INOTEK, Vol. 9
ISSN: 2580-3336 (Print) / 2549-7952 (Online)
Url: https://proceeding.unpkediri.ac.id/index.php/inotek/

Penerapan Algoritma Apriori untuk Analisis Asosiasi Film pada Dataset MovieLens

[1*]Rizal Syihab Saputra Adam, [2]M. Abdilah Saputra, [3]Erna Daniati
[1-3]Sistem Informasi, Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri
E-mail : [1]rizalsyihab07@gmail.com , [2]mabdilahsaputra@gmail.com , [3]ernadaniati@unpkediri.ac.id
Penulis Korespondens : Erna Daniati

Abstrak—Algoritma Apriori terbukti efektif dalam mengungkap pola asosiasi pada data transaksi. Penelitian ini menerapkan algoritma tersebut pada dataset MovieLens 32M untuk mengidentifikasi asosiasi antara film berdasarkan perilaku tontonan pengguna. Dataset difilter agar hanya mencakup film yang telah ditonton minimal oleh 500 pengguna dan pengguna yang menonton minimal 50 film, dengan sampel 5.000 pengguna aktif. Analisis dilakukan menggunakan Python di Google Colab dengan parameter minimum support 0,06 dan maksimal panjang itemset dua. Evaluasi menggunakan metrik lift mengungkap adanya asosiasi kuat antar film, terutama pada film yang tergabung dalam waralaba atau memiliki keterkaitan naratif. Hasil penelitian ini memberikan wawasan baru untuk pengembangan sistem rekomendasi film yang lebih cerdas dan personal, yang dapat meningkatkan pengalaman pengguna di platform streaming.
Kata Kunci—Apriori; data mining; asosiasi film; MovieLens; rekomendasi

Abstract*—The Apriori algorithm has proven effective in uncovering association patterns in transactional data. This study applies the algorithm to the MovieLens 32M dataset to identify film associations based on users' viewing behavior. The dataset was filtered to include films viewed by at least 500 users and users who have watched a minimum of 50 films, with a sample of 5,000 active users. The analysis was conducted using Python on Google Colab with a minimum support of 0.06 and a maximum itemset length of two. Evaluation with the lift metric revealed strong associations between films, particularly for those within a franchise or with related narratives. The findings offer new insights for developing smarter, personalized film recommendation systems that enhance user experiences on streaming platforms.*
Keywords—*Apriori; data mining; film association; MovieLens; recommendation*

I. PENDAHULUAN

Perkembangan teknologi informasi telah memungkinkan tersedianya data dalam jumlah besar yang tersebar di berbagai domain, termasuk industri hiburan seperti film. Salah satu tantangan utama dalam pengelolaan data tersebut adalah bagaimana mengekstrak informasi dan pola tersembunyi yang bermanfaat bagi pengguna. Dalam konteks sistem rekomendasi film, proses ini sangat penting untuk meningkatkan kepuasan pengguna melalui penyajian film-film yang relevan dengan preferensi mereka.

Dalam konteks analisis film, algoritma Apriori dapat dimanfaatkan untuk mengidentifikasi film-film yang sering ditonton bersamaan atau memiliki hubungan yang signifikan berdasarkan rating pengguna. Beberapa penelitian telah mencoba mengembangkan atau memodifikasi algoritma Apriori untuk meningkatkan performanya. Misalnya, Lu et al. dalam

penelitiannya menyajikan pendekatan pengoptimalan terhadap algoritma Apriori untuk mempercepat proses pencarian frequent itemsets [1]. Penelitian lainnya oleh Zubi, Elrowayati, dan Abu Fanas juga menunjukkan bahwa dengan rekomendasi pencarian film dengan menggunakan aturan asosiasi, efisiensi pencarian aturan asosiasi dapat ditingkatkan [2].

Data mining merupakan proses untuk menemukan pola tersembunyi dalam kumpulan data yang besar, guna mendukung pengambilan keputusan berbasis data. Salah satu metode data mining yang umum digunakan adalah association rule mining, yaitu pencarian hubungan antar item dalam suatu kumpulan data. Association rule berbentuk aturan if–then yang menyatakan bahwa jika suatu item muncul, maka item lain cenderung ikut muncul dalam transaksi yang sama[3]. Metode ini telah digunakan secara luas dalam berbagai domain, seperti analisis keranjang belanja, rekomendasi produk, dan perilaku pengguna.

