

Implementasi Algoritma Naïve Bayes untuk Klasifikasi Pemahaman Program MBKM bagi Mahasiswa

Ajif Yunizar Pratama Yusuf¹, Rafika Sari^{1,*}

¹ Informatika, Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan No.81 Margamulya, Kota Bekasi, Jawa Barat, Telp. (021) 7231948; e-mail: ajif.yunizar@dsn.ubharajaya.ac.id, rafika.sari@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: rafika.sari@dsn.ubharajaya.ac.id

Diterima: 12 Des 2022; Review: 29 Des 2022; Disetujui: 31 Des 2022; Diterbitkan: 31 Des 2022

Abstract

Outcome based education is the basis for the Informatics Study Program at Bhayangkara Jakarta Raya University (Ubhara Jaya) in formulating graduate profiles. Research to identify student knowledge related to MBKM. The analytical approach ranks 25 questions in the MBKM program which are the focus of students and lecturers. Through the survey conducted, it will provide an overview of whether the MBKM program has an attraction for students to develop their competencies, skills and soft skills as a provision for future graduates. In addition, do students and lecturers assess the MBKM program as having significant benefits in improving student abilities? The research was conducted by implementing data mining survey results of students and lecturers in depth. The method used is a classification learning ensemble. The classification process begins with collecting available data, preprocessing data in the form of feature selection, cleaning, integration, and transformation; continued with the process of making models on training data by applying 10-fold cross validation and the assembly learning method to handle imbalance classes; and finally evaluation of modeling results on data testing. The output target of this research is a policy recommendation for implementing the MBKM program for the Ubhara Jaya Informatics study program.

Keywords: Classification, *Data mining*, MBKM Program, Students, Survey

Abstrak

Outcome based education menjadi dasar pemikiran Program Studi Informatika Universitas Bhayangkara Jakarta Raya (Ubhara Jaya) dalam merumuskan profil lulusan. Penelitian untuk mengidentifikasi pengetahuan mahasiswa terkait MBKM. Pendekatan analitis meranking 25 pertanyaan program MBKM yang menjadi fokus mahasiswa dan dosen. Melalui survei yang dilakukan akan memberikan gambaran apakah program MBKM memiliki daya tarik bagi mahasiswa untuk mengembangkan kompetensi, keterampilan, dan *softskillnya* sebagai bekal bagi lulusan masa depan. Selain itu, apakah mahasiswa dan dosen menilai program MBKM memiliki manfaat yang signifikan dalam meningkatkan kemampuan mahasiswa. Penelitian dilakukan dengan mengimplementasikan *data mining* hasil survei mahasiswa dan dosen secara mendalam. Metode yang digunakan yaitu ensemble learning klasifikasi. Proses klasifikasi diawali dengan pengumpulan data yang sudah tersedia, preprocessing data berupa *feature selection*, *cleaning*, integrasi, dan transformasi; dilanjutkan proses pembuatan model pada data training dengan menerapkan *10-fold cross validation* serta metode *ensemble learning* untuk menangani imbalance class; dan terakhir evaluasi hasil pemodelan pada data testing. Target luaran dari penelitian ini yaitu suatu rekomendasi kebijakan penerapan program MBKM bagi program studi Informatika Ubhara Jaya.

Kata kunci: Data Maining, Klasifikasi, Mahasiswa, Program MBKM, Survei

1. Pendahuluan

Kebijakan baru dari Kementerian Pendidikan dan Kebudayaan (Kemendikbud) pada Perguruan Tinggi guna menghasilkan lulusan yang berkompeten dengan dunia kerja, industri serta perubahan pada masa depan, maka diberlakukan kebijakan baru melalui program Merdeka Belajar Kampus Merdeka (MBKM) (Direktorat Jenderal Pendidikan Tinggi Kemendikbud RI, 2020). Dengan adanya beberapa kategori program serta kriteria persyaratan yang dibutuhkan untuk dipenuhi oleh mahasiswa. Penerapan konsep baru Pendidikan ini masih banyak yang belum mengetahuinya, khususnya bagi para civitas academia kampus. Maka dari itu diperlukan survei terkait pengetahuan tentang program MBKM ini terhadap semua pelaku Pendidikan. Data hasil survei tersebut harus diolah agar menghasilkan informasi yang bernilai, khususnya bagi pemangku kebijakan di perguruan tinggi.

