

Klasifikasi Peminatan Topik Keilmuan Dalam Penyelesaian Studi Menggunakan Algoritma Naive Bayes

Waldi Setiawan*, Dedy Hartama, Muhammad Ridwan Lubis, Ihsan Syajidan, Agus Perdana Windarto

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: *waldisetiawan1999@gmail.com, dedyhartama@amiktunasbangsa.ac.id, muhammadridwanlubis76@gmail.com,

ihsan.syahidan03@gmail.com, agus.perdana@amiktunasbangsa.ac.id

Email Korespondensi: waldisetiawan1999@gmail.com

Abstrak–Keilmuan merupakan materi pembelajaran yang diajarkan di perguruan tinggi yang berguna untuk membantu mahasiswa dalam menyelesaikan penulisan skripsi sehingga studi keserjanaan dapat terselesaikan. Peminatan keilmuan yang dipilih sesuai visi-misi tiap prodi oleh mahasiswa dapat memberikan dampak baik terhadap perguruan tinggi. Keilmuan yang dipilih mahasiswa dalam menyelesaikan studi ada yang sesuai dan ada yang tidak sesuai visi-misi prodi. Variabel–variabel yang digunakan dalam penelitian ini adalah Ipk, MKRV1, MKRV2, dan Keilmuan. Tujuan dari penelitian ini adalah untuk mengetahui seberapa banyak mahasiswa yang memilih topik keilmuan sesuai visi–misi prodi terkhusus pada penelitian ini prodi Teknik Informatika dalam menyelesaikan studi. Algoritma Naive Bayes digunakan dalam penelitian ini. Akurasi yang diperoleh dari penelitian ini sebesar 98,11%. Pada penelitian ini dapat dilihat bahwa penelitian dapat memberikan kontribusi pada STIKOM Tunas Bangsa Pematang Siantar untuk mengetahui keilmuan yang dipilih mahasiswa prodi lain seberapa banyak memiliki kesesuaian terhadap visi–misi tiap prodi.

Kata Kunci: Data Mining; Klasifikasi; Naive Bayes; Prediksi; RapidMiner

Abstract–Academic expertise is a subject of study taught at the university level to assist students in completing their thesis writing, thereby enabling them to successfully complete their graduate studies. The chosen academic specialization aligns with the vision and mission of each program and can have a positive impact on the university. Students' chosen fields of expertise in completing their studies may either align or not align with the program's vision and mission. The variables used in this research are GPA, MKRV1, MKRV2, and Academic Expertise. The aim of this research is to determine how many students select an academic topic that aligns with the program's vision and mission, particularly in this case, the Computer Science program, as they complete their studies. The Naive Bayes algorithm is employed in this research, yielding an accuracy rate of 98.11%. This research can provide valuable insights for STIKOM Tunas Bangsa Pematang Siantar to understand the extent to which students from other programs choose academic expertise that aligns with the vision and mission of each program.

Keywords: Data Mining; Classification; Naive Bayes; Prediction; RapidMiner

1. PENDAHULUAN

Data Mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari database yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan database. Tujuan dari data mining adalah untuk mengekstrak informasi dengan metode cerdas dari kumpulan data kemudian mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut. Sistem terdiri dari tiga tahapan, yaitu input, proses, dan output[1]–[3]. Pada umumnya, suatu hal diawali dengan input data kemudian dilakukan proses hingga dihasilkan suatu output atau hasil. Hal yang menjadi pembeda pada data mining adalah input berupa set data, prosesnya adalah algoritma atau metode dalam data mining, serta output-nya berbentuk pola, cluster, decision tree, dan lain –lain [4]. Data mining atau kadang disebut *Knowledge Discovery in Database* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar.

Klasifikasi adalah salah satu model yang paling sering digunakan sebagai model pembelajaran di dalam data mining. Tujuannya membangun sebuah model untuk memprediksi perilaku konsumen di masa mendatang melalui klasifikasi record di dalam database ke dalam beberapa kelompok kelas berdasarkan kriteria tertentu. Klasifikasi merupakan proses pengumpulan data bersama – sama yang didasarkan atas sekumpulan kesamaan yang awalnya telah ditentukan oleh seorang analis sebelum analisis dimulai.

Naive Bayes Classifier merupakan metode klasifikasi yang formulasinya didasarkan pada Teorema Bayes. Keunggulan dari Naive Bayes adalah tingkat keakuratan dan kecepatannya yang tinggi dalam mengolah data yang berjumlah atau berukuran besar. Dari segi performa Naive Bayes lebih baik dari algoritma Decision Tree dan algoritma Selected Neural Network Classifier. Selain itu Naive Bayes dalam menentukan estimasi dari parameter yang dibutuhkan pada proses klasifikasi tidak memerlukan training data dalam jumlah yang besar. Dengan kata lain Naive Bayes hanya memerlukan jumlah training data yang kecil namun mampu menghasilkan klasifikasi yang akurat dan cepat [5]–[8].

Rapid Miner adalah suatu perangkat lunak untuk sains data yang dibentuk oleh rapidminer, inc yang menyajikan terintegrasi data, text mining, machine learning, dan analisis prediktif. Serta membantu langkah dalam suatu proses pembelajaran mesin seperti hasil visualisasi, optimasi, persiapan data dan validasi model. Secara sederhana rapidminer adalah sebuah aplikasi yang digunakan untuk mengolah sebuah data dengan berbagai teknik dan metode dalam data mining, sehingga data dapat menjadi informasi yang berguna. Rapidminer merupakan salah satu perangkat lunak yang bersifat opensource yang berguna untuk pengolahan suatu data mining [9].