Dalam hal ini, pemahaman terhadap perilaku pengguna dalam menonton film menjadi penting untuk meningkatkan sistem rekomendasi yang relevan dan personal. Salah satu pendekatan yang dapat digunakan untuk menggali pola ini adalah dengan menerapkan algoritma Apriori terhadap dataset historis tontonan pengguna[4].

Meskipun telah banyak penelitian yang menerapkan algoritma Apriori untuk analisis data transaksi seperti data belanja konsumen atau data sistem penjualan, penerapannya dalam domain film, khususnya menggunakan dataset MovieLens, masih memiliki ruang eksplorasi lebih lanjut. Salah satu kesenjangan yang diidentifikasi adalah minimnya studi yang secara spesifik memanfaatkan struktur data rating pengguna sebagai transaksi untuk menghasilkan asosiasi antar film.

Penelitian ini bertujuan untuk menerapkan algoritma Apriori pada dataset MovieLens 32M, yang berisi lebih dari 32 juta rating terhadap lebih dari 87.000 film oleh sekitar 200.000 pengguna. Untuk mengevaluasi kekuatan aturan asosiasi antar film, digunakan metrik lift, yang memberikan informasi seberapa besar dua film berkaitan melebihi kemungkinan kemunculan acak[5]. Pendekatan ini diimplementasikan menggunakan Python di Google Colab, dengan strategi optimasi seperti filtering user aktif dan film populer untuk menghindari kendala memori.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan tentang hubungan antar film yang sering ditonton bersama, yang selanjutnya dapat dimanfaatkan untuk mendukung sistem rekomendasi film berbasis asosiasi[6].

II. METODE

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari MovieLens versi 25M+ atau dikenal sebagai MovieLens 32M, yang dapat diakses melalui situs resmi GroupLens (https://grouplens.org/datasets/movielens/). Dataset ini dikumpulkan pada Oktober 2023 dan dirilis Mei 2024 yang terdiri atas lebih dari 32 juta rating yang diberikan oleh sekitar 200.000 pengguna terhadap lebih dari 87.000 film. Dua file utama yang digunakan adalah:

- ratings.csv, yang berisi userId, movieId, rating, dan timestamp;
- movies.csv, yang berisi movieId, title, dan genres.

B. Algoritma Apriori

Algoritma Apriori merupakan algoritma fundamental dalam frequent itemset mining yang bertujuan menemukan pola kombinasi item (itemset) yang sering muncul dalam suatu kumpulan data. Algoritma ini digunakan dalam proses association rule mining untuk menghasilkan aturan asosiasi dalam bentuk A → B, yang berarti jika A terjadi, maka B juga cenderung terjadi.

Apriori bekerja dengan menghitung nilai support dari setiap kombinasi item, lalu membentuk aturan asosiasi berdasarkan nilai confidence dan lift. Dalam penelitian ini, metrik yang digunakan untuk mengevaluasi aturan adalah lift, yang lebih tepat untuk menunjukkan kekuatan hubungan dibandingkan confidence saja. Dalam berbagai domain, seperti bidang kesehatan dan retail, algoritma Apriori telah digunakan untuk menemukan hubungan tersembunyi dalam data transaksi penjualan[7].

C. *Google Colab*

Seluruh proses analisis dilakukan menggunakan Google Colaboratory (Colab), sebuah layanan cloud gratis dari Google yang menyediakan akses ke komputasi berbasis Python, termasuk pustaka data mining seperti mlxtend dan pandas[8]. Colab dipilih karena kemudahan integrasi dengan Google Drive, serta kemampuannya menjalankan skrip berskala menengah tanpa memerlukan komputer lokal dengan spesifikasi tinggi.

D. Data *Preprocessing*

Tahap preprocessing dilakukan untuk menjaga efisiensi komputasi dan menghindari konsumsi memori yang berlebihan. Tahap-tahapnya meliputi[9]:

- Menggabungkan file ratings.csv dan movies.csv berdasarkan movieId.
- Menyaring hanya film yang telah ditonton oleh minimal 500 pengguna.
- Menyaring hanya pengguna yang telah menonton minimal 50 film.
- Mengambil sampel acak sebanyak 5.000 user aktif.
- Membentuk data transaksi dalam format userId → daftar film yang ditonton.

