

SKRIPSI

**KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI
ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST
BERBASIS WEB**

DISUSUN OLEH

ZAKI ZAIN RANGKUTI

2009010031



PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2025

**KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON
DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM
FOREST BERBASIS WEB
SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**ZAKI ZAIN RANGKUTI
NPM. 2009010031**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

LEMBAR PENGESAHAN

Judul Skripsi : KLASIFIKASI BARANG PROMOSI DAN HARGA
DISKON DI ALFAMIDI TURI MENGGUNAKAN
METODE RANDOM FOREST BERBASIS WEB

Nama Mahasiswa : ZAKI ZAIN RANGKUTI

NPM : 2009010031

Program Studi : SISTEM INFORMASI



Menyetujui
Dosen Pembimbing

(Halim Maulana, ST, M.Kom)
NIDN. 0121119102

Ketua Program Studi

(Dr. Firahmi Rizky, S.Kom., M.Kom)
NIDN. 0116079201



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

Unggul | Cerdas | Terpercaya

PERNYATAAN ORISINALITAS

KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST BERBASIS WEB

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Agustus 2025

Yang membuat pernyataan



Zaki Zain Rangkuti

NPM. 2009010031

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Zaki Zain Rangkuti
NPM	2009010031
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI
ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST
BERBASIS WEB**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Agustus 2025.

Yang membuat pernyataan



Zaki Zain Rangkuti

NPM. 2009010031

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Zaki Zain Rangkuti
Tempat dan Tanggal Lahir : Medan, 18 Agustus 2001
Alamat Rumah : JL. Brigjend Zein Hamid gg.Kelapa
kuning no.13 Medan
Telepon/Faks/HP : 0895333761020
E-mail : zakizain31@gmail.com

DATA PENDIDIKAN

SD : PERGURUAN HARAPAN MANDIRI (2007-2013)
SMP : PERGURUAN HARAPAN MANDIRI (2013-2016)
SMA : SMA NEGERI 13 MEDAN (2016-2019)

KATA PENGANTAR



Puji syukur Alhamdulillah kehadiran Allah Subhanahu Wa Ta'ala yang telah memberikan Nikmat dan Karunia-Nya sehingga penulis dapat menyelesaikan proposal skripsi ini dengan judul “KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST BERBASIS WEB”. Proposal ini disusun sebagai salah satu syarat untuk penulisan skripsi Program Studi Sitem Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Adapun yang perlu disampaikan penyusunan proposal ini tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ayahanda Prof. Dr. Agussani, M.AP. selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Alkhowarizmi., S.T., M.Kom selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
3. Bapak Halim Maulana, S.T., M.Kom selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
4. Bapak Lutfi Basit, S.Sos., M.I.Kom selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.

5. Ibu Dr. Firahmi Rizky. S.Kom.,M.Kom selaku Ketua Prodi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
6. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom selaku Sekretaris Prodi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
7. Bapak Halim Maulana, S.T., M.Kom selaku dosen pembimbing yang telah memberikan arahan dan bimbingan kepada penulis.
8. Bapak dan Ibu dosen Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara yang telah memberikan banyak bekal ilmu kepada penulis.
9. Bapak dan Ibu staff pegawai Biro Fakultas Ilmu Komputer dan Teknologi Informasi.
10. Bapak dan Ibu Pimpinan dan seluruh staff Pegawai yang ada di Alfamidi Turi yang telah memberikan izin dan informasi yang tidak bisa penulis sebut namanya satu persatu.
11. Teristimewa CH.Azhar Rangkuti dan Nurmala selaku Orangtua yang telah banyak memberikan dukungan moril dan materil, serta mendoakan penulis dalam setiap langkah dan usaha dalam menyelesaikan proposal ini.
12. Kakak dan Adik tersayang Amaris Adha Rangkuti dan Ridho Rizki yang telah banyak memberikan dukungan moril dan materil, serta mendoakan penulis dalam setiap langkah dan usaha dalam

menyelesaikan proposal ini.

13. Teman-teman seperjuangan Program Studi Sistem Informasi angkatan 2020 telah membantu dan selalu memberikan support yang tidak bisa penulis sebutkan namanya satu persatu
14. Dan semua pihak yang telah membantu dan tidak dapat disebut namanya satu persatu.

Penulis menyadari didalam penulisan proposal ini masih banyak kekurangan. Oleh karena itu, penulis menerima apabila ada kritik dan saran yang sifatnya dapat membantu agar penulisan ini bisa menjadi sempurna. Segala ucapan terima kasih tentunya belum cukup, semoga Allah Subhanahu Wa Ta'ala senantiasa membalas segala kebaikan Anda semua. Aamiin

KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST BERBASIS WEB

ABSTRAK

Strategi promosi dan pemberian diskon merupakan elemen penting dalam meningkatkan penjualan serta daya saing pada industri ritel. Namun, penentuan produk yang tepat untuk dipromosikan seringkali masih dilakukan secara manual dan subjektif, sehingga berpotensi menyebabkan ketidakefisienan dalam strategi pemasaran. Penelitian ini bertujuan untuk menerapkan metode *Random Forest* dalam proses klasifikasi barang promosi dan harga diskon di Alfamidi Turi serta mengimplementasikannya dalam bentuk sistem berbasis web. Data yang digunakan berasal dari laporan penjualan historis sebanyak 500 data produk. Metode *Random Forest* dipilih karena memiliki kemampuan tinggi dalam menangani data kompleks dan menghasilkan akurasi yang stabil. Sistem dirancang menggunakan arsitektur terintegrasi antara Python (Flask) sebagai pemroses model *machine learning*, PHP dan MySQL sebagai *backend*, serta HTML, CSS, dan JavaScript sebagai *frontend*. Berdasarkan hasil pengujian, model *Random Forest* mampu melakukan klasifikasi produk promosi dengan tingkat akurasi sebesar 92%. Hasil ini menunjukkan bahwa metode *Random Forest* efektif digunakan untuk membantu pengambilan keputusan promosi yang lebih objektif, efisien, dan berbasis data historis.

Kata Kunci: *Random Forest*, Klasifikasi, Barang Promosi, Harga Diskon, Sistem Berbasis Web.

CLASSIFICATION OF PROMOTIONAL PRODUCTS AND DISCOUNT PRICES AT ALFAMIDI TURI USING THE RANDOM FOREST METHOD BASED ON A WEB APPLICATION

ABSTRACT

Promotion strategies and discount pricing play an important role in increasing sales and competitiveness in the retail industry. However, determining the right products to promote is often done manually and subjectively, which can lead to inefficiency in marketing strategies. This study aims to apply the *Random Forest* method to classify promotional products and discount prices at Alfamidi Turi and to implement it through a web-based system. The dataset used consists of 500 historical sales records. The *Random Forest* algorithm was chosen for its high accuracy and ability to handle complex data. The system was developed using an integrated architecture consisting of Python (Flask) for machine learning processing, PHP and MySQL for the backend, and HTML, CSS, and JavaScript for the frontend. Based on the testing results, the *Random Forest* model achieved an accuracy of 92% in classifying promotional products. These findings indicate that the *Random Forest* method is effective in supporting more objective, efficient, and data-driven promotional decision-making processes.

Keywords: *Random Forest*, Classification, Promotional Products, Discount Prices, Web-Based System.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN	iv
RIWAYAT HIDUP	v
KATA PENGANTAR	vi
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian	5
BAB II LANDASAN TEORI	6
2.1 Sistem Informasi Berbasis Web.....	6
2.2 Konsep Dasar Klasifikasi	6
2.3 Decision Tree (Pohon Keputusan)	7
2.4 Algoritma Random Forest	8
2.5 HTML.....	12
2.6 PHP.....	12
2.7 MySQL.....	13
2.8 XAMPP	13
2.9 Visual Studio Code.....	13
2.10 Python dan Flask	13
2.11 Confussion Matrix	14
2.12 Integrasi Sistem	15
2.13 Unified Modelling Language (UML)	16
2.14 Penelitian Terdahulu.....	22
BAB III METODE PENELITIAN.....	26
3.1 Jenis dan Pendekatan Penelitian	26
3.2 Objek Penelitian	28
3.3 Sumber Data	30
3.4 Teknik Pengolahan Data.....	30
3.5 Simulasi Perhitungan <i>Random Forest</i>	33
3.5.1. Menentukan <i>Root Node</i> (Akar Pohon).....	35
3.5.2. Membangun Cabang Pohon (<i>Recursive Splitting</i>)	38
3.5.3. Struktur Pohon Keputusan dan Ekstraksi Aturan.....	39
3.5.4. Agregasi dan Prediksi Akhir (<i>Voting</i>).....	40
3.5.5. Evaluasi Model	40

3.6	<i>Flowchart</i> Aplikasi	41
3.7	Desain Sistem	43
3.8	Rancangan Basis Data	54
3.9	Desain <i>User Interface</i>	57
BAB IV HASIL DAN PEMBAHASAN		62
4.1	Perhitungan Manual Random Forest	62
4.2	Tampilan Berbasis Web.	67
4.3	Kelebihan dan Kelemahan Program	73
4.4	Tabel Data Pengujian.	74
BAB V KESIMPULAN DAN SARAN		85
5.1.	Kesimpulan.....	85
5.2.	Saran.....	86
DAFTAR PUSTAKA		87

DAFTAR GAMBAR

Gambar 2. 1 Komponen dalam Class Diagram.....	37
Gambar 3. 1 Flowchart Penelitian.....	42
Gambar 3. 2 Pohon Keputusan Terbentuk	43
Gambar 3. 3 Flowchart Aplikasi	44
Gambar 3. 4 Use Case Diagram	45
Gambar 3. 5 Activity Diagram Login	46
Gambar 3. 6 Activity Diagram Kelola Data Master	47
Gambar 3. 7 Activity Diagram Proses Random Forest.....	48
Gambar 3. 8 Activity Diagram Klasifikasi Produk Promosi.....	48
Gambar 3. 9 Activity Diagram Laporan.....	49
Gambar 3. 10 Activity Diagram Logout	50
Gambar 3. 11 Sequence Diagram Login	50
Gambar 3. 12 Sequence Diagram Kelola Data Master	51
Gambar 3. 13 Sequence Diagram Proses Random Forest	51
Gambar 3. 14 Sequence Diagram Klasifikasi Produk Promosi	52
Gambar 3. 15 Sequence Diagram Laporan	55
Gambar 3. 16 Sequence Diagram Logout	56
Gambar 3. 17 Class Diagram	56
Gambar 3. 18 Desain Halaman Login.....	57
Gambar 3. 19 Desain Halaman Kelola Data Master	57
Gambar 3. 20 Desain Halaman Tambah / Edit Produk.....	58
Gambar 3. 21 Desain Interface Hapus	58
Gambar 3. 22 Desain Halaman Proses Random Forest.....	59
Gambar 3. 23 Desain Halaman Klasifikasi	63
Gambar 3. 24 Desain Halaman Laporan	63
Gambar 3. 25 Desain Interface Logout	64
Gambar 4. 1 Tampilan Halaman Login.....	64
Gambar 4. 2 Tampilan Halaman Dashboard	65
Gambar 4. 3 Tampilan Halaman Data Produk	65
Gambar 4. 4 Tampilan Halaman Tambah Data Produk	66
Gambar 4. 5 Tampilan Halaman Penjualan	66
Gambar 4. 6 Tampilan Halaman Tambah Penjual	67
Gambar 4. 7 Tampilan Awal Proses Random Forest	67
Gambar 4. 8 Tampilan Agregasi Data.....	68
Gambar 4. 9 Bagian Distribusi Data	68
Gambar 4. 10 Hasil Pelatihan Model	72
Gambar 4. 11 Tampilan Awal Riwayat Klasifikasi	73
Gambar 4. 12 Contoh Detail Tampilan Klasifikasi.....	73

DAFTAR TABEL

Tabel 2. 1 Simbol Dalam Use Case	16
Tabel 2. 2 Simbol Activity Diagram	17
Tabel 2. 3 Tabel Simbol Sequence Diagram	18
Tabel 2. 4 Tabel Penelitian Terdahulu.....	22
Tabel 3. 1 Diskritisasi Fitur	31
Tabel 3. 2 Dataset Simulasi	33
Tabel 3. 3 Perhitungan Gini Gain untuk Pemilihan Root Node	36
Tabel 3. 4 Aturan Keputusan (Rules) dari Pohon Keputusan.....	37
Tabel 3. 5 Rancangan Tabel Admin	53
Tabel 3. 6 Rancangan Tabel Produk	53
Tabel 3. 7 Rancangan Tabel Model.....	54
Tabel 3. 8 Rancangan Tabel Klasifikasi	54
Tabel 4. 1 Tabel Representasi Data	64
Tabel 4. 2 Perhitungan Gini Impurity Kategori Penjualan	66
Tabel 4. 3 Perbandingan Gini Gain untuk Pemilihan Root Node.....	66
Tabel 4. 4 Hasil Pengujian Sistem (Percobaan 1)	75
Tabel 4. 5 Hasil Pengujian Sistem (Percobaan 2).....	75
Tabel 4. 6 Hasil Pengujian Sistem (Percobaan 3).....	76
Tabel 4. 7 Hasil Pengujian Sistem (Percobaan 4).....	76
Tabel 4. 8 Hasil Pengujian Sistem (Percobaan 5)	77
Tabel 4. 9 Hasil Pengujian Sistem (Percobaan 6).....	77
Tabel 4. 10 Hasil Pengujian Sistem (Percobaan 7)	78
Tabel 4. 11 Hasil Pengujian Sistem (Percobaan 8).....	78
Tabel 4. 12 Hasil Pengujian Sistem (Percobaan 9).....	79
Tabel 4. 13 Hasil Pengujian Sistem (Percobaan 10)	79

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Strategi promosi dan diskon berperan penting dalam meningkatkan penjualan serta daya saing di industri ritel. Namun, pemilihan produk yang tepat untuk diberikan dan diskon masih menjadi tantangan, terutama di Alfamidi Turi yang cenderung mengandalkan intuisi dan pengalaman subjektif dalam proses penentuan produk. Pendekatan manual ini kerap menimbulkan risiko salah sasaran, sehingga program promosi menjadi kurang efektif dan bisa menyebabkan penumpukan stok maupun kehilangan peluang penjualan (Pratiwi & Nugroho, 2024).

Permasalahan tersebut diperparah oleh naik turunnya permintaan yang tinggi dan keterbatasan analisis data penjualan. Tanpa analisis yang kuat, keputusan mengenai produk promosi dan harga diskon seringkali tidak didasarkan pada data historis penjualan atau tren konsumen. Hal ini menyebabkan terjadinya ketidakseimbangan stok yang akhirnya berdampak pada kenaikan biaya operasional dan menurunnya kepuasan pelanggan (Syahrul Efendi et al., 2024). Studi lain menunjukkan bahwa prediksi yang tidak akurat dalam pengelolaan promosi berisiko menurunkan profitabilitas dan daya saing perusahaan di Tengah kompetisi industri yang semakin ketat (Barus & Darmanto, 2024).

Untuk mengatasi permasalahan ini, dibutuhkan solusi berbasis teknologi yang mampu mengolah data penjualan secara komprehensif. Metode *Random Forest* merupakan salah satu algoritma *machine learning* yang unggul dalam

tugas klasifikasi produk dan penjualan di berbagai sektor ritel (Suci Amaliah et al., 2022). *Random Forest* merupakan metode *ensemble* berbasis pohon keputusan yang tidak hanya mampu meningkatkan akurasi, tetapi juga mampu menangani variabel yang banyak serta pola data yang kompleks (Mahmuda, 2024). Metode ini juga mampu mengidentifikasi faktor-faktor penting yang berpengaruh dalam penentuan produk promosi.

Penelitian Pratiwi & Nugroho (2024) menunjukkan bahwa penggunaan *Random Forest* mampu mengklasifikasikan produk terlaris di swalayan dengan akurasi 94,6%, sehingga membantu pihak toko dalam memfokuskan strategi promosi pada produk-produk yang berpotensi mendongkrak penjualan. Sementara itu, Syahrul Efendi et al. (2024) membuktikan bahwa pemanfaatan *Random Forest* untuk prediksi penjualan dan pengelolaan stok dapat menekan risiko kerugian akibat salah perhitungan persediaan hingga 85% lebih akurat dibandingkan pendekatan konvensional.

Untuk mengimplementasikan model *Random Forest* ini secara praktis dan mudah diakses oleh pihak Alfamidi Turi, maka diperlukan sebuah antarmuka pengguna yang fungsional. Oleh karena itu, penelitian ini akan mengembangkan sebuah sistem berbasis web. Pemilihan platform web didasarkan pada beberapa keunggulan utama: aksesibilitas, di mana sistem dapat diakses dari perangkat komputer manapun di lingkungan toko hanya dengan menggunakan peramban web tanpa perlu instalasi perangkat lunak khusus; sentralisasi, di mana data dan model tersimpan terpusat sehingga menjamin konsistensi; serta kemudahan penggunaan bagi staf yang mungkin tidak memiliki latar belakang teknis yang mendalam. Sistem ini akan dibangun dengan arsitektur modern yang mengintegrasikan

beberapa teknologi web. Bagian *frontend* atau tampilan antarmuka akan menggunakan HTML, CSS, dan JavaScript. Sementara itu, *backend* untuk logika aplikasi utama dan manajemen basis data akan menggunakan PHP dan MySQL , yang berjalan di atas server lokal XAMPP. Adapun proses komputasi inti dari metode *Random Forest* akan ditangani oleh layanan terpisah yang dibangun menggunakan Python dengan *framework* Flask, yang akan berkomunikasi dengan sistem utama.

Berdasarkan temuan-temuan tersebut, penelitian ini merancang sebuah sistem berbasis web yang mengimplementasikan metode *Random Forest* untuk klasifikasi barang promosi dan harga diskon di Alfamidi Turi. Sistem ini akan mengolah data penjualan, harga, dan atribut produk lainnya secara otomatis, sehingga hasil klasifikasi lebih objektif dan berbasis data historis. Diharapkan, sistem ini dapat mengoptimalkan keputusan promosi, mengurangi risiko kerugian akibat salah penentuan barang promosi, serta meningkatkan kepuasan pelanggan dan profitabilitas toko. Khususnya pada Alfamidi Turi, oleh karena itu, penelitian ini berjudul KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST BERBASIS WEB.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana penerapan metode *Random Forest* dalam melakukan klasifikasi barang promosi dan penetapan harga diskon di Alfamidi Turi?

2. Bagaimana perancangan sistem berbasis web yang dapat mengintegrasikan metode Random Forest untuk mendukung proses klasifikasi barang promosi dan harga diskon di Alfamidi Turi?

1.3 Batasan Masalah

Agar penelitian lebih terfokus dan terarah, maka ruang lingkup penelitian dibatasi pada hal-hal berikut :

1. Penelitian ini hanya membahas penerapan metode *Random Forest* untuk klasifikasi barang promosi dan penetapan harga diskon di Alfamidi Turi
2. Sistem yang dikembangkan adalah aplikasi berbasis web yang arsitekturnya mencakup *frontend* (HTML, CSS, JavaScript), *backend* aplikasi (PHP dan MySQL), serta integrasi dengan *service* terpisah (Python dan Flask) untuk komputasi model *Random Forest*
3. Analisis data dilakukan berdasarkan data penjualan historis yang diperoleh dari Alfamidi Turi, dengan jumlah dataset yang digunakan untuk pelatihan dan pengujian model mencakup sekitar 500 data.
4. Evaluasi model akan difokuskan pada metrik akurasi yang dihasilkan dari *Confusion Matrix*, seperti *precision*, *recall*, dan *F1-score*

1.4 Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah sebagai berikut :

1. Untuk menerapkan metode Random Forest dalam melakukan klasifikasi barang promosi dan penetapan harga diskon di Alfamidi Turi.
2. Untuk menganalisis hasil klasifikasi barang promosi dan harga diskon yang diperoleh melalui metode Random Forest berdasarkan data penjualan historis di Alfamidi Turi.

3. Untuk merancang dan membangun sistem berbasis web yang dapat mengintegrasikan metode Random Forest guna mendukung proses klasifikasi barang promosi dan harga diskon di Alfamidi Turi.

1.5 Manfaat Penelitian

Adapun yang menjadi manfaat dalam penelitian ini adalah sebagai berikut

1. Memberikan solusi berbasis web yang dapat membantu Alfamidi Turi dalam menentukan barang promosi dan harga diskon secara lebih objektif dan efisien.
2. Menjadi referensi penerapan metode Random Forest untuk klasifikasi produk di sektor ritel, khususnya dalam pengelolaan promosi dan diskon.
3. Memberikan kontribusi bagi pengembangan aplikasi machine learning berbasis web di bidang penjualan dan pemasaran.

BAB II

LANDASAN TEORI

2.1 Sistem Informasi Berbasis Web

Sistem Informasi dapat didefinisikan sebagai pengintegrasian unsur sistem dan unsur informasi yang merupakan elemen-elemen yang berinteraksi untuk menghasilkan informasi bagi penggunanya (Rahman et al., 2022). Sistem Informasi adalah sebuah sistem terintegrasi yang terdiri dari komponen-komponen yang saling berhubungan untuk mengumpulkan, menyimpan, memproses, dan mendistribusikan informasi guna mendukung pengambilan keputusan, koordinasi, dan kontrol dalam suatu organisasi. Dalam konteks modern, sistem informasi seringkali diimplementasikan dalam bentuk aplikasi berbasis web, yang memungkinkan aksesibilitas yang luas melalui internet.

2.2 Konsep Dasar Klasifikasi

Klasifikasi adalah salah satu tugas paling umum dalam *data mining* dan *machine learning*. Klasifikasi merupakan proses *Supervised Learning*, yang berarti algoritma belajar dari data yang telah memiliki label yang benar (Populix, 2023). Tujuan dari klasifikasi adalah membangun sebuah model (atau fungsi) yang dapat memetakan atribut-atribut dari suatu objek ke salah satu kelas yang telah didefinisikan sebelumnya. Setelah model ini dilatih menggunakan data historis, model tersebut dapat digunakan untuk memprediksi kelas dari data baru yang labelnya belum diketahui (Kompasiana, 2022).