Penelitian ini menggunakan data yang diperoleh dari database STIKOM Tunas Bangsa berdasarkan mahasiswa lulusan dari prodi Teknik Informatika (TI). Dari penelitian yang dilakukan sebelumnya algoritma Naive Bayes dapat

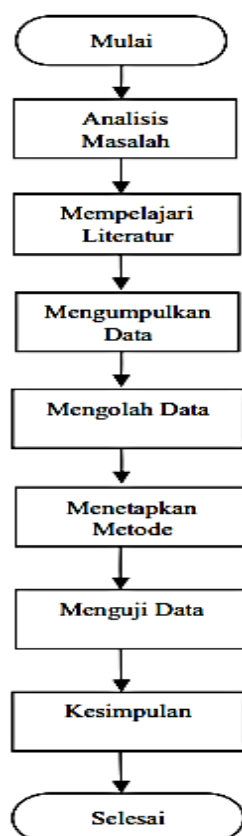
digunakan untuk mengklasifikasi. Hal ini dapat dibuktikan dengan beberapa penelitian dibawah ini yang dilakukan menggunakan metode Naïve Bayes dalam mengklasifikasi kualitas sperma[4]. Atribut yang akan digunakan dalam melakukan klasifikasi penduduk adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin/Belum Kiveawin). Kemudian penelitian yang dilakukan dalam mengklasifikasi lokasi pembangunan sumber air. Penggunaan teknik datamining khususnya klasifikasi menggunakan algoritma Naive Bayes dapat dilakukan prediksi terhadap kelayakan lokasi pembangunan sumber air bersih berdasarkan data yang ada. Implementasi data mining dengan algoritma naïve bayes untuk klasifikasi kelayakan penerima bantuan [10]–[13]. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang berkaitan dengan pengguna algoritma yang digunakan.

Dari hasil penelitian yang dibuktikan diatas bahwa algoritma Naïve Bayes dapat melakukan sebuah klasifikasi [14]–[17]. Hal tersebut menguatkan keyakinan penulis untuk menggunakan algoritma Naïve Bayes dalam mengklasifikasi peminatan topik keilmuan pada mahasiswa semester akhir di STIKOM Tunas Bangsa. Perbedaan penelitian ini dengan penelitian sebelumnya ialah penelitian ini mengklasifikasi mahasiswa yang memilih keilmuan yang digunakan untuk menyelesaikan studi sedangkan penelitian sebelumnya mengklasifikasi masyarakat miskin dan mengklasifikasi lokasi pembangunan sumber air.

2. METODOLOGI PENELITIAN

2.1 Rancangan Penelitian

Flowchart adalah sebuah diagram yang menjelaskan alur proses dari sebuah program. Dalam membangun sebuah program, flowchart berperan penting untuk menerjemahkan proses berjalannya sebuah program agar lebih mudah untuk dipahami. Rancangan penelitian digambarkan pada gambar 1 flowchart berikut ini :



Gambar 1. Rancangan Penelitian

Dari gambar 1 diatas, merupakan rancangan penelitian dari judul Klasifikasi peminatan topik keilmuan mahasiswa semester menggunakan algoritma Naïve Bayes dijelaskan bahwa :

1. Analisis Masalah
Pada tahap ini mencari pola minat mahasiswa semester akhir berdasarkan visi – misi untuk mengklasifikasi peminatan topik keilmuan mahasiswa semester akhir di STIKOM Tunas Bangsa.
2. Mempelajari Literatur
Tahap ini mengumpulkan informasi yang relevan dengan topik masalah yang terjadi pada objek penelitian, melengkapi pengetahuan teori – teori yang digunakan dalam penelitian baik dari buku maupun jurnal – jurnal terkait.
3. Mengumpulkan Data

Penulis mengumpulkan data yang dibutuhkan dari kantor pendidikan STIKOM Tunas Bangsa Pematang Siantar dalam bentuk softcopy Microsoft Excel 2007 yang berjudul “lulusan STIKOM”, “Matakuliah TI”, “Nilai TI”.

4. Mengolah Data

Pada penelitian ini pengolahan data dilakukan menggunakan algoritma Naïve Bayes dan Software Microsoft Excel 2007.

5. Menetapkan Metode

Penggunaan metode pada penelitian ini menggunakan algoritma Naïve Bayes untuk penyelesaian masalah.

6. Menguji Data

Pengujian data dilakukan dengan menggunakan software RapidMiner 9.9

7. Kesimpulan

Kesimpulan yang diharapkan dari penelitian yang berjudul klasifikasi peminatan topik keilmuan mahasiswa semester akhir bahwa hasil akurasi prediksi peminatan topik keilmuan mahasiswa semester menjadi acuan untuk pihak tenaga pengajar di STIKOM Tunas Bangsa dalam memfokuskan materi pembelajaran sesuai visi – misi prodi TI.

2.2 Analisis Data

Dalam melaksanakan penelitian, bagian terpenting yang diperlukan adalah data penelitian dimana data akan diolah sehingga menghasilkan informasi yang bermanfaat. Berikut ini merupakan data mentah yang diperoleh dari kantor pendidikan STIKOM Tunas Bangsa Pematang Siantar seperti pada tabel 1 berikut :

Tabel 1. Data Sampel Lulusan STIKOM Tunas Bangsa Prodi TI

No	NIM	Nama	Judul Skripsi	Huruf Mutu	Keterangan		Matakuliah KRIPTOGRAFI	Sistem Digital
1.	1501002	AGUNG BIMANTORO	Estimasi Pemberantasan Hama di Kebun Bah Jambi Dengan Menggunakan Algoritma Backpropagation	A	IPK angkaMutu		3.25	80.00
				B	IPK angkaMutu		79.00	
				C	IPK angkaMutu			
2.	1501003	ALAN BOY SANDY DAMANIK	Implementasi Algoritma Data Encryption Standard (DES) dalam Pengamanan Data Karyawan Ramayana Department Store	A	IPK angkaMutu		3.53 87.00	3.53 85.00
				B	IPK angkaMutu			
....								
8.	1501010	DANIEL SAPUTRA SITORUS	Implementasi Konfigurasi Quota Bandwidth untuk Akses Internet Menggunakan Mikrotik Router di Kantor Walikota Pematang Siantar	A	IPK angkaMutu			3.36 80.00
				B	IPK angkaMutu		3.36 79.00	
				C	IPK angkaMutu			
9.	1501013	EKO HARTATO	Analisis Algoritma AES dalam Mengamankan Data pada Kantor Walikota Pematang Siantar	A	IPK angkaMutu			3.70 82.00
				B	IPK angkaMutu		3.70 77.00	
10.	1501014	ENJEL VIENZA SINAGA	Sistem Pendukung Keputusan dalam Menyeleksi karyawan baru pada hotel	A	IPK angkaMutu			2.88 80.00
				B	IPK angkaMutu		2.88 78.00	
				C	IPK angkaMutu			

