



Artikel

Penerapan Data Mining Untuk Menganalisa Data Penjualan Barang Di Swalayan Dengan Menggunakan Algoritma Fp-Growth

Thomas Febrian¹, Rino², Hartana Wijaya³

^{1,2,3} Teknik Informatika, Universitas Buddhi Dharma, Banten, Indonesia

SUBMISSION TRACK

Received: Jan 22, 2023

Final Revision: March 12, 2023

Available Online: March 24, 2023

KEYWORD

Asosiasi, Data Mining, FP-Growth, Pola Penjualan, Swalayan.

KORESPONDENSI

Phone: 082112012897

E-mail: thomasfebrian1702@gmail.com

A B S T R A C T

Penjualan harian menghasilkan banyak transaksi penjualan, sehingga transaksi penjualan harian menyebabkan terjadinya penumpukan data. Contoh suatu cara untuk dapat meningkatkan pemasaran yang dilakukan adalah dengan menggunakan data penjualan yang telah ada. Data tersebut digunakan bukan hanya untuk arsip perusahaan saja, tetapi dapat diolah dan dimanfaatkan untuk menghasilkan suatu informasi yang berguna untuk meningkatkan penjualan suatu produk. Data operasional bukan satu-satunya data yang cukup untuk mengambil suatu keputusan, dibutuhkannya data yang berasal dari sumber lain. Analisa pada data diperlukan untuk mencari suatu kemungkinan pada informasi dari data. Pengambil keputusan mencoba menggunakan data yang tersedia untuk membuat keputusan di gudang data. Ini telah menghasilkan pengembangan pendekatan baru untuk mengekstraksi informasi. Ini dikenal sebagai penambangan data. Salah satu metode yang dapat menggali informasi ini adalah metode asosiasi dengan algoritma FP-Growth. Algoritma ini merupakan salah satu metode dalam penambangan data yang bertujuan untuk mengetahui produk-produk yang sering terjual dan keterkaitan antar produk satu dengan yang lainnya. Penelitian ini memanfaatkan data sekunder berasal dari situs kaggle yang berjudul groceries dataset dan data diambil berdasarkan atribut yang diperlukan. Pola penjualan kombinasi barang swalayan ini ditentukan berdasarkan nilai *support* dan nilai *confidence*, sehingga berdasarkan hasil yang diperoleh nilai *support* terbesar pada item *whole milk* dan *support* terkecil pada item *female sanitary product*.

PENDAHULUAN

Semakin ketatnya persaingan dalam dunia perdagangan, menuntut para pemilik usaha untuk menemukan cara atau strategi yang dapat meningkatkan penjualan. Transaksi penjualan yang dilakukan setiap hari akan mengakibatkan terjadinya penumpukan data, karena penjualan yang dilakukan setiap hari akan menyebabkan banyaknya transaksi penjualan. Salah satu strategi untuk dapat meningkatkan penjualan yang dilakukan adalah dengan memanfaatkan data penjualan yang telah ada. Data tersebut digunakan bukan hanya untuk arsip perusahaan saja, tetapi dapat diolah dan dimanfaatkan untuk menghasilkan suatu informasi yang berguna untuk meningkatkan penjualan suatu produk.

Dalam dunia bisnis, perusahaan perlu kompetitif dan mencari cara bagaimana mereka dapat terus tumbuh dan melebarkan jangkauan mereka. Pelaku ekonomi harus mengembangkan berbagai strategi guna meningkatkan penjualan produk. Salah satu cara untuk meningkatkan kinerja penjualan adalah dengan mengumpulkan semua data penjualan yang terjadi di dalam perusahaan. Data operasional bukan satu-satunya data yang cukup untuk mengambil suatu keputusan [1]. Data operasional bukan satu-satunya data yang cukup untuk mengambil suatu keputusan, dibutuhkannya data yang berasal dari sumber lain. Analisa pada data diperlukan untuk mencari suatu kemungkinan pada informasi dari data. Pengambil keputusan mencoba menggunakan data yang tersedia untuk membuat keputusan di gudang data. Kegiatan ini telah menghasilkan pengembangan pendekatan baru untuk mengekstraksi informasi. Ini dikenal sebagai penambangan data. Penggunaan datamining memiliki tujuan untuk mengeluarkan informasi baru yang sebelumnya terdapat di dalam gudang data sehingga dapat dibuat menjadi suatu informasi yang berharga dan dapat digunakan.

