

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Pendidikan

Menurut Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional yang dimaksud dengan pendidikan yaitu usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara.

Pendidikan dalam arti luas adalah segala pengalaman belajar yang berlangsung sepanjang hayat yang memberikan pengaruh positif pada pertumbuhan setiap orang dalam segala lingkungan dan situasi. Namun, dalam arti sempit, pendidikan adalah upaya lembaga untuk memberikan hasil kepada siswa dengan tujuan mereka memiliki kemampuan yang baik dan memahami hubungan dan masalah sosial (Pristiwanti dkk., 2022).

Pendidikan pada tingkat SMA terdapat beberapa ujian diantaranya Penilaian Tengah Semester Ganjil (PTS), Penilaian Akhir Semester (PAS), Penilaian Tengah Semester Genap (PTS), Penilaian Akhir Tahun (PAT), sedangkan pada beberapa mata pelajaran bisa terdapat ujian lain seperti ulangan harian setiap pergantian BAB.

2.1.2 Pedagogik Guru

Dalam Peraturan pemerintah republik indonesia nomor 19 tahun 2005 tentang standar nasional pendidikan yang dimaksud dengan kompetensi pedagogik adalah kemampuan mengelola pembelajaran peserta didik yang meliputi pemahaman terhadap peserta didik, perancangan dan pelaksanaan pembelajaran, evaluasi hasil belajar, dan pengembangan peserta didik untuk mengaktualisasikan berbagai potensi yang dimilikinya (PERATURAN PEMERINTAH REPUBLIK INDONESIA, t.t.).

Salah satu kompetensi yang membedakan guru dari profesi lain adalah kompetensi pedagogik. Guru yang memiliki kompetensi pedagogik yang baik akan menentukan keberhasilan proses pembelajaran dan diharapkan dapat berinteraksi dengan siswa dengan lebih baik maupun dalam menangani masalah. (Akbar, 2021; Maulana Haz dkk., 2022).

Ada tujuh aspek kompetensi pedagogik yang harus dikuasai oleh guru, yaitu:

1. Menguasai Karakteristik Peserta Didik
2. Menguasai Teori dan Prinsip Pembelajaran
3. Pengembangan Kurikulum
4. Kegiatan Pembelajaran
5. Pengembangan Potensi Peserta Didik
6. Komunikasi dengan Peserta Didik
7. Penilaian dan Evaluasi

2.1.3 Kecurangan Akademik

Kecurangan dalam bidang akademik didefinisikan sebagai tindakan atau upaya kecurangan yang dilakukan oleh siswa dengan cara yang tidak sah atau tidak dapat diterima di bidang akademik. Terdapat beberapa faktor pendorong perbuatan kecurangan akademik antara lain dipengaruhi oleh kontrol diri, efikasi diri, dan prestasi akademik. Serta Sebagian besar dipengaruhi oleh tingkat hukuman yang diberikan atas perilaku tersebut, pengaruh teman sebaya, persepsi terhadap materi, dan pengajar yang terdiri dari pengawasan saat ulangan, metode pembelajaran atau suasana pembelajaran, dan kedisiplinan guru. Kebiasaan kecurangan akademik akan berdampak negatif pada siswa. Siswa yang terbiasa melakukan kecurangan akademik akan senang bergantung pada orang lain atau metode tertentu daripada kemampuan diri mereka sendiri untuk mencapai hasil belajar mereka. (Aulia, 2015; Mushthofa dkk., 2021; PURNAMAWATI, 2016; Saputra dkk., 2021)

Yang dilakukan dalam perbuatan kecurangan akademik yaitu:

1. Bertanya teman.
2. Membuat catatan kecil.
3. Membuka internet menggunakan HP.
4. Membawa buku.

2.1.4 Populasi dan Sampel

a. Populasi

Populasi adalah keseluruhan elemen dalam penelitian yang meliputi objek dan subjek dengan karakteristik tertentu. Populasi dapat berupa semua anggota kelompok manusia, binatang, peristiwa, atau benda yang tinggal bersama dalam

suatu tempat secara terencana, seperti guru, siswa, kurikulum, fasilitas, lembaga sekolah, hubungan sekolah, masyarakat, dan sebagainya. Populasi bukan hanya orang, tetapi juga organisasi, binatang, hasil karya manusia, dan benda-benda lainnya. (Amin dkk., 2023).