Proses pengembangan model klasifikasi umumnya mengikuti alur kerja yang sistematis, yang terdiri dari beberapa tahapan utama:

1. **Persiapan Data:** Tahap ini melibatkan pengumpulan, pembersihan, dan pra-pemrosesan data untuk memastikan data siap digunakan. Ini termasuk menangani nilai yang hilang dan mengubah data kategorikal menjadi format numerik jika diperlukan.
2. **Pembagian Data:** Kumpulan data dibagi menjadi dua bagian: data latih (*training set*) dan data uji (*testing set*). Umumnya, proporsi yang digunakan adalah 80% untuk data latih dan 20% untuk data uji, seperti yang diterapkan dalam beberapa penelitian relevan.[1, 1]
3. **Pelatihan Model (*Model Training*):** Model klasifikasi (dalam hal ini, Random Forest) dibangun dengan menggunakan data latih. Selama tahap ini, algoritma "belajar" pola dan hubungan antara fitur-fitur input dan label kelas yang sesuai.
4. **Evaluasi Model (*Model Evaluation*):** Kinerja model yang telah dilatih diuji menggunakan data uji, yaitu data yang belum pernah "dilihat" oleh model sebelumnya. Tahap ini penting untuk menilai seberapa baik model dapat menggeneralisasi pengetahuannya ke data baru.
5. **Penerapan (*Deployment*):** Setelah model divalidasi dan dianggap memiliki kinerja yang memuaskan, model tersebut siap untuk digunakan dalam aplikasi nyata untuk mengklasifikasikan data baru.

2.3 Decision Tree (Pohon Keputusan)

Decision Tree merupakan teknik yang digunakan untuk melakukan prediksi atau klasifikasi dengan cara membagi kumpulan data besar menjadi kelompok-kelompok yang lebih kecil melalui penerapan serangkaian aturan Keputusan. (Hafizan & Putri, 2020).

Komponen-komponen dalam pohon keputusan pada umumnya adalah sebagai berikut :

1. Simpul (*Nodes*)

Dalam pohon keputusan terdapat tiga jenis simpul :

a. Simpul Akar

Simpul ini mewakili pilihan yang akan membagi semua data menjadi dua atau lebih kelompok

b. Simpul Internal

Simpul ini adalah pilihan yang tersedia pada suatu titik dalam pohon dan terhubung dengan simpul induk dan simpul anak

c. Simpul Daun

Simpul ini mewakili hasil akhir dari kombinasi keputusan atau peristiwa

2. Cabang (*Branches*)

Cabang menghubungkan satu simpul pohon ke simpul pohon lainnya dalam ruang lingkup pohon keputusan, dan cabang akan merepresentasikan aliran keputusan dari satu simpul pohon ke dalam satu simpul lainnya

Dengan kata lain, pohon keputusan bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan-aturan klasifikasi tertentu, salah satu algoritma yang menerapkan pohon keputusan adalah algoritma C.45 (Azwanti, 2019).

2.4 Algoritma Random Forest

Teknik *Machine Learning* yang disebut *Random Forest* akan menggabungkan data dari beberapa banyak pohon Keputusan untuk mendapatkan hasil yang lebih akurat. Dibandingkan dengan pendekatan *Classification and*

Regression Tree (CART), algoritma ini memiliki beberapa keunggulan, dikarenakan *Random Forest* membuat beberapa pohon keputusan berdasarkan data dan atribut yang dipilih secara acak, yang disebut sebagai “hutan” karena banyaknya pohon Keputusan. Langkah berikutnya, semua pohon Keputusan yang dibuat digunakan untuk menguji data di *Random Forest*, dan hasilnya ditentukan dengan melihat sebagian besar respon yang diberikan oleh pohon *Random Forest* (Handayani & Charis Fauzan, 2024).

Proses pembentukan pohon Keputusan di *Random Forest* meliputi penentuan atribut akar yang didasarkan pada nilai *information gain* yang biasanya dihitung dari *Entropy*. Proses tersebut diulangi pada setiap atribut sampai seluruh bagian *instance* mempunyai kelas yang sama. Adapun untuk perhitungan entropi dapat digunakan rumus dibawah ini.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots (2.1)$$

Dimana :

S : Kumpulan Kasus

N : Jumlah partisi S

Pi : proporsi dari Si terhadap S

Persamaan $Entropy(S)$ di atas digunakan untuk mengukur tingkat ketidakpastian atau keragaman dalam suatu kumpulan kasus (S). Entropy menentukan seberapa murni atau tidaknya suatu data berdasarkan distribusi label kelas yang ada. Berikut penjelasan lanjutan dari rumus tersebut:

1. S merujuk pada kumpulan data atau fenomena yang sedang diamati.

2. N adalah jumlah partisi atau kategori yang terdapat dalam kumpulan kasus S. Misalnya, jika ada dua kelas dalam data (seperti "positif" dan "negatif"), maka N bernilai 2.
3. P_i adalah proporsi kasus yang termasuk dalam partisi S_i terhadap keseluruhan kasus dalam S. Proporsi ini mengukur berapa banyak data yang berada dalam setiap kategori dibandingkan dengan total data.

Sedangkan perhitungan *Information Gain* dapat menggunakan rumus dibawah ini :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \dots \dots \dots (2.1)$$

Dimana :

S : Himpunan kasus

A : Atribut

n : jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S

Rumus entropy digunakan untuk menentukan pemilihan atribut dalam pohon keputusan. Himpunan kasus (S) merujuk pada total entitas atau data yang dianalisis dalam sebuah konteks. Atribut (A) adalah karakteristik atau variabel yang diukur atau diamati pada setiap kasus dalam S. Jumlah partisi atribut A (n) menunjukkan pembagian yang dibuat berdasarkan nilai-nilai berbeda dari atribut tersebut. $|S_i|$ menunjukkan jumlah kasus yang termasuk dalam partisi ke-i dari atribut A, sedangkan $|S|$ mewakili jumlah total kasus dalam himpunan S secara keseluruhan.

Random Forest dibangun di atas teknik yang disebut *Bootstrap Aggregating* atau *Bagging*. Bagging adalah metode ansambel yang dirancang khusus untuk meningkatkan stabilitas dan akurasi model *machine learning* dengan cara

mengurangi varians (IBM, 2024). Proses kerja Bagging adalah sebagai berikut (Boehmke & Greenwell, 2019):

1. **Bootstrap Sampling:** Dari dataset latih asli yang berukuran N , dibuat sejumlah B dataset baru (disebut sampel bootstrap), yang masing-masing juga berukuran N . Sampel-sampel ini dibuat dengan cara pengambilan sampel acak **dengan pengembalian** (*sampling with replacement*). Artinya, satu data observasi dapat terpilih lebih dari sekali dalam satu sampel bootstrap, sementara observasi lainnya mungkin tidak terpilih sama sekali.
2. **Out-of-Bag (OOB) Samples:** Data observasi yang tidak terpilih dalam suatu sampel bootstrap disebut sebagai sampel *Out-of-Bag* (OOB) untuk pohon yang dilatih pada sampel tersebut. Sampel OOB ini sangat berguna untuk validasi model internal.
3. **Pelatihan Model Paralel:** Sebuah model (dalam hal ini, pohon keputusan) dilatih secara independen pada setiap sampel bootstrap yang telah dibuat. Hal ini menghasilkan B model yang berbeda.
4. **Agregasi Prediksi:** Untuk melakukan prediksi pada data baru, data tersebut dimasukkan ke setiap dari B pohon. Hasil prediksi dari semua pohon kemudian diagregasikan. Untuk tugas **klasifikasi**, agregasi dilakukan melalui **pemungutan suara mayoritas** (*majority voting*), di mana kelas yang paling banyak dipilih oleh pohon-pohon menjadi prediksi akhir. Untuk tugas regresi, hasil prediksi dirata-ratakan.

Fungsi utama dari Bagging adalah untuk mengurangi varians tinggi yang sering ditemukan pada model seperti pohon keputusan. Dengan merata-ratakan prediksi dari banyak pohon yang dilatih pada subset data yang sedikit berbeda,

"kebisingan" dan kesalahan individual dari setiap pohon cenderung saling meniadakan, menghasilkan prediksi akhir yang lebih stabil dan andal.

2.5 HTML

HTML atau *Hyper Text Markup Language* adalah Bahasa markup standar yang digunakan untuk membuat halaman web. HTML berfungsi untuk membuat struktur dari sebuah website. HTML digunakan untuk menandai bagian mana yang akan menjadi judul artikel, bagian mana yang berfungsi sebagai isi artikel, atau bagian mana yang disajikan dalam bentuk tabel. HTML memungkinkan pengguna untuk membuat halaman web dengan struktur yang terorganisir dan mudah dimengerti oleh browser. Dengan menggunakan tag-tag HTML yang sesuai, pengguna dapat menentukan judul, paragraf, daftar, gambar, dan tautan dalam halaman web mereka. Selain itu, HTML juga memungkinkan pengguna untuk menambahkan elemen-elemen interaktif seperti formulir dan tombol. (Pratama, 2020).

2.6 PHP

PHP (*Hypertext PreProcessor*) merupakan bahasa pemrograman yang diproses di *server*, fungsi utama PHP dalam membangun website adalah untuk menerjemahkan baris kode program menjadi kode mesin yang dapat dimengerti oleh komputer yang bersifat *server-side* yang dapat ditambahkan ke dalam HTML. Selain itu, PHP juga berfungsi untuk pengelolaan data dalam database (Tumini & Fitria, 2021).

2.7 MySQL

MySQL adalah sebuah *database* atau media penyimpanan data yang mendukung *script* PHP, MySQL juga mempunyai *query* atau bahasa SQL (*Structured Query Language*) yang simple dan menggunakan *escape character* yang sama dengan PHP (Tumini & Fitria, 2021).

2.8 XAMPP

XAMPP adalah perangkat lunak open source yang mendukung berbagai sistem operasi dan merupakan kumpulan dari beberapa program. Fungsinya sebagai server mandiri (localhost) yang terdiri dari beberapa program, termasuk Apache HTTP Server, database MySQL, serta penerjemah bahasa untuk PHP dan Perl.(Hartiwati, 2022)

2.9 Visual Studio Code

Visual Studio Code (VS Code) adalah editor kode sumber open-source yang dikembangkan oleh Microsoft. Aplikasi ini mendukung berbagai bahasa pemrograman dan dilengkapi dengan fitur-fitur canggih seperti debugging, kontrol versi Git, penyorotan sintaks, IntelliSense (*autocompletion*), dan alat linting (Jennie et al., 2021) .

2.10 Python dan Flask

Python adalah bahasa pemrograman tingkat tinggi yang dikenal karena sintaksisnya yang jelas dan mudah dipelajari. Python banyak digunakan dalam pengembangan web, analisis data, kecerdasan buatan, dan lainnya.

Flask adalah kerangka kerja web mikro untuk Python yang memungkinkan pengembang untuk membuat aplikasi web dengan cepat dan sederhana. Flask menyediakan alat dan pustaka yang diperlukan untuk membangun aplikasi web,

termasuk dukungan untuk routing, template, dan sesi. Menurut Walecha (2021), Flask menawarkan fleksibilitas tinggi dan cocok untuk proyek kecil hingga menengah (Walecha, 2021).

2.11 *Confussion Matrix*

Confusion Matrix adalah tabel yang menyatakan klasifikasi jumlah uji yang benar dan jumlah uji yang salah (Normawati & Prayogi, 2021). *Confusion Matrix* sendiri memiliki empat istilah :

1. *True Negative* (TN) yang berarti model klasifikasi memprediksi data ada di kelas *negative* namun yang sebenarnya data memang ada di kelas *negative*.
2. *True Positive* (TP) yang berarti model klasifikasi memprediksi data ada di kelas positif namun sebenarnya memang data berada di kelas positif.
3. *False Negative* (FN) yang berarti model klasifikasi memprediksi data ada di kelas *negative* namun yang sebenarnya data ada di kelas positif.
4. *False Positive* (FP) yang berarti model klasifikasi memprediksi data ada di kelas positif namun yang sebenarnya data ada di kelas *negative*.

Dari definisi *confusion matrix*, beberapa poin poin dalam *confusion matrix* digunakan untuk menghitung *precision*, *recall*, dan *f1 score*. *Precision* adalah perbandingan antara *True Positive* (TP) dengan banyaknya data yang diprediksi positif, secara matematis dapat dilihat dibawah ini

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2.2)$$

Untuk *recall* sendiri adalah perbandingan antara *true positive* (TP) dengan banyaknya data yang sebenarnya positif. Dapat dinyatakan secara matematis seperti dibawah ini

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2.3)$$

Sedangkan *F1 Score* adalah nilai tengah dari *precision* dan *recall*. Nilai terbaik dari *F1 Score* adalah 1 dan nilai terburuknya adalah 0, secara matematis dapat dituliskan seperti dibawah ini

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right) \dots\dots\dots (2.4)$$

Nilai *F1 Score* yang baik menandakan bahwa model klasifikasi kita punya *precision* dan *recall* yang baik (S. Setiawan, 2020).

2.12 Integrasi Sistem

Dalam implementasinya, sistem pendukung keputusan berbasis web akan mengintegrasikan:

1. Frontend: Dibangun dengan HTML, CSS, dan JavaScript untuk antarmuka pengguna yang interaktif.
2. Backend: Menggunakan PHP untuk logika aplikasi dan MySQL sebagai basis data.
3. Komponen Python: Bertanggung jawab atas pemrosesan data dan perhitungan metode *Random Forest*, yang diakses melalui API yang dibangun menggunakan Flask.
4. Lingkungan Pengembangan: Visual Studio Code sebagai editor kode dan XAMPP sebagai server lokal untuk testing aplikasi.

Integrasi ini memastikan bahwa data mentah dari basis data dapat diproses secara efisien menggunakan Python, sedangkan hasilnya dikirim kembali ke antarmuka berbasis web untuk disajikan kepada pengguna secara real-time.

2.13 *Unified Modelling Language (UML)*

Unified Modelling Language (UML) adalah bahasa pemodelan visual yang digunakan untuk menentukan, memvisualisasikan, membangun berbagai aspek dari sebuah sistem perangkat lunak. UML sendiri memiliki fungsi sebagai alat untuk menangkap pemahaman mengenai sistem yang perlu dibangun.

Bayangkan sistem sebagai kelompok objek yang berbeda yang bekerja sama untuk melakukan pekerjaan yang berguna untuk pengguna. Bagian struktur statis mendefinisikan jenis objek yang penting untuk sistem dan bagaimana mereka terkait satu sama lain. Bagian perilaku dinamis menggambarkan bagaimana objek berubah seiring waktu dan berkomunikasi satu sama lain untuk mencapai tujuan tertentu. Dengan memodelkan sistem dari berbagai sudut pandang yang terkait, kita dapat memahami sistem tersebut untuk berbagai keperluan (Rumbaugh et al., 2021).

Alat bantu yang digunakan dalam perancangan sistem menggunakan Unified Modelling Language (UML) adalah sebagai berikut :

1. *Use Case Diagram*

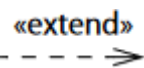
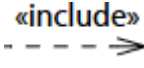
Use Case Diagram menunjukkan bagaimana perilaku suatu sistem ketika digunakan oleh orang lain, *Use Case Diagram* memecah fungsionalitas sistem menjadi tindakan yang bermakna bagi pengguna (*Actor*) yang menggunakan sistem. Pengguna (*Actor*) mencakup manusia, serta sistem dan proses komputer lainnya).

Sedangkan *Actor* adalah representasi ideal dari orang, proses, atau objek eksternal yang berinteraksi dengan suatu sistem, subsistem, atau kelas. Untuk identifikasi *Actor*, harus ditentukan tugas tugas yang berkaitan

dengan peran pada konteks sistem. Orang atau sistem bisa muncul dalam beberapa peran, perlu diketahui jika *Actor* berinteraksi dengan *use case*.

Selain Actor, terdapat simbol simbol lain yang bisa digunakan didalam *Use Case* untuk memodelkan fungsi apa saja dari sistem yang dibangun beserta hubungan antar fungsi nya. Adapun simbol-simbol dalam *Use Case* antara lain sebagai berikut :

Tabel 2. 1 Simbol Dalam Use Case

Gambar	Function	Keterangan
	Asosiasi	Asosiasi antara <i>actor</i> dan <i>use case</i> , digambarkan dengan garis tanpa panah yang menunjukkan siapa atau apa yang meminta interaksi secara langsung
	Extend	Perluasan dari <i>use case</i> lain ketika kondisi atau syarat terpenuhi
	Include	Merujuk pada penambahan perilaku tambahan dalam suatu <i>use case</i> yang secara eksplisit menjelaskan penambahan tersebut
	Use case generalization	Hubungan antara suatu <i>use case</i> umum dengan <i>use case</i> yang lebih spesifik yang mewarisi

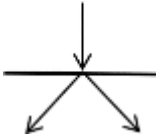
		dan menambahkan fitur fitur kepadanya
--	--	---------------------------------------

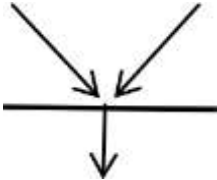
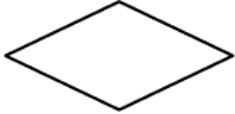
(Sumber : Rumbaugh et al., 2021).

2. Diagram Aktivitas (*Activity Diagram*)

Diagram aktivitas (*Activity Diagram*) menggambarkan aliran kerja dari sebuah sistem. Diagram aktivitas dapat mencakup cabang dan bercabangnya control dalam sebuah sistem yang berjalan secara bersamaan. Cabang cabang ini mewakili aktivitas yang dapat dilakukan secara bersamaan (Rumbaugh et al., 2021). Simbol-simbol yang digunakan dalam *Activity Diagram* adalah sebagai berikut :

Tabel 2. 2 Simbol Activity Diagram

Gambar	Function	Keterangan
	<i>Start Point</i>	Diletakkan pada pojok kiri atas dan merupakan awal aktifitas
	<i>End Point</i>	Akhir dari aktifitas
	<i>Activities</i>	Menggambarkan suatu proses / kegiatan dalam aplikasi
	<i>Fork</i>	Digunakan untuk menunjukkan kegiatan yang dilakukan secara parallel untuk menggabungkan

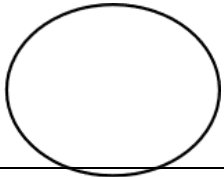
		dua kegiatan parallel menjadi satu
	<i>Join</i>	Digunakan untuk menunjukkan adanya dekomposisi
	<i>Decisions Point</i>	Menggambarkan pilihan untuk pengambilan keputusan, <i>true</i> , <i>false</i>

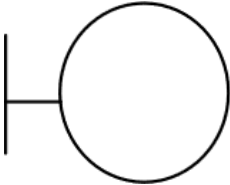
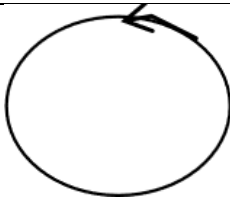
(Sumber : Rumbaugh et al., 2021).

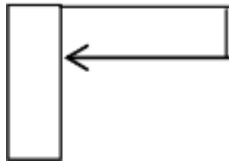
3. *Sequence Diagram*

Sequence Diagram menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objek dan pesan yang dikirimkan dan diterima antara objek. Simbol-simbol yang digunakan dalam *sequence diagram* yaitu :

Tabel 2. 3 Tabel Simbol *Sequence Diagram*

Gambar	Function	Keterangan
	<i>Entity Class</i>	Merupakan bagian dari system yang berisi kumpulan kelas berupa entitas-entitas

		yang membentuk gambaran awal sistem dan menjadi landasan untuk menyusun basis data
	<i>Boundary Class</i>	Berisi kumpulan kelas yang menjadi <i>interface</i> atau interaksi antara satu atau lebih <i>actor</i> dengan system, seperti tampilan form entry dan <i>form</i> cetak
	<i>Control Class</i>	Suatu objek yang berisi logika aplikasi yang tidak memiliki tanggung jawab kepada entitas, contohnya adalah kalkulasi dan aturan bisnis yang melibatkan berbagai objek
	<i>Message</i>	Simbol mengirim pesan antar <i>class</i>

	<i>Recursive</i>	Menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri
	<i>Activation</i>	<i>Activation</i> mewakili sebuah eksekusi operasi dari objek, Panjang kotak ini berbanding lurus dengan durasi aktivitas sebuah operasi
	<i>Lifeline</i>	garis titik-titik yang terhubung dengan objek, sepanjang <i>lifeline</i> terdapat <i>activation</i>

(Sumber : Rumbaugh et al., 2021).

4. *Class Diagram* (Diagram Kelas)

Class diagram adalah salah satu jenis diagram struktur pada UML yang menggambarkan dengan jelas struktur serta deksripsi *class*, atribut, metode, dan hubungan dari setiap objek. *Class Diagram* bersifat statis, dalam artian diagram ini tidak menjelaskan apa yang terjadi jika kelas-kelasnya berhubungan, melainkan menjelaskan hubungan apa yang terjadi.

Class Diagram memiliki tiga komponen penyusun. Dapat dilihat dalam gambar dibawah ini



Gambar 2. 1 Komponen dalam Class Diagram

(Sumber : R. Setiawan, 2021)

Berikut merupakan penjelasan komponen-komponen diatas :

1. Komponen Atas

Komponen ini berisikan nama *class*. Setiap class memiliki nama yang berbeda-beda

2. Komponen Tengah

Komponen ini berisikan atribut dari *class*, komponen ini digunakan untuk menjelaskan kualitas dari suatu kelas. Atribut ini dapat ditulis dengan detail, dengan cara memasukkan tipe nilai

3. Komponen Bawah

Komponen ini menyertakan operasi yang ditampilkan dalam bentuk daftar. Operasi ini dapat menggambarkan bagaimana suatu *class* dapat berinteraksi dengan data.

2.14 Penelitian Terdahulu

Pemilihan metode *Random Forest* untuk klasifikasi barang promosi didasarkan pada rekam jejak keberhasilannya yang telah terbukti dalam berbagai penelitian di domain yang relevan. Tinjauan literatur menunjukkan bahwa *Random Forest* secara konsisten memberikan kinerja yang unggul dalam tugas klasifikasi dan prediksi yang melibatkan data bisnis dan operasional. Akurasi tinggi yang

dilaporkan di berbagai studi memberikan justifikasi yang kuat untuk penerapannya dalam penelitian ini.