Data mining merupakan suatu penyelesaian masalah dengan melakukan analisis terhadap data yang disajikan dalam *database* (Sunhare et al., 2022). Selain itu *data mining* juga digunakan untuk mengetahui pola data, dimana setiap pola memiliki karakteristik masing-masing yang dapat memberikan informasi penting dari data tersebut. *Data mining* dapat diartikan sebagai berbagai macam cabang ilmu pengetahuan yang menjadi satu, terdiri atas sistem basis data, statistika, *machine learning*, *visualization*, dan informasi pengetahuan. *Data mining* telah berhasil diterapkan di berbagai bidang ilmu seperti bisnis, bioinformatika, genetika, kedokteran, pendidikan dan lain sebagainya (Annur, 2018).

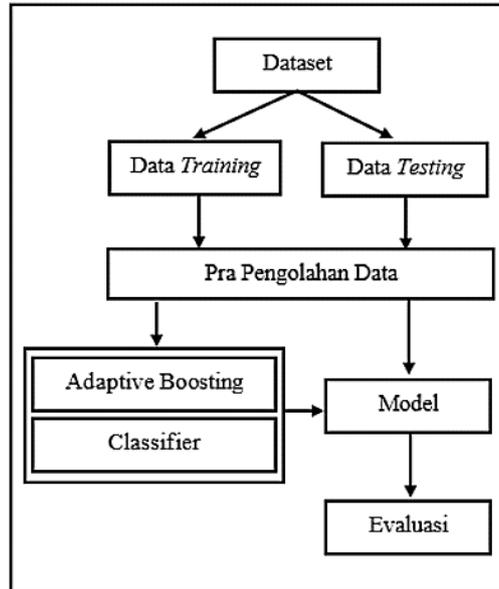
Klasifikasi merupakan proses untuk menemukan sebuah model atau pola yang dapat menggambarkan serta membedakan kelas pada suatu *Dataset*. Maka dari itu, penelitian ini mengidentifikasi pengetahuan pelaksanaan MBKM. Penelitian meranking 25 pertanyaan terkait MBKM yang menjadi fokus mahasiswa dan dosen. Berdasarkan 25 pertanyaan survei prioritas dilakukan analisis deskriptif dari hasil survei. Melalui survei yang dilakukan akan memberikan gambaran apakah program MBKM memiliki daya tarik bagi mahasiswa untuk mengembangkan kompetensi, keterampilan, dan *softskillnya* sebagai bekal bagi lulusan masa depan.

Penelitian tentang kasifikasi data telah banyak dilakukan, penelitian terkait yang dijadikan rujukan yaitu: a) Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors and Decision Tree (Rozaq et al., 2022), b) Analisis Kinerja Algoritma Naïve Bayes pada *Dataset* Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah (Susanto et al., 2022), c) Implementasi Profile Match Up Untuk Administrasi Pelaksanaan MBKM – Studi Kasus: Program Studi Sistem Informasi Universitas Kristen Duta Wacana (Simangunsong, 2021), d) Sistem Rekomendasi Tugas Akhir Mahasiswa pada AMIK Indonesia untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF) (Salam et al., 2022), d) Identifikasi Pengetahuan Mahasiswa Teknik Industri terhadap Implementasi Program MBKM dengan pendekatan Analytical Network Process (Nurhasanah et al., 2022).

2. Metode Penelitian

2.1. Tahapan Penelitian

Alur yang digunakan pada penelitian ini ditunjukkan pada Gambar 1 yang merupakan diagram dari langkah-langkah penelitian.



Sumber: Hasil Penelitian (2022)

Gambar 1. Diagram Alir Tahapan Penelitian

a. Dataset

Dataset yang digunakan pada penelitian ini adalah hasil survey MBKM terhadap mahasiswa Universitas Bhayangkara Jakarta Raya yang dilakukan serentak oleh Kemdikbud untuk semua perguruan tinggi di Indonesia sumber data dapat diperoleh dari link: <https://colab.research.google.com/drive/1rMH8-MzV5SaGWYSeFo7QcHiTuJcOfJag2>. *Dataset* tersebut merupakan *Dataset* publik yang dishare dari Kemdikbud tahun 2021. *Data User Knowledge Modeling Dataset* memiliki 403 *instance*, 5 *attributes* fitur dan 1 *attribute class*. Tabel 1 merupakan potongan data *User Knowledge Modeling Dataset* yang digunakan pada penelitian.

