



Analisis Data Mining Dalam Pemilihan Smartphone dan Klasifikasi di Berbagai Perangkat Menggunakan Random Forest

Ananda Aulia*, Sri Wahyuni

Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: anandaaulia30@gmail.com, sriwahyuni@dosen.pancabudi.ac.id

Email Penulis Korespondensi: anandaaulia30@gmail.com

Abstrak– Teknologi smartphone terus berkembang pesat, mendorong kebutuhan akan metode analisis yang efektif untuk membantu pengguna dalam memilih perangkat yang sesuai dengan kebutuhan mereka. Penelitian ini bertujuan untuk mengimplementasikan data mining menggunakan metode Random Forest dalam proses pemilihan smartphone dan klasifikasi perangkat berdasarkan spesifikasi teknisnya. Metode Random Forest dipilih karena kemampuannya yang andal dalam menangani data dengan jumlah atribut yang besar, menghasilkan model klasifikasi yang akurat, serta meminimalkan risiko overfitting. Dataset yang digunakan mencakup spesifikasi teknis berbagai smartphone, seperti resolusi kamera, chipset, kapasitas RAM, resolusi layar, dan dukungan untuk perekaman video 4K. Proses penelitian melibatkan pengumpulan data, pra-pemrosesan untuk menangani nilai yang hilang dan transformasi data, serta pelatihan model menggunakan algoritma Random Forest.

Hasil penelitian menunjukkan bahwa metode Random Forest mampu mengklasifikasikan perangkat dengan akurasi tinggi, membantu pengguna menentukan smartphone yang sesuai dengan kriteria mereka, seperti dukungan perekaman video 4K dan performa keseluruhan. Selain itu, penelitian ini memberikan wawasan tentang pentingnya atribut tertentu dalam pemilihan smartphone. Dengan demikian, implementasi data mining menggunakan Random Forest dapat menjadi solusi yang efektif dalam mendukung pengambilan keputusan berbasis data di bidang teknologi konsumen. Uji coba yang dilakukan mendapatkan hasil keputusan dengan menggunakan metode Random Forest dengan akurasi uji coba sebesar 99.94%.

Kata Kunci: Data Mining, Random Forest, Smartphone, Klasifikasi, Spesifikasi Teknis

Abstract– Smartphone technology continues to develop rapidly, driving the need for effective analysis methods to assist users in selecting devices that suit their needs. This research aims to implement data mining using the Random Forest method in the process of selecting smartphones and classifying devices based on their technical specifications. The Random Forest method was chosen because of its reliable ability to handle data with a large number of attributes, produce an accurate classification model, and minimize the risk of overfitting.

The dataset used includes technical specifications of various smartphones, such as camera resolution, chipset, RAM capacity, screen resolution, and support for 4K video recording. The research process involved data collection, pre-processing to handle missing values and data transformation, as well as model training using the Random Forest algorithm.

The research results show that the Random Forest method is able to classify devices with high accuracy, helping users determine smartphones that meet their criteria, such as support for 4K video recording and overall performance. Additionally, this research provides insight into the importance of certain attributes in smartphone selection. Thus, implementing data mining using Random Forest can be an effective solution in supporting data-based decision making in the field of consumer technology.

Keywords: Data Mining, Random Forest, Smartphone, Classification, Technical Specifications

I. PENDAHULUAN

Data mining telah menjadi salah satu teknologi penting dalam era big data, digunakan untuk mengekstrak informasi bernilai dari kumpulan data yang besar dan kompleks. Dalam konteks pemilihan smartphone, teknologi ini sangat relevan mengingat pesatnya perkembangan pasar perangkat seluler dalam lima tahun terakhir. Indonesia, misalnya, telah mengalami peningkatan penetrasi smartphone hingga hampir 2/3 populasi, dengan pengguna terbanyak berada di Pulau Jawa (86,60%) dan mayoritas pada kelompok usia 20-29 tahun [1][2], [3], [4].

