

Analisis Strategi Pemasaran Menggunakan *K-Means* dan Algoritma Apriori (Studi Kasus: Edulink.id)

Dede Irsandi¹, Petrus Sokibi², Willy Eka Septian³

¹ e-mail: dede.irsandi.ti.21@cic.ac.id, ² petrus.sokibi@cic.ac.id, ³ willy.eka.septian@cic.ac.id

¹ Mahasiswa, Prodi S1 Teknik Informatika Universitas Catur Insan Cendekia,
^{2,3} Dosen, Prodi S1 Teknik Informatika Universitas Catur Insan Cendekia,

Abstract— Educational technology is rapidly evolving with the rise of various Education Technology (EdTech) platforms. Edulink.id, an EdTech startup based in Cirebon, offers School Management Information Systems and Learning Management Systems. Although there is significant potential, schools still face challenges in understanding the benefits of such technology. This study aims to develop a more effective marketing strategy using a data mining approach. The *K-Means* method is employed to classify schools into two categories: potential customers and non-potential ones. Subsequently, the Apriori algorithm is applied to identify characteristic patterns within the potential group. The research findings are expected to provide recommendations for suitable marketing strategies to enhance awareness, interest, and partnerships between schools and Edulink.id.

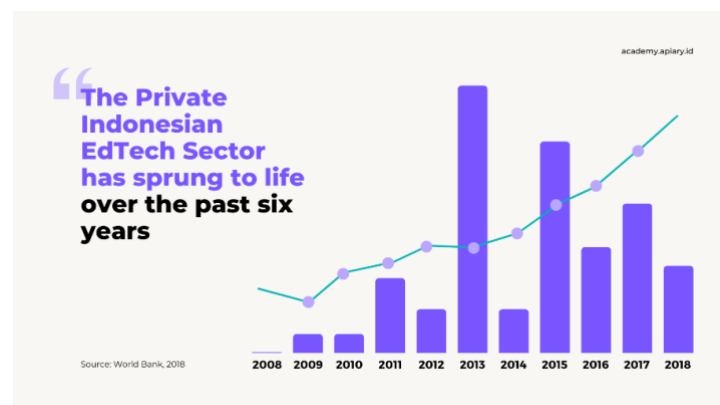
Intisari— Teknologi pendidikan berkembang pesat dengan munculnya berbagai platform Teknologi Pendidikan (EdTech). Edulink.id, sebuah startup EdTech yang berbasis di Cirebon, menawarkan Sistem Informasi Manajemen Sekolah dan Sistem Manajemen Pembelajaran. Meskipun terdapat potensi yang signifikan, sekolah masih menghadapi tantangan dalam memahami manfaat teknologi tersebut. Penelitian ini bertujuan untuk mengembangkan strategi pemasaran yang lebih efektif menggunakan pendekatan *data mining*. Metode *K-Means* digunakan untuk mengklasifikasikan sekolah menjadi dua kategori: calon pelanggan dan non-potensial. Selanjutnya, algoritma Apriori diterapkan untuk mengidentifikasi pola karakteristik dalam kelompok potensial. Temuan penelitian ini diharapkan dapat memberikan rekomendasi strategi pemasaran yang tepat untuk meningkatkan kesadaran, minat, dan kemitraan antara sekolah dan Edulink.id.

Kata Kunci— Strategi Pemasaran, Data Mining, Edulink.id, *K-Means Clustering*, Algoritma Apriori.

I. PENDAHULUAN

Kemajuan teknologi di bidang pendidikan terjadi dengan cepat berkat banyaknya platform *Education Technology* (EdTech). EdTech membuat pembelajaran lebih fleksibel dan mudah dengan menawarkan berbagai perangkat digital seperti kelas daring, Sistem Manajemen Pembelajaran (LMS), Sistem Manajemen Sekolah (SMS), Kursus Daring Terbuka Masif (MOOC), e-Learning, dan Sistem Informasi Manajemen Sekolah (SIMS).

Baik di dunia maupun di Indonesia, penggunaan teknologi pendidikan, atau EdTech, telah berkembang pesat, terutama sejak pandemi. Pandemi mendorong guru dan siswa untuk memanfaatkan perangkat pembelajaran daring dengan lebih baik. Jumlah orang yang menggunakan aplikasi EdTech meningkat dua kali lipat selama periode tersebut, terutama pada Maret 2020. Pada tahun 2021, sebuah studi pemerintah menemukan bahwa terdapat sekitar 210 *startup* EdTech di Indonesia [1]. Jumlah ini kemungkinan akan terus bertambah dalam sepuluh tahun ke depan. Bukti penelitian yang ditunjukkan pada grafik di bawah ini, pertumbuhan industri EdTech berkaitan dengan tingginya persentase penduduk di suatu negara atau wilayah yang memiliki akses internet.



Gambar 1.1. Grafik Perkembangan Sektor EdTech di Indonesia
Sumber gambar: <https://academy.apiary.id/blog/Perkembangan-Startup-Edukasi-Di-Indonesia>

Ini berarti orang-orang sangat antusias dengan metode pembelajaran baru ini. Banyak perusahaan dan situs web EdTech telah menyediakan berbagai hal seperti materi pembelajaran, program sekolah, dan perangkat pendidikan lainnya.

EduLink.id adalah perusahaan EdTech yang berbasis di Cirebon yang membantu sekolah-sekolah dengan Sistem Informasi Manajemen Sekolah, Sistem Manajemen Pembelajaran, dan solusi TI untuk membangun jaringan internet. Didirikan pada tahun 2020 oleh PT Alvarel Technology Innovation, perusahaan ini masih berupaya keras untuk mengembangkan basis pelanggannya. Meskipun perangkat-perangkat ini bermanfaat, tidak semua sekolah atau penyedia layanan pembelajaran ingin menggunakannya. Masalah besarnya adalah menemukan lebih banyak sekolah dan kelompok pendidikan yang bersedia menggunakan aplikasi ini dan bekerja sama dengan perusahaan.



Gambar 2. Logo EduLink.id

Pemasaran adalah tindakan yang dilakukan oleh perusahaan untuk meningkatkan kinerja bisnisnya. Ini karena pemasaran berhubungan langsung dengan konsumen. Pemasaran adalah tentang membuat dan memberi pelanggan potensial nilai unik untuk mendapatkan keunggulan kompetitif jangka panjang [2]. Selama beberapa dekade. Pemasaran selain sebagai sosialisasi tentang produk ke calon pelanggan juga sebagai evaluasi bisnis tentang standar kebutuhan pelanggan yang akan membantu sebagai landasan membuat keputusan bisnis yang strategis.

