



Penerapan *FP-Growth* dan *Random Forest* dalam Analisis Data Penjualan Makanan Ringan

Irfan Ricky Afandi¹, Irma Wahyuningtyas², Sewin Fathurrohman³,
Firman Noor Hasan^{4*}

^{1,4}*Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka,
Jl. Tanah Merdeka, Jakarta 13830, Indonesia*

²*Pendidikan Matematika, Universitas Muhammadiyah Prof. Dr. Hamka,
Jl. Tanah Merdeka, Jakarta 13830, Indonesia*

³*Sistem Informasi, Universitas Gunadarma,
Jl. Margonda, Depok 16431, Indonesia*

*Email Penulis Koresponden: firman.noorhasan@uhamka.ac.id

Abstrak:

Penelitian ini bertujuan untuk menganalisis pola pembelian produk makanan ringan serta memprediksi penjualan produk dengan menggunakan pendekatan *data mining* dan *machine learning*. Dalam industri makanan ringan yang semakin kompetitif pemahaman mendalam tentang pola perilaku konsumen dan tren penjualan produk sangat penting untuk pengambilan keputusan bisnis yang lebih efektif serta peningkatan profitabilitas perusahaan. Tantangan utama dalam penelitian ini adalah mengidentifikasi variabel yang relevan dalam dataset penjualan untuk mengungkap pola asosiasi antar produk dan menghasilkan prediksi penjualan yang akurat. Metodologi yang digunakan dalam penelitian ini melibatkan algoritma *FP-Growth* untuk menemukan asosiasi produk yang sering dibeli bersamaan serta algoritma *Random Forest* untuk memprediksi penjualan berdasarkan data historis. Hasil penelitian dari penerapan algoritma *FP-Growth* mampu mengidentifikasi sembilan aturan asosiasi yang potensial untuk diterapkan dalam sistem rekomendasi produk untuk menyediakan rekomendasi produk yang lebih personal kepada konsumen. Selain itu, model prediksi menggunakan *Random Forest* menunjukkan performa yang baik dengan nilai *Mean Absolute Error* (MAE) sebesar 23,54, *Root Mean Squared Error* (RMSE) sebesar 36,36 dan *R-squared* sebesar 0,86 dengan keseluruhan menunjukkan tingkat akurasi yang cukup baik. Penelitian ini memberikan kontribusi penting dalam optimasi stok dan strategi pemasaran berbasis data. Penelitian lanjutan disarankan menggunakan data yang lebih bervariasi dan periode waktu yang lebih panjang untuk meningkatkan akurasi prediksi.



Kata Kunci:

FP-Growth;
Random Forest;
Rekomendasi Produk;
Prediksi Penjualan

Riwayat Artikel:

Diserahkan 30 September, 2024
Direvisi 28 April, 2025
Diterima 02 Mei, 2025

DOI:

10.22441/incomtech.v15i1.30260

1. PENDAHULUAN

Kemajuan pesat dalam teknologi informasi serta peningkatan volume data telah menghadirkan tantangan sekaligus peluang baru di berbagai sektor industri termasuk pada industri makanan ringan. Data penjualan yang semakin kompleks dan heterogen membutuhkan pendekatan analitik yang lebih canggih untuk mendukung pengambilan keputusan bisnis yang strategis [1]. Jika tidak dikelola dengan tepat data tersebut bisa berpotensi menyulitkan perusahaan dalam mengidentifikasi pola perilaku konsumen dan memprediksi permintaan produk secara akurat, dimana pada akhirnya dapat berdampak negatif terhadap efisiensi operasional dan profitabilitas.

Pada era digital penggunaan metode *data mining* dan *machine learning* semakin diperlukan terutama untuk mengungkap pola tersembunyi dalam data penjualan dan memperkirakan tren di masa depan [2]. Salah satu tantangan utama dalam analisis data penjualan adalah banyaknya variabel yang harus dipertimbangkan dalam dataset transaksi yang beragam dan besar [3]. Pemilihan variabel yang tidak tepat dapat mengakibatkan bias dalam hasil analisis, sehingga mengurangi akurasi prediksi serta keandalan strategi bisnis yang dihasilkan [4]. Oleh karena itu, dibutuhkan pendekatan yang lebih efisien dan akurat dalam pemrosesan dan analisis data penjualan.

Salah satu pendekatan populer dalam analisis data penjualan adalah penerapan algoritma *FP-Growth (Frequent Pattern Growth)* untuk menemukan pola pembelian atau *frequent itemsets* yang sering muncul bersamaan dalam data transaksi [5]. Teknik ini mampu membantu dalam menghasilkan aturan asosiasi yang kemudian dapat dimanfaatkan untuk merekomendasikan produk kepada konsumen berdasarkan pola pembelian sebelumnya [6]. Penggunaan algoritma ini memungkinkan perusahaan untuk tidak hanya memahami tren pembelian tetapi juga memberikan rekomendasi produk yang lebih relevan kepada pelanggan, sehingga meningkatkan peluang penjualan silang (*cross-selling*) [7].

Penggunaan *FP-Growth* telah berhasil diimplementasikan dalam berbagai bidang. Sebagai contoh penelitian oleh Anil dan Venkatesh (2024) [8] menerapkan *FP-Growth* untuk mengidentifikasi pola pembelian pada data transaksi di sebuah restoran. Hasilnya menunjukkan bahwa *FP-Growth* dapat secara efisien menemukan asosiasi produk makanan dan minuman yang sering dibeli bersamaan, sehingga dapat mendukung strategi pemasaran melalui rekomendasi dan promosi produk yang relevan. Studi ini membuktikan bahwa algoritma *FP-Growth* mampu menangani dataset besar dengan efisiensi yang tinggi.

