

PENGARUH PROSES DISKRITISASI FAYYAD IRANI PADA AKURASI DATA MINING KLASIFIKASI NAÏVE BAYES

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Ahmad Halim Prabowo
09021281520112

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

2019

LEMBAR PENGESAHAN TUGAS AKHIR

PENGARUH PROSES DISKRITISASI FAYYAD IRANI PADA AKURASI DATA MINING KLASIFIKASI NAÏVE BAYES

Oleh :

AHMAD HALIM PRABOWO
NIM : 09021281520112

Palembang, 29 Juli 2019

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004

Pembimbing,



Rifkie Primartha, M.T.
NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 23 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Ahmad Halim Prabowo
N I M : 09021281520112
Judul : Pengaruh Proses Diskritisasi Fayyad Irani pada Akurasi Data Mining Klasifikasi Naïve Bayes

1. Pembimbing

Rifkie Primartha, M.T.
NIP. 197706012009121004


.....

2. Penguji I

Yunita, M.Cs.
NIP. 198306062015042002


.....

3. Penguji II

Desty Rodiah, M.T.
NIP. 1671016112890005


.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha S.T. M.T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Ahmad Halim Prabowo
NIM : 09021281520112
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh Proses Diskritisasi Fayyad Irani pada Akurasi
Data Mining Klasifikasi Naïve Bayes
Hasil Pengecekan Software *iThenticate/Turnitin* : 14%

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.

Palembang, 29 Juli 2019



Ahmad Halim Prabowo
NIM. 09021281520112

KATA PENGANTAR



Puji syukur kepada Allah SWT 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 guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah Subhanahu wa Ta'ala yang telah memberikan hamba keimanan, kesehatan, kecerdasan, kemudahan, dan kelancaran sehingga hamba dapat menyelesaikan tugas-tugas hamba sebagai seorang mahasiswa.
2. Nabi Muhammad Shallallahu 'Alaihi wa Sallam yang telah memberikan jalan yang terang dan memberikan ilmu yang bermanfaat kepada umatnya atas seizin Allah.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya dan sekaligus pembimbing yang telah mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir serta Ibu Hardini Novianti, M.T. selaku Sekretaris Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Kanda Januar, M.T. selaku Pembimbing Akademik pada Jurusan Teknik Informatika Universitas Sriwijaya yang selalu memotivasi dan membantu saya.
6. Ibu Yunita, M.Cs. selaku dosen penguji I dan Ibu 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. Orang tua saya, Bagiyo Budi S. dan Lismarita yang telah memberikan dorongan dan dukungan baik secara moril maupun materil, serta kepada adik-adik saya Rafif Imtisaldin dan Khairunisa Rahma Yani.
9. Mbak Winda dan Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Faiz Muhammad, Arief Rachmatullah, Gian Agnar, Mega Rizky, Taca Rosa, Arfah Anggraina, Ikrom Izzudin, M. Fachri, Agung Pribadi, Rahmad Tirta, Kerel Khalif, Misbahudin Munadi, Ardi Wasila, Genda Ananta, Albar, Dian Ayu serta seluruh teman-teman jurusan Teknik Informatika baik angkatan 2015, HMIF dan DPM yang telah berbagi keluh kesah, motivasi, dan canda tawa di masa-masa perkuliahan ini.

11. Ali Zainal, Irfan, Agung, Rafif, Rizky, Susetyo, dan Hafiz sebagai rekan & teman *gaming* yang telah menemani penulis dalam penulisan Tugas Akhir.
12. Bulldog's Kappa Club sebagai server discord yang telah menghibur penulis selama proses penelitian ini.

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.