b. Sampel

Sampel dapat diartikan sebagai bagian dari populasi yang menjadi sumber data yang dijadikan objek penelitian, terdapat beberapa cara untuk menentukan ukuran sampel salah satunya yaitu menggunakan rumus *slovin* yang biasanya digunakan ketika tidak tahu berapa proporsi populasi yang ideal untuk mewakili keseluruhan populasi (Dewi & Pardede, 2021), berikut notasi rumus *slovin* untuk menghitung sampel penelitian

$$n = \frac{N}{1 + N(e)^2} \quad (2.1)$$

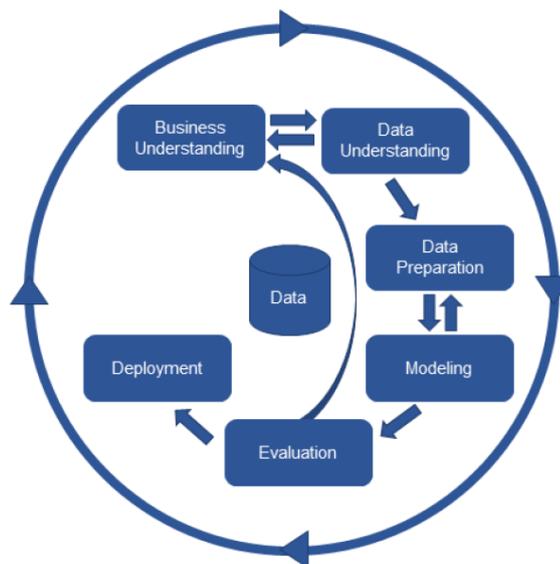
2.1.5 Data Mining

a. Pengertian

Data Mining adalah proses mencari pola atau informasi menarik dari suatu basis data. Untuk menghasilkan suatu informasi dari basis data dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery in Database (KDD). (Wanhari, 2018).

b. Proses Tahapan Data Mining

Pada tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, NCR, dan SPPSS, mengembangkan Cross-Industry Standard Process for Data Mining (CRISP-DM) untuk menyediakan proses standar dari data mining sebagai strategi pemecahan masalah secara umum dari bisnis maupun unit penelitian. Dalam CRISP-DM, sebuah proses data mining memiliki siklus yang terdiri dari enam fase. Berikut gambaran alur proses yang terjadi pada CRISP-DM pada Gambar 2.1:



Gambar 2.1 CRISP DM
Sumber (Munawwaroh & Primandari, 2022)

Fase-fase dalam CRISP-DM dijelaskan sebagai berikut :

1. Fase Pemahaman Bisnis (*Business Understanding Phase*), yang terdiri dari beberapa tahapan:
 - a. Mengidentifikasi tujuan bisnis

- b. Menilai situasi
 - c. Menentukan tujuan penambangan/analisis data
 - d. Membuat rencana proyek
2. Fase Pemahaman Data (*Data Understanding Phase*), yang terdiri dari beberapa tahapan:
- a. Mengumpulkan data
 - b. Mendeskripsikan data
 - c. Mengeksplorasi data
 - d. Memverifikasi kualitas data
3. Fase Persiapan Data (*Data Preparation Phase*), yang terdiri dari beberapa tahapan:
- a. Memilih data
 - b. Membersihkan data
 - c. Membangun data
 - d. Mengintegrasikan data
 - e. Memformat data
4. Fase pemodelan (*Modeling Phase*), yang terdiri dari beberapa tahapan:
- a. Memilih teknik pemodelan
 - b. Merancang pengujian
 - c. Membangun model
 - d. Menilai model

Pada fase ini kita sudah menerapkan teknik-teknik data mining yang sesuai untuk menyelesaikan permasalahan. Contoh melakukan klasifikasi dengan membuat sebuah model berupa decision tree.

5. Fase Evaluasi (*Evaluation Phase*) yang terdiri dari beberapa tahapan:
 - a. Mengevaluasi hasil
 - b. Meninjau proses
 - c. Menentukan langkah selanjutnya
6. Fase Penyebaran (*Deployment Phase*), yang terdiri dari beberapa tahapan:
 - a. Merencanakan *deployment* (metode kita untuk mengintegrasikan penemuan data mining ke dalam penggunaan)
 - b. Perencanaan pemantauan dan pemeliharaan
 - c. Melaporkan hasil akhir
 - d. Meninjau hasil akhir

2.1.6 Klasifikasi

Klasifikasi yaitu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Decision tree merupakan satu teknik klasifikasi yang paling populer karena mudah untuk interpretasi seperti Algoritma C4.5, ID3 dan lain-lain.

2.1.7 Decision Tree

Decision Tree adalah struktur pohon yang menyerupai diagram alur, di mana setiap *internal node* (*node* yang bukan *leaf* atau bukan *node* terluar)

menunjukkan pengujian pada suatu atribut, setiap cabang mewakili hasil pengujian, dan setiap *leaf node* (atau terminal *node*) mempunyai label kelas. *Node* paling atas dalam sebuah pohon adalah *node* akar (Han dkk., 2011). Decision Tree digunakan untuk mempelajari klasifikasi dan prediksi pola dari data dan menggambarkan relasi dari variabel atribut x dan variabel target y dalam bentuk pohon (Sutoyo, 2018).