Berbagai metode data mining, seperti klasifikasi, asosiasi, dan clustering, digunakan untuk menganalisis preferensi konsumen berdasarkan spesifikasi perangkat, seperti prosesor, RAM, kamera, baterai, dan harga. Metode klasifikasi, khususnya algoritme Random Forest, mampu memberikan hasil akurasi tinggi dengan memanfaatkan variabel data secara simultan untuk pengambilan keputusan [5], [6], [7]. Misalnya, teknik ini dapat digunakan untuk

mengelompokkan perangkat berdasarkan performa mereka atau menentukan perangkat terbaik sesuai kebutuhan pengguna .

Dalam lima tahun terakhir, lonjakan data akibat peningkatan konektivitas internet telah mempercepat adopsi teknologi ini, baik untuk mengidentifikasi pola pembelian konsumen maupun untuk meningkatkan strategi pemasaran produsen smartphone.[8] Dengan penerapan data mining, perusahaan dapat lebih memahami kebutuhan pasar dan merancang produk yang lebih relevan, memperkuat daya saing mereka di pasar yang dinamis [9], [10].

Pada penelitian [11], dari hasil penelitian yang dilakukan Random Forest pada data tanpa Skala Likert diperoleh akurasi sebesar 72% dan nilai Mtry optimal sebesar 2, sedangkan Random Forest dengan Skala Likert diperoleh akurasi sebesar 83,67% dan nilai Ntree optimal sebesar 100. Hasil penelitian ini membuktikan penerapan Skala Likert dapat meningkatkan akurasi terhadap klasifikasi tingkat kepuasan pelanggan agen BRILink menggunakan Random Forest .

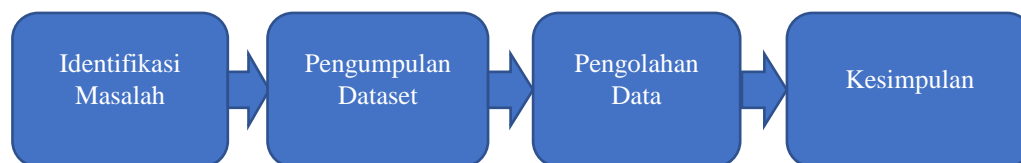
Dalam hasil penelitian [12] menunjukkan bahwa algoritma random forest dapat melakukan prediksi penyakit diabetes dengan kinerja yang baik. Nilai-nilai evaluasi kinerja algoritma random forest untuk prediksi penyakit diabetes adalah: akurasi sebesar 99.3 %, recall sebesar 99.5%, presisi sebesar 99.1%, F1-score sebesar 99.%, dan AUC sebesar 100% . Pada hasil penelitian[13] untuk mendapatkan nilai prediksi yang tinggi penelitian dilakukan secara berulang kali dengan 80% dataset untuk training dan 20% dataset untuk testing guna mencapai hasil prediksi yang tinggi. Hasil dari penelitian ini diperoleh bahwa algoritma random forest regression merupakan algoritma terbaik dalam penelitian ini dengan nilai akurasi sebesar 81,6%.

Menurut penelitian [14], objek penelitian dalam mengklasifikasikan anggur merah berdasarkan kualitas yang dipengaruhi oleh setiap anggur merah atau red wine berdasarkan pada kandungan setiap jenis wine, dari setiap atribut yang berisi komposisi dalam wine dilihat atribut mana yang paling mempengaruhi kualitas dari red wine, sehingga akan diketahui ingredients yang bisa meningkatkan kualitas dari wine tersebut, dalam penelitian ini dilakukan dengan penerapan Machine learning dengan membandingkan tiga algoritma data mining yaitu, Decision Tree, Random Forest dan Support Vector Machine (SVM), dari hasil penelitian yang telah dilakukan dengan membandingkan ketiga algoritma, dihasilkan Random Forest menghasilkan akurasi paling baik diantara algoritma lainnya yang telah diuji. Random Forest dengan hasil akurasi 0.7468 menjadikan algoritma ini paling baik digunakan untuk mengklasifikasikan kualitas red wine.

