

PENELITIAN PREDIKSI KELULUSAN MAHASISWA PROGRAM STUDI SISTEM INFORMASI DI STMIK ROYAL KISARAN DENGAN PENGUNAAN METODE NAIVE BAYES

Dinda Khairani¹, Septy Nur Intan^{1*}, Wiwin Handoko¹

¹Sistem Informasi, STMIK Royal Kisaran

*email: septynurintan05@gmail.com

Abstract: Universities are responsible for creating graduates who have adequate skills. Measuring student graduation rates is considered a critical parameter in achieving this goal. This research was conducted to determine the number of graduates from each classification, namely graduating on time or not accurately. Bearing in mind that STMIK Royal Kisaran already has B accreditation. The success of students in graduating on time has a significant impact on the assessment elements in the accreditation process of higher education institutions. The use of the Naïve Bayes Algorithm method in applying data mining can be a solution to overcome the problem of determining the number of graduations. Naïve Bayes is a technique or method in the field of Data Mining that uses probability and statistics to make predictions. This method is able to project opportunities for future events based on past experience. The dataset in this research is STMIK Royal Kisaran Student Graduation data in 2017 as many as 608 student data which were divided into 2 (two) with a ratio of 70:30, 425 training data and 183 testing data managed to obtain an accuracy value of 61%. The Naïve Bayes algorithm produces the following results: precision for the negative class is 45%, for the positive class is 96%; recall for the negative class was 97%, for the positive class was 44%; f1-score for the negative class is 62%, for the positive class is 61%; and accuracy of 61%.

Keywords: *Graduation; Data Mining; Naïve Bayes*

Abstrak: Perguruan tinggi bertanggung jawab untuk menciptakan lulusan yang memiliki keterampilan yang memadai. Pengukuran tingkat kelulusan mahasiswa dianggap sebagai parameter kritis dalam mencapai sasaran ini. Adapun penelitian ini dilakukan untuk menentukan jumlah kelulusan dari masing-masing klasifikasi yaitu Lulus tepat waktu atau tidak secara akurat. Mengingat bahwa STMIK Royal Kisaran sudah memiliki akreditasi B. Keberhasilan mahasiswa dalam lulus tepat waktu memiliki dampak signifikan terhadap penilaian unsur dalam proses akreditasi institusi pendidikan tinggi. Penggunaan metode Algoritma *Naïve Bayes* dalam penerapan data mining dapat menjadi solusi untuk mengatasi permasalahan penentuan jumlah kelulusan. *Naïve Bayes* adalah sebuah teknik atau metode dalam bidang *Data Mining* yang menggunakan probabilitas dan statistik untuk melakukan prediksi. Metode ini mampu memproyeksikan peluang kejadian di masa mendatang berdasarkan pengalaman masa lalu. Dataset dalam penelitian ini adalah data Kelulusan Mahasiswa STMIK Royal Kisaran pada tahun 2017 sebanyak 608 data mahasiswa yang dibagi menjadi 2 (dua) dengan rasio 70:30, 425 *data training* dan 183 *data testing* berhasil memperoleh nilai akurasi sebesar 61%. Algoritma *Naïve Bayes* menghasilkan hasil sebagai berikut: presisi untuk kelas negatif sebesar 45%, untuk kelas positif sebesar 96%; recall untuk kelas negatif sebesar 97%, untuk kelas positif sebesar 44%; f1-score untuk kelas negatif sebesar 62%, untuk kelas positif sebesar 61%; dan akurasi sebesar 61%.

Kata kunci: *Kelulusan; Data Mining; Naïve Bayes*

PENDAHULUAN

Kemajuan teknologi informasi terus berkembang pesat di berbagai bidang kehidupan, menghasilkan banyak data yang luas dari sektor-sektor seperti industri, ekonomi, ilmu dan teknologi, serta aspek kehidupan lainnya. Dalam dunia pendidikan, penerapan teknologi informasi juga menghasilkan sejumlah besar data terkait dengan siswa dan proses pembelajaran [1].

