

KLASIFIKASI KELAYAKAN MAHASISWA MASUK PROGRAM MSIB KAMPUS MERDEKA: SYSTEM LITERATURE REVIEW

Kurnia Hayati, Roni Habibi

Program Studi DIV Teknik Informatika, Sekolah Vokasi
Universitas Logistik dan Bisnis Internasional, Jalan Sariasih No.54, Bandung, Indonesia
kurniahayati907@gmail.com, roni.habibi@ulbi.ac.id

ABSTRAK

Kampus merdeka adalah suatu program diberlakukan oleh Menteri Nadiem Makarin untuk mengoptimalkan eksistensi dari perguruan tinggi agar ketegangan pendidikan tinggi. Berbagai jenis programnya, seperti magang. Banyak mahasiswa yang melamar untuk program magang maupun yang lainnya, namun banyak yang tidak memenuhi kualifikasi. Menggunakan pendekatan Systematic Literature Review, klasifikasi adalah salah satu pendekatan untuk mengatasi masalah ini. Penelitian ini menggunakan bantuan tools publish or perish, terdapat 18 artikel yang termasuk termasuk kriteria inklusif dan eksklusif. Sumber dataset yang banyak digunakan berasal dari USA dengan dataset Kaggle. Akurasi, presisi, recall, dan F1 semuanya ditentukan melalui evaluasi kinerja. Dengan nilai akurasi sebesar 98,82 persen, model SVM dan metode machine learning memiliki akurasi tertinggi menurut temuan penelitian ini.

Kata kunci: Klasifikasi, Machine Learning, Kuesioner, Kampus Merdeka, System Literature Review

1. PENDAHULUAN

Kampus merdeka merupakan program yang diberlakukan oleh Menteri Nadiem Makarin untuk mengoptimalkan eksistensi dari perguruan tinggi agar kesenjangan pendidikan tinggi, khususnya pada aspek kualitas serta pemerataannya dapat terselesaikan [1]. Kebijakan Kampus Merdeka artinya mengusahakan kemampuan lulusan, baik *soft skill* maupun *hard skill*, sehingga siap dan gigih terhadap kebutuhannya. [2]. Berbagai jenis program Kampus Merdeka, seperti magang, proyek pengabdian masyarakat, mengajar di unit sekolah, pertukaran pelajar, penelitian, kewirausahaan, studi/proyek akademik, dan berpartisipasi dalam layanan kemanusiaan [3]. Banyak sekali mahasiswa yang *apply* untuk program magang maupun *study independent*. Namun banyak juga yang tidak memenuhi kualifikasi. Klasifikasi adalah salah satu solusi untuk masalah tersebut. Dengan adanya mahasiswa yang mendaftar ini, peneliti ingin melakukan klasifikasi untuk mahasiswa yang layak masuk ke program MSIB Kampus Merdeka.

Suatu kelas data dideskripsikan dan dibedakan melalui klasifikasi ini [4]. Jenis algoritma klasifikasi yang digunakan secara luas secara keseluruhan adalah *Decision trees*, *Naïve Bayes classifiers*, *k-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines* (SVM) [5]. Beberapa fase dalam urutannya adalah pengembangan model, aplikasi model, dan penilaian [6]. Klasifikasi memanfaatkan pembelajaran mesin, yang termasuk dalam kategori pembelajaran terawasi [7]. Algoritma untuk mengoptimalkan fungsi yang memetakan input ke output yang diinginkan dikenal dengan pembelajaran diinginkan [8]. Metode klasifikasi

performa dibandingkan menggunakan akurasi, presisi, *recall*, dan *f-measure* atau *f1* sebagai ukuran *f1* [9]. *Sytem Literature Review* (SLR) merupakan metode penelitian yang digunakan untuk mengumpulkan serta menilai penelitian yang berhubungan dengan topik tertentu [10]. *Sytem Literature Review* (SLR) digunakan untuk mengidentifikasi yang bertujuan menjawab *question* peneliti, yaitu metode yang tepat untuk melakukan klasifikasi.

