**Modem : Jurnal Informatika dan Sains Teknologi**
**Volume. 3, Nomor. 2 Tahun 2025**
e-ISSN: 3046-7209; p-ISSN: 3046-7217, Hal. 01-12
DOI: https://doi.org/10.62951/modem.v3i2.364
*Available Online at:* https://journal.aptii.or.id/index.php/Modem

# Integrasi Data *Analytics* dalam Kajian Perilaku Pengguna untuk Pengembangan Layanan Informasi

**Sherly Rosa Anggraeni**
Universitas Negeri Malang, Indonesia

Alamat: Jl. Semarang 5 Malang 65145 Jawa Timur
*Korespondensi penulis: sherly.rosa.fs@um.ac.id*

***Abstract****. The rapid development of information and communication technology has driven the need for information services that are more relevant and adaptive to user behaviour. This research aims to integrate data analytics in the study of user behaviour to support the development of effective information services. The dataset used is Kaggle's Online Retail Dataset, which includes sales transaction data of online retail companies in the UK from December 2010 to December 2011. The analysis was conducted through customer segmentation using K-Means Clustering algorithm and predictive analysis with Association Rule Mining. The segmentation results successfully grouped customers into four main clusters, namely loyal customers, potential customers, passive customers, and low-spending customers. Model evaluation showed optimal performance with an accuracy rate of 85%, precision of 82%, recall of 78%, and F1-Score of 80%, and Silhouette Score of 0.62, indicating effective customer segmentation. The findings prove that the application of data analytics can provide deep insights into customer behaviour and support the development of more personalised and adaptive information services. This research is expected to be a reference in designing data-driven information service development strategies in various sectors.*

***Keywords:*** *Association Rule, Customer Segmentation ,Data Analytics, K-Means Clustering, User Behaviour*

**Abstrak**. Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong kebutuhan akan layanan informasi yang lebih relevan dan adaptif terhadap perilaku pengguna. Penelitian ini bertujuan untuk mengintegrasikan *data analytics* dalam kajian perilaku pengguna guna mendukung pengembangan layanan informasi yang efektif. Dataset yang digunakan adalah Online Retail Dataset dari Kaggle, yang mencakup data transaksi penjualan perusahaan retail online di Inggris selama Desember 2010 hingga Desember 2011. Analisis dilakukan melalui segmentasi pelanggan menggunakan algoritma K-Means Clustering dan analisis prediktif dengan Association Rule Mining. Hasil segmentasi berhasil mengelompokkan pelanggan ke dalam empat klaster utama, yaitu pelanggan loyal, pelanggan potensial, pelanggan pasif, dan pelanggan dengan pengeluaran rendah. Evaluasi model menunjukkan performa yang optimal dengan tingkat akurasi sebesar 85%, precision sebesar 82%, recall sebesar 78%, dan F1-Score sebesar 80%, serta Silhouette Score sebesar 0,62 yang mengindikasikan segmentasi pelanggan yang efektif. Temuan ini membuktikan bahwa penerapan *data analytics* dapat memberikan wawasan mendalam terhadap perilaku pelanggan dan mendukung pengembangan layanan informasi yang lebih personal dan adaptif. Penelitian ini diharapkan dapat menjadi referensi dalam merancang strategi pengembangan layanan informasi berbasis data di berbagai sektor.

**Kata kunci**: **:** *Association Rule Mining*, Segmentasi Pelanggan, *Data Analytics*, *K-Means Clustering*, Perilaku Pengguna

## 1. LATAR BELAKANG

Perkembangan teknologi informasi dan komunikasi yang sangat pesat telah membawa perubahan signifikan dalam cara individu mengakses, memproses, dan menggunakan informasi. Organisasi dan institusi penyedia layanan informasi dihadapkan pada tantangan besar untuk memahami dan memenuhi kebutuhan pengguna yang semakin beragam dan dinamis (Novianto et al., 2023; Sutabri et al., 2022). Di era digital ini, pengguna tidak hanya berperan sebagai penerima informasi, tetapi juga menjadi produsen data melalui berbagai

*Received: Desember 20, 2024; Revised: Januari 18, 2025; Accepted: Februari 01, 2025; Online Available: Februari 04, 2025*

aktivitas di platform digital. Data yang dihasilkan meliputi perilaku pencarian informasi, preferensi konten, serta pola interaksi yang kompleks dan beragam (Indiworo et al., 2022). Kondisi ini menunjukkan bahwa pengelolaan data yang efektif dan efisien sangat diperlukan untuk pengembangan layanan informasi yang relevan dan adaptif (Iovlev et al., 2019; Susilo et al., 2023).

