

Perbandingan Algoritma *Naïve Bayes* dan Algoritma C4.5 Dalam Klasifikasi Data

*Diajukan untuk Menyusun Tugas Akhir
di Jurusan Teknik Informatika Fakultas Ilmu Komputer Unsri*



Oleh:

Lidya A.Panjaitan

09021181320050

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2019

LEMBAR PENGESAHAN TUGAS AKHIR

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN ALGORITMA C4.5
DALAM KLASIFIKASI DATA

Oleh :

Lidya A. Panjaitan
NIM : 09021181320050

Palembang, Desember 2019

Pembimbing I

Pembimbing II

Rifkie Primartha, M.T
NIP. 19770601200912004

Dagny Mathew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

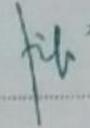
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 15 November 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Lidya A. Panjaitan
NIM : 09021181320050
Judul : Perbandingan Algoritma Naïve Bayes dan C4.5 Pada Klasifikasi Data

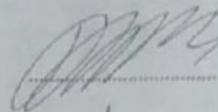
1. Pembimbing

Rifkie Primartha, M.T.
NIP. 19770601200912004



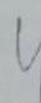
2. Pembimbing II

Danny Mathew Saputra, M.Sc
NIP. 198505102015041002



3. Penguji I

Yunita, M.Cs
NIP. 198306062015042002



4. Penguji II

Desty Rodiah, M.T
NIP.



Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 19770601200912004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Lidya A. Panjaitan
NIM : 09021181320050
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Algoritma Naïve Bayes dan C4.5
Pada Klasifikasi Data

Hasil Pengecekan Software *iThenticate/Turnitin* : 18 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek 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, Desember 2019



Lidya A. Panjaitan
NIM. 09021181320050

Motto :

- *When big problems come to you, remember that there is still a greater God than your problem who can help you*
- *Just try, maybe there will be a valuable opportunity coming to you. Believe it*

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer Universitas Sriwijaya*

COMPARISON OF NAÏVE BAYES ALGORITHM AND C4.5 ALGORITHM
FOR DATA CLASSIFICATION

By:

LIDYA A. PANJAITAN

09021181320050

ABSTRACT

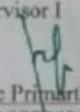
Data Mining is the analysis of the large data to find a clear correlation and conclude that is not known in advance by the current and easy way to understand and useful for the data owner. Classification is the process of grouping a variable into a class that has been determined. The study is comparing the Naïve Bayes algorithm and C4.5 algorithm to classify the data. Naïve Bayes algorithm is an algorithm that is based on probability, while C4.5 algorithm is an algorithm that is based on a tree in the establishment of the rule. This study produces a comparison of the results of accuracy and the execution time of each algorithm in carrying out the classification process of the data that has been determined. The study obtained the value of average accuracy of comparison to both methods, is Naïve Bayes methods at the Split Validation process in the amount of 92.89% and on the Cross Validation process amount of 91.73%. While for time execution of comparison of the two algorithms classification result in Naïve Bayes methods much faster than C4.5.

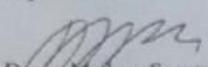
Keyword: Data Mining, Classification, Naïve Bayes, C4.5, Confussion Matrix

Indralaya, Desember 2019

Supervisor I

Supervisor II


Rifkie Primartha, M.T.
NIP. 19770601200912004


Dany Mathew Saputra, M.Sc
NIP. 198505102015041002

Approval,
Head of Department of Informatics Engineering


Rifkie Primartha, M.T.
NIP. 19770601200912004

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN C4.5 PADA
KLASIFIKASI DATA

Oleh:

LIDYA A. PANJAITAN

09021181320050

ABSTRAK

Data Mining adalah analisis terhadap data yang berukuran besar untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dan mudah dipahami dan berguna bagi pemilik data. Klasifikasi merupakan proses pengelompokan sebuah variabel kedalam kelas yang sudah ditentukan. Penelitian ini membandingkan algoritma *Naive Bayes* dan algoritma C4.5 untuk mengklasifikasi data. Algoritma *Naive Bayes* merupakan algoritma yang berbasis probabilitas, sedangkan algoritma C4.5 merupakan algoritma yang berbasis *tree* dalam pembentukan *rule*. Penelitian ini menghasilkan perbandingan hasil akurasi dan waktu eksekusi dari masing-masing algoritma dalam menjalankan proses klasifikasi terhadap data yang telah ditentukan. Pada pengujian ini didapatkan nilai rata-rata akurasi dari perbandingan kedua metode, yaitu metode *Naive Bayes* pada proses *Split Validation* sebesar 92,89% dan pada proses *Cross Validation* sebesar 91,73%. Sedangkan rata-rata waktu untuk waktu eksekusi dari perbandingan klasifikasi kedua algoritma menghasilkan metode *Naive Bayes* lebih cepat dibandingkan C4.5.

