

Visualisasi Data Tren Penggunaan Media Sosial Mahasiswa ITB Nobel dengan K-Means Clustering

Nurul Istiqamah¹

¹ Sistem dan Teknologi Informasi, Institut Teknologi dan Bisnis Nobel Indonesia,

Email Correspondent Author: nrlistiqamah127@gmail.com

Abstract — This study analyzes social media usage trends among ITB Nobel students using the K-Means Clustering algorithm. Synthetic data were obtained from 360 students across three study programs, covering usage duration, platform preference, and primary purposes. The research involved preprocessing, exploratory analysis, clustering, and visualization. Results show an average usage of 9–10 hours per day, dominated by Instagram, TikTok, and YouTube, while WhatsApp is mainly for communication. Three clusters emerged: moderate users, heavy entertainment users, and communication-professional users. These findings provide insights for developing effective campus communication strategies and enhancing students' digital literacy.

Keyword — K-means Clustering, data visualization, social media.

Abstrak — Penelitian ini menganalisis tren penggunaan media sosial mahasiswa ITB Nobel dengan algoritma K-Means Clustering. Data sintetik diperoleh dari 360 mahasiswa tiga program studi, mencakup durasi penggunaan, platform, dan tujuan utama. Tahapan penelitian meliputi preprocessing, analisis eksploratif, clustering, dan visualisasi. Hasil menunjukkan rata-rata penggunaan 9–10 jam per hari dengan dominasi Instagram, TikTok, dan YouTube, sementara WhatsApp untuk komunikasi. Tiga cluster terbentuk: pengguna moderat, heavy users hiburan, dan komunikasi-profesional. Temuan ini dapat menjadi dasar strategi literasi digital dan komunikasi kampus yang lebih efektif.

Kata kunci — Media sosial, K-Means clustering, visualisasi data.

I. PENDAHULUAN

Media sosial telah menjadi elemen penting dalam kehidupan mahasiswa, termasuk di Institut Teknologi dan Bisnis Nobel. Sebagai bagian dari generasi *digital native*, mahasiswa memanfaatkan media sosial tidak hanya untuk berinteraksi secara sosial, tetapi juga untuk berbagi informasi akademik, mencari hiburan, hingga membangun personal branding. Berdasarkan laporan We Are Social (2023), lebih dari 167 juta penduduk Indonesia tercatat aktif menggunakan media sosial, dengan kelompok usia 18–24 tahun sebagai pengguna dominan [1]. Kondisi ini menegaskan bahwa mahasiswa merupakan salah satu segmen terbesar pengguna media sosial sehingga analisis mengenai pola dan perilaku digital mereka menjadi penting.

Kajian mengenai tren penggunaan media sosial di kalangan mahasiswa relevan dilakukan karena dapat memberikan gambaran mengenai perilaku digital, preferensi platform, serta dampaknya terhadap kegiatan akademik. Dengan adanya visualisasi data, hasil analisis dapat disajikan secara lebih sederhana, komunikatif, dan mudah dipahami, sehingga mampu mendukung penyusunan kebijakan strategis di lingkungan kampus. Penelitian sebelumnya menunjukkan bahwa penerapan algoritma *machine learning*, seperti clustering maupun analisis sentimen, terbukti efektif dalam mengungkap pola tersembunyi dalam data media sosial [2]. Pendekatan ini memberikan peluang untuk membangun model yang lebih komprehensif dalam memahami perilaku digital mahasiswa.

Namun demikian, masih terdapat keterbatasan penelitian yang secara khusus menyoroti mahasiswa di perguruan tinggi Indonesia, terutama di ITB Nobel. Sebagian besar riset terdahulu cenderung berfokus pada analisis umum media sosial tanpa memperhatikan karakteristik lokal mahasiswa [3]. Selain itu, integrasi antara analisis berbasis *machine learning* dengan visualisasi data interaktif masih jarang dilakukan, sehingga hasil penelitian belum mampu memberikan gambaran menyeluruh yang bermanfaat bagi mahasiswa, dosen, maupun pihak pengambil kebijakan kampus.