Tabel 1. *Data User Knowledge Modeling Dataset*

STG	SCG	STR	LPR	PEG	UNS (Class)
0	0,1	0,5	0,26	0,05	<i>Very Low</i>
0,05	0,05	0,55	0,6	0,14	<i>Low</i>
0,08	0,18	0,63	0,6	0,85	<i>High</i>
0,2	0,2	0,68	0,67	0,85	<i>High</i>
0,22	0,22	0,9	0,3	0,9	<i>High</i>
0,14	0,14	0,7	0,5	0,3	<i>Low</i>
0,16	0,16	0,8	0,5	0,5	<i>Middle</i>

Sumber: Hasil Penelitian (2022)

Keterangan:

STG (The degree of study time for goal object materials).

SCG (The degree of repetition number of user for goal object materials)

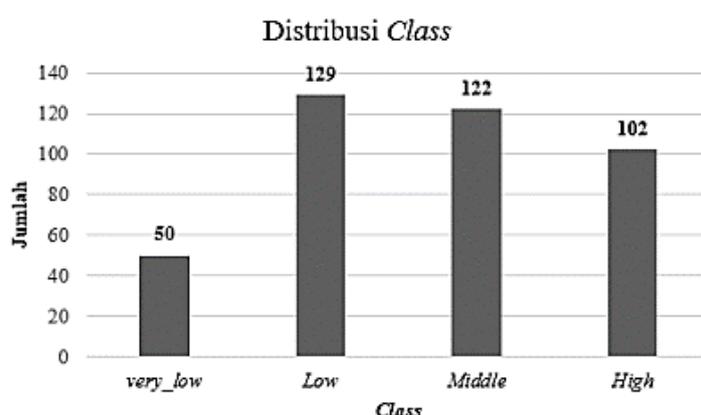
STR (The degree of study time of user for related objects with goal object)

LPR (The exam performance of user for related objects with goal object)

PEG (The exam performance of user for goal objects)

UNS (The knowledge level of user) merupakan *attribute class* pada *Dataset* tersebut.

Dataset User Knowledge Modeling memiliki 5 *attributes* bertipe *numeric*, dan 1 *attribute* kelas dengan jumlah kelas sebanyak empat yaitu (*very low*, *low*, *middle*, dan *high*). Selain itu *Dataset User Knowledge Modeling* memiliki distribusi kelas yang tidak seimbang. Gambar 2 merupakan ilustrasi distribusi kelas pada *Dataset* tersebut.



Sumber: Hasil Penelitian (2022)

Gambar 2. Ilustrasi Distribusi Kelas pada *Dataset*

Gambar 2 menunjukkan adanya distribusi kelas yang tidak seimbang dimana kelas *Very Low* sebanyak 50 *instance*, *Low* sebanyak 129 *instance*, *Middle* sebanyak 122 *instance* dan *High* sebanyak 130 *instance*, secara teori hal ini dapat mempengaruhi kinerja dari suatu algoritma klasifikasi.

b. Pra Pengolahan Data

Pada Gambar 2 menunjukkan adanya distribusi kelas yang tidak seimbang dimana kelas *Very Low* sebanyak 50 *instance*, *Low* sebanyak 129 *instance*, *Middle* sebanyak 122 *instance* dan *High* sebanyak 130 *instance*, secara teori hal ini dapat mempengaruhi kinerja dari suatu algoritma klasifikasi.

c. Klasifikasi dan Model

Pada tahap ini algoritma klasifikasi yang akan digunakan ialah *Naive Bayes*.

d. Evaluasi Kinerja

Pada tahap ini algoritma klasifikasi yang akan digunakan ialah *Naive Bayes*, menggunakan persamaan 1.

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
 Sensitivity &= \frac{TP}{TP+FN} \\
 Specificity &= \frac{TN}{TN+FP}
 \end{aligned}
 \tag{1}$$

$$G - Mean = \sqrt{Sensitivity * Specificity}$$

Keterangan:

True Positive adalah jumlah record positif yang diklasifikasikan sebagai positif.

False positif adalah jumlah record negatif yang diklasifikasikan sebagai positif.