“Data mining adalah salah satu ilmu dalam komputer yang digunakan untuk mengumpulkan berbagai informasi dari berbagai sumber, yang nantinya akan diolah dan akan menghasilkan informasi yang lebih

baik” dalam mengoperasikan data mining dibutuhkan yang namanya algoritma [2]. Salah satu metode yang dapat menggali informasi ini adalah metode asosiasi dengan algoritma FP-Growth. Algoritma ini merupakan salah satu metode dalam penambangan data yang bertujuan untuk mengetahui produk-produk yang sering terjual dan keterikatan antar produk. Algoritma FP-Growth merupakan salah satu metode asosiasi pada *data mining* yang menyatakan keterikatan antara barang, yang sering disebut sebagai analisa belanja keranjang [3]. Sistem ini nantinya bisa beroperasi dengan menganalisis dan menemukan suatu pola yang keterikatan antar produk-produk yang dibeli oleh konsumen.

I. METODE

Metode CRISP-DM telah banyak digunakan pada beberapa industri dalam mengelola data, para pebisnis membutuhkan proses data mining yang standar sebagai acuan untuk menyelesaikan permasalahan [4]. Dalam praktiknya, data mining dengan metode CRISP-DM telah banyak dipraktikkan, salah satunya yang terdapat jurnal “Business and Understanding Data” sebagai bagian dari pemodelan layout, dan layout pasar tradisional untuk meningkatkan margin penjualan dengan menggunakan “Aturan Asosiasi dan Metode Pohon Keputusan” (Studi Kasus Kota Bandung)” Tinjauan ini berkaitan dengan pemahaman tujuan Bisnis dan data yang akan digunakan selama penambangan data untuk menentukan tata letak dan tata letak pasar tradisional yang strategis untuk meningkatkan keuntungan penjualan.

Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase, yaitu:

1. Pemahaman Bisnis adalah memahami aktivitas data mining apa yang Anda lakukan dan apa kebutuhan Anda dari perspektif bisnis.
2. Pemahaman Data adalah tahap mengumpulkan data awal, memeriksa data untuk memahami data apa yang digunakan, dan mengidentifikasi masalah yang terkait dengan data tersebut[5].

3. Penyiapan Data adalah kegiatan menentukan tabel dan field mana yang akan diubah menjadi database baru untuk *data mining*.
4. Pemodelan adalah fase memutuskan teknik mana yang akan digunakan untuk data mining, memutuskan alat mana yang akan digunakan untuk melakukan data mining, metode apa yang digunakan dan memutuskan algoritma mana yang akan digunakan [6].
5. Evaluasi adalah tahap interpretasi hasil *data mining* yang disajikan pada proses pemodelan tahap sebelumnya.
6. *Deployment* adalah fase penyusunan laporan atau penyajian *insight* yang diperoleh setelah melakukan aktivitas *data mining* ini [7].

Algoritma FP-Growth adalah salah satu contoh algoritma yang cukup sering digunakan untuk mencari kumpulan data yang paling sering muncul dalam suatu basis data[8]. FP-Growth dapat menemukan hubungan itemset satu dengan yang lainnya hanya dengan beberapa kunjungan ke basis data asli. Algoritma FP-Growth dapat menghasilkan banyak frequent itemset atau yang disebut pola yang sering muncul untuk sebuah item pada sebuah basis data yang besar dengan melakukan *candidate generation* dan merekam basis data secara berulang [9]. Sehingga penulis, berkeinginan untuk menggunakan algoritma ini dengan melakukan penelitian bagaimana cara meminimalkan frequent itemset. Langkah-langkah algoritma FP-Growth adalah sebagai berikut :

1. Menentukan nilai minimum dari item set yang memiliki frekuensi kemunculan lebih tinggi dari nilai minimum yang ditentukan, dan diperoleh kandidat 1 item set.
2. Setelah hasil iterasi pertama didapatkan, pencarian dilanjutkan untuk kandidat 2 itemset. Jika tidak ada kecocokan yang ditemukan untuk nilai minimum yang ditentukan, proses berhenti.
3. Nilai support dan confidence dihitung dari kandidat itemset yang diterima.