Evaluasi program studi perlu memberikan perhatian serius pada mahasiswa, yang merupakan faktor kunci dalam menilai keberhasilan program studi. Durasi studi mahasiswa, yang mencakup waktu yang diperlukan untuk menyelesaikan program studi, menjadi indikator utama dalam evaluasi ini. Lama studi tidak hanya mencerminkan seberapa cepat mahasiswa menyelesaikan studinya, tetapi juga mencerminkan tingkat prestasi mereka dalam perjalanan akademis. Dalam cakupan yang lebih besar, lamanya durasi studi mahasiswa memiliki pengaruh terhadap kualitas keseluruhan program studi, sehingga menjadi salah satu kriteria evaluasi akreditasi yang digunakan oleh Badan Akreditasi Nasional Perguruan Tinggi [2].

Perguruan tinggi memiliki tanggung jawab untuk menghasilkan lulusan yang memiliki kompetensi. Evaluasi terhadap tingkat kelulusan mahasiswa dianggap sebagai indikator penting dalam mencapai tujuan tersebut [3]. Kelulusan mahasiswa merupakan unsur penilaian signifikan dalam proses akreditasi suatu perguruan tinggi. BAN-PT menetapkan poin khusus untuk evaluasi lulusan setiap perguruan tinggi. Oleh karena itu, jika sebuah perguruan tinggi memiliki tingkat kelulusan mahasiswa yang sesuai pada setiap semester, hal tersebut dapat memberikan kontribusi positif dalam mendukung proses penilaian akreditasi. Meskipun demikian, dalam praktiknya, tidak dapat dipastikan bahwa tingkat kelulusan mahasiswa selalu mencerminkan pencapaian lulus tepat waktu [4].

STMIK Royal terletak di Jl. Prof.H.M.Yamin No.173, Kisaran Naga, Kec. Kota Kisaran Timur, Kabupaten Asahan. Dalam perguruan tinggi mahasiswa yang tidak lulus tepat waktu dapat mempengaruhi unsur penilaian dalam proses akreditasi. Oleh karena itu, penting untuk melakukan klasifikasi mahasiswa yang lulus tepat waktu dan tidak lulus tepat waktu. Klasifikasi adalah hal yang sangat penting dalam mengelola data mining [5]. Klasifikasi juga berarti mengelompokkan data, di mana data yang digunakan memiliki kelas atau target yang perlu diidentifikasi [6]. sehingga, algoritma dapat digunakan dalam menyelesaikan masalah klasifikasi [7]. Dengan melakukan klasifikasi tersebut data yang diperoleh diharap mampu membantu kampus dalam meningkatkan proses pembelajaran yang lebih baik.

Data mining adalah upaya untuk mengungkap pola menarik dan mendapatkan pemahaman dari sejumlah besar data. [8]. *Berry & Linoff* (1997) menjelaskan data mining sebagai proses menggali dan menganalisis data dalam skala besar dengan tujuan mengungkap makna di balik pola dan aturan [9]. Dari berbagai teori yang dikemukakan oleh para ahli, dapat disimpulkan bahwa data *mining* merupakan kegiatan pencarian dan analisis pada *dataset* untuk menemukan pola menarik. Sasaran utamanya adalah mengekstraksi informasi dan pengetahuan yang tepat dan berpotensi, sehingga dapat dipahami dan bermanfaat dalam proses pengambilan keputusan.

Berdasarkan masalah diatas, STMIK Royal Kisaran membutuhkan suatu sistem yang dapat menentukan jumlah kelulusan dari masing-masing klasifikasi yaitu Lulus tepat waktu atau tidak secara akurat.

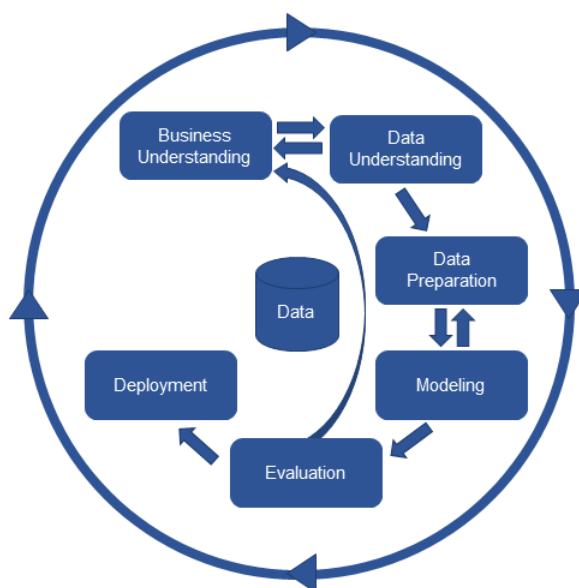
Penelitian yang dilakukan oleh [11], berjudul "Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma *Naïve Bayes*", menggunakan data pendidikan alumni D-3 STMIK YMI Tegal, Jurusan Manajemen Informatika dari tahun 1999 hingga 2014. Metode yang digunakan adalah *Naive Bayes* dengan kriteria NIM, IPS1, IPS2, IPS3, IPS4, tahun masuk, tahun lulus, jenis kelamin, status pekerjaan, dan status pernikahan. Penelitian ini mencapai tingkat akurasi sebesar 91,37%. Penelitian lainnya oleh [12] menyimpulkan bahwa setiap perguruan tinggi memiliki database yang menyimpan data akademik dan biodata mahasiswa. Informasi ini dapat diproses untuk mengkaji pola dan tingkah laku mahasiswa dengan tujuan mengurangi keterlambatan dalam kelulusan. Salah satu teknik yang dapat diterapkan untuk mengevaluasi basis data tersebut adalah data *mining*.