Kata Kunci: *Data Mining*, Klasifikasi, *Naive Bayes*, C4.5, *Confusion Matrix*

Indralaya, Desember 2019

Pembimbing I

Rifkie Priandita, M.T.
NIP. 19770601200912004

Pembimbing II

Danny Mathew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Priandita, M.T
NIP. 19770601200912004

vi

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini kepada:

1. Orangtua saya, Ir.Lammer Panjaitan dan Derismala Sinaga beserta saudara-saudara saya, Emma Panjaitan, Daniel panjaitan, Dian Panjaitan dan Josua Panjaitan yang selalu mendoakan dan memberikan dukungan baik moril ataupun materil. Terimakasih.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika dan sebagai pembimbing Tugas Akhir yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Danny Matthew Saputra, M.Sc. selaku dosen pembimbing yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses pengerjaan tugas akhir.
5. Yunita, M.Cs selaku dosen penguji I yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Desty Rodiah, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Seluruh staf administrasi dan pegawai yang telah membantu dan mendukung Penulis.
9. Kepada teman-teman seangkatan saya Agung 013 yang selalu memberikan dukungan.
10. Kepada Mantan serumah saya yaitu Rumah Marpen ada kak Desymel, kak Zetly, kak Octa, kak Laura, kak Ica, Novita, adek Fitri, aku mengasihi kalian gaiss.

11. Kepada adek-adek bedeng Nikson, Sutri, Fanti, Sofia, Vero, Friska, Tiffany, Vinny, Lamtiur, Friski dan Richmen dll yang selalu memberi tawa.
12. Kepada team olahraga saya setiap sore MIB yang juga membuat saya semangat dan tertawa, terimakasih gais.
13. Semua pihak yang telah membantu saya dalam penyelesaian tugas akhir ini dan tidak dapat disebutkan satu persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Desember 2019

Lidya A. Panjaitan

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
DAFTAR ISI.....	ii
DAFTAR TABEL.....	vi
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xii
BAB I.....	I – 1
1.1 Pendahuluan	I – 1
1.2 Latar Belakang	I – 1
1.3 Rumusan Masalah	I – 4
1.4 Tujuan Penelitian	I – 5
1.5 Manfaat Penelitian	I – 5
1.6 Batasan Masalah.....	I – 5
BAB II	II – 1
2.1 Pendahuluan	II – 1
2.2 Landasan Teori	II – 1
2.2.1 Data Mining	II – 1
2.2.2 Klasifikasi (<i>Classification</i>)	II – 2
2.2.3 Naïve Bayes	II – 3
2.2.4 Algoritma C4.5.....	II – 8
2.2.5 Cross Validation	II – 11

2.2.6 Split Validation	II – 13
2.2.7 Confusion Matrix	II – 13
2.2.8 Data yang Digunakan	II – 15
2.2.8.1 Data Iris	II – 15
2.2.8.2 Blood Transfusion	II – 17
2.2.8.3 Weather Nominal	II – 17
2.2.9 Penelitian Terkait	II – 18
BAB III	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 Dan Unit Penelitian	III – 2
3.4.2 Menentukan Dasar Teori Yang Berkaitan Dengan Penelitian ..	III – 3
3.4.3 Menetapkan Kriteria Pengujian	III – 3
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III – 9
3.4.5 Melakukan Pengujian Penelitian	III – 9
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian	III – 10
3.5 Metode Pengembangan Perangkat Lunak	III – 10
3.5.1 Fase Insepsi	III – 11
3.5.3 Fase Elaborasi	III – 11

