**Penerapan Algoritma Naïve Bayes Dalam Klasifikasi Penjurusan**

**di MAN 2 Lubuklinggau”.**

**Vitria Alanda, Davit Irawan, M.Kom2, Antoni Zulius, M.Kom3**

1Program Studi Sistem Keterangan, Universitas Bina Insan, Lubuklinggau

e-mail: **\*1** 18030019@gmail.com, [2 davit irawan@univbinainsan.ac.id](mailto:2%20davit_irawan@univbinainsan.ac.id), [3 antoni.zulius@univbinainsan.ac.id](mailto:3%20antoni.zulius@univbinainsan.ac.id)

**ABSTRAK**

Madrasah Aliyah Negeri (MAN) 2 Lubuklinggau merupakan salah satu madrasah yang terlibat dalam proses belajar mengajar. MAN 2 Lubuklinggau yang terletak di Jl. Yos Sudarso No. 50, Taba Jemekeh, Kecamatan Lubuklinggau Timur 1 Kota Lubuklinggau. Jurusan/minat yang tersedia di MAN 2 Lubuklinggau antara lain IPA, IPS, dan agama. Kesulitan penelitian ini dalam menentukan jurusan mahasiswa menjadi kendala bagi institusi dalam menghitung kriteria masing-masing mahasiswa dan mengurangi efisiensi waktu. Penetapan jurusan mahasiswa masih dilakukan secara manual dengan menginput data mahasiswa ke dalam Microsoft Excel sehingga tidak efektif. Mengamati dan merekam langsung di lokasi penelitian (observasi), mengajukan pertanyaan langsung dan menerima jawaban langsung dari sumber (wawancara), dan mendokumentasikan dengan membaca standar literatur digunakan untuk mengumpulkan data untuk penelitian ini. Sistem kategorisasi untuk jurusan mahasiswa yang menggunakan pendekatan Nave Bayes menggunakan R Software, yang dirancang untuk membantu penjurusan mahasiswa menganalisis hasil tes lebih tepat.

Kata Kunci : Penjurusan Siswa, Software R, Sistem Klasifikasi

***ABSTRACT***

*Madrasah Aliyah Negeri (MAN) 2 Lubuklinggau is one of the madrasas involved in the teaching and learning process. The location of MAN 2 Lubuklinggau is Jl. Yos Sudarso No. 50, Taba Jemekeh, East Lubuklinggau District 1 Lubuklinggau City. The majors/interests available at MAN 2 Lubuklinggau include science, social studies, and religion. The problem of this research is that the difficulty in determining student majors becomes an obstacle for institutions in calculating the criteria for each student and reduces time efficiency. Determination of student majors is still done manually by inputting student data into Microsoft Excel so it is not effective. Observing and recording directly at the research site (observation), asking direct questions and receiving answers directly from sources (interviews), and documenting by reading standard literature were used to collect data for this study. The categorization system for student majors using the Nave Bayes approach uses R Software, which is designed to help student majors analyze test results more precisely.*

*Keywords: Student Majors, Software R, Classification System*

# PENDAHULUAN

Kemajuan teknologi keterangan dan komunikasi (TIK) telah memberikan pengaruh terhadap pendidikan, khususnya pada cara orang belajar. Semua aspek kehidupan berkembang seiring berjalannya waktu, termasuk bidang sosial, budaya, ekonomi, seni, dan teknologi keterangan dan komunikasi (TIK). Perkembangan tercepat yang terjadi saat ini adalah di ranah TIK. [1] Wajar saja jika menganggap bahwa madrasah yang merupakan lembaga pendidikan Islam merupakan kelanjutan dari lembaga pendidikan Islam sebelumnya. Beberapa derajat madrasah kini terbagi dalam kelompok-kelompok di Indonesia, antara lain Raudhatul Athfal (RA), Madrasah Ibtidaiyah (MI), Madrasah Tsanawiyah (MTs), dan Madrasah Aliyah/Kejuruan (MA/MAK). [2]