Berdasarkan kunjungan pertama tim pemasaran EduLink.id ke berbagai sekolah di Cirebon dan sekitarnya, termasuk SD, SMP, dan SMA, mereka berdiskusi dengan sekolah-sekolah tentang penggunaan layanan Sistem Informasi Manajemen Sekolah dan Sistem Manajemen Pembelajaran. Dari Januari 2022 hingga Februari 2025, dari 248 kunjungan sekolah, 151 sekolah (sekitar 60%) memutuskan untuk bekerja sama dengan EduLink.id. Data ini berasal dari upaya tim pemasaran selama tiga tahun. Sebagian sekolah masih ragu terhadap manfaat teknologi, keterbatasan infrastruktur TI, serta adanya persaingan dari platform lain yang lebih populer. Hal ini menunjukkan perlunya strategi pemasaran yang lebih tepat sasaran agar dapat meningkatkan kesadaran dan minat sekolah untuk bermitra. Berbagai studi sebelumnya banyak membahas adopsi EdTech secara umum, tetapi belum menekankan penggunaan analisis data untuk memahami karakteristik sekolah sebagai calon pelanggan. Di sinilah letak gap penelitian ini: bagaimana memanfaatkan data mining untuk menghasilkan strategi pemasaran berbasis pola nyata dalam data.

Teori pemasaran strategis membantu organisasi menemukan nilai, memenuhi kebutuhan pelanggan, dan berkomunikasi dengan jelas [4]. Pengambilan keputusan pemasaran perlu didasarkan pada fakta nyata dan riset yang terencana dengan baik. Dalam studi ini, data dikumpulkan melalui wawancara dan diskusi dengan sekolah-sekolah di Cirebon dan sekitarnya antara Januari 2022 dan Februari 2025. Data tersebut mencakup berbagai hal seperti minat sekolah dalam menggunakan aplikasi sekolah, status akreditasinya, ketersediaan perangkat digital, penggunaan sistem pembelajaran lain, dan permasalahan infrastruktur serta budaya sekolah. Informasi ini kemudian dianalisis menggunakan metode penambangan data untuk mengembangkan strategi pemasaran yang terfokus.

Data mining membantu menemukan pola tersembunyi dalam kumpulan data besar yang sulit dilihat [5][6][7][8]. Dalam studi ini, algoritma K-Means digunakan untuk membagi sekolah menjadi dua kelompok: sekolah yang kemungkinan besar tertarik dan sekolah yang tidak tertarik. Pengelompokan ini membantu memahami sekolah mana yang memiliki peluang lebih baik, seperti minat yang tinggi, jumlah siswa yang lebih banyak, atau infrastruktur yang lebih baik [9]. Kemudian, algoritma Apriori digunakan untuk menemukan bagaimana fitur-fitur sekolah yang berbeda saling terhubung. Koneksi-koneksi ini membantu menciptakan rencana pemasaran yang lebih sesuai dengan kebutuhan sekolah yang kemungkinan besar tertarik. Untuk melakukan analisis, data diolah menggunakan Python di Google Collaboratory, dengan alat-alat seperti Pandas, Scikit-Learn, dan MLxtend. Hasilnya ditampilkan menggunakan Looker Studio untuk membantu tim pemasaran memahaminya dengan lebih baik. Dengan metode ini, studi ini bertujuan untuk menciptakan strategi pemasaran yang akurat dan praktis berdasarkan data, yang akan membantu meningkatkan kesadaran, membuat penawaran lebih menarik, dan mendorong lebih banyak kerja sama tim antara sekolah dan EduLink.id..

Berdasarkan penelitian terdahulu, metode data mining seperti *K-Means* dan Apriori telah diterapkan dalam berbagai bidang untuk meningkatkan strategi pemasaran dan memahami pola perilaku konsumen. Dalam penelitian "Strategi *Marketing Promosi* Penerimaan Mahasiswa Baru Menggunakan *K-Means Clustering*" (2023), Penggunaan algoritma *K-Means Clustering* telah terbukti membantu menciptakan strategi pemasaran yang lebih terarah dengan menganalisis ciri-ciri calon mahasiswa, yang dapat menghasilkan lebih banyak pelamar [10]. Metode ini mengelompokkan mahasiswa baru ke dalam beberapa kategori. Hasilnya menunjukkan bahwa mahasiswa dapat dibagi menjadi tiga kelompok: mereka yang memiliki prestasi akademik tinggi, mereka yang terlibat dalam organisasi, dan mereka yang berasal dari keluarga berpenghasilan menengah ke atas. Pengelompokan ini dapat membantu universitas menciptakan promosi yang tepat sasaran dan sesuai dengan kebutuhan serta minat masing-masing kelompok. Jika sebuah universitas ingin meningkatkan jumlah pendaftar dan mencapai target pendaftarannya, universitas tersebut dapat menggunakan *K-Means Clustering* untuk merancang strategi pemasaran berdasarkan data penerimaan mahasiswa.

Selain itu, penelitian "Penerapan Metode *K-Means* Dan Apriori Untuk Pemilihan Produk *Bundling*" (2022) Studi ini menunjukkan bahwa menggabungkan metode *K-Means* dan Apriori dapat menghasilkan rekomendasi produk yang lebih

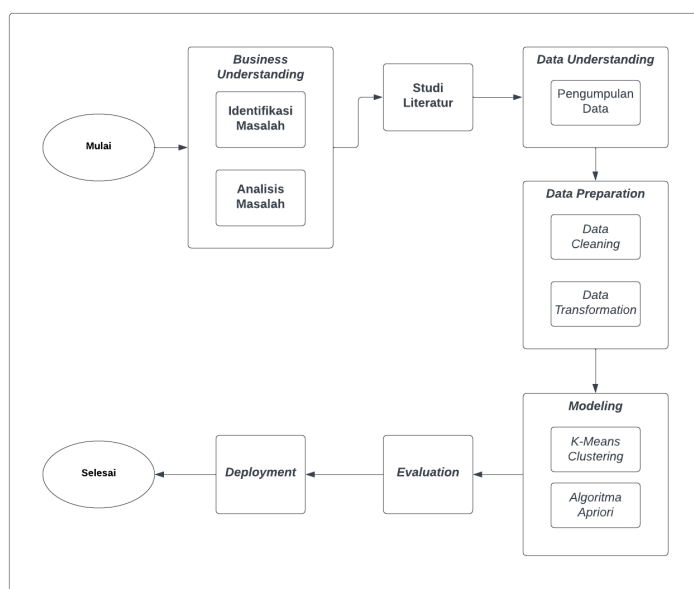
terstruktur. Dengan menganalisis pola pembelian, perusahaan dapat lebih memahami perilaku konsumen remaja, meningkatkan daya tarik mereka, dan mendorong penjualan melalui strategi bundling produk yang inovatif [11]. Dengan menggunakan hasil pengujian metode McCall, Cara penggunaan metode *K-Means* dan Apriori untuk memilih produk bundel menunjukkan bahwa sistem dapat memenuhi kebutuhan pengguna, dengan skor kinerja sebesar 83,58% yang dianggap baik.