Tabel 2. 4 Tabel Penelitian Terdahulu

No	Judul Penelitian & Penulis (Tahun)	Konteks/Objek Penelitian	Temuan Kunci & Akurasi Model	Relevansi dengan Penelitian Ini
1	<i>Implementasi Metode Random Forest untuk Klasifikasi Penjualan Produk Sabun Paling Laris (Pratiwi & Nugroho, 2024)</i>	Klasifikasi produk sabun di swalayan ke dalam 4 kategori (tidak laris, kurang laris, laris, paling laris) berdasarkan data penjualan.	Akurasi model 94,6%. Faktor penting: jumlah terjual, harga, merek, jenis. Membantu optimasi inventaris dan strategi pemasaran.	Sangat relevan. Menggunakan RF untuk klasifikasi produk ritel berdasarkan tingkat penjualan, sejalan dengan tujuan penelitian ini. Menjadi acuan dalam metrik evaluasi dan arah bisnis.
2	<i>Penerapan Algoritma Random Forest untuk Prediksi Penjualan dan Sistem Persediaan Produk (Syahrul Efendi et al., 2024)</i>	Prediksi penjualan dan optimasi persediaan produk Bolen Crispy untuk mengatasi fluktuasi permintaan dan risiko over/understock.	Akurasi prediksi meningkat hingga 85% dibanding metode konvensional, menghasilkan manajemen stok yang lebih efisien.	Relevan. Menunjukkan manfaat RF dalam prediksi penjualan yang berdampak langsung terhadap efisiensi pengelolaan persediaan.
3	<i>Implementasi Metode Random Forest untuk Memprediksi Penjualan Produk (Barus & Darmanto, 2024)</i>	Prediksi penjualan pelumas otomotif untuk mengatasi fluktuasi permintaan dan ketidakpastian stok.	R-Squared sebesar 0,99. Memberikan dasar kuat untuk strategi penjualan dan manajemen inventaris.	Relevan. Memperkuat keunggulan RF dalam menangani data penjualan kompleks dan fluktuatif.
4	<i>Penerapan Metode Klasifikasi Random Forest</i>	Klasifikasi nilai ujian nasional berdasarkan skor akreditasi serta	Akurasi 83,49%. Variabel penting:	Relevan secara metodologis. Menunjukkan kemampuan RF

	<i>dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan (Ramadhan et al., 2019)</i>	identifikasi variabel penentu mutu pendidikan.	kondisi laboratorium, kualifikasi guru (berdasarkan Mean Decrease Gini).	dalam analisis <i>feature importance</i> , yang juga digunakan dalam penelitian ini.
5	<i>Penerapan Algoritma Klasifikasi Random Forest untuk Penentuan Kelayakan Pemberian Kredit (Zailani & Hanun, 2020)</i>	Klasifikasi kelayakan kredit nasabah koperasi untuk memprediksi risiko kredit macet.	Akurasi 87,88%, AUC 0,907. Menunjukkan performa klasifikasi yang baik.	Relevan. Menjadi contoh penggunaan RF dalam klasifikasi risiko bisnis, yang dapat dianalogikan dengan kesalahan penentuan produk promosi.
6	<i>Implementasi Metode Random Forest pada Kategori Konten Kanal YouTube (Mahmuda, 2024)</i>	Klasifikasi konten kanal YouTube berdasarkan jumlah subscriber, video, dan penayangan.	Akurasi 77,27%. Variabel terpenting: jumlah subscriber (19,04%).	Relevan secara metodologis. Contoh lain penggunaan RF untuk klasifikasi dan analisis variabel penting.
7	<i>Penerapan Metode Random Forest untuk Klasifikasi Varian Minuman Kopi (Suci Amaliah et al., 2022)</i>	Klasifikasi varian minuman kopi paling diminati pelanggan di kedai kopi.	Akurasi 94,12%. Varian <i>coffee-based</i> lebih diminati dibanding <i>signature coffee</i> .	Relevan. Studi langsung tentang preferensi konsumen di sektor F&B, selaras dengan klasifikasi produk untuk promosi.
8	<i>Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB (Aini et al., 2024)</i>	Klasifikasi bidang studi independen (MSIB) berdasarkan nilai mata kuliah mahasiswa.	Akurasi 80%, presisi 80%, dan recall 82%.	Relevan sebagai studi klasifikasi multi-kelas dengan RF, menunjukkan fleksibilitas algoritma.

9	<i>Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum</i> (Abdi et al., 2024)	Klasifikasi kualitas air minum menggunakan dataset dari Kaggle.	Akurasi 69%. Analisis <i>feature importance</i> : pH, sulfat, dll.	Relevan secara metodologis. Menunjukkan proses lengkap mulai dari penerapan RF hingga evaluasi dan analisis fitur.

;

BAB III

METODE PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

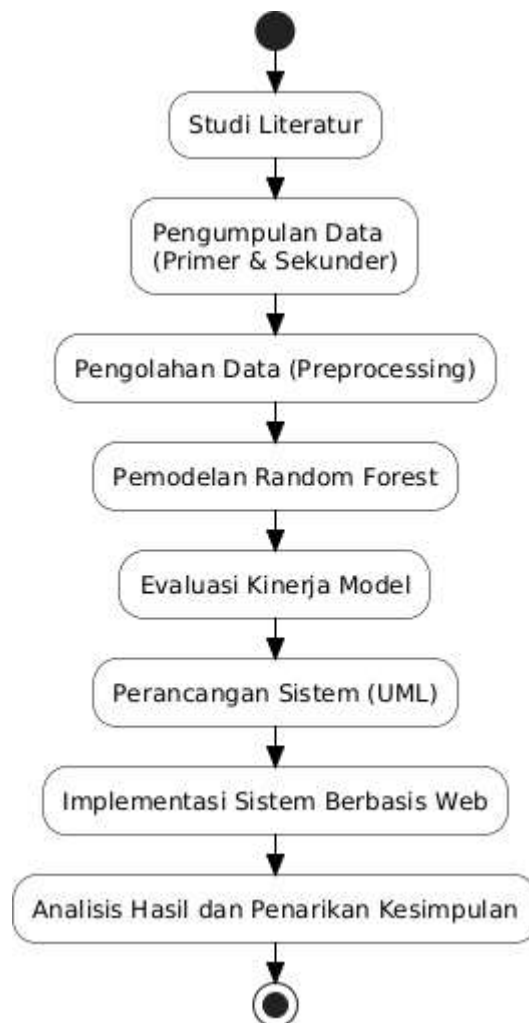
Penelitian ini menggunakan pendekatan kuantitatif dengan fokus pada pemodelan klasifikasi. Pendekatan ini dipilih karena penelitian ini secara fundamental bergantung pada analisis data historis numerik, seperti data penjualan, margin keuntungan dan jumlah stok untuk membangun sebuah model matematis yang dapat mengklasifikasikan produk.

Metodologi spesifik yang diimplementasikan adalah klasifikasi *supervised machine learning* (pembelajaran mesin terarah). Pendekatan ini bertujuan untuk membangun sebuah model yang dapat "belajar" pola dari data historis. Model yang telah terlatih ini kemudian dapat digunakan untuk mengklasifikasikan produk baru yang belum pernah dilihat sebelumnya ke dalam kategori yang telah ditentukan, yaitu "Layak Promosi" atau "Tidak Layak Promosi". Dengan demikian, penelitian ini tidak hanya menyelesaikan masalah pemilihan produk saat ini, tetapi juga menyediakan alat bantu keputusan yang dapat digunakan secara berkelanjutan.

Algoritma yang dipilih untuk tugas klasifikasi ini adalah *Random Forest*. *Random Forest* merupakan metode *ensemble learning* yang menggabungkan prediksi dari banyak pohon keputusan (*decision trees*) untuk menghasilkan prediksi akhir yang lebih akurat dan stabil. Metode ini dikenal memiliki keunggulan dalam hal akurasi yang tinggi, kemampuan menangani dataset yang kompleks dengan banyak variabel, serta ketahanannya terhadap *overfitting*,

menjadikannya sangat relevan untuk diterapkan pada data ritel yang dinamis dan beragam.

Tahapan-tahapan yang dilalui dalam penelitian ini dapat dilihat pada gambar dibawah ini.



Gambar 3. 1 *Flowchart* Penelitian

Secara ringkas, penelitian ini akan dilaksanakan melalui beberapa tahapan utama yang sistematis, yaitu:

1. Studi Literatur: Mengkaji landasan teori mengenai klasifikasi, algoritma Random Forest, dan pengembangan sistem berbasis web, serta meninjau penelitian terdahulu yang relevan.

2. Pengumpulan Data: Mengumpulkan data primer berupa laporan penjualan produk dari Alfamidi Turi dan data sekunder dari jurnal ilmiah.
3. Pengolahan Data: Melakukan persiapan data (*data preprocessing*) yang mencakup penggabungan dataset, pelabelan kelas, seleksi fitur, dan pembagian dataset menjadi data latih dan data uji.
4. Pemodelan dan Evaluasi: Membangun model klasifikasi menggunakan algoritma Random Forest pada data latih, kemudian mengevaluasi kinerjanya menggunakan data uji dengan metrik dari *Confusion Matrix*.
5. Perancangan dan Implementasi Sistem: Merancang arsitektur sistem menggunakan UML (*Use Case, Activity, Sequence, Class Diagram*) dan membangun aplikasi berbasis web yang mengintegrasikan model yang telah dibuat.
6. Analisis dan Penarikan Kesimpulan: Menganalisis hasil klasifikasi yang didapatkan dari sistem untuk ditarik kesimpulan dan saran.

3.2 Objek Penelitian

Objek dalam penelitian ini adalah produk-produk yang dijual di Alfamidi Turi, yang dianalisis sebagai kandidat untuk program promosi. Dalam konteks *machine learning*, objek-objek ini diperlakukan sebagai alternatif data, di mana setiap produk memiliki serangkaian atribut yang akan menjadi dasar bagi model untuk melakukan klasifikasi.

1. Alternatif (*Instances*) : Setiap produk individual yang tercatat dalam laporan penjualan dan performa Alfamidi Turi, seperti yang terdapat dalam dokumen

"Laporan 500 Item Paling Laris" dan "Laporan 500 Item Performance Terendah".

2. Kriteria (*Features*) : Atribut atau karakteristik kuantitatif dari setiap produk yang digunakan sebagai variabel prediktor (input) untuk model *Random Forest*.

Berdasarkan data yang tersedia, fitur-fitur utama yang akan digunakan adalah :

- a. Rata-rata penjualan bulanan (*average sales qty*) : Indikator utama popularitas produk dan permintaan dari konsumen.
- b. Margin Keuntungan (%Grs Mrg) : Ukuran Profitabilitas produk bagi toko
- c. Jumlah Stok : Indikator tekanan inventaris, stok yang tinggi mungkin memerlukan promosi untuk mempercepat perputaran barang.
- d. Umur Stok (Hari): Variabel potensial yang mengindikasikan seberapa cepat suatu produk terjual dari rak.

.

3. Kelas (*Target Variable*): Variabel target atau label yang ingin diprediksi oleh model. Ini adalah variabel kategorikal yang akan didefinisikan selama tahap pengolahan data. Kelasnya adalah :

- a. **Layak Promosi:** Produk yang berdasarkan data historis menunjukkan karakteristik yang ideal untuk dipromosikan (misalnya, penjualan tinggi, potensi keuntungan baik).
- b. **Tidak Layak Promosi:** Produk yang tidak menunjukkan karakteristik tersebut (misalnya, penjualan lambat, stok rendah).

3.3 Sumber Data

Data yang digunakan dalam penelitian ini berasal dari dua sumber utama, yaitu data primer dan data sekunder, yang keduanya krusial untuk membangun dan memvalidasi model klasifikasi.

1. Data Primer : Data primer merupakan data mentah yang menjadi fondasi utama dalam pelatihan model *machine learning*. Data ini diperoleh secara langsung dari Alfamidi Turi dalam bentuk laporan operasional internal. Secara spesifik, data primer yang digunakan adalah
 - a. Laporan 500 item paling laris : Laporan ini berisi daftar produk dengan volume penjualan tertinggi selama periode tertentu.
 - b. Laporan 500 item *Performance* terendah : Laporan ini berisi daftar produk dengan performa penjualan terendah atau perputaran stok yang lambat.
2. Data Sekunder : Data sekunder digunakan untuk membangun landasan teoretis yang kuat, terutama terkait metodologi *Random Forest*. Sumber ini mencakup tinjauan literatur dari berbagai publikasi ilmiah, seperti jurnal, artikel teknis, dan dokumentasi algoritma. Data sekunder ini memberikan justifikasi pemilihan metode, penjelasan matematis, metrik evaluasi, dan praktik terbaik dalam implementasi model klasifikasi.

3.4 Teknik Pengolahan Data

Tahap pengolahan data (atau *data preprocessing*) adalah langkah krusial dalam alur kerja *machine learning* untuk mengubah data mentah menjadi format yang bersih, terstruktur, dan siap digunakan untuk melatih model *Random Forest*. Proses ini memastikan kualitas dan relevansi data yang akan menjadi input bagi algoritma. Berikut adalah tahapan-tahapan yang dilakukan:

1. Pengumpulan dan Penggabungan Data : Langkah awal adalah mengonsolidasikan data dari dua sumber laporan primer. Data dari "Laporan 500 Item Paling Laris" dan "Laporan 500 Item Performance Terendah" digabungkan menjadi satu dataset tunggal. Proses ini menciptakan kumpulan data yang beragam, mencakup spektrum produk dari yang berkinerja sangat baik hingga yang sangat buruk.
2. Defenisi Variabel Target (*Labelling*) : Karena *Random Forest* adalah algoritma *supervised learning*, setiap baris data (setiap produk) harus memiliki label kelas yang telah ditentukan. Proses pelabelan ini dilakukan berdasarkan sumber laporan dari mana data tersebut berasal, yang secara inheren mencerminkan kelayakan promosi dari perspektif bisnis
 - a. Semua produk yang berasal dari "Laporan 500 Item Paling Laris" akan diberi label kelas 1, yang merepresentasikan Layak Promosi.
 - b. Sebuah produk yang berasal dari "Laporan 500 item *performance* terendah" akan diberi label kelas 0, yang merepresentasikan tidak layak promosi.
3. Seleksi Fitur (*Feature Selection*) : Tidak semua kolom dalam laporan relevan untuk prediksi. Berdasarkan analisis awal dan relevansi bisnis, fitur-fitur berikut dipilih dari dataset gabungan untuk menjadi variabel prediktor.
 - a. Avg Sales Qty(bh) (Rata-rata Penjualan)
 - b. %Grs Mrg (Margin Keuntungan)
 - c. Stock (QOH) (Jumlah Stok)

Fitur-fitur ini dipilih karena secara logis memiliki pengaruh kuat terhadap keputusan promosi: penjualan mencerminkan permintaan, margin mencerminkan profitabilitas, dan stok mencerminkan urgensi.

4. Pembersihan Data (*Data Cleaning*) : Dataset yang telah diseleksi kemudian diperiksa untuk memastikan kualitasnya. Ini melibatkan pengecekan nilai-nilai yang hilang (*missing values*) atau anomali data. Meskipun data laporan cenderung bersih, langkah ini tetap merupakan praktik standar untuk memastikan tidak ada data korup yang masuk ke dalam model.
5. Diskritisasi Fitur (Untuk Simulasi Manual): Khusus untuk tujuan demonstrasi pada perhitungan manual di sub-bab 3.5, fitur-fitur yang bersifat kontinu (seperti penjualan dan margin) akan diubah menjadi kategori diskrit. Langkah ini penting untuk menyederhanakan perhitungan matematis *Gini Impurity* dan membuat logika pohon keputusan lebih mudah dipahami secara intuitif. Aturan diskritisasi yang diusulkan adalah sebagai berikut

Tabel 3. 1 Diskritisasi Fitur

Fitur	Nilai Asli	Kategori Diskrit
Penjualan	> 10	Tinggi
	1 – 10	Sedang
	< 1	Rendah
Margin	> 30%	Tinggi
	15% – 30%	Sedang
	< 15%	Rendah
Stok	> 100	Tinggi
	10 – 100	Sedang
	< 10	Rendah

6. Pembagian Dataset (*Data Splitting*) : Sebelum melatih model, dataset lengkap dibagi menjadi dua bagian yang terpisah: data latih (*training set*) dan data uji

(*testing set*). Pembagian ini sangat penting untuk mengevaluasi kinerja model secara objektif dan menghindari *overfitting*, yaitu kondisi di mana model terlalu 'hafal' data latih tetapi tidak mampu menggeneralisasi dengan baik pada data baru. Proporsi pembagian yang umum digunakan dan akan diterapkan dalam penelitian ini adalah 80% untuk data latih dan 20% untuk data uji. Model akan dibangun dan "belajar" pola-pola dari data latih, sementara kinerjanya akan diukur menggunakan data uji yang belum pernah dilihat sebelumnya.

Sisa stok digunakan sebagai salah satu parameter karena jumlah stok berhubungan dengan kebutuhan promosi. Jika stok suatu produk masih banyak, maka promosi bisa membantu mempercepat penjualan agar tidak menumpuk terlalu lama. Sebaliknya, jika stok sudah sedikit, biasanya promosi tidak diperlukan karena produk akan habis terjual dengan sendirinya. Oleh karena itu, stok dianggap layak dijadikan parameter dalam sistem klasifikasi ini.

3.5 Simulasi Perhitungan *Random Forest*

Bagian ini menyajikan simulasi perhitungan manual untuk mendemonstrasikan mekanisme inti dari algoritma *Random Forest*. Karena *Random Forest* terdiri dari banyak pohon keputusan, simulasi ini akan berfokus pada pembangunan satu pohon keputusan (*Decision Tree*) representatif. Proses ini menggunakan metrik Ketakmurnian Gini (*Gini Impurity*) dan Gini Gain untuk menentukan pemisahan (*split*) terbaik pada setiap simpul (*node*).

Random Forest bekerja melalui proses yang disebut *Bootstrap Aggregating* atau *Bagging*. Proses ini terdiri dari tiga langkah utama:

1. *Bootstrap Sampling* : Membuat sejumlah B sampel data baru dari data latih asli. Setiap sampel (disebut sampel bootstrap) dibuat dengan mengambil data

secara acak *dengan pengembalian* (*with replacement*). Akibatnya, setiap sampel bootstrap akan sedikit berbeda satu sama lain.

2. Pelatihan Paralel : Sebuah pohon keputusan dilatih secara independen pada setiap sampel bootstrap. Untuk menambah keragaman, pada setiap simpul pohon, hanya sebagian kecil fitur yang dipilih secara acak untuk dipertimbangkan sebagai kriteria pemisahan.

3. Agregasi (Voting): Untuk mengklasifikasikan data baru, data tersebut dimasukkan ke semua pohon yang telah dibangun. Setiap pohon memberikan "suara" untuk satu kelas. Kelas yang menerima suara terbanyak (*majority vote*) menjadi prediksi akhir dari *Random Forest*.

Untuk simulasi ini, sebuah dataset kecil yang terdiri dari 10 produk diambil dari data primer.

Tabel 3. 2 Dataset Simulasi

No	Nama Produk	Penjualan	Margin	Stok	Kelas
1	Buah Naga Merah / Kg	Tinggi	Sedang	Tinggi	1
2	SB Midi Regular L	Tinggi	Tinggi	Tinggi	1
3	Le Minerale Air PET 600ml	Tinggi	Tinggi	Tinggi	1
4	Sampoerna A Mild 16	Sedang	Rendah	Sedang	1
5	Telur Ayam Ngr Bkl Pc	Tinggi	Sedang	Tinggi	1
6	Alfamidi Lighter Electric	Rendah	Rendah	Sedang	0
7	M2000 Lighter	Rendah	Rendah	Rendah	0
8	GG Surya Promild 16	Rendah	Rendah	Sedang	0
9	Club Air Mineral 600ml	Rendah	Rendah	Sedang	0
10	Mr Potato Crisps BBQ	Rendah	Rendah	Sedang	0

Dari tabel di atas, terdapat 5 produk Layak (Kelas 1) dan 5 produk Tidak Layak (Kelas 0).

3.5.1. Menentukan *Root Node* (Akar Pohon)

Root node adalah pemisahan pertama yang paling informatif. Untuk menentukannya, kita harus menghitung *Gini Gain* untuk setiap fitur.

A. Menghitung *Gini Impurity* Induk (*Gini Parent*)

Gini Impurity mengukur tingkat ketidakmurnian atau keragaman kelas dalam sebuah himpunan data. Nilainya berkisar dari 0 (semua data milik satu kelas) hingga 0.5 untuk dua kelas yang terdistribusi merata. Rumusnya adalah :

$$Gini = 1 - \sum_{i=1}^c (p_i)^2$$

di mana p_i adalah proporsi kelas ke- i .

Untuk dataset simulasi diatas (5 Layak, 5 Tidak Layak)

$$P_{layak} = \frac{5}{10} = 0,5$$

$$P_{tidak layak} = \frac{5}{10} = 0,5$$

$$Gini_{parent} = 1 - ((0.5)^2 + (0.5)^2) = 1 - (0.25 + 0.25) = 0.5$$

B. Menghitung *Gini Impurity* untuk Setiap Fitur

Kita perlu menghitung *Gini Impurity* tertimbang untuk setiap kemungkinan pemisahan.