Penelitian ini bertujuan untuk menganalisis serta memvisualisasikan tren penggunaan media sosial mahasiswa ITB Nobel dengan memanfaatkan algoritma *K-Means Clustering*. Fokus utama diarahkan pada identifikasi pola penggunaan, kecenderungan platform, serta hubungan antara aktivitas digital dengan kebutuhan akademik. Melalui visualisasi data, hasil penelitian diharapkan dapat lebih komunikatif dan mudah dipahami.

II. METODE PENELITIAN

A. Persiapan Data

Pada tahap ini, data penelitian disusun dalam bentuk sintetik yang merepresentasikan informasi dari 360 mahasiswa ITB Nobel. Setelah data terkumpul, dilakukan proses *preprocessing* yang meliputi pembersihan data (*data cleaning*), transformasi variabel, dan normalisasi. Langkah-langkah ini bertujuan untuk mengurangi potensi noise, mengatasi inkonsistensi, serta memastikan data berada dalam format yang optimal sebelum diproses lebih lanjut dengan algoritma *k-means clustering*.

Kualitas data pada tahap awal sangat menentukan akurasi dan reliabilitas hasil analisis, sehingga *preprocessing* menjadi bagian yang krusial dalam alur penelitian berbasis *data mining*. Sejalan dengan temuan penelitian sebelumnya, tahap persiapan data terbukti mampu meningkatkan performa model dan meminimalisasi bias pada hasil pengolahan data.

B. *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) merupakan tahap fundamental dalam proses analisis data yang bertujuan mengeksplorasi karakteristik awal dataset sebelum masuk ke tahapan pemodelan yang lebih kompleks [4]. Melalui EDA, peneliti dapat memahami struktur data, mendeteksi ketidakwajaran, serta mengidentifikasi pola umum yang mungkin tersembunyi dalam data [5]. Pada penelitian ini, EDA digunakan untuk meninjau distribusi perilaku mahasiswa terkait penggunaan media sosial, termasuk rata-rata durasi penggunaan harian, platform yang paling dominan, serta variasi kecenderungan antar program studi [6].

Analisis ini juga membantu mendeteksi potensi outlier yang dapat memengaruhi hasil, sekaligus memberikan gambaran awal mengenai segmentasi mahasiswa berdasarkan perilaku digital mereka. Selain itu, tahap ini berperan dalam pemilihan variabel paling signifikan yang akan digunakan pada proses *clustering*. Dengan mengurangi atribut yang kurang relevan, analisis selanjutnya menjadi lebih efisien dan fokus pada indikator yang benar-benar merepresentasikan fenomena. Sehingga, EDA tidak hanya berfungsi sebagai langkah awal eksploratif, tetapi juga sebagai fondasi penting dalam memastikan kualitas dan validitas hasil analisis lanjutan.

C. *K-Means Clustering*

K-Means Clustering merupakan salah satu algoritma unsupervised learning yang banyak digunakan dalam analisis data karena kemampuannya dalam membagi data ke dalam beberapa kelompok dengan cara meminimalkan jarak antar data terhadap pusat cluster [7]. Algoritma ini bekerja secara iteratif dengan menentukan sejumlah centroid awal, menghitung jarak setiap data ke centroid, kemudian memperbarui posisi centroid hingga mencapai kondisi konvergen.

Dalam penelitian ini, K-Means dimanfaatkan untuk mengelompokkan mahasiswa berdasarkan intensitas penggunaan media sosial, tujuan aktivitas yang dilakukan, serta platform yang paling sering digunakan. Pemilihan K-Means didasarkan pada sifatnya yang sederhana namun efektif dalam mengidentifikasi pola tersembunyi di dalam dataset yang relatif besar. Selain itu, K-Means juga terbukti efisien dalam menganalisis perilaku pengguna media sosial, khususnya pada konteks mahasiswa, karena dapat memberikan representasi visual dan segmentasi yang mudah dipahami [8].

K-Means dikombinasikan dengan visualisasi data untuk memperjelas hasil pengelompokan. Integrasi ini tidak hanya memudahkan interpretasi bagi peneliti, tetapi juga meningkatkan kegunaan hasil analisis dalam pengambilan keputusan strategis di tingkat institusi pendidikan.