2.2.1 Preprocessing Data

Dari tabel 1, dilakukan preprocessing data untuk membersihkan data yang tidak diperlukan sehingga tersisa data yang siap diolah kedalam software rapid miner. Tabel 2 berikut merupakan data awal penelitian yang merupakan data lulusan mahasiswa STIKOM Tunas Bangsa prodi Teknik Informatika sebelum dilakukan preprocessing data.

Tabel 2. Data Rata-Rata Matakuliah Visi-Misi

No	Nim	Nama	IPK	MKV1. 1	...	MKV2. 3	MKRV2	Judul Skripsi	Keilmuan	Keterangan
1.	1501002	AGUNG BIMANTORO	3,25	75	...	79	79,666666 67	Estimasi Pemberantasan Hama di Kebun Bah Jambi Dengan Menggunakan Algoritma Backpropagati on	JST	Sesuai
2.	1501003	ALAN BOY SANDY DAMANIK	3,53	78	...	87	83,333333 33	Implementasi Algoritma Data Encryption Standard (DES) dalam Pengamanan Data Karyawan Ramayana Department Store	Kriptografi	Tidak Sesuai
3.	1501005	ALVIN FUADI	3,47	78	...	75	79,666666 67	Rancang Bangun Alat Pemberi Pakan Hewan Ternak Ikan Otomatis Berdasarkan Mikrokontroler Arduino, Sensor Ultrasonik dan Module SIM800L	Robotika	Sesuai

Pada tabel 2 merupakan dataset mencari nilai rata-rata matakuliah visi – misi yang terbagi atas dua bagian, dimana terdapat nilai matakuliah rata-rata visi – misi 1 dengan tiga matakuliah yaitu : Data Mining, Jaringan Syaraf Tiruan, dan Kecerdasan Buatan. Matakuliah visi – misi 2 yaitu : Keamanan Jaringan Komputer, Sistem Digital, dan Kriptografi. Pada tabel 2 diatas juga dilakukan pencarian kesesuaian keilmuan yang digunakan pada judul skripsi apakah sesuai dengan visi – misi prodi TI.

Tabel 3. Data Transformasi

No	Nim	Nama	IPK	MKRV1	MKRV2	Judul Skripsi	Keilmuan	Keterangan
1.	1501002	AGUNG BIMANTORO	SM	B	B	Estimasi Pemberantasan Hama di Kebun Bah Jambi Dengan Menggunakan Algoritma Backpropagation	JST	Sesuai
2.	1501003	ALAN BOY SANDY DAMANIK	DP	A	A	Implementasi Algoritma Data Encryption Standard (DES) dalam Pengamanan Data Karyawan Ramayana Department Store	Kriptografi	Sesuai
3.	1501005	ALVIN FUADI	SM	A	B	Rancang Bangun Alat Pemberi Pakan Hewan Ternak Ikan Otomatis Berbasis Mikrokontroler Arduino, Sensor Ultrasonik dan Module SIM800L	Robotika	Sesuai
4.	1501006	ANDRIANO	SM	B	B	Sistem Pendukung Keputusan Penentuan Karyawan Tetap Menggunakan Metode Simple Additive Weighting (SAW) Pada PT. Bridgestone Sumatera Rubber Estate	Spk	Tidak Sesuai
5.	1501007	BIL KLINTON SIHOTANG	DP	B	A	Implementasi Access Control List pada Mikrotik dalam Mengamankan Koneksi Internet Koperasi Sumber Dana Mutiara	Jaringan Komputer	Sesuai

Pada tabel 3 merupakan hasil transformasi data dari data rata-rata matakuliah visi-misi awal yang bisa dilihat pada tabel 2.

Tabel 4. Data Riset

No	Nim	Nama	IPK	MKRV1	MKRV2	Keilmuan	Keterangan
1.	1501002	AGUNG BIMANTORO	SM	B	B	JST	Sesuai
2.	1501003	ALAN BOY SANDY DAMANIK	DP	A	A	Kriptografi	Sesuai
3.	1501005	ALVIN FUADI	SM	A	B	Robotika	Sesuai
4.	1501006	ANDRIANO	SM	B	B	Spk	Tidak Sesuai

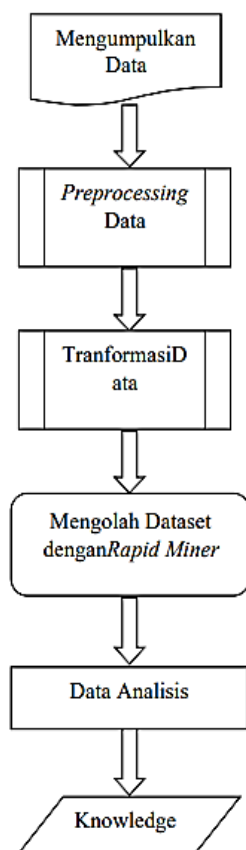
No	Nim	Nama	IPK	MKRV1	MKRV2	Keilmuan	Keterangan
5.	1501007	BIL KLINTON SIHOTANG	DP	B	A	Jaringan Komputer	Sesuai

