



Analisis Pola Pembelian Mobil di Indonesia Menggunakan Metode ID3

Maria Yohana Gabriela Sasi¹, Yampi R Kaesmetan²

^{1*,2}Sistem Informasi, STIKOM Uyelindo Kupang, Kupang, Indonesia

Email: ¹sasiyolan@gmail.com, ²kaesmetanyampi@gmail.com

Informasi Artikel

Diterima : 22-04-2025

Disetujui : 11-05-2025

Diterbitkan : 20-05-2025

ABSTRACT

Indonesia's geographical conditions consisting of islands encourage the need to develop efficient transportation, especially automotive, to improve the country's economy. Cars are seen as safer vehicles than motorbikes, and with the increasing number of automotive companies, competition in the market is getting tighter. Car sales, especially from brands such as Toyota, Honda, and Mitsubishi, have shown rapid growth, with Toyota recording sales growth of 6.5% and Honda increasing by 56% in 2023. This condition requires companies to understand consumer needs and implement the right business strategies. The comfort and quality of the car are the main priorities in choosing a vehicle, although price is also an important consideration. With the growth of marketing through social media, the expected data can be used for further analysis, so that manufacturers can improve the quality of their products. The purpose of using the ID3 Method (Iterative Dichotomiser 3) as an algorithm to explore patterns in car purchasing decisions in Indonesia, given its ability to form decision trees that can reflect alternative choices and decision results. This study aims to analyze car purchase data and provide more insight into consumer preferences in the context of the Indonesian automotive market

Keyword: : Car, Entropy, Gain, Indonesia, ID3 Method

ABSTRAK

Kondisi geografis Indonesia yang terdiri dari kepulauan mendorong perlunya pengembangan alat transportasi yang efisien, khususnya otomotif, untuk meningkatkan perekonomian negara. Mobil dipandang sebagai kendaraan yang lebih aman dibandingkan sepeda motor, dan dengan

meningkatnya jumlah perusahaan otomotif, persaingan di pasar semakin ketat. Penjualan mobil, khususnya dari merek-merek seperti Toyota, Honda, dan Mitsubishi, menunjukkan perkembangan pesat, dengan Toyota mencatatkan pertumbuhan penjualan 6,5% dan Honda mengalami kenaikan 56% pada tahun 2023. Kondisi ini menuntut perusahaan untuk memahami kebutuhan konsumen dan menerapkan strategi bisnis yang tepat. Kenyamanan dan kualitas mobil menjadi prioritas utama dalam pemilihan kendaraan, meskipun harga juga menjadi pertimbangan penting. Dengan pertumbuhan pemasaran melalui media sosial, data yang diharapkan dapat dimanfaatkan untuk analisis lebih lanjut, sehingga produsen dapat meningkatkan kualitas produk mereka. Tujuan menggunakan Metode ID3 (Iterative Dichotomiser 3) sebagai algoritma untuk menggali pola dalam keputusan pembelian mobil di Indonesia, mengingat kemampuannya dalam membentuk pohon keputusan yang dapat mencerminkan alternatif pilihan dan hasil keputusan. Penelitian ini bertujuan untuk menganalisis data pembelian mobil dan memberikan wawasan lebih mengenai preferensi konsumen dalam konteks pasar otomotif Indonesia