Setelah transaksi terbentuk, dilakukan one-hot encoding menggunakan TransactionEncoder dari pustaka mlxtend untuk mengubah data ke bentuk matriks boolean.

E. *Association Rule*

Setelah data dalam bentuk one-hot matrix, algoritma Apriori diterapkan dengan parameter sebagai berikut:

- min_support = 0.06
- max_len = 2 (hanya asosiasi dua item)
- metrik evaluasi: lift (min_threshold = 1.0)

Lift digunakan untuk mengevaluasi kekuatan hubungan antara dua film[10], [11]. Semakin tinggi nilai lift, semakin kuat keterkaitan antara antecedent dan consequent dalam aturan tersebut.

Rumus-rumus yang digunakan[12]:

1. Support

Tahapan ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut.

$$Support\ (A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ A}{Jumlah\ Total\ Transaksi}$$

2. Confidence
   Menunjukkan seberapa besar kemungkinan B terjadi jika A terjadi.

$$Confidence\ (A \rightarrow B) = \frac{Support(A \cup B)}{Support\ (A)}$$

3. Lift
   Mengukur kekuatan aturan dibandingkan dengan independensi antar item.

$$Lift\ (A \rightarrow B) = \frac{Confidence\ (A \cup B)}{Support\ (B)}$$

III. HASIL DAN PEMBAHASAN

A. Proses Analisis Data

Penelitian ini menerapkan algoritma Apriori untuk menemukan pola asosiasi antar film yang sering ditonton bersama oleh pengguna[13], menggunakan dataset MovieLens 32M. Dataset ini mencakup lebih dari 32 juta rating dari sekitar 200.000 pengguna terhadap lebih dari 87.000 film.Untuk menjaga efisiensi pemrosesan dan menghindari kelebihan penggunaan memori (RAM), dilakukan tahapan preprocessing sebagai berikut:

- Hanya film yang ditonton oleh minimal 500 pengguna yang disertakan.
- Hanya pengguna yang telah menonton minimal 50 film yang disertakan.
- Dari hasil penyaringan, diambil sampel sebanyak 5.000 pengguna secara acak untuk membentuk transaksi user → daftar film.

Transaksi tersebut kemudian dikonversi menjadi matriks one-hot encoding menggunakan TransactionEncoder, dan algoritma Apriori diterapkan dengan parameter sebagai berikut[14], [15]:

- min_support = 0.06 (aturan hanya dibentuk dari kombinasi film yang muncul di ≥6% transaksi)
- max_len = 2 (hanya aturan asosiasi antara 2 film)
- metric asosiasi = lift (min_threshold = 1.0)

Lift dipilih karena mampu menunjukkan kekuatan asosiasi antara dua film dibandingkan kemunculan acak, sehingga memberikan nilai yang lebih bermakna secara statistik.

B. Hasil Aturan Asosiasi Film

Dari proses Apriori yang dilakukan, ditemukan 50 aturan asosiasi film dengan nilai lift ≥ 6.7. Aturan-aturan ini menunjukkan bahwa jika pengguna menonton film A, maka

kemungkinan besar mereka juga menonton film B, dan hubungan tersebut jauh lebih kuat dibandingkan kemunculan acak[14].