1. Fitur : Penjualan

Penjualan = Tinggi (4 data) : 4 Layak, 0 Tidak

$$Gini_{penjualan=Tinggi} = 1 - \left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2 = 1 - 1 = 0$$

Penjualan = Sedang (1 data) : 1 Layak, 0 Tidak

$$Gini_{penjualan=Sedang} = 1 - \left(\frac{1}{1}\right)^2 + \left(\frac{0}{1}\right)^2 = 1 - 1 = 0$$

Penjualan = Rendah (5 data) : 0 Layak, 5 Tidak

$$Gini_{penjualan=Rendah} = 1 - \left(\frac{0}{5}\right)^2 + \left(\frac{5}{5}\right)^2 = 1 - 1 = 0$$

Gini tertimbang untuk fitur penjualan adalah :

$$Gini_{Split}(Penjualan) = \left(\frac{4}{10} \times 0\right) + \left(\frac{1}{10} \times 0\right) + \left(\frac{5}{10} \times 0\right) = 0$$

2. Fitur : Margin

Margin = Tinggi (2 data) : 2 Layak, 0 Tidak

$$Gini_{margin=Tinggi} = 1 - \left(\frac{2}{2}\right)^2 + \left(\frac{0}{2}\right)^2 = 1 - 1 = 0$$

Margin = Sedang (2 data) : 2 Layak, 0 Tidak

$$Gini_{margin=Sedang} = 1 - \left(\frac{2}{2}\right)^2 + \left(\frac{0}{2}\right)^2 = 1 - 1 = 0$$

Margin = Rendah (6 data) : 1 Layak, 5 Tidak

$$Gini_{margin=Rendah} = 1 - \left(\frac{1}{6}\right)^2 + \left(\frac{5}{6}\right)^2 = 1 - (0.027 + 0.694) = 0.278$$

Gini tertimbang untuk fitur margin adalah :

$$Gini_{Split}(Margin) = \left(\frac{2}{10} \times 0\right) + \left(\frac{2}{10} \times 0\right) + \left(\frac{6}{10} \times 0.278\right) = 0.1668$$

3. Fitur : Stok

Stok = Tinggi (4 data) : 4 Layak, 0 Tidak

$$Gini_{stok=Tinggi} = 1 - \left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2 = 1 - 1 = 0$$

Stok = Sedang (5 data) : 1 Layak, 4 Tidak

$$Gini_{stok=Sedang} = 1 - \left(\frac{1}{5}\right)^2 + \left(\frac{4}{5}\right)^2 = 1 - (0.04 + 0.64) = 0.32$$

Stok = Rendah (1 data) : 0 Layak, 1 Tidak

$$Gini_{stok=Rendah} = 1 - \left(\frac{0}{1}\right)^2 + \left(\frac{1}{1}\right)^2 = 0$$

Gini tertimbang untuk fitur stok adalah :

$$Gini_{Split}(Stok) = \left(\frac{4}{10} \times 0\right) + \left(\frac{5}{10} \times 0.32\right) + \left(\frac{1}{10} \times 0\right) = 0.1$$

C. Menghitung *Gini Gain* dan Memilih *Root Node*

Gini Gain mengukur seberapa besar penurunan ketidakmurnian setelah pemisahan. Fitur dengan *Gini Gain* tertinggi akan dipilih. Rumusnya adalah :

$$Gini\ Gain(Fitur) = Gini_{Parent} - Gini_{Split}(Fitur)$$

Perhitungan untuk Keseluruhan adalah

$$Gini\ Gain(Penjualan) = 0.5 - 0 = 0.5$$

$$Gini\ Gain(Margin) = 0.5 - 0.1668 = 0.3332$$

$$Gini\ Gain(Stok) = 0.5 - 0.16 = 0.34$$

Tabel 3. 3 Perhitungan Gini Gain untuk Pemilihan Root Node

Fitur	Gini Split Tertimbang	Gini Gain
Penjualan	0.0000	0.5000
Margin	0.1668	0.3332
Stok	0.1600	0.3400

Fitur **Penjualan** memiliki *Gini Gain* tertinggi (0.5), sehingga dipilih sebagai *root node*.

3.5.2. Membangun Cabang Pohon (*Recursive Splitting*)

Setelah *root node* ditentukan (Penjualan), dataset dibagi menjadi tiga cabang: Tinggi, Sedang, dan Rendah.

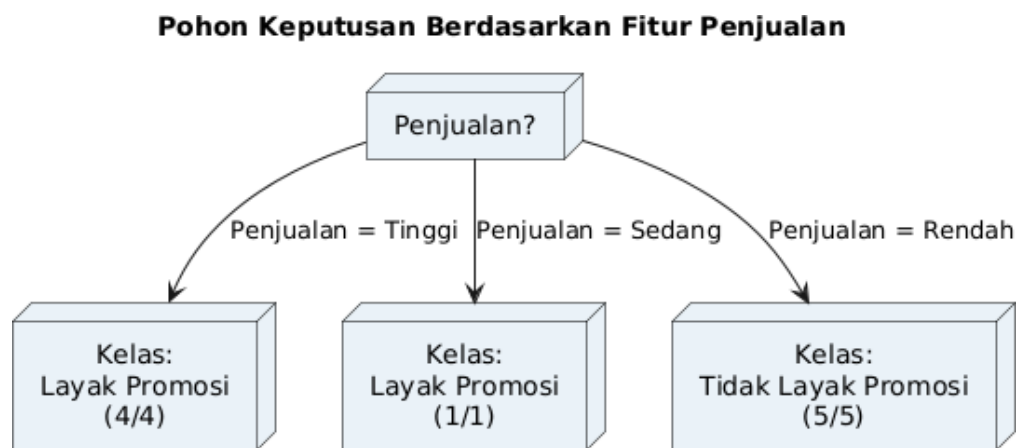
1. Cabang Penjualan = Tinggi: berisi 4 data, semuanya Layak. Node ini menjadi *leaf node* (daun) karena sudah murni ($Gini = 0$).
2. Cabang Penjualan = Sedang: berisi 1 data, yaitu Layak. Node ini juga menjadi *leaf node*.
3. Cabang Penjualan = Rendah: berisi 5 data, semuanya Tidak Layak. Node ini juga menjadi *leaf node*.

Dalam contoh simulasi ini, pohon keputusan selesai terbentuk hanya dalam satu level karena pemisahan pertama menggunakan fitur 'Penjualan' sudah menghasilkan cabang-cabang yang murni. Dalam kasus nyata yang lebih kompleks, jika sebuah cabang belum murni, proses perhitungan Gini Gain akan diulang pada cabang tersebut menggunakan fitur-fitur yang tersisa untuk menemukan pemisahan terbaik berikutnya. Proses ini berlanjut secara rekursif hingga semua cabang

menjadi murni atau kriteria pemberhentian lain terpenuhi (misalnya, kedalaman maksimum pohon).

3.5.3. Struktur Pohon Keputusan dan Ekstraksi Aturan

Berdasarkan perhitungan di atas, struktur pohon keputusan yang terbentuk sangat sederhana namun efektif untuk dataset simulasi ini.



Gambar 3. 2 Pohon Keputusan Terbentuk

Dari struktur pohon tersebut, dapat diekstraksi sejumlah aturan keputusan (*IF-THEN rules*) yang logis dan mudah diinterpretasikan.

Tabel 3. 4 Aturan Keputusan (*Rules*) dari Pohon Keputusan

No	Aturan (IF...THEN)	Keterangan
1	JIKA Penjualan = Tinggi MAKA Kelas = Layak Promosi	Produk dengan penjualan tinggi direkomendasikan untuk promosi.
2	JIKA Penjualan = Sedang MAKA Kelas = Layak Promosi	Produk dengan penjualan sedang juga layak untuk dipertimbangkan promosi.
3	JIKA Penjualan = Rendah MAKA Kelas = Tidak Layak Promosi	Produk dengan penjualan rendah tidak diprioritaskan untuk promosi.

3.5.4. Agregasi dan Prediksi Akhir (*Voting*)

Pohon keputusan yang ditunjukkan sebelumnya hanyalah satu dari banyak pohon yang dibangun oleh algoritma Random Forest. Setiap pohon dilatih menggunakan sampel *bootstrap* berbeda dan mempertimbangkan subset fitur yang berbeda di setiap simpul.

Sebagai ilustrasi, misalkan terdapat produk baru dengan karakteristik:

- a. Penjualan: Tinggi
- b. Margin: Rendah
- c. Stok: Sedang

Produk tersebut akan dievaluasi oleh seluruh pohon dalam *hutan* (forest).

Contoh prediksi dari beberapa pohon:

- a. Pohon 1 memprediksi: Layak
- b. Pohon 2 memprediksi: Layak
- c. Pohon 3 memprediksi: Layak
- d. ...
- e. Pohon 100 memprediksi: 95 Layak, 5 Tidak Layak

Dengan demikian, berdasarkan mayoritas suara (*voting*), produk tersebut akan diklasifikasikan sebagai Layak Promosi. Proses agregasi inilah yang menjadi kekuatan utama dari Random Forest dalam meningkatkan akurasi dan mengurangi *overfitting*.

3.5.5. Evaluasi Model

Evaluasi model adalah tahap fundamental untuk mengukur seberapa baik kinerja model klasifikasi yang telah dilatih. Evaluasi ini dilakukan menggunakan *testing set*, yaitu data yang tidak digunakan selama proses pelatihan, untuk

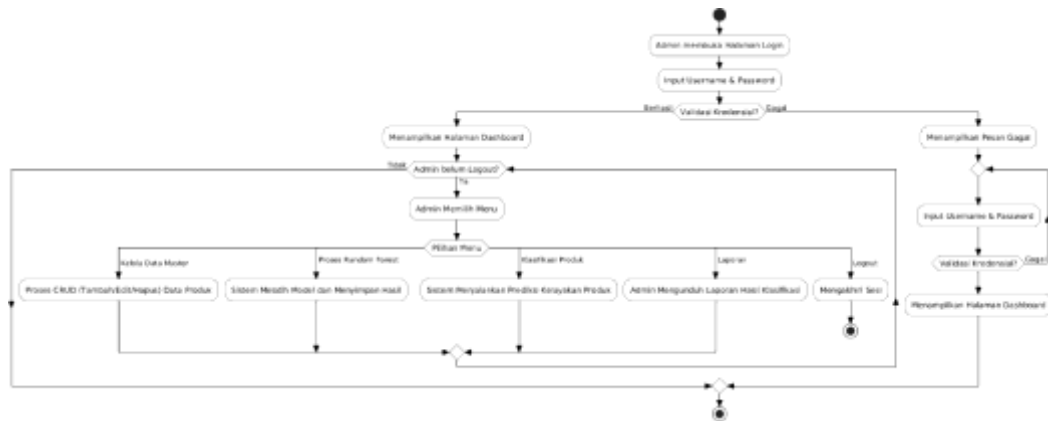
memberikan penilaian yang objektif terhadap kemampuan generalisasi model. Metrik utama yang digunakan berasal dari **Confusion Matrix**.

Confusion Matrix adalah tabel 2x2 yang memvisualisasikan performa model dengan membandingkan hasil prediksi dengan kelas aktual. Tabel ini memiliki empat komponen utama:

- a. **True Positive (TP)**: Jumlah produk yang secara aktual Layak Promosi dan diprediksi dengan benar sebagai Layak Promosi.
- b. **True Negative (TN)**: Jumlah produk yang secara aktual Tidak Layak Promosi dan diprediksi dengan benar sebagai Tidak Layak Promosi.
- c. **False Positive (FP)**: Jumlah produk yang secara aktual Tidak Layak Promosi tetapi salah diprediksi sebagai Layak Promosi (Kesalahan Tipe I).
- d. **False Negative (FN)**: Jumlah produk yang secara aktual Layak Promosi tetapi salah diprediksi sebagai Tidak Layak Promosi (Kesalahan Tipe II).

3.6 *Flowchart* Aplikasi

Untuk memberikan gambaran umum mengenai alur interaksi pengguna dengan fungsionalitas sistem, dibuatlah sebuah *flowchart* aplikasi. *Flowchart* ini memvisualisasikan langkah-langkah yang akan dilalui oleh admin, mulai dari proses autentikasi (*login*), mengakses fitur-fitur utama seperti pengelolaan data, pelatihan model, proses klasifikasi, hingga mengunduh laporan dan keluar dari sistem (*logout*). *Flowchart* terkait dapat dilihat pada gambar dibawah ini.



Gambar 3. 3 Flowchart Aplikasi

1. Autentikasi Pengguna: Proses dimulai saat Admin membuka halaman login untuk memasukkan *username* dan *password*. Sistem akan langsung melakukan validasi terhadap kredensial yang diinput.

2. Penanganan Hasil Login:

Jika Gagal: Sistem akan menampilkan pesan kegagalan, dan Admin diminta untuk kembali memasukkan kredensial hingga berhasil.

Jika Berhasil: Admin akan diarahkan ke halaman *dashboard* utama, yang menjadi pusat navigasi sistem.

3. Akses Fitur Utama: Selama Admin belum melakukan *logout*, ia dapat memilih berbagai menu fungsional yang tersedia di *dashboard*.

4. Pilihan Menu Fungsional: Berdasarkan menu yang dipilih Admin, sistem akan menjalankan fungsi spesifik:

5. Kelola Data Master: Admin dapat melakukan manajemen data produk, seperti menambah data baru, mengubah data yang ada, atau menghapus data.

6. Proses Random Forest: Admin bisa memulai proses pelatihan model *machine learning*. Sistem akan mengambil data, melatih model, dan menyimpannya untuk digunakan nanti.

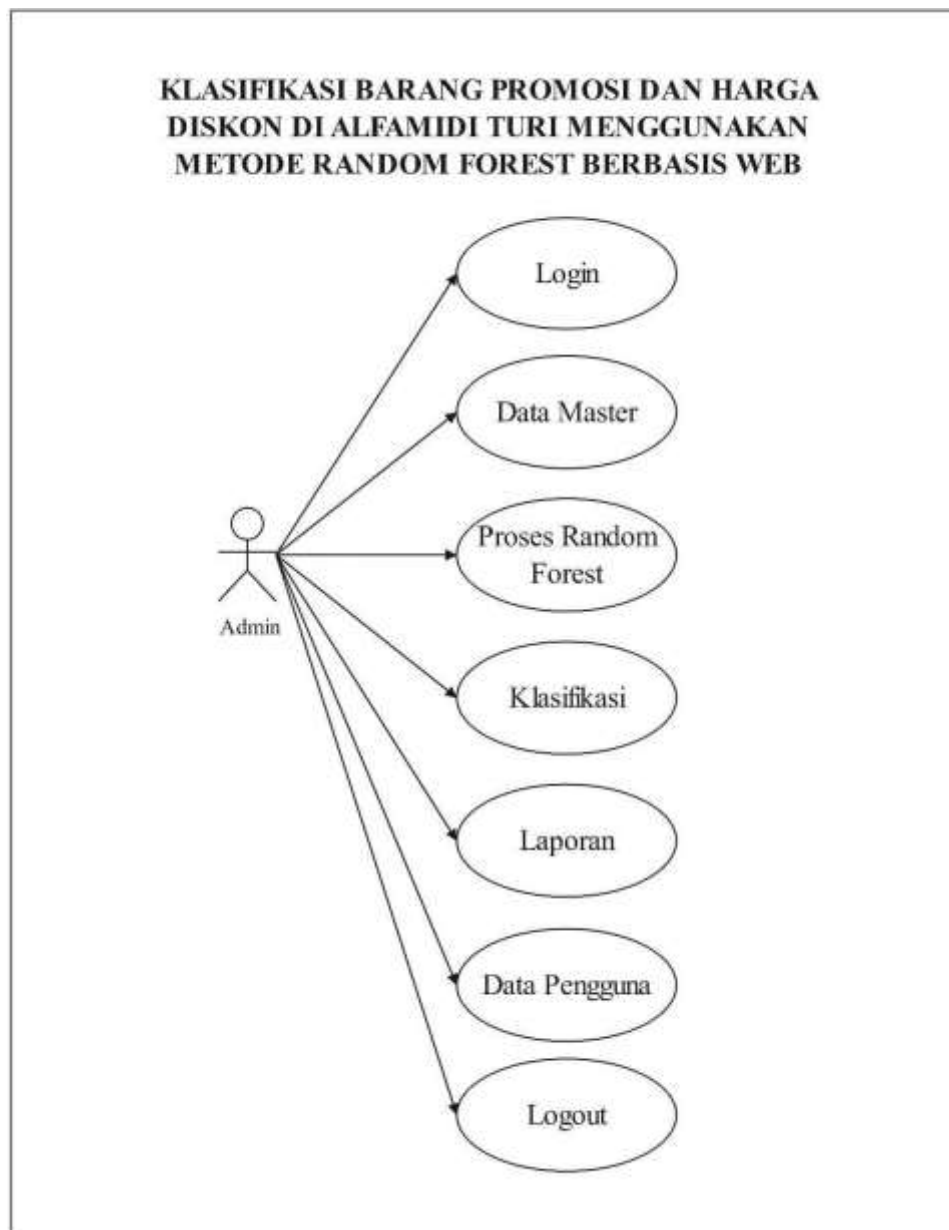
7. Klasifikasi Produk: Fitur ini digunakan untuk menerapkan model yang telah dilatih guna memprediksi kelayakan promosi dari produk-produk yang ada.
8. Laporan: Admin dapat mengunduh laporan hasil dari proses klasifikasi.
9. Akhir Sesi: Proses akan berakhir jika Admin memilih opsi Logout. Sistem akan mengakhiri sesi dan mengarahkan Admin kembali ke halaman login awal.

3.7 Desain Sistem

Proses ini berfokus pada perancangan seperti struktur data dan arsitektur perangkat lunak yang dibuat dengan pemodelan UML seperti *use case diagram*, *activity diagram*, *sequence diagram*, dan *class diagram*.

3.7.1 Use Case Diagram

Use Case Diagram dibawah ini menggambarkan interaksi antara Admin sebagai aktor tunggal dengan fungsionalitas utama sistem klasifikasi berbasis web. Admin dapat melakukan *Login* untuk mengakses sistem, kemudian mengelola Data Master dan Data Pengguna. Fungsi inti sistem adalah Proses Random Forest untuk melatih model, Klasifikasi untuk menerapkan model pada produk, serta Laporan untuk melihat hasil analisis, dan diakhiri dengan proses Logout untuk keluar dari sistem.



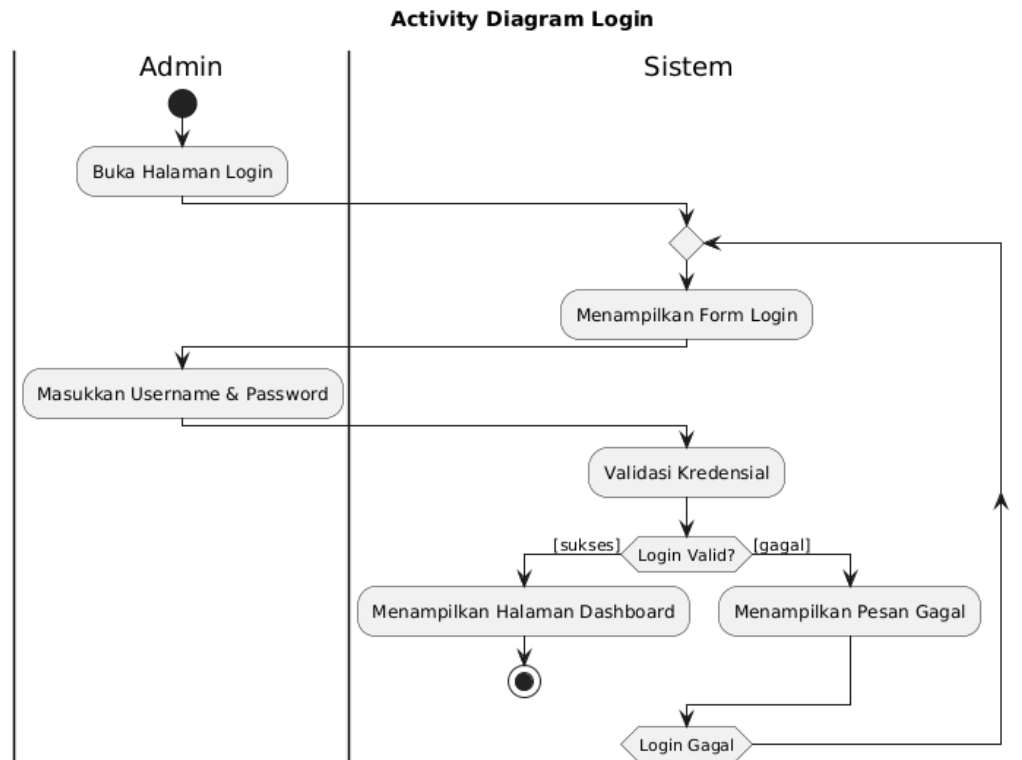
Gambar 3. 4 Use Case Diagram

3.7.2 Activity Diagram

1. Activity Diagram Login

Diagram ini menggambarkan alur kerja dimana Admin memasukkan username dan password untuk mengakses sistem. Sistem kemudian memvalidasi

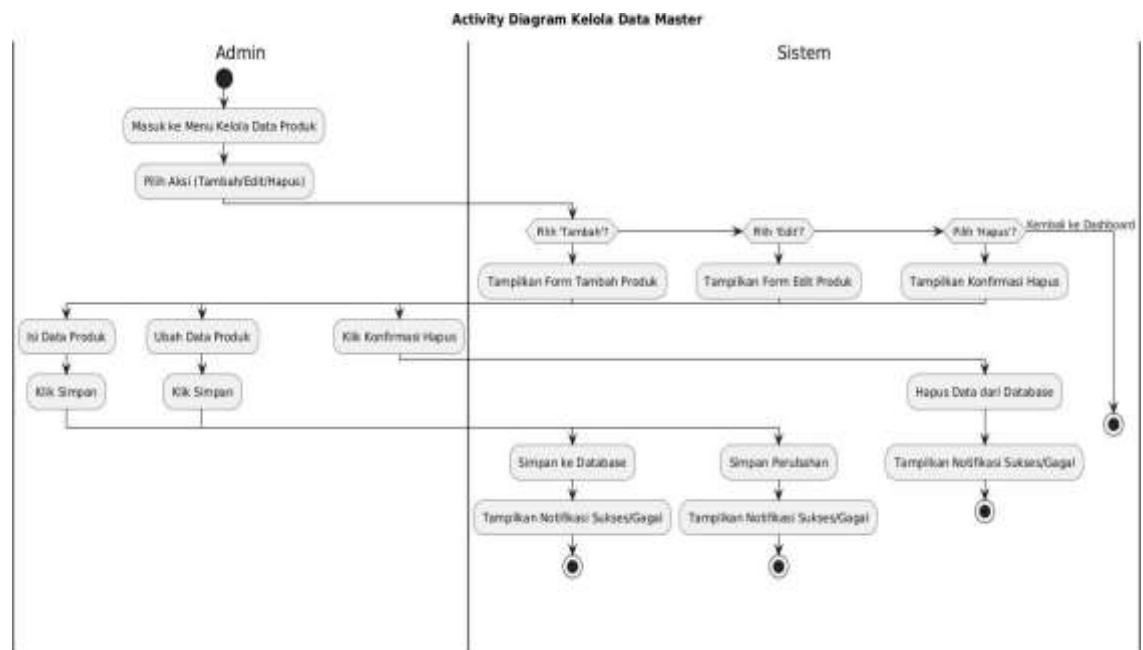
kredensial tersebut, dan jika berhasil akan menampilkan halaman dashboard, namun jika gagal akan menampilkan pesan kesalahan dan meminta login ulang.