Palembang, 29 Juli 2019

Ahmad Halim Prabowo

DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR TABEL.....	v
DAFTAR GAMBAR	vii
BAB I.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Perumusan Masalah.....	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-4
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II.....	II-1
2.1 Pendahuluan	II-1
2.1 Penelitian Terkait	II-1
2.2 Naïve Bayes.....	II-2
2.3 <i>Data Preprocessing</i>	II-5
2.4 Diskritisasi.....	II-7
2.5 Diskritisasi Fayyad – Irani	II-8
2.6 Confusion Matrix	II-10
2.7 Dataset Avila	II-11
2.8 Kesimpulan.....	II-12
BAB III	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.3 Tahapan Penelitian	III-1
3.3.1 Menentukan Ruang Lingkup dan Unit Penelitian	III-2
3.3.2 Menemukan Dasar Teori yang Berkaitan dengan Penelitian	III-2

3.3.3 Menetapkan Kriteria Pengujian.....	III-2
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-4
3.3.5 Melakukan Pengujian Penelitian.....	III-4
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan.....	III-5
Penelitian	III-5
3.4 Metode Pengembangan Perangkat Lunak	III-7
3.4.1 Fase Insepsi.....	III-8
3.4.2 Fase Elaborasi	III-8
3.4.3 Fase Konstruksi.....	III-9
3.4.4 Fase Transisi	III-9
3.5 Penjadwalan Penelitian	III-10
3.6 Kesimpulan.....	III-17
BAB IV	IV-1
4.1 Pendahuluan	IV-1
4.2 Rational Unified Process	IV-1
4.2.1 Fase Insepsi.....	IV-2
4.2.2 Fase Elaborasi	IV-15
4.2.3 Fase Konstruksi.....	IV-27
4.2.4 Fase Transisi	IV-31
4.3 Kesimpulan.....	IV-39
BAB V	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan/Penelitian	V-1
5.2.1 Percobaan	V-1
5.2.2 Hasil Percobaan/Penelitian	V-1
5.3 Analisis Hasil	V-7
5.4 Kesimpulan.....	V-8
BAB VI	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-1

DAFTAR PUSTAKA	viii
----------------------	------

DAFTAR TABEL

Tabel II-1. Confusion Matrix	II-10
Tabel III-1. Rancangan Tabel Hasil Klasifikasi.....	III-5
Tabel III-2. Rancangan Tabel Perbandingan Akurasi.....	III-6
Tabel III-3. Rancangan Tabel Perbandingan Waktu Rata-Rata.....	III-7
Tabel III-4. Tabel Penjadwalan Penelitian dalam Bentuk Work Breakdown Structure (WBS).....	III-11
Tabel IV-1. Tabel Kebutuhan Fungsional	IV-4
Tabel IV-2 Tabel Kebutuhan Non-Fungsional	IV-4
Tabel IV-3. Tabel Definisi Aktor.....	IV-8
Tabel IV-4. Tabel Definisi Use Case	IV-8
Tabel IV-5. Tabel Skenario Use Case Mengklasifikasi Dataset dengan Naïve Bayes.....	IV-9
Tabel IV-6. Tabel Use Case Mendiskritisasi Data.....	IV-12
Tabel IV-7. Tabel Implementasi Kelas	IV-29
Tabel IV-8. Tabel Pengujian Use-Case Mengklasifikasikan Dataset.....	IV-32
Tabel IV-9. Tabel Pengujian Use-Case Mendiskritisasi Dataset.....	IV-33
Tabel IV-10 Tabel Hasil Pengujian Use-Case Mengklasifikasi Dataset	IV-34
Tabel IV-11 Tabel Hasil Pengujian Use-Case Mendiskritisasi Dataset	IV-37
Tabel V-1. Tabel 20 Data Hasil Klasifikasi Menggunakan Gaussian Naïve Bayes .	V-2
Tabel V-2. Tabel Data Confusion Matrix Menggunakan Gaussian Naïve Bayes	V-3
Tabel V-3. Tabel 30 Data Hasil Diskritisasi Data Latih	V-4
Tabel V-4. Hasil Klasifikasi Menggunakan Diskritisasi dan Naïve Bayes	V-5

Tabel V-5. Tabel Data Confusion Matrix Menggunakan Diskritisasi dan Naïve Bayes	V-6
Tabel V-6. Tabel Hasil Klasifikasi.....	V-7
Tabel V-7. Tabel Presisi dan Recall Metode Naive Bayes dengan Diskritisasi	V-7
Tabel V-8. Tabel Waktu Proses pada Lima Percobaan	V-8