Pada tabel 4 merupakan dataset yang siap diolah kedalam rapidminer. Dimana terdapat 4 variabel yaitu Ipk, MKRV1, MKRV2, dan keilmuan. Dan terdapat 1 label yaitu: Keterangan. Pada gambar transformasi terdapat beberapa penjelasan sebagai berikut :

- IPK Nilai IPK dari mahasiswa yang tadinya berupa angka ditransformasi menjadi huruf DP, SM, CM, C, dan K. Dimana DP (Dengan Pujian) nilai IPK yang didapat dari 3,5 – 4,0 , SM (Sangat Memuaskan) nilai IPK yang didapat dari 3,0 – 3,49, CM (Cukup Memuaskan) nilai IPK yang didapat dari 2,7 – 2,99 , C (Cukup) nilai IPK yang didapat dari 2 – 2,69 , K (Kurang) nilai IPK yang didapat dari 1 – 1,99.
- MKRV1 Mata Kuliah Rata – rata Visi – misi 1 (MKRV1) merupakan nilai rata – rata dari hasil penggabungan 3 mata kuliah yaitu : Data Mining, Jaringan Syaraf Tiruan, dan Kecerdasan Buatan. Dimana Nilai untuk huruf A = 80 – 90, B = 70 – 79, C = 60 – 69, D = 50 – 59, E = 40 – 49.
- MKRV2 Mata Kuliah Rata – rata Visi – misi 2 (MKRV2) merupakan nilai rata – rata dari hasil penggabungan 3 mata kuliah yaitu : Keamanan Jaringan Komputer, Sistem Digital, dan Kriptografi. Dimana Nilai untuk huruf A = 80 – 90, B = 70 – 79, C = 60 – 69, D = 50 – 59, E = 40 – 49.

2.2.2 Instrumen Penelitian

Data yang digunakan dalam penelitian ini menggunakan file excel 2007 (xls). Transformasi data diperlukan untuk menjadi input software yang digunakan. Software RapidMiner 9.9 digunakan dalam melakukan tranformasi data. Instrumen penelitian dapat dilihat seperti pada gambar 2 berikut :

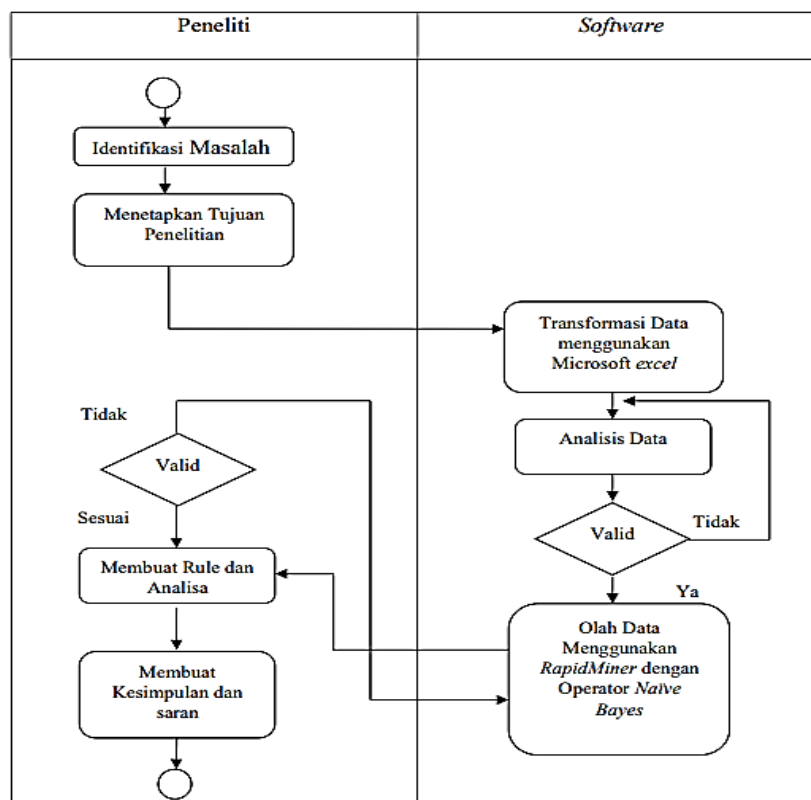


Gambar 2. Instrumen Penelitian

Pada gambar 2 instrumen penelitian diatas menjelaskan bahwa data yang digunakan dalam penelitian bersumber dari kantor pendidikan STIKOM Tunas Bangsa Pematang Siantar lulusan tahun 2014 sampai 2021, nilai matakuliah prodi TI, Matakuliah TI. Dataset yang telah terkumpul di cleaning untuk menyisakan data yang diperlukan kemudian data ditranformasi sesuai dan selanjutnya dimasukkan ke dalam RapidMiner 9.9 untuk diolah. Lalu dilakukan analisis dan akan menjadi knowledge atau pengetahuan.

2.2.3 Diagram Aktifitas Kerja Penelitian

Diagram aktifitas kerja penelitian yang dilakukan dapat dilihat seperti pada gambar 3 berikut :