Penelitian di STMIK Royal Kisaran bertujuan untuk menggunakan Algoritma *Naïve Bayes* dalam mengklasifikasikan mahasiswa yang lulus tepat waktu dan yang tidak. Harapannya, hasil dari penelitian ini dapat membantu meningkatkan jumlah mahasiswa yang lulus tepat waktu sesuai dengan peraturan Menteri Pendidikan.

METODE

Data mining bisa dijelaskan sebagai langkah-langkah untuk menemukan informasi dan pola yang bermanfaat dari kumpulan data yang besar. Proses ini melibatkan mengumpulkan, mengekstrak, menganalisis, dan menggunakan statistik pada data [10]. Data mining merupakan proses yang memanfaatkan matematika, statistik, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan menemukan informasi yang berharga dari berbagai sumber data besar dengan maksud menghasilkan pengetahuan baru [11]. Hasil dari data mining digunakan sebagai opsi solusi untuk memutuskan keputusan yang dapat meningkatkan keputusan di masa depan.

Langkah-langkah penelitian ini mengikuti *Cross-Industry Standard Process for Data Mining* (CRISP-DM), sebuah panduan proses standar untuk melakukan data mining secara umum dalam pemecahan masalah di berbagai bidang bisnis atau unit penelitian [12].



Gambar 1. Metode *Naïve Bayes*

Understanding the Business

Kelulusan mahasiswa adalah faktor penting dalam evaluasi akreditasi sebuah perguruan tinggi. Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) memiliki kriteria khusus untuk menilai lulusan setiap perguruan tinggi, termasuk STMIK Royal Kisaran. Fokus penelitian ini adalah untuk mengklasifikasi mahasiswa berdasarkan apakah mereka lulus tepat waktu atau tidak.

Understanding the Data

Data yang digunakan berasal dari Kampus STMIK Royal Kisaran. Data ini mencakup berbagai atribut seperti NIM, Nama, IPK, Gelombang Wisuda, Jalur Kelulusan, Tahun Masuk, Tahun Lulus, dan Status Kelulusan.

#	Column	Non-Null Count	Dtype
0	NO	608 non-null	int64
1	NIM	608 non-null	int64
2	NAMA	608 non-null	object
3	IPK	608 non-null	float64
4	WISUDA GELOMBANG	608 non-null	int64
5	JALUR KELULUSAN	608 non-null	object
6	TAHUN MASUK	608 non-null	int64
7	TAHUN LULUS	608 non-null	int64
8	LULUS TEPAT WAKTU	608 non-null	object

Gambar 2. Informasi Data

Data Preparation

Beberapa langkah yang perlu diambil dalam mempersiapkan data meliputi pemilihan data yang akan digunakan sebagai kriteria untuk menentukan kelulusan, penggabungan data, dan membersihkan data..

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 608 entries, 0 to 607
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   IPK                    608 non-null    float64
1   WISUDA GELOMBANG      608 non-null    int64
2   JALUR KELULUSAN       608 non-null    int32
3   TAHUN MASUK           608 non-null    int64
4   TAHUN LULUS           608 non-null    int64
5   LULUS TEPAT WAKTU     608 non-null    object
```

Gambar 3. Klasifikasi Data

Modeling

Naïve Bayes adalah algoritma klasifikasi yang memanfaatkan teorema Bayes dan mengasumsikan bahwa nilai antar variabel saling bebas (independen) pada suatu nilai output [13].

Dari 608 data, dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan porsi 70:30.