Salah satu lembaga pendidikan yang didirikan atas dasar agama Islam adalah Madrasah Aliyah Negeri (MAN) 2 Lubuklinggau. MAN 2 Lubuklinggau, yang dapat ditemukan di Taba Jemekeh, Kecamatan 1 Lubuklinggau Timur, Jl. Yos Sudarso No.50, Kota Lubuklinggau. Saat ini MAN 2 Lubuklinggau memiliki tiga jurusan/peminatan: IPA, IPS, dan agama. Siswa MAN 2 Lubuklinggau harus memilih salah satu dari tiga jurusan dengan menggunakan sistem yang telah ditetapkan oleh lembaga. Ini digunakan untuk mengklasifikasikan siswa untuk menetapkan jurusan dan bidang minat mereka. Meski siswa memiliki preferensi sendiri-sendiri, MAN 2 Lubuklingau harus mendistribusikan siswa sesuai dengan kemampuannya untuk menghindari komplikasi di kemudian hari dalam proses belajar mengajar.

Sementara beberapa ahli mendefinisikan aplikasi sebagai tindakan mempraktikkan teori, teknik, atau item lain untuk mencapai tujuan tertentu atau untuk kepentingan yang diinginkan oleh kelompok atau kelompok yang telah direncanakan dan disiapkan sebelumnya, yang lain mendefinisikan aplikasi sebagai tindakan melamar. [5]

Menemukan model atau fungsi yang menggambarkan atau membedakan konsep atau kelas data adalah proses klasifikasi. Tujuannya adalah untuk dapat memprediksi kelas suatu barang yang labelnya tidak diketahui. Untuk melakukan ini, proses klasifikasi membuat model yang dapat membagi data menjadi beberapa kelompok sesuai dengan aturan atau fungsi yang telah ditentukan. Aturan "jika-maka", pohon keputusan, atau rumus matematika semuanya dapat berfungsi sebagai model itu sendiri. [6]

Data mining, biasa disebut sebagai Knowledge Discovery in Databases (KDD), adalah proses memperoleh keterangan yang berpotensi berguna, implisit, dan ambigu dari kumpulan data. Penemuan pengetahuan dalam database termasuk mengubah hasil dari proses data mining (proses penggalian kecenderungan pola data) menjadi keterangan yang dapat dimengerti. [7]

Menemukan pola atau tautan dalam kumpulan data besar adalah teknik yang dikenal sebagai penambangan data, juga dikenal sebagai penemuan pengetahuan dalam basis data (KDD). Hasil dari proses data mining dapat digunakan untuk mengukur bagaimana pengambilan keputusan di masa depan akan berubah. [3]