DAFTAR GAMBAR

Gambar II-1. Pengelompokan Algoritma <i>Data Preprocessing</i> Paling Berpengaruh	II-6
Gambar III-1. Diagram Proses tanpa Diskritisasi	III-4
Gambar III-2. Diagram Proses dengan Diskritisasi	III-4
Gambar III-3. Diagram Tahapan Pengujian Penelitian.....	III-7
Gambar IV-1. Diagram Use Case	IV-7
Gambar IV-2. Diagram Aktivitas Mengklasifikasi Dataset dengan Naïve Bayes	IV-14
Gambar IV-3. Diagram Aktivitas Mendiskritisasi Data	IV-15
Gambar IV-4. Antarmuka Perangkat Lunak	IV-17
Gambar IV-5. Sequence Diagram Utama Pelatihan Model Algoritma Naive Bayes dan Gaussian Naive Bayes.....	IV-19
Gambar IV-6. Sequence Diagram Utama Pengujian Model Algoritma	IV-20
Gambar IV-7. Sequence Diagram Utama Diskritisasi	IV-21
Gambar IV-8 Sequence Diagram Latih Naive Bayes	IV-22
Gambar IV-9. Sequence Diagram Latih Gaussian Naive Bayes	IV-23
Gambar IV-10. Sequence Diagram Naive Bayes Uji	IV-24
Gambar IV-11. Sequence Diagram Gaussian Naive Bayes Uji.....	IV-25
Gambar IV-12. Sequence Diagram Latih Diskritisasi	IV-26
Gambar IV-13. Diagram Kelas Perangkat Lunak.....	IV-28
Gambar IV-14. Implementasi Antarmuka	IV-30

DAFTAR LAMPIRAN

Lampiran 1. Kode Program	L-1
--------------------------------	-----

PENGARUH PROSES DISKRITISASI FAYYAD IRANI PADA AKURASI
DATA MINING KLASIFIKASI NAÏVE BAYES

Oleh :
Ahmad Halim Prabowo
09021281520112

ABSTRACT

Data Mining Classification is a technique to classify data into a category or class. The classification process can have varying accuracy depending on the compatibility between the data and the algorithm used. One way to improve the accuracy of the classification process is to use data processing. Fayyad-Irani discretization is a pre-processing of data that converts numerical values to categorical. Fayyad-Irani's discretization has been proven to increase the accuracy of classifications using decision tree algorithms. Unlike the decision tree algorithm, the Naïve Bayes algorithm is a Bayesian probability based algorithm so that it can have a different effect on the accuracy and timing of the classification. Based on the experiments conducted in this study, it is known that Fayyad-Irani Discretization has succeeded in increasing the accuracy of the Naïve Bayes classification process significantly from only 29.31% to 93.26%, but with the processing time needed to be slower with time difference as much as 25.5 minutes slower than the process without discretization.

Keywords: Classification Data Mining, *Naïve Bayes*, Fayyad-Irani's Discretization

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004

Pembimbing,

Rifkie Primartha, M.T.
NIP. 197706012009121004

**PENGARUH PROSES DISKRITISASI FAYYAD IRANI PADA AKURASI
DATA MINING KLASIFIKASI NAÏVE BAYES**

Oleh :
Ahmad Halim Prabowo
09021281520112

ABSTRAK

Data Mining Klasifikasi merupakan suatu teknik untuk mengelompokkan suatu data kedalam suatu kategori atau kelas. Proses dalam klasifikasi dapat memiliki akurasi yang beragam tergantung dari kesesuaian antara data dan algoritma yang digunakan. Salah satu cara untuk meningkatkan akurasi dari proses klasifikasi adalah dengan menggunakan prapemrosesan data. Diskritisasi Fayyad-Irani adalah prapemrosesan data yang mengubah nilai numerik menjadi kategorikal. Diskritisasi Fayyad-Irani ini telah dibuktikan dapat meningkatkan akurasi dari klasifikasi yang menggunakan algoritma pohon keputusan. Berbeda dengan algoritma pohon keputusan, algoritma Naïve Bayes merupakan algoritma yang berbasis probabilitas Bayesian sehingga dapat memiliki pengaruh yang berbeda juga terhadap akurasi dan waktu klasifikasi. Berdasarkan percobaan yang telah dilakukan dalam penelitian ini, diketahui bahwa Diskritisasi Fayyad-Irani berhasil meningkatkan akurasi dari proses klasifikasi Naïve Bayes secara signifikan dari yang hanya sebesar 29,31% menjadi 93,26%, tetapi waktu proses yang dibutuhkan menjadi lebih lambat dengan selisih waktu sebanyak 25,5 menit lebih lambat dari proses tanpa diskritisasi.