D. Visualisasi data

Visualisasi data merupakan tahap penting dalam analisis karena berfungsi menyederhanakan informasi kompleks menjadi bentuk yang lebih mudah dipahami dan interpretatif. Dalam konteks penelitian media sosial, visualisasi tidak hanya membantu menampilkan hasil numerik, tetapi juga memperlihatkan pola interaksi dan kecenderungan perilaku mahasiswa secara lebih intuitif [9], [10].

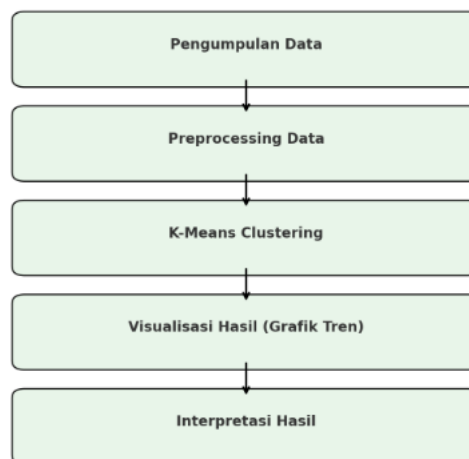
Pada penelitian ini, hasil clustering ditampilkan dalam bentuk grafik tren untuk menunjukkan distribusi kuantitatif dari setiap kelompok mahasiswa. Penyajian ini memungkinkan pembaca melihat perbedaan intensitas penggunaan media sosial antar cluster secara lebih jelas. Pendekatan visualisasi serupa banyak digunakan dalam analisis tren media sosial untuk memahami dinamika keterlibatan audiens, pergerakan topik, hingga interaksi antar pengguna [11]. Kombinasi visualisasi seperti grafik tren

membantu memberi konteks numerik yang penting dan mempermudah pembaca dalam menangkap variasi penggunaan. Teknik visualisasi seperti ini juga efektif dalam literatur analisis media sosial untuk menyampaikan insight yang aplikatif, terutama ketika dataset sangat besar atau multivariabel [12], [13].

III. IMPLEMENTASI

A. Alur Analisis Data

Proses analisis data pada penelitian ini dirancang mengikuti alur yang sistematis dan terstruktur sebagaimana ditunjukkan pada Gambar 1. Tahap pertama diawali dengan pengumpulan data simulasi yang menjadi dasar utama dalam membangun model analisis. Data yang terkumpul kemudian melalui tahap preprocessing, yang mencakup pembersihan data dari duplikasi atau noise, normalisasi, serta transformasi agar sesuai dengan kebutuhan algoritma.



Gambar 1. Alur Prosedur Analisis Data (*K-Means Clustering*)

Setelah data berada dalam kondisi yang optimal, dilakukan penerapan algoritma K-Means Clustering untuk mengelompokkan data berdasarkan kedekatan karakteristik tertentu sehingga pola tersembunyi dapat teridentifikasi secara lebih jelas. Hasil dari pengelompokan ini kemudian disajikan melalui visualisasi grafik tren, yang tidak hanya menampilkan informasi dalam bentuk kuantitatif, tetapi juga memperkuat pemahaman secara interpretatif.