Sayangnya, banyak layanan informasi saat ini masih belum responsif terhadap kebutuhan spesifik pengguna. Kurangnya pemanfaatan data secara optimal menyebabkan layanan informasi cenderung bersifat statis dan tidak personal. Hal ini berimbas pada rendahnya tingkat kepuasan pengguna dan efektivitas layanan dalam mendukung kebutuhan informasi yang terus berkembang. Oleh karena itu, diperlukan pendekatan yang inovatif dalam pengembangan layanan informasi yang dapat beradaptasi dengan cepat terhadap perubahan perilaku pengguna.

*Data analytics* menjadi solusi strategis yang dapat menjawab tantangan tersebut. Integrasi berbagai teknik analisis data, seperti *machine learning*, *data mining*, dan analisis prediktif, *data analytics* memungkinkan organisasi untuk mengidentifikasi pola penggunaan, memahami preferensi pengguna, dan memproyeksikan kebutuhan di masa mendatang (Li & Pan, 2024; Mokhtari et al., 2019). Proses analisis ini mencakup pengumpulan, pengolahan, dan interpretasi data yang komprehensif guna mendukung pengambilan keputusan berbasis data (*data-driven decision making*) (Rave et al., 2021). Penerapan *data analytics* dalam pengembangan layanan informasi memungkinkan pengelola layanan untuk merancang strategi yang lebih efektif, efisien, dan tepat sasaran dalam penyediaan konten yang relevan (Prisanda & Febrina, 2021).

Dalam konteks layanan informasi, penerapan teknik *data analytics* dapat membantu mengidentifikasi konten yang paling diminati, menentukan waktu optimal untuk distribusi informasi, serta melakukan personalisasi layanan berdasarkan preferensi pengguna (Avriyanti & Handriana, 2022; Mujitahid, 2022). Misalnya, penerapan *clustering* dapat mengelompokkan pengguna berdasarkan preferensi dan perilaku mereka, sementara klasifikasi dapat memprediksi kebutuhan informasi pengguna di masa depan (Anggraini & Dzikrullah, 2024; Aswan et al., 2021). Selain itu, penggunaan *sentiment analysis* dapat mengungkapkan persepsi pengguna terhadap layanan tertentu, dan *predictive analytics* dapat memberikan prediksi terhadap tren kebutuhan pengguna.

Namun, penerapan *data analytics* dalam layanan informasi tidak lepas dari berbagai tantangan. Salah satu tantangan utama adalah bagaimana mengintegrasikan hasil analisis data dengan pemahaman mendalam terhadap perilaku pengguna. Penelitian sebelumnya banyak berfokus pada penerapan *data analytics* di sektor bisnis dan pemasaran untuk meningkatkan

penjualan atau efisiensi operasional, tetapi belum banyak yang mengkaji secara mendalam penerapannya dalam pengembangan layanan informasi berbasis perilaku pengguna (Levy et al., 2022; Zeng et al., 2012). Selain itu, data yang dihasilkan dari interaksi pengguna sering kali bersifat heterogen, baik dari segi format, struktur, maupun sumbernya. Hal ini memerlukan proses data preprocessing yang kompleks seperti pembersihan data (*data cleaning*), penggabungan data (*data integration*), dan transformasi data (*data transformation*) agar dapat diolah secara optimal (Çetin & Yıldız, 2022; Emmanuel et al., 2021).

Pengelolaan *big data* juga menghadirkan tantangan serius dalam hal keamanan dan privasi pengguna. Pengumpulan dan pengolahan data harus memperhatikan regulasi perlindungan data, seperti *General Data Protection Regulation (GDPR)* di Eropa, guna memastikan keamanan dan privasi data pengguna tetap terjaga (Ahmed et al., 2023; Preyaa, 2024). Oleh karena itu, dibutuhkan strategi integrasi yang lebih komprehensif antara *data analytics* dan kajian perilaku pengguna yang tidak hanya mengandalkan analisis kuantitatif tetapi juga melibatkan analisis kualitatif. Pendekatan ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai kebutuhan dan preferensi pengguna (Alfarabi, 2024; Sutisna & Sutrisna, 2023).