2. TINJAUAN PUSTAKA

2.1. Research Question

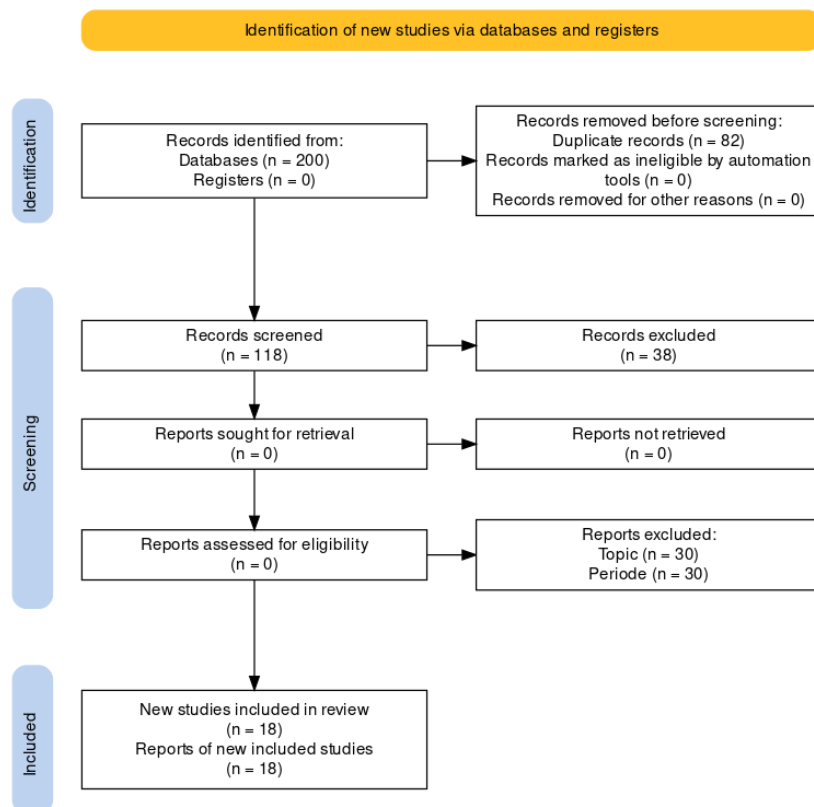
Tabel 1. Research Question

ID	Research Question	Motivation
RQ1	Jurnal mana yang paling banyak berhubungan dengan klasifikasi?	Identifikasi jurnal yang signifikan pada klasifikasi
RQ2	Model pembelajaran mesin apa yang tepat untuk digunakan dalam klasifikasi?	Identifikasi performa model pembelajaran yang tepat

2.2. Searching for the Literature

Penelitian ini dilakukan pada bulan Mei 2023 di Bandung. Artikel atau jurnal dicari dengan menggunakan tools, yaitu *publish or perish* untuk membantu dalam pencarian. *Publish or perish* adalah sebuah *software* yang banyak digunakan oleh berbagai peneliti untuk mendapatkan informasi terkait topik tertentu yang diinginkan dengan mengambil sumber database dari scopus. Pencarian artikel dengan topik yang terkait dengan menggunakan kata kunci: *Classification*, *Algorithm*, *Naive Bayes*.

3. METODE PENELITIAN



Gambar 1. Prisma Diagram System Literature Review

Dari pencarian menggunakan *publish or perish* mendapatkan 200 artikel. Pemilihan kriteria artikel ada dua, yaitu kriteria inklusi merupakan kriteria yang disesuaikan, serta kriteria eksklusi adalah kriteria yang tidak termasuk atau dikecualikan. Kriteria inklusi dari penelitian ini adalah rentang waktu maksimal dari tahun 2017, lalu berhubungan dengan klasifikasi, dan menggunakan metode. Kriteria eksklusi dari penelitian ini adalah artikel yang diterbitkan sebelum tahun 2017 dan artikel yang tidak memiliki hasil. Dari hasil pencarian terdapat 82 artikel duplikat dan 39 artikel yang tidak termasuk. Selanjutnya terdapat 30 artikel yang tidak sesuai dengan topik serta 30 artikel yang tidak masuk ke periode. Setelah melalui tahapan identification dan screening, artikel yang termasuk adalah sebanyak 18 artikel.