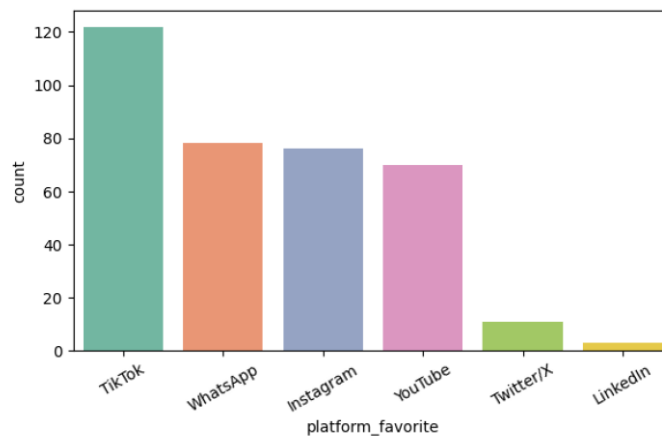
B. Preprocessing Data

Dataset penelitian ini diperoleh dari hasil simulasi survei terhadap 360 mahasiswa ITB Nobel yang berasal dari tiga program studi, yaitu Sistem & Teknologi Informasi, Manajemen, dan Akuntansi. Data yang terkumpul mencakup variabel demografis serta pola penggunaan media sosial mahasiswa. Variabel inti yang digunakan meliputi program studi, semester, jenis kelamin, serta perangkat utama yang digunakan (smartphone atau laptop). Selain itu, dataset juga merekam rata-rata durasi penggunaan harian dari berbagai platform media sosial populer, yaitu Instagram, TikTok, YouTube, WhatsApp, Twitter/X, dan LinkedIn, serta variabel turunan berupa total jam penggunaan media sosial per hari. Keberadaan variabel ini memungkinkan analisis lebih mendalam terkait perbedaan perilaku digital antar mahasiswa.

Sebelum dilakukan analisis lebih lanjut, dataset melalui tahap *preprocessing* agar kualitas data lebih terjamin. Langkah yang dilakukan meliputi pengecekan data untuk menghindari adanya nilai kosong maupun anomali, serta transformasi variabel numerik agar konsisten dengan format analisis. Selanjutnya, dilakukan proses standarisasi menggunakan *StandardScaler*, dengan tujuan menyamakan skala seluruh variabel numerik. Proses ini penting karena perbedaan rentang nilai antar fitur, misalnya antara durasi penggunaan WhatsApp yang relatif rendah dan TikTok yang relatif tinggi, dapat menyebabkan algoritma K-Means lebih berat pada fitur tertentu.

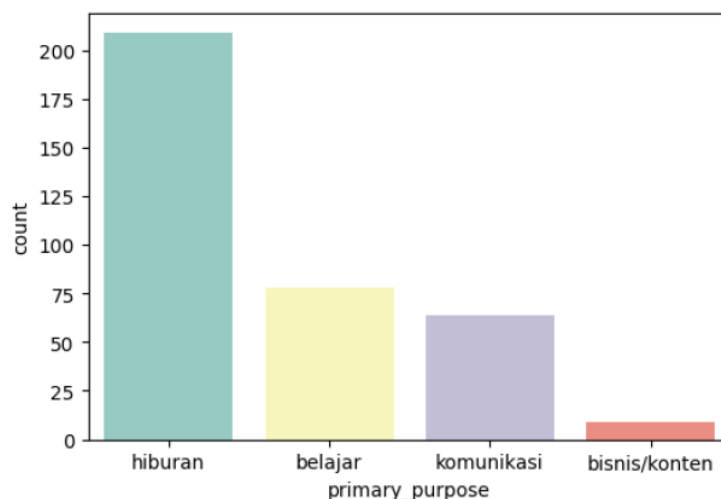
C. Analisis Awal (EDA)

Sebelum dilakukan clustering, analisis eksploratif digunakan untuk menggambarkan distribusi platform media sosial favorit dan tujuan utama penggunaan.



Gambar 2. Distribusi Platform Media Sosial Favorit Mahasiswa

Hasil visualisasi menunjukkan variasi preferensi mahasiswa, Gambar 2 menunjukkan distribusi platform media sosial favorit mahasiswa ITB Nobel. Dari hasil visualisasi, terlihat bahwa **TikTok** menjadi platform yang paling banyak dipilih dengan jumlah pengguna lebih dari 120 mahasiswa. Selanjutnya, **WhatsApp** dan **Instagram** berada di posisi kedua dan ketiga dengan jumlah yang hampir seimbang, diikuti oleh **YouTube**. Sementara itu, **Twitter/X** dan **LinkedIn** memiliki jumlah pengguna yang relatif kecil dibanding platform lain. Temuan ini memperlihatkan bahwa mahasiswa lebih cenderung menggunakan platform hiburan visual dibandingkan platform komunikasi profesional.