Penelitian lainnya, "Penerapan Algoritma *K-Medoids* dan Algoritma Apriori Untuk Strategi Pemasaran Buku" (2023), mengungkapkan bahwa penggunaan algoritma Apriori bersama *K-Medoids* memungkinkan pengelompokan produk yang lebih akurat. Hasilnya adalah strategi pemasaran yang lebih terarah dengan penawaran diskon yang relevan, yang dapat meningkatkan daya saing bisnis [12]. Dengan menggunakan dua strategi promosi, hasil penelitian menunjukkan bahwa iklan dapat mempromosikan dua jenis produk dalam kategori Al-Qur'an, Zikir, dan Doa. Strategi ini mendapatkan skor dukungan 3%, tingkat keyakinan 44%, dan melibatkan promosi penjualan melalui dua paket produk terkait. Kelemahan studi ini adalah buku-buku tersebut dikelompokkan terlebih dahulu untuk mengidentifikasi hubungan di antara mereka.

Studi ini mengkaji riset yang sudah ada untuk merancang rencana pemasaran yang membuat Edulink.id lebih menarik bagi remaja. Studi ini melampaui penelitian sebelumnya dengan tidak hanya memilah calon pelanggan berdasarkan karakteristik mereka, tetapi juga menemukan tren umum dalam karakteristik tersebut. Wawasan ini akan membantu menciptakan rencana pemasaran yang lebih baik yang akan membuat orang lebih menyadari manfaat menggunakan Edulink.id. Dengan menggunakan metode data mining, studi ini ingin membangun pendekatan pemasaran yang lebih cerdas dan efektif yang dapat membantu Edulink.id tetap kompetitif di bidang EdTech.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining*, adalah proses standar untuk penambangan data di berbagai industri, untuk menganalisis dan mengelola data. Penelitian ini dibagi menjadi beberapa langkah, sebagai berikut:



Gambar 3. Alur Penelitian

Deskripsi Gambar 3 tentang alur penelitian:

A. *Bussiness Understanding*

Pahami sepenuhnya situasi bisnis Edulink.id. Langkah ini berarti mencermati tujuan utama perusahaan, kesulitan yang dihadapi, dan bagaimana penyelesaiannya dapat menghasilkan hasil yang baik. Pada tahap ini, analis bekerja sama dengan tim bisnis untuk mengidentifikasi secara jelas masalah yang perlu dipecahkan menggunakan data. Fase ini mencakup beberapa langkah yang lebih kecil:

1. Identifikasi Masalah

Mencoba menemukan permasalahan di Edulink.id dan mencari cara mengukur permasalahan tersebut serta menghubungkannya dengan proses penelitian.

2. Analisis Masalah

Proses penggalian mendalam untuk memahami suatu masalah, mencari tahu mengapa hal itu terjadi, sehingga dapat menemukan solusi yang baik dan memutuskan data mana yang akan digunakan untuk penelitian.

B. Studi Literatur

Dalam tinjauan pustaka ini, peneliti meninjau berbagai artikel dan jurnal untuk mendukung topik dan metode penelitian. Berdasarkan temuan dari tinjauan ini, penelitian ini akan menggabungkan dua teknik penambangan data: pengelompokan dan aturan asosiasi, serta menggunakan Google Colab untuk memproses data.

C. *Data Understanding*

Data Understanding adalah pemahaman data berarti memikirkan untuk apa data tersebut akan digunakan dalam penelitian. Data tersebut berasal dari Edulink.id antara Januari 2022 dan Februari 2025. Tahap ini memiliki langkah-langkah yang lebih kecil, seperti yang dijelaskan di bawah ini:

1. Pengumpulan Data

Cara mendapatkan informasi/fakta dari berbagai sumber untuk membantu analisis, pengambilan keputusan, atau pemecahan masalah. Semua data dimasukkan ke dalam satu lembar data. Data yang digunakan berjumlah 97 sekolah yang dianalisis hanya mencakup data internal hasil wawancara dan diskusi dengan sekolah yang telah dikunjungi tim pemasaran, mencakup variabel-variabel seperti:

- a. Minat atau ketertarikan terhadap penggunaan aplikasi Edulink.id.
- b. Akreditasi sekolah.
- c. Keinginan sekolah terhadap sistem digitalisasi.
- d. Penggunaan aplikasi kompetitor
- e. Kendala biaya, budaya, dan SDM.
- f. Status penggunaan aplikasi digital saat ini.

D. *Data Preparation*

Data Preparation (Persiapan Data) adalah proses mengubah data mentah menjadi format yang bersih, terorganisir, dan siap dianalisis. Ini merupakan bagian penting dari ilmu data dan analisis data karena hasil analisis sangat bergantung pada kualitas data. Langkah ini mencakup beberapa bagian yang lebih kecil:

1. *Data Cleaning*

Proses Memperbaiki atau menghilangkan informasi yang salah, hilang, berulang, atau tidak berguna dalam suatu kumpulan data agar data tersebut menjadi lebih baik. Langkah ini membantu memastikan data yang digunakan dalam studi dan model tetap konsisten dan dapat dipercaya.

2. *Data Transformation*

Data transformation yaitu mengubah data ke bentuk yang sesuai dengan penambangan data melibatkan beberapa langkah. Salah satu langkah umum adalah pengkodean, yang mengubah teks atau informasi berbasis kategori menjadi angka. Langkah lainnya adalah memilih fitur mana yang paling penting. Selain itu, data sering kali dibagi ke dalam beberapa kelompok atau rentang. Metode-metode ini membantu mempermudah pemrosesan dan analisis data oleh algoritma.

E. *Modeling*

Peran ini melibatkan pembuatan model analitis atau pembelajaran mesin untuk menemukan pola dalam data atau membuat prediksi menggunakan data yang ada. Peran ini memeriksa apakah data dapat memberikan wawasan atau prakiraan yang jelas dan akurat. Temuan ini membantu mendukung keputusan yang diambil bisnis.

F. *Evaluation*

Untuk memastikan hasil model selaras dengan tujuan bisnis dan mengonfirmasi data mencerminkan tujuan ini, hasil yang diperoleh dari perhitungan yang dilakukan menggunakan Google Colab dengan kedua algoritma ini akan dibandingkan dengan hasil yang dihasilkan secara manual menggunakan Microsoft Excel, untuk menilai keakuratan model.