4. Didapatkan suatu aturan yang dapat digunakan oleh pengguna untuk suatu informasi [10].

Contoh Perhitungan:

1. Rumus Asosiasion rule: Nilai *support* suatu item yang diperoleh dengan rumus:

$$\text{Support (A)} = \frac{\sum \text{Transaksi Mengandung A}}{\sum \text{Transaksi}} \times 100\%$$

2. Sedangkan untuk nilai support dari 2 item yang diperoleh dengan rumus: *Support P(A|B)*

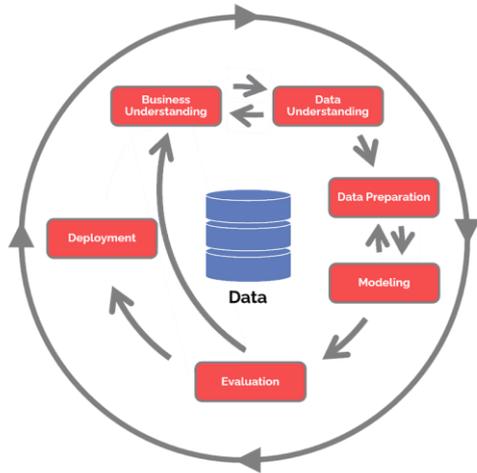
$$= \frac{\sum \text{Mengandung A dan B}}{\sum \text{Transaksi}} \times 100$$

3. Dan nilai confidence diperoleh dengan rumus:

$$\text{Confidence (A|B)} = \frac{\sum \text{Transaksi Mengandung } A \cap B}{\sum \text{Transaksi A}} \times 100$$

II. HASIL

Berdasarkan hasil tinjauan pustaka yang penulis dapatkan, maka saya pun mengikuti kerangka kerja dari metode CRISP-DM untuk mendapatkan pola perilaku konsumen dalam membeli produk di Toko Swalayan dengan menggunakan algoritma FP-Growth. Adapun tahap-tahap dari Crisp-DM adalah sebagai berikut:



Gambar 1: Langkah-langkah Crisp-DM

1. Business Understanding

Pemahaman bisnis adalah tingkat pertama dalam CRISP-DM dan termasuk bagian yang cukup vital [11]. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun. Tujuan bisnis yang ingin dicapai pada penelitian kali ini adalah untuk mendapatkan pola perilaku konsumen dalam membeli produk di Toko Kelontong, agar pemilik Toko Kelontong dapat menjual produk yang ia miliki lebih efektif lagi. Baik itu dengan cara penempatan produk yang paling sering dibeli, maupun membundling produk yang paling banyak dibeli oleh konsumen menjadi satu paket. Tujuan data mining yang ingin dicapai pada penelitian ini:

Mengetahui hubungan atau pola dari pembelian suatu produk dengan produk yang lainnya, menggunakan algoritma FP-Growth.

2. Data Understanding

Tahapan pemahaman data dimulai dengan pengumpulan data awal dan hasil kegiatan dalam rangka untuk membiasakan diri dengan data untuk mengidentifikasi masalah data dan mendeteksi subset menarik untuk membentuk hipotesis untuk informasi yang tersembunyi[12].

-Pengumpulan Data

Pada penelitian kali ini, penulis menggunakan dataset “Groceries_dataset.csv” yang penulis dapatkan melalui website Kaggle di url berikut ini

<https://www.kaggle.com/heeraldedhia/groceries-dataset>

-Pengenalan Dataset

Dataset memiliki 18 baris pesanan pembelian orang dari toko swalayan. Pesanan ini dapat dianalisis dan aturan asosiasi dapat dihasilkan menggunakan Analisis Keranjang Pasar dengan algoritma seperti Algoritma *FP-Growth*. Dataset yang penulis dapatkan berisikan 3 buah baris, yaitu:

Tabel 1. Tabel Atribut Dataset

Member _number	Berisi nomor unik pembeli, sebagai pembeda antara pembeli satu dengan yang lainnya
Date	Berisi tanggal, bulan dan tahun pembeli melakukan transaksi
ItemDescription	Berisi nama produk

3. Data Preparation

Tahap persiapan dan seleksi data yang telah dikumpulkan dan diubah menjadi bentuk yang dapat diolah dalam model yang ditentukan selanjutnya [13].

