

KLASIFIKASI PENCAPAIAN NILAI MAHASISWA BERDASARKAN JENIS MATA KULIAH MENGGUNAKAN ALGORITMA NAIVE BAYESIAN CLASSIFISIER

Charles Rasyid Batuallo¹, Alfrina Mewengkang², Daniel Riano Kaparang³

^{1,2,3}Jurusan Pendidikan Teknologi Informasi dan Komunikasi, Fakultas Teknik,
Universitas Negeri Manado

e-mail: ¹charlesrasyid17@gmail.com, ²mewengkangalfrina@unima.ac.id,

³drkaparang@unima.ac.id

ABSTRAK

Tujuan penelitian ini untuk mengelompokkan data nilai mahasiswa berdasarkan atribut yang ada serta menentukan kebijakan dalam menentukan bakat dan kemampuan mahasiswa berdasarkan hasil klasterisasi. Metode yang digunakan pada penelitian ini adalah deskriptif dengan pendekatan kuantitatif. Penelitian deskriptif mengkaji antar dua variabel atau lebih, variabel pada penelitian ini bertujuan untuk melihat tingkat prestasi akademik mahasiswa Nilai Mahasiswa Universitas Negeri Manado Fakultas Teknik Jurusan Pendidikan Teknologi Informasi Dan Komunikasi. Hasil penelitian yang diperoleh dari penelitian ini (1) Impelementasi Naive bayes menggunakan Aplikasi WEKA dapat menelusuri karakteristi katribut dari dataset dengan Lulus dan Tidak Lulus. Metode Klasifikasi Naive Bayes menggunakan data training untuk menciptakan probabilitas setiap kriteria untuk berbagai kelas yang berbeda, dengan tujuan agar nilai-nilai probabilitas dari kriteria ini dapat ditingkatkan untuk memprediksi status data nilai mahasiswa proses klasifikasi yang dilakukan dengan metode Klasifikasi Naive Bayes. (2) Berdasarkan data nilai mahasiswa yang dijadikan data testing, metode Naive Bayes berhasil mengklasifikasikan 3.929 data yang diuji. Di mana dari 3.928 data mahasiswa, untuk Correctly Classified Instance sebanyak 99,9745 % sementara persentase untuk Incorrectly Classified Instance adalah sebesar 0,0255 %. Dimana dari 3.928 data klasifikasi pencapaian nilai mahasiswa dengan menggunakan Naive Bayesian Classifisier berhasil diklasifikasikan dengan benar dan sebanyak 1 data penggunaan klasifikasi pencapaian nilai mahasiswa tidak berhasil diklasifikasikan dengan benar.

Kata Kunci: Klasifikasi, Nilai Mahasiswa, Naive Bayesian Classifisier.

PENDAHULUAN

Perkembangan teknologi informasi telah mencapai tingkat pesat dalam berbagai aspek kehidupan saat ini. Kemajuan ini membawa banyak kemudahan dalam mengakses berbagai data dan informasi, terutama dalam bidang ekonomi, industri, dan ilmu pengetahuan. Selain itu, teknologi informasi juga memiliki peran penting dalam dunia pendidikan. Di samping itu, bidang komputasi numeris juga mengalami perkembangan yang signifikan seiring dengan kemajuan teknologi. Pendekatan statistik dan penggunaan

algoritma tertentu kini digunakan secara luas dalam memecahkan masalah yang melibatkan ketidakpastian (Kurdhi dkk, 2023).

Pendidikan tinggi memiliki peran yang sangat penting sebagai lembaga pendidikan yang memberikan layanan kepada masyarakat (Rabiah, 2019). Kualitas pendidikan tinggi berpengaruh langsung terhadap kemajuan dan kesejahteraan bangsa Indonesia. Salah satu faktor yang memengaruhi kredibilitas institusi perguruan tinggi adalah kualitas mahasiswanya. Untuk meningkatkan kualitas pendidikan, institusi-institusi tersebut harus menyesuaikan diri dengan perkembangan teknologi saat ini, sehingga pendidikan yang diberikan dapat mencapai hasil yang optimal. Selain itu, pemilihan konsentrasi yang tepat juga berpengaruh terhadap minat belajar mahasiswa, serta keterlibatan dan ketekunan mereka dalam mengerjakan tugas dan ujian selama perkuliahan, yang pada akhirnya berdampak pada pencapaian Indeks Prestasi Akademik (IPK) yang optimal.

Data yang diperoleh dari kelompok nilai yang dimaksud bisa juga menjadi salah satu referensi Perguruan Tinggi dalam menilai kinerja dosen pengampu mata kuliah. Parameter data yang dimaksud dalam penelitian ini meliputi mata kuliah, kode mata kuliah jumlah SKS, nomor masiswa. Pengelompokan ini mengabaikan latar belakang Mahasiswa dan dosen yang tentunya berbeda – beda sehingga hasil akhir dari pendekatan ini tidak menjadi benar-benar ideal tetapi diharapkan mendekati kesimpulan secara umum. Data ini nantinya akan dikelola oleh pihak perguruan tinggi untuk berbagai kepentingan administrasi, atau sebagai acuan identifikasi terhadap mahasiswa dan dosen untuk membantu pihak Perguruan Tinggi menentukan kebijakan-kebijakan kedepan, sehingga data klasifikasi yang dihasilkan menjadi penting bagi Perguruan Tinggi.