Kata Kunci: Data Mining Klasifikasi, Naïve Bayes, Diskritisasi Fayyad-Irani.

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

Pembimbing,

Rifkie Primartha, M.T.
NIP. 197706012009121004

BAB I

PENDAHULUAN

1.1 Pendahuluan

Dalam bab ini yang akan dibahas adalah tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat dari penelitian, juga batasan masalah dan menjelaskan keseluruhan gambaran penelitian secara umum.

1.2 Latar Belakang

Menurut KBBI (Kamus Besar Bahasa Indonesia) data adalah keterangan yang benar dan nyata yang dapat dijadikan dasar kajian (analisis atau kesimpulan). Suatu data dapat memiliki nilai dan atribut tertentu. Ketika data diolah, data dapat menghasilkan yang namanya informasi. Pengolahan data tersebut harus dilakukan dengan cara yang tepat dan benar agar informasi yang didapatkan juga baik dan akurat.

Pada saat ini dunia sudah memiliki banyak sekali data dikarenakan pertumbuhan teknologi yang sangat pesat. Tentu saja dengan semakin banyaknya data tingkat kesulitan dalam mengolahnya juga akan bertambah sehingga hampir tidak mungkin atau membutuhkan waktu yang lama untuk menghasilkan suatu informasi jika dikerjakan secara manual. Data yang sebenarnya bisa dimanfaatkan malah tidak terpakai, sehingga pada akhirnya pengambilan atau penentuan keputusan ditentukan hanya dengan intuisi saja. (Aggarwal, 2015)

Penggalan *data mining* adalah proses menemukan pola dalam kumpulan data yang besar yang melibatkan metode pembelajaran mesin, statistik, dan sistem

database (Aggarwal, 2015). Proses pengolahan data pada *data mining* dilakukan menggunakan perangkat lunak sehingga lebih akurat dan cepat. Dengan *data mining* kita dapat mengolah data yang banyak sehingga dapat membantu dalam proses pengambilan keputusan.

Salah satu teknik dalam *data mining* adalah teknik klasifikasi. Klasifikasi adalah suatu proses pengelompokkan objek – objek ke dalam suatu kategori, atau beberapa kelas (Widiastuti, Rainarli, & Dewi, 2017). Beberapa algoritma yang bisa digunakan untuk klasifikasi adalah seperti C4.5, Naive Bayes, K-Nearest Neighbour dan lain-lain.

Proses dalam *data mining* klasifikasi memiliki akurasi yang beragam. Tingkat akurasi dapat tergantung dari kecocokan algoritma yang digunakan dengan kasus data, atau kualitas data itu sendiri. Untuk meningkatkan akurasi beberapa modifikasi dapat dilakukan, seperti modifikasi terhadap algoritma, modifikasi data sebelum di proses (*data preprocessing*) dan lain-lain.

Diskritisasi atau disebut juga *binning* adalah proses pengkonversian atribut dari yang bersifat numerik menjadi kategorikal atau nominal. Diskritisasi bisa digunakan untuk metode *data mining* yang tidak dapat memproses atribut numerik atau untuk membantu mengurangi banyaknya nilai atau *value* dari suatu atribut (Aggarwal, 2015).

