

## **Review Paper Data Mining Klasifikasi Data Mining**

**Agustina Srirahayu<sup>1)</sup>, Laras Setya Pribadie<sup>2)</sup>**

<sup>1)2)</sup> Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

Jln. Bhayangkara No.55 Surakarta, Jawa Tengah

Email : [agustina@udb.ac.id](mailto:agustina@udb.ac.id)<sup>1)</sup> [laras\\_setya@udb.ac.id](mailto:laras_setya@udb.ac.id)<sup>2)</sup>

### **ABSTRACT**

*The process of combining statistical techniques, mathematical calculations, Artificial Intelligence (AI) and machine learning to extract useful and interrelated information from large amounts of data. Data mining is commonly used to analyze and explore big data to get useful information. There is a lot of information that can be extracted from processing using data mining, such as analyzing consumer purchases or making decisions regarding the production process. Science has often been implemented to solve problems that arise from existing circumstances. In addition to the knowledge needed in order to solve problems or make strategic decisions in dealing with problems that arise, past experience or data that has been obtained at the time of the lights can also be used as a reference for making decisions. Data mining is the process of extracting (mining) information from a set of past data which is then displayed as knowledge to be used in accordance with the desired needs. Research on data mining has been carried out by many researchers to date related to the application of data mining to solve the problems they face, even though the problems are different in type and designation. This paper was written to review existing papers regarding data mining, especially classification. And to get information and map from research that has been done to be used as literature on the author's research plan. Through this review, it is hoped that later you can choose the right method to process data using data mining for the best results.*

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

### **ABSTRAK**

Proses menggabungkan teknik statistik, perhitungan matematika, Artificial Intelligence (AI) dan machine learning untuk menggali informasi yang saling terkait dan bermanfaat dari jumlah data yang besar. Data mining biasa digunakan untuk menganalisis dan mengeksplorasi data yang besar untuk mendapatkan informasi yang berguna. Ada banyak informasi yang bisa diambil dari pengolahan menggunakan data mining, seperti menganalisa pembelian konsumen atau penentuan terkait proses produksi Ilmu pengetahuan saat ini telah sering diimplementasikan untuk menyelesaikan masalah yang timbul dari suatu keadaan yang ada. Selain pengetahuan dibutuhkan dalam rangka menyelesaikan masalah atau mengambil sebuah keputusan strategis dalam menghadapi permasalahan yang muncul, pengalaman atau data lampau yang telah didapatkan pada waktu lampu dapat juga dijadikan sebagai acuan untuk menentukan keputusan. Data mining merupakan proses untuk menggali (menambang) informasi dari sekumpulan data – data lampau yang kemudian ditampilkan sebagai pengetahuan untuk digunakan sesuai dengan kebutuhan yang diinginkan. Penelitian mengenai data mining telah banyak dilakukan oleh peneliti hingga saat ini terkait dengan penerapan data mining untuk menyelesaikan masalah yang mereka hadapi, meskipun masalah tersebut berbeda – beda tipe dan peruntukannya. Paper ini ditulis untuk melakukan review atas paper – paper yang ada mengenai data mining terutama klasifikasi. Dan untuk mendapatkan informasi dan memetakan dari penelitian yang sudah dilakukan untuk digunakan sebagai literatur pada rencana penelitian penulis. Melalui review ini diharapkan nantinya dapat memilih metode yang tepat untuk memproses data dengan menggunakan data mining untuk hasil yang terbaik.

**Kata kunci :** data mining, klasifikasi, dataset



#### **Article History**

Received : 09/02/2022  
Revised : 24/06/2022  
Accepted : 29/07/2022  
Online : 30/04/2023



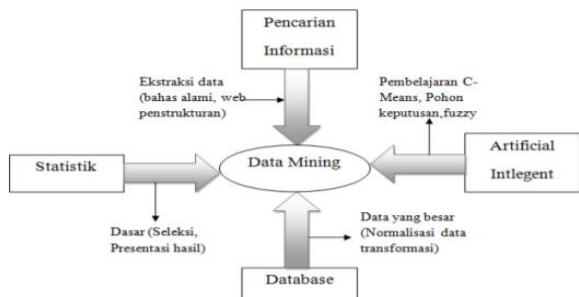
This is an open access article under the  
**CC BY-SA 4.0 License**

1. Pendahuluan

Seiring dengan lonjakan era digital dan internet yang semakin luas digunakan oleh banyak pihak, membuat semakin banyak pula data yang dihasilkan dari penggunaan itu, baik dari data personal maupun data transaksi yang dilakukan. Database yang dihasilkan ini menjadi data yang sangat berharga bagi pihak yang mempunyai kepentingan terhadap data tersebut. Untuk mendapat informasi dari database tersebut diperlukan sebuah cara untuk mengambil sebuah informasi dari database yang relatif mempunyai ukuran besar.