Standar penilaian pada Standar Nasional Pendidikan Tinggi (SN-DIKTI) mencakup lama masa studi sebagai salah satu faktor dalam mengukur akreditasi perguruan tinggi. Menurut PERMENRISTEKDIKTI Nomor 44 Tahun 2015, masa studi maksimal untuk program sarjana adalah 7 tahun akademik dengan beban belajar minimal 144 SKS. Mahasiswa sarjana menempuh masa studi antara 8 hingga 14 semester, dan mereka yang melebihi 8 semester dianggap menempuh masa studi yang lama.

Data mengenai mahasiswa dan jumlah kelulusan mereka memberikan informasi berharga, termasuk jumlah kelulusan tahunan, profil, dan pencapaian akademik selama proses belajar mengajar di Pendidikan Teknologi Informasi dan Komunikasi Fakultas Teknik Unima. Informasi mengenai lama masa studi mahasiswa menjadi penting sebagai panduan bagi manajemen Fakultas Teknik Unima dalam pengambilan keputusan berikutnya.

Dalam data mining, klasifikasi adalah masalah utama. Dalam klasifikasi, pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan sebelumnya. Tingkat ketepatan biasanya digunakan untuk mengukur kinerja pengklasifikasi. Digunakan untuk memprediksi kemungkinan keanggotaan dalam kelas tertentu, pengklasifikasi Bayesian adalah teknik statistik. Perhitungan probabilitas hipotesis, probabilitas kelas dari kelompok atribut yang ada, dan penentuan kelas yang paling optimal adalah semua bagian dari pendekatan ini.

Oleh karena itu, maka peneliti berasumsi bahwa perlu adanya analisis terhadap sekumpulan data nilai mahasiswa dengan mengkombinasikan beberapa atribut yang

memenuhi syarat. Pada Universitas Negeri Manado, khususnya dalam pemanfaatan data capaian nilai mahasiswa. Penelitian ini sama sekali belum pernah dilakukan. Hasil analisis diharapkan dapat memberikan informasi mengenai Daya serap mahasiswa yang terwujud dalam nilai mahasiswa untuk masing-masing mata kuliah.

Berdasarkan masalah di atas, penelitian ini akan berkonsentrasi pada meningkatkan akurasi klasifikasi nilai siswa dengan menggunakan algoritma klasifikasi naive bayesian yang digunakan dengan metode software WEKA. Penulis tertarik untuk mengadakan penelitian dengan judul : *Klasifikasi pencapaian nilai mahasiswa berdasarkan jenis mata kuliah menggunakan algoritma naive bayesian classifisier* dengan Fakultas Teknik Universitas Negeri Manado Sebagai studi kasus penelitian.

KAJIAN TEORI

Nilai Mahasiswa

Dalam pendidikan tinggi, mahasiswa sangat penting. Dalam sistem pendidikan formal, perguruan tinggi adalah institusi pendidikan tertinggi (Dimiyati, 2018). Perguruan tinggi dibagi menjadi dua kategori: negeri dan swasta berdasarkan bagaimana mereka dimiliki.

Prestasi merupakan hasil dari suatu usaha yang kompleks (Ekayani, 2017). Prestasi dapat didefinisikan sebagai pencapaian seseorang dalam melakukan suatu tindakan. Setiap orang dapat menginterpretasikan prestasi dengan cara yang berbeda. Sebagai contoh, seorang penyapu jalanan dapat dianggap meraih prestasi jika dia mampu membersihkan jalanan tanpa menyisakan sampah sedikit pun (Nurdiaman, 2019).

Prestasi akademik memiliki beberapa fungsi penting. Pertama, prestasi akademik digunakan sebagai acuan untuk menilai kualitas dan kuantitas pengetahuan yang dikuasai oleh individu, serta sebagai indikator untuk mengukur kemampuan peserta didik. Kedua, prestasi akademik berperan sebagai motivasi untuk meningkatkan pengetahuan dan kinerja akademik, dan menjadi simbol dari semangat keingintahuan. Ketiga, prestasi akademik berfungsi sebagai indikator baik secara internal maupun eksternal bagi lembaga pendidikan untuk mengevaluasi keberhasilan dalam mencapai tujuan pendidikan (Pradiptasari, 2016).

Beberapa hal yang harus diperhatikan dalam pembelajaran siswa untuk mencapai keberhasilan akademik. Secara umum faktor-faktor yang mempengaruhi prestasi akademik mahasiswa terbagi menjadi dua kategori, yaitu faktor internal dan eksternal (Abdullah dkk, 2019).

Data Mining

Penggunaan pembelajaran mesin untuk menganalisis dan mengekstrak data secara otomatis dikenal sebagai penambangan data. Ini membutuhkan identifikasi informasi penting dari informasi yang ada dalam basis data untuk menghasilkan informasi yang berharga. Metode ini menggunakan AI, matematika, statistik, dan pembelajaran mesin untuk menemukan data dalam basis data yang sangat besar (Utomo dan Mesran, 2020). Penambangan data berfungsi sebagai alat untuk menganalisis kumpulan data, mengenali pola, dan membuat prediksi. Tujuan meliputi prediksi, deskripsi, klasifikasi,

pengelompokan, deteksi perubahan dan anomali, regresi dan peringkasan data. Penambangan data diperlukan untuk menemukan informasi dan wawasan berharga saat memproses data dalam jumlah besar.