False negatif adalah jumlah record positif yang diklasifikasikan sebagai negatif.

True negatif adalah jumlah record negatif yang diklasifikasikan sebagai negatif.

2.2. Algoritma Naïve Bayes

Naive Bayes adalah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan (Undamayanti et al., 2022). Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau non-interdependen diberikan oleh nilai variabel kelas. Algoritma Naïve Bayes adalah metode probabilistik yang digunakan untuk mengklasifikasikan kelas dari suatu data. Garis Besar Metode Naïve Bayes adalah suatu analisis statistik dimana probabilitas awal (*prior probability*) diestimasi dari data training. Untuk setiap parameter probabilitas ditentukan berdasarkan probabilitas awal. Secara matematis algoritma ini dapat dituliskan seperti pada persamaan 2.

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{2}$$

Dimana:

$P(H|E)$: Probabilitas posisional dari probabilitas (probabilitas bersyarat)

$P(E|H)$: Probabilitas parameter E pada hipotesis H

$P(H)$: Probabilitas sebelumnya (sebelumnya) hipotesis H

$P(E)$: Probabilitas awal (sebelumnya) parameter E

Keuntungan *Naive Bayesian*:

- a. Menangani data kuantitatif dan diskrit.
- b. Hanya membutuhkan sedikit data pelatihan untuk mengestimasi parameter (rata-rata dan varian variabel) yang diperlukan untuk klasifikasi.
- c. Tangani nilai yang hilang dengan mengabaikan agensi selama perkiraan perhitungan peluang.
- d. Efisiensi cepat dan ruang.
- e. Kuat terhadap atribut yang tidak relevan Kekurangan *Naive Bayesian* : 1) Tidak berlaku jika probabilitas bersyarat adalah nol, jika nol maka probabilitas yang diprediksi akan menjadi nol juga, 2) Asumsikan variabel independen.

2.3. Pembobotan Nilai

Pada setiap respon mahasiswa diberikan bobot nilai dengan menggunakan Skala Likert (Budiaji, 2013) dimana dengan bobot tiap jawaban akan memberikan data numerik yang kemudian akan diproses dengan menggunakan Metode *Naïve Bayes*. Pada proses pembobotan nilai akan dipilih pertanyaan-pertanyaan yang memiliki pilihan variasi jawaban. Berikut pertanyaan-pertanyaan yang diberikan bobot nilai dengan Skala Likert.

Tabel 2. Pembobotan Nilai

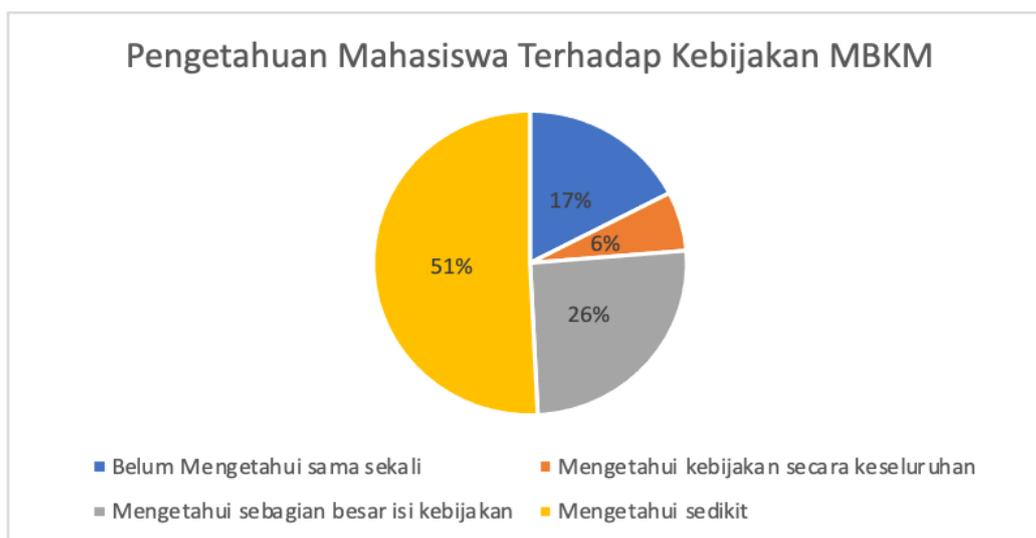
No.	Pertanyaan	Pilihan Jawaban	Skala Likert
1	Apakah dokumen kurikulum, panduan dan prosedur operasional untuk mengikuti kegiatan MBKM sudah ada pada program studi saudara?	Tidak Tahu	0
		Belum	1
		Sudah	2
2	Apakah Saudara sudah menyiapkan diri untuk menjadi bagian dalam kegiatan MBKM?	Tidak Berminat	0
		Belum	1
		Sudah	2
3	Menurut Saudara, apakah kegiatan pembelajaran di luar program studi akan berimplikasi pada masa studi?	Tidak Tahu	0
		Masa studi menjadi lama	1
		Tetap tepat waktu	2
4	Menurut Saudara, apakah kegiatan pembelajaran di luar kampus akan memberikan kompetensi tambahan seperti keterampilan dalam menyelesaikan permasalahan nyata yang kompleks, keterampilan dalam menganalisis, etika profesi, dll?	Tidak Tahu	0
		Mungkin	1
		Ya	2
5	Menurut Saudara, belajar di program studi lain akan memperluas perspektif dan memberikan kompetensi tambahan yang dibutuhkan?	Tidak Tahu	0
		Mungkin	1
		Ya	2