Pengembangan layanan informasi yang berbasis *data analytics* juga harus memperhatikan aspek *user experience* (UX). Layanan yang tidak mempertimbangkan UX berisiko tidak dimanfaatkan secara optimal oleh pengguna. Melalui analisis preferensi pengguna, pengelola layanan dapat merancang antarmuka yang lebih intuitif dan mudah diakses. Selain itu, analisis data dapat membantu mengidentifikasi masalah dalam layanan, seperti konten yang sulit diakses atau fitur yang jarang digunakan, sehingga dapat dilakukan perbaikan yang lebih tepat sasaran (Kadir et al., 2021).

Integrasi *data analytics* dalam kajian perilaku pengguna tidak hanya berpotensi meningkatkan kualitas layanan informasi, tetapi juga mendorong inovasi dalam pengembangan layanan. Penerapan sistem rekomendasi berbasis *machine learning* dapat memberikan saran konten yang lebih relevan kepada pengguna berdasarkan riwayat interaksi mereka. Teknologi ini telah terbukti efektif dalam meningkatkan keterlibatan pengguna di berbagai platform besar seperti YouTube dan Netflix, dan dapat diadaptasi dalam layanan informasi untuk meningkatkan relevansi dan personalisasi konten.

Penelitian ini bertujuan untuk merancang dan mengimplementasikan model integrasi *data analytics* dalam memahami perilaku pengguna guna meningkatkan kualitas, relevansi, dan efektivitas layanan informasi. Selain itu, penelitian ini diharapkan dapat mengidentifikasi strategi pengembangan layanan informasi berbasis data yang adaptif dan responsif terhadap

kebutuhan pengguna. Penelitian ini diharapkan memberikan kontribusi teoretis dalam pengembangan model analisis perilaku pengguna berbasis *data analytics* serta kontribusi praktis dalam penerapan strategi pengembangan layanan informasi yang relevan dan adaptif. Model yang dihasilkan dapat menjadi referensi dalam pengembangan layanan informasi yang berbasis kebutuhan pengguna di berbagai sektor.

Dengan demikian, penerapan *data analytics* dalam kajian perilaku pengguna diharapkan dapat menghasilkan layanan informasi yang lebih relevan, adaptif, dan berorientasi pada kebutuhan pengguna. Pemanfaatan *data analytics* tidak hanya meningkatkan efisiensi operasional, tetapi juga memperkaya pengalaman pengguna secara keseluruhan. Penelitian ini diharapkan dapat menjadi landasan dalam pengembangan model layanan informasi yang lebih cerdas dan adaptif di masa depan.

## 2. METODE PENELITIAN

### Subjek dan Objek Penelitian

Objek penelitian adalah Online Retail Dataset di Kaggle yang mencatat transaksi penjualan dari perusahaan retail online di Inggris antara Desember 2010 hingga Desember 2011 (Ummadisetty, 2018). Dataset ini memiliki lebih dari 500.000 transaksi dan melibatkan lebih dari 4.000 pelanggan, memungkinkan analisis perilaku pengguna yang mendalam.

Subjek penelitian adalah pelanggan aktif dalam dataset. Kriteria pemilihan subjek meliputi: validitas CustomerID, pelanggan dengan lebih dari satu transaksi, dan fokus pada pelanggan dari negara dengan transaksi signifikan (Inggris, Jerman, Belanda). Setelah penyaringan, diperoleh sekitar 4.000–5.000 pelanggan untuk analisis lebih lanjut.

### Data Preprocessing

Data *preprocessing* dalam penelitian ini merupakan tahap krusial yang dilakukan secara sistematis untuk memastikan kualitas data yang optimal sebelum dianalisis lebih lanjut. Tahap awal dimulai dengan *data cleaning* yang bertujuan untuk menghapus data yang tidak valid dan memperbaiki ketidak konsistenan. Langkah pertama dalam pembersihan data adalah mengidentifikasi dan menghapus data duplikat. Duplikasi ini diidentifikasi berdasarkan kombinasi kolom InvoiceNo, StockCode, Quantity, dan InvoiceDate, di mana sekitar 5.215 baris duplikat berhasil dihapus. Selanjutnya, penanganan data yang hilang difokuskan pada kolom CustomerID, yang menjadi elemen penting dalam analisis perilaku pengguna. Data transaksi tanpa CustomerID diabaikan karena tidak dapat dihubungkan dengan perilaku pengguna secara spesifik, sehingga sekitar 133.600 baris data