Proses pengolahan data yang belakangan ini diminati adalah data mining (Hendrian and Senna, 2018). Data mining merupakan sebuah proses yang menggabungkan teknik statistic, matematika, kecerdasan buatan dan *machine learning* (istilah yang digunakan untuk mesin yang bisa bekerja sendiri tanpa arahan dari penggunannya) untuk menggali informasi yang saling terkait dan bermanfaat dari jumlah data yang besar. Data mining biasa digunakan untuk menganalisis dan mengeksplorasi data yang besar untuk mendapatkan informasi yang berguna. Ada banyak informasi yang bisa diambil dari pengolahan menggunakan data mining, seperti menganalisis pembelian konsumen atau penentuan terkait proses produksi (Azahari *et al.*, 2020). Pengetahuan yang didapatkan dari pengolahan data mining berupa rumus, aturan atau model (Sumarni and Suhardi, 2020). Informasi yang dihasilkan dari data mining dengan menggunakan algoritma tertentu sering digunakan sebagai bentuk pengambilan keputusan (SPK) atau *Decision Support System* (DSS). Melalui DSS ini diharapkan akan lebih mudah dalam menyelesaikan masalah yang ditimbulkan terkait dengan apa yang sedang dihadapi saat ini atau apa yang akan terjadi dimasa mendatang (prediksi).

Data mining digunakan untuk mendapatkan pengetahuan / informasi yang baru. Prinsip kerjanya melakukan pemisahan suatu data yang besar menjadi data yang lebih kecil atau intisarinya dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data yang cukup besar (DP Utomo and Mesran, 2020). Proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya, penganalisisan data dan penemuan pola yang tersembunyi menggunakan arti yang otomatis dan semi-otomatis. Seperti pola untuk pencarian informasi, statistik, database dan *artificial intelligence* tersaji dalam asal usul dari bidang ilmu data mining yang tergambar sesuai gambar 1 dibawah ini.



Gambar 1. Bidang Ilmu Data Mining

Dalam memperoleh informasi pada data mining menggunakan teknik tertentu yang diantaranya deskripsi, estimasi, clustering, asosiasi, prediksi dan klasifikasi (Tri N *et al.*, 2021). Salah satu data mining yang sering digunakan adalah klasifikasi. Klasifikasi merupakan proses analisis data yang digunakan untuk menentukan label kelas dari sampel data yang ingin diklasifikasikan (Y Rado, L Marifni and S Irma, 2020). Klasifikasi sendiri termasuk dalam *supervised learning*, yang membutuhkan sampel data masukan dan hasilnya untuk menemukan hasil hubungan dari data sudah didefinisikan sebelumnya dengan tujuan meningkatkan kehandalan hasil yang diperoleh dari data yang ada (Budiman *et al.*, 2015a). Contoh dari klasifikasi yaitu klasifikasi masa studi, penjurusan atau kelulusan mahasiswa yang merupakan penerapan di dalam bidang pendidikan.

2. Pembahasan

A. Dataset

Kumpulan data yang berelasi dan berkaitan satu dengan yang lain dengan sifat data yang lebih fokus pada satu bidang spesifik atau fokus pada satu kasus tertentu biasa disebut dataset (A Rachmat and Y Lukoto, 2016). Ciri dataset yang bagus dan baik yaitu lengkap dari segi data yang dimiliki, selain itu data harus terupdate dan konsisten dalam mewakili data yang bermakna. Konten besaran variable yang jelas dan transparan, tidak ada keributan, data harus menarik dan mudah dipahami sesuai maksud juga masuk dalam ciri dataset yang baik (F Hosseinkhak *et al.*, 2009).