```
x_train.shape, x_test.shape
((425, 5), (183, 5))
```

Gambar 4. Split Data

Evaluation

Seringkali, alat ini dipakai untuk mengevaluasi seberapa baik pengklasifikasi dapat mengidentifikasi berbagai kelas tupel. Evaluasi menggunakan Confusion Matrix. Confusion matrix adalah sebuah matriks yang terdiri dari kelas negatif dan kelas positif yang dibagi [14].

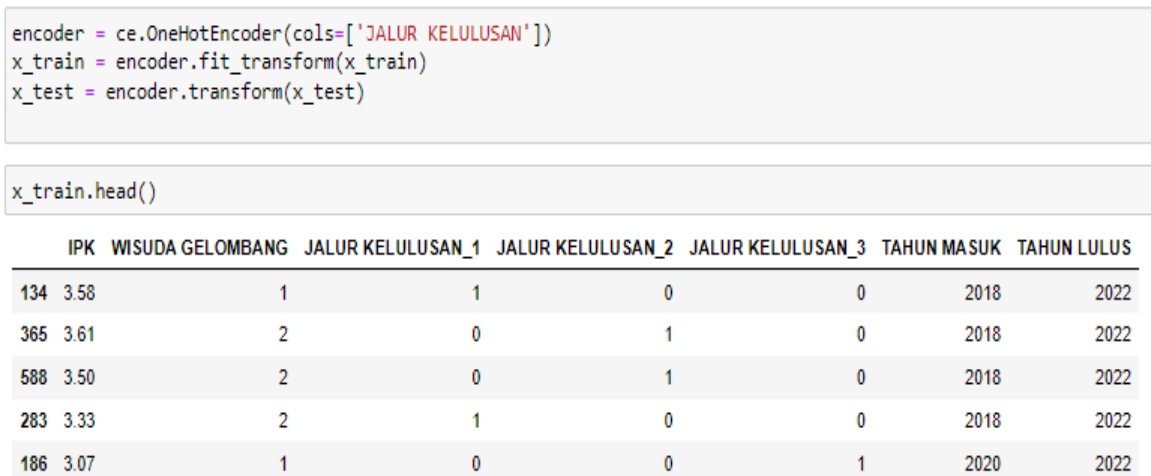
HASIL DAN PEMBAHASAN

Dalam melakukan pengklasifikasian data mahasiswa menggunakan metode *naïve bayes*, tahapan awal untuk analisis yaitu mengimport modul yang diperlukan serta dataset Kelulusan Mahasiswa.

	IPK	WISUDA GELOMBANG	JALUR KELULUSAN	TAHUN MASUK	TAHUN LULUS	LULUS TEPAT WAKTU
0	3.65	1	REGULER	2017	2021	YA
1	3.35	1	REGULER	2017	2021	YA
2	3.61	1	REGULER	2017	2021	YA
3	3.75	1	REGULER	2017	2021	YA
4	3.56	1	REGULER	2017	2021	YA

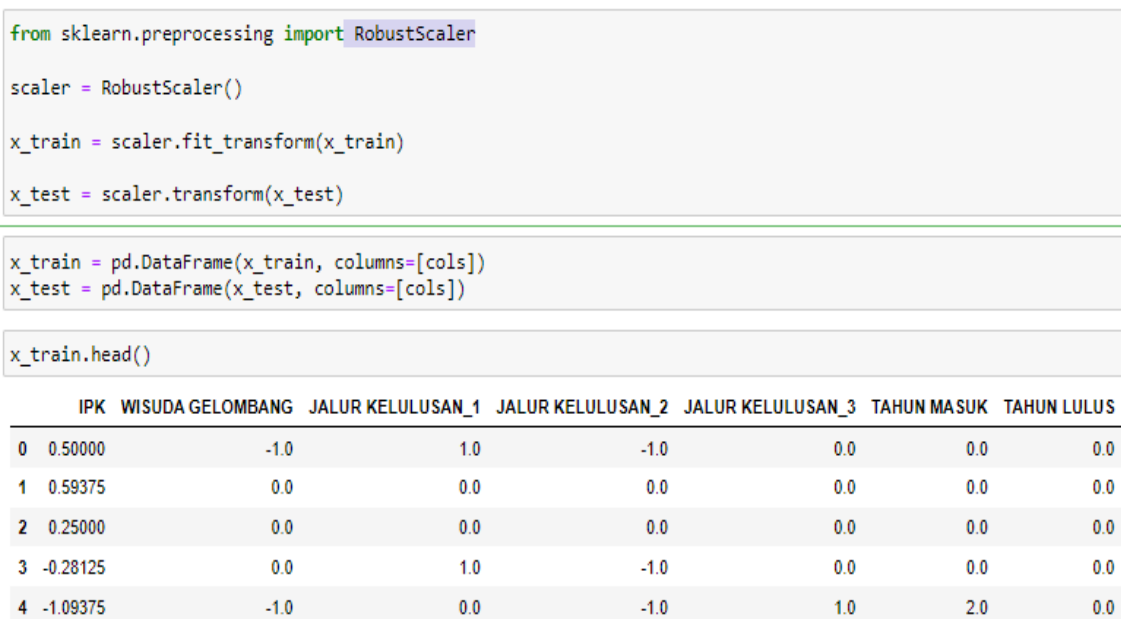
Gambar 5. Dataset Kelulusan Mahasiswa

Untuk mengklasifikasikan data, perlu untuk mengubah data kategorikal menjadi numerik agar dapat membantu memastikan konsistensi dalam model. Dengan menyatakan fitur-fitur kategorikal dalam bentuk numerik dapat menghindari masalah yang mungkin muncul karena label kategori yang berbeda-beda.

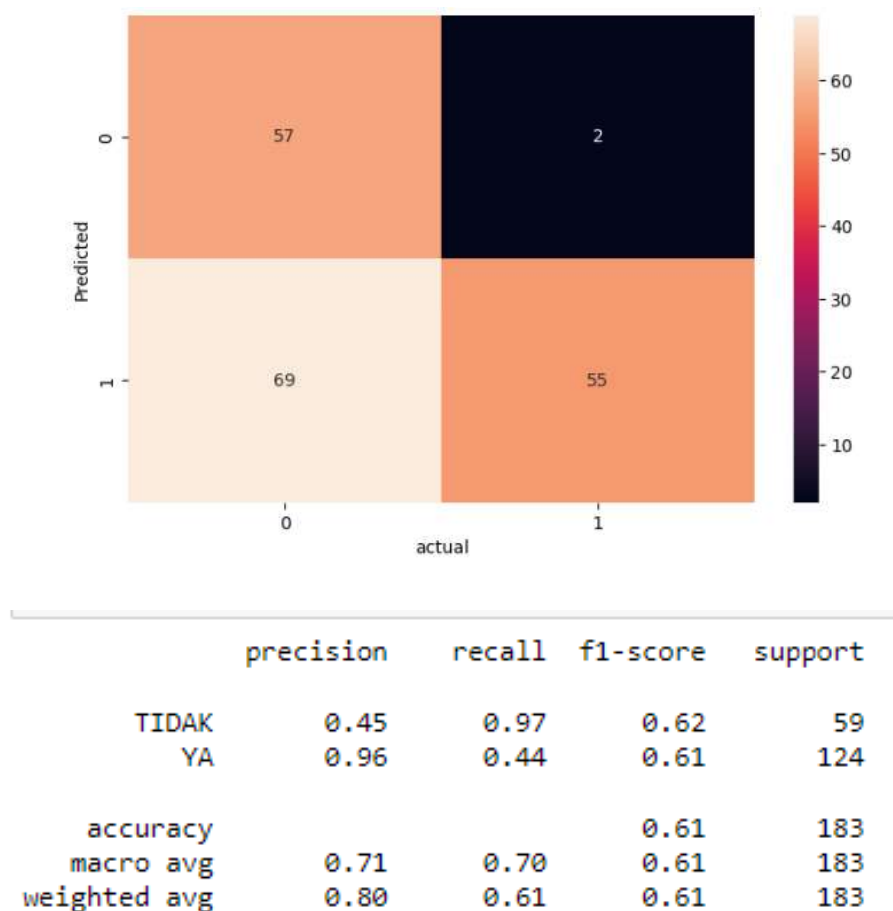


Gambar 6. Encoder Data

Dalam data tersebut terdapat angka-angka yang sangat besar atau kecil yang bisa mempengaruhi hasil prediksi. Untuk membuat data lebih seimbang dan dapat diandalkan, gunakan *RobustScaler* untuk membantu model agar tidak terlalu dipengaruhi oleh nilai-nilai ekstrem dalam data untuk model Naive Bayes.



Gambar 7. RobustScaler



Gambar 8. Tingkat Akurasi