Selain memahami pola pembelian, penting juga bagi perusahaan untuk dapat memprediksi permintaan produk di masa mendatang [9]. *Random Forest* sebuah algoritma berbasis *ensemble learning* yang terkenal dengan performa prediktif yang efektif dalam menangani dataset yang kompleks dan menghasilkan prediksi yang akurat [10]. Dalam konteks ini *Random Forest* dapat digunakan untuk memprediksi penjualan setiap produk di masa mendatang yang dimana dapat membantu perusahaan dalam mengelola persediaan stok secara lebih efisien [11]. Prediksi yang tepat dapat meminimalkan risiko kekurangan atau kelebihan stok yang sering kali menjadi masalah dalam pengelolaan rantai pasok pada produk [12].

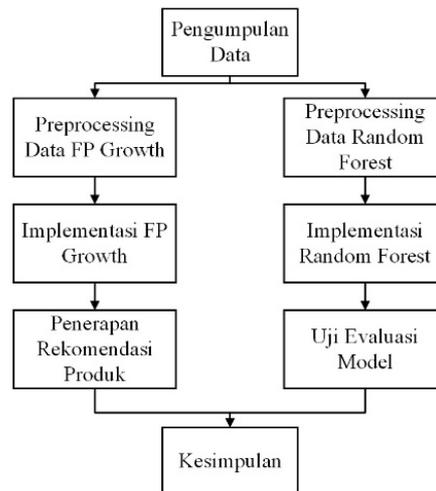
Penelitian sebelumnya oleh Huishan dan Zihao (2024) [13] menunjukkan bahwa algoritma *Random Forest* berhasil diterapkan dalam memprediksi volume

pendapatan dan strategi harga sayuran. *Random Forest* digunakan untuk menggabungkan berbagai fitur prediktif sehingga mampu mengatasi data yang mengandung *noise* dan risiko *overfitting* dan menghasilkan prediksi yang lebih akurat. Hasil penelitian ini membuktikan bahwa *Random Forest* efektif dalam meningkatkan akurasi prediksi pendapatan dan membantu supermarket memaksimalkan keuntungan melalui strategi penetapan harga yang lebih optimal.

Berdasarkan penelitian terdahulu, proses menerapkan kombinasi algoritma *FP-Growth* dan *Random Forest* dalam analisis data penjualan pada industri makanan ringan dengan variabel yang berbeda masih belum banyak dijelajahi dalam literatur sebelumnya. Oleh karena itu, penelitian ini bertujuan untuk mengaplikasikan kombinasi algoritma *FP-Growth* dan *Random Forest* dalam analisis data penjualan makanan ringan. Penelitian ini akan mengeksplorasi bagaimana algoritma *FP-Growth* dapat digunakan untuk mengidentifikasi pola pembelian dalam dataset transaksi makanan ringan, serta bagaimana algoritma *Random Forest* dapat memprediksi penjualan di masa mendatang. Penggabungan kedua metode ini diharapkan dapat mengatasi keterbatasan masing-masing pendekatan dan memberikan analisis yang lebih komprehensif terhadap data penjualan dalam bidang industri makan ringan. Urgensi penelitian ini semakin signifikan mengingat dinamika persaingan di industri makanan ringan, di mana penggunaan data yang efisien menjadi kunci untuk memenangkan persaingan pasar. Analisis yang lebih mendalam terhadap pola penjualan dan prediksi permintaan produk yang akurat akan membantu perusahaan dalam mengelola persediaan dengan lebih efisien, meminimalkan risiko kekurangan maupun kelebihan stok, serta mengoptimalkan keuntungan melalui strategi pemasaran yang lebih tepat sasaran. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoritis bagi pengembangan metode analisis data penjualan tetapi juga memberikan manfaat praktis bagi industri makanan ringan dalam pengambilan keputusan yang berbasis data.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif berbasis *data mining* dan *machine learning* untuk menganalisis data penjualan makanan ringan. Analisis ini melibatkan penerapan algoritma *FP-Growth* untuk menemukan pola asosiasi antar produk serta algoritma *Random Forest* untuk memprediksi penjualan di masa mendatang. Penelitian dilakukan pada platform *Google Colab* menggunakan bahasa pemrograman *Python* dengan *library mlxtend* dan *scikit-learn* sebagai instrumen utama dalam implementasi kedua algoritma. Proses penelitian dilakukan secara berurutan seperti diagram alir pada Gambar 1 yang dimulai dari pengumpulan data hingga pembuatan kesimpulan penelitian.