Sebuah penelitian berhasil meningkatkan akurasi klasifikasi yang menggunakan algoritma C4.5 dengan proses diskritisasi metode Fayyad-Irani.. Akurasi yang didapatkan meningkat sebesar 2.63% pada dataset pertama dan 10.53% pada dataset kedua. Dalam penelitian tersebut dataset yang digunakan

adalah dataset ekspresi gen untuk menentukan tipe kanker darah pada seorang pasien. Dataset ekspresi gen memiliki 7125 atribut tetapi hanya memiliki 38 sampel pada dataset pertama dan 34 sampel pada dataset kedua. (Chandrasekar, Qian, Shahriar, & Bhattacharya, 2017). Oleh karena itu dalam penelitian ini peneliti akan mencoba menerapkan *data mining* klasifikasi dengan proses diskritisasi berbasis entropi Fayyad-Irani pada algoritma Naive Bayes untuk dataset Avila dengan harapan untuk meningkatkan akurasi serta mengetahui berapa lama waktu yang dibutuhkan dari data mining klasifikasi.

Naive Bayes dipilih dikarenakan perbedaannya dengan algoritma pohon keputusan, dimana Naive Bayes merupakan *classifier* yang berbasis probabilitas Bayesian (Farid, Zhang, Mofizur, Hossain, & Strachan, 2014). Selain itu dalam mengelola data numerik, Naïve Bayes biasanya menggunakan *Gaussian Distribution*, sehingga menarik untuk dilihat pengaruh dari diskritisasi terhadap hasil dari pengklasifikasian.

Dataset Avila adalah dataset dari gambar “*Avila Bible*” sebuah salinan raksasa bahasa Latin dari Bible yang disalin oleh 12 orang berbeda. Setiap halaman yang dibuat oleh penyalin tidak memiliki jumlah halaman yang sama. Dataset memiliki 10 fitur atau atribut dan 12 macam kelas yang mewakili tiap penyalin. Dataset ini digunakan dikarenakan oleh jumlah sampel yang banyak dan seluruh atribut yang digunakan adalah numerik sehingga membantu proses diskritisasi. (De Stefano, Maniaci, Fontanella, & Scotto di Freca, 2018)

1.3 Perumusan Masalah

Berdasarkan latar belakang yang telah disampaikan di atas dapat disimpulkan bahwa rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apa pengaruh proses diskritisasi berbasis entropi Fayyad-Irani terhadap akurasi dari *data mining* klasifikasi yang menggunakan algoritma Naive Bayes?
2. Apa pengaruh proses diskritisasi terhadap lama waktu proses klasifikasi?

1.4 Tujuan Penelitian

Dari permasalahan yang ada, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui pengaruh dari diskritisasi Fayyad-Irani terhadap akurasi klasifikasi algoritma Naive Bayes untuk dataset Avila.
2. Mengetahui pengaruh dari diskritisasi Fayyad-Irani terhadap lama waktu klasifikasi algoritma Naive Bayes untuk dataset Avila.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah dapat digunakan sebagai referensi untuk pengembangan perangkat lunak yang memanfaatkan algoritma Naive Bayes dan diskritisasi Fayyad-Irani untuk penelitian selanjutnya.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Dataset yang digunakan adalah dataset Avila. Dataset avila adalah dataset dari gambar “*Avila Bible*” dimana setiap atributnya bersifat kontinyu atau numerik. (De Stefano et al., 2018)
2. Dataset yang digunakan memiliki 12 kelas dan 10 atribut, dengan jumlah data sebanyak 20867 sampel.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut.

BAB I PENDAHULUAN

Bab ini akan membahas tentang latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup, metodologi penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini membahas seluruh dasar-dasar teori yang digunakan mulai dari definisi sistem, informasi mengenai domain, dan semua yang digunakan pada tahapan analisis, perancangan, dan implementasi.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai tahap-tahap yang akan diterapkan pada penelitian. Setiap rencana dari tahapan penelitian dideskripsikan secara rinci

berdasarkan kerangka kerja. Dilanjutkan dengan perancangan manajemen proyek dalam pelaksanaan penelitian.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas perancangan dan lingkungan implementasi, berupa analisis dari masalah yang dihadapi dalam penelitian serta perancangan perangkat lunak untuk meningkatkan akurasi algoritma *Naïve Bayes* menggunakan diskritisasi *Fayyad-Irani* pada klasifikasi dataset Avila sebagai alat penelitian.