1. Pemilihan Informasi
2. Pemrosesan Data
3. Konversi Data
4. Evaluasi

Teknik Klasifikasi

Proses klasifikasi melibatkan pencarian model atau fungsi yang dapat membedakan kelas data atau konsep untuk memprediksi kelas objek yang tidak memiliki nama kelas. Klasifikasi melibatkan klasifikasi catatan baru ke dalam kelas atau kategori yang telah ditentukan Menurut (Andita dkk, 2019).

Ada dua jenis model klasifikasi :

1. Pemodelan Deskriptif: Model klasifikasi ini digunakan sebagai alat penjelas untuk membedakan item dari berbagai kelas.
2. Pemodelan Prediktif: Label kelas dari daftar yang tidak diketahui diprediksi dengan menggunakan model klasifikasi.

Algoritma Klasifikasi

Algoritma Naive Bayes adalah teknik klasifikasi probabilistik yang melihat kemungkinan bahwa suatu objek dengan atribut tertentu akan termasuk dalam kelompok atau kelas tertentu (Mukminin dan Riana, 2017). Tujuan dari metode ini adalah untuk meminimalkan variasi antara data dalam satu kelompok (cluster) dan memaksimalkan variasi antara data dalam kelompok lain.

Proses pengelompokan Naive Bayesian adalah metode klasifikasi statistik berdasarkan teorema Bayes. Teori keputusan Bayesian adalah pendekatan statistik dasar untuk pengenalan pola, dan penggunaan algoritma ini dalam klasifikasi bergantung pada aspek statistik dari masalah yang dihadapi.

Waikato Environment for Knowledge (WEKA)

Lingkungan Waikato untuk Analisis Pengetahuan (WEKA) adalah sistem penambangan data sumber terbuka yang dibuat oleh Universitas Waikato di Selandia Baru. Itu menggunakan bahasa pemrograman Java dan tersedia secara bebas di bawah lisensi umum GNU. WEKA mendukung seluruh proses penambangan data, termasuk penilaian pembelajaran, persiapan input data, rencana statistik, visualisasi input data, dan hasil pembelajaran. Karena sistem WEKA mendukung teknik seperti pengolahan data, clustering, klasifikasi, regresi, visualisasi, dan reduksi fitur, metode atau teknik yang digunakan dalam WEKA bersifat deskriptif dan prediktif (Witten et al., 2011).

WEKA merupakan software data mining open source yang populer sejak tahun 1994. Dikembangkan oleh University of Waikato di Selandia Baru, software ini menawarkan beberapa keunggulan seperti menyediakan banyak algoritma data mining dan machine learning, mudah digunakan dan selalu up to date. dengan perkembangan.

algoritma terakhir. WEKA tidak hanya digunakan dalam konteks akademik, tetapi juga sering digunakan oleh perusahaan untuk memprediksi jalannya bisnis perusahaannya. WEKA mendukung beberapa format file input, antara lain:

METODOLOGI PENELITIAN

Pengumpulan Data

Pada langkah ini, peneliti berhasil mengumpulkan data yang diperlukan. Data utama yang dikumpulkan adalah data nilai mahasiswa Fakultas Teknik, Universitas Negeri Manado dari tahun 2018 hingga 2020 untuk jenjang S1. Jumlahnya mencakup 99 data nilai mahasiswa dan 111 mata kuliah. Data ini diperoleh secara langsung dari operator Jurusan PTIK Universitas Negeri Manado untuk memenuhi kebutuhan penelitian. Peneliti memilih data berdasarkan atribut yang relevan dengan kurikulum tahun 2018.

Atribut yang didapatkan antara lain:

- Nomor Induk Mahasiswa,
- Nama mata kuliah,
- Bobot nilai,
- Kode Mata kuliah,
- Data Nilai Mahasiswa.

Persiapan Data

Untuk memulai proses pemilihan data yang akan dianalisis dengan metode Naive Bayes, yang pertama adalah membaca data latih. Data didapatkan secara riil dari operator Jurusan PTIK Universitas Negeri Manado untuk memenuhi kebutuhan penelitian, maka peneliti mengambil data berdasarkan atribut yang memiliki kesamaan yang menggunakan kurikulum 2018.

Data pada tabel 1 diperoleh data-data statistik sebagai berikut: Jumlah = 11.358; Mean = 2,89; Median = 3; Modus = 4; Simpangan baku 1,30; Variance = 1,68.

Tabel 1. Data Statistik

Jumlah (Σ)	11.358
Mean (\bar{x})	2,89
Nilai Maksimum	4
Nilai Minimum	0
Simpangan Baku	1,30
Varians	1,68
Median	3
Modus	4

Melalui tabel 2 dapat dilihat bahwa Lulus = 3281/3929 Jumlah data nilai mahasiswa yang lulus dengan dibagi dengan jumlah *data training*. Tidak Lulus = 648/3929 Jumlah data tidak lulus dengan atribut nilai dengan jumlah *data training*.