Gambar 1. Diagram Alir Penelitian

Pada Gambar 1 di atas menunjukkan diagram alir yang menjelaskan tahapan utama yang harus dilakukan dalam penelitian ini. Proses penelitian diawali dengan pengumpulan data penjualan makanan ringan yang diambil dari dataset *Kaggle*. Setelah data terkumpul, dilakukan dua jalur analisis paralel yaitu untuk penerapan algoritma *FP-Growth* dan lainnya untuk algoritma *Random Forest*. Pada jalur *FP-Growth* data penjualan yang telah di-*preprocessing* digunakan untuk menemukan aturan asosiasi antar produk. Data yang tidak relevan akan dihapus dan semua transaksi disatukan untuk mempersiapkan data yang akan dianalisis menggunakan algoritma *FP-Growth*. Hasil dari penerapan algoritma ini berupa aturan asosiasi yang kemudian digunakan untuk membangun sistem rekomendasi produk. Sementara itu, jalur *Random Forest* berfokus pada prediksi penjualan. Setelah data di-*preprocessing* dengan menyederhanakan variabel-variabel yang relevan, model prediksi penjualan dibangun menggunakan algoritma *Random Forest*. Data akan dipecah menjadi data pelatihan dan pengujian untuk memastikan model dapat memprediksi penjualan di bulan mendatang secara akurat. Setelah kedua jalur analisis selesai, hasil dari kedua pendekatan ini dikombinasikan untuk memberikan rekomendasi produk berdasarkan pola asosiasi dan memprediksi penjualan di masa mendatang. Tahap akhir penelitian ini adalah menarik kesimpulan dari hasil analisis yang dilakukan dengan harapan memberikan wawasan yang berguna bagi strategi promosi produk dan optimalisasi manajemen stok produk.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari platform *Kaggle*, mencakup transaksi penjualan makanan ringan dari Januari hingga Agustus 2024. Dataset berisi 1678 data transaksi dengan variabel seperti ID Transaksi, Tanggal Pembelian, Nama Produk, Varian Rasa, Jumlah Terjual, Harga per Unit, dan Total Harga. Produk yang dijual dalam dataset ini terdiri dari 23 jenis makanan ringan, mulai dari Keripik Tempe hingga makanan olahan lainnya. Data ini digunakan sebagai dasar untuk analisis asosiasi dan prediksi penjualan. Tabel 1 menyajikan contoh lima data transaksi penjualan produk makanan ringan yang ada pada dataset untuk memberikan gambaran lebih jelas mengenai variasi transaksi penjualan yang digunakan dalam penelitian ini.

Tabel 1. *Sample Data* Transaksi Penjualan Makanan Ringan

ID Transaksi	ID Pelanggan	Tanggal Pembelian	Nama Produk	Varian Rasa	Jumlah Terjual	Harga per Unit	Total Harga
TR0001	PL0001	2024-01-01	Kacang Goreng		5	10.000	50.000
TR0024	PL0016	2024-01-16	Basreng	Original	2	15.000	30.000
TR0024	PL0016	2024-01-16	Keripik Bawang	Balado	3	10.000	30.000
TR0031	PL0027	2024-02-08	Kacang Atom		4	10.000	40.000
TR0031	PL0027	2024-02-08	Basreng	Pedas	1	15.000	15.000
TR0037	PL0005	2024-04-08	Makaroni Goreng	Balado	2	15.000	30.000
TR0037	PL0005	2024-04-08	Sistik	Pedas	1	15.000	15.000
TR0037	PL0005	2024-04-08	Sistik	Balado	3	15.000	45.000
TR0066	PL00041	2024-03-27	Kacang Atom		2	10.000	20.000
TR0066	PL00041	2024-03-27	Kacang Goreng		1	10.000	10.000
TR0066	PL00041	2024-03-27	Keripik Tempe		1	10.000	10.000
TR0066	PL00041	2024-03-27	Keripik Singkong	Pedas	5	15.000	75.000

2.2 Preprocessing Data *FP-Growth*

Tahap *preprocessing* untuk algoritma *FP-Growth* dilakukan dengan menghilangkan kolom yang tidak relevan seperti ID Pelanggan, Tanggal Pembelian, dan Total Harga. Kolom Nama Produk dan Varian Rasa digabungkan menjadi satu kolom baru yaitu kolom Produk Lengkap guna merepresentasikan produk secara lebih spesifik. Setiap transaksi disatukan dalam satu daftar produk yang dibeli untuk mempersiapkan data dalam format yang sesuai bagi algoritma *FP-Growth* dalam menemukan pola asosiasi antar produk. Tabel 2 menunjukkan hasil dari *preprocessing* data untuk penerapan algoritma *FP-Growth*.

Tabel 2. Hasil *Preprocessing* Data untuk *FP-Growth*

ID Transaksi	Produk Lengkap
TR0001	Kacang Goreng
TR0024	Basreng Original, Keripik Bawang Balado
TR0031	Kacang Atom, Basreng Pedas
TR0037	Makaroni Goreng Balado, Sistik Pedas, Sistik Balado
TR0066	Kacang Atom, Kacang Goreng, Keripik Tempe, Keripik Singkong Pedas

2.3 Implementasi *FP-Growth*

Penerapan algoritma *FP-Growth* dilakukan menggunakan *library mlxtend* pada *Python*. Setelah data di *preprocessing* dilakukan *encoding* pada data transaksi untuk mengubahnya ke dalam format numerik yang dapat diproses oleh algoritma. Algoritma *FP-Growth* kemudian diterapkan dengan menggunakan parameter minimum *support* sebesar 20%, *confidence* 60%, dan *lift* 1. *Support*, *confidence* dan

lift adalah metrik penting dalam algoritma *FP-Growth* [5]. *Support* mengukur seberapa sering *itemset* muncul dalam dataset dan diperlukan untuk menemukan pola yang sering terjadi [14]. Dalam penelitian ini minimum *support* ditetapkan sebesar 20% untuk memastikan hanya aturan yang cukup umum yang akan dipertimbangkan. *Confidence* mengukur kekuatan aturan asosiasi [15], yaitu seberapa sering *consequents* (produk yang direkomendasikan) muncul saat *antecedents* (produk awal) terjadi. Minimum *confidence* sebesar 60% dipilih untuk memastikan aturan yang dihasilkan memiliki tingkat kepercayaan yang tinggi. *Lift* mengukur kekuatan hubungan antara *antecedents* dan *consequents* dibandingkan dengan asumsi acak [16] dan minimum *lift* sebesar 1 memastikan hanya aturan dengan dampak nyata dan saling terkait yang dipertimbangkan. Kombinasi nilai-nilai minimum ini memastikan aturan yang dihasilkan relevan dan dapat diandalkan. Hasil dari *FP-Growth* berupa aturan asosiasi yang disajikan dalam tabel yang mencantumkan kolom *antecedents* dan *consequents* beserta kolom nilai *support*, *confidence* dan *lift* untuk setiap aturan.