2.1.8 Algoritma C4.5

Algoritma C4.5 merupakan salah satu solusi yang sering digunakan dalam pemecahan masalah dalam klasifikasi. Fungsi utama dari algoritma C4.5 adalah membangun model pohon keputusan yang dapat digunakan untuk melakukan klasifikasi atau prediksi. Model pohon keputusan ini digunakan untuk mengambil keputusan berdasarkan atribut-atribut yang terdapat pada data. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *records* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. (Berry & Linoff, 2004)

Algoritma C4.5 merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan dari ID3 (*Iterative Dichotomiser 3*). Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan yang dapat dipahami dengan mudah dengan bahasa alami. Dan pohon keputusan juga dapat diekspresikan dalam bentuk bahasa

basis data seperti *Structured Query Language* untuk mencari *records* pada kategori tertentu (Hasibuan, 2021)

a. Pembuatan Pohon Keputusan

(Haryati dkk., 2015), sebagai berikut :

1. Menyiapkan *data training*. *Data training* biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Pilih atribut sebagai akar, untuk memilih atribut sebagai akar didasarkan pada nilai *gain ratio* tertinggi dari atribut-atribut yang ada. Sebelum menghitung nilai *gain ratio* dari atribut, hitung dahulu nilai *entropy*.
3. Menghitung nilai *entropy*.

Dalam algoritma C4.5, *entropy* adalah konsep yang digunakan untuk mengukur seberapa banyak ketidakakasian atau ketidakpastian yang terkandung dalam data. *Entropy* menggambarkan seberapa banyak informasi yang diperlukan untuk memprediksi kelas suatu data berdasarkan atribut yang ada. Semakin kecil nilai *entropy*, semakin informatif atribut tersebut dalam menghasilkan keputusan dan semakin baik atribut tersebut dalam mengklasifikasikan data.

Entropy dapat dihitung dengan menggunakan persamaan:

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (2.2)$$

Keterangan :

S : Himpunan Kasus

N : jumlah partisi S

P_i : proporsi S_i terhadap S

4. Kemudian hitung nilai *gain*

Gain pada algoritma C4.5 adalah metrik yang digunakan untuk mengevaluasi seberapa baik suatu atribut dapat memisahkan data menjadi kelas-kelas yang berbeda dengan mempertimbangkan perbedaan antara *entropy* sebelum dan setelah pemisahan data menggunakan atribut tertentu. *Gain* memiliki bias terhadap atribut dengan banyak nilai (kategori) yang membagi data secara lebih detail.

Gain dapat dihitung dengan persamaan:

$$Gain(S, A) = Entropi(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2.3)$$

Keterangan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke i

$|S|$: Jumlah kasus dalam S

5. Menghitung nilai *Split Info*

Split Info pada algoritma C4.5 adalah konsep yang membantu dalam pemilihan atribut yang paling informatif untuk membangun pohon keputusan. *Split Info* menyatakan *entropy* atau informasi potensial, semakin rendah nilai *Split Info*, semakin baik atribut tersebut dalam memisahkan data.

Split Info dapat dihitung dengan menggunakan rumus:

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (2.4)$$

Keterangan:

S : Himpunan kasus

A : Atribut

S_i : jumlah sampel dalam subset yang memiliki nilai atribut A sama dengan nilai ke-i.

6. Menghitung nilai *Gain Ratio*

Gain Ratio adalah modifikasi dari *gain* yang mengurangi biasnya terhadap atribut dengan banyak nilai dengan mempertimbangkan kompleksitas atribut (jumlah nilai) dalam memilih atribut yang paling informatif.

Gain Ratio dapat dihitung dengan menggunakan rumus:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \quad (2.5)$$

Keterangan:

S : Himpunan kasus

A : Atribut

Gain(S,A) : Information gain pada atribut A

Split Info(S,A) : Split information pada atribut A

7. Buat cabang untuk masing-masing nilai

8. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama/semua *node* terpartisi

b. Kelebihan dan Kekurangan Algoritma C4.5

Algoritma C4.5 merupakan evolusi dari ID3 yang dikembangkan oleh J. Ross Quinlan pada tahun 1993 dan telah membuat beberapa perubahan untuk meningkatkan algoritma ID3 (Barito dkk., 2022). Algoritma C4.5 merupakan salah satu metode yang sering digunakan dalam klasifikasi data mining. layaknya teknik klasifikasi lainnya algoritma C4.5 memiliki kelebihan dan kekurangan, adapun kelebihan algoritma C4.5 adalah sebagai berikut:

1. Mampu menangani *variable* yang memiliki *missing value*
2. Dapat mengelola atribut dengan tipe diskrip maupun maupun kontinu
3. Daerah pengambilan lebih simpel dan spesifik.
4. Dapat memangkas cabang dari pohon keputusan
5. Algoritma C4.5 atau pohon keputusan lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian (Elisa, 2017).
6. Eliminasi perhitungan-perhitungan tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sampel diuji hanya berdasarkan kriteria atau kelas tertentu.