G. *Deployment* (Penerapan)

Menunjukkan hasil analisis yang dilakukan dengan Algoritma Pengelompokan *K-Means* dan Algoritma Apriori kepada tim pemasaran, lalu menggunakan informasi tersebut untuk menyusun rencana pemasaran. Selain itu, membuat bagan dan laporan yang menunjukkan analisis Anda dengan cara yang jelas dan sederhana sehingga setiap orang di perusahaan dapat memahaminya dengan mudah.

III. LANDASAN TEORI

A. *K-Means*

Algoritma *K-Means* mengelompokkan data ke dalam kluster berbeda sehingga data yang serupa berakhir di kelompok yang sama dan data yang berbeda berada di kelompok terpisah. [22]. Lebih spesifiknya, algoritma *k-means* bekerja seperti ini:

1. Pilih K sebagai jumlah grup yang dibuat.
2. Pilih posisi acak untuk titik awal setiap grup (disebut sentroid).
3. Hitung jarak terdekat setiap titik data dan setiap sentroid dengan rumus jarak *Euclidean*. Rumus untuk jarak *Euclidean* adalah:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Keterangan:

x_i : kriteria

μ_j : sentroid kluster ke- j

4. Klasifikasikan setiap titik data berdasarkan jaraknya ke pusat massa, pilih titik dengan jarak terpendek. Kemudian, perbarui nilai pusat massa. Pusat massa yang baru dihitung dengan mengambil rata-rata semua titik data dalam kluster tersebut, menggunakan rumus yang diberikan:

$$C_k = \frac{\sum d_i}{n_k}$$

Keterangan:

n_k : Jumlah data kluster k

d_i : Jumlah dari nilai jarak yang masuk dalam kluster.

5. Ulangi langkah 2 - 5 hingga tidak ada lagi perubahan pada setiap kelompok.
6. Setelah Anda menyelesaikan langkah terakhir, gunakan nilai pusat kelompok (μ_j) dari putaran terakhir untuk membantu mengklasifikasikan data.

Memahami Pengelompokan *K-Means*, K adalah angka yang menunjukkan berapa banyak kelompok yang ingin kita bentuk. Istilah "Means" mengacu pada rata-rata sekelompok titik data, dan setiap kelompok disebut sebagai kluster. Pengelompokan adalah teknik yang digunakan untuk mengelompokkan data dan merupakan bagian dari analisis data atau penambahan data. Metode ini tidak memerlukan data berlabel, sehingga dikenal sebagai pembelajaran tanpa pengawasan. Algoritma *K-Means* digunakan untuk membagi data menjadi beberapa kelompok, di mana setiap kelompok memiliki karakteristiknya sendiri yang berbeda tetapi juga memiliki beberapa kesamaan dengan kelompok lain:

1. Cari nilai k , yang menunjukkan berapa banyak grup atau kluster yang perlu kita buat.
2. Mulailah dengan memilih k titik acak sebagai pusat awal kluster.
3. Ukur jarak setiap titik data dari setiap pusat menggunakan rumus Jarak *Euclidean*:

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

4. Masukkan setiap titik data ke dalam kelompok yang paling dekat dengan pusat klusternya.
5. Temukan lokasi baru pusat kluster (k).
6. Kembali ke langkah 3 jika pusat kluster yang baru berbeda dari yang lama.

Pengelompokan *K-Means* adalah teknik non-hierarkis untuk pengelompokan data. Algoritma ini membagi data menjadi satu atau beberapa kelompok, menempatkan data yang serupa dalam kelompok yang sama dan data yang berbeda dalam kelompok terpisah. Metode ini populer karena mudah digunakan. Namun, salah satu masalah dengan metode ini adalah cara pembentukan kluster sangat bergantung pada bagaimana titik awal dipilih.

B. Algoritma Apriori

Algoritma Apriori digunakan untuk menemukan aturan asosiasi. Algoritma ini menciptakan pola itemset yang sering muncul melalui langkah-langkah analisis data yang berulang. Dalam proses ini, algoritma mengikuti dua langkah utama: pertama, algoritma menghasilkan kemungkinan kombinasi item, yang disebut kandidat, berdasarkan tingkat dukungan minimum. Kemudian, kandidat-kandidat ini diperiksa terhadap data lain untuk menghitung nilai keyakinan, yang menentukan seberapa kuat aturan asosiasi tersebut. Untuk membentuk aturan asosiasi, nilai keyakinan harus lebih tinggi dari tingkat keyakinan minimum. Berikut langkah-langkah untuk menemukan aturan asosiasi:

1. Temukan tingkat dukungan terendah yang harus dipenuhi oleh item.

2. Buat kemungkinan kombinasi item, mulai dari item tunggal hingga kelompok yang lebih besar, pastikan semuanya memenuhi tingkat dukungan minimum. Proses ini terus berlanjut hingga tidak ada lagi kombinasi yang dapat dibuat.
3. Hitung tingkat keyakinan untuk set item terakhir.

IV. IMPLEMENTASI DAN PENGUJIAN

Tampilan program adalah bentuk atau susunan dari bagian-bagian yang terlihat oleh pengguna ketika menjalankan suatu program komputer. Dalam konteks *data mining* menggunakan bahasa pemrograman Python di Google Colab, tampilan program merujuk pada struktur atau isi dari kode yang dituliskan dalam *notebook*, serta bagaimana hasil dari kode tersebut ditampilkan. Tampilan program biasanya mencakup beberapa hal berikut:

A. Kode Program:

Serangkaian langkah yang ditulis oleh orang-orang yang menggunakan Python untuk melakukan berbagai tugas seperti mendapatkan data, membersihkannya, menggunakan alat khusus untuk menemukan pola, dan menampilkan hasil akhir.

B. Struktur Proses:

Kode ini menunjukkan langkah-langkah proses penambangan data secara berurutan, dimulai dengan mendapatkan data, mengolahnya, dan akhirnya menampilkan hasilnya. Hal ini memudahkan pengguna untuk mengikuti cara kerja program.

C. *Output* Hasil:

Hasilnya berupa teks, tabel, atau gambar yang muncul setelah kode dieksekusi. Hasil ini menampilkan hasil dari setiap langkah, seperti hasil klasifikasi, pengelompokan, atau representasi visual data.