Berdasarkan tujuan hasil dari penelitian [15], yaitu menerapkan metode random forest pada klasifikasi penipuan transaksi kartu kredit berdasarkan evaluasi ketepatan klasifikasi seperti seperti confusion matrix, akurasi, sensitivitas, presisi, f-measure dan nilai AUC. Hasil dari penelitian didapatkan bahwa penerapan metode random forest memberikan hasil klasifikasi yang sangat baik dalam mengklasifikasikan penipuan transaksi kartu kredit .

II. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

1. Identifikasi Masalah

Pada tahapan ini dilakukan identifikasi masalah untuk mendefinisikan masalah yang ada juga sebagai acuan data yang akan diolah.

2. Pengumpulan Data



Pengumpulan data dilakukan dengan pencarian dataset yang diambil dari situs kaggle.com

3. Pengolahan Data Menggunakan Metode Random Forest

Pengolahan data yang sudah terkumpul akan diuji dengan metode Random Forest

4. Kesimpulan

Kesimpulan berisi data yang telah selesai diolah dalam penelitian.

2.2 Data Mining

Data mining adalah proses analisis data yang bertujuan untuk menemukan pola, tren, atau informasi tersembunyi yang bernilai dari kumpulan data yang besar dan kompleks [16]. Proses ini menggunakan metode statistik, algoritme kecerdasan buatan, pembelajaran mesin, dan teknik lainnya untuk mengekstrak pengetahuan yang dapat mendukung pengambilan keputusan. Data mining sering diterapkan di berbagai bidang, seperti pemasaran, kesehatan, pendidikan, keuangan, dan teknologi, untuk memahami perilaku konsumen, memprediksi hasil, atau meningkatkan efisiensi operasional [17]. Data mining juga merupakan metode untuk menemukan pola tertentu dari kumpulan data yang berjumlah besar. Meskipun banyak dipelajari pada bidang ilmu komputer dan statistika, data mining adalah metode yang bisa diterapkan dan mempermudah pekerjaan di bidang lainnya juga [Prastiwi, 2022]. Metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi [18].

2.3 Klasifikasi

Klasifikasi adalah salah satu teknik utama dalam data mining yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut-atributnya [19]. Metode ini menggunakan algoritme untuk membangun model dari data berlabel (data pelatihan), kemudian menggunakan model tersebut untuk memprediksi kategori data baru yang belum diketahui labelnya (data uji) [20]. Klasifikasi juga merupakan suatu proses yang dilakukan untuk menemukan sebuah model dengan tujuan untuk memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Salah satu permasalahan yang dihadapi pada klasifikasi adalah tentang ketidakseimbangan kelas (imbalance Class) yang mana suatu dataset terdapat jumlah kelas yang datanya tidak merata Sehingga memberikan dampak yang tidak baik pada hasil klasifikasi [21].

2.4 Random Forest

Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang digunakan untuk tugas klasifikasi dan regresi [22], [23]. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan (decision trees) selama pelatihan, kemudian menggabungkan hasilnya (melalui rata-rata untuk regresi atau voting untuk klasifikasi) untuk membuat prediksi yang lebih akurat dan stabil [24], [25]. Random Forest (RF) juga merupakan metode yang dapat meningkatkan hasil akurasi dalam membangkitkan atribut untuk setiap node yang dilakukan secara acak [26].

III. HASIL DAN PEMBAHASAN

Untuk memperoleh keputusan pada algoritma Random Forest, langkah-langkah yang dilakukan adalah:

1. Memilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut-atribut yang ada.
2. Membuat cabang untuk masing-masing nilai,
3. Membagi kasus dalam cabang, dan
4. Mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama. Agar dapat diolah dalam Software Rapidminer, pertama kali data yang sudah di olah sebelumnya disimpan dalam format microsoft excell file berekstensi *.csv.