Gambar 3. 5 Activity Diagram *Login*

2. *Activity Diagram* Kelola Data Master

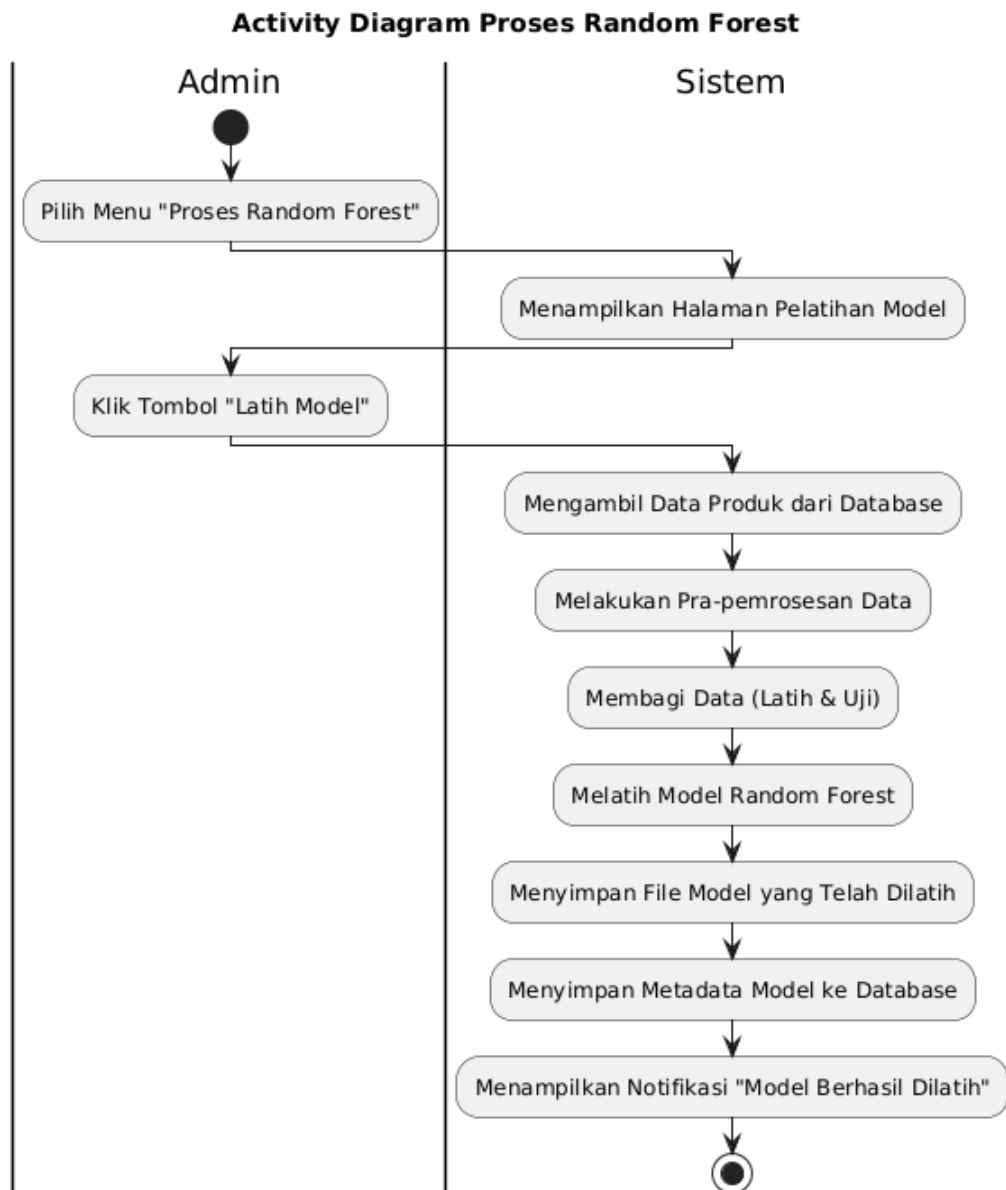
Diagram ini menunjukkan proses Admin dalam mengelola data produk dengan memilih salah satu aksi yaitu Tambah, Edit, atau Hapus. Sistem akan merespons sesuai aksi yang dipilih, memprosesnya ke database, dan memberikan notifikasi kembali kepada Admin.



Gambar 3. 6 Activity Diagram Kelola Data Master

3. *Activity Diagram Proses Random Forest*

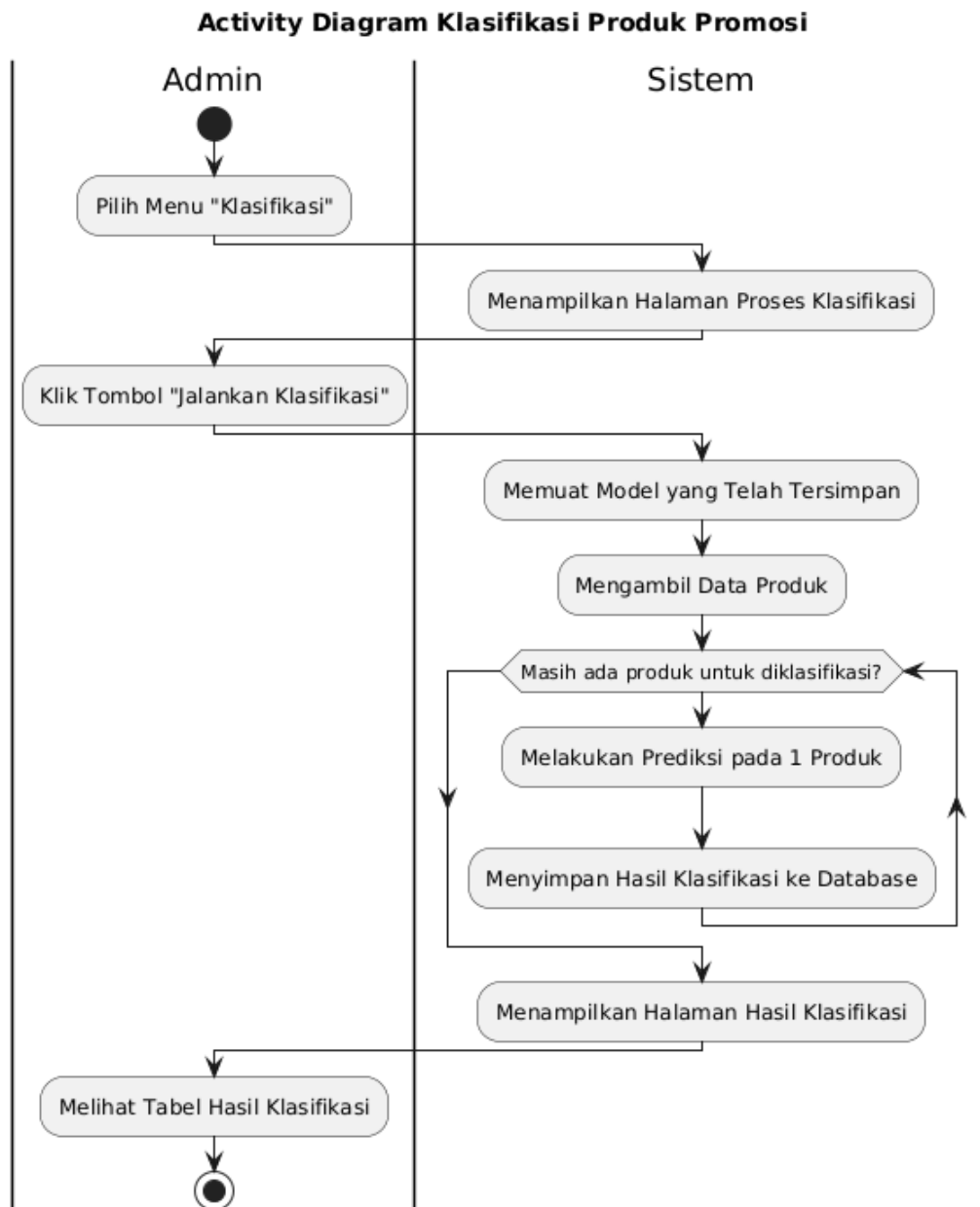
Diagram ini menjelaskan alur kerja saat Admin memulai proses pelatihan model berdasarkan data produk yang tersedia. Sistem akan melakukan seluruh langkah teknis mulai dari pengambilan data, pelatihan, hingga penyimpanan model untuk digunakan pada proses klasifikasi.



Gambar 3. 7 Activity Diagram Proses Random Forest

4. *Activity Diagram* Klasifikasi Produk Promosi

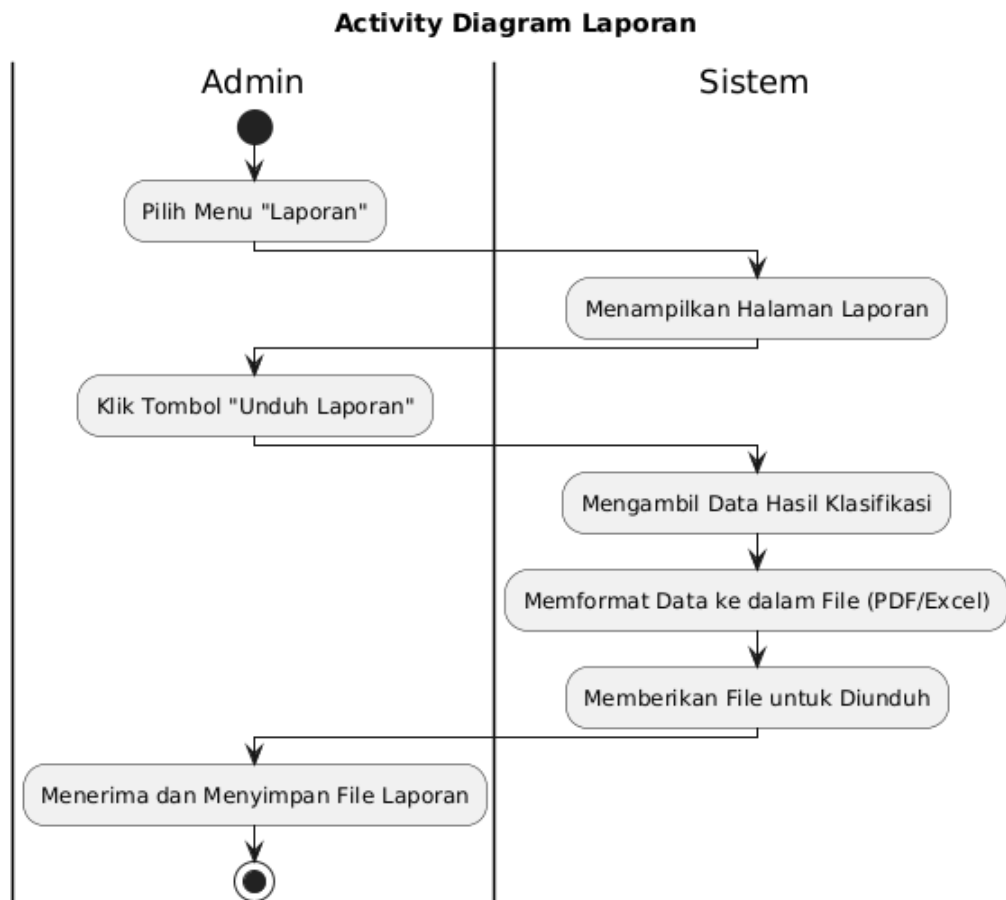
Diagram ini menggambarkan bagaimana Admin menjalankan proses klasifikasi untuk menentukan kelayakan promosi produk. Sistem akan menggunakan model yang sudah terlatih untuk memprediksi setiap produk dan menyimpan hasilnya ke dalam database.



Gambar 3. 8 Activity Diagram Klasifikasi Produk Promosi

5. Activity Diagram Laporan

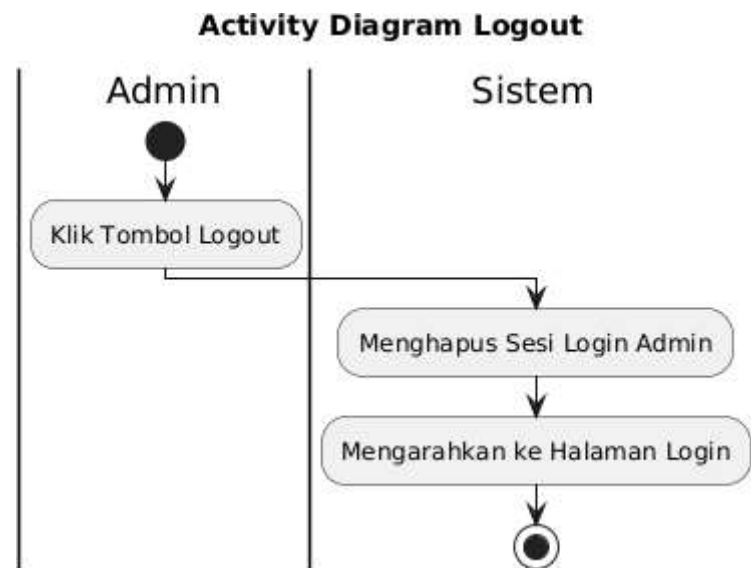
Diagram ini menunjukkan proses saat Admin ingin mengunduh laporan dari hasil klasifikasi yang telah dilakukan. Sistem akan mengambil data yang relevan, memformatnya menjadi sebuah file seperti PDF atau Excel, lalu memberikannya kepada Admin.



Gambar 3. 9 Activity Diagram Laporan

6. *Activity Diagram Logout*

Diagram ini menjelaskan proses sederhana dimana Admin memilih untuk keluar dari sistem. Sistem kemudian akan mengakhiri sesi yang sedang berjalan dan mengarahkan Admin kembali ke halaman Login.



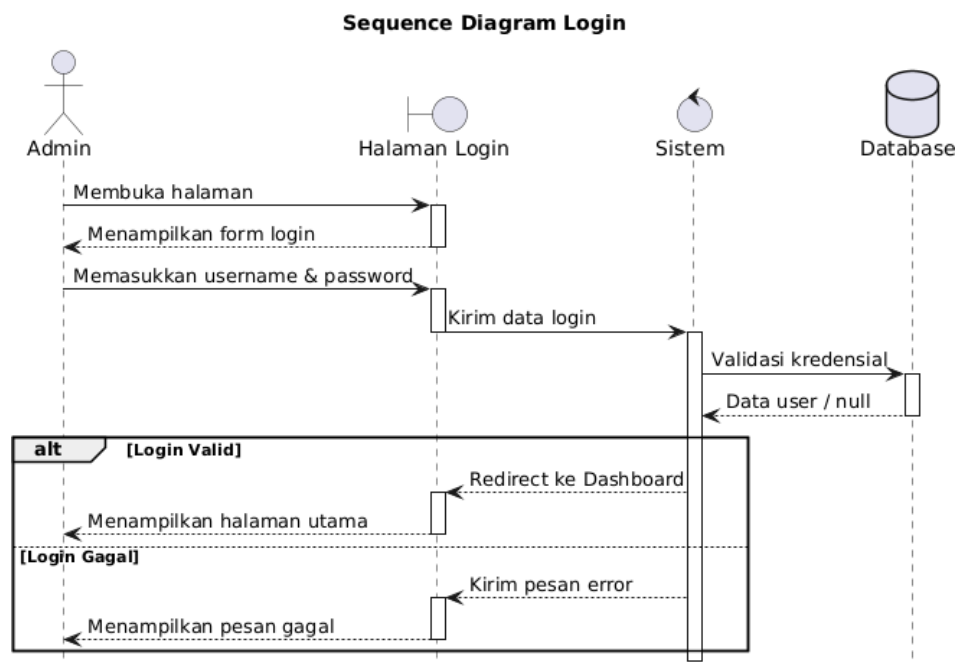
Gambar 3. 10 Activity Diagram Logout

3.7.3 Sequence Diagram

1. Sequence Diagram Login

Diagram ini menggambarkan urutan proses saat Admin melakukan login.

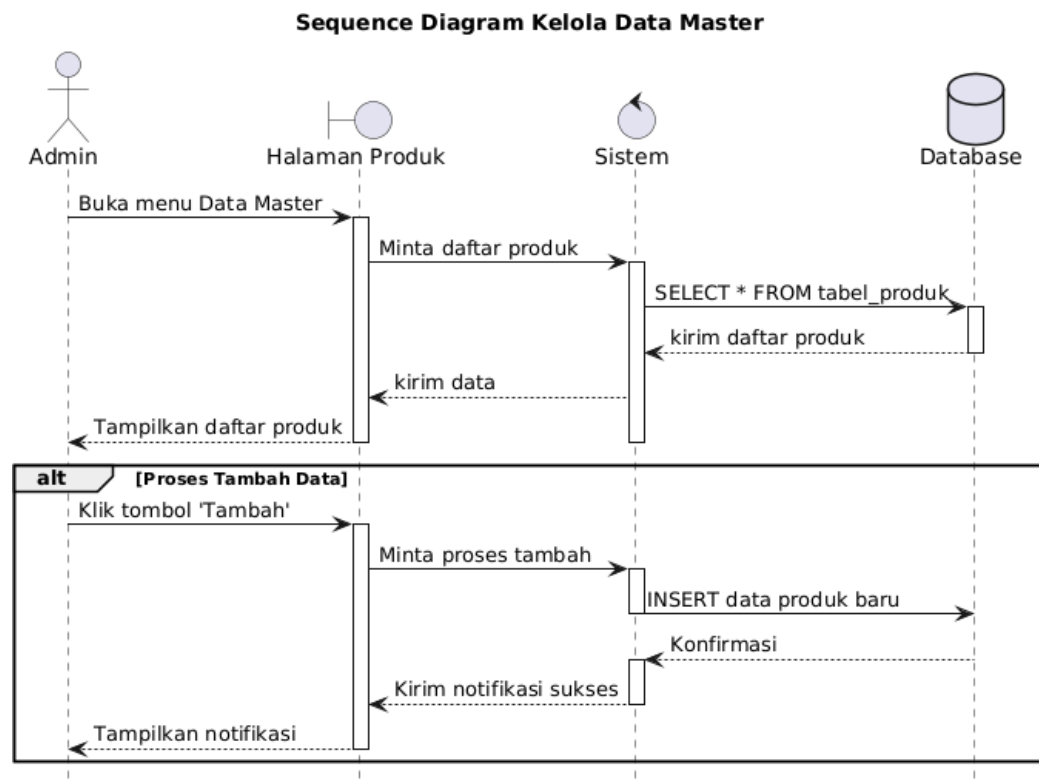
Admin berinteraksi dengan Halaman Login



Gambar 3. 11 Sequence Diagram Login

2. Sequence Diagram Kelola Data Master

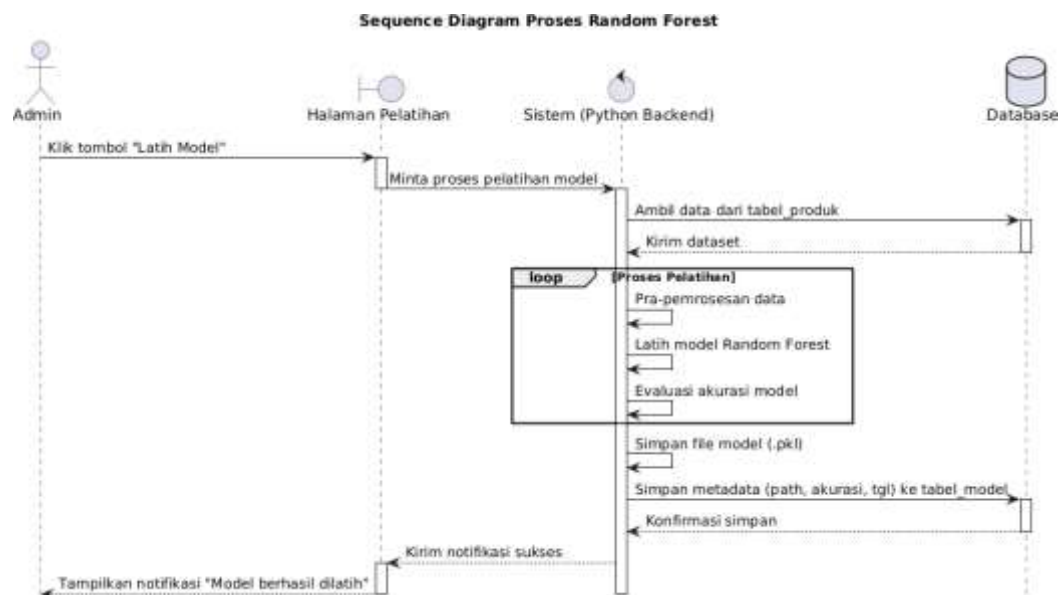
Diagram ini menunjukkan interaksi saat Admin mengelola data produk. Prosesnya mencakup menampilkan data, serta alur untuk menambah, mengedit, dan menghapus data yang semuanya berinteraksi dengan Sistem dan *Database*.