D. Komentar dan Penjelasan:

Baris teks dalam kode yang memberikan informasi atau menjelaskan sesuatu, membantu orang lain memahami fungsi setiap bagian program. Baris ini biasanya diawali dengan tanda pagar (#).

```
[ ] # 7. Jalankan K-Means (dengan asumsi 2 cluster)
kmeans = KMeans(n_clusters=2, random_state=42, n_init='auto')
df = df.copy()
df['cluster'] = kmeans.fit_predict(df_selected)

[ ] # Simpan hasil K-Means
df.to_csv('hasil_clustering.csv', index=False)
df
```

No	Kelompok	Tanggal Kunjungan	Jenjang	Nama Sekolah	Tertarik	Akreditasi	Jumlah Siswa	Status	Tim Kunjungan	Digitalisasi	Penggunaan Produk Kompetitor	Punya Aplikasi Sendiri	Kendala Harga	Kendala SDM	Sarana Pendukung	Kendala Budaya	Cluster
0	1	Kabupaten Cirebon	25/01/2023	MI	MI Al Washliyah	Tertarik	A	644	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Tidak	Ada	Tidak	0
1	2	Kabupaten Cirebon	26/01/2023	SD	SDIT Sabuk Huda	Tertarik	A	500	Connected	Irfan	Ya	Tidak	Tidak	Tidak	Ada	Tidak	1
5	6	Kabupaten Cirebon	14/02/2023	SMA	SMAS PGRI Babakan	Tertarik	A	110	Hold/Tertarik	Irfan	Ya	Ya	Tidak	Ya	Tidak	Ada	1
6	7	Kabupaten Cirebon	15/02/2023	SMK	SMK Uluwatu Sukanan	Tertarik	B	75	Hold/Tertarik	Irfan	Ya	Tidak	Tidak	Ya	Tidak	Ada	1
8	9	Kabupaten Cirebon	21/02/2023	SMP	SMP ISLAM TERPADU SALMAN	Tertarik	B	285	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Ya	Tidak	Ada	1
12	13	Kabupaten Cirebon	22/02/2023	SMK	SMK Nasyrul Uluwatu Gegesik	Tertarik	B	500	Hold/Tertarik	Irfan & Kurniawan	Ya	Tidak	Ya	Tidak	Ada	Tidak	1
15	16	Kabupaten Cirebon	21/02/2023	SMP	SMP DARBIYATUL KHAN	Tertarik	B	112	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Ya	Tidak	Ada	1
16	17	Kabupaten Cirebon	22/02/2023	SMA	SMA 1 WALED	Tertarik	A	900	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Tidak	Ada	Tidak	0
20	21	Kabupaten Cirebon	24/02/2023	SMA	SMA NEGERI 1 CILEDUG	Tertarik	A	880	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Tidak	Ada	Tidak	0
22	23	Kabupaten Cirebon	25/02/2023	SMA	SMA 1 GESESIK	Tertarik	A	980	Connected	Irfan & Kurniawan	Ya	Tidak	Ya	Tidak	Ada	Tidak	0
23	24	Kabupaten Cirebon	26/02/2023	SMA	SMA NEGERI 9 KOTA CIREBON	Tertarik	A	1040	Connected	Irfan & Kurniawan	Ya	Tidak	Tidak	Tidak	Ada	Tidak	0

Gambar 4. Tampilan input dan output hasil Algoritma *K-Means*

Gambar 4. menampilkan *coding* algoritma *K-Means* dengan asumsi 2 klaster, hasil dari proses *K-Means* disimpan pada file CSV. Dari data 97 sekolah yang digunakan 28 sekolah termasuk klaster prospektif (klaster 0) dan 69 klaster non-prospektif (klaster 1), penentuan data masuk dalam klaster prospektif ada beberapa poin pelanggan bisa disebut prospektif antara lain : tertarik untuk melakukan digitalisasi di sekolahnya, memiliki cukup banyak siswa karena dalam penyewaan aplikasi dalam tagihan dibayar sesuai dengan jumlah siswa yang ada jadi semakin banyak siswa disekolah semakin banyak juga keuntungan yang didapatkan perusahaan, tidak memiliki kendala pada alat elektronik maupun siswa dan guru dan staff lainnya untuk menggunakan aplikasi.

```
[ ] # Load data cluster
cluster_apriori = pd.read_csv('/content/cluster_0.csv')

# Pilih kolom yang relevan
selected_columns = ['Tertarik', 'Akreditasi', 'Digitalisasi', 'Menggunakan Produk Kompetitor',
                    'Punya Aplikasi Sendiri', 'Kendala Harga', 'Kendala SDM', 'Sarana Pendukung', 'Kendala Budaya']
df_apriori = cluster_apriori[selected_columns].copy()

# One-hot encoding
df_apriori_encoded = pd.get_dummies(df_apriori)

# --- FILTER SUPPORT ---
min_support = 0.7

frequent_itemsets = apriori(df_apriori_encoded, min_support=min_support, use_colnames=True)

# Simpan hasil frequent itemsets
frequent_itemsets.to_csv('hasil_1_support_only.csv', index=False)
frequent_itemsets
```

	support	itemsets
0	1.000000	(Tertarik_Tertarik)
1	0.857143	(Akreditasi_A)
2	1.000000	(Digitalisasi_Ya)
3	0.785714	(Menggunakan Produk Kompetitor_Tidak)
4	0.750000	(Punya Aplikasi Sendiri_Tidak)
...	...	...
202	0.714286	(Tertarik_Tertarik, Kendala Budaya_Tidak, Meng...
203	0.785714	(Digitalisasi_Ya, Kendala Budaya_Tidak, Kendal...
204	0.714286	(Digitalisasi_Ya, Kendala Budaya_Tidak, Menggu...
205	0.785714	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala B...
206	0.714286	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala B...