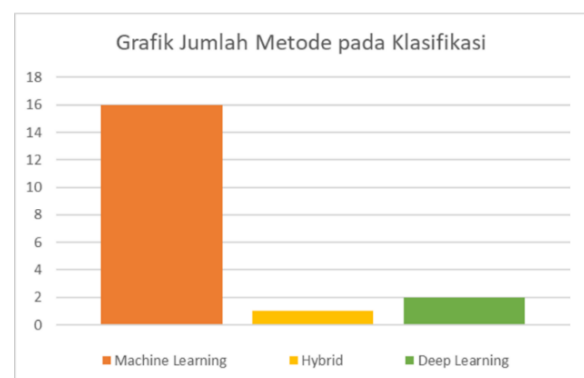
4. HASIL DAN PEMBAHASAN

Hasil dan pembahasan yang menggunakan *System Literature Review* ini, terdapat 18 artikel yang terpilih sebelumnya akan menjelaskan artikel yang didapat terkait metode yang digunakan, model pembelajaran, tahun artikel, dataset yang didapat, dan evaluasi kinerja dengan pengujian akurasi, presisi, *recall*, dan *f1*.

3.1. Metode yang digunakan pada Klasifikasi

Dari artikel yang telah diamati, metode yang paling banyak digunakan pada klasifikasi adalah

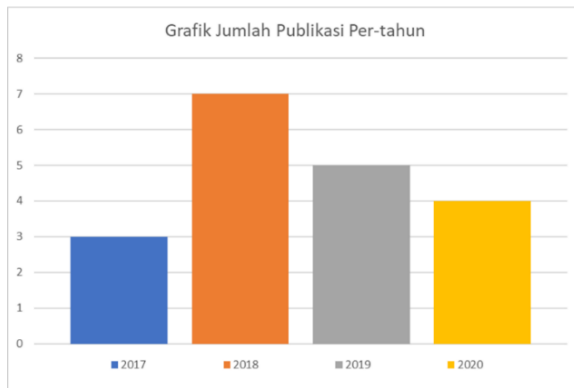
machine learning. Berdasarkan beberapa penelitian yang menggunakan *machine learning* dapat menganalisis dan memproses data yang lebih banyak dan rumit dengan waktu yang singkat. Seperti yang terlihat pada gambar 2, *machine learning* lebih banyak digunakan dari keseluruhan artikel.



Gambar 2. Grafik Jumlah Metode yang digunakan pada Klasifikasi

3.2. Tahun Artikel pada Klasifikasi

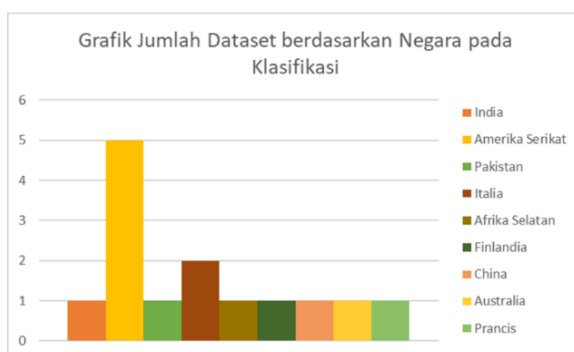
Berdasarkan artikel yang telah diamati, tahun artikel yang paling banyak dari 2017 hingga 2020 didominasi tahun 2018 dengan jumlah 7 artikel. Selanjutnya untuk tahun yang lain seperti tahun 2017 berjumlah 3 artikel, tahun 2019 sebanyak 5 artikel, dan yang terakhir tahun 2020 sebanyak 4 artikel.