Berdasarkan data yang tergambar pada Gambar 9, algoritma Naïve Bayes menghasilkan hasil sebagai berikut: presisi untuk kelas negatif sebesar 45%, untuk kelas positif sebesar 96%; recall untuk kelas negatif sebesar 97%, untuk kelas positif sebesar 44%; f1-score untuk kelas negatif sebesar 62%, untuk kelas positif sebesar 61%; dan akurasi sebesar 61%.

KESIMPULAN

Hasil analisis dengan menggunakan Algoritma Naïve Bayes menunjukkan tingkat akurasi sebesar 61%. Oleh karena itu, dapat dianggap bahwa metode Naïve Bayes adalah model yang efektif karena mampu memprediksi probabilitas berdasarkan pengalaman historis sebelumnya. Algoritma Naïve Bayes menghasilkan hasil sebagai berikut: presisi untuk kelas negatif sebesar 45%, untuk kelas positif sebesar 96%; recall untuk kelas negatif sebesar 97%, untuk kelas positif sebesar 44%; f1-score untuk kelas negatif sebesar 62%, untuk kelas positif sebesar 61%; dan akurasi sebesar 61%. Kemampuan ini mempermudah proses klasifikasi mahasiswa yang lulus tepat waktu dan yang tidak dengan cara yang cepat dan efisien.

DAFTAR PUSTAKA

- [1] N. Y. Septian, “Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro,” *J. Semant.* 2013, pp. 1–11, 2009.
- [2] M. Zainuddin, “Metode Klasifikasi Berbasis Particle Swarm Optimization (Pso) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa,” *J. Ilm. Teknol. Inf. Asia*, vol. 13, no. 1, p. 1, 2018, doi: 10.32815/jitika.v13i1.247.
- [3] N. Khasanah and A. Salim, “Rachman Komarudin 4) , Yana Iqbal Maulana 5) 1) Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 2,3) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika 4) Sistem Informasi, Fakultas Teknologi Inf,” *Fak. Teknol. Inf.*, vol. 13, no. 3, pp. 207–214, 2022.
- [4] Y. Kusnia, “Pengaruh Karakteristik Gender dan Motivasi Belajar Terhadap Prestasi Belajar Matematika Siswa Kelas X IPA 1 di MAN 2 Semarang,” *Semin. Nas. Pendidikan, Sains dan Teknol.*, pp. 398–405, 2017.