Tabel 2. Atribut Nilai

Nilai	Lulus	Tidak Lulus	Jumlah
A	1647	-	1647
B	1264	-	1264
C	370	-	370
D	-	238	238
E	-	410	410
Total	3281	648	3929

Tabel 3. Atribut Predikat

Predikat	Lulus	Tidak Lulus	Jumlah
Sangat Memuaskan	1647	-	1647
Memuaskan	1264	-	1264
Kurang Memuaskan	370	-	370
Kurang	-	238	238
Sangat Kurang	-	410	410
TOTAL	3281	648	3929

Dari tabel 3 didapati bahwa Lulus = $3281 / 3929$ Jumlah data nilai mahasiswa yang lulus dengan atribut nilai dibagi dengan jumlah *data training*. Tidak Lulus = $648 / 3929$ Jumlah data tidak lulus dengan atribut nilai dengan jumlah *data training*.

Dengan mengacu pada data nilai mahasiswa yang tertera pada Tabel 1 dan Tabel 2, dapat disimpulkan bahwa jumlah data latih (*data training*) yang tersedia sebanyak 3.929 data. Dari jumlah tersebut, 1.647 data memenuhi Kriteria A (Sangat Memuaskan).

Tabel 4. Probabilitas Kriteria

Nilai	Predikat	Lulus	Probabilitas
A	Sangat Memuaskan	1.647	0,419
B	Memuaskan	1.264	0,322
C	Kurang Memuaskan	370	0,094
D	Kurang	238	0,061
E	Sangat Kurang	410	0,104

Untuk mengestimasi probabilitas akhir dari setiap kelas, dapat menggunakan data *training* yang terdapat pada tabel 1 dan tabel 2 Data ini kemudian akan diubah menjadi nilai yang telah ditentukan dalam proses perhitungan probabilitas untuk setiap atribut. Probabilitas kelas akan dikalikan dengan nilai atribut yang relevan. Selanjutnya, hasil yang telah ditentukan untuk setiap kelas akan dibandingkan, dan nilai yang paling tinggi atau mendekati 1 akan dipilih. Jika probabilitas kelas tepat waktu memiliki nilai yang mendekati 1, maka hasilnya akan diklasifikasikan sebagai tepat waktu, begitu pula sebaliknya.

Berdasarkan data nilai mahasiswa pada tabel 2 diketahui jumlah data latih (*data training*) adalah sebanyak 3.929 data, dimana dari 3281 yang memperoleh Kriteria Lulus

(A; B; C) dan 648 yang memperoleh Kriteria Tidak Lulus (D, E). Kontribusi nilai secara probabilitas (Lulus) = $3281 / 3.929 = 0,835$; dengan persentase $0,835 \times 100\% = 83,51\%$. Sedangkan kontribusi nilai secara probabilitas Tidak Lulus = $648 / 3.929 = 0,165$, dengan persentase $0,165 \times 100\% = 16,49\%$.

Pemodelan model data, model olah data Algoritma *Naive Bayes*

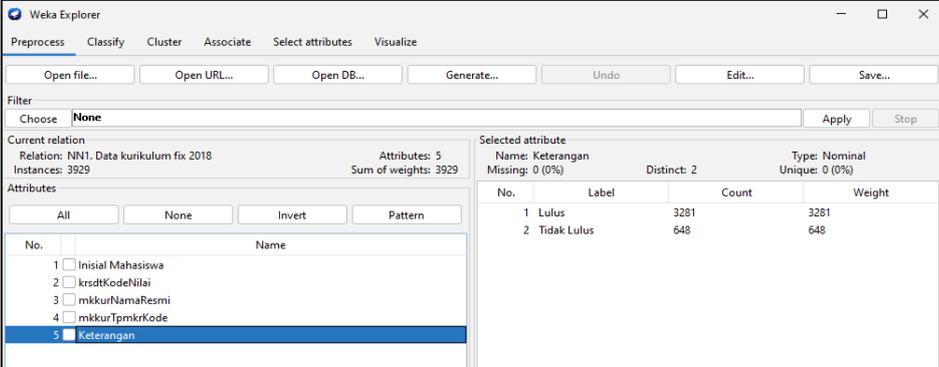
Aplikasi Weka akan menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokan pilihan diklasifikasikan Pencapaian Nilai Mahasiswa dilakukan berdasarkan atribut terpilih yaitu Kode Nilai, Mata Kuliah, Kode Mata Kuliah.

Pada *Weka Classifier* juga dapat dilihat rincian kelompok pencapaian nilai mahasiswa dengan *Naive Bayesian Classifier*. Probabilitas LULUS dan Probabilitas TIDAK LULUS pada masing-masing mata kuliah dan didapatkan hasil bahwa *class* Probabilitas LULUS yang memperoleh nilai A, B dan C lebih tinggi dibandingkan *class* Probabilitas TIDAK LULUS yang memperoleh nilai D dan E.

Proses dataset dilakukan dengan menggunakan teknik klasifikasi *Naive Bayes Updateable* untuk mengklasifikasikan pencapaian nilai mahasiswa. Dalam hal ini, digunakan jenis uji coba yang disebut sebagai training set, karena tujuannya adalah untuk menemukan pola baru dalam data yang dapat menghubungkan pola data yang sudah ada dengan data baru. Panel pengklasifikasi memungkinkan konfigurasi dan implementasi salah satu metode pengklasifikasian yang dipilih pada dataset yang tersedia.