2.4 Penerapan Rekomendasi Produk

Berdasarkan aturan asosiasi yang dihasilkan oleh algoritma *FP-Growth*, dibangun sebuah sistem rekomendasi produk. Sistem ini bekerja dengan mencocokkan produk yang dibeli dalam setiap transaksi dengan aturan asosiasi untuk merekomendasikan produk tambahan yang relevan. Hasil rekomendasi diuji pada dataset penjualan sebelumnya dengan menambahkan kolom baru yaitu Rekomendasi Produk dan hasilnya dievaluasi secara manual untuk memastikan efektivitas sistem berdasarkan aturan asosiasi yang ada.

2.5 Preprocessing Data Random Forest

Pada penerapan algoritma *Random Forest*, *preprocessing* data melibatkan penghapusan kolom yang tidak relevan seperti ID Transaksi, ID Pelanggan, Harga Per Unit, Dan Total Harga. Kolom Nama Produk dan Varian Rasa juga digabungkan untuk memberikan identifikasi produk yang lebih tepat. Selain itu, kolom Tanggal Pembelian diubah menjadi bulan dan tahun guna menyederhanakan analisis berdasarkan waktu. Penjumlahan jumlah penjualan per produk setiap bulan dilakukan sebagai langkah persiapan data untuk prediksi penjualan bulan berikutnya. Tabel 3 menyajikan hasil dari *preprocessing* data untuk penerapan algoritma *Random Forest*.

Tabel 3. Hasil *Preprocessing* Data untuk *Random Forest*

Bulan Pembelian	Produk Lengkap	Jumlah Terjual
2024-01	Basreng Original	373
2024-01	Basreng Pedas	304
2024-01	Kacang Atom	653
2024-01	Kacang Goreng	678
2024-01	Keripik Bawang Balado	303

2.6 Implementasi Random Forest

Algoritma *Random Forest* diterapkan menggunakan *library scikit-learn* melalui modul *RandomForestRegressor*. Dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian di mana pembagian ini dilakukan untuk memberikan model cukup data untuk belajar sambil menjaga akurasi evaluasi pada data yang belum

pernah dilihat. Model dibangun untuk memprediksi jumlah penjualan pada bulan September 2024. Prediksi ini digunakan untuk mengantisipasi kebutuhan stok produk di masa depan. Hasil prediksi ditampilkan dalam kolom baru yaitu kolom Prediksi Terjual yang menggambarkan hasil prediksi yang sudah dibuat oleh model.

2.7 Uji Evaluasi Model

Evaluasi model dilakukan menggunakan tiga metrik utama yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared*. MAE mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi [17], sementara RMSE memberikan bobot lebih besar pada kesalahan besar [18] dan *R-squared* digunakan untuk mengukur seberapa baik variabel independen menjelaskan variabilitas variabel dependen dengan nilai mendekati 1 menunjukkan prediksi yang lebih akurat [19]. Kombinasi metrik ini memberikan gambaran menyeluruh tentang kinerja model prediksi penjualan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi *FP-Growth*

Setelah melalui tahapan *preprocessing* data dan penerapan algoritma *FP-Growth* dengan menggunakan bahasa pemrograman *Python* dan *library mlxtend*, penelitian ini menghasilkan aturan asosiasi antara produk yang sering dibeli bersamaan dalam transaksi penjualan. Hasil penerapan algoritma *FP-Growth* disajikan dalam Tabel 4 yang mencakup kolom *antecedents*, *consequents*, *support*, *confidence*, dan *lift*. Setiap kolom memberikan informasi mengenai hubungan antara produk dalam transaksi.

Tabel 4. Hasil Aturan Asosiasi Algoritma *FP-Growth*

Antecedents	Consequents	Support	Confidence	Lift
Kacang Atom	Kacang Goreng	0.2843	0.6252	1.3229
Kacang Atom	Keripik Pisang	0.2771	0.6094	1.3044
Kacang Atom	Keripik Tempe	0.2759	0.6068	1.3173
Kacang Atom, Keripik Pisang	Kacang Goreng	0.2068	0.7462	1.5790
Kacang Atom, Keripik Tempe	Kacang Goreng	0.2050	0.7430	1.5722
Kacang Goreng	Kacang Atom	0.2843	0.6015	1.3229
Kacang Goreng, Kacang Atom	Keripik Pisang	0.2068	0.7275	1.5570
Kacang Goreng, Kacang Atom	Keripik Tempe	0.2050	0.7212	1.5655
Kacang Goreng, Keripik Pisang	Kacang Atom	0.2068	0.7462	1.6411
Kacang Goreng, Keripik Tempe	Kacang Atom	0.2050	0.7511	1.6518