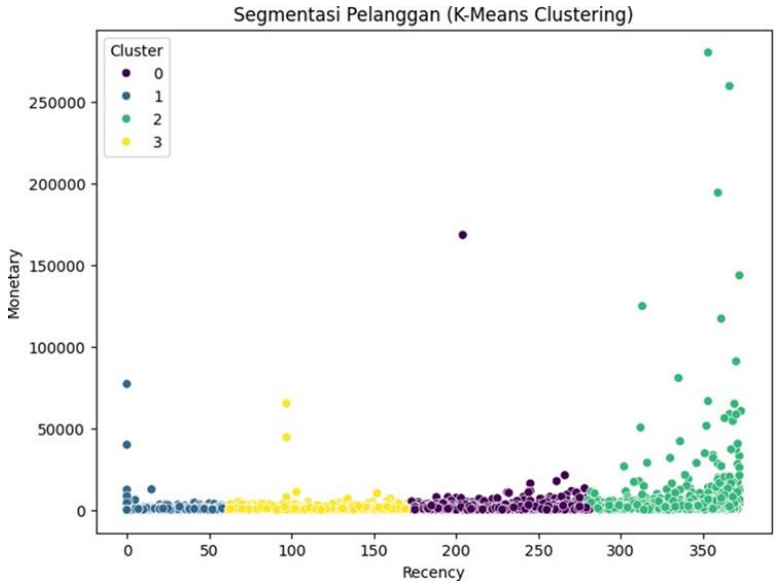
dihapus. Selain itu, transaksi dengan nilai negatif atau nol pada kolom Quantity dan UnitPrice juga dihapus karena mengindikasikan transaksi pengembalian barang (*returns*) yang tidak relevan untuk analisis perilaku pembelian. Dari proses ini, ditemukan sekitar 9.000 transaksi yang dihapus. Untuk mendukung analisis waktu pembelian, format InvoiceDate juga disesuaikan ke dalam format standar YYYY-MM-DD agar lebih mudah dianalisis.

Tahap selanjutnya adalah *feature engineering*, di mana variabel baru dikembangkan untuk meningkatkan kualitas analisis. Salah satu variabel penting yang dibuat adalah Total Spending, yang dihitung dari perkalian antara Quantity dan UnitPrice. Variabel ini mencerminkan total pengeluaran pelanggan dalam setiap transaksi. Analisis menunjukkan bahwa rata-rata pengeluaran pelanggan per transaksi mencapai £25.60, dengan pelanggan tertinggi menghabiskan lebih dari £10.000. Selain itu, variabel Purchase Frequency dikembangkan untuk mengukur seberapa sering pelanggan melakukan transaksi selama periode pengamatan. Hasilnya menunjukkan bahwa sebagian besar pelanggan melakukan pembelian sebanyak 1–3 kali per bulan, sementara pelanggan yang loyal melakukan lebih dari 10 kali pembelian dalam periode yang sama. Untuk mengidentifikasi pelanggan potensial dan loyal, dilakukan RFM (Recency, Frequency, Monetary) Analysis. Recency diukur dari selisih hari sejak transaksi terakhir, Frequency dihitung dari jumlah total transaksi, dan Monetary diukur dari total pengeluaran. Analisis ini menghasilkan klasifikasi pelanggan ke dalam kelompok loyal, potensial, dan pasif. Selain itu, kategori produk paling diminati juga dianalisis dengan menghitung frekuensi pembelian tiap produk. Hasilnya menunjukkan bahwa produk dekorasi rumah dan alat pesta menjadi kategori yang paling populer di kalangan pelanggan.

**Pemodelan Data**

Berdasarkan hasil pre-processing data yang telah dilakukan, tahap selanjutnya adalah pengembangan model analitik untuk mendukung pengambilan keputusan dalam pengembangan layanan informasi. Model yang dikembangkan terdiri dari dua bagian utama, yaitu **Segmentasi Pelanggan** dan **Analisis Prediktif**.