Kata Kunci: Entropy, Gain, Indonesia, Metode ID3

1. PENDAHULUAN

Indonesia adalah salah satu negara dengan kepemilikan penduduk yang berjumlah banyak, dalam mobilitas sehari-hari. Mobil adalah modal transportasi yang mayoritas dipakai oleh masyarakat. Seiring dengan meningkatnya jumlah pengguna mobil di Indonesia, banyak industri otomotif yang membangun pabrik di Indonesia sehingga mobil yang digunakan di Indonesia merupakan gabungan dari mobil buatan Indonesia dan mobil yang diimpor dari berbagai negara dan merupakan industri dimana banyak menampung angkatan kerja yang ada (Retno, et.al., 2021). Jumlah perusahaan otomotif yang semakin meningkat, dapat memicu persaingan antar perusahaan untuk mendapatkan pasar yang lebih besar. Jutaan kendaraan banyak diproduksi dengan varian baru tiap tahunnya, hal tersebut membuat mereka mengalami jatuh bangun untuk mencapai puncak kejayaan, Agar tidak kalah bersaing, perusahaan dalam negeri mengembangkan pasar dan menyesuaikan perkembangan yang terjadi saat ini dengan menerapkan strategi bisnis yang didapatkan dari menganalisis kebutuhan pembeli (Pratama, et.al., 2021). Penjualan kendaraan bermotor khususnya mobil memiliki perkembangan yang sangat pesat ini seperti toyota, daihatsu, honda, mitsubishi, suzuki yang menguasai pasar otomotif di Indonesia pada tahun 2023-2024. (Gakindo, 2024).

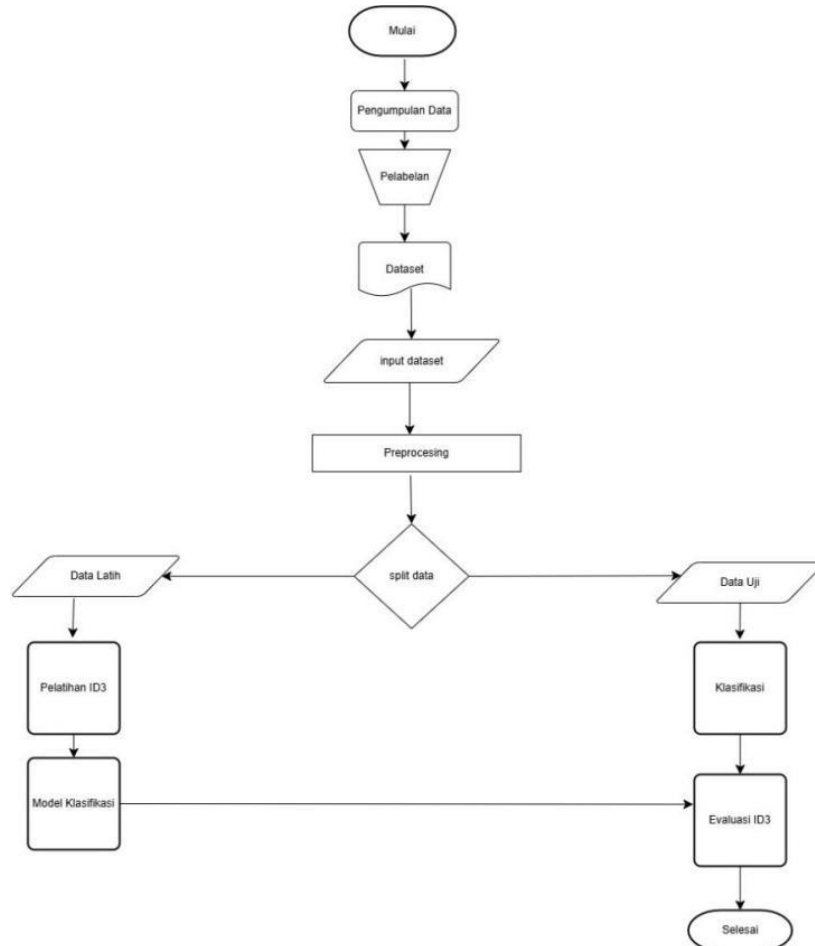
Berdasarkan data Gabungan Industri Kendaraan Bermotor Indonesia pada tahun 2023 Toyota berkontribusi ritel 156.83 unit dan mengalami pertumbuhan penjualan sebesar 6,5%. Pada tahun 2023 Honda mengalami kenaikan ritel 56% tahun 2024 sebesar 10,64% penjualan ritel Honda sebanyak 7.609 unit pada September 2024 kendati demikian, penjualan itu turun 10,64% secara bulanan, dibandingkan capaian Agustus sebanyak 8.515 unit. Secara keseluruhan, sepanjang 9 bulan pertama 2024, Honda mencatatkan penjualan mobil secara ritel tembus 75.514 unit, dengan pangsa pasar sebesar 11,5%. penjualan secara ritel juga tercatat menurun 11,53 persen dari 7.238 unit pada Mei 2023 menjadi 6.403 unit. (Gaikindo, 2023). Begitupun dengan Daihatsu yang mengeluarkan penjualan ritel sebanyak 194.108 unit dan mengalami kenaikan 2,9%. Model yang paling laris seperti Sigr, Gran Max PU (Pick-Up), dan Terios (Daihatsu, 2023). Selanjutnya pada tahun 2023 Suzuki mengeluarkan 82.244 unit dan mengalami kenaikan 25% di akhir tahun (Satutempo, 2023). Penjualan mobil secara retail atau dari dealer ke konsumen pun mengalami penurunan. Berdasarkan data Gaikindo, retail sales mobil baru sepanjang semester pertama tahun 2024 hanya sebanyak 431.987 unit. Angka itu turun 14% dibanding Januari-Juni 2023 yang mencatatkan angka penjualan sebanyak 502.533 unit (oto, 2024).