Adapun kekurangan algoritma C4.5 adalah sebagai berikut:

1. Pada skalabilitas yaitu *data training* hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori
2. Kesulitan dalam mendesain pohon keputusan yang optimal
3. Hasil kualitas keputusan yang didapat sangat tergantung bagaimana pohon keputusan tersebut didesain. Sehingga jika pohon keputusan yang dibuat

kurang optimal, maka akan berpengaruh pada kualitas dari keputusan yang didapat.

4. Terjadi *overlap* terutama jika kelas-kelas kriteria yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
5. Pengakumulasian jumlah *error* dari setiap tingkat dalam sebuah pohon keputusan yang besar.

2.1.9 *Confusion Matrix*

Confusion Matrix merupakan pengujian model penelitian yang mempresentasikan hasil evaluasi model dalam melakukan prediksi maupun klasifikasi dengan menggunakan *table matrix*. Jika dataset terdiri dari 2 kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Evaluasi menggunakan *confusion matrix* menghasilkan nilai *Accuracy*, *Precision*, *Recall*, serta *F1-Measure*. Akurasi dalam klasifikasi merupakan presentasi ketepatan *record* data diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. *Precision* merupakan proposikasi yang diprediksi positif yang juga positif benar pada data sebenarnya. *Recall* merupakan proporsi kasus positif yang sebenarnya diprediksi positif secara benar (Putra & Wibowo, 2020).

True Positive (TP) merupakan jumlah *record* positif dalam dataset yang diklasifikasikan positif. *True Negative* (TN) merupakan jumlah *record negative* dalam dataset yang diklasifikasikan positif. *False Positive* (FP) merupakan jumlah *record* negatif dalam *dataset* yang diklasifikasikan positif. *False Negative* (FN)

merupakan jumlah *record positive* dalam dataset yang diklasifikasikan *negative*, dapat dilihat pada Gambar 2.2.

| | | Actual Value | |
|-----------------|--------------|---|---|
| | | Cat | Bukan kucing |
| Predicted Value | Cat |  TP |  FP |
| | Bukan kucing |  FN |  TN |

Gambar 2.2 *Convusion Matrix*

Sumber : <https://www.trivusi.web.id>

Berikut adalah persamaan model *Confusion Matrix* :

Accuracy adalah jumlah perbandingan data yang benar dengan jumlah keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.6)$$

Precision digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif.

$$Precision = \frac{TP}{FP + TP} \quad (2.7)$$

Recall digunakan untuk menunjukkan presentase kelas data positif yang berhasil diprediksi benar dari keseluruhan data kelas positif.

$$Recall = \frac{TP}{FN + TP} \quad (2.8)$$

F1-Score digunakan untuk menghitung ukuran keseimbangan antara *precision* dan *recall*

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.9)$$

2.2 State of The Art

Penelitian ini memiliki beberapa penelitian terkait yang akan digunakan sebagai bahan acuan dan perbandingan hasil penelitian. Berikut adalah beberapa hasil penelitian yang relevan dapat dilihat pada Tabel 2.1 dibawah ini:

Tabel 2.1 *State Of The Art 1*

| | | |
|---|---------------|--|
| 1 | Judul | Model Klasifikasi Berbasis <i>Machine Learning</i> Untuk Perpanjangan Masa Jabatan Kepala Sekolah Menggunakan Algoritma C4.5 |
| | Nama Peneliti | Puji Catur Siswipraptini, Ahmad Saputra Fadiarora, Hengki Sikumbang |
| | Tahun | 2023 |
| | Penerbit | Jurnal Indonesia : Manajemen Informatika dan Komunikasi |
| | Sumber | MODEL KLASIFIKASI BERBASIS MACHINE LEARNING UNTUK PERPANJANGAN MASA JABATAN KEPALA SEKOLAH MENGGUNAKAN ALGORITMA C4.5 Jurnal Indonesia : Manajemen Informatika dan Komunikasi (amikindonesia.ac.id) |
| | Metode/Solusi | Model klasifikasi berbasis <i>Machine Learning</i> menggunakan algoritma C4.5 Decision Tree. Model yang diusulkan di evaluasi menggunakan metode <i>Confusion Matrix</i> yang merupakan dasar pengukuran klasifikasi |
| | Hasil Utama | Model yang dihasilkan memiliki 15 <i>rules</i> yng dijadikan dasar untuk membentuk Pohon Keputusan. Pengukuran validasi model di uji menggunakan metode <i>Confusion Matrix</i> dan memiliki hasil akurasi sebesar 83,33 % |