Tahap Evaluasi

Menggunakan *Tools Weka Naive Bayes Classifier* menjadi solusi yang sangat efektif untuk membantu menentukan kebijakan-kebijakan berikutnya. Setelah mengamati, memilah dan memproses data, sistem harus memberikan masukan yang akan membantu user mengambil kesimpulan. Dengan menggunakan aplikasi Weka, akan dilakukan analisis terhadap atribut-atribut dalam dataset Nilai Mahasiswa. Pengelompokan pilihan lulus akan dilakukan berdasarkan atribut yang terpilih, yaitu nilai A, nilai B, nilai C, nilai D, dan nilai E. Data yang akan diuji sejumlah 3.929 data dan akan diproses menggunakan tools Weka untuk menghasilkan klasifikasi pencapaian nilai mahasiswa menggunakan metode *Naive Bayesian Classifier* seperti yang terlihat pada gambar 2.



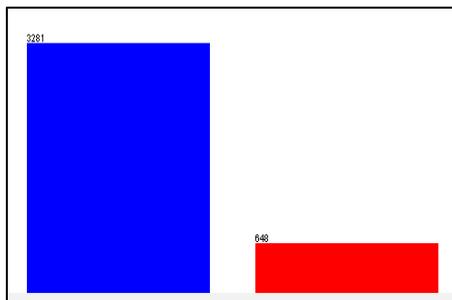
The screenshot shows the Weka Explorer interface. The 'Selected attribute' table is displayed, showing the distribution of 'Keterangan' values. The table has columns for 'No.', 'Label', 'Count', and 'Weight'.

No.	Label	Count	Weight
1	Lulus	3281	3281
2	Tidak Lulus	648	648

Gambar 2. Keterangan Lulus / Tidak Lulus

Hasil dari klasifikasi pencapaian nilai mahasiswa berdasarkan jenis mata kuliah menggunakan algoritma *naive bayesian classifisier* diklasifikasikan dengan Kriteria A (Sangat Memuaskan); Probabilitas Kriteria B (Memuaskan); Probabilitas Kriteria C (Kurang Memuaskan); Probabilitas Kriteria D (Kurang); Probabilitas Kriteria E (Sangat Kurang). Divisualisasikan dengan diagram batang dengan klasifikasi pencapaian nilai mahasiswa yang terdiri dari 2 kategori pilihan. Diklasifikasikan dimana dari 3.281 yang memperoleh Kriteria **Lulus** (A; B; C) dan 648 yang memperoleh Kriteria **Tidak Lulus** (D, E). Kontribusi nilai secara probabilitas (Lulus) = $3.281 / 3.929 = 0,835$, dengan persentase $0,835 \times 100\% = 85,67\%$. Kontribusi nilai secara probabilitas Tidak Lulus = $648 / 3.929 = 0,143$, dengan persentase $0,143 \times 100\% = 13,33\%$.

Proses klasifikasi dipengaruhi oleh atribut-atribut terpilih yang mendukung untuk menentukan kelompok Kriteria Lulus (A; B; C), dan Kriteria Tidak Lulus (D, E). Hasil dari klasifikasi divisualisasikan dengan diagram batang yang dapat dilihat pada gambar 3. Hasil dari diagram akan menampilkan *class* Kriteria Lulus, dan Tidak Lulus (D, E).



Gambar 3. Diagram Klasifikasi Nilai Mahasiswa

```

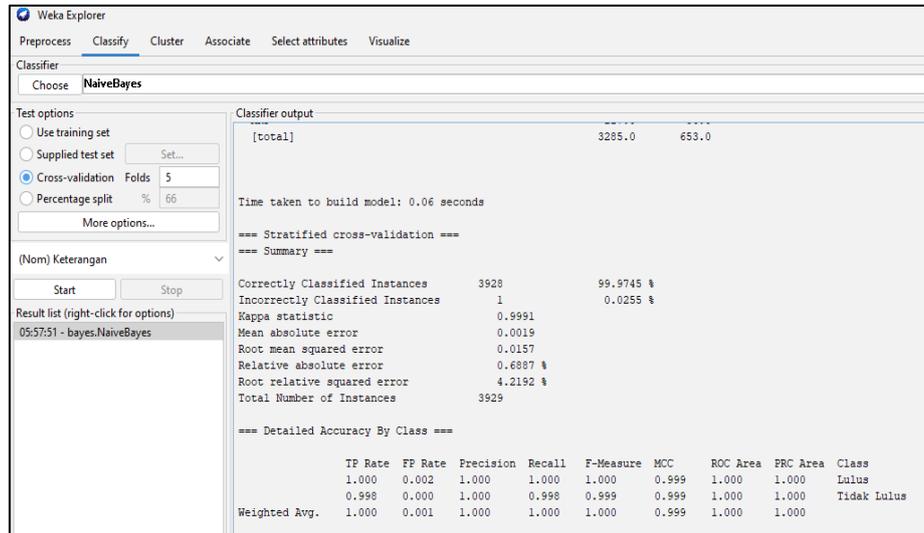
=== Run information ===
Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    ML1_Data_kurikulum_sis_2018
Instances:   3929
Attributes:  5
              Inisial Mahasiswa
              kramasKodeFile1
              skKurSiswaKewal
              skKurTipeKode
              Keterangan
Test mode:   5-fold cross-validation

=== Classifier model (full training set) ===
Naive Bayes Classifier

Attribute                                     Class
                                             Lulus Tidak Lulus
                                             (0.83) (0.17)
-----
Inisial Mahasiswa
Mahasiswa 1                                15.0    2.0
Mahasiswa 2                                12.0    1.0
Mahasiswa 3                                17.0    1.0
Mahasiswa 4                                17.0    1.0
Mahasiswa 5                                 9.0     2.0
Mahasiswa 6                                19.0   12.0
Mahasiswa 7                                10.0   10.0
Mahasiswa 8                                 9.0     0.0
Mahasiswa 9                                 6.0     0.0
Mahasiswa 10                               18.0   14.0
Mahasiswa 43                                14.0    5.0
Mahasiswa 44                                13.0    5.0
Mahasiswa 45                                17.0    2.0
Mahasiswa 46                                 4.0     7.0
Mahasiswa 47                                 2.0     9.0
Mahasiswa 48                                 3.0     0.0
Mahasiswa 49                                42.0    1.0
Mahasiswa 70                                10.0   12.0
    
```

