Volume 5, Nomor 2, April 2021, Page 513-523 ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i2.2925



Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu

Muhammad Gunawan^{1,*}, Muhammad Zarlis², Roslina³

¹ Fakultas Teknik dan Ilmu Komputer, Program Studi Magister Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia Email: ^{1,*}gunawan6451kz@gmail.com, ²m.zarlis@yahoo.com, ³roslinanich@gmail.com Email Penulis Korespondensi: m.rifqi@mercubuana.ac.id

Abstrak—Mahasiswa merupakan salah satu pilar penting dalam siklus hidup suatu perguruan tinggi. Dalam proses berkembangnya, suatu perguruan tinggi dapat dipengaruhi oleh berapa banyak jumlah lulusan strata-1 (S1) dari perguruan tinggi tersebut. Jumlah kelulusan suatu perguruan tinggi terkadang memiliki rasio yang rendah jika dibandingkan dengan jumlah mahasiswa yang diterima pada tahun ajaran yang sama. Tingkat kelulusan mahasiswa yang rendah ini dapat diakibatkan oleh beberapa faktor, seperti banyaknya kegiatan kemahasiswaan yang diikuti, faktor ekonomi, dan beberapa faktor tidak terduga lainnya. Hal ini membuat suatu perguruan tinggi harus memiliki suatu skema atau suatu rumus yang dapat memprediksi apakah mahasiswa tersebut dapat lulus tepat waktu. Normalnya, seorang mahasiswa strata-1 (S1) menempuh pendidikan selama 8 semester. Tetapi dengan adanya beberapa faktor yang telah disebutkan dapat membuat waktu menempuh pendidikan S1 menjadi lebih banyak, atau bahkan gagal lulus. Penelitian ini akan mencoba untuk membandingkan hasil Analisa dua metode dalam algoritma klasifikasi untuk memprediksi kelulusan mahasiswa. Algoritma yang digunakan ialah Algoritma K-Nearest Neighbour dan Naïve Bayes. Penelitian ini juga bertujuan mengidentifikasi algoritma terbaik di antara dua pilihan algoritma klasifikasi tersebut. Penelitian ini menghasilkan kesimpulan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang sama dengan algoritma KNN dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran yaitu sebesar 90 %.

Kata Kunci: Naïve Bayes; K-Nearest Neighbor; Kelulusan Mahasiswa; Prediksi; Perbandingan Algoritma

Abstract—Students are one of the important pillars in the life cycle of a university. In the process of developing, a university can be influenced by how many bachelor degree (S1) graduates from the university are. The number of graduations of a college sometimes has a low ratio when compared to the number of students admitted in the same school year. This low passing rate of students can be caused by several factors, such as the number of student activities that are participated in, economic factors, and several other unexpected factors. This makes a university must have a scheme or a formula that can predict whether the student can graduate on time. Normally, a bachelor (S1) student takes 8 semesters of education. But the existence of several factors that have been mentioned can make the time to take S1 education to be more, or even fail to graduate. This study will try to compare the results of the analysis of the two methods in the classification algorithm to predict student graduation. The algorithm used is the K-Nearest Neighbor and Naïve Bayes Algorithm. This study also aims to identify the best algorithm among the two classification algorithm choices. This research concluded that the Naïve Bayes algorithm has the same level of accuracy as the KNN algorithm in predicting the graduation of students in the Medical Education study program, which is 90%.

Keywords: Naïve Bayes; K-Nearest Neighbor; Student Graduation; Prediction; Algorithm Comparison

1. PENDAHULUAN

Mahasiswa merupakan salah satu pilar penting dalam siklus hidup suatu perguruan tinggi. Dalam proses berkembangnya, suatu perguruan tinggi dapat dipengaruhi oleh berapa banyak jumlah lulusan strata-1 (S1) dari perguruan tinggi tersebut. Jumlah kelulusan suatu perguruan tinggi terkadang memiliki rasio yang rendah jika dibandingkan dengan jumlah mahasiswa yang diterima pada tahun ajaran yang sama. Tingkat kelulusan mahasiswa yang rendah ini dapat diakibatkan oleh beberapa faktor, seperti banyaknya kegiatan kemahasiswaan yang diikuti, faktor ekonomi, dan beberapa faktor tidak terduga lainnya. Hal ini membuat suatu perguruan tinggi harus memiliki suatu skema atau suatu rumus yang dapat memprediksi apakah mahasiswa tersebut dapat lulus tepat waktu. Normalnya, seorang mahasiswa strata-1 (S1) menempuh pendidikan selama 8 semester. Tetapi dengan adanya beberapa faktor yang telah disebutkan dapat membuat waktu menempuh pendidikan S1 menjadi lebih banyak, atau bahkan gagal lulus.

Penelitian ini akan mencoba untuk membandingkan hasil Analisa dua metode dalam algoritma klasifikasi untuk memprediksi kelulusan mahasiswa. Algoritma yang digunakan ialah Algoritma K-Nearest Neighbour dan Naïve Bayes. Penelitian ini juga bertujuan mengidentifikasi algoritma terbaik di antara dua pilihan algoritma klasifikasi tersebut.

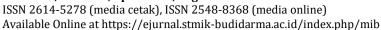
Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (*neighbor*) terdekat [1]. K-NN adalah rumus yang paling sederhana yang sering digunakan dalam pengimplementasian pencarian jarak. Rumus yang digunakan dalam perhitungan K-NN adalah rumus Euclidean Distance [2].