Di dalam database yang besar bisa berisi berbagai macam tipe ada. Masing – masing data yang mempunyai mempunyai cara pengolahan sendiri – sendiri di dalam data mining. Saat ini data bisa didapatkan dari berbagai sumber dengan bermacam format atau tipenya. Berdasarkan tipenya data bisa dibedakan menjadi tiga dengan rincian sebagai berikut:

1) Tipe Data Biner

Data biner adalah tipe data yang hanya mempunyai 2 kategori atau nilai. Nilai dalam data biner dinotasikan dalam angka 0 dan 1. Dimana angka 1 merepresentasikan nilai benar/ada, dan angka 0 merepresentasikan nilai salah/tidak ada seperti dalam operasi Boolean atau logika. Tipe data ini sering dipakai pada data hasil dari sebuah *record* dataset apakah masuk dalam kategori atau tidak. Sebagai contoh mahasiswa pada kelas lulus diberi nilai 1 dan yang tidak lulus diberi nilai 0.

2) Tipe Data Angka

Data angka adalah data numerik, baik yang merepresentasikan nilai tersebut langsung ke dalam nilai angka atau sebagai kode (R Ikhsan and PB Manggala, 2020). Angka yang langsung memberikan nilai contohnya adalah nilai matakuliah atau IPK mahasiswa. Sedangkan untuk kode biasanya digunakan untuk data yang berhubungan dengan kategori lain seperti NIM atau nomor KTP mahasiswa.

3) *Tipe data teks*

Data teks berisi adalah data yang mengandung huruf, baik yang keseluruhan berupa huruf maupun campuran dari huruf dan angka. Data teks juga bisa memiliki nilai dan panjang yang beragam. Data yang berupa huruf akan dikonversikan ke dalam data angka dalam teknik tertentu di data mining, kecuali jika akan menganalisis data teks itu sendiri maka tidak akan dikonversikan. Hal ini dilakukan untuk bisa menghitung setiap atribut yang dipakai ke dalam rumus untuk menentukan kategori kelasnya (R Ikhsan and PB Manggala, 2020).

Review paper bertujuan untuk mengetahui sejauh mana kedalaman atau keahlian/*expert* pembahasan dari masing-masing artikel yang dipilih. Selain itu Review paper juga untuk memahami isi dari paper dengan mengenali beberapa poin penting dan apa yang dituliskan dalam bentuk argumen oleh masing - masing penulis paper (RS Perdana, 2015). Berikut adalah judul artikel dan data peneliti yang direview :

1. Azahari, dkk dengan judul Komparasi Data Mining Naive Bayes Dan Neural Network Memprediksi Masa Studi Mahasiswa S1
2. Budiman, dkk dengan judul Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi
3. Hendrian dan Senna dengan judul Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan.
4. Jananto Arief, dkk dengan judul Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining
5. Khoirunnisa, dkk dengan judul Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi
6. Mardi Yuli dengan judul Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika
7. Novika Tri dkk dengan judul Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika.
8. Permana Baiq Andrisca Candra dan Dewi Intan Komala dengan judul Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes
9. Ristianti Fadma Dita dan Suparman dengan judul Komparasi Algoritma Klasifikasi pada Data Mining
10. Romli Ikhsan dan Ahmad Turmudi Zy dengan judul Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5.
11. Romli Ikhsan dan Putra Boy Manggala dengan judul Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes
12. Sugriyono dan Siregar Maria Ulfah dengan judul Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi
13. Sumarni dan Rustam Suhardi dengan judul Klasifikasi Topik Tugas Akhir Mahasiswa

menggunakan Algoritma Particle Swarm Optimization dan K-Nearest Neighbor.

14. Yendra Rado, dkk dengan judul Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode Naïve Bayes.

Setelah dilakukan review dari data penelitian diatas sejumlah 14 judul, diperoleh dataset yang dipakai tersaji dalam tabel 1 dibawah ini.

**Tabel 1.** Dataset Yang Dipakai

Peneliti	Dataset/Atribut
3	Data siswa SMA sebanyak 254 record, 37 atribut
4	Data induk mahasiswa yang telah lulus pada data wisuda berupa nim, nama mahasiswa, jenis kelamin, kota lahir, tanggal lahir, nama sekolah asal selama 5 tahun yang terdistribusi dalam setiap semester sebanyak total 1151 record
14	menggunakan data peserta sebanyak 284 dengan 227 sebagai data training dan 57 sebagai data testing
8	Dataset yang digunakan merupakan dataset primer terdiri atas 520 data pasien dan 17 atribut meliputi (age, gender, olyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, obesity, class)
7	Data kuisioner dengan atribut Minat Siswa, Cara Belajar Siswa, Motivasi Siswa, Cara Guru Mengajar, Media Pembelajaran, Sarana Prasarana sebanyak 100 koresponden
2	63 dataset, 10 atribut dan 2 label
1	3229 data mahasiswa, 1769 data training, 321 data testing, dan 1139 data target
12	6847 instance, dengan 18 atribut dan tiga kelas
5	158 Data siswa dengan 9 atribut