| | | |
|---|---------------|---|
| 2 | Judul | Analisis Faktor Kompetensi Pedagogik Guru Dalam Proses Pembelajaran Menggunakan Algoritma C4.5 |
| | Nama Peneliti | Ingkana Devi Cahaya, Heru Satria Tambunan, Harly Okprana |
| | Tahun | 2021 |
| | Penerbit | BRAHMANA: Jurnal Penerapan Kecerdasan Buatan |
| | Sumber | https://tunasbangsa.ac.id/pkm/index.php/brahmana/article/view/69 |
| | Metode/Solusi | Penerapan algoritma C4.5 dapat digunakan untuk menghasilkan decision tree yang akan diambil penentuan berupa <i>rules</i> atau aturan-aturan yang akan dijadikan pohon keputusan. |
| | Hasil Utama | yang menjadi kriteria terbaik dalam proses pembelajaran adalah mengevaluasi hasil pembelajaran, nilai akurasi sebesar 87,90%, pada prediksi label sedang sebesar 83,33% dan prediksi label baik sebesar 91,67%. |
| 3 | Judul | Penerapan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan |
| | Nama Peneliti | Ahmad Muhajir Hasibuan |
| | Tahun | 2021 |
| | Penerbit | UIN Suska Riau |
| | Sumber | https://repository.uin-suska.ac.id/54986/1/BAB1-3.pdf |
| | Metode/Solusi | Klasifikasi data mining menggunakan algoritma C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH) |
| | Hasil Utama | Penerapan algoritma C4.5 untuk mengklasifikasi penerima bantuan program keluarga harapan (PKH) telah berhasil dilakukan. Pengujian menggunakan <i>confusion matrix</i> memperoleh akurasi terbaik dengan perbandingan 90% |

| | | |
|---------------|--|---|
| | | data latih dan 10% data uji sebesar 86,37% dan termasuk pada <i>range good classification</i> . |
| 4 | Judul | Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes |
| | Nama Peneliti | Detrinal Putra, Arief Wibowo |
| | Tahun | 2020 |
| | Penerbit | SENARIS |
| | Sumber | Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes Putra Prosiding Seminar Nasional Riset Information Science (SENARIS) (tunasbangsa.ac.id) |
| | Metode/Solusi | Penelitian ini menggunakan algoritma Naïve Bayes dengan teknik pengklasifikasi <i>data mining</i> untuk memperkirakan atau memprediksi keputusan minat penjurusan di SMA. |
| 5 | Hasil Utama | Setelah perhitungan akurasi menggunakan <i>Confusion Matrix</i> , diperoleh nilai akurasi model Naïve Bayes sebesar 93.75%, <i>precision</i> sebesar 83.33%, dan <i>recall</i> sebesar 100%. |
| | Judul | Analisis Indikator Kinerja Dosen Terhadap Prestasi Mahasiswa Semester Satu dengan Menggunakan Decision Tree |
| | Nama Peneliti | Rofilde Hasudungan |
| | Tahun | 2018 |
| | Penerbit | JURTI , Vol.2 No.2, Desember 2018, ISSN: 2579-8790 |
| | Sumber | (PDF) Analisis Indikator Kinerja Dosen Terhadap Prestasi Mahasiswa Semester Satu dengan Menggunakan Decision Tree (researchgate.net) |
| Metode/Solusi | Pada penelitian ini akan membahas penggunaan data mining khususnya teknik klasifikasi dengan menggunakan pohon keputusan untuk memprediksi prestasi mahasiswa. | |

| | | |
|---|---------------|--|
| | | Pada penelitian ini juga menyediakan suatu faktor-faktor pedagogik yang sangat berpengaruh terhadap prestasi mahasiswa didalam perkuliahan |
| | Hasil Utama | Dari hasil yang didapatkan dengan menggunakan decision tree diperoleh bahwa dari 28 indikator yang ditanyakan diperoleh bahwa hanya 5 faktor yang memiliki pengaruh terhadap prestasi mahasiswa tersebut. Model yang dibangun juga memiliki tingkat akurasi sebesar 74% dengan standar deviasi 18.41%. |
| 6 | Judul | <i>Pedagogue Performance Assessment (PPA) using data mining Techniques</i> |
| | Nama Peneliti | Ashily M Baby |
| | Tahun | 2018 |
| | Penerbit | IOP Conf. Series: Materials Science and Engineering 396 (2018) 012024 doi:10.1088/1757-899X/396/1/012024 |
| | Sumber | Pedagogue Performance Assessment (PPA) using Data mining Techniques - IOPscience |
| | Metode/Solusi | Menggunakan tiga algoritma klasifikasi yaitu decision tree (C4.5), support vector machines (SMO), dan artificial neural networks (MLP) untuk evaluasi dan prediksi kinerja pedagogik di Institusi perguruan tinggi. |
| | Hasil Utama | Hasilnya menunjukkan keakuratan model pengklasifikasi yang memprediksi kinerja pendidik. Decision tree J48 memiliki akurasi lebih tinggi 94,37% dibandingkan SMO dan MLP |
| 7 | Judul | Perbandingan Algoritma C4.5 dan Naïve Bayes Untuk Klasifikasi <i>Mustahik</i> |
| | Nama Peneliti | Awliya Wanhari |
| | Tahun | 2018 |
| | Penerbit | Uinjkt |