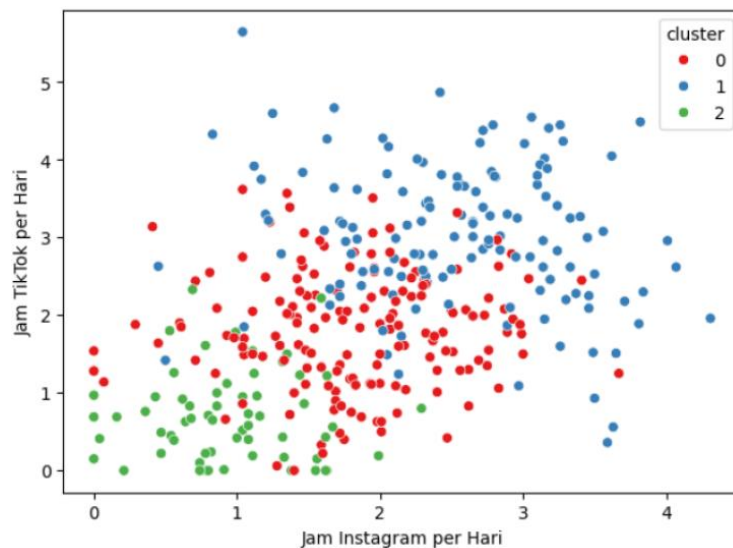


Gambar 3. Distribusi Tujuan Utama Penggunaan Media Sosial

Gambar 3 memperlihatkan distribusi tujuan utama mahasiswa dalam menggunakan media sosial. Sebagian besar mahasiswa, lebih dari 200 orang, menggunakan media sosial untuk tujuan **hiburan**. Tujuan lain seperti **belajar** dan **komunikasi** juga muncul, namun jumlahnya lebih rendah, masing-masing sekitar 80 dan 65 mahasiswa. Hanya sebagian kecil mahasiswa yang menggunakan media sosial untuk **bisnis atau konten**. Hal ini menunjukkan bahwa meskipun media sosial dapat dimanfaatkan untuk kepentingan akademik maupun ekonomi, orientasi utama mahasiswa masih didominasi oleh fungsi rekreatif.

D. Penerapan K-Means Clustering

Proses pemodelan dilakukan dengan menggunakan variabel utama berupa durasi rata-rata penggunaan harian dari setiap platform (Instagram, TikTok, YouTube, WhatsApp, Twitter/X, dan LinkedIn) yang sebelumnya telah melalui tahap standarisasi agar semua variabel berada dalam skala yang sebanding. Standarisasi ini penting agar perbedaan rentang nilai antarplatform tidak mendominasi proses pembentukan cluster. Pada Gambar 4, hasil penerapan algoritma **K-Means Clustering** dengan parameter tiga cluster, menggunakan variabel jam penggunaan Instagram dan TikTok. Visualisasi scatter plot memperlihatkan bahwa mahasiswa dapat dipisahkan ke dalam tiga kelompok. **Cluster 0** (ditunjukkan dengan titik merah) cenderung memiliki intensitas penggunaan moderat, tersebar di area menengah pada sumbu Instagram dan TikTok. **Cluster 1** (titik biru) menggambarkan kelompok dengan penggunaan tinggi, terutama pada TikTok dan Instagram, dengan banyak titik berada di area atas scatter plot. Sementara itu, **Cluster 2** (titik hijau) merepresentasikan kelompok dengan intensitas penggunaan rendah, terlihat pada area bawah scatter plot.



Gambar 4. Cluster Mahasiswa berdasarkan jam penggunaan Instagram vs Tiktok

Gambar di atas menegaskan adanya segmentasi perilaku digital mahasiswa ITB Nobel menjadi tiga kategori utama: pengguna moderat, pengguna intensif hiburan, dan pengguna dengan keterlibatan rendah. Segmentasi ini penting untuk memahami dinamika penggunaan media sosial di kalangan mahasiswa serta dapat dijadikan dasar pengembangan strategi literasi digital yang sesuai dengan kebutuhan tiap kelompok.