Sumber: Hasil Penelitian (2022)

3. Hasil dan Pembahasan

Hasil penelitian diawali dengan menunjukkan pengetahuan mahasiswa di Universitas Bhayangkara yakni sebanyak 51 persen mengetahui sedikit, 26 persen mahasiswa mengetahui sebagian besar isi kebijakan secara keseluruhan, sebanyak 17 persen mahasiswa belum mengetahui sama sekali, serta sebanyak 6 persen mahasiswa mengetahui kebijakan secara keseluruhan. Hasil ini sejalan dengan penelitian yang dilakukan oleh (Laga et al., 2021). Hal ini

dikarenakan oleh ketidakmerataan informasi yang menjadi penghambat upaya peningkatan pemahaman mahasiswa secara keseluruhan tentang kebijakan Merdeka Belajar Kampus Merdeka (MBKM) di Universitas Bhayangkara Jakarta Raya.

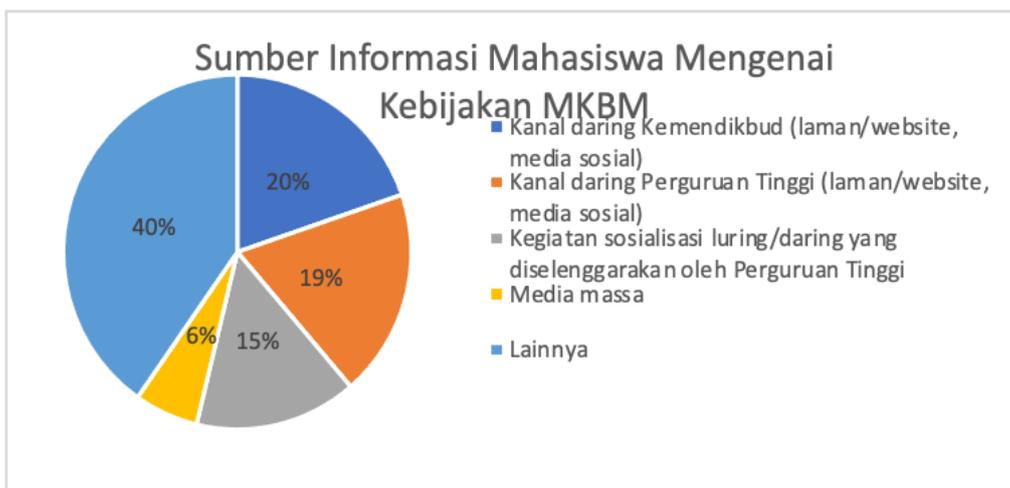


Sumber: Hasil Penelitian (2022)

Gambar 3. Persepsi Mahasiswa Terhadap Kebijakan MBKM

Ketimpangan pemahaman ini juga berdampak pada informasi penerapan semester yang disetarakan, banyak responden yang belum tahu informasi tersebut yang seharusnya 4 (empat) semester sesuai dengan ketentuan peraturan di tingkat Universitas, berdasarkan data hanya 5,3 persen mahasiswa yang menjawab 4 (empat), sedangkan 44,77 persen menjawab 3 (tiga) semester dan 33 persen menjawab 2 (dua) semester serta 16,93 persen. Penyetaraan Kebijakan MBKM pada hasil survei terdapat 35,67 persen mahasiswa yang menjawab sesuai dengan ketentuan yakni 40 sks.