-Tampilan Dataset

Berikut ini adalah tampilan dataset “Groceries_dataset.csv” sebelum dilakukan proses yang Data Preparation yang selanjutnya:

Tabel 2. Dataset sebelum dilakukan pembersihan

Row No.	Member _number	Date	ItemDescription
1	1808	Jul 21, 2015	Tropical fruit
2	2552	Jan 5, 2015	Whole milk
3	2300	Sep 19, 2015	Pip fruit
4	1187	Dec 12, 2015	Other vegetables

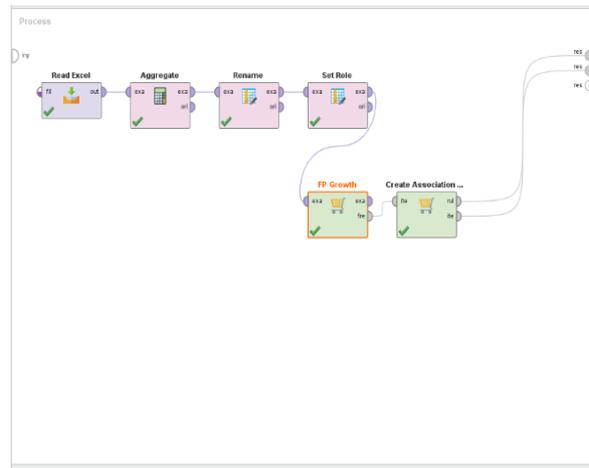
5	3037	Feb 1, 2015	Whole milk
6	4941	Feb 14, 2015	Rolls/buns
7	4501	May 8, 2015	Other vegetables
8	3803	Dec 23, 2015	Pot plants
9	2762	Mar 20, 2015	Whole milk
10	4119	Feb 12, 2015	Tropical fruit
11	1340	Feb 24, 2015	Citrus fruit
12	2193	Apr 14, 2015	beef
13	1997	Jul 12, 2015	Frankfurter
14	4546	Sep 3, 2015	Chicken
15	4736	Jul 21, 2015	Butter
16	1959	Mar 30, 2015	Fruit/vegetable juice
17	1974	May 3, 2015	Packaged fruit
18	2421	Sep 2, 2015	Chocolate

1997	Frankfurter
4546	Chicken
4736	Butter
1959	Fruit/vegetable juice
1974	Packaged fruit
2421	Chocolate

Jadi penulis hanya menggunakan kolom Member_number dan itemDescription pada dataset “Groceries_dataset.csv”.

4. Modeling

Dalam tahap ini dilakukan pemilihan dan implementasi model, jika dibutuhkan data kembali ke proses pengolahan data untuk menjadikan data yang sesuai untuk kebutuhan *data mining* [15].



Gambar 2: Pemodelan pada Rapidminer

a. Data cleaning bertujuan untuk membersihkan data-data yang tidak diperlukan [14]. Di dataset yang penulis gunakan tidak ditemukan data bernilai null atau data yang ambigu sehingga pada dataset “Groceries_dataset.csv” ini, hanya menghapus kolom yang tidak mengalami ketergantungan yaitu kolom Date.