Tahap pertama yang dilakukan yaitu melakukan sesi pemilihan cell untuk di Import, hal ini dilakukan agar memudahkan penelitian untuk menyusun data cell yang dipakai saat melakukan uji coba. Proses ini ditunjukkan pada Gambar 2 berikut.

Row No.	Model	prediction(M)	confidence_1	confidence_2	confidence_3	confidence_4	confidence_5	confidence_6	confidence_7	confidence_8	confidence_9	confidence_10
1	Apple iPhone 1	Apple iPhone 1	1	0	0	0	0	0	0	0	0	0
2	Apple iPhone 1	Apple iPhone 1	1	0	0	0	0	0	0	0	0	0
3	Apple iPhone 1	Apple iPhone 1	1	0	0	0	0	0	0	0	0	0
4	Apple iPhone 1	Google Pixel 8	0	0.902	0	0	0	0	0	0	0	0
5	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
6	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
7	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
8	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
9	Apple iPhone X	Apple iPhone X	0	0	0	0	1	0	0	0	0	0
10	Apple iPhone X	Apple iPhone X	0	0	0	0	1	0	0	0	0	0
11	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0.143	0	0	0	0.143
12	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0.143	0	0	0	0.143
13	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
14	Apple iPhone 1	Apple iPhone 1	0	0	0.902	0.902	0	0	0.902	0	0	0
15	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0	0	1	0	0
16	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0	0	0	0	0.143
17	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0	0.143	0	0	0.143
18	Apple iPhone X	Apple iPhone X	0	0	0	0	0	0	0.143	0	0	0.143

Gambar 2. Pemilihan Cell untuk di Import

Tahap kedua yang dilakukan yaitu melakukan sesi pemilihan variable untuk di proses hal ini dilakukan agar memudahkan penelitian untuk memilih variable data yang dipakai saat melakukan uji coba. Proses ini ditunjukkan pada Gambar 3 berikut.

Name	Type	Missing	Statistics
Model	Polynomial	0	Label: vivo Y70 5G (1) Target: Apple iPhone 15 Pro (12), Google Pixel 8 Pro (11)
prediction(Model)	Polynomial	0	Label: vivo G00 12 (0) Target: Google P [] (1463) Target: Google Pixel 8 Pro (1463), Apple iPhone 15 Pro (116)
confidence(Apple iPhone 16 Pro)	Real	0	Min: 0 Max: 1 Average: 0.905
confidence(Apple iPhone 16 P...)	Real	0	Min: 0 Max: 0.902 Average: 0.902
confidence(Apple iPhone 12 m...)	Real	0	Min: 0 Max: 0.952 Average: 0.904
confidence(Apple iPhone 13 m...)	Real	0	Min: 0 Max: 0.952 Average: 0.904
confidence(Apple iPhone XR)	Real	0	Min: 0 Max: 1 Average: 0.904
confidence(Apple iPhone XS ...)	Real	0	Min: 0 Max: 0.143 Average: 0.904
confidence(Apple iPhone 12 P...)	Real	0	Min: 0 Max: 0.952 Average: 0.904

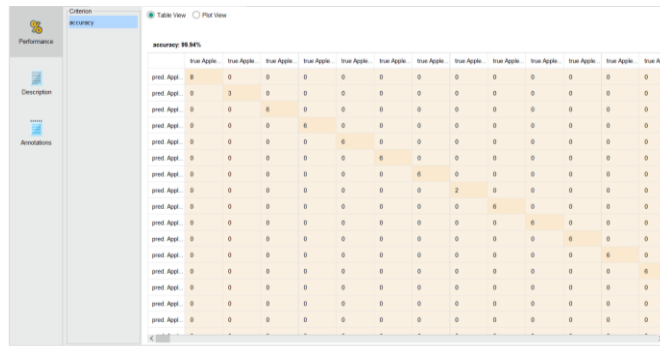
Gambar 3. Statistik Data

Pada tahap ketiga menampilkan halaman operator validation yang terbagi atas training dan testing, sesaat setelah data yang diuji ditraining selanjutnya akan dilakukan testing terhadap data testing ditunjukkan pada Gambar 4 berikut.