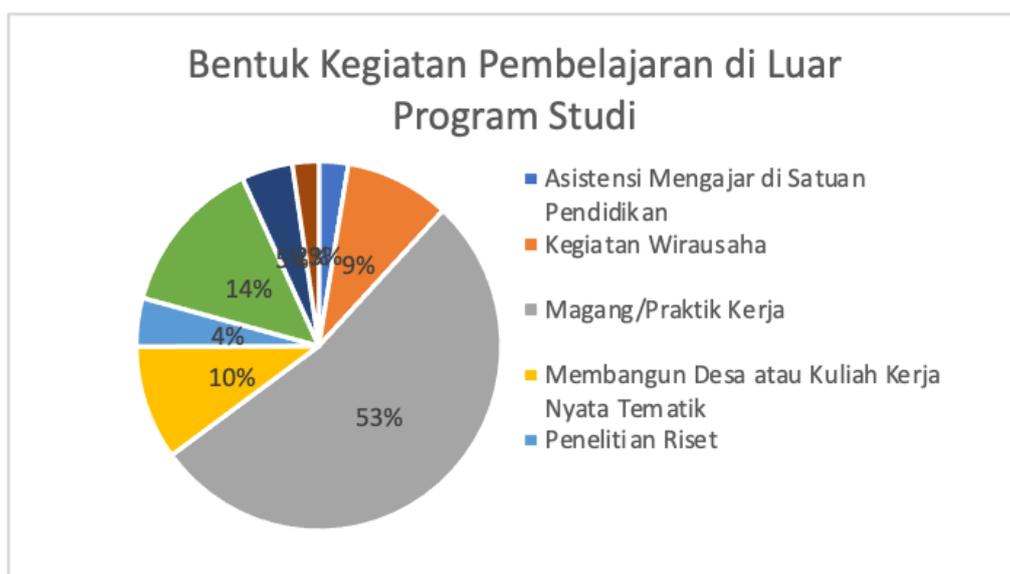
Sebagaimana terlihat pada Gambar 4 sumber informasi yang diperoleh mahasiswa Program Studi Manajemen terdapat berkaitan dengan kegiatan Merdeka Belajar Kampus Merdeka, terdapat 20 persen dari Kanal daring Kemendikbud (laman/website, media sosial), 19 persen dari Kanal daring Perguruan Tinggi (laman/website, media sosial), sebanyak 15 persen diperoleh dari Kegiatan sosialisasi luring/daring yang diselenggarakan oleh Perguruan Tinggi, sebanyak 6 persen dari Media Massa dan sebanyak 40 persen yang menjawab Lainnya. Sehingga dari respon mahasiswa tersebut dapat terlihat adanya pengaruh yang sangat signifikan terhadap kegiatan sosialisasi baik secara luring maupun daring yang diselenggarakan oleh Perguruan Tinggi sangat membantu dalam memahami isi kebijakan kegiatan MBKM. Adapun hasil survei membuktikan Kanal Daring baik oleh Kemendikbud ataupun Perguruan Tinggi masih menjadi sumber informasi teratas dalam mengetahui kegiatan MBKM oleh mahasiswa.



Sumber: Hasil Penelitian (2022)

Gambar 4. Sumber Informasi Mahasiswa Mengenai Kebijakan MBKM

Selanjutnya menganalisis mengenai respon mahasiswa pada 8 bentuk kegiatan pembelajaran diluar Program Studi mana yang menjadi favorit untuk dipilih. Dari Gambar 5 terlihat mahasiswa di Universitas Bhayangkara Jakarta Raya sangat menyukai kegiatan Magang/Praktik Kerja yakni sebanyak 53 persen. Selebihnya dari tujuh kegiatan lain masih kurang adanya minat mahasiswa. Ini sejalan juga dengan penelitian oleh (Undamayanti et al., 2022).