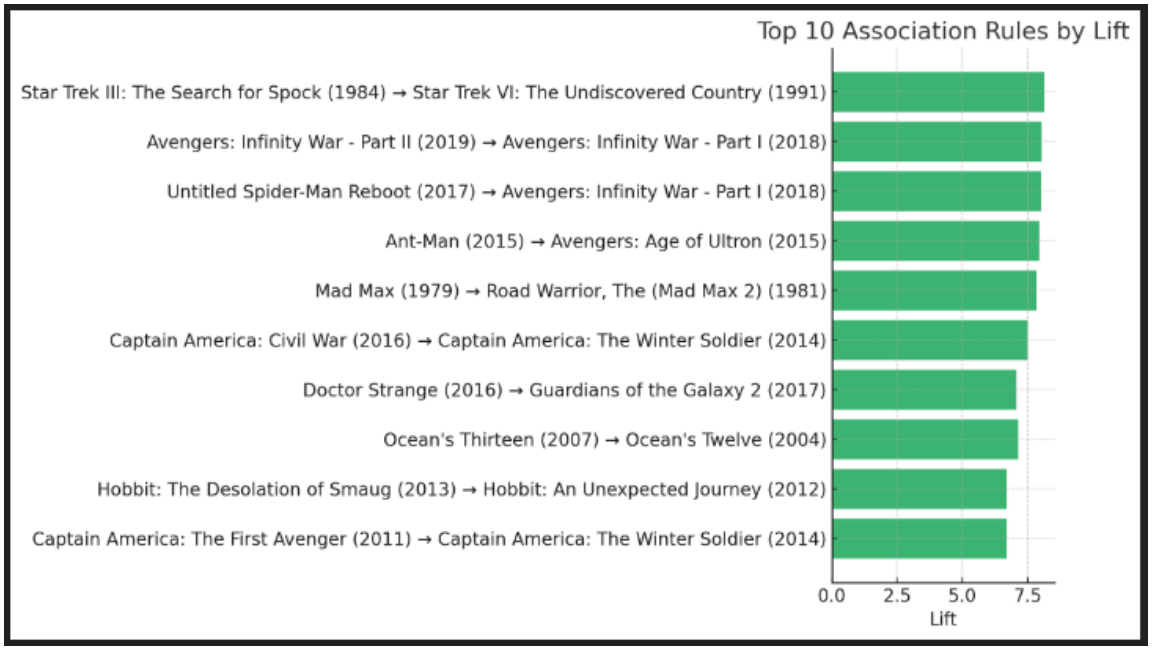
Beberapa aturan teratas ditampilkan dalam **Tabel 1** berikut:

Tabel 1 – Aturan Asosiasi Film Berdasarkan Lift Tertinggi

No	Antecedent (Jika Menonton)	Consequent (Maka Kemungkinan Menonton)	Support (%)	Confidence (%)	Lift
1	Star Trek III: The Search for Spock (1984)	Star Trek VI: The Undiscovered Country (1991)	6.26	71.14	8.14
2	Avengers: Infinity War - Part II (2019)	Avengers: Infinity War - Part I (2018)	7.38	85.81	8.04
3	Untitled Spider-Man Reboot (2017)	Avengers: Infinity War - Part I (2018)	6.02	85.51	8.01
4	Ant-Man (2015)	Avengers: Age of Ultron (2015)	6.68	76.78	7.96
5	Mad Max (1979)	Road Warrior, The (Mad Max 2) (1981)	6.56	71.77	7.85
6	Captain America: Civil War (2016)	Captain America: The Winter Soldier (2014)	7.00	80.83	7.51
7	Doctor Strange (2016)	Guardians of the Galaxy 2 (2017)	7.00	74.95	7.08
8	Ocean's Thirteen (2007)	Ocean's Twelve (2004)	6.00	74.44	7.14
9	Hobbit: The Desolation of Smaug (2013)	Hobbit: An Unexpected Journey (2012)	7.78	81.21	6.71
10	Captain America: The First Avenger (2011)	Captain America: The Winter Soldier (2014)	7.88	72.16	6.71

Hasil di atas menunjukkan bahwa film-film yang memiliki hubungan naratif (sekuel, prekuel, atau dalam satu waralaba) cenderung memiliki asosiasi yang sangat kuat. Hal ini terlihat dari nilai lift yang jauh di atas 1, bahkan mencapai lebih dari 8. Artinya, peluang penonton menonton kedua film tersebut bersamaan hingga 8 kali lebih besar dibandingkan jika tidak ada hubungan[16].

Untuk memperjelas pola asosiasi film yang ditemukan, dilakukan visualisasi terhadap 10 aturan asosiasi dengan nilai lift tertinggi. **Gambar 1** memperlihatkan kekuatan hubungan antar film dalam bentuk diagram batang horizontal.



Gambar 1 – Visualisasi Top 10 Aturan Asosiasi Berdasarkan Lift

Grafik menunjukkan bahwa film Star Trek III: The Search for Spock (1984) memiliki asosiasi paling kuat dengan Star Trek VI: The Undiscovered Country (1991) dengan nilai lift sebesar 8.14. Selanjutnya, beberapa pasangan film dari Marvel seperti Avengers, Captain America, Ant-Man, dan Doctor Strange juga menempati posisi teratas.