3.5.4 Fase Konstruksi	III – 11
3.5.5 Fase Transisi	III – 12
3.6 Penjadwalan Penelitian	III – 12
BAB IV	IV – 1
4.1 Pendahuluan	IV – 1
4.2 Fase Inseepsi	IV – 1
4.2.1 Pemodelan Bisnis	IV – 2
4.2.2 Kebutuhan	IV – 3
4.2.3 Analisis Dan Desain	IV – 5
4.2.3.1 Analisis Perangkat Lunak	IV – 5
4.2.4 Implementasi	IV – 6
4.3 Fase Elaborasi	IV – 17
4.3.1 Pemodelan Bisnis	IV – 17
4.3.1.1 Perancangan Data	IV – 17
4.3.1.2 Perancangan Antar Muka	IV – 17
4.3.2 Kebutuhan	IV – 18
4.3.3 Diagram Sequence	IV – 18
4.3.4 Implementasi	IV – 22
4.4 Fase Konstruksi	IV – 22
4.4.1 Kebutuhan	IV – 22
4.4.2 Analisa dan Desain.....	IV – 23
4.4.3 Implementasi	IV – 25
4.4.3.1 Implementasi Kelas	IV – 25

4.4.3.2 Implementasi Antarmuka	IV – 27
4.4.3.3 Implementasi Perancangan Data	IV – 30
4.4.4 Pengujian	IV – 33
4.4.4.1 Rencana Pengujian White Box	IV – 33
4.4.4.2 Pengujian White Box	IV – 34
4.5 Fase Transisi	IV – 38
4.5.1 Pemodelan Bisnis	IV – 38
4.5.2 Kebutuhan	IV – 38
4.5.3 Analisa dan Desain	IV – 38
4.5.4 Implementasi	IV – 40
4.5.5 Pengujian (Black Box)	IV – 42
BAB V.....	V – 1
5.1 Pendahuluan	V – 1
5.2 Data Hasil Pengujian.....	V – 1
5.2.1 Pembagian Data K-Fold Cross Validation.....	V – 1
5.2.2 Hasil Pengujian Metode Naïve Bayes.....	V – 2
5.2.3 Hasil Pengujian Metode C4.5	V – 5
5.2.4 Analisis Hasil Akurasi Confusion Matrix Pada Perbandingan Pengujian Metode Naive Bayes dan Metode C4.5	V – 7
5.2.5 Analisis Perbandingan Hasil Waktu Komputasi Pengujian Metode Naive bayes dan Metode C4.55	V – 10
BAB VI.....	VI – 1
6.1 Pendahuluan	VI – 1

6.2 Kesimpulan VI – 1

6.3 Saran VI – 2

DAFTAR PUSTAKA vii

DAFTAR GAMBAR

	Halaman
Gambar II – 1 Flowchart C4.5	II – 11
Gambar II – 2 Tiga Jenis Bunga Iris	II – 16
Gambar II – 3 Bagian – Bagian Bunga	II – 16
Gambar III – 2. Flowchart C4.5	II – 14
Gambar III – 1. Diagram Tahapan Penelitian	III – 2
Gambar III – 2. Flowchart Naïve Bayes	III – 6
Gambar III – 3. Flowchart C4.5	III – 8
Gambar IV – 1. Diagram Use Case	IV – 6
Gambar IV – 2. Diagram Aktivitas Klasifikasi dengan C4.5	IV – 13
Gambar IV – 3. Diagram Aktivitas Klasifikasi dengan Naïve Bayes.....	IV – 14
Gambar IV – 4. Diagram Aktivitas Import Data	IV – 15
Gambar IV – 5. Diagram Aktivitas Menentukan Jenis Atribut Data.....	IV – 15
Gambar IV – 6. Diagram Aktivitas Melakukan Split Validation	IV – 16
Gambar IV – 7. Diagram Aktivitas Melakukan Cross Validation.....	IV – 16
Gambar IV – 8. Diagram Sequence Import Data.....	IV – 18
Gambar IV – 9. Diagram Sequence Menentukan Jenis Atribut Data	IV – 19
Gambar IV – 10. Diagram Sequence Melakukan Split Validation	IV – 19
Gambar IV – 11. Diagram Sequence Melakukan Cross Validation	IV – 20

Gambar IV – 12. Diagram Sequence C4.5.....	IV – 21
Gambar IV – 13. Diagram Sequence Naïve Bayes	IV – 22
Gambar IV – 14. Diagram Kelas.....	IV – 24
Gambar IV – 15. Rancangan Antarmuka Menu Utama.....	IV – 27
Gambar IV – 16. Rancangan Utama Antarmuka Menu Dashboard	IV – 28
Gambar IV – 17. Rancangan Utama Antarmuka Menu Pengaturan Data	IV – 29
Gambar IV – 18. Rancangan Utama Antarmuka Menu Telusuri Data	IV – 30
Gambar IV – 19. Antarmuka Menu Utama.....	IV – 40
Gambar IV – 20. Antarmuka Menu Dashboard	IV – 41
Gambar IV – 21. Antarmuka Menu Pengaturan Data.....	IV – 41
Gambar IV – 22. Antarmuka Menu Telusuri Data	IV – 42
Gambar V – 1. Perbandingan waktu komputasi proses split validation	V – 11
Gambar V – 2. Perbandingan waktu komputasi proses cross validation	V – 11