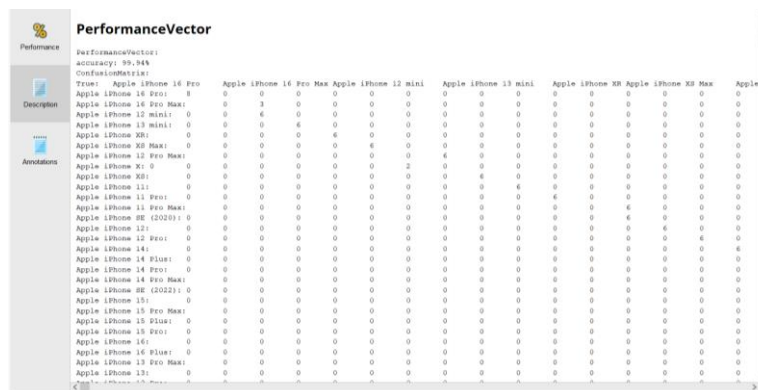
Gambar 4. Hasil Validasi

Pada tahap ke empat didapatkan hasil performa training dan testing data dengan akurasi 99.94% yang ditunjukkan pada Gambar 5 berikut.



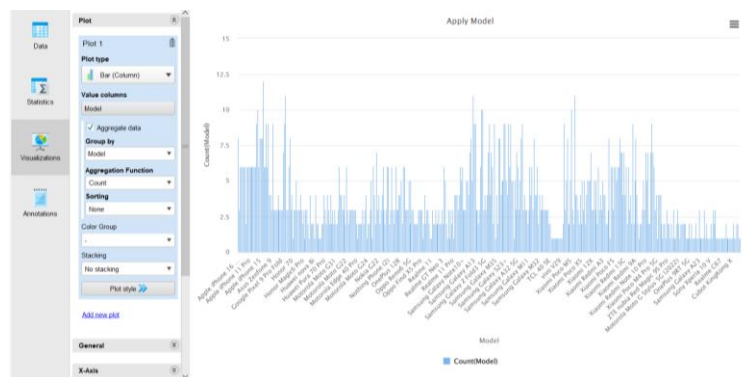
Gambar 5. Hasil Performa

Pada tahap ke lima didapatkan hasil perhitungan dengan akurasi 99.94% yang ditunjukkan pada Gambar 6 berikut.



Gambar 6. Hasil perhitungan

Berdasarkan rangkaian test yang telah dilakukan pada RapidMine, maka didapatkan hasil keputusan pada views Result. Hasilnya berbentuk apply model seperti pada Gambar 7 berikut.



Gambar 7. Hasil Keputusan Random Forest

IV. KESIMPULAN

Dari grafik diperoleh pada pengujian dengan software RapidminerStudio, Penelitian ini berhasil mengidentifikasi pola pemilihan smartphone dengan klasifikasi berdasarkan spesifikasi yang paling diminati oleh banyak pengguna yaitu pada Apple iPhone 15. Uji coba ini berdasarkan proses yang telah dilakukan mulai dari: Sesi Pemilihan Cell untuk di Import, Sesi pemilihan variable, halaman operator validation yang terbagi atas training dan testing, Hasil performa dengan akurasi 99.94%, Hasil perhitungan dengan akurasi 99.94%, dan Hasil Keputusan Random Forest. Uji coba ini didapatkan pada hasil keputusan random forest yang ditampilkan pada Gambar 7 dengan akurasi uji coba yaitu 99.94%.