Algoritma Naïve Bayes merupakan algoritma data mining yang dapat digunakan untuk melakukan klasifikasi, dan merupakan algoritma yang akan memberikan hasil kemungkinan yang akan terjadi di masa akan datang dari hasil klasifikasi. Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidak tergantungan yang kuat. Selain itu naive bayes juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang [3].

Muhammad Gunawan, Copyright ©2021, MIB, Page 513 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

DOI 10.30865/mib.v5i2.2925





Berdasarkan masalah yang diuraikan maka penelitian ini membahas analisis komparasi algoritma naive bayes dan k-nearest neighbor. Tujuan diadakannya komparasi ini adalah diharapkan akan diketahui metode yang mendapatkan akurasi lebih besar. Berikutnya, metode yang terbaik diantara keduanya dipakai dalam menentukan prediksi kelulusan mahasiswa, yaitu antara tepat waktu atau terlambat, hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan metode pengujian Confusion Matrix dalam tool weka.

Penelitian yang dilakukan oleh Yuliandi dan Yondi menghasilkan kesimpulan bahwa metode K-Nearest Neighbor (KNN) efektif digunakan dalam Sistem Pendukung Keputusan Bantuan Siswa Miskin, karena KNN lebih tepat dalam perhitungan nilai [1].

Penelitian yang dilakukan oleh Nuari, dkk menghasilkan kesimpulan bahwa implementasi metode K-Nearest Neighbor dalam memprediksi varietas padi pada lahan pertanian dapat digunakan sebagai sarana untuk menentukan varietas padi yang cocok untuk ditanam disuatu lahan [4].

Penelitian yang dilakukan oleh Guntur, dkk menghasilkan kesimpulan bahwa algoritma Naïve Bayes dapat digunakan untuk memprediksi harga emas dengan nilai ketepatan sebesar 75 % [5].

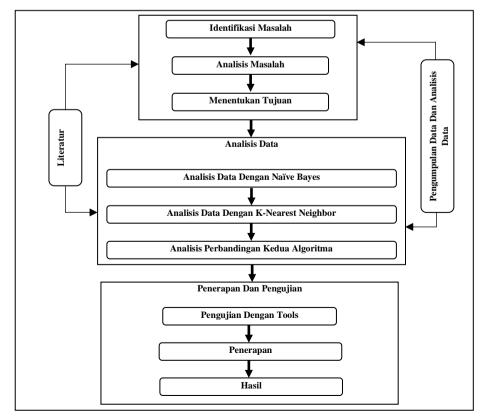
Penelitian yang dilakukan oleh Fadlan, dkk menghasilkan kesimpulan bahwa Algoritma Naive Bayes sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga memudahkan pengurus desa dalam menentukan keluarga yang berhak menerima beras rastra [6].

Penelitian yang sedang dilakukan mengambil data dari indeks prestasi mahasiswa yang diperoleh dari pengambilan matakuliah yang ditawarkan pada setiap semesternya. Kelulusan tepat waktu merupakan indikator keberhasilan mahasiswa dalam proses pendidikan dilihat dari nilai yang diperoleh, indeks prestasi serta faktor lain yang dapat menyebabkan kegagalan masa studi mahasiswa sehingga perlu adanya tool yang mampu memprediksi. Penekanan pada penelitian ini adalah mahasiswa Fakultas Kedokteran Universitas Muhammadiyah sumatera Utara sebagai training set yang akan diprediksi kelulusannya ditinjau dari usia, jenis kelamin dan indeks prestasi selama 4 semester pertama yaitu semester 1, semester 2, semester 3 dan semester 4. Selanjutnya dilakukan analisis komparasi algoritma klasifikasi yaitu Algoritma K-Nearest Neighbor dan Naïve Bayes sehingga dapat diketahui algoritma yang paling akurat untuk memprediksi kelulusan mahasiswa tepat waktu pada Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan tahapan-tahapan yang akan dilakukan dalam rangka penyelesaian masalah yang akan dibahas. Adapun kerangka kerja dari penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Muhammad Gunawan, Copyright ©2021, MIB, Page 514 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib

DOI 10.30865/mib.v5i2.2925



1. Identifikasi Masalah

Pada tahap ini dilakukan identifikasi masalah yang akan diteliti. Pada identifikasi masalah akan dilakukan pengenalan terhadap masalah atau inventarisasi masalah yang akan diteliti. Identifikasi masalah dilakukan dengan cara studi literatur dan observasi langsung di Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara. Masalah yang ditemukan dalam proses identifikasi masalah ialah masih banyak mahasiswa fakultas kedokteran yang tidak lulus tepat waktu.

2. Analisis Masalah

Permasalahan yang ditemukan kemudian akan dianalisa. Langkah dalam proses analisa masalah adalah langkah untuk memahami masalah yang telah ditentukan. Dengan menganalisa permasalahan yang telah ditentukan tersebut, maka diharapkan masalah tersebut dapat dipahami dengan baik. Dalam proses analisis masalah juga telah menemukan faktor-faktor penilaian apa saja yang akan digunakan untuk memprediksi kelulusan mahasiswa.

3. Menentukan Tujuan

Berdasarkan pemahaman dari permasalahan yang telah dianalisa, langkah berikutnya adalah menentukan tujuan yang akan dicapai dalam penelitian ini. Pada tujuan ini target yang akan dicapai adalah membandingkan dua metode untuk memprediksi kelulusan mahasiswa.

Setelah menentukan tujuan, maka selanjutnya dilakukan pengumpulan data. Pengumpulan data dilakukan dengan 2 cara yaitu:

a. Pengumpulan Data Primer.