DAFTAR TABEL

	Halaman
Tabel II – 1 Tabel Confussion Matrix.....	II – 13
Tabel III – 1 Tabel Hasil Perbandingan Metode.....	III – 10
Tabel III – 2 Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III – 12
Tabel IV – 1 Kebutuhan Fungsional	IV – 4
Tabel IV – 2 Kebutuhan Non-Fungsional.....	IV – 4
Tabel IV – 3 Defenisi Aktor.....	IV – 7
Tabel IV – 4 Deskripsi Use Case	IV – 7
Tabel IV – 5 Skenario Use Case Klasifikasi dengan C4.5.....	IV – 8
Tabel IV – 6 Skenario Use Case Klasifikasi Naïve Bayes.....	IV – 9
Tabel IV – 7 Skenario Use Case Import Data.....	IV – 10
Tabel IV – 8 Skenario Use Case Menentukan Jenis Atribut Data	IV – 11
Tabel IV – 9 Skenario Use Case Split Validation.....	IV – 12
Tabel IV – 10 Skenario Use Case Cross Validation	IV – 12
Tabel IV – 11 Implementasi Kelas.....	IV – 25
Tabel IV – 12 Tabel Data Iris	IV – 30
Tabel IV – 13 Tabel Data Transfusion.....	IV – 31
Tabel IV – 14 Tabel Data Weather Nominal	IV – 32
Tabel IV – 15 Rencana Pengujian White Box Klasifikasi C4.5	IV – 33
Tabel IV – 16 Rencana Pengujian White Box Klasifikasi Naïve Bayes.....	IV – 33

Tabel IV – 17 Rencana Pengujian White Box Import Data.....	IV – 34
Tabel IV – 18 Rencana Pengujian White Box Memilih Jenis Atribut Data	IV – 34
Tabel IV – 19 Rencana Pengujian White Box Melakukan Validasi	IV – 34
Tabel IV – 20 Pengujian White Box Klasifikasi dengan C4.5	IV – 35
Tabel IV – 21 Pengujian White Box Klasifikasi dengan Naïve Bayes.....	IV – 35
Tabel IV – 22 Pengujian White Box Import Data.....	IV – 36
Tabel IV – 23 Pengujian White Box Menentukan Jenis Atribut Data.....	IV – 36
Tabel IV – 24 Pengujian White Box Melakukan Validasi.....	IV – 37
Tabel IV – 25 Rencana Pengujian Black Box Klasifikasi C4.5.....	IV – 39
Tabel IV – 26 Rencana Pengujian Balck Box Klasifikasi Naïve Bayes	IV – 39
Tabel IV – 27 Rencana Pengujian Black Box Mengatur Data.....	IV – 39
Tabel IV – 28 Pengujian Black Box Klasifikasi dengan C4.5.....	IV – 43
Tabel IV – 29 Pengujian Black Box Klasifikasi dengan Naïve Bayes	IV – 44
Tabel IV – 30 Pengujian Black Box Import Data	IV – 46
Tabel IV – 31 Pengujian Black Box Menentukan Jenis Atribut Data	IV – 46
Tabel IV – 31 Pengujian White Box Melakukan Validasi.....	IV – 47
Tabel V – 1 Pembagian Data Latih dan Data Uji.....	V – 2
Tabel V – 2 Hasil Akurasi dari Pengujian Confussion Matrix Pada Proses Split Validation.....	V – 3
Tabel V – 3 Hasil Akurasi dari Pengujian Confussion Matrix Pada Proses Cross Validation.....	V – 4
Tabel V – 4 Hasil Akurasi dari Pengujian Confussion Matrix Pada Proses Split Validation.....	V – 5