Sumber: Hasil Penelitian (2022)

Gambar 5. Bentuk Kegiatan Pembelajaran di Luar Program Studi

4. Kesimpulan

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma klasifikasi *Naïve Bayes* memiliki performa yang baik dalam menentukan periode prediksi penelitian. Studi ini menunjukkan bahwa *Naïve Bayes* mampu mengklasifikasikan pengujian data dengan benar.

Penentuan data latih dan data uji dapat mempengaruhi hasil uji, karena pola data latih akan digunakan sebagai aturan untuk menentukan kelas pada data uji. Dapat disimpulkan dari hasil pengujian algoritma *Naïve Bayes* memiliki rata-rata akurasi sebesar 86,16% dan *error* sebesar 13,84%. Dari analisis berdasarkan data tersebut dapat digunakan oleh institusi Universitas untuk lebih meningkatkan sosialisasi program MBKM terpusat dari Kemdikbud agar kuota penerimaan mahasiswa yang lolos seleksi program MBKM tiap semesternya meningkat yang nantinya akan berkaitan dengan pencapaian Capaian Pembelajaran Lulusan (CPL) program studi dan berkontribusi terhadap kualitas penilaian akreditasi program studi di Universitas.

Daftar Pustaka

- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode *Naïve Bayes*. *ILKOM Jurnal Ilmiah*, 10(2), 160–165.
- Budiaji, W. (2013). The Measurement Scale and The Number of Responses in Likert Scale. *Jurnal Ilmu Pertanian dan Perikanan Desember*, 2(2), 127–133.
- Direktorat Jenderal Pendidikan Tinggi Kemendikbud RI. (2020). *Buku Panduan Merdeka Belajar-Kampus Merdeka* (pp. 1–33). <http://dikti.kemdikbud.go.id/wp-content/uploads/2020/04/Buku-Panduan-Merdeka-Belajar-Kampus-Merdeka-2020>
- Laga, Y., Nona, R. V., Langga, L., & Jamu, M. E. (2021). Persepsi Mahasiswa Terhadap Kebijakan Merdeka Belajar Kampus Merdeka (MBKM). *Edukatif: Jurnal Ilmu Pendidikan*, 4(1), 699–706. <https://doi.org/10.31004/edukatif.v4i1.1951>
- Nurhasanah, N., Aribowo, B., Purwandari, A. T., Sumantri, D., Maulana, S., Yasmin, M. A., Raghdawulan, R., & Shity, C. (2022). Identifikasi Pengetahuan Mahasiswa Teknik Industri terhadap Implementasi Program MBKM dengan pendekatan Analytical Network Process. *Urnal Al-Azhar Indonesia Seri Sains Dan Teknologi*, 7(2), 72. <https://doi.org/10.36722/sst.v7i2.1016>
- Rozaq, A., Yunitasari, Y., Sussolaikah, K., Sari, E. R. N., & Syahputra, R. I. (2022). Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors dan Decision Tree. *Jurnal Media Informatika Budidarma*, 6(2), 746. <https://doi.org/10.30865/mib.v6i2.3554>
- Salam, A., Albahri, F. P., & Fathurrahmad. (2022). Sistem Rekomendasi Tugas Akhir Mahasiswa pada AMIK Indonesia untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF). *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(2), 281–288. <https://doi.org/10.35870/jtik.v6i2.420>
- Simangunsong, T. M. (2021). *Implementasi Profile Match Up Untuk Administrasi Pelaksanaan MBKM - Studi Kasus: Program Studi Sistem Informasi Universitas Kristen Duta Wacana*. Universitas Kristen Duta Wacana.
- Sunhare, P., Chowdhary, R. R., & Chattopadhyay, M. K. (2022). Internet of things and *Data mining*: an Application Oriented Survey. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 3569–3590. <https://doi.org/10.1016/j.jksuci.2020.07.002>

- Susanto, E. B., Christianto, P. A., Maulana, M. R., & Binabar, S. W. (2022). Analisis Kinerja Algoritma Naïve Bayes pada *Dataset* Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah. *Jurnal Computer Science and Information Technology (CoSciTech)*, 3(3), 234–241. <https://doi.org/10.37859/coscitech.v3i3.4343>
- Undamayanti, E., Hermanto, T. I., & Kaniawulan, I. (2022). Analisis Sentimen Menggunakan Metode *Naive Bayes* Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(2), 916–930. <https://doi.org/10.30645/j-sakti.v6i2.502>