Gambar 3. 12 Sequence Diagram Kelola Data Master

3. Sequence Diagram Proses Random Forest

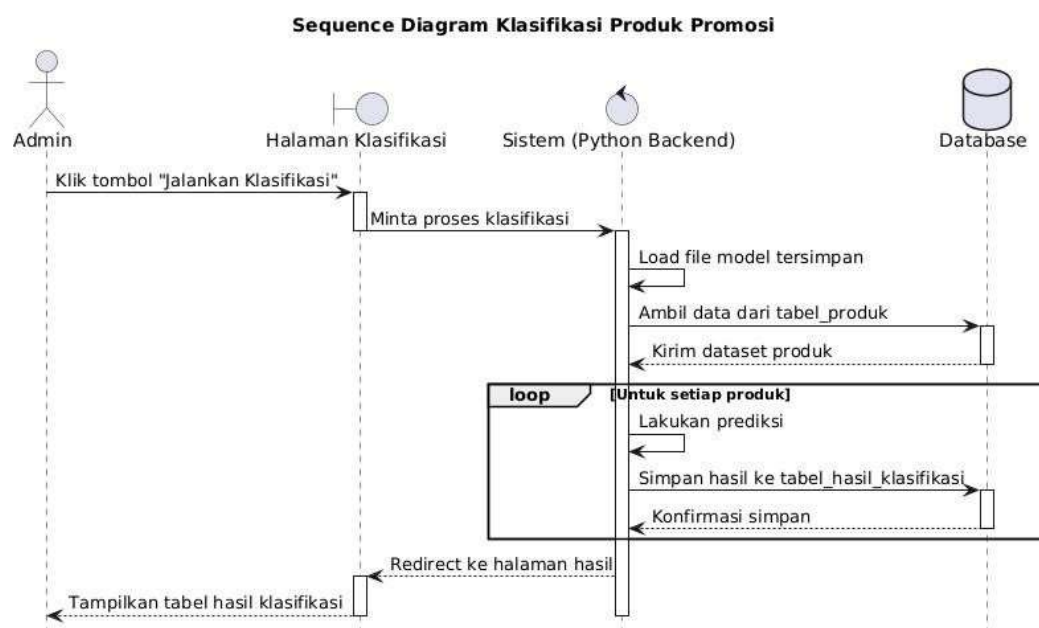
Diagram ini merinci urutan langkah teknis saat Admin memulai pelatihan model. Sistem akan mengambil data dari Database, melakukan serangkaian proses internal (pra-pemrosesan, latih, evaluasi), lalu menyimpan file model dan metadatanya kembali ke *Database*.



Gambar 3. 13 Sequence Diagram Proses Random Forest

4. Sequence Diagram Klasifikasi Produk Promosi

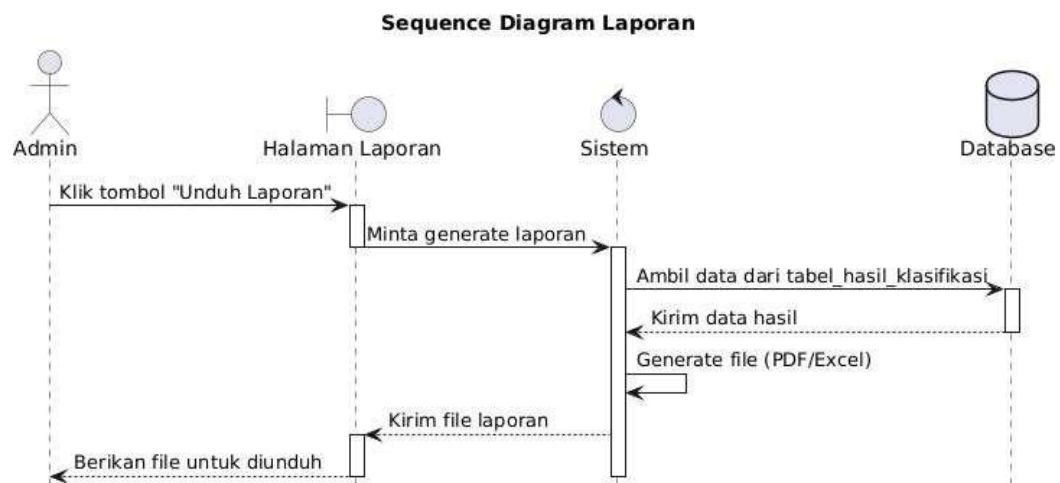
Diagram ini menjelaskan alur saat Admin menjalankan klasifikasi. Sistem memuat model yang ada, mengambil data produk, lalu melakukan prediksi untuk setiap produk dan menyimpan hasilnya satu per satu ke dalam Database.



Gambar 3. 14 Sequence Diagram Klasifikasi Produk Promosi

5. *Sequence Diagram Laporan*

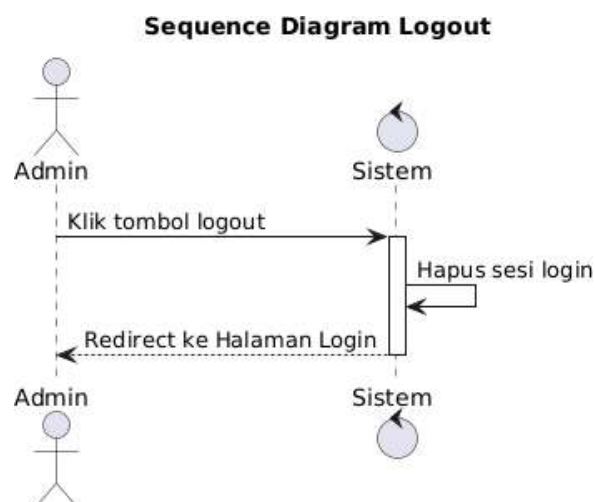
Diagram ini menggambarkan urutan interaksi saat Admin mengunduh laporan. Sistem mengambil data hasil klasifikasi dari Database, memformatnya menjadi sebuah file, dan mengirimkannya ke Admin.



Gambar 3. 15 *Sequence Diagram Laporan*

6. *Sequence Diagram Logout*

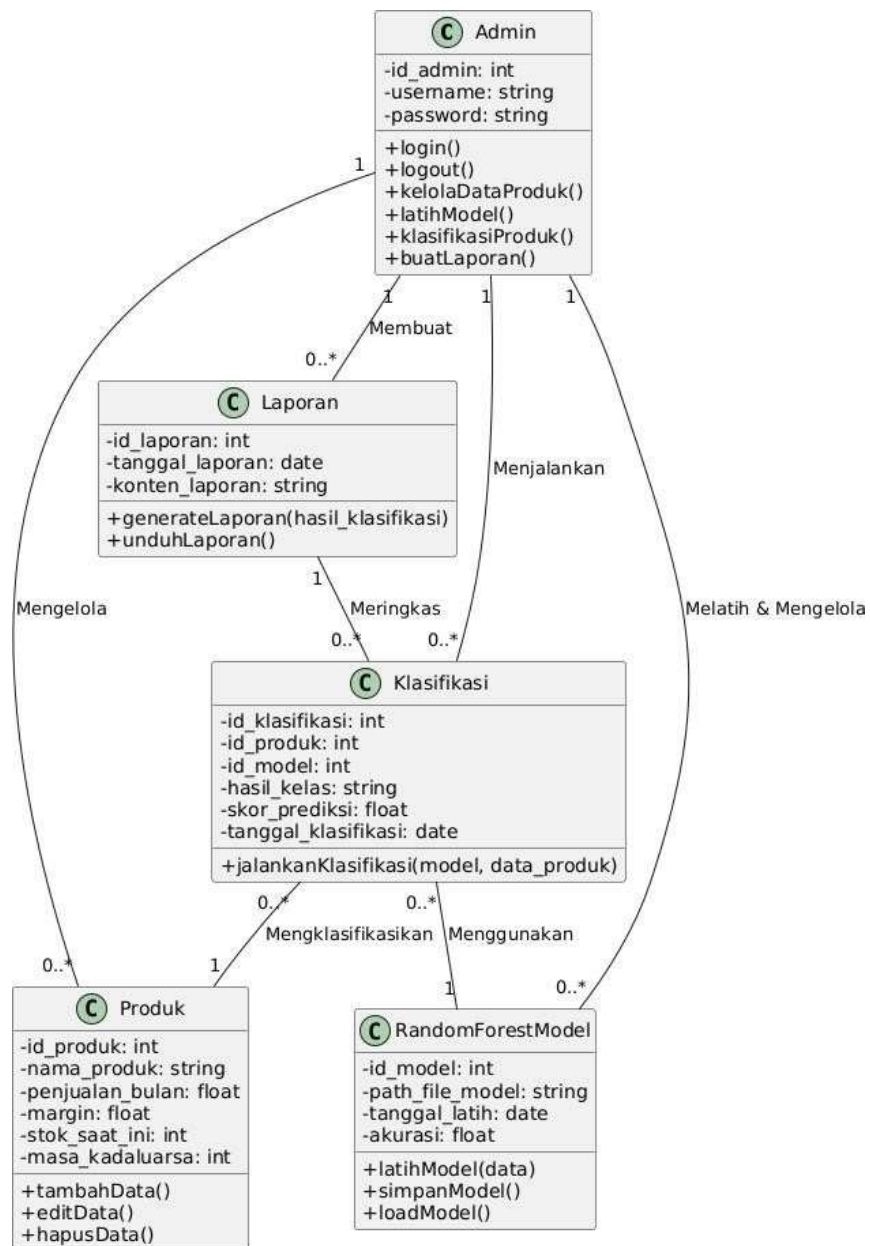
Diagram ini menunjukkan interaksi sederhana saat Admin keluar. Sistem akan menghapus sesi login dan mengarahkan Admin kembali ke Halaman Login.



Gambar 3. 16 *Sequence Diagram Logout*

3.7.4 Class Diagram

Class Diagram ini menunjukkan struktur sistem. dapat dilihat pada gambar dibawah ini.



Gambar 3. 17 Class Diagram

3.8 Rancangan Basis Data

1. Rancangan Tabel Admin

Tabel ini menyimpan data kredensial untuk admin yang dapat mengakses dan mengelola sistem.

Tabel 3. 5 Rancangan Tabel Admin

Field	Tipe Data	Keterangan
id_admin	INT (PK)	Primary key, ID unik untuk admin
username	VARCHAR(50)	Nama pengguna admin
password	VARCHAR(255)	Kata sandi terenkripsi

2. Rancangan Tabel Produk

Tabel ini berisi data mentah mengenai setiap produk, yang akan menjadi *fitur* dalam proses pelatihan dan klasifikasi.

Tabel 3. 6 Rancangan Tabel Produk

Field	Tipe Data	Keterangan
id_produk	INT (PK)	Primary key, ID unik produk
nama_produk	VARCHAR(100)	Nama produk
Penjualan_bulan	FLOAT	Rata-rata jumlah unit terjual per bulan
margin	FLOAT	Persentase margin keuntungan produk
stok_saat_ini	INT	Jumlah stok prod
masa_kadaluarsa	INT	Sisa umur produk dalam satuan hari

3. Rancangan Tabel Model

Tabel ini berfungsi untuk menyimpan jejak dan metadata dari setiap model *Random Forest* yang dilatih.

Tabel 3. 7 Rancangan Tabel Model

Field	Tipe Data	Keterangan
id_model	INT (PK)	Primary key, ID unik untuk setiap versi model yang disimpan.
path_file_model	VARCHAR (255)	Lokasi penyimpanan file model
Tanggal_latih	DATETIME	Waktu kapan model dilatih
akurasi	FLOAT	Nilai akurasi model saat diuji setelah pelatihan
id_admin	INT (FK)	<i>Foreign Key</i> ke tabel_admin untuk mencatat siapa yang melatih

4. Rancangan Tabel Hasil Klasifikasi

Tabel ini mencatat setiap output dari proses klasifikasi untuk setiap produk, sehingga riwayatnya dapat dilihat kembali.

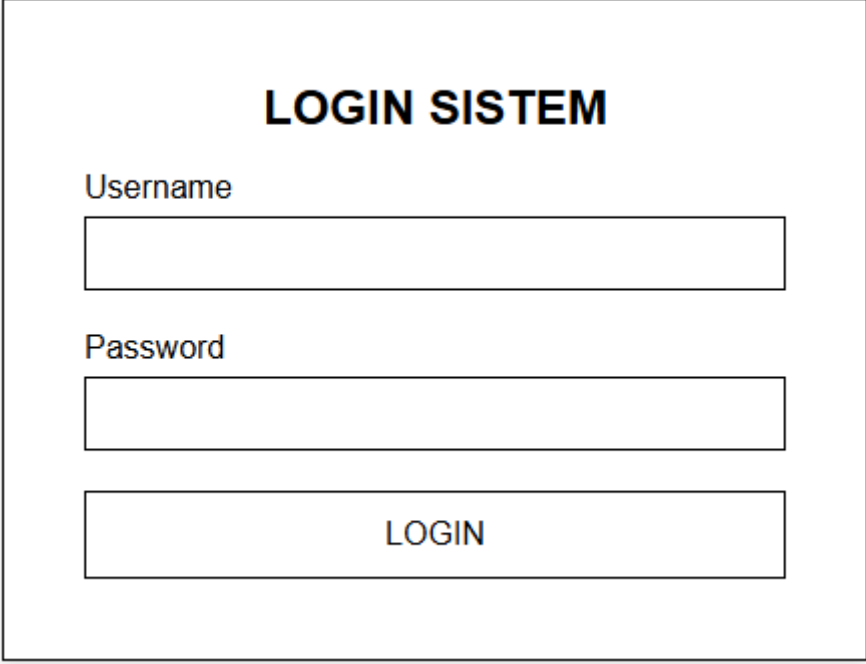
Tabel 3. 8 Rancangan Tabel Klasifikasi

Field	Tipe Data	Keterangan
id_klasifikasi	INT (PK)	<i>Primary Key</i> , ID unik untuk setiap record hasil klasifikasi
Id_produk	INT (FK)	<i>Foreign Key</i> ke tabel_produk untuk menunjuk produk yang diklasifikasi.
Id_model	INT (FK)	<i>Foreign Key</i> ke tabel_model untuk menunjuk model yang digunakan.
Hasil_kelas	VARCHAR(50)	Hasil klasifikasi (misal: "Layak Promosi", "Tidak Layak Promosi").
Skor_prediksi	FLOAT	Skor kepercayaan (probabilitas) dari prediksi model.
Tanggal_klasifikasi	DATETIME	Waktu kapan proses klasifikasi ini dijalankan.

3.8 Desain *User Interface*

1. Desain Halaman *Login*

Halaman ini merupakan gerbang utama sistem yang meminta Admin untuk memasukkan *username* dan *password*.



The image shows a login interface design within a rectangular frame. At the top center, the text "LOGIN SISTEM" is displayed in a bold, black, sans-serif font. Below this title, there are two input fields. The first is labeled "Username" in a standard black font, followed by a horizontal rectangular input box. The second is labeled "Password" in a standard black font, followed by another horizontal rectangular input box. At the bottom of the form, there is a wide rectangular button with the word "LOGIN" centered on it in a bold, black, sans-serif font.

Gambar 3. 18 Desain Halaman *Login*

2. Desain Halaman Kelola Data Master

Halaman ini berfungsi sebagai pusat untuk melihat dan mengelola semua data produk yang ada di sistem. Terdapat tabel yang menampilkan daftar produk beserta tombol aksi untuk menambah, mengedit, atau menghapus data.

No	Nama Produk	Penjualan/Bulan	Aksi
1	Indomie Mi Goreng	3200	Edit Hapus

Gambar 3. 19 Desain Halaman Kelola Data Master

3. Desain Halaman Tambah / Edit Data Produk

Halaman ini digunakan untuk mengisi formulir, baik untuk menambah data baru maupun mengedit data lama.

Gambar 3. 20 Desain Halaman Tambah / Edit Produk

4. Desain Notifikasi Hapus Produk

Interface ini bukanlah halaman penuh, melainkan sebuah dialog atau *modal* yang muncul di atas halaman untuk meminta konfirmasi.



Gambar 3. 21 Desain Interface Hapus

5. Desain Halaman Proses *Random Forest*

Halaman ini adalah pusat kendali untuk melatih model *Random Forest*.



Gambar 3. 22 Desain Halaman Proses *Random Forest*

6. Desain Halaman Klasifikasi

Halaman ini menampilkan hasil dari proses klasifikasi untuk menentukan produk layak atau tidak layak dipromosikan.

**SISTEM
KLASIFIKASI**
Dashboard
Kelola Data Master
Proses Random Forest
Klasifikasi
Laporan
Logout

Klasifikasi Produk Promosi

Jalankan Klasifikasi Sekarang

Hasil Klasifikasi

No	Nama Produk	Hasil Klasifikasi
1	Indomie Mi Goreng	Layak Promosi
2	Sabun Lifebuoy 85g	Tidak Layak Promosi

Gambar 3. 23 Desain Halaman Klasifikasi

7. Desain Halaman Laporan

Halaman ini didedikasikan untuk fungsi mengunduh laporan, dapat dilihat pada gambar dibawah ini.

**SISTEM
KLASIFIKASI**
Dashboard
Kelola Data Master
Proses Random Forest
Klasifikasi
Laporan
Logout

Laporan Hasil Klasifikasi

Gunakan form di bawah ini untuk mengunduh laporan

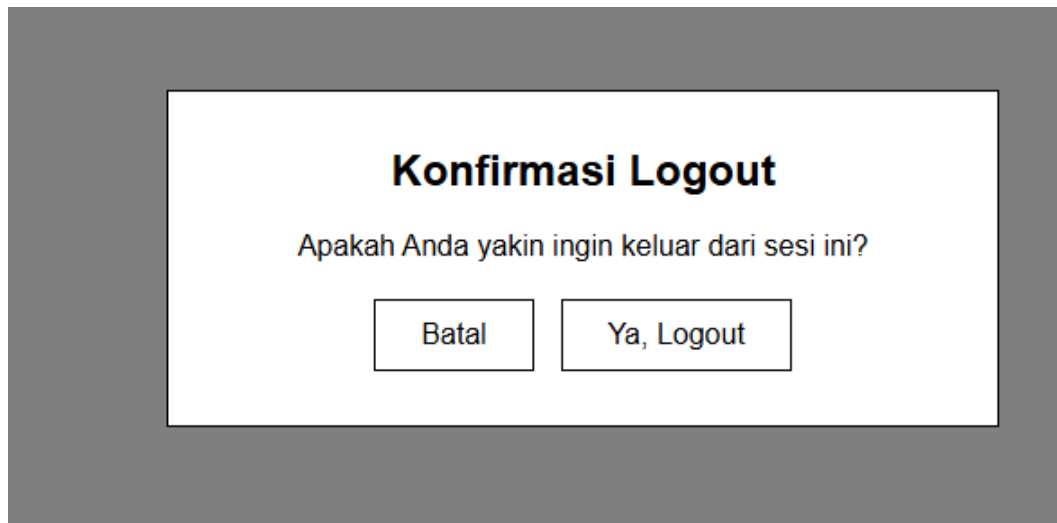
Filter Laporan

Unduh Laporan

Gambar 3. 24 Desain Halaman Laporan

8. Desain *Interface Logout*

Desain *Interface* ini muncul Ketika admin ingin *logout* dari sistem, dapat dilihat pada gambar dibawah ini.



Gambar 3. 25 Desain *Interface Logout*

BAB IV

HASIL DAN PEMBAHASAN

4.1. Perhitungan Manual *Random Forest*

Sesuai dengan tujuan penelitian yang telah diuraikan pada Bab 1, implementasi metode *Random Forest* bertujuan untuk membangun sebuah model klasifikasi yang mampu membedakan produk mana yang layak dan tidak layak untuk dipromosikan berdasarkan data historis. Pendekatan ini dirancang untuk menggantikan proses pengambilan keputusan yang bersifat subjektif dengan analisis berbasis data yang objektif dan sistematis.

Dalam pemodelan ini, variabel-variabel yang digunakan sebagai prediktor (fitur) adalah atribut kuantitatif yang secara logis memengaruhi keputusan promosi, sebagaimana telah diidentifikasi pada Bab 3. Fitur-fitur tersebut meliputi :

1. Penjualan : rata-rata penjualan bulanan, yang merefleksikan tingkat permintaan dan popularitas produk
2. Margin : Persentase margin keuntungan, yang mengindikasikan profitabilitas produk bagi toko.
3. Stok : Jumlah stok saat ini, yang menunjukkan urgensi perputaran barang

Variabel target yang ingin diprediksi oleh model adalah **Kelas**, sebuah variabel kategorikal yang memiliki dua kemungkinan nilai :

1. Layak Promosi (diberi label '1'): Produk yang memiliki karakteristik ideal untuk dimasukkan ke dalam program promosi.
2. Tidak Layak Promosi (diberi label '0'): Produk yang tidak menunjukkan potensi untuk memberikan hasil optimal jika dipromosikan.

Proses ini merupakan tugas klasifikasi *supervised learning*, di mana algoritma akan "belajar" pola dari data yang telah dilabeli untuk kemudian dapat mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Untuk memberikan pemahaman yang mendalam mengenai cara kerja algoritma *Random Forest*, bagian ini menyajikan simulasi perhitungan manual secara bertahap. *Random Forest* adalah sebuah metode *ensemble* yang terdiri dari banyak pohon keputusan (*decision tree*). Demi kejelasan dan kemudahan pemahaman, simulasi ini akan berfokus pada proses pembangunan satu pohon keputusan representatif, yang merupakan unit fundamental dari sebuah *forest*. Proses ini akan menggunakan metrik Ketakmurnian Gini (*Gini Impurity*) untuk mengevaluasi kualitas pemisahan pada setiap simpul pohon.

Simulasi ini menggunakan subset data kecil yang terdiri dari 10 produk, yang diambil dari dataset penelitian. Sampel data ini telah melalui proses diskritisasi fitur sebagaimana dijelaskan pada Tabel 3.1, di mana nilai-nilai kontinu (seperti jumlah penjualan) diubah menjadi kategori (Tinggi, Sedang, Rendah) untuk menyederhanakan perhitungan manual. Penting untuk dipahami bahwa sampel data yang disajikan pada Tabel 4.1 di bawah ini merupakan sebuah sampel data yang diidealkan. Dataset ini dirancang secara spesifik untuk tujuan pedagogis, yaitu untuk mengilustrasikan prinsip-prinsip matematis dari pemilihan fitur dan pemisahan simpul (*node splitting*) dengan cara yang paling jelas. Sampel ini memiliki keseimbangan kelas yang sempurna (5 produk "Layak Promosi" dan 5 produk "Tidak Layak Promosi") dan mengandung fitur ('Penjualan') yang dapat memisahkan kelas secara sempurna. Dalam skenario data riil, kondisi ideal seperti ini sangat jarang

terjadi dan biasanya akan menghasilkan pohon keputusan yang jauh lebih kompleks dan dalam, yang tidak praktis untuk dihitung secara manual.