Tabel 2. Tabel Dataset Cleaning

Member_number	ItemDescription
1808	Tropical fruit
2552	Whole milk
2300	Pip fruit
1187	Other vegetables
3037	Whole milk
4941	Rolls/buns
4501	Other vegetables
3803	Pot plants
2762	Whole milk
4119	Tropical fruit
1340	Citrus fruit
2193	beef

Fungsi tiap operator:

1. *Read Excel* = Operator ini membaca data dari excel.
2. *Aggregate* = Operator ini melakukan fungsi yang diketahui dari SQL. Operator ini menyediakan banyak fungsi dalam format yang sama yang disediakan oleh fungsi agregasi SQL.
3. *Rename* = Operator ini berfungsi untuk mengganti nama attribute lain.
4. *FP-Growth* = Operator ini berfungsi untuk menghitung semua itemset yang sering muncul dalam Exampleset dan menggunakan struktur data pohon FP
5. *Create Association Rules* = Operator ini menghasilkan beberapa aturan asosiasi dari himpunan frequent itemset. Data set di

read menggunakan Read Excel, lalu ditab Aggregate attributes masukan ItemDescription dan aggregation function menjadi concatenation. Untuk Tab Rename old name concat (itemDescription) new item menjadi (item). Pada Tab Set Role set attribute name menjadi Member_member dan target role id. FP growth dengan parameter input format = item list in a column, Min requirement = support, min support 0.1 dan min item per itemset 3 dan max item per itemset 5. Pada tab create association rules criterion berupa confidence dan min confidence 0.01.

5. Evaluation

Tahapan ini untuk melakukan validasi dan pengukuran keakuratan hasil pengolahan dengan Algoritma pada Software RapidMiner kemudian melakukan Analisis terhadap hasil yang diperoleh dari Penelitian ini [16]. Dari hasil model diatas didapatkan bahwa item yang memiliki jumlah support tertinggi dalam gabungan 3 item adalah whole milk, other vegetables, rolls/buns dengan support 0.082 dan support terkecil dengan item whole milk, other vegetables, herbs dengan support 0.010. Hasil ditampilkan pada gambar dibawah:

Size	Support ↓	Item 1	Item 2	Item 3
3	0.082	whole milk	other vegetables	rolls/buns

Size	Support ↑	Item 1	Item 2	Item 3
3	0.010	whole milk	other vegetables	herbs

Gambar 3 : Hasil pemodelan 3 item

Size	Support	Item 1
1	0.458	whole milk

Size	Support ↑	Item 1
1	0.010	female sanitary products
1	0.010	spices

Gambar 4 : Hasil pemodelan 1 item

Dari gambar diatas ini menunjukkan bahwa whole milk sebagai item dengan jumlah support terbesar dengan 0.458 dan support terkecil dengan 0.010 dengan item female sanitary products dan spices (1 item).

6. Deployment

Berdasarkan hasil dari proses evaluasi, maka ditemukan pola pola dengan tingkat keyakinan/confidence yang tinggi yaitu:

1. Jika seseorang membeli *domestic egg* dan *meat* secara bersamaan maka berkemungkinan akan membeli *whole milk*.
2. Jika seseorang membeli *fruit/vegetable juice* dan *chocolate* secara bersamaan maka berkemungkinan akan membeli *whole milk*.
3. Jika seseorang membeli *yogurt, bottled water, pip fruit* secara bersamaan, maka berkemungkinan akan membeli *whole milk*.

Ketiga premis ini memiliki nilai keyakinan/confidence diatas 0.7 dimana angka tersebut merupakan nilai yang sangat baik untuk dijadikan acuan.

Langkah bisnis yang dapat diambil adalah:

1. Meletakkan kombinasi item/barang yang dijual diatas etalase yang berdekatan.
2. Memberikan penawaran secara langsung kepada pelanggan disaat pelanggan membeli salah satu item/barang.
3. Membuat *Bundle item* yang memiliki kemungkinan pembelian terbesar.

III. KESIMPULAN

Berdasarkan data dan pembahasan tersebut maka dapat diambil kesimpulan sebagai berikut: Aturan yang dihasilkan memberikan wawasan baru tentang perilaku konsumen dalam pembelian barang. Dengan pengetahuan ini, dapat digunakan untuk membuat keputusan dalam persediaan barang dan menawarkan diskon untuk pola pembelian produk dengan nilai *rules* yang tertinggi; Algoritma fp-growth dapat digunakan untuk mengetahui keterkaitan antara satu barang dengan yang lainnya. Dari informasi tersebut bisa digunakan untuk menentukan penempatan rak barang yang memiliki nilai *support* tertinggi.