Data primer yaitu data kelulusan atau data alumni mahasiswa program studi pendidikan kedokteran Universitas Muhammadiyah Sumatera Utara tahun angkatan 2015. Data primer tersebut diperoleh langsung dari Prodi Pendidikan Kedokteran UMSU.

b. Studi Literatur

Studi literatur dilakukan untuk mengkaji berbagai teori yang sudah pernah ada melalui jurnal. Pemahaman terhadap metode-metode yang akan dilakukan pada penalitan ini.

4. Analisis Data Dengan Naïve Bayes

Dalam tahap ini, data mahasiswa yang telah diperoleh akan diolah dengan menggunakan algoritma naïve bayes. Proses analisis dengan algoritma naïve bayes dimulai dengan penentuan faktor-faktor penilaian kemudian melakukan pemodelan dengan rumus-rumus yang dimiliki oleh algoritma naïve bayes.

5. Analisis Data Dengan K-Nearest Neighbor

Dalam tahap ini, data mahasiswa yang telah diperoleh akan diolah dengan menggunakan algoritma K-Nearest Neighbor. Proses analisis dengan algoritma K-Nearest Neighbour dimulai dengan penentuan faktor-faktor penilaian kemudian melakukan pemodelan dengan rumus-rumus yang dimiliki oleh algoritma K-Nearest Neighbour.

6. Analisis Perbandingan Kedua Algoritma

Pada tahap ini, hasil analisis dari kedua algoritma akan dibandingkan dengan data kelulusan mahasiswa Prodi Pendidikan Kedokteran Angkatan tahun 2016 untuk mengetahui perbedaan hasil yang diperoleh dari kedua algoritma ini.

7. Pengujian Dengan Tools

Pada tahap ini dilakukan pengujian data dengan menggunakan tools weka untuk melihat tingkat akurasi perhitungan algoritma yang telah dilakukan.

8. Penerapan

Pada tahap ini, algoritma akan diterapkan untuk memprediksi mahasiswa yang lulus tepat waktu atau tidak.

9. Hasil

Pada tahap ini diketahui hasil penelitian, yaitu penentuan algoritma mana antara algoritma naïve bayes dan algoritma k-nearest neighbor yang dapat digunakan untuk memprediksi kelulusan mahasiswa Prodi Pendidikan Kedokteran UMSU.

2.2 Klasifikasi

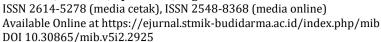
Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM) [7].

2.3 Algoritma Naïve Bayes

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian [8]. Naïve Bayes juga merupakan sebuah pengelompokan statistik yang bisa di dipakai untuk memprediksi probabilitas anggota suatu class [9].

Muhammad Gunawan, Copyright ©2021, MIB, Page 515 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523



Naïve Bayes berasumsi bahwa efek dari suatu pada kelas yang diberikan adalah independen terhadap nilai atirbut yang lainnya. Asumsi ini biasa disebut dengan class conditional inedependence. Itu dibuat untuk menyederhanakan komputasi yang tekait dan dalam hal ini disebut sebagai "naïve" [10]. Naïve Bayes Classifier merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidak tergantungan (independent) yang tinggi [11]. Keuntungan penggunaan Naïve Bayes Classifier adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [10].

Naive bayes merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [12]. Dalam metode ini perhitungan ditunjukkan melalui Persamaan 1 [13]:

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \tag{1}$$

Keterangan

X : Data dengan class yang belum diketahuiH : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X): Probabilitas X

2.4 K-Nearest Neighbor

Algoritma k-NN adalah suatu metode yang menggunakan algoritma supervised Perbedaan antara supervised learning dengan unsupervised learning adalah pada supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru [14]. Sedangkan pada unsupervised learning, data belum memiliki pola apapun, dan tujuan unsupervised learning untuk menemukan pola dalam sebuah data. Tujuan dari algoritma k-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples [15]. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN [16]. Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori [17]. Algoritma k-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru [4].

K-Nearest Neighbor menggunakan rumus perhitungan jarak. Jarak yang digunakan adalah jarak Euclidean Distance. Jarak Euclidean adalah jarak yang paling umum digunakan pada data numerik. Euclidean distance didefinisikan sebagai berikut [4]:

$$d(x_{i,}x_{j}) = \sqrt{\sum_{r=1}^{n} (a_{r}(x_{i}) - a_{r}(x_{j}))^{2}}$$
(2)

Keterangan

 $d(x_i,x_i)$: Jarak Euclidean (Eucledean Distance).

 (x_i) : record ke- i (x_j) : record ke- j (a_r) : data ke-r i,j :1,2,3,...n

2.5 Weka

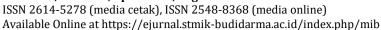
Weka adalah aplikasi data mining open source berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi /formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

2.6 Komparasi Akurasi Prediksi Kelulusan Mahasiswa Tepat Waktu

Menggunakan Algoritma Naïve Bayes dan Decision Tree Kerangka pemikiran dari penelitian ini bersumber dari fenomena akademik pada Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara yaitu ada ketidaksesuaian jumlah mahasiswa yang masuk dan yang lulus tepat waktu pada perguruan tinggi sehingga perlunya sebuah solusi yang tepat agar terdapat kesinambungan proses akademik yang relevan dengan mengevaluasi transaksi aktifitas kuliah mahasiswa 5 (lima) semester pertama.

Muhammad Gunawan, Copyright ©2021, MIB, Page 516 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523





Untuk mendapatkan nilai akurasi dari masa studi mahasiswa berdasarkan atribut tersebut menggunakan Algoritma Naïve Bayes kemudian mengkombinasikannya dengan algoritma Decision Tree yang bertujuan menemukan pola yang terdapat pada data mahasiswa berdasarkan data Indeks Pretasi Sementara (IPS) selama 5 (lima) semester pertama untuk memprediksi kelulusan mahasiswa tepat waktu.