*B. Knowledge Discovery in Database (KDD)*

Data yang sudah berhasil dikumpulkan dari berbagai sumber diperlukan analisis data yang tepat agar didapatkan pengambilan keputusan yang sesuai (Dita RF and Suparman, 2019). Dalam menganalisis data di data mining terdapat beberapa prosedur dan pemakaian algoritma yang harus dilakukan untuk bisa memperoleh suatu pola pengetahuan dari data yang relatif sangat besar (Arief J, 2020). *Knowledge Discovery in Database (KDD)* merupakan metode untuk memperoleh pengetahuan dari database yang ada. Di dalam database terdapat tabel – tabel yang saling berhubungan/berelasi. Hasil dari proses KDD dapat digunakan sebagai data dasar untuk pemrosesan lebih lanjut.

*Knowledge Discovery in Database (KDD)* dan data mining merupakan proses yang dilaksanakan secara berdampingan untuk menggali informasi dari data yang

berjumlah besar dan lengkap (Khoirunnisa *et al.*, 2021). Walaupun keduanya memiliki konsep yang hampir sama namun masih tetap berkaitan satu dengan yang lainnya. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (M Yuli, 2021):

1. *Data selection*

Merupakan tahap pemilihan data (*sample*) yang akan digunakan dalam pengolahan data mining sesuai dengan informasi atau tujuan yang ingin diperoleh. Data yang digunakan adalah data yang sudah ada sebelumnya berdasar dari proses yang sudah berjalan di waktu lampau. Data yang sudah dipilih, akan digunakan dalam tempat yang berbeda dengan tempat penyimpanan data yang masih digunakan untuk operasional. Data ini sendiri nantinya akan dibagi dua untuk melakukan pemodelan dan untuk pengujian model.

2. *Pre-processing / Cleaning*

Di dalam database yang sudah didapatkan, tidak semua data bisa langsung digunakan untuk perhitungan model. Terdapat data double atau data yang salah satu datanya tidak diisi atau terisi dengan data yang tidak sesuai. Hal ini bisa terjadi karena pada saat penyimpanan terjadi *error* sehingga data tidak tersimpan atau kesalahan input oleh pengguna manusia yang tidak terlepas dari human error. Data yang tidak lengkap akan dilakukan perbaikan data terlebih dahulu sehingga tidak ada informasi yang hilang atau tidak lengkap tentunya dengan berdasar pada data – data yang relevan/data yang sudah ada. Sehingga data menjadi bersih dan siap dipakai untuk tahapan selanjutnya (Sugriyono and Ulfah SM, 2020)

3. *Transformation*

Data yang sudah siap kemudian ditransformasikan agar data tersebut bisa diproses dengan data mining. Transformasi merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Hal ini dilakukan untuk mempermudah dalam pembuatan model di data mining. Data yang serupa ditransformasikan atau dirubah ke dalam bentuk yang lebih sederhana, namun tidak merubah isi dasar dari basis data tersebut.

4. *Data Mining*

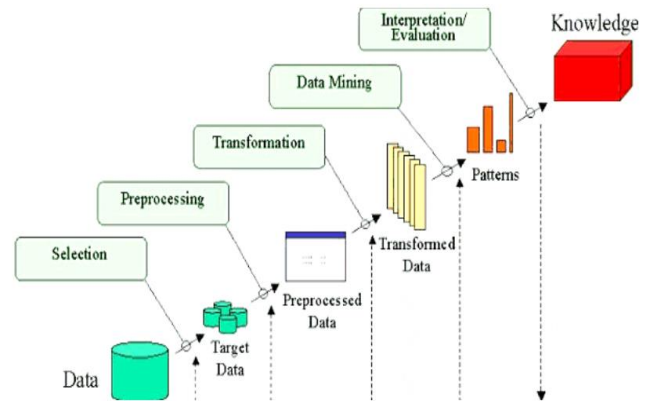
Data mining adalah proses pemodelan mencari hubungan antara data masukan dengan hasil dari basis data yang sudah ada dengan menggunakan Teknik atau metode tertentu (Budiman *et al.*, 2015a). Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi

yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Dari penjelasan 5 aliran informasi diatas dapat diringkas informasinya sesuai dengan gambar 2 tentang Proses Aliran Penggalian Informasi Data Mining (Arief J, 2020) dibawah ini:



Gambar 2. Proses Aliran Penggalian Informasi Data Mining

C. *Metode Data Mining*

Pengolahan data yang besar dengan menggunakan data mining memiliki tujuan yang berbeda – beda dalam mendeteksi masalah. Metode yang sering digunakan adalah klasifikasi atau klustering dan yang akan dibahas lebih lanjut adalah klasifikasi.

Klasifikasi adalah bagaimana melaksanakan proses penggerombolan dengan mengacu pada karakteristik tertentu. Klasifikasi merupakan sebuah proses training (pembelajaran) suatu fungsi yang memiliki target yang digunakan untuk memetakan tiap himpunan objek ke satu dari label kelas yang sudah didefinisikan sebelumnya. Klasifikasi termasuk tipe analisis yang dapat digunakan untuk membantu menemukan kelas label dari sejumlah sampel yang akan diklasifikasikan (Budiman *et al.*, 2015b). Klasifikasi merupakan bentuk dasar dari analisa data. Klasifikasi juga sering disebut sebagai teknik untuk menentukan kategori kelas dari data – data yang telah tersedia (Andriska Candra, no date). Teknik Klasifikasi ini cocok digunakan dalam mendeskripsikan dataset dengan tipe data dari suatu himpunan data yaitu biner atau nominal. Adapun kekurangan dari teknik ini yaitu tidak tepat untuk himpunan data ordinal karena pendekatan - pendekatan yang digunakan secara implisit dalam kategori data.

Didalam klasifikasi ada beberapa teknik atau algoritma yang dipakai untuk pemecahan suatu kasus/model data, diantaranya yaitu algoritma *C4.5/decision tree*, *naive bayes*, *neural network*, *Knearest neighbor*, *support vector machine*. Dari beberapa penelitian dalam klasifikasi data mining data set / *data training* menghasilkan role pengetahuan baru berupa pohon keputusan:

1. *Algoritma C4.5*

Algoritma C4.5 merupakan algoritma yang sering dipakai dalam pemecahan masalah yang menggunakan teknik klasifikasi. Algoritma C4.5 merupakan teknik

klasifikasi data dengan cara membuat metode berupa pohon keputusan dari atas ke bawah. Dimana atribut akar akan diletakan di bagian paling atas diikuti dengan atribut lain sebagai daunnya (R Ikhsan and PB Mangala, 2020):

Algoritma dasar dari algoritma C4.5 adalah : 1). Pohon yang dihasilkan berupa pohon terbalik, 2). Pada tahap awal, semua contoh training adalah akar. 3). Atribut adalah kategori. 4) Contoh di partisi secara berulang berdasarkan atribut yang dipilih. 5). Atribut tes dipilih dari data heuristic atau pengukuran statistik.

2. Naïve Bayesian

Naïve Bayes merupakan teknik klasifikasi dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan asal Inggris Thomas Bayes, yaitu dengan memprediksi peluang yang akan terjadi masa yang akan datang dengan melihat pengalaman yang ada di masa sebelumnya.

Metode Naïve Bayes Classifier menggunakan prinsip dasar probabilitas sebagai acuannya. Metode ini bertujuan untuk mengklasifikasikan kelas satu objek dengan cara menghitung probabilitas dari masing-masing karakteristik objek tersebut berdasarkan kumpulan data yang tersedia (Y Rado, L Marifni and S Irma, 2020).

3. Algoritma K-Nearest Neighbor (KNN)

Algoritma merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut, cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya *Euclidian distance* dengan menggunakan sampel-sampel dari data testing (Sumarni and Suhardi, 2020)

Berikut tersaji hasil pengujian dengan metode algoritma yang dipakai dalam masing-masing paper dan hasil akurasi dari 14 artikel yang sudah direview dapat dilihat dalam tabel 2 dibawah ini.