Gambar 3. Grafik Jumlah Publikasi Artikel Per-tahun

3.3. Sumber Dataset Berdasarkan Negara pada Klasifikasi

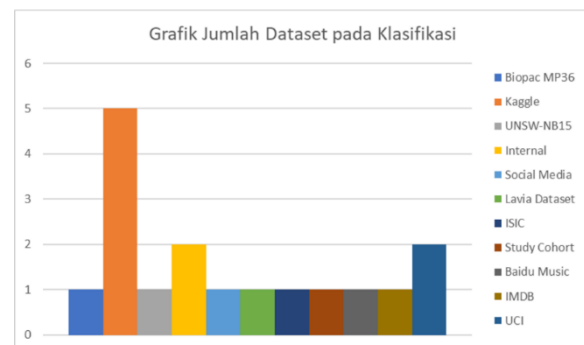
Berdasarkan artikel yang telah diamati, lima artikel merupakan mayoritas sumber dataset dan berasal dari Amerika Serikat. Umumnya artikel yang paling banyak diterbitkan berasal dari Amerika Serikat dan China. Selain itu diikuti oleh Italia sebanyak 2 artikel, sedangkan negara yang lain seperti India, Pakistan, Afrika Selatan, Finlandia, China, Australia, dan Prancis hanya didapat dari satu artikel.



Gambar 4. Dataset berdasarkan Negara pada Klasifikasi

3.4. Sumber Dataset yang digunakan pada Klasifikasi

Berdasarkan artikel yang telah diamati, sumber dataset yang paling banyak digunakan diantara yang lain pada klasifikasi adalah berasal dari *kaggle*. Umumnya dataset yang paling banyak digunakan untuk analisis data termasuk klasifikasi adalah *kaggle* karna aksesnya mudah didapat. Seperti pada gambar 5 dibawah ini, terlihat dataset yang banyak digunakan setelah *Kaggle* adalah dataset dari UCI dan Internal.



Gambar 5. Sumber Dataset yang digunakan pada Klasifikasi

3.5. Model yang memiliki Akurasi Tertinggi pada Klasifikasi

Berdasarkan artikel yang telah diamati, model yang memiliki performa tertinggi adalah model *Support Vector Machine* (SLR). Seperti pada tabel 3.1 memperlihatkan bahwa SLR mempunyai tingkat akurasi, presisi, dan *recall* lebih tinggi dibandingkan dengan model yang lain, yaitu Hutan Acak, Pohon Keputusan, KNN, dan *Naïve Bayes*.

Tabel 2. Evaluation Matrix

Artikel	Model	Metode	Akurasi	Presisi	Recall	F1	AUC
[11]	Random Forest	Deep Learning	95.93%	-	-	-	-
[12]	Decision Tree	Machine Learning	94%	-	-	-	-
[13]	SVM	Machine Learning	98.82%	98.41%	98.41%	-	-
[14]	KNN	Machine Learning	98.75	97.6%	98.8%	-	-
[15]	Naïve Bayes	Hybrid	94.97%	86.10%	78.50%	-	-

3.6. Pembahasan

Pembahasan ini dilakukan untuk menjawab *research question* pada penelitian agar lebih memahami informasi atau detail pada artikel yang telah dipilih. Rangkuman artikel dipilih berdasarkan model, metode, negara, dan dataset. Selain itu juga

terpilih berdasarkan performa dari tingkat akurasi, presisi, *recall*, F1, dan AUC. Umumnya artikel pada klasifikasi menggunakan performa akurasi. Seperti pada tabel 3.2 terlihat semua rangkuman dari artikel-artikel yang telah dipilih.

Tabel 3. Ringkasan Artikel Terpilih

No	Area Penelitian	Model	Metode	Performanc Evaluation				
				Akurasi	Presisi	Recall	F1	AUC
1.	A deep learning method for detection system	Random Forest	Deep Learning	95.93%	-	-	-	-
2.	Human Activity Recognition	Decision	Machine	94%	-	-	-	-