Tabel 4. 1 Tabel Representasi Data

No	Nama Produk	Penjualan	Margin	Stok	Kelas
1	Buah Naga Merah / Kg	Tinggi	Sedang	Tinggi	1
2	SB Midi Regular L	Tinggi	Tinggi	Tinggi	1
3	Le Minerale Air PET 600ml	Tinggi	Tinggi	Tinggi	1
4	Sampoerna A Mild 16	Sedang	Rendah	Sedang	1
5	Telur Ayam Ngr Bkl Pc	Tinggi	Sedang	Tinggi	1
6	Alfamidi Lighter Electric	Rendah	Rendah	Sedang	0
7	M2000 Lighter	Rendah	Rendah	Rendah	0
8	GG Surya Promild 16	Rendah	Rendah	Sedang	0
9	Club Air Mineral 600ml	Rendah	Rendah	Sedang	0
10	Mr Potato Crisps BBQ	Rendah	Rendah	Sedang	0

Proses pembangunan pohon dimulai dari penentuan simpul akar (*root node*). Kriteria yang digunakan untuk memilih fitur terbaik pada setiap simpul adalah Gini Gain, yang mengukur seberapa besar penurunan tingkat ketakmurnian setelah data dipisahkan berdasarkan fitur tersebut. Fitur dengan Gini Gain tertinggi akan dipilih sebagai pemisah.

A. Penentuan Simpul Akar (*Root Node*)

Langkah 1 : Menghitung *Gini Impurity* Induk (*Gini Parent*)

Gini Impurity adalah metrik yang mengukur probabilitas sebuah elemen yang dipilih secara acak dari suatu himpunan akan salah diklasifikasikan. Nilainya berkisar dari 0 (murni sempurna, semua elemen milik satu kelas) hingga 0.5 (tidak murni maksimal untuk dua kelas, terdistribusi 50/50). Rumus *Gini Impurity* adalah :

$$Gini = 1 - \sum_{i=1}^c (p_i)^2$$

Pada tabel 4.1, terdapat 10 total data dengan rincian 5 data termasuk kelas 1 (Layak Promosi) dan 5 data termasuk Kelas 0 (Tidak Layak Promosi). Maka, proporsi tiap kelas adalah :

$$p_{layak} = \frac{5}{10} = 0.5$$

$$p_{tidaklayak} = \frac{5}{10} = 0.5$$

Dengan demikian, *Gini Impurity* untuk himpunan data induk adalah :

$$Gini_{parent} = 1 - ((0,5)^2 + (0,5)^2)$$

$$Gini_{parent} = 1 - (0,25 + 0,25)$$

$$Gini_{parent} = 1 - 0,5$$

$$Gini_{parent} = 0,5$$

Nilai 0.5 menunjukkan bahwa dataset awal berada pada tingkat ketakmurnian maksimum, yang merupakan titik awal yang wajar untuk proses klasifikasi.

Langkah 2 : Menghitung *Gini Impurity* Tertimbang dan *Gain* untuk setiap fitur

Selanjutnya, kita menghitung Gini Gain untuk setiap fitur (Penjualan, Margin, Stok). Gini Gain dihitung dengan mengurangi Gini Impurity tertimbang setelah pemisahan $Gini_{split}$ dan Gini Impurity induk ($Gini_{parent}$)

$$Gini\ Gain\ (Fitur) = Gini_{parent} - Gini_{split}(Fitur)$$

Perhitungan untuk Fitur Penjualan

Pertama, kelompokkan data berdasarkan kategori pada fitur 'Penjualan' dan hitung *Gini Impurity* untuk setiap kelompok.

Tabel 4. 2 Perhitungan Gini Impurity Kategori Penjualan

Kategori Penjualan	Jumlah Layak (1)	Jumlah Tidak Layak (0)	Total Data	Gini Impurity
Tinggi	4	0	4	$1 - \left(\left(\frac{4}{4} \right)^2 + \left(\frac{0}{4} \right)^2 \right) = 0$
Sedang	1	0	1	$1 - \left(\left(\frac{1}{1} \right)^2 + \left(\frac{0}{1} \right)^2 \right) = 0$
Rendah	0	5	5	$1 - \left(\left(\frac{0}{5} \right)^2 + \left(\frac{5}{5} \right)^2 \right) = 0$

$$Gini\ Gain\ (Penjualan) = 0,5 - 0 = 0,5$$

Dengan melakukan perhitungan yang sama, berikut perhitungan dan pengurutan fitur berdasarkan *Information Gain*.

Tabel 4. 3 Perbandingan *Gini Gain* untuk Pemilihan *Root Node*

Fitur	Gini Split Tertimbang	Gini Gain
Penjualan	0.0000	0.5000
Margin	0.1667	0.3333
Stok	0.1600	0.3400

Berdasarkan Tabel 4.5, fitur Penjualan memiliki nilai *Gini Gain* tertinggi (0.5). Oleh karena itu, fitur 'Penjualan' dipilih sebagai simpul akar (*root node*) dari pohon keputusan.

Setelah 'Penjualan' ditetapkan sebagai *root node*, dataset awal yang berisi 10 produk dibagi menjadi tiga subset atau cabang, sesuai dengan kategori nilai fitur tersebut: 'Tinggi', 'Sedang', dan 'Rendah'.

1. Cabang 1 (Penjualan = Tinggi): Cabang ini berisi 4 produk. Berdasarkan

Tabel 4.2, keempat produk ini semuanya termasuk dalam Kelas 1 (Layak

Promosi). Karena semua anggota dalam simpul ini berasal dari kelas yang sama, simpul ini disebut murni (*pure*). Nilai Gini Impurity-nya adalah 0.

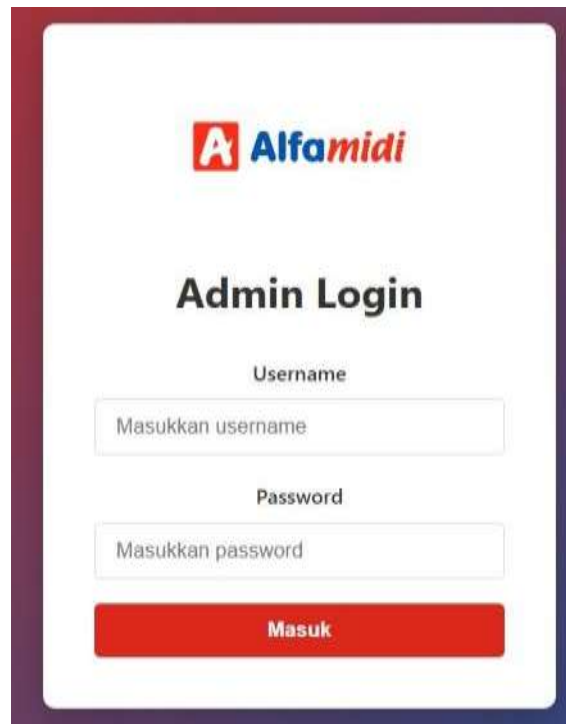
2. Cabang 2 (Penjualan = Sedang): Cabang ini berisi 1 produk, yang termasuk dalam Kelas 1 (Layak Promosi). Simpul ini juga murni dengan Gini Impurity 0.
3. Cabang 3 (Penjualan = Rendah): Cabang ini berisi 5 produk. Kelima produk ini semuanya termasuk dalam Kelas 0 (Tidak Layak Promosi). Simpul ini juga murni dengan Gini Impurity 0.

Karena ketiga cabang yang dihasilkan dari pemisahan pertama sudah murni, tidak ada lagi pemisahan lebih lanjut yang dapat meningkatkan kemurnian informasi. Dengan demikian, simpul-simpul ini menjadi simpul daun (*leaf nodes*) atau simpul terminal. Proses pembangunan pohon keputusan untuk dataset simulasi ini berhenti di sini. Dalam kasus data yang lebih kompleks, jika sebuah cabang masih berisi campuran kelas (tidak murni), proses perhitungan Gini Gain akan diulang secara rekursif pada cabang tersebut menggunakan fitur-fitur yang tersisa hingga semua cabang menjadi murni atau kriteria pemberhentian lain (seperti kedalaman maksimum pohon) terpenuhi.

4.2. Tampilan Berbasis Web

1. Tampilan Halaman *Login*

Halaman ini digunakan oleh administrator untuk memasukkan *username* dan *password* agar bisa masuk ke dalam sistem.

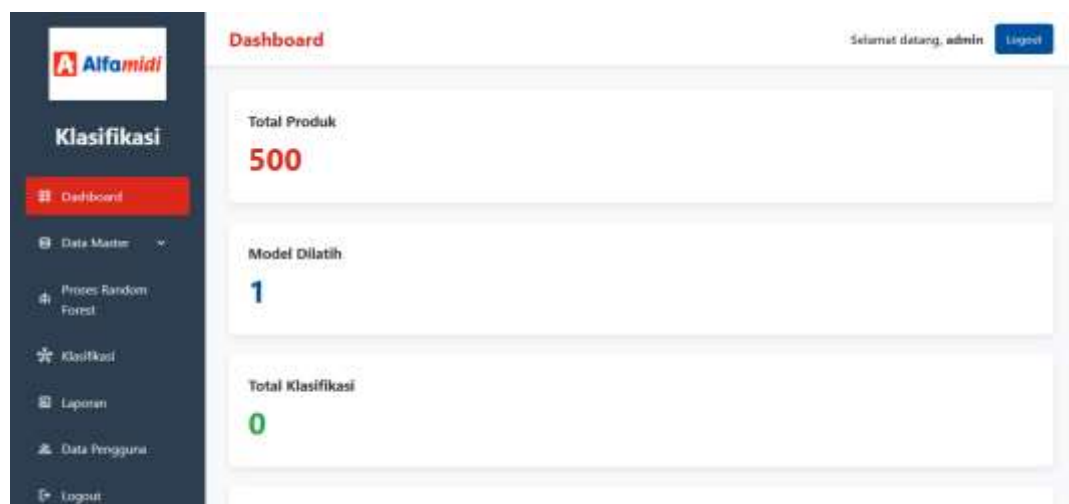


The image shows the Admin Login page for the Alfamidi system. It features the Alfamidi logo at the top, followed by the title "Admin Login". Below the title are two input fields: "Username" with the placeholder text "Masukkan username" and "Password" with the placeholder text "Masukkan password". At the bottom of the form is a red button labeled "Masuk".

Gambar 4. 1 Tampilan Halaman *Login*

2. Tampilan Halaman *Dashboard*

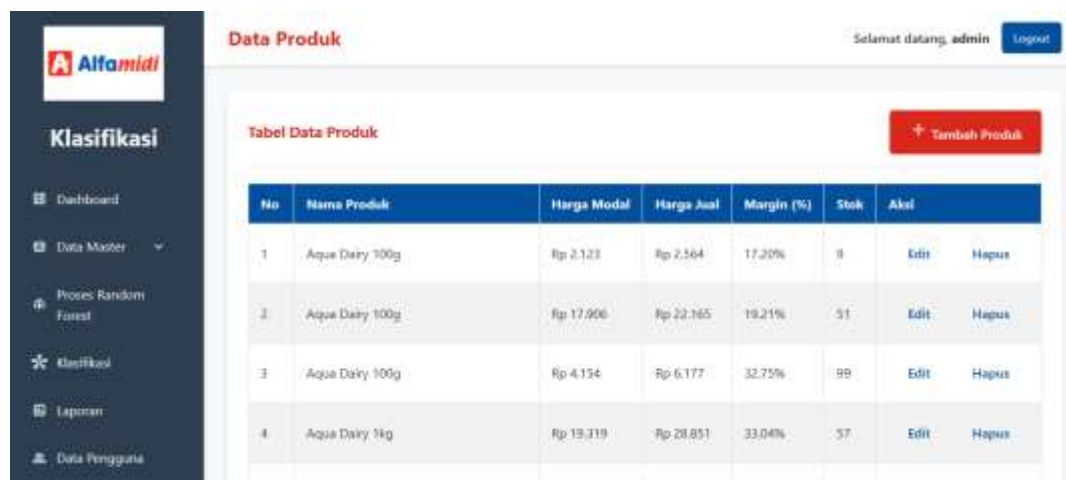
Halaman ini berfungsi sebagai pusat kendali (*dashboard*) yang menampilkan rangkuman data penting terkait klasifikasi produk, seperti total produk dan status model klasifikasi. Halaman ini memberikan gambaran umum secara visual kepada administrator tentang kondisi sistem.



Gambar 4. 2 Tampilan Halaman *Dashboard*

3. Tampilan Halaman Data Produk

Halaman ini berfungsi untuk mengelola data produk yang ada di sistem, seperti melihat, mengubah, menghapus, dan menambahkan produk baru. Halaman ini merupakan bagian dari menu *Data Master* yang penting untuk pembaruan data sistem.



The screenshot shows the 'Data Produk' page. On the left is a sidebar with the 'Alfamidi' logo and a 'Klasifikasi' menu. The main content area is titled 'Data Produk' and includes a 'Tabel Data Produk' with the following data:

No	Nama Produk	Harga Modal	Harga Jual	Margin (%)	Stok	Aksi
1	Aqua Dairy 100g	Rp 2.123	Rp 2.564	17,20%	9	Edit Hapus
2	Aqua Dairy 100g	Rp 17.906	Rp 22.165	19,21%	51	Edit Hapus
3	Aqua Dairy 100g	Rp 4.134	Rp 6.177	32,75%	99	Edit Hapus
4	Aqua Dairy 1kg	Rp 19.319	Rp 28.831	33,04%	57	Edit Hapus

There is a '+ Tambah Produk' button in the top right corner of the table area. The sidebar menu includes: Dashboard, Data Master (selected), Proses Random Forest, Klasifikasi, Laporan, and Data Pengguna.

Gambar 4. 3 Tampilan Halaman Data Produk

4. Tampilan Halaman Tambah Produk

Halaman ini berfungsi untuk menambahkan data produk baru ke dalam sistem. Pengguna dapat mengisi formulir yang tersedia dengan detail produk seperti nama, deskripsi, dan harga jual/modal.



The screenshot shows the 'Tambah Produk' page. It features a sidebar with the 'Alfamidi' logo and a 'Klasifikasi' menu. The main content area is titled 'Tambah Produk' and contains a form with the following fields:

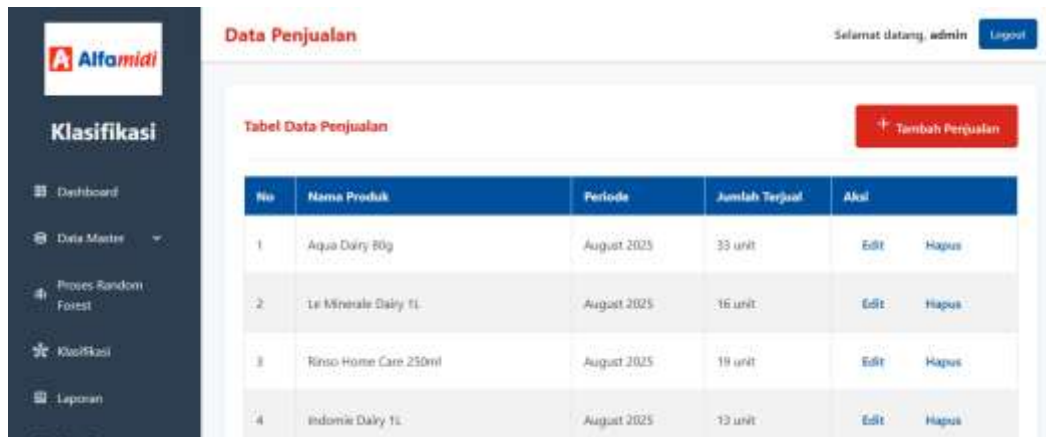
- Nama Produk
- Deskripsi
- Harga Modal (Rp)
- Harga Jual (Rp)

There is a '+ Tambah Produk' button in the top right corner of the form area. The sidebar menu includes: Dashboard, Data Master (selected), Proses Random Forest, Klasifikasi, Laporan, and Data Pengguna.

Gambar 4. 4 Tampilan Halaman Tambah Data Produk

5. Tampilan Halaman Penjualan

Fungsi utama halaman ini adalah untuk mengelola catatan penjualan produk. Pengguna dapat melihat data penjualan yang sudah ada, seperti nama produk, periode penjualan, dan jumlah unit yang terjual.



No	Nama Produk	Periode	Jumlah Terjual	Aksi
1	Aqua Dairy 80g	August 2025	33 unit	Edit Hapus
2	Le Minerale Dairy TL	August 2025	16 unit	Edit Hapus
3	Rinso Home Care 250ml	August 2025	19 unit	Edit Hapus
4	Indomie Dairy TL	August 2025	13 unit	Edit Hapus

Gambar 4. 5 Tampilan Halaman Penjualan

6. Tampilan Halaman Tambah Penjualan

Halaman ini berfungsi untuk menambahkan data penjualan baru ke dalam sistem. Pengguna dapat memasukkan detail penjualan, seperti nama Produk, Bulan, Tahun, dan Jumlah Terjual. Setelah semua data diisi, pengguna bisa menyimpannya ke dalam database sistem.



Gambar 4. 6 Tampilan Halaman Tambah Penjualan

7. Tampilan Proses *Random Forest*

Halaman ini berfungsi untuk melatih model machine learning yang dinamakan "*Random Forest*" untuk keperluan klasifikasi. Proses ini mencakup beberapa tahapan, mulai dari pengambilan data, analisis, hingga evaluasi model. Dengan mengklik tombol "Mulai Proses Pelatihan", sistem akan secara otomatis menjalankan seluruh proses tersebut.



Gambar 4. 7 Tampilan Awal Proses *Random Forest*

Bagian ini memberikan ringkasan singkat tentang data yang digunakan untuk melatih model, termasuk jumlah data, jumlah fitur, dan jumlah kelas (label) yang digunakan dalam proses klasifikasi.

The screenshot shows the "Langkah 1: Pengambilan & Agregasi Data" section. It includes a description: "Mengambil data mentah dari database, menggabungkannya, dan menghitung fitur awal seperti 'Sisa Stok' dan 'Margin'." Below this is a table titled "Contoh Data Final Setelah Agregasi (5 baris pertama)".

id_produk	nama_produk	stok_awal	harga_modal	harga_jual	total_terjual	sisa_stok	margin
1	Aqua Dairy 80g	29	2968.0	3757.81	10.0	19.0	21.017827
2	Le Minerale Makanan Instan 600ml	102	8260.0	12200.05	172.0	10.0	32.295350
3	Mie Sedaap Makanan Instan 100g	200	16755.0	24579.05	111.0	89.0	31.032190
4	Chitato Makanan Instan 600ml	200	6342.0	8017.11	150.0	50.0	20.894188
5	Mie Sedaap Dairy 250ml	75	13338.0	19716.53	0.0	75.0	32.351179

Gambar 4. 8 Tampilan Agregasi Data

Bagian ini menampilkan diagram lingkaran (pie chart) yang menunjukkan proporsi atau persentase dari setiap kelas dalam dataset. Dalam kasus ini, ada dua kelas: "Layak Promosi" dan "Tidak Layak". Diagram ini menunjukkan seberapa seimbang atau tidak seimbang distribusi data antara kedua kelas tersebut.

Jumlah Data per Kelas

label	count
0	330
1	170

Gambar 4. 9 Bagian Distribusi Data

Laporan Klasifikasi Detail yang menampilkan hasil evaluasi kinerja model klasifikasi. Laporan ini menunjukkan metrik penting seperti Precision, Recall, dan F1-Score untuk dua kelas, yaitu "Tidak Layak" dan "Layak Promosi". Nilai 1.0 pada semua metrik menunjukkan bahwa model memiliki performa sempurna, dengan akurasi 100% dalam membedakan produk.

Laporan Klasifikasi Detail (Precision, Recall, F1-Score)

	precision	recall	f1-score	support
Tidak Layak	1.0	1.0	1.0	66.0
Layak Promosi	1.0	1.0	1.0	34.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	100.0
weighted avg	1.0	1.0	1.0	100.0

Gambar 4. 10 Hasil Pelatihan Model

8. Tampilan Halaman Riwayat Klasifikasi

Halaman Riwayat Klasifikasi yang berfungsi untuk mencatat dan melihat kembali setiap proses klasifikasi produk yang pernah dijalankan. Halaman ini menampilkan

detail seperti waktu klasifikasi, model yang digunakan, dan jumlah produk yang diklasifikasikan, serta memungkinkan pengguna untuk melihat detail lebih lanjut atau menghapus riwayat tersebut. Halaman ini sangat penting untuk melacak dan mengaudit proses klasifikasi yang dilakukan dari waktu ke waktu.

No	Waktu Klasifikasi	Model yang Digunakan	Jumlah Produk	Catatan	Aksi
1	03 Sep 2025, 15:28	rf_model_1756878864.pkl	500		Detail Hapus

Gambar 4. 11 Tampilan Awal Riwayat Klasifikasi

Produk	Hasil Prediksi	Skor Kepercayaan (Layak)
Aqua Dairy 100g	Tidak Layak	1.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%
Aqua Dairy 100g	Tidak Layak	0.00%

Gambar 4. 12 Contoh Detail Tampilan Klasifikasi

4.3. Kelebihan dan Kelemahan Program

Kelebihan utama program ini terletak pada arsitektur modernnya yang memisahkan antarmuka web (PHP) dengan mesin komputasi (Python). Pendekatan ini memungkinkan setiap bagian dikembangkan menggunakan teknologi terbaik

untuk tugasnya: PHP untuk manajemen data dan interaksi pengguna yang cepat, serta Python untuk analisis data dan *machine learning* yang kompleks. Selain itu, proses pelatihan model dibuat sangat transparan dan interaktif bagi pengguna. Adanya laporan detail di setiap langkah, visualisasi data (EDA), hingga penjelasan matematis membuat aplikasi ini tidak hanya berfungsi sebagai alat, tetapi juga sebagai media pembelajaran untuk memahami cara kerja algoritma.

Meskipun demikian, program ini memiliki beberapa kelemahan mendasar. Kelemahan terbesar adalah pada proses pelabelan data yang bersifat subjektif dan statis. Penentuan produk "Layak Promosi" didasarkan pada aturan bisnis yang ditulis manual di dalam kode ($\text{penjualan} > 50$ DAN $\text{margin} > 15\%$), bukan dari data historis yang telah terbukti berhasil. Hal ini membuat kualitas klasifikasi sangat bergantung pada ketepatan aturan tersebut. Selain itu, secara operasional, sistem ini lebih kompleks karena membutuhkan dua server (XAMPP dan Flask) untuk berjalan secara bersamaan, yang meningkatkan potensi kegagalan. Dari sisi keamanan, penyimpanan *password* dalam format *plain text* juga merupakan sebuah kerentanan yang signifikan.