Tabel 2. Hasil Pengujian

Peneliti	Metode/Algoritma	Hasil Akurasi
3	algoritma C4.5	98,80%
4	Algoritma CART	63%
14	Naïve Bayesian Clasification	81%
8	Decision Tree Naïve Bayesian	95,58% 87,69%
7	Algoritma C4.5	96%
2	Algoritma C4.5	73,33%
13	KNN-PSO	89,86%
5	Decision Tree Naïve Bayesian KNN	95,60% 92,40% 94,96%
12	K Means - Euclidean K Means - Manhattan KNN	98,42% 97,76% 94,96%

Peneliti	Metode/Algoritma	Hasil Akurasi
1	Naïve Bayesian Neural Network	57,63% 72,58%

3. Kesimpulan

Dari review paper ini berkaitan dengan klasifikasi data mining dapat disimpulkan bahwa :

1. Klasifikasi data mining dapat digunakan untuk memperoleh informasi atau pembuatan model dengan menggunakan data yang memiliki atribut dan kelas yang sudah sebelumnya.
2. Dalam pengolahan data mining diperlukan tahapan pemilihan dan pre-processing data yang baik, baik itu untuk klasifikasi ataupun untuk fungsi yang lainnya.
3. Tahapan pre-processing diperlukan untuk membantu mempermudah dalam pemodelan data karena data yang hilang atau kurang informasinya akan diperbaiki terlebih dahulu sebelum digunakan.
4. Klasifikasi menggunakan algoritma C4.5 mempunyai prosentase akurasi tertinggi.

Daftar Pustaka

Andriska Candra, BP., & K. di (no date) “Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes,” *Jurnal Informatika dan Teknologi Vol. 4 No. 1 Hal. 63-69* [Preprint].

A Rachmat and Y Lukoto (2016) “Sentipol: Dataset Sentimen Komentar Pada Kampanye Pemilu Presiden Indonesia 2014 dari Facebook Page,” *Konferensi Nasional Teknologi Informasi dan Komunikasi (KNASTIK 2016)* [Preprint], (2338–7718).

Arief J (2020) “Data Induk Mahasiswa sebagai Prediktor Ketepatan Waktu Lulus Menggunakan Algoritma CART Klasifikasi Data Mining,” *Jurnal Sisfokom*, 10.

Azahari *et al.* (2020) “Komparasi Data Mining Naive Bayes Dan Neural Network Memprediksi Masa Studi Mahasiswa S1,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(3), pp. 443–452.

Budiman *et al.* (2015a) “Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi,” *Jurnal Jupiter*, 7.

Budiman *et al.* (2015b) “Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi,” *Jurnal Jupiter*, 7.

Dita RF and Suparman (2019) “Komparasi Algoritma Klasifikasi pada Data Mining,” in *Proceedings of The 1st Steem*, pp. 148–156.

DP Utomo and Mesran (2020) “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data set Penyakit Jantung,” *Jurnal Media Informatika Budidarma*, 4(2), pp. 437–444.

- F Hosseinkhak *et al.* (2009) "Challenges in Data Mining on Medical Databases. In M. Khosrow-Pour (Ed.),", *Encyclopedia of Information Science and Technology, Second Edition*, pp. 502–511.
- Hendrian and Senna (2018) "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan."
- Khoirunnisa *et al.* (2021) "Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi," *Jurnal Informatika*, 8(1), pp. 26–33.
- M Yuli (2021) "Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, 2, pp. 213–219.
- R Ikhsan and PB Manggala (2020) "Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes," *Jurnal Pelita Industri*, 1(1), pp. 36–45.
- RS Perdana (2015) "Menulis Review Sebuah Paper atau Artikel Ilmiah," *Faculty of Computer Science, Brawijaya University* [Preprint].
- Sugriyono and Ulfah SM (2020) "Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi," *Jurnal Teknologi dan Sistem Komputer*, 8(4), pp. 311–316.
- Sumarni and Suhardi (2020) "Klasifikasi Topik Tugas Akhir Mahasiswa menggunakan Algoritma Particle Swarm Optimization dan K-Nearest Neighbor," *ILKOM Jurnal Ilmiah*, 12(2), pp. 168–175.
- Tri N *et al.* (2021) "Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika," *Jurnal Media Informatika Budidarma*, 5(1), pp. 9–17.
- Y Rado, L Marifni and S Irma (2020) "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode Naïve Bayes," *Jurnal Sains Matematika dan Statistika*, 6.