## METODOLOGI PENELITIAN

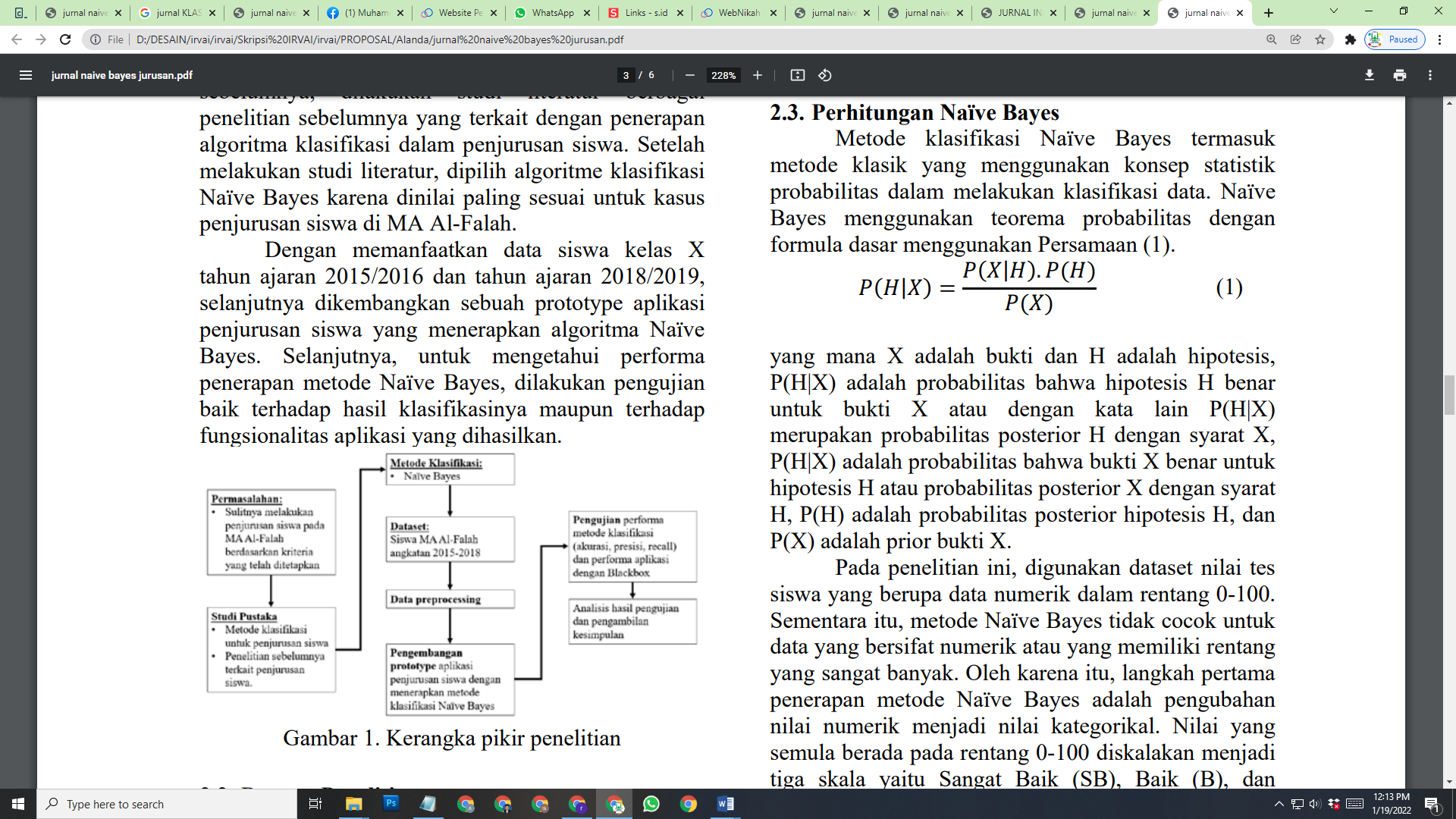
## Di antara algoritma lainnya, algoritma Naive Bayes merupakan teknik data mining yang masuk dalam 10 besar klasifikasi data mining [4].

## 

## Kelas dilambangkan dengan C, dan ciri-ciri prosedur klasifikasi dengan F1 sampai Fn. Rumus tersebut selanjutnya menjelaskan bahwa peluang suatu sampel memiliki karakteristik dari kelas C (posterior) sama dengan peluang munculnya kelas C (sebelum dimasukkannya sampel, disebut juga prior) dikalikan dengan peluang karakteristik sampel tersebut. terjadi di kelas C (juga disebut kemungkinan), dibagi dengan peluang karakteristik sampel yang muncul secara global (juga disebut bukti). Berikut ini dapat diturunkan dari penjelasan ini:

Saat mengevaluasi bukti untuk satu kelas, nilainya konsisten di semua sampel. Klasifikasi sampel akan ditentukan dengan membandingkan nilai posterior dengan nilai posterior kelas lain.

Untuk klasifikasi data, pendekatan Naive Bayes adalah teknik standar yang didasarkan pada gagasan probabilitas dalam statistik. Nave Bayes menerapkan teorema probabilitas menggunakan rumus dasar:



Untuk memperjelas, P(H|X) adalah probabilitas posterior dari H dengan suku X, P(H|X) adalah probabilitas posterior dari X yang diberikan H, P(H) adalah probabilitas posterior dari H, P(H|X ) adalah probabilitas posterior dari X jika diberikan H, dan P(X) adalah probabilitas sebelumnya dari X jika diberikan H.