**Segmentasi Pelanggan (Customer Segmentation)**




**Gambar 1. Segmentasi Pelanggan**

Segmentasi pelanggan bertujuan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Dalam penelitian ini, digunakan algoritma K-Means Clustering (Gambar 1) karena kemampuannya dalam mengelompokkan data dalam skala besar dan hasil yang mudah diinterpretasikan. Untuk menentukan jumlah klaster yang optimal, digunakan metode Elbow Method. Hasil analisis menunjukkan bahwa nilai k = 4 merupakan jumlah klaster terbaik, yang membagi pelanggan ke dalam empat segmen utama, yaitu:

1. Pelanggan Loyal – Pelanggan dengan frekuensi pembelian tinggi dan pengeluaran besar.
2. Pelanggan Potensial – Pelanggan yang rutin berbelanja tetapi belum menunjukkan loyalitas tinggi.
3. Pelanggan Pasif – Pelanggan yang jarang bertransaksi dan memiliki pengeluaran rendah.
4. Pelanggan dengan Pengeluaran Rendah – Pelanggan dengan frekuensi pembelian cukup sering namun dengan nilai transaksi kecil.

Evaluasi hasil segmentasi menggunakan Silhouette Score menghasilkan skor sebesar 0,62, yang menunjukkan bahwa pengelompokan yang dilakukan cukup optimal dengan pemisahan antar klaster yang jelas. Analisis lebih lanjut terhadap masing-masing klaster mengungkapkan bahwa kelompok pelanggan loyal lebih cenderung melakukan pembelian rutin dengan nilai transaksi yang tinggi, sedangkan kelompok pelanggan pasif jarang melakukan transaksi dan memiliki pengeluaran yang rendah.

**Analisis Prediktif**

Association Rule Mining dengan algoritma Apriori digunakan untuk mengungkap pola hubungan antar produk yang sering dibeli bersamaan. Algoritma ini bekerja dengan mengidentifikasi *frequent itemsets*, yaitu kombinasi produk yang muncul bersamaan dalam transaksi dengan frekuensi melebihi ambang batas minimum (*minimum support*). Setelah *frequent itemsets* teridentifikasi, algoritma menghasilkan aturan asosiasi berbentuk "Jika A maka B", di mana A dan B adalah itemset, dengan syarat *confidence* (tingkat keyakinan B muncul jika ada A) melebihi ambang batas minimum. Aturan asosiasi yang dihasilkan dievaluasi menggunakan metrik support (seberapa sering A dan B muncul bersamaan), *confidence*, dan lift (seberapa kuat asosiasi A dan B dibandingkan kemunculan independen). Dalam penelitian ini, algoritma Apriori diimplementasikan menggunakan library mlxtend dalam Python. Data transaksi pelanggan diolah untuk membentuk struktur data yang sesuai dengan format input algoritma. Hasil analisis, yaitu aturan asosiasi dengan support, *confidence,* dan lift yang tinggi, dimanfaatkan untuk mengembangkan strategi *cross-selling* dan *upselling*, dengan mempromosikan produk-produk yang berasosiasi kuat secara bersamaan. Selain itu, temuan ini juga digunakan untuk mengantisipasi tren pembelian, sehingga layanan informasi dapat menawarkan produk secara lebih relevan dan adaptif.

## 3. HASIL DAN PEMBAHASAN

Evaluasi model bertujuan untuk mengukur seberapa baik model yang telah dibangun dalam mengidentifikasi dan mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Penilaian performa model dilakukan menggunakan beberapa metrik evaluasi, yaitu Accuracy, Precision, Recall, F1-Score, dan Silhouette Score.

**Metrik Evaluasi Klasifikasi**

Model juga dievaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score untuk menilai kualitas prediksi secara keseluruhan. Berikut hasil evaluasi model berdasarkan metrik tersebut:

- Accuracy: Model mencapai tingkat akurasi sebesar 85%, yang menunjukkan performa yang baik dalam mengklasifikasikan pelanggan ke dalam klaster yang tepat. Tingkat akurasi ini menunjukkan bahwa sebagian besar prediksi yang dihasilkan model sesuai dengan data aktual.