Menghitung tingkat akurasi metode naive bayes dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung. Sedangkan Menghitung tingkat akurasi metode decision tree dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Kebutuhan Data

DOI 10.30865/mib.v5i2.2925

Sebelum melakukan proses perhitungan algoritma, terlebih dahulu dilakukan pengumpulan data. Data yang digunakan dalam penelitian ini ialah Data Alumni Mahasiswa Fakultas Kedokteran UMSU Tahun Angkatan 2015. Data mahasiswa yang digunakan adalah **100** data mahasiswa. Data penelitian yang diambil memiliki atribut input indeks prestasi semester (IPS) 1 hingga 5 dan atribut output ialah Kelulusan. Adapun data penelitian yang diperoleh dapat dilihat pada tabel 1 sebagai berikut:

Tabel 1. Data Mahasiswa Fakultas Kedokteran Tahun Angkatan 2015

No	Nama	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	Kelulusan
1	Nuryani	3,26	3,29	3,33	3,1	3,5	<= 4 Tahun
2	Yufi Yuwarditra	3,22	3,26	3,26	3,25	3,3	<= 4 Tahun
3	Rima Rahmi Putri Harahap	3,8	3,69	3,5	3,62	3,95	<= 4 Tahun
4	Ida Nuyani	3,23	3,05	3,21	3,19	3,3	<= 4 Tahun
5	Masyithah Pratiwi	2,89	2,9	2,6	3,02	2,65	> 4 Tahun
6	Muhammad Teguh Syahputra	3,34	3,19	2,98	2,45	2,55	<= 4 Tahun
7	Yuni Valentri L.Tobing	3,39	3,29	3,26	3,64	3,5	<= 4 Tahun
8	Fayan Nadya Shahiba. S	3,04	2,93	2,93	3,38	3,38	<= 4 Tahun
9	Fahrul Fadhli Panjaitan	2,89	3,17	3,26	3,4	3,28	<= 4 Tahun
10	Khairidho Rezeki Sembiring	3,57	3,17	1,48	2,74	1,4	> 4 Tahun
				•••			
99	Muhammad Al Anas	3,84	4	3,9	3,81	3,7	<= 4 Tahun
100	Adinda Nadira Larasati	3,41	3,33	2,64	3,36	3,09	> 4 Tahun

3.2 Transformasi Data

Setelah proses analisis kebutuhan data, maka langkah selanjutnya adalah melakukan proses transformasi data. Berdasarkan data alumni tersebut, data yang akan ditransformasi adalah IPS1, IPS2, IPS3, IPS4, dan IPS5. Data tersebut merupakan data yang akan dijadikan atribut *input* yang akan digunakan untuk analisis dengan algoritma naïve bayes dan k-nearest neighbour. Proses transformasi dilakukan dengan cara membuat klasifikasi pada masingmasing atribut *input*. Adapun klasifikasi atribut input dapat dilihat pada tabel 2 sebagai berikut.

Tabel 2. Klasifikasi Atribut Input

< 3,00 (Rendah)
$3,00 \le IPK \le 3,50 $ (Sedang)
>= 3,50 (Tinggi)
< 3,00 (Rendah)
$3,00 \le IPK \le 3,50 $ (Sedang)
>= 3,50 (Tinggi)
< 3,00 (Rendah)
$3,00 \le IPK \le 3,50 $ (Sedang)
>= 3,50 (Tinggi)
< 3,00 (Rendah)
$3,00 \le IPK \le 3,50 $ (Sedang)
>= 3,50 (Tinggi)
< 3,00 (Rendah)
$3,00 \le IPK \le 3,50 $ (Sedang)
>= 3,50 (Tinggi)

Pada tabel 2 di atas menjelaskan bahwa ketujuh atribut yang digunakan sebagai *input* sudah memiliki klasifikasi masing-masing sehingga memudahkan dalam proses *mining* untuk memprediksi kelulusan mahasiswa

Muhammad Gunawan, Copyright ©2021, MIB, Page 517 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)

Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib

DOI 10.30865/mib.v5i2.2925



berdasarkan kelima atribut tersebut. Lima atribut prediktor tersebut akan di analisa menggunakan algoritma naïve bayes dan algoritma k-nearest neighbour untuk menentukan tepat waktu atau tidak tepat waktunya kelulusan mahasiswa. Adapun atribut target dari data dapat dilihat pada tabel 3 sebagai berikut.

Tabel 3. Klasifikasi Atribut Output

Atribut Target	Klasifikasi
Tepat Waktu	Kelulusan <= 4 Tahun
Tidak Tepat Waktu	Kelulusan > 4 Tahun

Data yang akan digunakan untuk dianalisa terdiri dari 5 atribut input dan 1 atribut target. Untuk analisis data, akan digunakan 100 sample data yang diambil secara acak dari data alumni, setelah itu 100 sample data alumni tersebut ditransformasi sehingga menghasilkan keputusan (lulus tepat waktu dan tidak tepat waktu) seperti terlihat pada tabel 4 sebagai berikut.