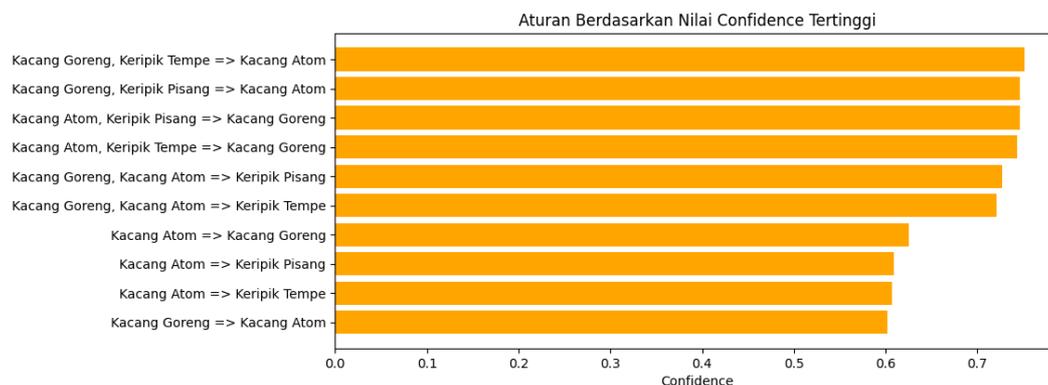
Tabel 4 menampilkan sembilan aturan asosiasi yang menunjukkan produk-produk yang sering dibeli bersamaan. Sebagai contoh aturan pada kolom pertama menunjukkan bahwa Kacang Atom dan Kacang Goreng sering dibeli bersamaan dengan nilai *support* sebesar 0,2843, *confidence* sebesar 0,6252 dan *lift* sebesar 1,3229. Nilai *support* mencerminkan frekuensi kemunculan kedua produk dalam transaksi, sedangkan *confidence* menunjukkan seberapa sering Kacang Goreng dibeli setelah Kacang Atom. Nilai *lift* yang lebih besar dari 1 menunjukkan adanya hubungan positif yang signifikan antara kedua produk tersebut yang berarti mereka lebih mungkin dibeli bersama daripada secara acak. Selain itu dalam mempermudah proses analisis data, peneliti juga membuat diagram batang yang

memvisualisasikan nilai *support* yang ditampilkan pada Gambar 2, *confidence* yang ditampilkan pada Gambar 3 dan *lift* yang ditampilkan pada gambar 4 untuk setiap aturan asosiasi dan diurutkan berdasarkan nilai tertinggi dari masing-masing metrik tersebut.



Gambar 2. Diagram Batang Nilai *Support*

Diagram batang yang divisualisasikan dalam Gambar 2 menggambarkan nilai *support* untuk setiap aturan asosiasi. Aturan dengan pasangan Kacang Goreng dan Kacang Atom memiliki nilai *support* tertinggi yaitu 0,2843 yang menunjukkan bahwa sekitar 28,43% dari seluruh transaksi melibatkan kedua produk ini. Visualisasi ini membantu mengidentifikasi produk yang sering muncul bersamaan serta memberikan wawasan untuk strategi promosi dan pengelompokan produk.



Gambar 3. Diagram Batang Nilai *Confidence*

Pada Gambar 3 yaitu diagram batang yang memperlihatkan nilai *confidence* untuk masing-masing aturan asosiasi. Aturan dengan *confidence* tertinggi melibatkan Kacang Goreng dan Keripik Tempe sebagai *antecedents*, yang memprediksi Kacang Atom sebagai *consequents* dengan nilai *confidence* sebesar 0,7511. Ini berarti bahwa pada 75,11% transaksi yang melibatkan Kacang Goreng dan Keripik Tempe dimana juga akan membeli Kacang Atom. Hubungan yang kuat ini mengindikasikan potensi tinggi produk tersebut dalam paket promosi.