| | | |
|---|---------------|--|
| | Sumber | https://repository.uinjkt.ac.id/dspace/bitstream/123456789/53608/1/AWLIYA%20WANHARI-FST.pdf |
| | Metode/Solusi | Menganalisa perbandingan algoritma C4.5 dan <i>Naïve Bayes</i> untuk klasifikasi penentuan <i>mustahik</i> |
| | Hasil Utama | Penggunaan algoritma C4.5 lebih baik dari pada algoritma <i>Naïve Bayes</i> , hal ini dibuktikan tingkat <i>accuracy</i> dimulai dari 75% - 100%, sedangkan algoritma <i>Naïve Bayes</i> dimulai dari 50% - 100% dan <i>execution time</i> masing-masing sama yaitu 0 s, didapatkan melalui <i>tools</i> Rapid Miner Studio dengan menggunakan <i>split validation</i> . |
| 8 | Judul | Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik |
| | Nama Peneliti | Imam Sutoyo |
| | Tahun | 2018 |
| | Penerbit | PILAR Nusa Mandiri |
| | Sumber | Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik Jurnal Pilar Nusa Mandiri |
| | Metode/Solusi | Metodologi yang digunakan adalah CRISP-DM. Algoritma Decision Tree yang diujicoba adalah C4.5 dan Random Forest. Validasi dilaksanakan menggunakan <i>10-Fold Cross Validation</i> untuk dicari algoritma yang memberikan akurasi paling tinggi. |
| | Hasil Utama | Decission Tree C.45 memberikan hasil terbaik dengan akurasi 96,73 %. Oleh karena itu, pada tahap <i>Deployment</i> dari metodologi CRISP-DM, model dan <i>rule</i> dari C4.5 digunakan untuk membuat aplikasi untuk klasifikasi ini |
| 9 | Judul | Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti |
| | Nama Peneliti | Erlin Elisa |
| | Tahun | 2017 |

| | | |
|--------|---------------|--|
| | Penerbit | JOIN |
| | Sumber | Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti Jurnal Online Informatika (uinsgd.ac.id) |
| | Metode/Solusi | Analisis Data Mining dengan menggunakan Algoritma C4.5 untuk mengidentifikasi penyebab terjadinya kecelakaan kerja |
| | Hasil Utama | Metode Algoritma C4.5 atau pohon keputusan lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian. Dari hasil penelitian faktor-faktor yang menjadi penyebab terjadinya kecelakaan kerja kontruksi yang sering terjadi adalah Lingkungan Tempat Kerja, Rambu-Rambu Keselamatan dan Pekerja dan Cara kerja Sistem yang diimplementasikan menggunakan salah satu metode Data Mining algoritma C4.5 dan telah diuji dengan <i>software</i> Data Mining Weka Gui Chooser dalam mengidentifikasi faktor-faktor kecelakaan kerja konstruksi memiliki hasil pengklasifikasian sama |
| 1 0 | Judul | Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu) |
| | Nama Peneliti | Siska Haryati, Aji Sudarsono, Eko Suryana |
| | Tahun | 2015 |
| | Penerbit | Jurnal Media Infotama Vol. 11 No. 2, September 2015 |
| | Sumber | IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: |