207 rows x 2 columns

Gambar 5. Tampilan input dan output itemset yang lolos minimum nilai *support*

Gambar 5. Kode ini menunjukkan cara menemukan nilai *support* minimum dan juga menghitung *support* untuk setiap kelompok item yang dibuat dalam Algoritma Apriori. Nilai *support* minimum yang digunakan adalah 0,7 atau 70%. Setiap kelompok item yang memenuhi atau melampaui nilai *support* minimum ini akan disimpan dalam berkas CSV.

```
[ ] # Baca ulang hasil tahap 1
frequent_itemsets = pd.read_csv('hasil_1_support_only.csv')

# Ubah kolom 'itemsets' menjadi frozenset
frequent_itemsets['itemsets'] = frequent_itemsets['itemsets'].apply(eval).apply(frozenset)
# Hitung aturan asosiasi dari frequent itemsets
rules_conf = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.9)

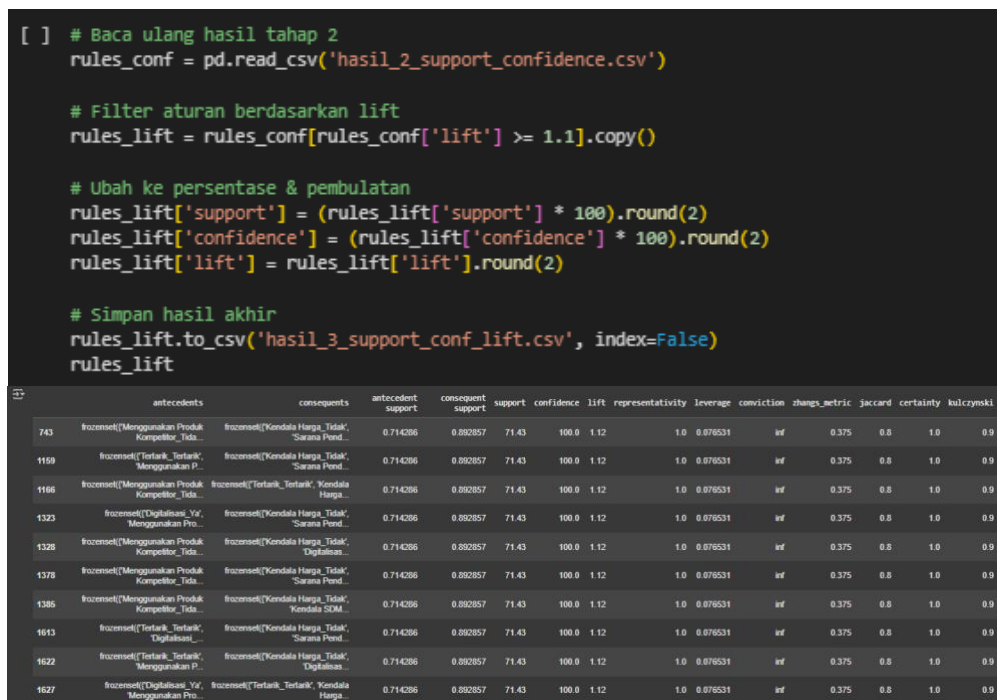
# Simpan hasil filter confidence
rules_conf.to_csv('hasil_2_support_confidence.csv', index=False)
rules_conf
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certainty	kulczynski
0	(Akreditasi_A)	(Tertarik_Tertarik)	0.857143	1.000000	0.857143	1.000000	1.000000	1.0	0.000000	inf	0.000000	0.857143	0.000000	0.820571
1	(Tertarik_Tertarik)	(Digitalisasi_Ya)	1.000000	1.000000	1.000000	1.000000	1.000000	1.0	0.000000	inf	0.000000	1.000000	0.000000	1.000000
2	(Digitalisasi_Ya)	(Tertarik_Tertarik)	1.000000	1.000000	1.000000	1.000000	1.000000	1.0	0.000000	inf	0.000000	1.000000	0.000000	1.000000
3	(Menggunakan Produk Kompetitor_Tidak)	(Tertarik_Tertarik)	0.785714	1.000000	0.785714	1.000000	1.000000	1.0	0.000000	inf	0.000000	0.785714	0.000000	0.892857
4	(Punya Aplikasi Sendiri_Tidak)	(Tertarik_Tertarik)	0.750000	1.000000	0.750000	1.000000	1.000000	1.0	0.000000	inf	0.000000	0.750000	0.000000	0.875000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1935	(Kendala Harga_Tidak, Menggunakan Produk Kompetitor_Tidak)	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala B...)	0.750000	0.928571	0.714286	0.952381	1.025641	1.0	0.017857	1.500000	0.100000	0.740741	0.333333	0.860806
1936	(Digitalisasi_Ya, Menggunakan Produk Kompetitor_Tidak)	(Tertarik_Tertarik, Kendala Budaya_Tidak, Kendal...	0.785714	0.892857	0.714286	0.909091	1.018182	1.0	0.012755	1.178571	0.083333	0.740741	0.151515	0.854545
1937	(Menggunakan Produk Kompetitor_Tidak, Sarana P...)	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala B...)	0.750000	0.892857	0.714286	0.952381	1.066667	1.0	0.044643	2.250000	0.250000	0.708231	0.555556	0.876190
1938	(Menggunakan Produk Kompetitor_Tidak, Kendala ...)	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala S...)	0.714286	0.892857	0.714286	1.000000	1.120000	1.0	0.078531	inf	0.375000	0.800000	1.000000	0.900000
1939	(Menggunakan Produk Kompetitor_Tidak)	(Tertarik_Tertarik, Digitalisasi_Ya, Kendala B...)	0.785714	0.892857	0.714286	0.909091	1.018182	1.0	0.012755	1.178571	0.083333	0.740741	0.151515	0.854545

1940 rows x 14 columns

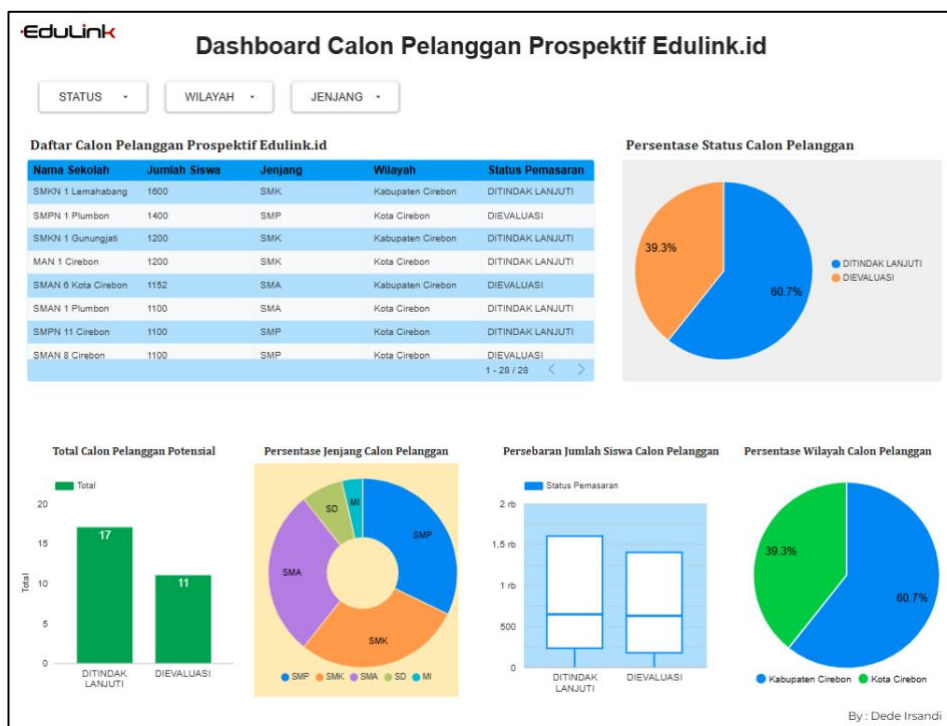
Gambar 6. Tampilan input dan output itemset yang lolos minimum nilai *confidence*

Gambar 6. Kode ini menunjukkan cara menemukan tingkat *confidence* terendah dan menghitung *confidence* untuk setiap kelompok item yang dibuat dengan metode Apriori. Kode ini menggunakan tingkat *confidence* minimum 0,9 atau 90%. Setiap kelompok yang memenuhi atau melampaui tingkat *confidence* ini akan disimpan dalam berkas CSV.



Gambar 7. Tampilan input dan output itemset yang lolos minimum nilai *lift*

Gambar 7. Kode ini menunjukkan cara menemukan nilai *lift* terendah dan juga menghitung *lift* untuk setiap kelompok item yang dibuat dalam Algoritma Apriori. Kode ini mencari kelompok yang memiliki nilai *lift* minimal 1,1 dan menyimpan kelompok tersebut ke dalam berkas CSV.



Gambar 8. Dashboard Calon Pelanggan Prospektif Edulink.id

Link untuk mengakses dashboard : <https://lookerstudio.google.com/reporting/896f829a-c888-4070-9a46-c0596b7e4435>

Gambar 8. menampilkan dashboard sekolah calon pelanggan Edulink.id pada bagian kiri atas dashboard menampilkan daftar nama-nama sekolah, jumlah siswanya, level jenjang sekolah calon pelanggan, wilayah tiap-tiap calon pelanggan dan juga status calon pelanggan prospektif data yang akan muncul pada daftar sesuai filter yang sedang digunakan. Bagian kanan atas *pie chart* berisi persentase jumlah keseluruhan dari status data berdasarkan status ditindaklanjuti dan dievaluasi. Pada bagian kiri bawah terdapat *bar chart* untuk menampilkan total dari tiap status ditindaklanjuti dan dievaluasi. Bagian bawah