Gambar 4. Diagram Klasifikasi Nilai Mahasiswa

Setelah mendapatkan nilai probabilitas seperti yang tercantum di pada gambar 4, data akan diuji menggunakan sejumlah 3.928 data dan dianalisis dengan menggunakan aplikasi Weka. Hasil dari analisis tersebut akan menghasilkan klasifikasi penggunaan listrik, sebagaimana ditampilkan pada gambar 5.



Gambar 5. Hasil Evaluasi

Berdasarkan gambar 5, dapat dilihat persentase untuk diklasifikasikan secara benar (Correct Classified Instances) adalah sebesar 99,9745 % sementara persentase secara tidak benar untuk (Incorrectly Classified Instance) adalah sebesar 0,0255 %. Dimana dari 3.928 data klasifikasi pencapaian nilai mahasiswa dengan menggunakan Naive Bayesian Classifier berhasil diklasifikasikan dengan benar dan sebanyak 1 data penggunaan klasifikasi pencapaian nilai mahasiswa tidak berhasil diklasifikasikan dengan benar.

Untuk mengevaluasi tingkat akurasi dari perhitungan yang telah dilakukan dalam penelitian ini, dilakukan evaluasi menggunakan confusion matrix. Data testing telah diolah, dan hasilnya memberikan akurasi pada model yang telah dibuat. Akurasi model ini dihitung dengan menggunakan confusion matrix, dengan metode naive bayes classifier. Tabel berikut menjelaskan confusion matrix, dengan huruf a dan b yang masing-masing menunjukkan class a = LULUS dan class b = TIDAK LULUS. Pengolahan ini menggunakan total 3.929 rekaman data.

```
=== Confusion Matrix ===

  a   b  <-- classified as
3281  0  |   a = Lulus
  1 647 |   b = Tidak Lulus
```

Gambar 6. Confusion Matrix

Berdasarkan confusion matrix pada gambar 6, Untuk menghitung persentase akurasi dari data yang diolah, digunakan formula berikut sebagai metode pengukuran akurasi (accuracy):

$$\text{Presentase Akurasi} = \frac{\text{Banyaknya Prediksi Yang Lulus}}{\text{Total Bayaknya Prediksi}} \times 100 \%$$

$$\text{Presentase Akurasi} = \frac{3.281 + 647}{(3.281 + 0) + (647 + 1)} \times 100 \%$$

$$= \frac{3928}{3929} \times 100 \%$$

$$= 99,9745 \%$$

Persentase keakuratan menunjukkan tingkat efektivitas dari dataset klasifikasi pencapaian nilai mahasiswa yang diterapkan menggunakan metode Naive Bayes Classification, dengan tingkat akurasi mencapai 99,9745%.

Untuk menghitung persentase tingkat kesalahan (error rate) dari data yang diolah, digunakan formula berikut:

$$\text{Presentase Akurasi} = \frac{\text{Banyaknya Prediksi Yang Tidak Lulus}}{\text{Total Bayaknya Prediksi}} \times 100 \%$$

$$= \frac{1}{(3281 + 0) + (647 + 1)} \times 100 \%$$

$$= \frac{1}{3929} \times 100 \%$$

$$= 0,0255 \%$$

Berdasarkan analisis terhadap model klasifikasi yang telah dibangun, dapat disimpulkan bahwa metode Naive Bayes Classifier menghasilkan klasifikasi yang baik untuk data nilai mahasiswa. Evaluasi menggunakan confusion matrix menunjukkan tingkat akurasi sebesar 99,9745%. Oleh karena itu, dapat disimpulkan bahwa model klasifikasi yang telah dibuat layak untuk diterapkan dalam konteks ini.

Pembahasan

Pendekatan klasifikasi Bayesian merupakan pendekatan statistik yang didasarkan pada Teorema Bayes. Metode ini digunakan dalam pengenalan pola dengan tujuan melakukan klasifikasi berdasarkan statistik yang dapat diamati dari data. Naive Bayes, sebagai salah satu implementasi dari metode Bayesian, didasarkan pada asumsi sederhana bahwa atribut-atribut memiliki ketergantungan yang bebas satu sama lain jika diberikan nilai output. Keuntungan penggunaan metode Naive Bayes adalah kemampuannya untuk bekerja dengan dataset yang relatif kecil dalam menentukan estimasi parameter yang diperlukan untuk proses klasifikasi.