No	Area Penelitian	Model	Metode	Performanc Evaluation				
				Akurasi	Presisi	Recall	F1	AUC
		Tree	Learning					
3.	Techniques for Classification Using Machine Learning to Diagnose Breast Cancer	SVM	Machine Learning	98.82%	98.41%	98.4%	-	-
4.	Computerized examination of visual leaf shape highlights for plant order	KNN	Machine Learning	98.75	97.6%	98.8%	-	-
5.	Method for selecting features that is hybrid	Naïve Bayes	Hybrid	94.97%	86.10%	78.5%	-	-
6.	Machine learning algorithms and high-resolution time-frequency methods for detection	Random Forest	Machine Learning	90%	-	-	-	-
7.	Correlation of regulated AI grouping methods in forecast of locoregional repeats in early oral tongue disease	Decision Tree	Machine Learning	81%	-	-	78%	-
8.	Lyric-Based Naive Bayes Classifiers for Music Emotion Classification	Naïve Bayes	Machine Learning	68%	-	-	-	-
9.	Penggunaan Strategi Pengelompokan Tetangga Terdekat K untuk Kumpulan Data Diabetes	KNN	Machine Learning	39%	65%	36%	46%	-
10.	Random Forest Classification Using a Bolasso-Based Consistent Feature Selection Algorithm: An Application to the Evaluation of Credit Risk	Random Forest	Machine Learning	82%	-	-	-	61%
11.	Social media extremism sentiment analysis using textual data		Machine Learning	82%	81%	82%	81%	-
12.	Predicting the Recurrence of Breast Cancer		Machine Learning	75.75%	-	-	-	52%
13.	Skin Lesion Classification and Segmentation Using Deep Learning		Machine Learning	85.19%	45.62%	50%	46%	-
14.	Utilizing machine learning classifications, sentiment analysis can be improved.		Machine Learning	91.3%	97%	-	92.34%	-
15.	Diabetes Prediction with Classification Algorithms		Machine Learning	76.30%	75.9%	76.3%	76%	-
16.	Prediction of Type 2 Diabetes Using Classification Methods from Machine Learning		Machine Learning	94.1%	97.6%	-	81.3%	100%
17.	A Framework for Feature Selection in Heart Disease Classification Using Machine Learning		Machine Learning	84.36%	-	-	-	-
18.	A method for identifying abnormalities that is based on a Random Forest classifier		Machine Learning	93.58%	-	-	75%	89%

Pada tabel 4 memperlihatkan bahwa *performance evaluation* yang banyak digunakan dari semua penelitian adalah performa akurasi, sedangkan pada tabel 5 menunjukan bahwa penelitian dengan akurasi terbaik adalah model SVM.

Tabel 4. KPI Description

Performance Evaluation	Referensi Artikel
Akurasi	[11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]
Presisi	[13], [14], [15], [19], [21], [23], [24], [25], [26],
Recall	[13], [14], [15], [19], [21], [23], [25],
F1	[17], [19], [21], [23], [24], [25], [26], [27]
AUC	[20], [22], [26],

Tabel 5. Evaluation Matrix

No	Area Penelitian	Akurasi	Presisi	Recall	F1	AUC
1.	<i>A deep learning method for detection system</i>	95.93%	-	-	-	-
2.	<i>Human Activity Recognition</i>	94%	-	-	-	-
3.	<i>Techniques for Classification Using Machine Learning to Diagnose Breast Cancer</i>	98.82%	98.41%	98.4%	-	-
4.	<i>Computerized examination of visual leaf shape highlights for plant order</i>	98.75	97.6%	98.8%	-	-
5.	<i>Method for selecting features that is hybrid</i>	94.97%	86.10%	78.5%	-	-
6.	<i>Machine learning algorithms and high-resolution time-frequency methods for detection</i>	90%	-	-	-	-
7.	<i>Correlation of regulated AI grouping methods in forecast of locoregional repeats in early oral tongue disease</i>	81%	-	-	78%	-
8.	<i>Lyric-Based Naive Bayes Classifiers for Music Emotion Classification</i>	68%	-	-	-	-
9.	Penggunaan Strategi Pengelompokan Tetangga Terdekat K untuk Kumpulan Data Diabetes	39%	65%	36%	46%	-
10.	<i>Random Forest Classification Using a Bolasso-Based Consistent Feature Selection Algorithm: An Application to the Evaluation of Credit Risk</i>	82%	-	-	-	61%
11.	<i>Social media extremism sentiment analysis using textual data</i>	82%	81%	82%	81%	-
12.	<i>Predicting the Recurrence of Breast Cancer</i>	75.75%	-	-	-	52%
13.	<i>Skin Lesion Classification and Segmentation Using Deep Learning</i>	85.19%	45.62%	50%	46%	-
14.	<i>Utilizing machine learning classifications, sentiment analysis can be improved.</i>	91.3%	97%	-	92.34%	-
15.	<i>Diabetes Prediction with Classification Algorithms</i>	76.30%	75.9%	76.3%	76%	-
16.	<i>Prediction of Type 2 Diabetes Using Classification Methods from Machine Learning</i>	94.1%	97.6%	-	81.3%	100%
17.	<i>A Framework for Feature Selection in Heart Disease Classification Using Machine Learning</i>	84.36%	-	-	-	-
18.	<i>A method for identifying abnormalities that is based on a Random Forest classifier</i>	93.58%	-	-	75%	89%