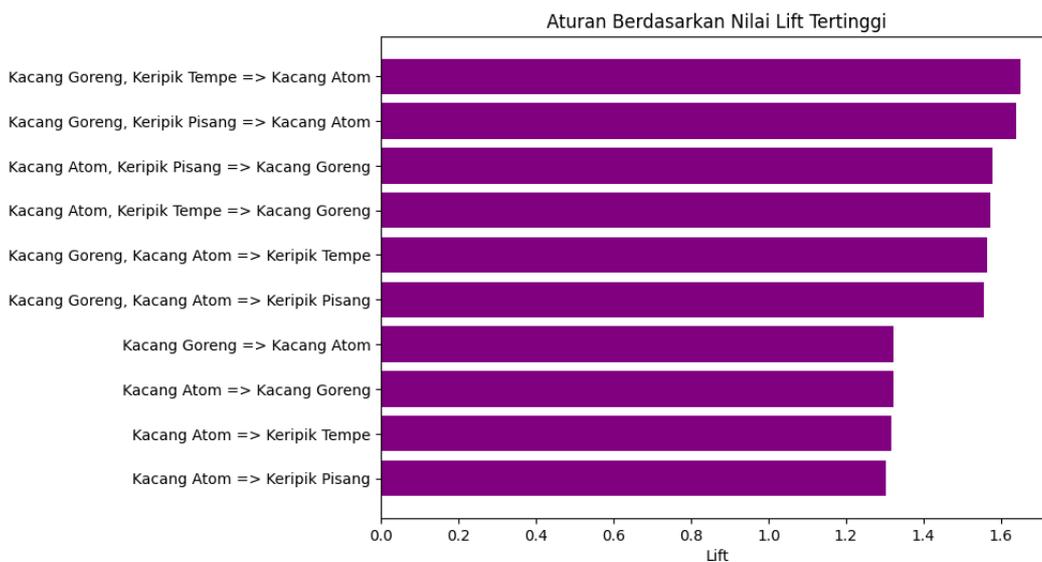
Gambar 4. Diagram Batang Nilai *Lift*

Diagram batang pada Gambar 4 menunjukkan nilai *lift* untuk setiap aturan asosiasi. Aturan dengan *lift* tertinggi adalah Kacang Goreng, Keripik Tempe sebagai *antecedents* yang memprediksi Kacang Atom sebagai *consequents* dengan nilai *lift* sebesar 1.6518. Nilai *lift* yang lebih dari 1 ini menunjukkan adanya hubungan positif antara produk-produk tersebut, di mana pembelian Kacang Goreng dan Keripik Tempe bersama-sama meningkatkan kemungkinan pembelian Kacang Atom. Semakin tinggi nilai *lift*, semakin kuat hubungan antara produk-produk tersebut dalam transaksi.

3.2 Hasil Penerapan Rekomendasi Produk

Setelah menghasilkan aturan asosiasi melalui algoritma *FP-Growth*, sistem rekomendasi produk diimplementasikan pada dataset penjualan yang sudah melalui tahap *preprocessing*. Sistem ini menggunakan aturan asosiasi yang dihasilkan untuk memberikan rekomendasi produk tambahan berdasarkan produk yang ditemukan dalam transaksi. Hasil rekomendasi produk ditampilkan pada Tabel 5 yang memuat kolom ID Transaksi, Produk Lengkap, dan Rekomendasi Produk.

Tabel 5. Hasil Penerapan Rekomendasi Produk

ID Transaksi	Produk Lengkap	Rekomendasi Produk
TR0001	Kacang Goreng	Kacang Atom
TR0024	Basreng Original, Keripik Bawang Balado	Tidak Ada Rekomendasi
TR0031	Kacang Atom, Basreng Pedas	Kacang Goreng, Keripik Pisang, Keripik Tempe
TR0037	Makaroni Goreng Balado, Sistik Pedas, Sistik Balado	Tidak Ada Rekomendasi
TR0066	Kacang Atom, Kacang Goreng, Keripik Tempe, Keripik Singkong Pedas	Tidak Ada Rekomendasi

Berdasarkan Tabel 5 sistem rekomendasi bekerja dengan baik sesuai dengan aturan asosiasi. Pada transaksi TR0001 yang berisi Kacang Goreng dimana sistem merekomendasikan Kacang Atom karena kedua produk sering dibeli bersamaan

berdasarkan aturan asosiasi. Sebaliknya, transaksi TR0024 dan TR0037 tidak menghasilkan rekomendasi karena produk yang dibeli tidak terhubung dengan aturan asosiasi yang ada. Pada transaksi TR0031 yang mencakup Kacang Atom, sistem merekomendasikan beberapa produk seperti Kacang Goreng, Keripik Pisang dan Keripik Tempe berdasarkan aturan asosiasi yang mendukung *multiple consequents*. Transaksi TR0066 tidak menghasilkan rekomendasi baru karena produk yang dibeli sudah mencakup semua produk yang terkait dalam aturan asosiasi.

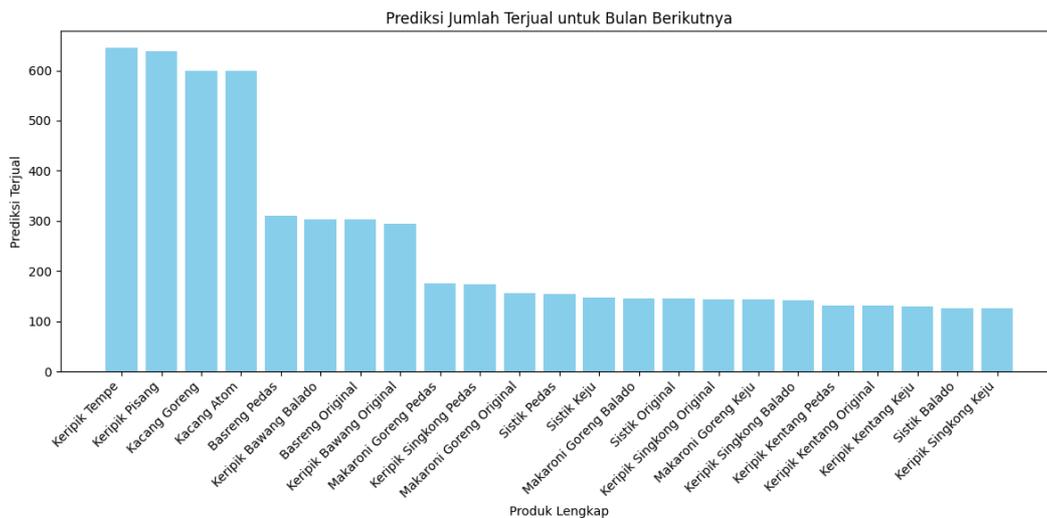
3.3 Hasil Implementasi *Random Forest*

Proses prediksi penjualan dilakukan menggunakan algoritma *Random Forest* melalui library *scikit-learn* pada bahasa pemrograman *Python*, setelah data sudah di-*preprocessing*. Model prediksi dibangun untuk memperkirakan jumlah penjualan produk pada bulan berikutnya. Hasil prediksi ditampilkan dalam Tabel 6 yang berisi kolom Produk Lengkap dan Prediksi Jumlah Terjual.