- [5] Y. Apridiansyah, N. D. M. Veronika, and E. D. Putra, “Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes,” *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, pp. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
- [6] H. Bhavsar and A. Ganatra, “A Comparative Study of Training Algorithms for Supervised Machine Learning,” *Int. J. Soft Comput. Eng.*, vol. 2, no. 4, pp. 74–81, 2012.
- [7] J. J. Aripin, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi pada BPR Pantura,” 2019.
- [8] W. Dwi Septiani, “Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes untuk Prediksi Penyakit Hepatitis,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.
- [9] C. F. Chien, W. C. Wang, and J. C. Cheng, “Data mining for yield enhancement in semiconductor manufacturing and an empirical study,” *Expert Syst. Appl.*, vol. 33, no. 1, pp. 192–198, 2007, doi: 10.1016/j.eswa.2006.04.014.
- [10] N. E. Putria, “Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel,” *Comput. Based Inf. Syst. J.*, vol. 6, no. 1, pp. 29–39, 2018.
- [11] “JUTSI (Jurnal Teknologi dan Sistem Informasi) BANGUNAN TERLARIS DENGAN MENGGUNAKAN METODE K-MEANS DI UD MAJU BERSAMA Mahasiswa Prodi Sistem Informasi , STMIK Royal Dosen Prodi Sistem Informasi , STMIK Royal JUTSI (Jurnal Teknologi dan Sistem Informasi ,” vol. 2, no. 2, pp. 93–102, 2022.
- [12] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, “Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa,” *Semin. Nas. SISFOTEK*, pp. 303–308, 2023.
- [13] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020,

doi: 10.24076/citec.2019v6i1.178.

- [14] L. A. Utami, “Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine dan K-Nearest Neighbor Berbasis Particle Swarm Optimization,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 103–112, 2017.