Tabel 4. Transformasi Data Penelitian

No	Nama	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	Kelulusan
1	Nuryani	Sedang	Sedang	Sedang	Sedang	Tinggi	Tepat Waktu
2	Yufi Yuwarditra	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
3	Rima Rahmi Putri Harahap	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Tepat Waktu
4	Ida Nuyani	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
5	Masyithah Pratiwi	Rendah	Rendah	Rendah	Sedang	Rendah	Tidak Tepat Waktu
6	Muhammad Teguh Syahputra	Sedang	Sedang	Rendah	Rendah	Rendah	Tepat Waktu
7	Yuni Valentri L.Tobing	Sedang	Sedang	Sedang	Tinggi	Tinggi	Tepat Waktu
8	Fayan Nadya Shahiba. S	Sedang	Rendah	Rendah	Sedang	Sedang	Tepat Waktu
9	Fahrul Fadhli Panjaitan	Rendah	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
10	Khairidho Rezeki Sembiring	Tinggi	Sedang	Rendah	Rendah	Rendah	Tidak Tepat Waktu
	•••						•••
99	Muhammad Al Anas	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Tepat Waktu
100	Adinda Nadira Larasati	Sedang	Sedang	Rendah	Sedang	Sedang	Tidak Tepat Waktu

3.3 Analisis Algoritma Naïve Bayes

Perhitungan probabilitas data yang belum diketahui lulus tepat waktu atau tidak tepat waktu dapat dihitung dengan algoritma naïve bayes sebagai berikut:

X (IPS1: Tinggi, IPS2: Sedang, IPS3: Tinggi, IPS4: Tinggi, IPS5: Tinggi).

Selanjutnya dilakukan perhitungan:

1. Menghitung nilai P(H)

P (Tepat Waktu) =
$$\frac{76}{100}$$
 = 0,76

Menghitung iniai P(11) P (Tepat Waktu) = $\frac{76}{100}$ = 0,76 Dimana: nilai 76 diperoleh dari jumlah Target Tepat Waktu, sementara 100 merupakan jumlah seluruh data. P (Tidak Tepat Waktu) = $\frac{24}{100}$ = 0,24 P (Tidak Tepat Waktu) = $\frac{24}{100}$ = 0,24 P (Tidak Tepat Waktu) = $\frac{24}{100}$ = 0,24

P (Tidak Tepat Waktu) =
$$\frac{24}{100}$$
 = 0,24

 $2. \ \ Menghitung \ nilai \ P(X|Hi) \ untuk \ i=1, \ 2, \ dst.$

a. P (IPS 1 = "Tinggi" | Target = "Tepat Waktu") =
$$\frac{14}{76}$$
 = 0,184

Dimana: nilai 14 diperoleh dari jumlah IPS1 Tinggi yang lulus tepat waktu, sementara 76 merupakan jumlah lulus tepat waktu.

P (IPS 1 = "Tinggi" | Target = "Tidak Tepat Waktu") =
$$\frac{1}{24}$$
 = 0,042

Dimana: nilai 1 diperoleh dari jumlah IPS 1 Tinggi yang lulus tidak tepat waktu, sementara 24 merupakan jumlah lulus tidak tepat waktu.

b. P (IPS 2 = "Sedang" | Target = "Tepat Waktu") =
$$\frac{39}{76}$$
 = 0,513

Dimana: nilai 39 diperoleh dari jumlah IPS 2 Sedang yang lulus tepat waktu, sementara 76 merupakan jumlah lulus tepat waktu.

P (IPS 2 = "Sedang" | Target = "Tidak Tepat Waktu") =
$$\frac{12}{24}$$
 = 0,5

Dimana: nilai 12 diperoleh dari jumlah IPS 2 Sedang yang lulus tidak tepat waktu, sementara 24 merupakan jumlah lulus tidak tepat waktu.

c. P (IPS 3 = "Tinggi" | Target = "Tepat Waktu") =
$$\frac{17}{76}$$
 = 0,223

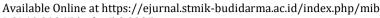
Dimana: nilai 17 diperoleh dari jumlah IPS 3 Tinggi yang lulus tepat waktu, sementara 76 merupakan jumlah lulus tepat waktu.

P (IPS 3 = "Tinggi" | Target = "Tidak Tepat Waktu") =
$$\frac{0}{24}$$
 = 0

Muhammad Gunawan, Copyright ©2021, MIB, Page 518 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)



DOI 10.30865/mib.v5i2.2925



Dimana: nilai 0 diperoleh dari jumlah IPS 3 Tinggi yang lulus tidak tepat waktu, sementara 24 merupakan jumlah lulus tidak tepat waktu.

d. P (IPS 4 = "Tinggi" | Target = "Tepat Waktu") = $\frac{20}{76}$ = 0,263

Dimana: nilai 20 diperoleh dari jumlah IPS 4 Tinggi yang lulus tepat waktu, sementara 76 merupakan jumlah lulus tepat waktu.

P (IPS 4 = "Tinggi" | Target = "Tidak Tepat Waktu") =
$$\frac{0}{24}$$
 = 0

Dimana: nilai 0 diperoleh dari jumlah IPS 4 Tinggi yang lulus tidak tepat waktu, sementara 24 merupakan jumlah lulus tidak tepat waktu.

e. P (IPS 5 = "Tinggi" | Target = "Tepat Waktu") = $\frac{30}{76}$ = 0,395 Dimana: nilai 20 diperoleh dari jumlah IPS 5 Tinggi yang lulus tepat waktu, sementara 76 merupakan jumlah lulus tepat waktu.

P (IPS 5 = "Tinggi" | Target = "Tidak Tepat Waktu") =
$$\frac{0}{24}$$
 = 0

Dimana: nilai 0 diperoleh dari jumlah IPS 5 Tinggi yang lulus tidak tepat waktu, sementara 24 merupakan iumlah lulus tidak tepat waktu.