Hasil Rangkuman, grafik, dan tabel yang dipilih menunjukkan bahwa model SVM memiliki akurasi kinerja tertinggi. Keakuratan dan tingkat kepercayaan model dalam mengklasifikasikan data baru umumnya diuji dalam penelitian. Rumus tersebut dapat digunakan untuk mencari akurasi suatu model. Model klasifikasi mengatur pemrosesan objek ke dalam kelas, prosedur, atau kategori [28]. Pola berdasarkan hasil model yang telah dibuat dapat digunakan untuk mengidentifikasi suatu objek dengan melakukan prosedur ini [29]. Tahapan klasifikasi dibagi menjadi dua kategori yaitu data latih dan data uji seperti yang dikemukakan pada [30]. Kumpulan data dengan atribut atau kelas yang digunakan mesin untuk mengidentifikasi data guna membuat model data disebut data pelatihan. Pengujian data, di sisi lain, adalah kumpulan data yang sama dengan data pelatihan yang memiliki atribut dan kelas dan digunakan untuk menguji akurasi model dalam mengklasifikasikan data pengujian [31]. Proses membangun model dari data pelatihan dilakukan selama klasifikasi ini menggunakan SVM, dan model tersebut kemudian digunakan dalam klasifikasi data baru [32]. Klasifikasi biasanya melibatkan identifikasi karakteristik tertentu [4]. Konstruksi model, implementasi model, dan evaluasi adalah tiga fase pertama klasifikasi. Dari cara paling umum membentuk model karakterisasi yang telah selesai,

sistem pengelompokannya layak untuk diselesaikan [33]. Selain itu, pengujian data telah dimasukkan untuk aplikasi model SVM [34]. Setelah itu, hasil klasifikasi dilihat untuk menentukan nilai akurasi yang perlu dilihat dan untuk melihat apakah model klasifikasi yang dibuat dapat digunakan [35]. Model dapat dilanjutkan dengan analisis data lebih lanjut jika tingkat akurasi yang diinginkan tercapai [6]. Teknik yang relatif baru untuk membuat prediksi, termasuk klasifikasi, adalah *Support Vector Machine* (SVM). *Supervised Learning* termasuk dalam SVM [36]. SVM membuat model untuk setiap kelas dengan menerapkan fungsi *hyperlane* ke data [37]. Standar fungsi SVM adalah mencari ruang partisi terbaik dari dataset di berbagai kelas [38]. *Hyperlane* harus diselesaikan secara linier dan akurat menggunakan pergeseran nonlinier ke dimensi yang lebih tinggi dalam contoh ini [39]. *Support vector* adalah data yang paling dekat dengan *hyperplane* [35]. Mengubah dokumen menjadi vektor sangat penting saat menggunakan SVM untuk mengklasifikasikannya. Variasi nilai digunakan untuk mendapatkan tingkat ketelitian terbaik [40]. Pencarian *hyperplane* terbaik dari semua kemungkinan adalah yang paling krusial dari metode ini [41].

5. KESIMPULAN

Pada penelitian ini menggunakan metode System Literature Review yang membantu peneliti mengkaji artikel dengan tools publish or perish untuk menemukan metode dan model yang tepat untuk klasifikasi. Sebanyak 200 artikel ditemukan. Dari seluruh artikel yang ditemukan, tersisa 17 artikel yang telah dilakukan identifikasi dan *screening* yang memenuhi kriteria inklusif dan eksklusif dibuat dengan diagram prisma. Hasil dari tinjauan literatur memperlihatkan bahwa klasifikasi banyak digunakan dengan metode *machine learning*. Selanjutnya untuk jumlah yang banyak pada sumber dataset yang digunakan berasal dari USA dengan dataset *Kaggle*. Akurasi, presisi, *recall*, dan F1 semuanya ditentukan melalui evaluasi kinerja. Dengan akurasi keseluruhan sebesar 98,82 persen, *Support Vector Machine* (SVM) merupakan model dengan tingkat akurasi tertinggi.