Kurva penjualan mobil mengalami penurunan dan peningkatan setiap tahun. Perusahaan juga harus memperhatikan kebutuhan konsumen selain fokus pada produk yang terjual. Banyak jenis mobil yang *trend* di Indonesia namun tidak sesuai kebutuhan, sehingga hal ini menjadi faktor pertimbangan dalam membeli mobil seperti kualitas dari konsumen (Hendra, 2017). Untuk kebutuhan sehari-hari jumlah kapasitas penumpang, jenis mobil, warna, kualitas mobil. Data yang diunggah di media sosial mengandung banyak kegunaan yang bisa diteliti dan dianalisis untuk menemukan informasi baru. yang bermanfaat. Data yang tersedia di media sosial bisa dimanfaatkan untuk mendapatkan informasi khusus, sehingga menarik untuk analisis lebih mendalam. (Liebenlito, et.al., 2024).

Data mining merupakan ekstraksi pola yang menarik dari kumpulan data dalam jumlah besar. Alasan utama mengapa data mining diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan pengetahuan yang berguna (Widiyati, et.al., 2018). Metode ID3 (*Iterative Dichotomiser 3*), sebagai algoritma pembentukan pohon keputusan, sangat berguna untuk menemukan pola tersembunyi dalam pembelian mobil di Indonesia. Pohon keputusan merupakan sebuah pohon dimana masing-masing cabang dari simpul merepresentasikan alternatif pilihan dan masing-masing ujung simpul/node merepresentasikan keputusan. Algoritma ini di implementasikan pada Dataset mobil (Pratama, et.al., 2021). Berdasarkan uraian permasalahan, peneliti melakukan analisis dengan topik penelitian “Analisis Pola Pembelian Mobil Di Indonesia Menggunakan Id3”. Sehingga pembeli dapat menemukan rekomendasi mobil yang trend saat ini sesuai kebutuhan.

2. METODE

2.1 Prosedur Penelitian



Gambar 1. *Flowchart* Penelitian

Berdasarkan Gambar 1, perancangan sistem dimulai dengan pengambilan data respon pengguna dari media sosial X menggunakan tweet Harvest, yang kemudian diklasifikasikan menjadi dua kelas: positif dan negatif, melalui Google Colab dengan perintah Python. Opini positif mencerminkan reaksi yang menyenangkan, sedangkan opini negatif menciptakan kesan buruk terhadap produk. Setelah proses pelabelan, data mengalami preprocessing yang meliputi *cleaning*, *casefolding*, *normalisasi*, *tokenizing*, dan *stopword removal*. Selanjutnya, data dibagi menjadi dua bagian yaitu data training dan data testing dengan ketentuan 80% yaitu 160 data dan data training dan 20% yaitu 40 data. Tahap berikutnya adalah pemberian bobot pada setiap term di data training menggunakan TF-IDF, yang membantu meningkatkan akurasi model analisis sentimen. Akhirnya, pelatihan data dibor menggunakan algoritma ID3 untuk mengidentifikasi pola dalam sentimen.