- 3. Kalikan semua atribut input yang memiliki Target Tepat Waktu dan Tidak Tepat Waktu.
 - a. P(X|Target = "Tepat Waktu") = 0.184 * 0.513 * 0.223 * 0.263 * 0.395 = 0.002186
 - b. P (X|Target = "Tidak Tepat Waktu") = 0.042 * 0.5 * 0 * 0 * 0 = 0
- 4. Lakukan perhitungan dengan rumus naïve bayes P(H|X) = P(X|H) * P(H)
 - a. P(X I Target="Tepat Waktu") P(Target="Tepat Waktu") = 0,002186 * 0,76 = 0,00167
 - b. P(X I Target="Tidak Tepat Waktu") P(Target="Tidak Tepat Waktu") = 0 * 0,76 = 0
- 5. Berdasarkan perhitungan pada nomor 4 dapat diperoleh bahwa nilai Tepat Waktu lebih besar dibandingkan dengan nilai Tidak Tepat Waktu, maka dari itu untuk data X(IPS1: Tinggi, IPS2: Sedang, IPS3: Tinggi, IPS4: Tinggi, IPS5: Tinggi) memiliki Target= "Tepat Waktu".

Setelah dilakukan perhitungan untuk 30 data baru, maka diperoleh hasil sebagai berikut:

Tabel 5. Hasil Perhitungan Algoritma Naïve Bayes Untuk Data Baru

No	Nama	IPS1	IPS2	IPS3	IPS4	IPS5	Kelulusan
1	Leni andini	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Tepat Waktu
2	Rizky syahriani	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
3	Hafiz anugrah mursyid	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
4	Arman maulana	Rendah	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
5	Atika rahmi	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
6	Reyna cintiya	Sedang	Rendah	Sedang	Sedang	Sedang	Tepat Waktu
7	M, fahriza winaldha nst	Sedang	Sedang	Sedang	Tinggi	Sedang	Tepat Waktu
8	Rika rianingsih	Sedang	Sedang	Sedang	Sedang	Sedang	Tepat Waktu
9	Muhammad alip meruza	Sedang	Sedang	Sedang	Tinggi	Sedang	Tepat Waktu
	salim						
10	Diah indah arizka	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Tepat Waktu
	•••					•••	
29	Kartika handayani	Tinggi	Sedang	Tinggi	Tinggi	Tinggi	Tepat Waktu
30	Andri muhammad idris	Rendah	Rendah	Rendah	Rendah	Rendah	Tidak Tepat Waktu

3.4 Analisis Algoritma K-Nearest Neighbour

Perhitungan probabilitas data yang belum diketahui lulus tepat waktu atau tidak tepat waktu dapat dihitung dengan algoritma k-nearest neighbour sebagai berikut:

X (IPS1: 3.7, IPS2: 3.318, IPS3: 3.549, IPS4: 3.361, IPS5: 3.695).

Langkah-langkah perhitungan algoritma sebagai berikut:

1. Hitung jarak data testing ke semua data training. Adapun hasil perhitungan jarak Euclidean adalah sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}$$

$$d(X_i, X_j) = \sqrt{((3,26 - 3,7)^2) + ((3,29 - 3,318)^2) + ((3,3 - 3,549)^2) + ((3,1 - 3,631)^2) + ((3,5 - 3,695)^2))}$$

$$d(X_i, X_j) = 0,749887$$

Muhammad Gunawan, Copyright ©2021, MIB, Page 519 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)

Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i2.2925



- 2. Menentukan nilai k, dalam hal ini k yang digunakan = 5, kerena 5 merupakan k yang paling optimal.
- 3. Mengurutkan data jarak (dari kecil ke besar) dan mengambil sebanyak k (5) data teratas, sehingga didapatkan 5 deret data sebagai berikut:

Tabel 6. Hasil Uji Pada Data Baru

No	Ips 1	Ips 2	Ips	Ips	Ips	Baru 1	Kelulusan
12	3,34	3,4	3,38	3,5	3,6	0,437117	Tepat Waktu
24	3,75	3,69	3,57	3,79	3,53	0,440262	Tepat Waktu
3	3,8	3,69	3,5	3,62	3,95	0,464684	Tepat Waktu
7	3,39	3,29	3,26	3,64	3,5	0,467452	Tepat Waktu
97	3,36	3,45	3,26	3,43	3,68	0,50712	Tepat Waktu

4. Pilih data kondisi terbanyak. Dalam hal ini terlihat pada data mayoritas atau 5 dari 5 data menunjukan bahwa kelulusan Tepat Waktu.

Setelah dilakukan perhitungan untuk 30 data baru, maka diperoleh hasil sebagai berikut:

Tabel 7. Hasil Uji Pada 30 Data Baru

No	Nama	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	Kelulusan
1	Leni andini	3,7	3,318	3,549	3,631	3,695	Tepat Waktu
2	Rizky syahriani	3,201	3,36	3,2	3,4	3,223	Tepat Waktu
3	Hafiz anugrah mursyid	3,301	3,249	3,117	3,499	3,305	Tepat Waktu
4	Arman maulana	2,899	3,117	3,2	3,417	3,055	Tepat Waktu
5	Atika rahmi	3,349	3,167	3,334	3,417	3,497	Tepat Waktu
6	Reyna cintiya	3,334	2,833	3,051	3,252	3,332	Tepat Waktu
7	M, fahriza winaldha nst	3,25	3,217	3,25	3,7	3,388	Tepat Waktu
8	Rika rianingsih	3,201	3,050	3,184	3,433	3,195	Tepat Waktu
9	Muhammad alip meruza salim	3,118	3,25	3,102	3,535	3,332	Tepat Waktu
10	Diah indah arizka	3,85	3,785	3,851	3,868	3,835	Tepat Waktu
	•••		•••	• • •	•••	•••	
29	Kartika handayani	3,666	3,418	3,535	3,598	3,665	Tepat Waktu
30	Andri muhammad idris	1,633	0,566	0	0	0	Tidak Tepat Waktu