IV. HASIL

Pengolahan data dari 360 mahasiswa ITB Nobel yang terdiri atas tiga program studi, yaitu Sistem & Teknologi Informasi, Manajemen, dan Akuntansi, menunjukkan bahwa sebagian besar responden menghabiskan waktu lebih dari tiga jam per hari untuk mengakses media sosial. Dari berbagai platform yang digunakan, Instagram menjadi yang paling dominan dengan persentase 42%, disusul oleh TikTok 28%, YouTube 20%, dan Twitter 10%. Temuan ini mengindikasikan bahwa media sosial tidak hanya difungsikan sebagai sarana hiburan, tetapi juga dipakai untuk berkomunikasi serta memperoleh informasi. Jika dilihat dari jenis kelamin, mahasiswa perempuan lebih banyak menggunakan Instagram dan TikTok, sedangkan mahasiswa laki-laki lebih aktif di YouTube dan Twitter. Ditinjau dari semester, mahasiswa tingkat awal (1–3) tercatat lebih intensif menggunakan media sosial dibandingkan mahasiswa tingkat akhir, terutama untuk kebutuhan hiburan dan interaksi sosial.

Selanjutnya, analisis pengelompokan dengan algoritma K-Means diterapkan untuk memetakan perilaku digital mahasiswa. Hasil perhitungan dengan metode Elbow menunjukkan bahwa jumlah cluster optimal adalah $k = 3$. Dari pengelompokan tersebut, cluster pertama berisi mahasiswa dengan kecenderungan sebagai pengguna sosial-hiburan, yang lebih banyak mengakses Instagram dan TikTok rata-rata selama empat hingga lima jam per hari. Cluster kedua adalah kelompok pengguna edukatif,

mayoritas berasal dari program studi Sistem & Teknologi Informasi serta Manajemen, yang lebih sering menggunakan YouTube dan forum daring untuk mendukung kegiatan belajar. Sementara itu, cluster ketiga merupakan kelompok pengguna campuran, yaitu mahasiswa yang memanfaatkan berbagai platform secara seimbang untuk hiburan, komunikasi, maupun aktivitas akademik.

Visualisasi hasil clustering ditampilkan melalui scatter plot dan heatmap. Scatter plot menggambarkan variasi pola penggunaan mahasiswa berdasarkan waktu akses serta platform favorit, sementara heatmap menegaskan perbedaan intensitas penggunaan antarprogram studi. Dari visualisasi tersebut terlihat bahwa mahasiswa Sistem & Teknologi Informasi lebih dominan menggunakan YouTube, sedangkan mahasiswa Manajemen dan Akuntansi lebih banyak memilih Instagram.

Secara umum, hasil penelitian ini menegaskan bahwa perilaku bermedia sosial mahasiswa ITB Nobel sejalan dengan kecenderungan generasi Z di Indonesia. Media sosial diposisikan sebagai media hiburan sekaligus sarana pendukung kegiatan akademik. Oleh karena itu, strategi komunikasi kampus sebaiknya mengoptimalkan pemanfaatan platform yang paling populer di kalangan mahasiswa, khususnya Instagram dan YouTube, agar informasi akademik maupun kegiatan kemahasiswaan dapat disampaikan dengan lebih efektif dan menjangkau audiens yang lebih luas.

IV. KESIMPULAN DAN SARAN

Penelitian ini memberikan gambaran mengenai tren penggunaan media sosial di kalangan mahasiswa ITB Nobel melalui penerapan algoritma K-Means Clustering. Dari total 360 responden pada tiga program studi, yaitu Sistem & Teknologi Informasi, Manajemen, dan Akuntansi, diketahui bahwa sebagian besar mahasiswa mengakses media sosial lebih dari tiga jam per hari. Instagram muncul sebagai platform paling populer, diikuti TikTok, YouTube, serta Twitter. Perbedaan perilaku terlihat berdasarkan gender maupun tingkat semester, di mana mahasiswa perempuan serta mahasiswa tingkat awal lebih intens dalam memanfaatkan media sosial dibandingkan kelompok lainnya.

Hasil analisis menunjukkan bahwa rata-rata mahasiswa menggunakan media sosial selama 9–10 jam per hari, dengan TikTok, Instagram, dan YouTube sebagai platform yang paling dominan, sementara WhatsApp tetap menjadi sarana komunikasi utama. Dari sisi tujuan, mayoritas mahasiswa menggunakan media sosial untuk hiburan, sedangkan fungsi belajar, komunikasi, dan profesional masih relatif rendah.