Di MAN 2 Lubuklinggau, memilih jurusan siswa melibatkan beberapa langkah.

* 1. Pengumpulan data

Mencari bacaan yang relevan (buku, artikel, dll) dan melakukan pengumpulan data awal di lokasi penelitian. Hasil tes siswa MAN 2 Lubuklinggau digunakan untuk analisis ini. Variabel penelitian meliputi hasil tes matematika, bahasa Indonesia, bahasa Inggris, keyakinan agama, dan UN dan rata-rata keseluruhan rapor. Kumpulan data akan dibagi menjadi dua: set pelatihan dan pengujian.

* 1. *Prapemrosesan*

Tahap persiapan selesai sebelum tahap implementasi dimulai. Dimungkinkan untuk mengumpulkan 339 catatan sekaligus, tetapi hanya sebagian dari catatan tersebut dan karakteristik terkaitnya yang digunakan dalam analisis; hal ini karena data harus terlebih dahulu menjalani langkah pengolahan data awal, yang juga dikenal sebagai persiapan data. Membersihkan, mengubah, dan mengompresi data adalah semua tahapan yang diperlukan dalam perjalanan menuju keterangan berkualitas tinggi.

* 1. *Kumpulan data (diproses sebelumnya)*

*Dataset* ( *preprocessed* ) Dataset yang telah mengalami preprocessing, seperti transformasi dan reduksi, sehingga cocok untuk digunakan dalam pelatihan data dan pengujian data

* 1. *Pelatihan* Data dan *Pengujian Data*

Ini sering digunakan sebagai pola data dalam data pelatihan untuk memandu pengembangan model penambangan data. Ketika model klasifikasi diuji, itu diterapkan pada data pengujian.

* 1. Membangun Nave Model *B* ya

Berdasarkan data pelatihan yang telah diproses sebelumnya untuk digunakan dalam mengembangkan model Naive Bayes Classifier, untuk tujuan menyelesaikan masalah klasifikasi kelas atau atribut dalam data.

* 1. Penerapan Model Naïve Bayes

Model Naive Bayes Classifier memanfaatkan model yang sudah ada sebelumnya untuk melabeli data baru sesuai dengan properti atau klasifikasinya.

* 1. Hasil Akurasi

Metode Naive Bayes Classifier menghasilkan angka akurasi, yang digunakan untuk menentukan hasilnya. Kami menggunakan Matriks Kebingungan, Akurasi, dan Skor F1 untuk menentukan seberapa andal pola tersebut.

Penulis menggunakan pendekatan waterfall untuk penelitian ini. Ini adalah pendekatan umum yang digunakan oleh analis sistem. Sederhananya, pendekatan air terjun mengharuskan tugas sistem diselesaikan secara berurutan. Dengan kata lain, tidak dapat melanjutkan ke langkah kedua jika yang pertama belum selesai. Jika langkah 1 dan 2 sudah dilakukan, maka secara otomatis langkah 3 akan dijalankan.

Penjelasan Metodologi *Air Terjun :*

1. Analisis data

Dengan menggunakan metode statistik, siswa kelas X di MAN 2 Lubuklinggau dapat mempelajari mata pelajaran apa yang paling mereka minati. Data lokasi penelitian digunakan untuk membuat kumpulan data pelatihan dan kumpulan data pengujian. Tabel 5 menampilkan data atribut yang digunakan dalam analisis ini:

Tabel 3.1. Tabel Atribut

|  |  |
| --- | --- |
| **Atribut** | **Keterangan** |
| matematika nilai | Skor Matematika |
| Bahasa Inggris Nilai Bahasa Inggris | Nilai Indonesia |
| Nilai bahasa | Kelas Bahasa Inggris |
| Nilai agama | Nilai Religius |
| Rata rata UN | Rata-rata Skor UN |
| Rata rata laporan | Skor Laporan Rata-rata |