Tabel V – 5 Hasil Akurasi dari Pengujian Confussion Matrix Pada Proses Cross Validation.....	V – 6
Tabel V – 6 Lama Waktu Komputasi Pada Proses Split Validation Metode Naïve Bayes.....	V – 7
Tabel V – 7 Lama Waktu Komputasi Pada Proses Split Validation Pada Metode C4.5	V – 8
Tabel V – 8 Lama Waktu Komputasi Pada Proses Cross Validation Metode Naïve Bayes.....	V – 8
Tabel V – 9 Lama Waktu Komputasi Pada Proses Split Validation Pada Metode C4.5	V – 9
Tabel V – 10 Hasil Rata-rata Waktu Komputasi Proses Split Validation.....	V – 10
Tabel V – 11 Hasil Rata-rata Waktu Komputasi Proses Cross Validation.....	V – 10

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah. Pada bab ini juga menjelaskan secara umum tentang keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan singkat mengenai pengertian data mining dan penjelasan tentang algoritma C4.5 dan *Naïve Bayes* yang digunakan untuk mengklasifikasi beberapa data dengan melakukan perbandingan kedua algoritma tersebut.

1.2 Latar Belakang

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dari sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Larose, 2005). Menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT data mining adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini yang mudah dipahami dan berguna bagi pemilik data tersebut. Data mining juga merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Salah satu teknik data mining adalah teknik klasifikasi. Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasi suatu spesies tertentu, seperti yang dilakukan oleh Carolus von Linne (atau dikenal dengan

nama Carolus Linnaeus) yang pertama kali mengklasifikasi spesies berdasarkan karakteristik fisik. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Beberapa algoritma yang digunakan dalam teknik klasifikasi adalah algoritma C4.5 dan *Naïve Bayes*. Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Adapun kelebihannya yaitu dapat mengolah data numerik (*kontinyu*) dan *diskret*, dapat menangani nilai atribut yang hilang menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut aktor dan kategorial (Khoirunnisa, A, *et al.*, 2016).

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip dengan yang lainnya. Proses pada pohon keputusan adalah mengubah bentuk data menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*.

Sedangkan klasifikasi *Naïve Bayes* merupakan pengklasifikasian berbasis probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris . Dengan kata lain, dalam *Naïve Bayes* model yang digunakan adalah model fitur independen. Dalam *Naïve Bayes* maksud independensi adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Metode ini memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana (Kang,*et al.*, 2006), mudah diimplementasikan, metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan waktu *training* yang lebih cepat dibandingkan metode lainnya. Adapun kekurangan dari *naïve bayes* ialah tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga.

Algoritma *Naïve Bayes* bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Model statistik merupakan salah satu model terpercaya dan sangat andal sebagai pendukung pengambilan keputusan. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain.

Beberapa penelitian terkait sebelumnya mengenai klasifikasi telah dilakukan, diantaranya ialah *Accuracy Evaluation Of C4.5 And Naïve Bayes Classifiers Using Attribute Ranking Method* (Sivakumari, S., etc 2009). Penelitian tersebut bertujuan mengklasifikasi set data kanker payudara menggunakan pohon keputusan C4.5 dan *Naïve Bayes*, hasilnya menunjukkan bahwa pada pengklasifikasian C4.5 menghasilkan hasil yang sangat akurat dalam

mempertimbangkan atribut yang menjadi gain tertinggi pada klasifikasi setelah dilakukan beberapa kali validasi silang. Sementara kinerja Naïve Bayes tidak menunjukkan beberapa perbaikan yang signifikan.

Penelitian terkait lainnya mengenai Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naïve Bayes Dan Algoritma C4.5 Untuk Prediksi Ketepatan Waktu Kelulusan (F, Gian. 2013), berdasarkan hasil penelitian yang telah dilakukan tingkat kelulusan mahasiswa dengan menggunakan pemodelan Decision Tree didapatkan tingkat akurasi tertinggi sebesar 82.43%, sedangkan pada pemodelan Naïve Bayes memperoleh rata-rata tingkat akurasi tertinggi sebesar 74.09%. Oleh karena itu dapat disimpulkan berdasarkan tingkat akurasi, bahwa pemodelan metode Decision Tree lebih baik dalam melakukan prediksi ketepatan kelulusan mahasiswa pada data penelitian mahasiswa strata 1 (S1) Fakultas Ilmu Komputer Universitas Dian Nuswantoro.

Berdasarkan uraian diatas, penelitian ini akan menganalisis perbandingan metode C4.5 dan *Naïve Bayes* dalam klasifikasi data untuk mengetahui metode mana yang menghasilkan klasifikasi terbaik.