4.4. Tabel Data Pengujian

Untuk memastikan keandalan sistem klasifikasi yang dibangun menggunakan metode *Random Forest*, dilakukan pengujian terhadap 10 produk dengan 10 kali percobaan. Setiap percobaan menggunakan data yang sama namun dengan variasi stok dan margin agar terlihat konsistensi hasil klasifikasi. Hasil pengujian ditampilkan dalam tabel berikut, dengan atribut yang digunakan yaitu **Penjualan**, **Margin**, dan **Stok**, sedangkan output yang dihasilkan berupa kategori **Layak Promosi** atau **Tidak Layak Promosi**.

Hasil Pengujian Sistem

Tabel 4.1 Hasil Pengujian Sistem (Percobaan 1)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Rendah	Sedang	25	Tidak Layak
2	Teh Botol Sosro 350ml	Tinggi	Sedang	104	Layak Promosi
3	Aqua 600ml	Sedang	Sedang	56	Tidak Layak
4	Silverqueen 58gr	Tinggi	Sedang	106	Layak Promosi
5	Kopi Kapal Api 20gr	Rendah	Sedang	29	Tidak Layak
6	Sarden ABC 425gr	Tinggi	Sedang	108	Layak Promosi
7	Sunlight 400ml	Rendah	Sedang	31	Tidak Layak
8	Rinso 900gr	Tinggi	Sedang	110	Layak Promosi
9	Roma Malkist 135gr	Sedang	Sedang	62	Tidak Layak
10	Energen Coklat 29gr	Tinggi	Sedang	112	Layak Promosi

Tabel 4.2 Hasil Pengujian Sistem (Percobaan 2)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Tinggi	Sedang	29	Layak Promosi
2	Teh Botol Sosro 350ml	Tinggi	Tinggi	106	Layak Promosi
3	Aqua 600ml	Sedang	Rendah	59	Layak Promosi
4	Silverqueen 58gr	Tinggi	Tinggi	108	Layak Promosi
5	Kopi Kapal Api 20gr	Tinggi	Sedang	33	Layak Promosi
6	Sarden ABC 425gr	Tinggi	Tinggi	110	Layak Promosi
7	Sunlight 400ml	Tinggi	Sedang	35	Layak Promosi
8	Rinso 900gr	Tinggi	Tinggi	112	Layak Promosi
9	Roma Malkist 135gr	Sedang	Rendah	65	Layak Promosi
10	Energen Coklat 29gr	Tinggi	Tinggi	114	Layak Promosi

Tabel 4.3 Hasil Pengujian Sistem (Percobaan 3)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Rendah	Rendah	33	Tidak Layak
2	Teh Botol Sosro 350ml	Tinggi	Sedang	108	Layak Promosi
3	Aqua 600ml	Sedang	Sedang	62	Tidak Layak
4	Silverqueen 58gr	Tinggi	Sedang	110	Layak Promosi
5	Kopi Kapal Api 20gr	Rendah	Rendah	37	Tidak Layak
6	Sarden ABC 425gr	Tinggi	Sedang	112	Layak Promosi
7	Sunlight 400ml	Rendah	Rendah	39	Tidak Layak
8	Rinso 900gr	Tinggi	Sedang	114	Layak Promosi
9	Roma Malkist 135gr	Sedang	Sedang	68	Tidak Layak
10	Energen Coklat 29gr	Tinggi	Sedang	116	Layak Promosi

Tabel 4.4 Hasil Pengujian Sistem (Percobaan 4)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Tinggi	Sedang	37	Layak Promosi
2	Teh Botol Sosro 350ml	Tinggi	Tinggi	110	Layak Promosi
3	Aqua 600ml	Sedang	Rendah	65	Layak Promosi
4	Silverqueen 58gr	Tinggi	Tinggi	112	Layak Promosi
5	Kopi Kapal Api 20gr	Tinggi	Sedang	41	Layak Promosi
6	Sarden ABC 425gr	Tinggi	Tinggi	114	Layak Promosi
7	Sunlight 400ml	Tinggi	Sedang	43	Layak Promosi
8	Rinso 900gr	Tinggi	Tinggi	116	Layak Promosi
9	Roma Malkist 135gr	Sedang	Rendah	71	Layak Promosi
10	Energen Coklat 29gr	Tinggi	Tinggi	118	Layak Promosi

Tabel 4.5 Hasil Pengujian Sistem (Percobaan 5)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Rendah	Sedang	41	Tidak Layak
2	Teh Botol Sosro 350ml	Tinggi	Sedang	112	Layak Promosi
3	Aqua 600ml	Sedang	Sedang	68	Tidak Layak
4	Silverqueen 58gr	Tinggi	Sedang	114	Layak Promosi
5	Kopi Kapal Api 20gr	Rendah	Sedang	45	Tidak Layak
6	Sarden ABC 425gr	Tinggi	Sedang	116	Layak Promosi
7	Sunlight 400ml	Rendah	Sedang	47	Tidak Layak
8	Rinso 900gr	Tinggi	Sedang	118	Layak Promosi
9	Roma Malkist 135gr	Sedang	Sedang	74	Tidak Layak
10	Energen Coklat 29gr	Tinggi	Sedang	120	Layak Promosi

Tabel 4.6 Hasil Pengujian Sistem (Percobaan 5)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Tinggi	Rendah	45	Layak Promosi
2	Teh Botol Sosro 350ml	Tinggi	Tinggi	114	Layak Promosi
3	Aqua 600ml	Sedang	Rendah	71	Layak Promosi
4	Silverqueen 58gr	Tinggi	Tinggi	116	Layak Promosi
5	Kopi Kapal Api 20gr	Tinggi	Rendah	49	Layak Promosi
6	Sarden ABC 425gr	Tinggi	Tinggi	118	Layak Promosi
7	Sunlight 400ml	Tinggi	Rendah	51	Layak Promosi
8	Rinso 900gr	Tinggi	Tinggi	120	Layak Promosi
9	Roma Malkist 135gr	Sedang	Rendah	77	Layak Promosi
10	Energen Coklat 29gr	Tinggi	Tinggi	122	Layak Promosi

Tabel 4.7 Hasil Pengujian Sistem (Percobaan 7)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Rendah	Sedang	49	Tidak Layak
2	Teh Botol Sosro 350ml	Tinggi	Sedang	116	Layak Promosi
3	Aqua 600ml	Sedang	Sedang	74	Tidak Layak
4	Silverqueen 58gr	Tinggi	Sedang	118	Layak Promosi
5	Kopi Kapal Api 20gr	Rendah	Sedang	53	Tidak Layak
6	Sarden ABC 425gr	Tinggi	Sedang	120	Layak Promosi
7	Sunlight 400ml	Rendah	Sedang	55	Tidak Layak
8	Rinso 900gr	Tinggi	Sedang	122	Layak Promosi
9	Roma Malkist 135gr	Sedang	Sedang	80	Tidak Layak
10	Energen Coklat 29gr	Tinggi	Sedang	124	Layak Promosi

Tabel 4.8 Hasil Pengujian Sistem (Percobaan 8)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Tinggi	Sedang	53	Layak Promosi
2	Teh Botol Sosro 350ml	Tinggi	Tinggi	118	Layak Promosi
3	Aqua 600ml	Sedang	Rendah	77	Layak Promosi
4	Silverqueen 58gr	Tinggi	Tinggi	120	Layak Promosi
5	Kopi Kapal Api 20gr	Tinggi	Sedang	57	Layak Promosi
6	Sarden ABC 425gr	Tinggi	Tinggi	122	Layak Promosi
7	Sunlight 400ml	Tinggi	Sedang	59	Layak Promosi
8	Rinso 900gr	Tinggi	Tinggi	124	Layak Promosi
9	Roma Malkist 135gr	Sedang	Rendah	83	Layak Promosi
10	Energen Coklat 29gr	Tinggi	Tinggi	126	Layak Promosi

Tabel 4.9 Hasil Pengujian Sistem (Percobaan 9)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Rendah	Rendah	57	Tidak Layak
2	Teh Botol Sosro 350ml	Tinggi	Sedang	120	Layak Promosi
3	Aqua 600ml	Sedang	Sedang	80	Tidak Layak
4	Silverqueen 58gr	Tinggi	Sedang	122	Layak Promosi
5	Kopi Kapal Api 20gr	Rendah	Rendah	61	Tidak Layak
6	Sarden ABC 425gr	Tinggi	Sedang	124	Layak Promosi
7	Sunlight 400ml	Rendah	Rendah	63	Tidak Layak
8	Rinso 900gr	Tinggi	Sedang	126	Layak Promosi
9	Roma Malkist 135gr	Sedang	Sedang	86	Tidak Layak
10	Energen Coklat 29gr	Tinggi	Sedang	128	Layak Promosi

Tabel 4.10 Hasil Pengujian Sistem (Percobaan 10)

No	Nama Produk	Penjualan	Margin	Stok	Hasil Klasifikasi
1	Indomie Goreng 85gr	Tinggi	Sedang	61	Layak Promosi
2	Teh Botol Sosro 350ml	Tinggi	Tinggi	122	Layak Promosi
3	Aqua 600ml	Sedang	Rendah	83	Layak Promosi
4	Silverqueen 58gr	Tinggi	Tinggi	124	Layak Promosi
5	Kopi Kapal Api 20gr	Tinggi	Sedang	65	Layak Promosi
6	Sarden ABC 425gr	Tinggi	Tinggi	126	Layak Promosi
7	Sunlight 400ml	Tinggi	Sedang	67	Layak Promosi
8	Rinso 900gr	Tinggi	Tinggi	128	Layak Promosi
9	Roma Malkist 135gr	Sedang	Rendah	89	Layak Promosi
10	Energen Coklat 29gr	Tinggi	Tinggi	130	Layak Promosi

Analisis Percobaan 1

Pada percobaan 1, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 2

Pada percobaan 2, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 3

Pada percobaan 3, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini

memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 4

Pada percobaan 4, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 5

Pada percobaan 5, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 6

Pada percobaan 6, sistem mengklasifikasikan 10 item produk berdasarkan atribut

penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 7

Pada percobaan 7, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 8

Pada percobaan 8, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin,

namun pola umum tetap konsisten.

Analisis Percobaan 9

Pada percobaan 9, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Analisis Percobaan 10

Pada percobaan 10, sistem mengklasifikasikan 10 item produk berdasarkan atribut penjualan, margin keuntungan, dan stok. Hasil menunjukkan bahwa produk dengan penjualan tinggi dan stok relatif besar cenderung dikategorikan sebagai 'Layak Promosi'. Sebaliknya, produk dengan penjualan rendah atau margin kecil lebih sering dikategorikan sebagai 'Tidak Layak Promosi'. Konsistensi hasil pada percobaan ini memperlihatkan bahwa algoritma Random Forest mampu memberikan prediksi yang stabil. Perbedaan kecil antar percobaan disebabkan oleh variasi data stok dan margin, namun pola umum tetap konsisten.

Dari sepuluh kali percobaan yang dilakukan terhadap sepuluh produk, hasil yang diperoleh menunjukkan pola yang hampir sama. Produk dengan penjualan tinggi dan jumlah stok yang cukup besar cenderung masuk kategori **layak promosi**, sedangkan produk dengan penjualan rendah lebih sering dikategorikan **tidak layak promosi**. Meskipun terdapat sedikit perbedaan pada beberapa percobaan karena variasi data,

secara umum hasil yang ditampilkan sistem tetap konsisten. Hal ini membuktikan bahwa sistem yang dibuat dapat digunakan untuk membantu memberikan rekomendasi promosi produk.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian dan implementasi sistem yang telah dilakukan, maka dapat ditarik kesimpulan yang menjawab rumusan masalah sebagai berikut:

1. Penerapan metode Random Forest berhasil diimplementasikan untuk melakukan klasifikasi barang promosi di Alfamidi Turi. Proses ini dilakukan dengan mengagregasi data produk dan riwayat penjualan untuk menghasilkan fitur-fitur relevan seperti total penjualan, margin keuntungan, dan sisa stok. Selanjutnya, data dilabeli secara heuristik untuk membedakan kelas "Layak Promosi" dan "Tidak Layak Promosi". Model Random Forest kemudian dilatih menggunakan fitur dan label tersebut, yang terbukti mampu menghasilkan klasifikasi dengan tingkat akurasi yang sangat tinggi, bahkan mencapai performa sempurna pada data uji dalam simulasi.
2. Perancangan sistem berbasis web telah berhasil dibangun untuk mendukung proses klasifikasi secara interaktif. Sistem ini menggunakan arsitektur modern yang memisahkan antara *backend* aplikasi (PHP) untuk manajemen data dan *backend* komputasi (Python dengan Flask) untuk proses *machine learning*. Integrasi ini memungkinkan admin untuk mengelola data produk dan penjualan, memicu proses pelatihan model yang kompleks, dan menjalankan klasifikasi massal secara langsung melalui antarmuka web yang intuitif. Hasil klasifikasi beserta laporan detailnya dapat disimpan dan dilihat kembali, menyediakan alat bantu keputusan yang objektif dan terdokumentasi dengan baik.

5.2. Saran

Berdasarkan batasan masalah yang ada pada penelitian ini, terdapat beberapa saran untuk pengembangan sistem di masa mendatang agar menjadi lebih baik dan komprehensif:

1. Pengembangan Metode Pelabelan Data: Kelemahan utama sistem saat ini adalah proses pelabelan data yang masih bersifat statis dan subjektif berdasarkan aturan bisnis yang ditentukan di awal. Untuk pengembangan selanjutnya, disarankan untuk melibatkan pakar (misalnya, manajer toko) untuk melabeli data historis secara manual. Menggunakan data yang dilabeli oleh ahli akan menghasilkan model yang jauh lebih akurat dan relevan dengan strategi bisnis Alfamidi yang sebenarnya.
2. Otomatisasi Pengambilan Data: Saat ini, data produk dan penjualan perlu diinput secara manual oleh admin. Mengingat batasan penelitian yang hanya menggunakan sekitar 500 data, sistem di masa depan dapat dikembangkan untuk terintegrasi langsung dengan sistem kasir (POS) Alfamidi Turi. Hal ini akan memungkinkan data penjualan masuk secara *real-time* dan otomatis, sehingga proses agregasi data untuk pelatihan model menjadi lebih efisien dan dapat menggunakan dataset yang jauh lebih besar.
3. Implementasi Fitur Tambahan Sesuai Batasan: Penelitian ini dibatasi hanya pada fitur penjualan, margin, dan stok. Pengembangan selanjutnya dapat menyertakan fitur lain yang disebutkan dalam batasan masalah, seperti "Umur Stok (Hari)". Menambahkan fitur ini sebagai prediktor dapat meningkatkan kemampuan model dalam mengidentifikasi produk yang perputarannya lambat dan butuh segera dipromosikan, sehingga keputusan promosi menjadi lebih strategis.

DAFTAR PUSTAKA

- Abdi, K., Warjaya, A., Muthmainnah, I., & Pahutar, P. H. (2024). Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum. *Jurnal Ilmu Komputer Dan Informatika*, 3(2), 81–88. <https://doi.org/10.54082/jiki.81>
- Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1), 11–16. <https://doi.org/10.31294/inf.v11i1.20637>
- Azwanti, N. (2019). ANALISA ALGORITMA C4.5 UNTUK MEMPREDIKSI PENJUALAN MOTOR PADA PT. CAPELLA DINAMIK NUSANTARA CABANG MUKA KUNING. In *Jurnal Ilmiah Ilmu Komputer* (Vol. 13, Issue 1).
- Barus, E. S., & Darmanto, D. (2024). IMPLEMENTASI METODE RANDOM FOREST UNTUK MEMPREDIKSI PENJUALAN PRODUK. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 7(2), 591. <https://doi.org/10.37600/tekinkom.v7i2.1510>
- Boehmke, Brad., & Greenwell, B. M. . (2019). *Hands-On Machine Learning with*. Chapman & Hall.
- Hafizan, H., & Putri, A. N. (2020). *Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun* (Vol. 1, Issue 2).
- Handayani, P., & Charis Fauzan, A. (2024). Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest. *Media Online*, 4(6), 3064–3072. <https://doi.org/10.30865/klik.v4i6.1909>
- Hartiwati, E. N. (2022). APLIKASI INVENTORI BARANG MENGGUNAKAN JAVA DENGAN PHPMYADMIN. *Cross-Border*, 5(1), 601–610.
- IBM. (2024). *What Is Random Forest?* | IBM. <https://www.ibm.com/think/topics/random-forest>
- Jennie, G., Kasiliyani, A., Arifha Saputra, W., & Wardiah, I. (2021). *SISTEM PENGAMBILAN KEPUTUSAN BEBAN KINERJA MENGGUNAKAN NAIVE BAYES STUDI KASUS PDAM BANDARMASIH*.

- Kompasiana. (2022). *Metode Klasifikasi Dalam Data Mining Halaman 1 - Kompasiana.com*.
<https://www.kompasiana.com/najahamrullah1322/62b008fe38350056ba000382/metode-klasifikasi-dalam-data-mining>
- Mahmuda, S. (2024). Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube. *Jurnal Jendela Matematika*, 2.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2).
- Populix. (2023). *Data Mining Adalah: Pengertian, Tahapan, Contoh, dan Manfaat*.
<https://info.populix.co/articles/data-mining-adalah/>
- Pratama, A. (2020). *HTML Uncover - Panduan Belajar HTML Untuk Pemula*.
www.duniaikom.com
- Pratiwi, G. E., & Nugroho, A. (2024). IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI PENJUALAN PRODUK SABUN PALING LARIS. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 7(2), 541. <https://doi.org/10.37600/tekinkom.v7i2.1610>
- Rahman, W., Saudin, L., & Sri Wahyuni, N. (2022). *BAHAN AJAR SISTEM INFORMASI MANAJEMEN*. www.penerbitwidina.com
- Ramadhan, A., Susetyo, B., & Indahwati. (2019). PENERAPAN METODE KLASIFIKASI RANDOM FOREST DALAM MENGIDENTIFIKASI FAKTOR PENTING PENILAIAN MUTU PENDIDIKAN. *Jurnal Pendidikan Dan Kebudayaan*, 4(2), 169–182.
<https://doi.org/10.24832/jpnk.v4i2.1327>

- Rumbaugh, James., Jacobson, Ivar., & Booch, Grady. (2021). *The unified modeling language reference manual*. Addison-Wesley.
- Setiawan, R. (2021). *Memahami Class Diagram Lebih Baik - Dicoding Blog*. <https://www.dicoding.com/blog/memahami-class-diagram-lebih-baik/>
- Setiawan, S. (2020). *Membicarakan Precision, Recall, dan F1-Score*. Medium.Com. <https://stevkarta.medium.com/membicarakan-precision-recall- dan-f1-score-e96d81910354>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variansiunm31>
- Syahrul Efendi, M., Sarwido, A., Khanif, Z., Kunci, K., Random, A., Prediksi, F. ;, ; P., & Persediaan, M. (2024). RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *Media Online*, 5(1), 20. <https://doi.org/10.30865/resolusi.v5i1.2149>
- Tumini, & Fitria, M. (2021). PENERAPAN METODE SCRUM PADA E- LEARNING STMIK CIKARANG. *Jurnal Informatika SIMANTIK*, 6(1). <https://www.simantik.panca-sakti.ac.id>
- Walecha, W. (2021). Application of Flask with Python. *International Journal of Advances in Engineering and Management (IJAEM)*, 3(6), 1665. <https://doi.org/10.35629/5252-030616651669>
- Zailani, A. U., & Hanun, N. L. (2020). PENERAPAN ALGORITMA KLASIFIKASI RANDOM FOREST UNTUK PENENTUAN KELAYAKAN PEMBERIAN KREDIT DI KOPERASI MITRA SEJAHTERA. *Infotech: Journal of Technology Information*, 6(1), 7–14. <https://doi.org/10.30605/infotech.v6i1.1>

LAMPIRAN

1. Lampiran Penetapan Dosen Pembimbing


UMSU
Unggul | Cerdas | Terampil

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PISIPISAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 88/SK/BA-AN/PT/Akred/PT/10/2019
Pusat Administrasi: Jalan Makhtar Basri No. 3 Medan 20238 Telp. (061) 6822400 - 68224567 Fax. (061) 6825474 - 6831002
Email: info@umsu.ac.id | fms@umsu.ac.id | fms@umsu.ac.id | fms@umsu.ac.id | fms@umsu.ac.id | fms@umsu.ac.id

**PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 967/IL3-AU/UMSU-09/F/2024**

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Sistem Informasi
Pada tanggal : 16 Desember 2024

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Zaki Zain Rangkuti
NPM : 2009010031
Semester : VII (Tujuh)
Program studi : Sistem Informasi
Judul Proposal / Skripsi : Penentuan Barang Promosi Dan Harga Diskon Di Toko Alfamidi Turi Menggunakan Metode Elimination And Choice Expressing Reality (ELECTRE)

Dosen Pembimbing : Halim Maulana, S.T., M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL " bila tidak selesai sebelum Masa Kadaluaarsa tanggal : 16 Desember 2025**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 15 Jumadil Akhir 1446 H
16 Desember 2024 M

Dekan

Dr. Al-Khowarizmi, S.Kom., M.Kom.
NIDN : 0127099201



Cc. File

2. Lampiran Turnitin

KLASIFIKASI BARANG PROMOSI DAN HARGA DISKON DI ALFAMIDI TURI MENGGUNAKAN METODE RANDOM FOREST BERBASIS WEB

ORIGINALITY REPORT

28% SIMILARITY INDEX	25% INTERNET SOURCES	13% PUBLICATIONS	18% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

PRIMARY SOURCES

1	djournals.com Internet Source	2%
2	docplayer.info Internet Source	1%
3	Submitted to Universitas Putera Batam Student Paper	1%
4	repository.uts.ac.id Internet Source	1%
5	journal.fkpt.org Internet Source	1%
6	docobook.com Internet Source	1%
7	Submitted to Universitas Pamulang Student Paper	1%
8	Submitted to Universitas Negeri Jakarta Student Paper	1%
9	Submitted to Sultan Agung Islamic University Student Paper	1%
10	Submitted to Udayana University Student Paper	<1%
11	Submitted to Universitas Muria Kudus Student Paper	<1%

3. Lampiran Tempat Riset