Referensi

- [1] R. Meng and W. He, "Data Mining," in *Flavoromics*, Boca Raton: CRC Press, 2023, pp. 83–122. doi: 10.1201/9781003268758-8.
- [2] D. Yates and M. Z. Islam, "Data Mining on Smartphones: An Introduction and Survey," *ACM Comput Surv*, vol. 55, no. 5, 2022, doi: 10.1145/3529753.
- [3] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, 2018, doi: 10.14569/IJACSA.2018.090630.
- [4] S. R. Durugkar, R. Raja, K. K. Nagwanshi, and S. Kumar, "Introduction to data mining," 2022. doi: 10.1002/9781119792529.ch1.
- [5] M. A. Ibrahim, "Jenis, Klasifikasi dan Karakteristik Media Pembelajaran," *Journal of Economic Perspectives*, vol. 2, no. 1, 2022.
- [6] A. P. Wibawa, M. Guntur, A. Purnama, M. Fathony Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, 2018.
- [7] M. Wibowo and Muh. R. F. Djafar, "Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 1, 2023, doi: 10.30865/mib.v7i1.5182.
- [8] S. Jamil, M. Ur Rahman, and Fawad, "A Comprehensive Survey of Digital Twins and Federated Learning for Industrial Internet of Things (IIoT), Internet of Learning for Industrial Internet of Things (IIoT), Internet of Vehicles (IoV) and Internet of Drones (IoD)," *Applied System Innovation*, vol. 5, no. 3, 2022, doi: 10.3390/asi5030056.
- [9] R. L. Kumar, F. Khan, S. Kadry, and S. Rho, "A Survey on blockchain for industrial Internet of Things: Blockchain for Internet of Things," *Alexandria Engineering Journal*, vol. 61, no. 8, 2022, doi: 10.1016/j.aej.2021.11.023.
- [10] S. Villamil, C. Hernández, and G. Tarazona, "An overview of internet of things," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 5, 2020, doi: 10.12928/TELKOMNIKA.v18i5.15911.
- [11] A. A. Santika, T. H. Saragih, and M. Muliadi, "Penerapan Skala Likert pada Klasifikasi Tingkat Kepuasan Pelanggan Agen Brilink Menggunakan Random Forest," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 3, 2023, doi: 10.26418/justin.v11i3.62086.
- [12] Sriyanto and A. Ria Supriyatna, "Teknika 17 (1): 163-172 Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest," *IJCCS*, vol. x, No.x, 2023.
- [13] C. Haryanto, N. Rahaningsih, and F. Muhammad Basysyar, "KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMREDIKSI HARGA RUMAH," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, 2023, doi: 10.36040/jati.v7i1.6343.
- [14] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis : Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [15] A. Ramadhan, B. Susetyo, and Indahwati, "PENERAPAN METODE KLASIFIKASI RANDOM FOREST DALAM MENGIDENTIFIKASI FAKTOR PENTING PENILAIAN MUTU PENDIDIKAN," *Jurnal Pendidikan dan Kebudayaan*, vol. 4, no. 2, 2019, doi: 10.24832/jpnk.v4i2.1327.
- [16] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc Sci Res*, vol. 110, 2023, doi: 10.1016/j.ssresearch.2022.102817.



- [17] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans Knowl Data Eng*, vol. 26, no. 1, 2014, doi: 10.1109/TKDE.2013.109.
- [18] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda)," *Jurnal Ilmu Komputer*, vol. 9, no. 2, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [19] O.- Pahlevi, A.- Amrin, and Y.- Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *Jurnal Infortech*, vol. 5, no. 1, 2023, doi: 10.31294/infortech.v5i1.15829.
- [20] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika: Jurnal Sistem Komputer*, vol. 11, no. 1, 2022, doi: 10.34010/komputika.v11i1.4350.
- [21] A. Syukron, S. Sardiarinto, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *Jurnal Teknologi Informasi dan Terapan*, vol. 10, no. 1, 2023, doi: 10.25047/jtit.v10i1.313.
- [22] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, 2001, doi: 10.1023/A:1010933404324.
- [23] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," 2019, doi: 10.1016/j.eswa.2019.05.028.
- [24] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata Journal*, vol. 20, no. 1, 2020, doi: 10.1177/1536867X20909688.
- [25] A. Nugroho, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings of Engineering*, vol. 10, No. 2, no. 2, 2023.
- [26] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, 2022, doi: 10.35580/variansiunm31