3.5 Perbandingan Algoritma Naïve Bayes Dengan Algoritma K-Nearest Neighbour

Hasil dari analisis algoritma naïve bayes dan algoritma k-nearest neighbor kemudian akan dibandingkan dengan data real yang ada. Data yang akan digunakan untuk perbandingan ialah data 30 mahasiswa Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara (UMSU) Tahun Angkatan 2016. Adapun hasil perbandingan dari kedua algoritma tersebut adalah sebagai berikut:

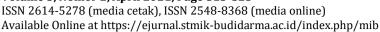
Tabel 8. Perbandingan Algoritma Naïve Bayes Dengan K-Nearest Neighbor

No	Nama	Naïve Bayes	K-Nearest Neighbour	Data Angkatan 2016
1	Leni andini	Tepat Waktu	Tepat Waktu	Tepat Waktu
2	Rizky syahriani	Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
3	Hafiz anugrah mursyid	Tepat Waktu	Tepat Waktu	Tepat Waktu
4	Arman maulana	Tepat Waktu	Tepat Waktu	Tepat Waktu
5	Atika rahmi	Tepat Waktu	Tepat Waktu	Tepat Waktu
6	Reyna cintiya	Tepat Waktu	Tepat Waktu	Tepat Waktu
7	M, fahriza winaldha nst	Tepat Waktu	Tepat Waktu	Tepat Waktu
8	Rika rianingsih	Tepat Waktu	Tepat Waktu	Tepat Waktu
9	Muhammad alip meruza salim	Tepat Waktu	Tepat Waktu	Tepat Waktu
10	Diah indah arizka	Tepat Waktu	Tepat Waktu	Tepat Waktu
		•••		
29	Kartika handayani	Tepat Waktu	Tepat Waktu	Tepat Waktu
30	Andri muhammad idris	Tidak Tepat Waktu	Tidak Tepat Waktu	Tidak Tepat Waktu

Berdasarkan hasil perbandingan yang telah dilakukan terhadap 30 data testing maka dapat ditarik kesimpulan bahwa algoritma naïve bayes dan algoritma k-nearest neighbor memiliki tingkat keberhasilan prediksi yang sama yaitu 90%.

Muhammad Gunawan, Copyright ©2021, MIB, Page 520 Submitted: **13/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

Volume 5, Nomor 2, April 2021, Page 513-523





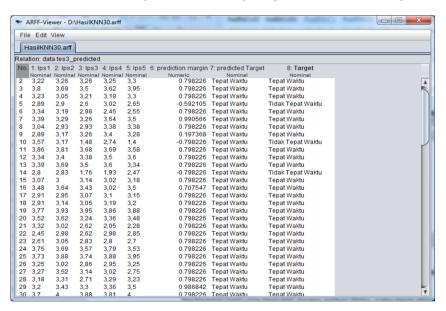
3.6 Hasil Pengujian

DOI 10.30865/mib.v5i2.2925

3.6.1 Hasil Pengujian Algoritma K-Nearest Neighbour Dengan WEKA

Hasil yang diperoleh dari perhitungan sebelumnya yaitu algoritma knn dengan nilai n=5 dan data testing berjumlah 30 data, memiliki tingkat akurasi 90 % untuk memprediksi mahasiswa yang Lulus Tepat Waktu atau Tidak. Aplikasi WEKA dapat digunakan untuk melakukan klasifikasi dengan algoritma knn. Penggunaan aplikasi ini bertujuan untuk melihat seberapa akurat perhitungan manual yang telah dilakukan. Adapun hasil yang diperoleh setelah melakukan analisis dengan aplikasi weka dapat dilihat pada gambar berikut:

Gambar 2. Proses Analisis Algoritma KNN Dengan Aplikasi WEKA Terhadap 30 Data Testing



Gambar 3. Hasil Analisis Algoritma KNN Dengan Aplikasi WEKA Terhadap 30 Data Testing

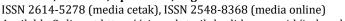


Gambar 4. Diagram Kedekatan Data Dengan Algoritma KNN

Berdasarkan hasil yang diperoleh dengan aplikasi Weka, maka dapat diketahui bahwa tingkat keberhasilan algoritma knn dalam memprediksi kelulusan mahasiswa adalah 90 %.

Muhammad Gunawan, Copyright ©2021, MIB, Page 521 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

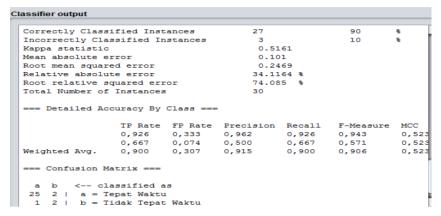




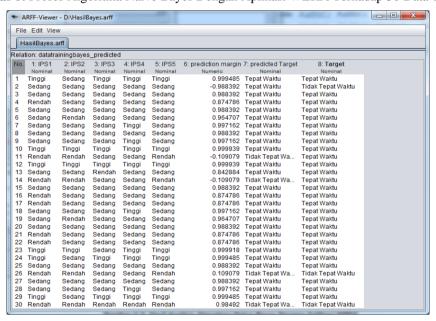
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i2.2925