C. Interpretasi Hasil

Secara umum, aturan asosiasi yang dihasilkan dari algoritma Apriori dengan metrik lift menggambarkan konsistensi perilaku pengguna dalam menonton film dalam satu rangkaian cerita atau universe. Sebagai contoh:

- Penonton Avengers: Infinity War - Part I (2018) sangat besar kemungkinan juga menonton Part II (2019), dan sebaliknya.
- Hubungan antar film Star Trek atau Captain America juga sangat kuat.
- Bahkan film dari universe berbeda seperti Mad Max, Hobbit, dan Ocean's Series menunjukkan asosiasi kuat secara internal.

Pemilihan metrik lift terbukti efektif untuk menilai hubungan yang tidak hanya sering terjadi, tetapi juga "lebih kuat dari ekspektasi acak". Jika hanya menggunakan confidence, aturan populer dengan film "pasaran" bisa bias — namun lift membantu menormalkan pengaruh popularitas.

D. Kesimpulan Sementara

Berdasarkan hasil yang diperoleh:

- Apriori dengan lift berhasil mengidentifikasi asosiasi film yang kuat dan relevan.
- Film dalam satu waralaba (MCU, Star Trek, Mad Max, dll.) sangat mendominasi aturan asosiasi terkuat[17].
- Strategi filtering pengguna dan film aktif, serta penggunaan metrik lift, membantu menghindari crash memori sekaligus menghasilkan aturan yang bermakna.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Apriori dapat diterapkan secara efektif untuk menemukan pola asosiasi antar film dalam dataset berskala besar seperti MovieLens 32M. Dengan menerapkan proses preprocessing yang ketat—yaitu menyaring film yang ditonton oleh minimal 500 pengguna dan pengguna yang telah menonton minimal 50 film—serta menggunakan teknik one-hot encoding dan pemilihan metrik lift, dihasilkan aturan-aturan asosiasi yang relevan dan bermakna secara statistik.

Hasil eksperimen mengungkapkan bahwa film-film dalam satu waralaba atau yang memiliki kesinambungan cerita, seperti Marvel Cinematic Universe, Star Trek, The Hobbit, dan Ocean's Series, cenderung memiliki keterkaitan kuat dalam perilaku tontonan pengguna. Hal ini dibuktikan dengan nilai lift tinggi yang menunjukkan kekuatan asosiasi yang signifikan, bahkan hingga delapan kali lebih besar dibandingkan kemungkinan acak.

Penggunaan Google Colab sebagai platform komputasi terbukti mampu mendukung analisis berskala menengah tanpa kendala teknis berarti. Strategi filtering dan pemilihan metrik yang tepat juga berhasil menghindari beban komputasi berlebih sekaligus menjaga kualitas hasil.

Dengan demikian, temuan ini tidak hanya memberikan gambaran pola perilaku pengguna dalam menonton film, tetapi juga memiliki implikasi praktis dalam pengembangan sistem rekomendasi film yang lebih personal, cerdas, dan kontekstual. Pendekatan berbasis asosiasi seperti ini dapat menjadi alternatif atau pelengkap dari model rekomendasi berbasis pembelajaran mesin yang lebih kompleks.

UCAPAN TERIMAKASIH

Terima kasih khusus kami sampaikan kepada instansi, perusahaan, dan lembaga yang memberikan dukungan finansial, serta kepada Ibu Erna Daniati, M.Kom selaku dosen Metodologi Penelitian atas bimbingan dan arahannya. Kami juga menghargai dukungan akademik dan fasilitas dari Universitas Nusantara PGRI Kediri, khususnya Program Studi Sistem Informasi. Ucapan terima kasih juga kami tujukan kepada semua yang telah membantu, baik secara langsung maupun tidak langsung. Semoga penelitian ini bermanfaat dan menjadi referensi di bidang analisis data dan teknologi informasi.

DAFTAR PUSTAKA

[1] S. Tirumalasetty, A. Jadda, and S. Reddy Edara, "An Enhanced Apriori Algorithm for Discovering Frequent Patterns with Optimal Number of Scans." doi: https://doi.org/10.48550/arXiv.1506.07087.