Tabel 6. Hasil Prediksi Penjualan Produk Algoritma *Random Forest*

Produk Lengkap	Prediksi Terjual
Keripik Tempe	646
Keripik Pisang	638
Kacang Goreng	600
Kacang Atom	599
Basreng Pedas	311
Keripik Bawang Balado	304
Basreng Original	303
Keripik Bawang Original	295
Makaroni Goreng Pedas	176
Keripik Singkong Pedas	174
Makaroni Goreng Original	156
Sistik Pedas	154
Sistik Keju	147
Makaroni Goreng Balado	145
Sistik Original	145
Keripik Singkong Original	144
Makaroni Goreng Keju	144
Keripik Singkong Balado	141
Keripik Kentang Pedas	132
Keripik Kentang Original	131
Keripik Kentang Keju	130
Sistik Balado	126
Keripik Singkong Keju	126

Tabel 6 menunjukkan bahwa produk seperti Keripik Tempe, Keripik Pisang dan Kacang Goreng diprediksi akan memiliki jumlah penjualan tertinggi, sementara produk seperti Keripik Kentang Keju, Sistik Balado, dan Keripik Singkong Keju diprediksi akan memiliki prediksi penjualan terendah. Peneliti juga membuat gambar diagram batang yang ditampilkan pada Gambar 5 berdasarkan data pada tabel prediksi untuk memperlihatkan jumlah prediksi produk terjual yang diurutkan dari yang tertinggi hingga terendah.



Gambar 5. Diagram Batang Prediksi Penjualan Produk

Gambar 5 menampilkan diagram batang yang menggambarkan jumlah produk yang diprediksi akan terjual dari yang tertinggi hingga yang terendah, dimana Keripik Tempe diprediksi akan terjual tertinggi yaitu dengan jumlah 646 unit sementara Keripik Singkong Keju diperkirakan akan terjual terendah yaitu dengan jumlah 126 unit. Prediksi ini memberikan wawasan yang berguna bagi perencanaan stok dan distribusi produk pada bulan mendatang, di mana produk dengan penjualan tinggi dapat diprioritaskan dalam penyediaan stok.

3.4 Hasil Uji Evaluasi Model

Evaluasi model prediksi dilakukan menggunakan beberapa metrik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared*. Berdasarkan evaluasi memperoleh nilai MAE sebesar 23.535 menunjukkan rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual. Artinya, prediksi model bisa meleset sekitar 23 unit dari nilai penjualan sebenarnya. Evaluasi model juga menghasilkan nilai RMSE sebesar 36,364 menunjukkan seberapa jauh prediksi model dari nilai aktual dengan kesalahan rata-rata sekitar 36 unit, mengindikasikan bahwa model dapat memberikan gambaran realistis mengenai kesalahan prediksi. Selain itu evaluasi model juga menghasilkan nilai *R-squared* sebesar 0,8575 menunjukkan bahwa 85,76% variasi dalam data penjualan dapat dijelaskan oleh model. Nilai ini mendekati 1 yang menandakan bahwa model *Random Forest* yang digunakan memiliki kinerja yang cukup baik dalam memprediksi penjualan produk untuk bulan berikutnya.

4. KESIMPULAN

Penelitian ini berhasil mengaplikasikan algoritma *FP-Growth* dan *Random Forest* dalam analisis data penjualan produk makanan ringan, dimana secara efektif dapat mengidentifikasi pola asosiasi antar produk dan memberikan prediksi penjualan di masa mendatang. Hasil implementasi *FP-Growth* menunjukkan adanya aturan asosiasi yang signifikan antara beberapa produk yang sering dibeli

bersamaan yang dapat dimanfaatkan untuk mengembangkan sistem rekomendasi produk yang lebih relevan dengan preferensi konsumen. Temuan ini menekankan pentingnya memanfaatkan data transaksi untuk mendukung pengambilan keputusan strategis dalam pemasaran dan penyusunan *bundling* produk. Pada aspek prediksi penjualan, algoritma *Random Forest* terbukti memberikan estimasi yang akurat untuk jumlah produk yang akan terjual pada bulan berikutnya. Evaluasi performa model dengan menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE) dan *R-squared* menunjukkan hasil yang cukup baik di mana nilai MAE sebesar 23,54, RMSE sebesar 36,36 dan *R-squared* sebesar 0,86 mencerminkan bahwa model mampu menjelaskan sebagian besar variabilitas dalam data penjualan. Model ini memberikan gambaran penting bagi perencanaan stok dan strategi distribusi produk yang lebih baik.

Secara keseluruhan, penelitian ini menunjukkan bahwa integrasi metode *data mining* dan *machine learning* dalam analisis penjualan dapat memberikan wawasan yang mendalam untuk pengelolaan stok dan optimalisasi strategi pemasaran. Sistem rekomendasi berbasis aturan asosiasi yang dikembangkan dapat meningkatkan relevansi dalam rekomendasi produk, sementara model prediksi yang dibangun mampu mendukung perencanaan permintaan dengan lebih tepat. Sebagai langkah selanjutnya, disarankan untuk memperluas analisis dengan mempertimbangkan data yang lebih variatif, serta melakukan evaluasi pada periode yang lebih panjang guna meningkatkan akurasi model. Penggunaan metode *ensemble* yang menggabungkan beberapa algoritma prediktif juga dapat menjadi strategi untuk memperbaiki performa model dalam memprediksi permintaan produk secara lebih komprehensif.