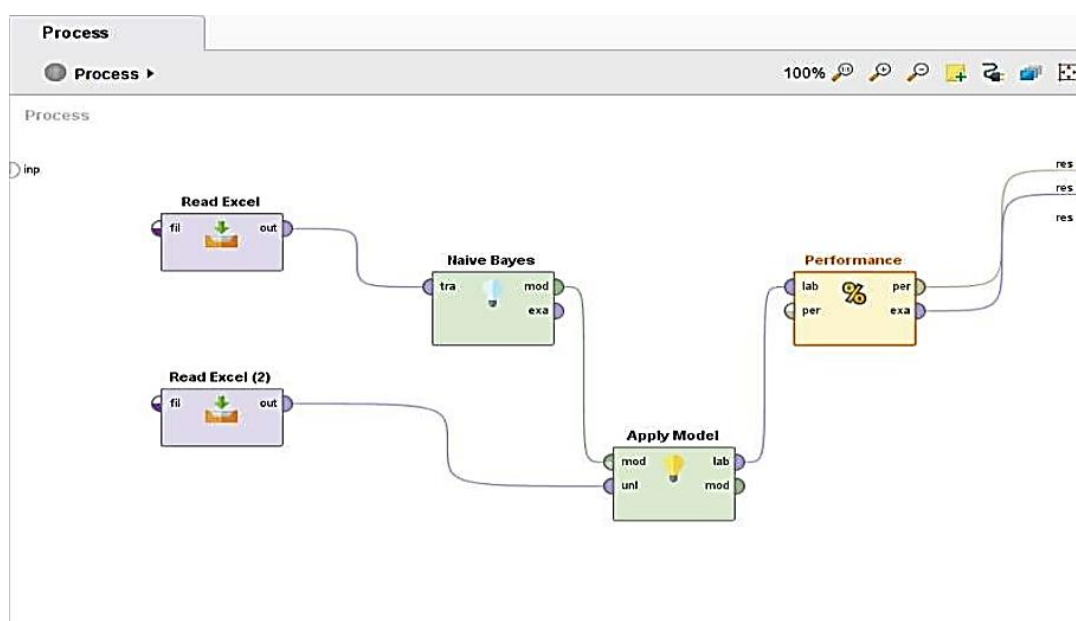


Gambar 3. Diagram Aktivitas Kerja Penelitian

Pada gambar 3 dijelaskan bahwa dilakukan proses mengidentifikasi masalah dan tujuan penelitian yang akan dilakukan, melakukan preprocessing data mentah agar menjadi dataset yang sesuai dengan penelitian. Data dianalisis terlebih dahulu untuk memastikan datanya sudah sesuai, Dataset yang sudah sesuai kemudian diolah kedalam RapidMiner 9.9 dengan operator Naïve Bayes. Apabila hasil pengolahan dari RapidMiner menghasilkan akurasi yang tinggi kemudian dilanjutkan membuat rule dan menganalisis data secara manual. Setelah proses secara manual selesai kemudian membuat kesimpulan dan saran.

2.2.4 Pemodelan Metode

Pemodelan metode Naïve Bayes pada RapidMiner dapat dilihat seperti pada gambar 4 berikut:



Gambar 4. Pemodelan Algoritma Naïve Bayes Pada RapidMiner

Pada gambar 4 menampilkan pemodelan dengan menggunakan operator seperti Apply Model, dan Performance dan terdapat Read Excel yang terbagi menjadi data training dan data testing. Data training digunakan oleh algoritma Naïve

Bayes untuk sebagai klasifikasi data sehingga menghasilkan kesimpulan. Pada data testing menggunakan operator Apply Model berfungsi untuk hasil dari algoritma Naïve Bayes dan menggunakan operator Performance yang berfungsi untuk mendapat tingkat akurasi pada data yang digunakan.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Data didapatkan dari kantor pendidikan STIKOM Tunas Bangsa Pematang Siantar dalam bentuk softcopy Microsoft Excel 2007 yang berjudul “lulusan STIKOM”, “Matakuliah TI”, “Nilai TI” mahasiswa angkatan 2015 dan 2016 yang terdata untuk mengklasifikasi peminatan topik keilmuan. Algoritma Naïve Bayes digunakan sebagai pemodelan data yang diperlihatkan oleh aplikasi Rapidminer 9.9. Data yang digunakan dalam penelitiann ini sebanyak 153 data. Data terbagi menjadi data training yang berjumlah 100 data dan data testing dengan jumlah 53 data. Data yang diperoleh diolah menggunakan software rapidminer 9.9 dan menggunakan operator performance untuk mendapatkan tingkat keakuratan data. Dibawah ini merupakan perhitungan manual untuk mengklasifikasi peminatan topik keilmuan dalam menyelesaikan studi di STIKOM Tunas Bangsa menggunakan algoritma Naïve Bayes :

a. Perhitungan pada jumlah kelas/class

Pada Tabel 5 berikut ditampilkan data training(latih) yang digunakan sebanyak 100 data.

Tabel 5. Tabel Data Training(latih)

No	Nama	IPK	MKRV1	MKRV 2	Keilmuan	Keterangan
1	Agung Bimantoro	SM	B	B	JST	Sesuai
2	Alan Boy Sandy Damanik	DP	A	A	Kriptografi	Sesuai
3	Alvin Fuadi	SM	A	B	Robotika	Sesuai
4	Andriano	SM	B	B	Spk	Tidak Sesuai
5	Bil Klinton Sihotang	DP	B	A	Jaringan Komputer	Sesuai
6	Dana Alpian Rianto	SM	B	B	Data Mining	Sesuai
7	Daniel Marpaung	SM	B	A	JST	Sesuai
8	Daniel Saputra Sitorus	SM	B	A	Jaringan Komputer	Sesuai
9	Eko Hartato	DP	A	A	Kriptografi	Sesuai
10	Enjel Vienza Sinaga	C	B	A	Spk	Tidak Sesuai
		M				
11	Fikri Yatussa`Ada	SM	B	B	Spk	Tidak Sesuai
12	Handri Al Fani	SM	B	A	Robotika	Sesuai
13	Imelda Asih Rohani	DP	A	A	Kriptografi	Sesuai
	Simbolon					
14	Juli Wahyuni	SM	B	B	Kriptografi	Sesuai
15	Mhd. Ade Dharmawan	SM	B	B	Robotika	Sesuai
	Nasution					
...						
87	Roma Marbun	DP	A	B	Lavenberg Marquardt	Tidak Sesuai
88	Sapta Darma Simarmata	SM	B	B	Robotika	Sesuai
89	Syafri Maradu Manurung	SM	B	A	Robotika	Sesuai
90	Syahri Ramadhan	SM	B	B	Robotika	Sesuai
91	Venny Vidya Utari	DP	B	A	JST	Sesuai
92	Vio Sri Zuliyanti	DP	A	A	LVQ	Tidak Sesuai
93	Widya Tri Charisma Gultom	DP	A	B	Lavenberg Marquardt	Tidak Sesuai
94	Yusuf Agung Pratama	SM	B	A	Robotika	Sesuai
95	Zepri Pratama	DP	C	A	Dijkstra	Tidak Sesuai
96	Afriani Oktavia	SM	B	A	Robotika	Sesuai
97	Alberth David Hetharua	SM	A	B	Robotika	Sesuai
98	Andi Sanggam Sidabutar	SM	B	A	Kecerdasan Buatan	Sesuai
99	Anggi Martua Valentino	DP	A	B	Robotika	Sesuai
	Sianipar					
100	Bellarmino Wira Permana	DP	A	B	Robotika	Sesuai