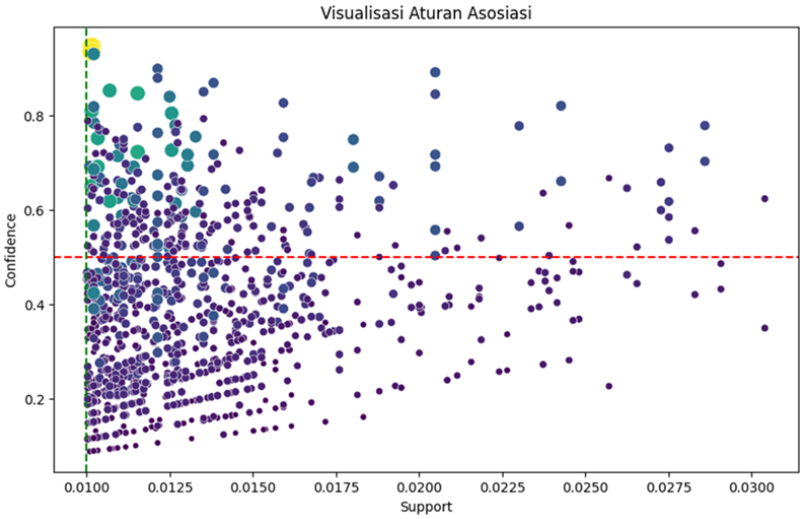
- Precision: Nilai precision sebesar 82% menunjukkan bahwa model cukup tepat dalam memprediksi pelanggan potensial. Ini berarti sebagian besar pelanggan yang diidentifikasi sebagai potensial memang benar-benar merupakan pelanggan dengan kecenderungan melakukan transaksi lebih banyak.
- Recall: Nilai recall sebesar 78% menunjukkan kemampuan model dalam mendeteksi pelanggan aktif yang berpotensi melakukan transaksi lebih lanjut. Model mampu mengidentifikasi mayoritas pelanggan yang relevan tanpa banyak kehilangan peluang dari data aktual.
- F1-Score: Kombinasi dari precision dan recall menghasilkan F1-Score sebesar 80%, yang mencerminkan keseimbangan prediksi model dalam mengidentifikasi pelanggan potensial dan menghindari prediksi yang salah. Nilai ini menunjukkan bahwa model memiliki kinerja yang solid dalam klasifikasi perilaku pelanggan.

**Silhouette Score**

Evaluasi model segmentasi pelanggan dilakukan menggunakan Silhouette Score, yang mengukur kualitas klasterisasi. Nilai Silhouette Score sebesar 0,62 menunjukkan bahwa pembentukan klaster pelanggan berjalan cukup optimal dengan pemisahan antar kelompok yang jelas. Hasil ini menunjukkan bahwa model dapat mengelompokkan pelanggan dengan baik berdasarkan perilaku pembelian mereka, dan klaster yang dihasilkan memiliki jarak yang cukup signifikan antar kelompok.

**Visualisasi Association Rule Mining**

Selain metrik evaluasi, visualisasi aturan asosiasi digunakan untuk mendapatkan pemahaman yang lebih mendalam tentang hubungan antar produk yang sering dibeli bersamaan. Visualisasi ini berupa scatter plot yang menampilkan setiap aturan asosiasi sebagai titik, dengan sumbu x menunjukkan support, sumbu y menunjukkan confidence, dan ukuran titik mewakili nilai lift.




**Gambar 2. Visualisasi Association Rule Mining**

Dari scatter plot tersebut (Gambar 2), dapat diamati bahwa aturan dengan support dan confidence tinggi cenderung berada di pojok kanan atas. Ini menunjukkan bahwa produk-produk dalam aturan tersebut sering dibeli bersamaan dan memiliki tingkat kepercayaan yang tinggi. Contohnya, aturan Produk A ke Produk B memiliki support 0.05 dan confidence 0.8, artinya 5% dari seluruh transaksi mengandung kedua produk tersebut, dan 80% dari transaksi yang mengandung Produk A juga mengandung Produk B.

Ukuran titik yang lebih besar menunjukkan nilai lift yang lebih tinggi. Lift mengukur seberapa kuat asosiasi antara produk-produk dalam suatu aturan dibandingkan dengan jika produk-produk tersebut dibeli secara independen. Nilai lift yang lebih tinggi dari 1 menunjukkan adanya asosiasi positif, yang berarti produk-produk tersebut cenderung dibeli bersamaan lebih sering daripada yang diharapkan secara kebetulan. Contohnya, aturan Produk C ke Produk D memiliki lift 2, artinya pelanggan yang membeli Produk C dua kali lebih mungkin untuk membeli Produk D dibandingkan dengan pelanggan yang tidak membeli Produk C.