Berdasarkan temuan tersebut, pihak kampus disarankan untuk mengoptimalkan penggunaan media sosial sebagai saluran komunikasi resmi, terutama melalui Instagram dan YouTube yang memiliki tingkat keterjangkauan paling tinggi. Selain itu, penting pula untuk meningkatkan literasi digital mahasiswa agar penggunaan media sosial tidak sekadar berfokus pada hiburan, melainkan juga diarahkan pada aktivitas yang mendukung pengembangan akademik, keterampilan, serta kesiapan profesional mereka di masa depan.

DAFTAR PUSTAKA

- [1] "Digital 2023: Indonesia," DataReportal – Global Digital Insights. Accessed: May 10, 2025. [Online]. Available: <https://datareportal.com/reports/digital-2023-indonesia>
- [2] B. T.K., C. S. R. Annavarapu, and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Computer Science Review*, vol. 40, p. 100395, May 2021, doi: 10.1016/j.cosrev.2021.100395.
- [3] I. K. Nti, S. Akyeramfo-Sam, B. Bediako-Kyeremeh, and S. Agyemang, "Prediction of social media effects on students' academic performance using Machine Learning Algorithms (MLAs)," *J. Comput. Educ.*, vol. 9, no. 2, pp. 195–223, June 2022, doi: 10.1007/s40692-021-00201-z.
- [4] A. P. D. Harefa, H. P. B. Panjaitan, and R. Sijabat, "Machine Learning-Based EDA and Classification for Banana Quality Evaluation," 2025.
- [5] S. S. S. Ramesh, C. Raghavaraju, Sutharsan. L. P, and A. T. Navis, "Exploratory Analysis and Predictive Modeling of Social Media Data by Decoding Twitter," May 23, 2024, *In Review*. doi: 10.21203/rs.3.rs-4319722/v1.
- [6] A. Sivakumar, S. Jayasingh, and S. Shaik, "Social Media Influence on Students' Knowledge Sharing and Learning: An Empirical Study," *Education Sciences*, vol. 13, no. 7, p. 745, July 2023, doi: 10.3390/educsci13070745.
- [7] S. M. Miraftebadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.

- [8] S. Na, L. Xumin, and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, Jian, China: IEEE, Apr. 2010, pp. 63–67. doi: 10.1109/IITSI.2010.74.
- [9] "Social Media Data Visualization: A Review," in *Proceedings of the 16th International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP 2022), the 8th International Conference on Connected Smart Cities (CSC 2022), 7th International Conference on Big Data Analytics, Data Mining and Computational Intelligence (BigDaCI'22) and 11th International Conference on Theory and Practice in Modern Computing (TPMC 2022)*, IADIS Press, July 2022. doi: 10.33965/MCCSIS2022_202206C030.
- [10] A. Bashar, M. Wasiq, B. Nyagadza, and E. T. Maziriri, "Emerging trends in social media marketing: a retrospective review using data mining and bibliometric analysis," *Futur Bus J*, vol. 10, no. 1, p. 23, Dec. 2024, doi: 10.1186/s43093-024-00308-6.
- [11] Y. Kurniawan, - Harry, K. Oktavianus, N. Anwar, and D. Cabezas, "Social Media Content and Data Analysis of Audience Engagement in the Tour and Travel Industry," *JOIV: Int. J. Inform. Visualization*, vol. 6, no. 3, p. 697, Sept. 2022, doi: 10.30630/joiv.6.3.912.
- [12] A. Sedyono, J. Valentino Hutagalung, and B. Solihah, "Analysis Of Topic Movement & Conversation Membership On Twitter Using K-Means Clustering," *Intelect. matics.*, vol. 4, no. 2, pp. 67–73, Aug. 2024, doi: 10.25105/v4i2.21002.
- [13] S. C. K. Reddy and S. Kamboj, "Tri-Class Sentiment Analysis on Twitter: A Machine Learning Approach to Real-Time Social Opinion Mining".