3. HASIL DAN PEMBAHASAN

3.1 *Crawling* Data

Crawling data yang diambil dari media sosial X dengan kata kunci “Honda, Toyota, Suzuki, Mitsubishi, Daihatsu” *crawling* data yang diambil dalam bentuk teks yang merupakan *tweet* dari pengguna aplikasi X. Data yang diambil pada tanggal Desember 2023 – 31 Desember 2024 seluruh data 1000 *tweet* Dengan Pembagian setiap merek yg di *crawling* berjumlah 200. Dalam *crawling* data, penelitian ini menggunakan *tweet harvest* yang disimpan dalam bentuk file csv. Bagian pertama pada kode program *crawling* data dengan pengaturan pada otentikasi untuk mendapatkan akses API twitter.

```
#title Twitter Auth Token
twitter_auth_token = 'a995d1529bdd30aef719e01d82f6edcf75b4678b' # change this auth token
```

Gambar 2. *Crawling* Data

```
# Crawl Data
filename = 'honda.csv'
search_keyword = 'gibran since:2023-12-31 until:2024-12-31 lang:id'
limit = 200

Inpx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {"86b9146eac2fcc151ch3501160f59"}
```

Gambar 3. Memasukkan Keyword

Penjelasan *source code* pada Gambar 1 pada baris pertama untuk mengumpulkan data dari Twitter yang relevan dengan kata kunci tertentu, dalam hal ini Mobil Indonesia, dalam bahasa Indonesia, dalam rentang waktu tertentu. Data yang diambil berupa tweet-tweet yang memenuhi kriteria pencarian yang ditentukan, dan hasilnya disimpan dalam sebuah file CSV dengan nama 'Honda.csv'. Kemudian, kata kunci pencarian yang ditentukan yaitu 'Honda' since:2023-12-31 until:2024-12-31 lang:id", yang berarti data yang dicari adalah tweet-tweet yang mengandung kata mobil 'Honda' dalam bahasa Indonesia dan diposting antara Desember 2023 – 31 Desember 2024. Batasan jumlah data yang akan digunakan ditetapkan sejumlah 200 tweet.

conversation_id_str	created_at	favorite_count	full_text	id_str
1873630419129929998	Mon Dec 30 22:17:45 +0000 2024	6	@RWWReborn Di mobil² ada weight sensor & safety belt sensor dan itu bakal buzzing kalau safety beltnya gak dipakai. Di Honda City 2000 awal malah kalau gak pakai safety belt gk bakal bisa di starter mobilnya. Cuman semua firur ini dianggap mengganggu kenyamanan jadi pada diakalin.	1873856025167864129
1873845321916416387	Mon Dec 30 21:35:13 +0000 2024	0	Andas con Tokio honda y Kawasaki	1873845321916416387

Gambar 4. Hasil *Crawling*

3.2 Labeling Sentimen

komentar pada postingan pada aplikasi X dilakukan dengan cara menggunakan Perintah *python* di Google Colab .

```
read_time = end_time_read - start_time_read

# 2. Menampilkan nama kolom
print("Kolom-kolom yang tersedia:", df.columns)

# 3. Daftar kata positif dan negatif
kata_positif = [
    'bagus', 'mantap', 'puas', 'cepat', 'top', 'recommended', 'baik', 'oke',
    'suka', 'cocok', 'keren', 'mulus', 'senang', 'bahagia',
    'rekomendasi', 'top', 'luar biasa', 'nyaman', 'terpercaya', 'hebat',
    'bagus banget', 'super',
    'terbaik', 'worth it', 'puas banget',
    'cocok', 'irit', 'murah', 'rekomendasi', 'traveling', 'tangguh', 'kokoh', 'trend'
]

kata_negatif = [
    'jelek', 'buruk', 'lambat', 'tidak puas', 'mengecewakan',
    'parah', 'benci', 'rusak', 'tidak cocok',
    'tidak merekomendasikan', 'berhenti',
    'mobil tua', 'tidak suka', 'kecewa', 'tidak nyaman', 'menyesal', 'palsu', 'bohong',
    'tidak berguna', 'membuang uang',
    'hancur', 'kebakaran', 'kotor', 'mahal', 'lecek'
]
```