Naive Bayes adalah metode klasifikasi yang menggunakan probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Dalam klasifikasi pencapaian nilai mahasiswa, penggunaan metode Naive Bayes Classification dapat menggambarkan efektivitas dataset dengan menghitung presentase keakuratan.

Untuk melaksanakan analisis menggunakan metode Naive Bayes, langkah awal yang harus dilakukan adalah membaca data latih yang akan digunakan dalam proses klasifikasi. Data didapatkan secara riil dari operator Jurusan PTIK Universitas Negeri Manado untuk memenuhi kebutuhan penelitian, maka peneliti mengambil data berdasarkan atribut yang memiliki kesamaan menggunakan kurikulum 2018.

Atribut yang didapatkan antara lain:

- Nomor Induk Mahasiswa,
- Kode Mata kuliah,
- Nama mata kuliah,
- Bobot nilai,
- Data Nilai Mahasiswa.

Data tersebut diperoleh data-data statistik sebagai berikut: Jumlah = 11.358; Mean = 2,89; Median = 3; Modus = 4; Simpangan baku 1,30; Variance = 1,68. Lulus = 3281 / 3.929 Jumlah data nilai mahasiswa yang lulus dengan atribut nilai dibagi dengan jumlah data training. Tidak Lulus = 648 / 3.929 Jumlah data tidak lulus dengan atribut nilai dengan jumlah data training. Data nilai mahasiswa diketahui jumlah data latih (data training) adalah sebanyak 3.929 data, dimana dari 1.647 yang memperoleh Kriteria A (Sangat Memuaskan); dari 1.264 yang memperoleh Kriteria B (Memuaskan); dari 370 yang memperoleh Kriteria C (Kurang Memuaskan); dari 370 yang memperoleh Kriteria C (Kurang Memuaskan); dari 238 yang memperoleh Kriteria D (Kurang) dan dari 410 yang memperoleh Kriteria E (Sangat Kurang).

Data nilai mahasiswa diketahui jumlah data latih (data training) adalah sebanyak 3.929 data, dimana dari 3281 yang memperoleh Kriteria Lulus (A; B; C) dan 648 yang memperoleh Kriteria Tidak Lulus (D, E). Kontribusi nilai secara probabilitas (Lulus) = $3.281 / 3.929 = 0,833$, dengan persentase $0,833 \times 100\% = 83,31\%$. Sedangkan kontribusi nilai secara probabilitas Tidak Lulus = $648 / 3.929 = 0,165$, dengan persentase $0,165 \times 100\% = 16,49\%$.

Aplikasi Weka akan menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokkan pilihan diklasifikasikan Pencapaian Nilai Mahasiswa dilakukan berdasarkan atribut terpilih yaitu Kode Nilai, Mata Kuliah, Kode Mata Kuliah. Probabilitas Lulus dan Probabilitas Tidak Lulus pada masing-masing mata kuliah dan didapatkan hasil bahwa class Kriteria Lulus (A; B; C), dan Kriteria Tidak Lulus (D, E). Hasil dari klasifikasi divisualisasikan dengan diagram batang. Hasil dari diagram akan menampilkan class Kriteria Lulus, dan Tidak Lulus (D, E).

Hasil dari klasifikasi pencapaian nilai mahasiswa berdasarkan jenis mata kuliah menggunakan algoritma naive bayesian classifisier divisualisasikan dengan diagram batang dengan klasifikasi pencapaian nilai mahasiswa yang terdiri dari 2 kategori pilihan. Diklasifikasikan dimana dari 3.281 yang memperoleh Kriteria Lulus (A; B; C) dan 648 yang memperoleh Kriteria Tidak Lulus (D, E). Kontribusi nilai secara probabilitas

(Lulus) = $3.281 / 3.929 = 0,835$, dengan persentase $0,835 \times 100\% = 85,67\%$. Kontribusi nilai secara probabilitas Tidak Lulus = $648 / 3.929 = 0,143$, dengan persentase $0,143 \times 100\% = 13,33\%$.

Pada pengolahan data ini, terdapat total 803 record yang digunakan. Hasil dari confusion matrix menunjukkan bahwa terdapat 3.281 record pada class a yang diprediksi dengan tepat sebagai class a. Tidak ada record yang diprediksi tidak tepat sebagai kelompok data class a, sehingga semuanya terklasifikasikan dengan benar. Selanjutnya, dari 647 record pada class b, semuanya diprediksi dengan tepat sebagai class b. Namun, terdapat 1 record yang diprediksi tidak tepat sebagai kelompok data class b, karena diklasifikasikan sebagai class Tidak Lulus.

Untuk menghitung persentase akurasi dari data yang telah diolah, digunakan formula pengukuran akurasi (accuracy). Nilai persentase keakuratan menunjukkan seberapa efektif dataset klasifikasi pencapaian nilai mahasiswa yang telah diterapkan menggunakan metode Naive Bayes Classification, yang mencapai tingkat akurasi sebesar 99,9745%.

Dalam menghitung persentase akurasi dari data yang telah diolah, digunakan formula pengukuran akurasi (accuracy). Namun, perlu diperhatikan bahwa terdapat kesalahan pada kalimat sebelumnya, yaitu akurasi (accuracy) yang Tidak Lulus. Sebagai hasil evaluasi menggunakan confusion matrix, metode Naive Bayes Classifier menunjukkan hasil yang baik dalam mengklasifikasikan data nilai mahasiswa. Nilai akurasi yang diperoleh adalah sebesar 0,0255%.