REFERENSI

- [1] F. N. Hasan and R. Ariyansah, "Utilization of the FP-Growth Algorithm on MSME Transaction Data: Recommendations for Small Gifts from The Padang Region," *J. Tek. Inform.*, vol. 17, no. 1, p. 71, 2024, doi: 10.15408/jti.v17i1.37966.
- [2] U. Arfan and N. Paraga, "The Comparison of K-Means, Naïve Bayes and Decision Tree Algorithm in Predicting Fuel Oil Sales," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, p. 1379, 2024, doi: <https://doi.org/10.57152/malcom.v4i4.1566>.
- [3] X. Han, J. Sun, and W. Wang, "Machine learning-based analysis of factors influencing the popularity of vitamin online sales," in *Proceedings of the 2023 4th International Conference on Big Data Economy and Information Management*, 2023, p. 819. doi: <https://doi.org/10.1145/3659211.3659352>.
- [4] Y. Li, H. Zhao, Y. Wang, and B. Yu, "Research on Vegetable Sales Data Analysis and Cost Pricing Modeling," in *3rd International Conference on Governance of Accounting and Global Business Management (GAGBM 2024)*, 2024, vol. 29, p. 155. doi: 10.54097/pz7q1302.
- [5] M. A. Zarkhasy and C. Satria, "Determining and Managing Stock of Goods Based on Purchasing Patterns Using the Frequent Pattern Growth Algorithm," *Int. J. Eng. Comput. Sci. Appl.*, vol. 3, no. 1, p. 4, 2023, doi: 10.30812/ijecsa.v3i1.3416.
- [6] Y. Liu, "Big Data Mining Method of New Retail Economy Based on Association Rules," in *Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022)*, 2023, p. 1584. doi: 10.2991/978-94-6463-030-5_159.
- [7] H. Hairani and J. Ximenes Guterres, "Exploring Customer Purchasing Patterns: A Study Utilizing FP-Growth Algorithm on Supermarket Transaction Data," *Int. J. Eng. Comput. Sci. Appl.*, vol. 3, no. 1, p. 34, 2024, doi: 10.30812/ijecsa.v3i1.3874.
- [8] A. K. Sah and V. K., "Predictive Modeling for Restaurant Menu Customization: An FP-Growth Algorithm-Based Solution," in *2024 IEEE International Students' Conference on*

- Electrical, Electronics and Computer Science (SCEECS)*, 2024, p. 5. doi: 10.1109/SCEECS61402.2024.10482175.
- [9] R. Ajay, R. S. Joel, and P. G. O. Prakash, "Analyzing and Predicting the Sales Forecasting using Modified Random Forest and Decision Tree Algorithm," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, 2023, p. 1645. doi: 10.1109/ICCES57224.2023.10192723.
- [10] A. Zikri, A. Nazir, S. Sanjaya, E. Haerani, and I. Afrianty, "The Random Forest algorithm for classifying stunting in toddlers based on anthropometric data," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 5, no. 3, p. 932, 2024, doi: 10.54660/ijmrge.2024.5.3.931-937.
- [11] C. Jia, "Prediction method of product market demand based on Prophet random forest," *Int. J. Prod. Dev.*, vol. 28, no. 1–2, p. 61, 2024, doi: <https://doi.org/10.1504/IJPD.2024.137814>.
- [12] S. Wu, Z. Zhang, and Y. Ru, "Research on Product Demand Forecasting Based on Random Forest and ARIMA Time Series: Precision Forecasting Method for Data-Scarce Environments," in *Transactions on Computer Science and Intelligent Systems Research*, 2024, p. 1328. doi: <https://doi.org/10.62051/g9r9ca46>.
- [13] H. Zhang and Z. Yan, "Vegetable Price Forecasting Based on ARIMA Model and Random Forest Prediction," *J. Educ. Humanit. Soc. Sci.*, vol. 25, no. 1, p. 82, 2024, doi: 10.54097/3afpgv27.
- [14] R. Xu, "the Evaluation of Ethnic Costume Courses Based on Fp-Growth Algorithm," *Scalable Comput. Pract. Exp.*, vol. 25, no. 1, p. 316, 2024, doi: 10.12694/scpe.v25i1.2297.
- [15] X. Li and M. F. Rosas, "Career Recommendation System Design Based on FP-growth Algorithm," in *Proceedings of the 3rd International Conference on Internet Technology and Educational Informatization, ITEI 2023*, 2024, p. 4. doi: 10.4108/eai.24-11-2023.2343709.
- [16] H. Patil, T. Mukherjee, K. Pandit, H. Jani, P. K. Jha, and V. Agarwal, "Enhancing Retail Strategies through Apriori, ECLAT& FP Growth Algorithms in Market Basket Analysis," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 9, p. 3836, 2023, doi: 10.17762/ijritcc.v11i9.9637.
- [17] S. M. K. Sistla, G. Krishnamoorthy, J. Jeyaraman, and B. K. Konidena, "Machine Learning for Demand Forecasting in Manufacturing," *Int. J. Multidiscip. Res.*, vol. 6, no. 1, p. 6, 2024, doi: 10.36948/ijfmr.2024.v06i01.14204.
- [18] J. E. A. Cabanlit, K. D. Dela Cruz, and Mathematics, "Root Mean Square Error of the Maximum Likelihood Estimate of the Parameters of Pareto Distribution," *Int. J. Sci. Adv.*, vol. 4, no. 5, p. 715, 2023, doi: 10.51542/ijscia.v4i5.6.
- [19] L. Diane and P. Brijlal, "Forecasting Stock Market Realized Volatility using Random Forest and Artificial Neural Network in South Africa," *Int. J. Econ. Financ. Issues*, vol. 14, no. 2, p. 12, 2024, doi: 10.32479/ijefi.15431.