| | | |
|---|---------------|---|
| | | UNIVERSITAS DEHAZEN BENGKULU) JURNAL MEDIA INFOTAMA (unived.ac.id) |
| | Metode/Solusi | Mengklasifikasikan kelululan mahasiswa digunakan teknik data mining dengan algoritma C4.5 dan diImplementasikan ke Rapid Miner, |
| | Hasil Utama | Algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analis mahasiswa. Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa tingkat ketepatan waktu mahasiswa menyelesaikan masa studinya. |
| 1 | Judul | Tanggapan Siswa Terhadap Kompetensi Pedagogik Guru Jurusan Administrasi Perkantoran SMK Negeri Klaten |
| 1 | Nama Peneliti | Adilina Hasibuan |
| | Tahun | 2013 |
| | Penerbit | Universitas Negeri Yogyakarta |
| | Sumber | TANGGAPAN SISWA TERHADAP KOMPETENSI PEDAGOGIK GURU JURUSAN ADMINISTRASI PERKANTORAN SMK NEGERI 1 KLATEN. - Lumbung Pustaka UNY |
| | Metode/Solusi | Mengetahui tanggapan siswa terhadap kompetensi pedagogik guru jurusan Administasi Perkantoran SMK N 1 Klaten. Teknik analisis data menggunakan analisis deskriptif kuantitatif dengan persentase. |
| | Hasil Utama | Hasil penelitian dari masing-masing indikator menunjukkan, (1) kemampuan memahami siswa berada pada kategori kurang memahami dengan jumlah frekuensi sebanyak 41 (45,56%), (2) kemampuan mengembangkan siswa berada pada kategori kurang baik dengan jumlah frekuensi sebanyak 42 (46,67%), (3) kemampuan merencanakan pembelajaran belum baik, (4) kemampuan |

| | | |
|--|--|---|
| | | melaksanakan proses pembelajaran berada pada kategori buruk dengan jumlah frekuensi sebanyak 45 (50%), dan (5) kemampuan menilai pembelajaran berada pada kategori kurang baik dengan jumlah ferekuensi sebanyak 50 (55,56%). |
|--|--|---|

2.3 *Matrix* Penelitian

Matrix penelitian menunjukkan perbedaan dan persamaan antara penelitian yang sudah dilakukan dengan penelitian yang akan dilakukan. *Matrix* penelitian dapat dilihat dari Table 2.2 sampai Tabel 2.8.

Tabel 2.2 *Matrix* Penelitian

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|--|---|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naïve Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 1 | Puji Catur Siswipraptini, Ahmad Saputra Fadiarora, Hengki Sikumbang (2023) | Model Klasifikasi Berbasis Machine Learning Untuk Perpanjangan Masa Jabatan Kepala Sekolah Menggunakan Algoritma C4.5 | ✓ | - | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | ✓ |

Tabel 2.3 *Matrix* Penelitian (Lanjutan 1)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|---|---|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naïve Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 2 | Ingkana Devi Cahaya, Heru Satria Tambunan, Harly Okprana (2021) | Analisis Faktor Kompetensi Pedagogik Guru Dalam Proses Pembelajaran Menggunakan Algoritma C45 | ✓ | - | - | ✓ | ✓ | ✓ | ✓ | ✓ | - | - | ✓ |
| 3 | Ahmad Muhajir Hasibuan (2021) | Penerapan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan | ✓ | - | - | ✓ | ✓ | - | - | ✓ | ✓ | - | ✓ |

Tabel 2.4 *Matrix* Penelitian (Lanjutan 2)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|-------------------------------------|--|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naïve Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 4 | Detrinal Putra, Arief Wibowo (2020) | Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes | - | ✓ | - | ✓ | ✓ | - | - | - | ✓ | - | - |

Tabel 2.5 *Matrix* Penelitian (Lanjutan 3)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|---------------------------|---|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naive Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 5 | Rofilde Hasudungan (2018) | Analisis Indikator Kinerja Dosen Terhadap Prestasi Mahasiswa Semester Satu dengan Menggunakan Decision Tree | ✓ | - | - | ✓ | - | - | ✓ | ✓ | ✓ | - | ✓ |
| 6 | Ashily M Baby (2018) | Pedagogue Performance Assessment (PPA) using Data mining Techniques | ✓ | - | - | - | - | - | ✓ | ✓ | - | ✓ | ✓ |

Tabel 2.6 *Matrix* Penelitian (Lanjutan 4)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|-----------------------|--|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naïve Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 7 | Awliya Wanhari (2018) | Perbandingan Algoritma C4.5 dan <i>Naïve Bayes</i> Untuk Klasifikasi <i>Mustahik</i> | ✓ | ✓ | - | - | - | - | - | - | - | ✓ | ✓ |
| 8 | Imam Sutoyo (2018) | Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik | ✓ | - | ✓ | - | ✓ | - | - | - | ✓ | ✓ | ✓ |

Tabel 2.7 *Matrix* Penelitian (Lanjutan 5)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | | |
|----|--------------------|--|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|---|
| | | | Metode | | | | Objek | | | Tujuan | | | | |
| | | | Algoritma C4.5 | Naive Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi | |
| 9 | Erlin Elisa (2017) | Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti | ✓ | - | - | - | - | - | - | - | ✓ | ✓ | - | ✓ |