Kelas sesuai sebanyak 83 data dan kelas tidak sesuai sebanyak 17 data. Sehingga menghitung nilai probabilitas kelas sesuai dalam mengklasifikasi peminatan topik keilmuan dalam menyelesaikan studi dapat dilihat pada persamaan (1) :

$$P(\text{sesuai}) = \frac{83}{100} = 0,83 \quad (1)$$

Untuk menghitung nilai probabilitas kelas tidak sesuai dalam mengklasifikasi peminatan topik keilmuan dalam menyelesaikan studi dapat dilihat pada persamaan (2) :

$$P(\text{Tidak sesuai}) = \frac{17}{100} = 0,17 \quad (2)$$

b. Perhitungan jumlah kasus per Kelas

Menghitung nilai probabilitas untuk masing – masing variabel. Perhitungan nilai probabilitas pada variabel ipk didapat dengan rumus perhitungan jumlah kriteria ipk yang sesuai dibagi jumlah data sesuai dan jumlah kriteria ipk tidak sesuai dibagi dengan jumlah data yang tidak sesuai untuk lebih jelasnya sebagai berikut : $\text{IpK DP(sesuai)} / \text{Jumlah(sesuai)}$, $\text{IpK DP(Tidak Sesuai)} / \text{Jumlah(Tidak Sesuai)}$.

c. Perkalian semua hasil variabel sesuai dan tidak sesuai

Jika nilai probabilitas setiap kriteria diketahui, selanjutnya melakukan perkalian semua hasil variabel sesuai dan tidak sesuai berdasarkan data testing ke 101 sampai 153. Berdasarkan data testing pada data ke 101 sampai 153 dilakukan klasifikasi kedalam kelas “sesuai” dapat dilihat pada persamaan (3) :

$$1 P X \text{ Kelas } P V \text{ Kelas } (|) (|) = x 2 P V \text{ Kelas } x x (|) \dots\dots (|) P V \text{ Kelas } n \dots \dots \text{Pers} \quad (3)$$

Sehingga untuk menghitung nilai sesuai pada data ke 101 sampai 153 sebagai berikut :

$$\begin{aligned} P(\text{Sesuai} | 101) &= P(\text{IpK} = \text{SM} | \text{Sesuai}) \times P(\text{MKRV1}) \\ &= A | \text{Sesuai}) \times P(\text{MKRV2}) \\ &= B | \text{Sesuai}) \times P(\text{Keilmuan}) \\ &= \text{Data Mining} | \text{Sesuai}) \\ &= 0,5783 \times 0,1686 \times 0,5301 \times 0,0722 \\ &= 0,00373169177 \end{aligned}$$

Sedangkan, perhitungan untuk nilai Tidak Sesuai pada data ke 101 sampai 153 dapat menerapkan rumus persamaan (4) sehingga menghasilkan nilai sebagai berikut :

$$\begin{aligned} P(101 | \text{Tidak Sesuai}) &= P(\text{IpK} = \text{SM} | \text{Tidak Sesuai}) \times P(\text{MKRV1}) \\ &= A | \text{Tidak Sesuai}) \times P(\text{MKRV2}) \\ &= B | \text{Tidak Sesuai}) \times P(\text{Keilmuan}) \\ &= \text{Data Mining} | \text{Tidak Sesuai}) \\ &= 0,5882 \times 0,4117 \times 0,5294 \times 0 \\ &= 0 \end{aligned}$$

Setelah nilai sesuai dan tidak sesuai didapatkan, langkah berikutnya menghitung pemaksimalan sesuai dan tidak sesuai. Perhitungan pemaksimalan nilai sesuai berdasarkan data testing ke 101 sampai ke 153 adalah sebagai berikut :

$$\begin{aligned} P(\text{Sesuai} | C) &= P(Dn | C) \times P(\text{Sesuai}) \\ &= (101 | C) \times P(0,83) \\ &= 0,00373169177 \times 0,83 \\ &= 0,00309730417 \end{aligned}$$

Sedangkan untuk perhitungan pemaksimalan tidak sesuai dapat dilihat pada perhitungan dibawah ini :

$$\begin{aligned} P(\text{Tidak Sesuai} | C) &= P(Dn | C) \times P(\text{Tidak Sesuai}) \\ &= (101 | C) \times P(0,17) \\ &= 0 \times 0,17 \\ &= 0 \end{aligned}$$

d. Membandingkan Hasil Kelas Sesuai dan Tidak Sesuai

Setelah nilai pemaksimalan sesuai dan tidak sesuai didapat, selanjutnya melakukan perbandingan nilai sesuai dan tidak sesuai, sehingga akan diketahui klasifikasi tersebut tergolong ke dalam sesuai atau tidak sesuai.