tengah ada *pie chart* menampilkan persentase dari tiap-tiap jenjang sekolah calon pelanggan Edulink.id dan juga *box plot* yang menampilkan persebaran berdasarkan jumlah siswa calon pelanggan. Bagian kanan bawah menampilkan persentase berdasarkan wilayah terdapat dua wilayah yaitu kabupaten Cirebon dan kota Cirebon. Dan pada bagian kotak paling kanan atas itu adalah tombol filter untuk memilih data mana yang ingin ditampilkan berdasarkan status calon pelanggan, berdasarkan wilayah calon pelanggan dan juga berdasarkan jenjang sekolah calon pelanggan prospektif. Dari dashboard tersebut bisa disimpulkan bahwa dari data pelanggan prospektif menghasilkan 2 tindakan: ditindaklanjuti yang artinya bisa langsung dilakukan promosi dengan menggunakan strategi yang sama, sedangkan tindakan dievaluasi perlu penyesuaian strategi yang digunakan untuk masing-masing pelanggan.

Tingkat akurasi yang dihasilkan berdasarkan perbandingan data hasil dari *tools* Google Colab dan hasil perhitungan manual dengan *tools* Microsoft Excel pada proses menggunakan Algoritma *K-Means* menghasilkan akurasi 100% berarti data antara kedua *tools* sama mulai dari anggota kluster yang dihasilkan dan juga nilai dari jarak ke masing-masing titik sentroidnya dengan menggunakan nilai sentroid yang digunakan sama pada 2 *tools* tersebut. Untuk akurasi dari Algoritma Apriori adalah 98% perbandingannya dilakukan pada itemset yang lolos nilai minimum confidence terdapat 6 item data di data hasil perhitungan manual tidak ada di data hasil perhitungan dengan *tools* Google Colab.

V. KESIMPULAN

Dari isu dan diskusi yang telah kita bicarakan sebelumnya, berikut adalah poin-poin utama yang dapat kita rangkum:

1. Algoritma *K-Means* berhasil mengelompokkan sekolah menjadi dua kluster utama: prospektif dan non-prospektif. Kluster prospektif memiliki karakteristik sekolah dengan jumlah siswa besar, akreditasi A, minat tinggi terhadap digitalisasi, serta tanpa hambatan signifikan dalam penggunaan aplikasi. Variabel yang dianalisis terbatas pada hasil wawancara internal dengan sekolah, tanpa mempertimbangkan faktor eksternal seperti kebijakan pemerintah atau tren industri EdTech secara umum.
2. Algoritma Apriori mengidentifikasi kombinasi karakteristik terkuat dari calon pelanggan potensial, yang dapat dijadikan dasar dalam penyusunan strategi promosi yang lebih terarah.
3. Dari kluster prospektif, terdapat 28 sekolah calon pelanggan, di mana 17 sekolah termasuk kategori paling potensial untuk segera ditindaklanjuti dengan strategi promosi yang seragam, sedangkan 11 sekolah lainnya memerlukan penyesuaian strategi sesuai karakteristik masing-masing. Data hanya mencakup sekolah di wilayah Cirebon dan sekitarnya, sehingga temuan belum tentu mewakili kondisi nasional.
4. Proses perhitungan Algoritma *K-Means* di Google Colab ini 100% akurat dengan perhitungan manual yang dilakukan di Microsoft Excel yang dijadikan sebagai acuan keakuratannya.
5. Tingkat akurasi dari proses perhitungan Algoritma Apriori menggunakan Google Colab terhadap perhitungan manual menggunakan Microsoft Excel (sebagai acuan kebenaran) adalah 98%.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada Tim JTII atas kontribusinya. Dan kami juga mengucapkan terima kasih kepada PT Alvarel Technology Innovation yang telah menjadikan perusahaan mereka sebagai fokus penelitian ini.

REFERENSI

- [1] detikedu. (2024). Mengenal Jenis dan Perkembangan Edutech Indonesia, dari <https://www.detik.com/edu/detikpedia/d-7158547/mengenal-jenis-dan-perkembangan-edutech-indonesia>.
- [2] T. P. H. Atmoko, "Strategi pemasaran untuk meningkatkan volume penjualan di Cavinton Hotel Yogyakarta," J. Indones. Tourism, Hospitality and Recreation, vol. 1, no. 2, pp. 83–96, 2018, doi: 10.17509/jithor.v1i2.13769.
- [3] "Colaboratory," FAQ, Google Research, 2025. [Online]. Available: <https://research.google.com/colaboratory/intl/id/faq.html> [<https://research.google.com/colaboratory/intl/id/faq.html>]. [Accessed: Apr. 24, 2025].
- [4] M. Isoraité, "Marketing mix theoretical aspects," 2016.
- [5] A. Amrullah, I. Purnamasari, B. N. Sari, Garno, and A. Voutama, "Analisis cluster faktor penunjang pendidikan menggunakan algoritma K-Means (studi kasus: Kabupaten Karawang)," J. Inform. Rekayasa Elektron, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [6] Mustika et al., Data Mining dan Aplikasinya Bandung, Indonesia, 2021.
- [7] Suyanto, Data Mining untuk Klasifikasi dan Klusterisasi Data. Bandung, Indonesia: Informatika, 2017.
- [8] G. Grand, "Penerapan algoritma Apriori untuk menemukan hubungan data murid dengan nilai sekolah," Ikraith Inform., vol. 2, no. 18, pp. 7–12, 2018.