Tabel 2.8 *Matrix* Penelitian (Lanjutan 6)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|--|--|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naive Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 10 | Siska Haryati, Aji Sudarsono, Eko Suryana (2015) | Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu) | ✓ | - | - | ✓ | - | - | - | - | ✓ | - | - |

Tabel 2.9 *Matrix* Penelitian (Lanjutan 7)

| No | Penulis dan Tahun | Judul | Ruang Lingkup | | | | | | | | | | |
|----|-------------------------|---|----------------|-------------|----------|------------------|-------|------|-----------|----------|--------------|--------------|-------------|
| | | | Metode | | | | Objek | | | Tujuan | | | |
| | | | Algoritma C4.5 | Naive Bayes | CRISP-DM | Confusion Matrix | Siswa | Guru | Pedagogik | Analisis | Implementasi | Perbandingan | Klasifikasi |
| 11 | Adilina Hasibuan (2013) | Tanggapan Siswa Terhadap Kompetensi Pedagogik Guru Jurusan Administrasi Perkantoran SMK Negeri Klaten | - | - | - | - | ✓ | ✓ | ✓ | - | - | - | - |

Pada penelitian sebelumnya yang hampir mendekati merupakan penelitian yang berjudul “Analisis Faktor Kompetensi Pedagogik Guru Dalam Proses Pembelajaran Menggunakan Algoritma C4.5” yang bertujuan menghasilkan decision tree yang akan diambil penentuan berupa *rules* atau aturan-aturan yang akan dijadikan pohon keputusan. Data yang di gunakan adalah data yang diperoleh dari MTs Al-Hurriyah Panei Tongah dengan menyebarkan angket untuk menilai kompetensi pedagogik guru kepada seluruh

siswa siswi di MTs Al-Hurriyah Panei Tengah tahun ajaran 2019 / 2020 semester ganjil dengan hasil akhir cukup, sedang atau baik yang dibagi berdasarkan enam atribut antara lain kemampuan mengelola pembelajaran, pemahaman peserta didik, perancangan pembelajaran, pelaksanaan pembelajaran yang mendidik dan dialogis, pemanfaatan teknologi dan evaluasi belajar. Penelitian tersebut menghasilkan pohon keputusan dengan beberapa kriteria dan yang menjadi kriteria terbaik dalam proses pembelajaran adalah mengevaluasi hasil pembelajaran. Untuk mengukur performa menggunakan *performance vector* menghasilkan nilai akurasi sebesar 87,90%, pada prediksi label sedang sebesar 83,33% dan prediksi label baik sebesar 91,67% yang diperoleh dengan pengaturan pada operator *split validation* dengan nilai *split ratio* = 0,5, dan *sampling type* = *shuffled sampling*.

Pada penelitian yang lain dengan judul “Analisis Indikator Kinerja Dosen Terhadap Prestasi Mahasiswa Semester Satu dengan Menggunakan Decision Tree” penelitian ini mengungkapkan pengaruh pedagogik dosen terhadap prestasi mahasiswa pada mata kuliah dasar pemrograman pada mahasiswa program studi informatika Universitas Muhammadiyah Kalimantan Timur. Pada penelitian ini melakukan kuesioner kepada mahasiswa berdasarkan boring evaluasi dosen terhadap mata kuliah dasar pemrograman. Terhadap 48 mahasiswa semester satu yang terlibat pada penelitian ini dengan indikator penilaian sebanyak 28. Analisis data menggunakan teknik pohon keputusan menghasilkan dari 28 indikator yang ada, hanya 5 indikator yang berpengaruh terhadap prestasi mahasiswa. Dengan menggunakan rasio *data training* dan *data testing* 70:30 dan menggunakan *cross-validation* untuk melakukan

pengukuran performa. Hasil evaluasi dengan menggunakan cross-validation menghasilkan tingkat akurasi dari model ialah 74% dengan standar deviasi 18.41%.

Pada penelitian ini berfokus pada kinerja pedagogik guru dan tingkat kecurangan siswa yang bertujuan menghasilkan suatu model data mining berupa pohon keputusan yang digunakan untuk menjelaskan pola atau hubungan dalam data kinerja pedagogik guru dan tingkat kecurangan siswa. Data yang di gunakan adalah data yang diperoleh dengan menyebarkan kuesioner kepada responden penelitian yaitu siswa kelas 11 Madrasah Aliyah Negeri 1 Kota Tasikmalaya tahun ajaran 2023/2024 semester ganjil, pada kuesioner tingkat kecurangan memiliki 6 atribut berupa jenis ujian dan pada kuesioner memiliki 7 atribut berupa aspek pedagogik guru. Dengan perbandingan rasio *data training* 80:20 dan 50:50, untuk mengukur performa menggunakan metode *Confusion Matrix*.