1.3 Perumusan Masalah

Perumusan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana model metode C4.5 dan algoritma *Naïve Bayes* untuk klasifikasi data.
2. Bagaimana tingkat keakuratan dan waktu eksekusi yang didapat dari kedua metode tersebut dalam klasifikasi data.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membandingkan metode C4.5 dan *Naïve Bayes* dari segi akurasi dan waktu eksekusi.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah mengetahui performa perbandingan antara metode C4.5 dan *Naïve Bayes* dalam melakukan klasifikasi.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data yang digunakan sebagai obyek penelitian adalah jenis data sekunder berupa data iris, data *blood transfusion*, data *weather nominal* yang diperoleh dari UCI Repository (<http://archive.ics.uci.edu/ml/>).
2. Hasil klasifikasi digunakan sebagai pembandingan antara metode C4.5 dan *Naïve Bayes*.

DAFTAR PUSTAKA

- Assiroj, P. (2018). KAJIAN PERBANDINGAN TEKNIK KLASIFIKASI ALGORITMA C4 . 5 , NAÏVE BAYES DAN CART UNTUK PREDIKSI KELULUSAN MAHASISWA (STUDI KASUS : STMIK ROSMA KARAWANG) (STUDI KASUS : STMIK ROSMA KARAWANG), (March). <https://doi.org/10.5281/zenodo.1184054>
- Bustami. (2014). Penerapan Algoritma Naive Bayes, 8(1), 884–898.
- Dan, A., Perbandingan, I., & Dengan, A. C. (2016). Analisis dan implementasi perbandingan algoritma c.45 dengan naïve bayes untuk prediksi penawaran produk, 3(3), 5029–5035.
- Fitria, A., & Widowati, H. (2017). Implementasi metode rational unified process dalam pengembangan sistem administrasi kependudukan. *Jurnal Teknologi Rekayasa*, 22, 27–36.
- Goyal, A., & Mehta, R. (2012). Performance comparison of Naïve Bayes and J48 classification algorithms. *International Journal of Applied Engineering Research*, 7(11 SUPPL.), 1389–1393.
- Huda, N. M. (2010). Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro). *Universitas Stuttgart*, 1(1), 1–87.
- Iskandar, D., & Suprpto, Y. K. (2013). Perbandingan akurasi klasifikasi tingkat kemiskinan antara algoritma C4 . 5 dan Naïve Bayes Clasifier, 11(1), 14–17.

- Khadafy, A. R. (2015). Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree, *1*(2), 136–142.
- Pada, O. L. I., & Roda, K. (2017). Prosiding SNST ke-8 Tahun 2017 Fakultas Teknik Universitas Wahid Hasyim Semarang 1, 1–6.
- Prakoso, S. A., & Tutik, E. T. (2017). Komparasi algoritma C4.5 dengan naive bayes untuk klasifikasi kelulusan mahasiswa tepat waktu di PTS “KZX.” *Komputaki*, *3*(1), 103–132.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *7*(1), 59–64.
- Saleh, A., & Potensi, U. (2016). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, (June).
- Singh, G., Bagwe, K., Shanbhag, S., Singh, S., & Devi, S. (2017). Heart disease prediction using Naïve Bayes. *International Research Journal of Engineering and Technology(IRJET)*, *4*(3), 4–6. Retrieved from <https://irjet.net/archives/V4/i3/IRJET-V4I3212.pdf>
- Sivakumari, S. (2009). ACCURACY EVALUATION OF C4 . 5 AND NAÏVE BAYES CLASSIFIERS USING ATTRIBUTE RANKING METHOD Received : 15-05-2008 Revised : 21-01-2009, *2*(1), 60–68.
- Thomas Wiga Heru Prasetya. (2016). Naive Bayesian Classifier Method

Implementation for Diabetes Melitus Diagnose Classification,
(125314074).

Yuli, M. (2017). Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika. *Jurnal Edik Informatika*, 2(2), 213–219.

Zhang J, Kang D-K, Silvescu A, Honavar V (2006) Learning accurate and concise naïve Bayes classifiers from attribute value taxonomies and data. *Knowl Inf Syst* 9(2):157–179

Zhang, S. and R. Jing (2011). "Dimension Reduction Based on Modified Maximum Margin Criterion for Tumor Classification." Fourth International Conference on Information and Computing (ICIC): 552-554.

Zul, M. I. (2013). Perbandingan Akurasi k-NN dan Naive Bayes untuk Algoritma Sistem Prediksi Nilai Akhir Mahasiswa. Riau: Jurusan Komputer Politeknik Caltex Riau Pekanbaru.