DAFTAR PUSTAKA

- [1] A. Nanggala and K. Suryadi, "Analisis Konsep Kampus Merdeka Dalam Perspektif Pendidikan Kewarganegaraan," *Glob. Citiz.*, vol. 9, no. 2, pp. 10–23, 2020.
- [2] N. N. Nofia, "Analisis Tantangan Implementasi Kebijakan 'Merdeka Belajar Kampus Merdeka' Pada Perguruan Tinggi Islam Negeri Di Indonesia," *PRODU Prokurasi Edukasi J. Manaj. Pendidik. Islam*, vol. 1, no. 2, pp. 61–72, 2020, doi: 10.15548/p-prokurasi.v1i2.3328.
- [3] R. Rochana, R. M. Darajatun, and M. A. Ramdhany, "Pengaruh Implementasi Kebijakan Kampus Merdeka terhadap Minat dan Keterlibatan Mahasiswa," *J. Bus. Manag. Educ.*, vol. 6, no. 3, pp. 11–21, 2021, [Online]. Available: <https://ejournal.upi.edu/index.php/JBME/article/view/40165>
- [4] N. Indriani, E. Rainarli, and K. E. Dewi, "Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *J. Infotel*, vol. 9, no. 4, p. 416, 2017, doi: 10.20895/infotel.v9i4.312.
- [5] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [6] A. C. Imanda, N. Hidayat, and M. T. Furqon, "Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2392–2399, 2018.
- [7] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [8] F. S. Pamungkas, B. D. Prasetya, and I. Kharisudin, "Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 692–697, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37875>
- [9] H. Azis, F. Tangguh Admojo, and E. Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah Performance Comparison Analysis of Classification Methods on the Multiclass Dataset of Bows," *Agustus*, vol. 19, no. 3, pp. 286–294, 2020.
- [10] E. Triandini, S. Jayanatha, A. Indrawan, G. Werla Putra, and B. Iswara, "Metode Systematic Literature Review untuk Identifikasi Platform dan Metode Pengembangan Sistem Informasi di Indonesia," *Indones. J. Inf. Syst.*, vol. 1, no. 2, p. 63, 2019, doi: 10.24002/ijis.v1i2.1916.
- [11] S. M. Kasongo and Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," *Comput. Secur.*, vol. 92, 2020, doi: 10.1016/j.cose.2020.101752.
- [12] G. De Leonardis et al., "Human Activity Recognition by Wearable Sensors : Comparison of different classifiers for real-time applications," *MeMeA 2018 - 2018 IEEE Int. Symp. Med. Meas. Appl. Proc.*, vol. 3528725544, pp. 1–6, 2018, doi: 10.1109/MeMeA.2018.8438750.
- [13] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine Learning Classification Techniques for Breast Cancer Diagnosis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 495, no. 1, 2019, doi: 10.1088/1757-899X/495/1/012033.
- [14] G. Saleem, M. Akhtar, N. Ahmed, and W. S. Qureshi, "Automated analysis of visual leaf shape features for plant classification," *Comput. Electron. Agric.*, vol. 157, no. February 2018, pp. 270–280, 2019, doi: 10.1016/j.compag.2018.12.038.
- [15] S. P. Rajamohana and K. Umamaheswari, "Hybrid approach of improved binary particle swarm optimization and shuffled frog leaping for feature selection," *Comput. Electr. Eng.*, vol. 67, pp. 497–508, 2018, doi: 10.1016/j.compeleceng.2018.02.015.
- [16] P. A. Karthick, D. M. Ghosh, and S. Ramakrishnan, "Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms," *Comput. Methods Programs Biomed.*, vol. 154, pp. 45–56, 2018, doi: 10.1016/j.cmpb.2017.10.024.
- [17] R. O. Alabi et al., "Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer," *Int. J. Med. Inform.*, vol. 136, no. December 2019, p. 104068, 2020, doi: 10.1016/j.ijmedinf.2019.104068.
- [18] Y. An, S. Sun, and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," *Proc. - 16th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2017*, no. 1, pp. 635–638, 2017, doi: 10.1109/ICIS.2017.7960070.
- [19] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [20] N. Arora and P. D. Kaur, "A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit

- risk assessment,” *Appl. Soft Comput. J.*, vol. 86, p. 105936, 2020, doi: 10.1016/j.asoc.2019.105936.
- [21] M. Asif, A. Ishtiaq, H. Ahmad, H. Aljuaid, and J. Shah, “Sentiment analysis of extremism in social media from textual information,” *Telemat. Informatics*, vol. 48, no. January, p. 101345, 2020, doi: 10.1016/j.tele.2020.101345.
- [22] A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, “Predicting breast cancer recurrence using effective classification and feature selection technique,” *19th Int. Conf. Comput. Inf. Technol. ICCIT 2016*, pp. 310–314, 2017, doi: 10.1109/ICCITECHN.2016.7860215.
- [23] R. D. Seeja and A. Suresh, “Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM),” *Asian Pacific J. Cancer Prev.*, vol. 20, no. 5, pp. 1555–1561, 2019, doi: 10.31557/APJCP.2019.20.5.1555.
- [24] J. Singh, G. Singh, and R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” *Human-centric Comput. Inf. Sci.*, vol. 7, no. 1, 2017, doi: 10.1186/s13673-017-0116-3.
- [25] D. Sisodia and D. S. Sisodia, “Prediction of Diabetes using Classification Algorithms,” *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018, doi: 10.1016/j.procs.2018.05.122.
- [26] N. P. Tigga and S. Garg, “Prediction of Type 2 Diabetes using Machine Learning Classification Methods,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 706–716, 2020, doi: 10.1016/j.procs.2020.03.336.
- [27] A. R. Chowdhury, T. Chatterjee, and S. Banerjee, “A Random Forest classifier-based approach in the detection of abnormalities in the retina,” *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 193–203, 2019, doi: 10.1007/s11517-018-1878-0.
- [28] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, “Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
- [29] M. R. A. Yudianto, K. Kusriani, and H. Al Fatta, “Analisis Pengaruh Tingkat Akurasi Klasifikasi Citra Wayang dengan Algoritma Convolutional Neural Network,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 182–191, 2020, doi: 10.36294/jurti.v4i2.1319.
- [30] A. Indriani, “Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data,” *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020, doi: 10.46984/sebatik.v24i1.909.
- [31] W. Musu, A. Ibrahim, and Heriadi, “Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5,” *Pros. Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. X, no. 1, pp. 186–195, 2021.
- [32] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [33] L. B. Ilmawan and M. A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [34] A. A. Annur, A. A. Murtopo, and N. Fadilah, “Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector Machine Dan ...,” *Indones. J. Informatics Res.*, pp. 9–17, 2021, [Online]. Available: <http://journal.universitasmulia.ac.id/index.php/sem-inastika/article/view/236%0Ahttp://journal.universitasmulia.ac.id/index.php/seminastika/article/download/236/199>
- [35] H. N. Irmanda and Ria Astriratma, “Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.
- [36] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features Twitter event detection View project Human Detection and Tracking View project,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1(12), no. October, pp. 1725–1732, 2017, [Online]. Available: <https://www.researchgate.net/publication/320234928>
- [37] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [38] O. Arifin and T. B. Sasongko, “Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier,” *Semin. Nas. Teknol. Inf. dan Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018.
- [39] Marthin Luter Laia and Yudi Setyawan, “Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM dan NBC,” *J. Stat. Ind. dan Komputasi*, vol. 5, no. 2, pp. 51–61, 2020.
- [40] R. A. Wati, H. Irsyad, and M. E. A. R. Rivan, “Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine,” *J. Algoritma*, vol. 1, no. 1, pp. 21–32, 2020, [Online]. Available: <http://download.garuda.kemdikbud.go.id/article.php?article=2652619&val=24585&title=KLASIFIKASI PNEUMONIA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE>
- [41] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.