BAB V HASIL DAN ANALISIS PENELITIAN

Bab ini membahas implementasi dari hasil analisis dan perancangan yang sudah dilakukan sebelumnya. Hasil analisis berupa kesimpulan yang dapat diambil dari penelitian. Melakukan pengujian perangkat lunak dan pengujian data penelitian.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi semua kesimpulan dari uraian-uraian yang telah dibahas sebelumnya, dan saran yang diharapkan dapat berguna untuk pengembangan penelitian lebih lanjut.

1.8 Kesimpulan

Jumlah data semakin lama semakin banyak. Untuk memanfaatkan data tersebut proses *data mining* dapat dilakukan. Akurasi adalah hal yang penting dalam proses *data mining*. Salah satu teknik data mining adalah klasifikasi. Untuk meningkatkan akurasi dari proses data mining dapat dilakukan dengan

diskritisasi. Dalam penelitian ini peneliti mencoba untuk menerapkan dan menguji pengaplikasian diskritisasi Fayyad-Irani terhadap algoritma Naive Bayes untuk klasifikasi dengan tujuan untuk meningkatkan akurasi dan mengetahui lama waktu dari klasifikasi.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-14142-8>
- Burdi, F., Setianingrum, A. H., & Hakiem, N. (2016). Application of the Naive Bayes Method to a Decision Support System to Provide Discounts (Case Study: PT. Bina Usaha Teknik). In *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)* (pp. 281–285). IEEE. <https://doi.org/10.1109/ICT4M.2016.064>
- Chandrasekar, P., & Qian, K. (2016). The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (pp. 618–619). IEEE. <https://doi.org/10.1109/COMPSAC.2016.205>
- Chandrasekar, P., Qian, K., Shahriar, H., & Bhattacharya, P. (2017). Improving the Prediction Accuracy of Decision Tree Mining with Data Preprocessing. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)* (pp. 481–484). IEEE. <https://doi.org/10.1109/COMPSAC.2017.146>
- De Stefano, C., Maniaci, M., Fontanella, F., & Scotto di Freca, A. (2018). Reliable writer identification in medieval manuscripts through page layout features: The “Avila” Bible case. *Engineering Applications of Artificial Intelligence*, 72, 99–110. <https://doi.org/10.1016/J.ENGAPPAI.2018.03.023>
- Farid, D., Zhang, L., Mofizur, C., Hossain, M. A., & Strachan, R. (2014). Expert Systems with Applications Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems With Applications*, 41(4), 1937–1946. <https://doi.org/10.1016/j.eswa.2013.08.089>
- Fayyad, U., & Irani, K. (1993). Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. Retrieved from <https://trs.jpl.nasa.gov/handle/2014/35171>
- García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98, 1–29. <https://doi.org/10.1016/J.KNOSYS.2015.12.006>
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining : concepts and techniques*. Elsevier Science.
- Netti, K., & Radhika, Y. (2015). A novel method for minimizing loss of accuracy in Naive Bayes classifier. In *2015 IEEE International Conference on*

Computational Intelligence and Computing Research (ICCIC) (pp. 1–4).
IEEE. <https://doi.org/10.1109/ICCIC.2015.7435801>

Netti, K., & Radhika, Y. (2016). An efficient Naïve Bayes classifier with negation handling for seismic hazard prediction. In *2016 10th International Conference on Intelligent Systems and Control (ISCO)* (pp. 1–4). IEEE. <https://doi.org/10.1109/ISCO.2016.7726906>

S. Pressman, R., & Maxim, B. (2014). *Software Engineering: A Practitioner's Approach, 8th Ed.*

Ting, K. M. (2017). Confusion Matrix BT - Encyclopedia of Machine Learning and Data Mining. In C. Sammut & G. I. Webb (Eds.) (p. 260). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-7687-1_50

Widiastuti, N. I., Rainarli, E., & Dewi, K. E. (2017). Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *JURNAL INFOTEL*, 9(4), 416. <https://doi.org/10.20895/infotel.v9i4.312>