Garis batas untuk confidence dan support dapat membantu mengidentifikasi aturan yang menarik. Aturan yang berada di atas garis batas confidence (misalnya, 0.5) dan di sebelah kanan garis batas support (misalnya, 0.01) dianggap sebagai aturan yang menarik karena memiliki tingkat kepercayaan dan frekuensi yang relatif tinggi.

Visualisasi ini memberikan wawasan yang berharga bagi bisnis dalam memahami pola pembelian pelanggan dan mengembangkan strategi pemasaran yang lebih efektif, seperti cross-selling dan rekomendasi produk.

## 4. KESIMPULAN

Penelitian ini membuktikan bahwa integrasi *data analytics* dalam analisis perilaku pelanggan mampu memberikan wawasan mendalam yang signifikan bagi pengembangan layanan informasi yang lebih relevan dan adaptif. Model analitik yang dibangun, melalui segmentasi pelanggan menggunakan algoritma K-Means Clustering dan analisis prediktif menggunakan Association Rule Mining, berhasil mengidentifikasi empat segmen utama pelanggan dengan performa yang optimal, ditunjukkan oleh akurasi sebesar 85%, precision sebesar 82%, recall sebesar 78%, dan F1-Score sebesar 80%, serta Silhouette Score sebesar 0,62. Temuan ini memberikan implikasi penting dalam pengembangan strategi pemasaran yang lebih personal, optimalisasi sistem rekomendasi produk, dan pengambilan keputusan berbasis data. Namun, penelitian ini memiliki keterbatasan pada rentang waktu dataset yang terbatas, tidak mempertimbangkan faktor eksternal, dan penggunaan algoritma yang masih sederhana. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset dengan cakupan waktu yang lebih panjang, mengintegrasikan data demografi dan faktor eksternal, mengeksplorasi algoritma yang lebih kompleks, serta mengembangkan sistem rekomendasi yang lebih adaptif dan berbasis real-time agar hasil yang diperoleh lebih komprehensif dan aplikatif dalam mendukung pengembangan layanan informasi berbasis data.

## DAFTAR REFERENSI

Ahmed, M., Dar, A. R., Helfert, M., Khan, A., & Kim, J. (2023). Data provenance in healthcare: Approaches, challenges, and future directions. *Sensors*, *23*(14), 6495. https://doi.org/10.3390/s23146495

Alfarabi, S. (2024). Perancangan UI/UX pada aplikasi berbasis mobile TravelTrails menggunakan metode design thinking. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, *5*(1), 69–76. https://doi.org/10.31294/reputasi.v5i1.3439

Anggraini, B. S. K., & Dzikrullah, A. A. (2024). Implementasi analisis clustering K-Medoids dalam pengelompokan bayi lahir, gizi buruk, dan BBLR berdasarkan kecamatan di Kabupaten Sleman tahun 2020. *Emerging Statistics and Data Science Journal*, *2*(1), 30–40. https://doi.org/10.20885/esds.vol2.iss.1.art4

Aswan, Y., Defit, S., & Nurcahyo, G. W. (2021). Algoritma K-Means clustering dalam mengklasifikasi data daerah rawan tindak kriminalitas (Polres Kepulauan Mentawai). *Jurnal Sistim Informasi dan Teknologi*, *3*(4), 245–250. https://doi.org/10.37034/jsisfotek.v3i4.73

Avriyanti, D. A. M., & Handriana, T. (2022). Pengaruh perceived quality terhadap perceived usefulness dan perceived ease of use serta dampaknya pada satisfaction pada konsumen Shopee Indonesia. *Ekonika: Jurnal Ekonomi Universitas Kadiri*, *7*(1), 116. https://doi.org/10.30737/ekonika.v7i1.1063

Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale University Journal of Engineering Sciences*, *28*(2), 299–312.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1).