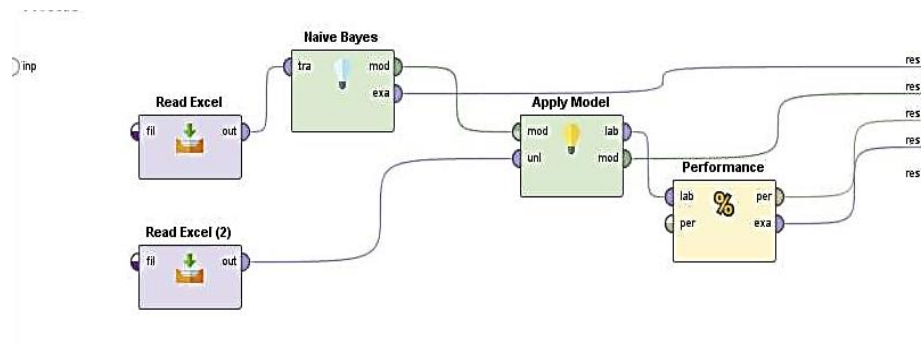
$$D101 = \text{Sesuai} \geq \text{Tidak Sesuai}$$

$$= 0,00309730417 \geq 0$$

$$= 0,00309730417 (\text{Sesuai})$$

Dari keseluruhan perbandingan diatas dapat diketahui bahwa probabilitas yang tertinggi ialah pada kelas sesuai. Sehingga dapat diketahui bahwa keilmuan yang digunakan dalam menyelesaikan studi ternasuk dalam klasifikasi sesuai.

3.2 Hasil Percobaan



Gambar 5. Tampilan operator pada Process View

Pada gambar 5 diatas tampilan operator yang digunakan dalam menghasilkan model Niave Bayes. Operator – operator yang digunakan adalah 2 buah operator read excel (data training dan data testing), Naïve Bayes, Apply Model,dan performance. Data training dihubungkan dengan Naïve Bayes untuk menampilkan model klasifikasi, mod.Naive Bayes dihubungkan ke mod.Apply Model,exa Naïve Bayes dihubungkan ke result process guna manampilkan data table (data training) pada result workspace. data testing dihubungkan ke port unlabelled data Apply Model guna memastikan kekonsistenan data testing. Port labelled Apply Model dihubungkan ke port labelled operator performance untuk memperoleh akurasi data yang diolah. Port model Apply Model dihubungkan ke result process guna mendapatkan model dari data training dan memperoleh prediksi hasil klasifikasi dari unlabelled data (data testing) Selanjutnya port performance dan example set dihubungkan ke port result process. Setelah semua port dihubungkan ke result process, maka klik ikon play sehingga muncul tampilan seperti pada gambar 6 berikut :

SimpleDistribution

Distribution model for label attribute Keterangan

Class Sesuai (0.830)
4 distributions

Class Tidak Sesuai (0.170)
4 distributions

Gambar 6. Tampilan Simple Distribution

Gambar 6 menjelaskan bahwa kelas sesuai memiliki nilai probabilitas 0,830 dan kelas tidak sesuai memiliki nilai probabilitas 0,170, adapun Distribution Table dapat dilihat seperti pada gambar 7 berikut :

Description	Attribute	Parameter	Sesuai	Tidak Sesuai
Simple Charts	lpk	value=SM	0.578	0.587
Distribution Table	lpk	value=DP	0.325	0.294
Annotations	lpk	value=CM	0.084	0.059
	lpk	value=C	0.012	0.059
	lpk	value=unknown	0.000	0.001
	MKRV1	value=B	0.735	0.470
	MKRV1	value=A	0.169	0.411
	MKRV1	value=C	0.096	0.118
	MKRV1	value=unknown	0.000	0.001
	MKRV2	value=B	0.530	0.529
	MKRV2	value=A	0.470	0.470
	MKRV2	value=unknown	0.000	0.001
	Keilmuan	value=JST	0.361	0.001
	Keilmuan	value=Kriptografi	0.060	0.001

Gambar 7. Tampilan Distribution Table

Pada gambar 7 telah didapat nilai probabilitas dari masing – masing kriteria dimana nilai probabilitas sama seperti Tabel 2, 3, 4, dan Tabel 5. Adapun class prediction dapat dilihat seperti pada gambar 8 berikut:

Data	Row No.	Keterangan	prediction(K...	confidence[...	confidence[...	lpk	MKRV1	MKRV2	Keilmuan
Statistics	1	Sesuai	Sesuai	0.996	0.004	SM	A	B	Data Mining
Visualizations	2	Sesuai	Sesuai	0.999	0.001	SM	A	B	Robotika
Annotations	3	Sesuai	Sesuai	0.996	0.004	DP	A	A	Kriptografi
	4	Sesuai	Sesuai	1.000	0.000	CM	B	B	JST
	5	Sesuai	Sesuai	1.000	0.000	DP	B	A	JST
	6	Sesuai	Sesuai	0.999	0.001	DP	A	B	Robotika
	7	Sesuai	Sesuai	0.996	0.004	DP	A	B	Kriptografi
	8	Sesuai	Sesuai	1.000	0.000	SM	B	B	Robotika
	9	Sesuai	Sesuai	1.000	0.000	SM	B	B	Robotika
	10	Sesuai	Sesuai	1.000	0.000	SM	C	B	Robotika
	11	Sesuai	Tidak Sesuai	0.315	0.685	DP	A	B	Kriptografi
	12	Sesuai	Sesuai	0.999	0.001	DP	A	A	Robotika
	13	Sesuai	Sesuai	0.996	0.004	SM	A	B	Data Mining
	14	Sesuai	Sesuai	0.996	0.004	DP	A	B	Data Mining

Gambar 8. Tampilan Class Prediction Data Testing

Pada gambar 8 diperoleh class prediction yang sama dengan perhitungan manual, dari 53 data testing yang digunakan memperoleh class prediction yang tergolong sesuai. Hasil Accuracy pengujian model algoritma Naïve Bayes dapat dilihat seperti pada gambar 9 berikut :

accuracy: 98.11%			
	true Sesuai	true Tidak Sesuai	class precision
pred. Sesuai	52	0	100.00%
pred. Tidak Sesuai	1	0	0.00%
class recall	98.11%	0.00%	