3.6.2 Hasil Pengujian Algoritma Naïve Bayes Dengan WEKA

Sama halnya dengan algoritma knn, algoritma naïve bayes juga akan diuji dengan 30 data testing. Data testing ini akan diuji menggunakan aplikasi Weka untuk melihat seberapa tingkat keberhasilan algoritma naïve bayes dalam memprediksi kelulusan mahasiswa. Adapun hasil dari analisis dengan aplikasi Weka dapat dilihat pada gambar berikut:



Gambar 5. Proses Algoritma Naïve Bayes Dengan Aplikasi WEKA Terhadap 30 Data Testing



Gambar 6. Hasil Analisis Algoritma Naïve Bayes Dengan Aplikasi WEKA Terhadap 30 Data Testing



Gambar 7. Grafik Data Dengan Algoritma Naïve Bayes

Berdasarkan hasil analisa dengan menggunakan aplikasi weka maka diperoleh hasil tingkat keberhasilan algoritma naïve bayes mencapai 90%. Hasil ini sama dengan hasil yang diperoleh dengan perhitungan secara manual.

Algoritma Naïve Bayes memiliki tingkat akurasi yang sama dengan algoritma KNN dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran yaitu sebesar 90%. Dalam proses analisis, algoritma naïve bayes memiliki proses yang lebih panjang karena harus melakukan transformasi data terlebih dahulu, sedangkan algoritma k-nearest neighbor memiliki proses yang lebih sedikit dalam proses analisanya. Sehingga

> Muhammad Gunawan, Copyright ©2021, MIB, Page 522 Submitted: 13/03/2021; Accepted: 14/04/2021; Published: 25/04/2021

Volume 5, Nomor 2, April 2021, Page 513-523

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib

DOI 10.30865/mib.v5i2.2925



dapat ditarik kesimpulan bahwa algoritma k-nearest neighbor lebih baik karena memiliki proses yang lebih sedikit dan tingkat akurasi yang tinggi.

4. KESIMPULAN

Berdasarkan dari penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang sama dengan algoritma KNN dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran yaitu sebesar 90 %. Hasil ini diperoleh dari analisis yang dilakukan dengan aplikasi Weka terhadap 30 data testing dan 100 data training. Dalam proses analisis, algoritma naïve bayes memiliki proses yang lebih panjang karena harus melakukan transformasi data terlebih dahulu, sedangkan algoritma k-nearest neighbor memiliki proses yang lebih sedikit dalam proses analisanya. Dalam hal ini, maka dapat ditarik kesimpulan bahwa algoritma k-nearest neighbor lebih unggul dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

REFERENCES

- [1] E. Yulianti and Y. A. Nurdin, "SISTEM PENDUKUNG KEPUTUSAN PENERIMAAN BANTUAN SISWA MISKIN (BSM) BERBASIS ONLINE DENGAN METODE KNN (K-NEAREST NEIGHBOR) (Studi kasus: SMPN 1 Koto XI Tarusan)," *TEKNOIF*, vol. 6, no. 1, pp. 12–17, 2018.
- [2] N. L. G. P. Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil," Techno, vol. 16, no. 2, pp. 120–131, 2017.
- [3] R. N. Devita, H. W. Herwanto, and A. Wibawa Prasetya, "PERBANDINGAN KINERJA METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI ARTIKEL BERBAHASA INDONESIA," *JTIIK*, vol. 5, no. 4, pp. 427–434, 2018.
- [4] R. Nuari, A. Apriliyani, Juwari, and Kusrini, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK MEMPREDIKSI VARIETAS PADI YANG COCOK UNTUK LAHAN PERTANIAN," *INFORMA*, vol. 4, no. 2, pp. 2–8, 2018.
- [5] M. Guntur, J. Santony, and Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko," *RESTI*, vol. 2, no. 1, pp. 354–360, 2018.
- [6] C. Fadlan, S. Ningsih, and A. P. Windarto, "PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KELAYAKAN KELUARGA PENERIMA BERAS RASTRA," *JUTIM*, vol. 3, no. 1, pp. 1–8, 2018.
- [7] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES," Ilk. J. Ilm., vol. 10, no. 2, pp. 160–165, 2018.
- [8] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, 2020.
- [9] R. W. Pratiwi and Y. S. Nugroho, "Prediksi Rating Film Menggunakan Metode Naïve Bayes," *Tek. Elektro*, vol. 8, no. 2, pp. 60–63, 2016.
- [10] Y. I. Kumiawan, "PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING," JTIIK, vol. 5, no. 4, pp. 455–464, 2018.
- [11] V. Vangara and S. P. Vangara, "Opinion Mining Classification using Naive Bayes Algorithm," *IJITEE*, vol. 9, no. 5, 2020.
- [12] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)," *SAINTIKOM*, vol. 15, no. 2, pp. 81–92, 2016.
- [13] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *J. RESTI*, vol. 1, no. 10, pp. 95–101, 2021.
- [14] A. P. Giovani, Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," J. TEKNOINFO, vol. 14, no. 2, pp. 116– 124, 2020.
- [15] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A New K-Nearest Neighbors Classifier for Big Data Based on E ffi cient Data Pruning," *Mathematics*, vol. 8, no. 286, pp. 1–12, 2020.
- [16] L. Farokhah, "IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BUNGA IMPLEMENTATION OF K-NEAREST NEIGHBOR FOR FLOWER CLASSIFICATION WITH EXTRACTION OF RGB COLOR FEATURES," *JTIIK*, vol. 7, no. 6, pp. 1129–1136, 2020.
- [17] Muhathir, T. T. S. Sibarani, and Al-Khowarizmi, "Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction," *AloCSIT*, vol. 1, no. 1, pp. 33–38, 2020.

Muhammad Gunawan, Copyright ©2021, MIB, Page 523 Submitted: **13/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**