Naive Bayes merupakan sebuah algoritma klasifikasi yang mampu memprediksi kemungkinan atau probabilitas di masa depan berdasarkan data yang telah ada di masa sebelumnya. Algoritma ini telah terbukti memberikan hasil yang baik saat diterapkan pada berbagai kasus dengan jumlah data yang besar. Semakin banyak data training yang digunakan, semakin tinggi pula tingkat akurasi yang dapat dicapai dalam proses klasifikasi. Dengan demikian, algoritma Naive Bayes memiliki kemampuan yang handal dalam mengklasifikasikan data dengan tingkat akurasi yang tinggi.

KESIMPULAN

Berdasarkan hasil penelitian di Jurusan Pendidikan Teknologi Informasi Dan Komunikasi maka dapat disimpulkan hasil penelitian sebagai berikut:

1. Metode Naive Bayes memungkinkan pengklasifikasian data dengan memanfaatkan probabilitas dan statistik. Algoritma ini dapat memprediksi peluang atau probabilitas di masa depan berdasarkan pengalaman dari data di masa sebelumnya. Tingkat keakuratan yang dihasilkan oleh klasifikasi juga mencerminkan sejauh mana dataset pencapaian nilai mahasiswa yang diterapkan mampu memberikan efektivitas dalam metode Naive Bayes Classification.
2. Dalam mengimplementasikan Naive Bayes menggunakan aplikasi WEKA, dapat dilakukan penelusuran terhadap karakteristik atribut dari dataset yang mencakup kelas Lulus dan Tidak Lulus. Metode klasifikasi Naive Bayes memanfaatkan data training untuk menghasilkan probabilitas dari setiap atribut untuk berbagai kelas yang berbeda. Dengan demikian, nilai probabilitas dari atribut-atribut ini dapat ditingkatkan untuk

melakukan prediksi terhadap status data nilai mahasiswa melalui proses klasifikasi menggunakan metode Naive Bayes Classification. Setelah melakukan pengujian terhadap data nilai mahasiswa menggunakan metode Naive Bayes, ditemukan bahwa metode ini berhasil mengklasifikasikan sebanyak 3.929 data yang diuji. Dari total 3.928 data mahasiswa, sebanyak 99,9745% di antaranya diklasifikasikan dengan benar sebagai Correctly Classified Instance. Namun, terdapat 1 data penggunaan klasifikasi pencapaian nilai mahasiswa yang tidak berhasil diklasifikasikan dengan benar, yang merupakan Incorrectly Classified Instance. Dengan hasil ini, dapat disimpulkan bahwa metode Naive Bayes memiliki tingkat keberhasilan yang tinggi dalam mengklasifikasikan data nilai mahasiswa, namun masih terdapat sedikit kesalahan dalam proses klasifikasi.

DAFTAR PUSTAKA

- Abdullah, N., Yusof, N., Lau, W. J., Jaafar, J., & Ismail, A. F. (2019). Recent trends of heavy metal removal from water/wastewater by membrane technologies. *Journal of Industrial and Engineering Chemistry*, 76, 17-38.
- Andita, V., Hermawat, W., & Hartati, N. S. (2019). Pengaruh Jumlah Pelayanan Rawat Jalan, Rawat Igd Dan Rawat Inap Terhadap Tingkat Pendapatan Rumah Sakit Umum Daerah (Rsud) Cideres Kabupaten Majalengka. *Jurnal Ekonomi Manajemen*, 14(2), 370-378.
- Dimiyati, T. R. (2018). Pembentukan karakter mahasiswa dalam sistem pendidikan tinggi Islam. *TADRIS: Jurnal Pendidikan Islam*, 13(1), 17-32.
- Ekayani, P. (2017). Pentingnya penggunaan media pembelajaran untuk meningkatkan prestasi belajar siswa. *Jurnal Fakultas Ilmu Pendidikan Universitas Pendidikan Ganesha Singaraja*, 2(1), 1-11.
- Kurdhi, N. A., Saksono, H., Rihfenti Ernayani, S. E., Ak, M., Bakri, A. A., SE, M., ... & MM, C. (2023). *Manajemen Sains*. Cendikia Mulia Mandiri.
- Mukminin, A., & Riana, D. (2017). Komparasi Algoritma C4. 5, Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah. *Jurnal Informatika*, 4(1).
- Nurdiaman, A. 2019. *Pendidikan Kewarganegaraan- Kecakapan berbangsa dan bernegara*. Jakarta: PT. Pribumi Mekar
- Pradiptasari, L. P. M. (2016). *Hubungan Antara Intensitas Penggunaan Media Sosial dan Interaksi Sosial dengan Prestasi Akademik di Kalangan Mahasiswa Pendidikan Ekonomi-FKIP Universitas Kristen Satya Wacana Salatiga*. Salatiga: Universitas Kristen Satya Wacana.
- Rabiah, S. (2019). Manajemen Pendidikan tinggi dalam meningkatkan mutu pendidikan. *Jurnal Sinar Manajemen*, 6(1), 58-67.
- Utomo, D. P., & Mesran, M. (2020). Analisis komparasi metode klasifikasi data mining dan reduksi atribut pada data set penyakit jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437-444.