DAFTAR PUSTAKA

- [1] L. Z. Saputri and S. Sugiyono, “Penerapan data mining pembuatan produk baru tinta menggunakan metode algoritma Apriori,” *Jurnal Manajemen Informatika Jayakarta*, vol. 1, no. 4, p. 320, 2021, doi: 10.52362/jmijayakarta.v1i4.567.
- [2] E. D. Sikumbang, “Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori,” vol. 4, no. 1, pp. 156–161, 2018.
- [3] N. D. Andriai, T. W. Utami, R. Wasodo, and M. Si, “IMPLEMENTASI ALGORITMA FP-GROWTH DALAM MARKET BASKET ANALYSIS UNTUK MENGANALISIS POLA BELANJA KONSUMEN PADA DATA TRANSAKSI PENJUALAN.” [Online]. Available: <http://repository.unimus.ac.id>
- [4] A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 1, no. 3, pp. 260–270, 2015, doi: 10.28932/jutisi.v1i3.406.
- [5] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [6] D. Feblian and D. U. Daihani, “IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X,” *JURNAL TEKNIK INDUSTRI*, vol. 6, no. 1, Feb. 2017, doi: 10.25105/jti.v6i1.1526.
- [7] I. Budiman, T. Prahasto, and Y. Christyono, “Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma,” *Jurnal Sistem Informasi Bisnis*, vol. 1, no. 3, pp. 15–16, 2014, doi: 10.21456/vol1iss3pp129-134.
- [8] E. Munanda and S. Monalisa, “Penerapan Algoritma Fp-Growth Pada Data Transaksi Penjualan Untuk Penentuan Tataletak,” *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 7, no. 2, pp. 173–184, 2021, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/13253>
- [9] Y. D. Lestari, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA FP-TREE DAN FP-GROWTH PADA DATA TRANSAKSI PENJUALAN OBAT.”
- [10] M. Tahir and N. Sitompul, “PENERAPAN ALGORITMA FP-GROWTH DALAM MENENTUKAN KECENDERUNGAN MAHASISWA MENGAMBIL MATA KULIAH PILIHAN,” 2021.
- [11] I. Purnama, R. Saputra, and A. Wibowo, “IMPLEMENTASI DATA MINING MENGGUNAKAN CRISP-DM PADA SISTEM INFORMASI EKSEKUTIF DINAS KELAUTAN DAN PERIKANAN PROVINSI JAWA TENGAH.”
- [12] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [13] B. A. Kennardi, “PENERAPAN DATA MINING UNTUK PREDIKSI NILAI TUKAR PETANI TANAMAN PANGAN DI INDONESIA DENGAN METODE LINEAR REGRESSION DAN SUPPORT VECTOR MACHINE Kennardi 1) dan Akhmad Budi 2).” [Online]. Available: <http://www.kwikkiangie.ac.id>
- [14] N. Putu, A. Widiari, M. Agus, D. Suarjaya, and D. Putra Githa, “Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali.”
- [15] M. A. Wiratama and W. M. Pradnya, “Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes,” *Jurnal Nasional Pendidikan*

- Teknik Informatika (JANAPATI)*, vol. 11, no. 1, p. 1, Apr. 2022, doi: 10.23887/janapati.v11i1.45282.
- [16] P. Purnamasari Wahid Suyitno, R. Eko Indrajit, M. Fauzi, P. STMIK Nusa Mandiri Jakarta, A. Institute Perbanas, and S. Bumigora Mataram, "PENERAPAN DATA MINING DALAM MENANGANI KEMACETAN DI JAKARTA," 2017.

BIOGRAFI

Thomas Febrian, lahir di Tangerang, Banten pada tanggal 17 Februari 2000. Menyelesaikan kuliah Strata I (S1) pada tahun 2022 pada Fakultas Sains dan Teknologi, program studi Teknik Informatika di Universitas Buddhi Dharma, Tangerang.

Rino, Saat ini bekerja sebagai dosen tetap pada Program Studi Teknik Informatika di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

Hartana Wijaya, Saat ini bekerja sebagai dosen tetap pada Program Studi Teknik Informatika di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.