- [9] M. R. A. Fernanda, P. Sokibi, and R. Fahrudin, "Sistem prediksi ketepatan kelulusan mahasiswa menggunakan K-Means (studi kasus: Universitas Catur Insan Cendekia)," J. Digit, vol. 11, no. 1, pp. 89–100, May 2021.
- [10] M. A. Melda, Sutedi, A. S. K. Arman, and R. S. Erlinda, "Strategi marketing promosi penerimaan mahasiswa baru menggunakan K-Means clustering," Teknika, vol. 18, no. 2, 2023.
- [11] A. Syifa, D. Mulyadi, and A. Setiawan, "Penerapan metode K-Means dan Apriori untuk pemilihan produk bundling," J. Cerita, vol. 8, no. 1, Feb. 2022.
- [12] H. Mulyani, "Penerapan algoritma K-Medoids dan Apriori untuk strategi pemasaran buku," UIN Sultan Syarif Kasim Riau, 2022.
- [13] M. Kaur and S. Kang, "Market basket analysis: Identify the changing trends of market data using association rule mining," Procedia Comput. Sci, vol. 85, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.
- [14] F. R. Arrafii, "Analisis pola perilaku konsumen menggunakan metode Market Basket Analysis pada toko sayur keluarga (Tosaga) di masa pandemi Covid-19," 2021. [Online]. Available: <https://dspace.uir.ac.id/handle/123456789/37774> [<https://dspace.uir.ac.id/handle/123456789/37774>].
- [15] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data transaksi bongkar muat di Provinsi Riau," J. Ilm. Rekayasa Manaj. Sist. Inf, vol. 5, no. 1, pp. 119–125, 2019.
- [16] S. Assauri, Manajemen Pemasaran. Jakarta, Indonesia: PT Raja Grafindo Persada, 2017.
- [17] P. Kotler and K. Keller, Manajemen Pemasaran Jilid 1. Jakarta, Indonesia: Erlangga, 2008.

- [18] S. Noor, "Penerapan analisis SWOT dalam menentukan strategi pemasaran Daihatsu Luxio di Malang," J. INTEKNA: Informasi Teknik dan Niaga, vol. 14, no. 2, 2014.
- [19] S. Atsari, H. Afina, and I. Muhammad, "Education start-up business ecosystem," Int. Res. J. Innov. Eng. Technol, vol. 7, no. 11, pp. 710–715, Nov. 2023.
- [20] N. Maula, "Praktik sumber daya manusia pada perusahaan startup Edutech di CV. Big Edu Indonesia," Creative Res. J, vol. 7, no. 1, 2021, doi: 10.34147/crj.v7i01.292.
- [21] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi pelanggan produk digital service Indihome menggunakan algoritma K-Means berbasis Python," J. Manaj. Inform. (JAMIKA), 2023.
- [22] N. Rohmawati, S. Defiyanti, and M. Jajuli, "Implementasi algoritma K-Means dalam pengklasteran mahasiswa pelamar beasiswa," J. Ilm. Teknol. Inf. Terapan vol. 1, no. 2, pp. 62–67, 2015.
- [23] F. Nasari and S. Dharma, "Penerapan K-Means clustering pada data penerimaan mahasiswa baru (studi kasus: Universitas Potensi Utama)," in Seminar Teknologi Informasi dan Multimedia, Yogyakarta, 2015, pp. 73–78.
- [24] E. Umar, D. Manongga, and A. Iriani, "Market basket analysis menggunakan association rule dan algoritma Apriori pada produk penjualan Mitra Swalayan Salatiga," J. Media Inform. Budidarma, vol. 6, no. 3, p. 1367, 2022.
- [25] A. Aquila, C. Pabendon, and H. D. Purnomo, "Penerapan algoritma Apriori dan FP-Growth untuk market basket analisis pada data transaksi non-promo," J. Media Inform. Budidarma, vol. 7, pp. 975–984, 2023.
- [26] M. Nurkholis, G. N. D. P. Gitama, and S. Pranata, "Mengukur kinerja strategi pemasaran berdasarkan persepsi konsumen Toko Salsa," J. Witana, vol. 2, no. 1, pp. 1–7, 2024.
- [27] "Pandas 2.2.3." Project description, 20 September 2024, [Online]. Tersedia: <https://pypi.org/project/pandas/> [Diakses: 25 April 2025].
- [28] M. Waskom, "Seaborn." Seaborn: statistical data visualization [Online]. Tersedia: <https://seaborn.pydata.org/> [Diakses: 25 April 2025].
- [29] "Matplotlib 3.10.1" Project description, 28 Februari 2025, [Online]. Tersedia: <https://pypi.org/project/matplotlib/> [Diakses: 25 April 2025].
- [30] Pavan Vadapalli, "Scikit-learn in Python: Features, Prerequisites, Pros & Cons," upGrad, 11 Juni 2020. <https://www.upgrad.com/blog/scikit-learn-python/> (diakses 28 Juli 2023).
- [31] "Microsoft Excel: Definisi, Fungsi, dan Rumus Penting," PasarTrainer Blog, 26 Maret 2024. [Online]. Tersedia: <https://pasartrainer.com/blog/microsoft-excel-definisi-fungsi-dan-rumus-penting> [Diakses: 25 April 2025].
- [32] "Smart Science and Technology," Telkom University, [Online]. Tersedia: <https://bpe.telkomuniversity.ac.id/apa-itu-flowchart/> [Diakses: 25 April 2025].
- [33] "Pengertian Google Spreadsheet dan Kelebihannya," Coding Studio, [Online]. Tersedia: <https://codingstudio.id/blog/pengertian-google-spreadsheet-adalah/> [Diakses: 25 April 2025].
- [34] D. Sunardi, C. Rozali, "Pelatihan Visualisasi Data Kependudukan Menggunakan Looker Studio Pada Kelurahan Duren Seribu," JIPM : Jurnal Informasi Pengabdian Masyarakat, Vol. 3, No. 1, Hal: 59-63 Januari 2025. S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.