1. Pengumpulan data

Penelitian ini meneliti 339 data nilai mentah siswa dari tahun ajaran 2018–2019, 2019–2020, dan 2020–2021 dari MAN 2 Lubuklinggau. Itu digunakan untuk menghasilkan data pelatihan. Untuk menentukan dengan menggunakan teknik Naive Bayes maka perlu dibuat model perhitungan dari data latih. Tiga metode pengambilan yang berbeda digunakan untuk pengumpulan data: 1) 70% data pelatihan dan 30% data pengujian, 2) 30% data pelatihan dan 70% data pengujian, dan 3) 50% data pelatihan dan 50% data pengujian

Data yang diperoleh dari siswa berasal dari tes standar yang telah diselesaikan di jurusan siswa. Nilai matematika, nilai bahasa Indonesia, nilai bahasa Inggris, nilai agama, nilai rata-rata UN, dan nilai rata-rata rapor semuanya menjadi pertimbangan dalam proses pemilihan jurusan.

Tabel 3.2 Skala Penilaian

|  |  |  |
| --- | --- | --- |
| No | Skala Nilai | Nama Skala |
| 1 | 90-100 | Sangat bagus |
| 2 | 70-89 | Bagus |
| 3 | 50-69 | Cukup |
| 4 | 20-49 | Tidak cukup |
| 5 | 0-19 | Sangat kurang |

Nilai tes siswa digunakan sebagai data numerik dalam penelitian ini, dengan rentang 0-100. Namun, data kuantitatif atau sangat luas tidak cocok untuk pendekatan Naive Bayes. Untuk mulai menggunakan pendekatan Naive Bayes, perlu untuk mengubah data kontinu ke dalam kategori diskrit. Tabel 6 menunjukkan bagaimana suatu bilangan dengan rentang awal 0–100 diubah menjadi lima kategori: Sangat Baik (SB), Baik (B), Cukup (C), Kurang (K), dan Sangat Kurang (SK).

Contoh dataset penelitian yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3. Tabel 8 menampilkan konversi data dari Tabel 3 ke skala 5 poin sesuai dengan kriteria yang tercantum dalam Tabel 1.

Hal berikutnya yang harus dilakukan adalah mencari tahu seberapa besar kemungkinan setiap kategori. IPA = 10/20 = 0,5, IPS = 4/20 = 0,2, dan Agama = 6/20 = 0,3 adalah peluang masing-masing jurusan berdasarkan data pelatihan pada tabel ilustrasi 8.

## HASIL DAN PEMBAHASAN

## Pembahasan bab ini adalah tampilan isi dari sistem “Penerapan Algoritma Naïve Bayes Pada Klasifikasi Penjurusan mahasiswa di MAN 2 Lubuklinggau”.

1. Admin Memanggil Paket dan menampilkan Script untuk pembuatan
2. sistem klasifikasi jurusan
3. Masukan Dataset
4. Memasukkan Data Pelatihan
5. Tampilkan Data Pelatihan
6. Masukan Data Pengujian
7. Menampilkan Hasil Prediksi Matriks Kebingungan
8. Memasukkan Data Baru dan Menampilkan Data Baru

Tampilan Skrip Sistem Klasifikasi Jurusan

Pada gambar di bawah ini, setelah admin melakukan paket-paket, admin akan menampilkan script untuk melanjutkan pembuatan sistem klasifikasi jurusan mahasiswa.

|  |
| --- |
|  |

**Jurusan Klasifikasi Sistem Tampilan Skrip**

Tampilan kumpulan data

Dataset tersebut merupakan data mentah dari MAN 2 Lubuklinggau yaitu data nilai siswa periode 2018-2019, 2019-2020, dan 2020-2021 dengan jumlah siswa sebanyak 339 siswa. Data dalam dataset berisi 3 kriteria jurusan yaitu IPA, IPS, dan AGAMA, yang digunakan untuk mempersiapkan data pelatihan.