[2] Z. S. Zubi, A. A. Elrowayati, and I. S. Abu Fanas, "A Movie Recommendation System Design Using Association Rules Mining and Classification Techniques," *WSEAS TRANSACTIONS ON COMPUTERS*, vol. 21, pp. 189–199, Jun. 2022, doi: 10.37394/23205.2022.21.24.

[3] D. Listriani, A. H. Setyaningrum, and F. Eka, "PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN (Studi Kasus Toko Buku Gramedia Bintaro)," vol. 9, no. 2.

[4] A. R. Riszky and M. Sadikin, "Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 3, pp. 103–108, Jul. 2019, doi: 10.14710/jtsiskom.7.3.2019.103-108.

[5] Y. Ji, A. Sun, J. Zhang, and C. Li, "A Re-visit of the Popularity Baseline in Recommender Systems," in *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, Inc, Jul. 2020, pp. 1749–1752. doi: 10.1145/3397271.3401233.

[6] T. Tran and K. Lee, "Regularizing matrix factorization with user and item embeddings for recommendation," in *International Conference on Information and Knowledge Management, Proceedings*, Association for Computing Machinery, Oct. 2018, pp. 687–696. doi: 10.1145/3269206.3271730.

[7] E. N. Salamah and N. Ulinnnuha, "Analisis Pola Pembelian Obat dan Alat Kesehatan di Klinik Ibu dan Anak Graha Amani dengan Menggunakan Algoritma Apriori," *Jurnal INFORM*, vol. xx No.xx.

[8] A. Fauzi, A. B. H. Yanto, and N. Indriyani, "Data Mining Implementation of Organic Vegetable Sales Using Apriori Algorithm," *JURNAL TEKNOLOGI DAN OPEN SOURCE*, pp. 98–109, Jun. 2023, doi: 10.36378/jtos.v6i1.3049.

[9] R. Candra Purnama Putra, K. Fahrizal Dzatama, A. Tristan Syafa, E. Daniati, A. Ristyawan, and N. PGRI Kediri, "Analisis Kombinasi Produk Bakery Menggunakan Apriori FP-Growth Penulis Korespondensi: Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 423," Online, 2024.

[10] D. Dwiputra, A. Mulyo Widodo, H. Akbar, and G. Firmansyah, "Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis to Discover Rules of Item Combinations," *Journal of World Science*, vol. 2, no. 8, pp. 1229–1248, Aug. 2023, doi: 10.58344/jws.v2i8.403.

[11] A. Hermawan, "SMART PRODUCT RECOMMENDATIONS IN THE E-COMMERCE WEBSITE: UTILIZING THE APRIORI ALGORITHM FOR MARKET BASKET ANALYSIS," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems*, vol. x, No.x, pp. 1–5, doi: 10.22146/ijccs.xxxx.

[12] F. Harahap, N. E. Saragih, E. D. P. Situmeang, E. Tuti, E. Ginting, and W. Fahrozi, "Implementasi Data Mining dalam Memprediksi Stok Herbal menggunakan Algoritma Apriori," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 1159, Apr. 2022, doi: 10.30865/mib.v6i2.3937.

[13] "Implementasi Data Mining Menggunakan Algoritma".

[14] M. Ali Ridla, Fajriyanto, and Misbahul Marzuqi, "Implementasi Algoritma Apriori untuk Menentukan Pola Transaksi Penjualan Berbasis Web," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 5, no. 3, pp. 196–207, Sep. 2023, doi: 10.35746/jtim.v5i3.399.

[15] I. Riadi, H. Herman, F. Fitriah, S. Suprihatin, A. Muis, and M. Yunus, "Implementation of association rule using apriori algorithm and frequent pattern growth for inventory control," *JURNAL INFOTEL*, vol. 15, no. 4, pp. 369–378, Dec. 2023, doi: 10.20895/infotel.v15i4.980.

[16] O. Liansyah and H. Destiana, "The Use of Apriori Algorithm in the Formation of Association Rule at Lotteria Cibubur," *SinkrOn*, vol. 4, no. 2, p. 76, Mar. 2020, doi: 10.33395/sinkron.v4i2.10526.

[17] M. Fauzy, K. W. Rahmat Saleh, I. Asror, J. Telekomunikasi No, and T. Buah Batu Bandung, "PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN ALGORITMA APRIORI PADA SIMULASI PREDIKSI HUJAN WILAYAH KOTA BANDUNG," 2016.