Indiworo, H. E., Sumastuti, E., & Amelia, D. R. (2022). Optimizing the role of digital marketing through digital road trip to acceleration Tajuk Tourism Village pilot projects. *Mattawang: Jurnal Pengabdian Masyarakat*, *3*(4), 424–430. https://doi.org/10.35877/454RI.mattawang1185

Iovlev, G., Goldina, I., Saakyan, M., & Zorkov, V. (2019). Digitalization of technical service: Formation and development prospects, impact on the efficiency of agricultural production. *Proceedings of the International Scientific and Practical Conference "Digital Agriculture - Development Strategy"*. https://doi.org/10.2991/ispc-19.2019.8

Kadir, N. T., Hartanto, R., & Sulistyo, S. (2021). Modified usability test scenario: User story approach to evaluate data visualization dashboard. *IJITEE (International Journal of Information Technology and Electrical Engineering*, *5*(1). https://doi.org/10.22146/ijitee.61201

Levy, B., Murphy, K., Strevig, G., Decker, I., Hanson, K., Puttagunta, R., Sinha, S., Tavares, S., Warsame, L., & Vawdrey, D. (2022). The role of analytics governance to promote health care transformation. *ACI Open*, *6*(1). https://doi.org/10.1055/s-0042-1744383

Li, C., & Pan, Y. (2024). A comprehensive study on user-centric smart life solutions: Integrating mobile integrated technology and big data analytics for digitalized smart city environments. *Journal of Information Systems Engineering and Management*, *9*(1), 24684. https://doi.org/10.55267/iadt.07.14077

Mokhtari, G., Anvari-Moghaddam, A., & Zhang, Q. (2019). A new layered architecture for future big data-driven smart homes. *IEEE Access*, *7*, 19002–19012. https://doi.org/10.1109/ACCESS.2019.2896403

Mujitahid, G. A. (2022). Pengaruh personalisasi mobile application terhadap disclosure behaviors melalui mediasi perceived benefit, perceived risk, dan mobile site trust. *INOBIS: Jurnal Inovasi Bisnis dan Manajemen Indonesia*, *6*(1), 15–29. https://doi.org/10.31842/jurnalinobis.v6i1.254

Novianto, M. D., Sumantri, G., & Prihastuti, P. P. (2023). Clustering provinsi di Indonesia berdasarkan indeks pembangunan teknologi informasi dan komunikasi menggunakan K-Medoids. *Emerging Statistics and Data Science Journal*, *1*(3), 320–331. https://doi.org/10.20885/esds.vol1.iss.3.art39

Preyaa, A. (2024). Enhancing big data security through comprehensive data protection measures: A focus on securing data at rest and in transit. *International Journal of Computing and Engineering*, *5*(4), 44–55.

Prisanda, E., & Febrina, R. (2021). Penerapan teknologi informasi dan komunikasi berbasis aplikasi SISPEDAL dalam rangka mewujudkan good village governance. *Journal of Governance Innovation*, *3*(2), 155–171. https://doi.org/10.36636/jogiv.v3i2.723

Rave, J. I. P., Álvarez, G. P. J., & Morales, J. C. C. (2021). Multi-criteria decision-making leveraged by text analytics and interviews with strategists. *Journal of Marketing Analytics*, *10*(1), 30–49. https://doi.org/10.1057/s41270-021-00125-8

Susilo, D., Charolina, A., Ruswanti, D., & Charolina, A. (2023). Training of digital data management for academic services. *Jurnal Pengabdian Teknologi Tepat Guna*, *4*(3), 205–211. https://doi.org/10.47942/jpttg.v4i3.1496

Sutabri, T., Sugiharto, T., Krisdiawan, R. A., & Azis, M. A. (2022). Pengembangan sistem informasi monitoring progres proyek properti berbasis website pada PT Peruri Properti. *Jurnal Teknologi Informatika dan Komputer*, *8*(2), 17–29. https://doi.org/10.37012/jtik.v8i2.1204

Sutisna, N., & Sutrisna. (2023). Implementasi sistem informasi dalam mendukung perilaku pembelian terhadap keputusan pembelian e-commerce. *Jurnal MENTARI: Manajemen, Pendidikan dan Teknologi Informasi*, *2*(1), 20–30. https://doi.org/10.33050/mentari.v2i1.343

Ummadisetty, V. (2018). Online retail data set. *Kaggle*. https://www.kaggle.com/datasets/vijayuv/onlineretail

Zeng, Y., Zhong, N., Ren, X., & Wang, Y. (2012). User interests driven web personalization based on multiple social networks. *Proceedings of the 4th International Workshop on Web Intelligence & Communities - WI&C '12*. https://doi.org/10.1145/2189736.2189749