Systems. *Decision Support Systems and Intelligent Systems*.