|  |
| --- |
|  |

Pada gambar data latih berupa data excel yang diolah, selanjutnya admin menginput data latih setelah menginput data tersebut software akan menampilkan halaman data latih sesuai dengan data yang diinputkan.

|  |
| --- |
|  |

Pada gambar data pengujian berupa data excel yang diolah kemudian admin menginput data pengujian setelah menginput data software akan menampilkan halaman data pengujian sesuai dengan data yang dimasukkan

|  |
| --- |
|  |

Tampilan Hasil Prediksi Matriks Kebingungan

Pada gambar dibawah ini dapat dijelaskan bahwa admin akan menampilkan hasil klasifikasi dari model yang telah dibentuk dengan 80% data latih dan 20% data uji.

|  |
| --- |
|  |

# KESIMPULAN

# Berdasarkan pembahasan yang telah dilakukan oleh penulis, maka dapat disimpulkan tentang penerapan algoritma Nave Bayes dalam klasifikasi jurusan di MAN 2 Lubuklinggau, dengan menggunakan Software R, yaitu sebagai berikut.

# 1. Dengan penerapan algoritma nave Bayes dalam klasifikasi jurusan di MAN 2 Lubuklinggau akan sangat membantu pihak sekolah dalam menentukan jurusan siswa

# 2. Dalam pengelolaan data siswa tidak perlu lagi digunakan secara manual yaitu karena adanya klasifikasi pengelola data siswa dengan hasil seleksi yang dilakukan oleh pendamping yang sudah tersistem dalam software.

# 3. Tingkat akurasi tes yang ditentukan oleh evaluasi Confusion Matrix adalah 0,9467 persen, termasuk dalam kelompok Klasifikasi Baik.

# V. SARAN

Beberapa saran yang dapat diberikan oleh penulis antara lain:

1. Sistem yang dibuat sebaiknya menggunakan lebih dari satu algoritma agar mendapatkan hasil yang optimal

2. Akurasi bisa ditingkatkan lagi agar lebih presisi (posisi yang tepat)

# DAFTAR PUSTAKA

[1] Abdaul, Irkham, “Perkembangan Teknologi Keterangan dan Komunikasi (TIK) Terhadap Kualitas Pembelajaran di Sekolah Dasar”, J. Education and Counseling, 2020.

[2] A. Zainul, M, A. Solichin, “Penerapan Metode Nave Bayes Classifier Bagi Siswa Jurusan Madrasah Aliyah Al-Falah Jakarta”, Jurnal Fountain of Informatics, Vol.5 No. 1 Mei 2020,

[3] Putria, NE 2018. “Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori di Terminal Tiket Tour & Travel Batam”. Jurnal Sistem Keterangan Berbasis Komputer, Vol.06 (01), 29-38

[4] Saputra, RA, Taufik, AR, Ramdhani, LS, Oktapian, R., & Marsusanti, E. (2018). Sistem Pendukung Keputusan Penentuan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes. SNIT 2018, 106–111.

[5] “Aplikasi,” 2021, [Online]. Tersedia: https://dspace.uii.ac.id/bitstream/handle/123456789/10672/05.2 bab 2.pdf?sequence=5&isAllowed=y#:~:text=Menurut Kamus Besar Bahasa Indonesia (KBBI)%2C aplikasi pemahaman adalah, oleh kelompok atau kelas.

[6] Arddy H. Hailitik dkk, “Klasifikasi Jurusan menggunakan metode Naïve Bayes di Sekolah Menengah Atas (SMAN) 1 Fatuleu Tengah”. J-ICON., Vol. 5., TIDAK. 2 Oktober 2017, hlm 21-27.

[7] Marsono, “Pelaksanaan Data Mining pada Pengaturan Pola Tata Letak Barang di

Berkah Self-Service untuk Strategi Penjualan Menggunakan Algoritma Apriori” *J*

*Teknologi Informatika dan Jaringan Nasional.* , jilid. 3, tidak. 2, 2019.