Gambar 9. Tampilan Accuracy Performance

Pada gambar 9 terdapat hasil Accuracy Performance dengan jumlah prediksi Sesuai dan kenyataan true Sesuai adalah 52 record. Jumlah prediksi Sesuai dan kenyataannya true Tidak Sesuai adalah 0 record, jumlah prediksi Tidak Sesuai dan kenyataannya true Sesuai adalah 1 record, jumlah prediksi Tidak Sesuai dan kenyataannya true Tidak Sesuai adalah 0 record. Nilai accuracy sebesar 98,11%, class prediction pada kelas Sesuai sebesar 100%, class prediction pada kelas Tidak Sesuai 0,00%, class recall pada true Sesuai Sebesar 98,11% dan class recall pada true Tidak Sesuai sebesar 0,00%.

3.3 Pembahasan

Hasil yang diperoleh dari perhitungan manual menggunakan algoritma Naïve Bayes menghasilkan nilai probabilitas pada kelas Sesuai sebesar 0,830 dan nilai probabilitas kelas Tidak Sesuai sebesar 0,170. Nilai probabilitas dari masing – masing

kriteria menggunakan algoritma Naïve Bayes telah diperoleh seperti pada Tabel 5. Hasil perbandingan antara nilai kelas Sesuai dan Tidak Sesuai atau class prediction diperoleh bahwa dari 53 data testing, 52 data termasuk dalam kelas Sesuai dan 1 data Tidak Sesuai. Hasil yang didapatkan dari Rapidminer 9.9 diperoleh nilai probabilitas pada kelas Sesuai sebesar 0,830 dan nilai probabilitas Tidak Sesuai sebesar 0,170 seperti yang terlihat pada gambar 4.5. Nilai probabilitas pada setiap kriteria yang ditampilkan oleh Rapidminer 9.9 dapat dilihat pada gambar 4.6. Hasil class prediction pada data testing yang ditampilkan Rapidminer 9.9 diperoleh bahwa dari 53 data, 52 data termasuk dalam kelas Sesuai dan 1 data kedalam kelas Tidak Sesuai. Sehingga dapat dikatakan hasil yang diperoleh melalui perhitungan manual dan yang ditampilkan oleh Rapidminer tingkat kesamaan dan kesesuaiannya lebih tinggi.

4. KESIMPULAN

Berdasarkan pada hasil penelitian dan pembahasan yang sudah dilakukan, dapat dikatakan bahwa algoritma Naïve Bayes dapat digunakan untuk melakukan klasifikasi peminatan topik keilmuan dalam menyelesaikan studi di STIKOM Tunas Bangsa dengan faktor yang digunakan adalah Ipk, MKRV1, MKRV2, dan Keilmuan. Berdasarkan pada 100 data training yang diuji melalui Software Rapidminer didapat 83 data yang Sesuai dengan nilai probabilitas 0,830 dan 17 data yang Tidak Sesuai dengan nilai probabilitas 0,170 dan diperoleh hasil akurasi sebesar 98,11%, class prediction pada kelas Sesuai sebesar 100%, class prediction pada kelas Tidak Sesuai 0,00%, class recall pada true Sesuai Sebesar 98,11% dan class recall pada true Tidak Sesuai sebesar 0,00%.

REFERENCES

- [1] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*, 1st ed. Yogyakarta: Deepublish, 2020.
- [2] A. Widiastari, Solikhun, and Irawan, "Analisa Datamining dengan Metode Klasifikasi C4 . 5 Sebagai Faktor Penyebab Tanah Longsor," *J. Comput. Syst. Informatics*, vol. 2, no. 3, pp. 247–255, 2021.
- [3] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. 2015.
- [4] S. Hosana and D. S. D. Putra, "Penerapan Data Mining Untuk Prediksi Kualitas Sperma (Fertility) Berbasis Android," *Algor Univ. Buddhi Dharma*, vol. 1, no. 1, pp. 64–71, 2019.
- [5] Y. Lukito and A. R. Chrismanto, "Recurrent neural networks model for WiFi-based indoor positioning system," *Proceeding 2017 Int. Conf. Smart Cities, Autom. Intell. Comput. Syst. ICON-SONICS 2017*, vol. 2018–Janua, no. 118, pp. 121–125, 2018.
- [6] S. M. Dewi, A. P. Windarto, and D. Hartama, "Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud.Selamat Selular," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 617–621, 2019.
- [7] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020.
- [8] R. H. Sukarna and Y. Ansori, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022.
- [9] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [10] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 219–225, 2021.
- [11] J. Ipmawati, Kusriani, and E. Taufiq Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *Indones. J. Netw. Secur.*, 2017.
- [12] Triase and Samsudin, "Implementasi Data Mining Dalam Mengklasifikasikan Ukt (Uang Kuliah Tunggal) Pada Uin Sumatera Utara Medan," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 370–376, 2020.
- [13] Baskoro, Sriyanto, and L. Setya Rini, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," *Semin. Nas. Has. Penelit. dan Pengabd. Masy. Inst. Inform. dan Bisnis Darmajaya*, pp. 87–94, 2021.
- [14] Sunardi, A. Fadlil, and Suprianto, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Pada Angket Mahasiswa," *Saintekbu*, vol. 10, no. 2, pp. 1–9, 2018.
- [15] K. Ayu Milati Nur Azizah, "Sistem Pendukung Keputusan Penyeleksian Mahasiswa Penerima Beasiswa Menggunakan Metode Naïve Bayes Berbasis Web," *Antivirus J. Ilm. Tek. Inform.*, vol. 10, no. 1, pp. 6–10, 2016.
- [16] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021.
- [17] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1 : Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," *J. Media Inform. Budidarma*, vol. 4, no. 3, pp. 855–862, 2020.