Gambar 5. Proses *labeling* di google colab

Gambar 5 menunjukkan proses *labeling* sentimen menggunakan Google Colab, dimulai dari pengunggahan file *Excel* berisi data ulasan. Setelah file dibaca, sistem menampilkan nama-nama kolom untuk memastikan struktur data. Dua daftar kosakata disiapkan untuk mendeteksi sentimen positif dan negatif. Setiap ulasan dianalisis berdasarkan kemunculan kata-kata dalam daftar tersebut, dan diberikan label sentimen secara otomatis. Berikut gambar tampilan dari hasil labeling di Google Colab yang tersimpan dengan tipe dokumen excel.

	ulasan	jumlah_positif	jumlah_negatif	label	sentimen
1					
2	Mobil yang paling umum di pedesaan Indonesia biasanya mobil serbaguna dan tan	1	0	1	Positif
3	Mobil terlaris sepanjang masa di Indonesia didominasi oleh Toyota Avanza Kijang (	0	0	1	Positif
4	Mobil Toyota Avanza lebih cocok dijual di Indonesia karena ada pasar tersendiri di	2	0	1	Positif
5	Toyota kembali mempertahankan posisinya sebagai produsen mobil terbesar di dui	0	0	-1	Negatif
6	Toyota Hilux perawatannya cukup simpel dan mahal jika kondisi normal cukup gan	0	1	-1	Negatif
7	Rizor Fuel Power (RFP) Solusi performa dan efisiensi untuk mobil bermesin diesel ti	0	0	1	Positif
8	Mobil Toyota Land Cruiser termahal dan terkecil di Indonesia. Kalo ga salah pas aw	0	1	-1	Negatif
9	Mobil Toyota akan menggeser pasar honda daihatsu hyundai dll. Di jalanan mobil t	0	0	-1	Negatif
10	perawatan Mobil Toyota hilux mahal Tidak baik ya Sahabat menyalakan rokok di d	1	1	1	Positif
11	musik speaker di mobil Toyota gr supra sangat besar lagi di mobil tim nyalain mus	0	1	-1	Negatif
12	Rizor Fuel Power (RFP) Solusi performa dan efisiensi untuk mobil bermesin diesel ti	0	0	1	Positif
13	Saat ini mungkin hanya bisa lihat-lihat di show room mobil suatu saat pasti bisa ke	0	0	-1	Negatif
14	Di umur segitu Sahabat mau pilih mobil Toyota tipe apa nih,bagus sih kalau Toyota	1	0	1	Positif
15	Ngomong-ngomong soal resolusi 2025. Sahabat punya target umur berapa nih pun	3	0	1	Positif
16	Maximizer 28 Solusi Turbo Lag dan backpressure untuk mobil bermesin bensin dan	0	0	-1	Negatif
17	Mobil Toyota yaris dari Surabaya tujuan ke Makassar dengan kapal roro estimasi pi	0	0	-1	Negatif
18	Toyota Camry menutup 2024 dengan penjualan lebih dari 168 ribu mobil di pasar I	0	0	1	Positif
19	Resolusi punya mobil Toyota 86 tapi tiap ada tanggal kembar kalap check out kerai	0	0	-1	Negatif
20	Toyota Corolla juga menunjukkan performa yang baik di pasar otomotif nasional pi	1	0	1	Positif
21	Semoga tahun ini bisa menjadi tahun keberuntungan Sahabat untuk beli mobil Toy	0	0	-1	Negatif
22	Mobil Terlaris Sepanjang 2024 di Indonesia Toyota Innova	0	0	-1	Negatif
23	hinda brio Kalaupun dijual di Indonesia tetap laku cuma tetap kalah saing dengan	0	0	1	Positif
24	Amiilin semoga impian untuk punya mobil Toyota Yaris impian Sahabat bisa segera	0	0	-1	Negatif
25	Kalau mobil Toyota favorit kamu apa nih Sahabat tahun ini toyota yaris sih menur.	1	0	1	Positif
26	Toyota Hilux memang Tahun Depan Sudah Punya Mobil Toyota Hilux Untuk Keluar	0	0	1	Positif

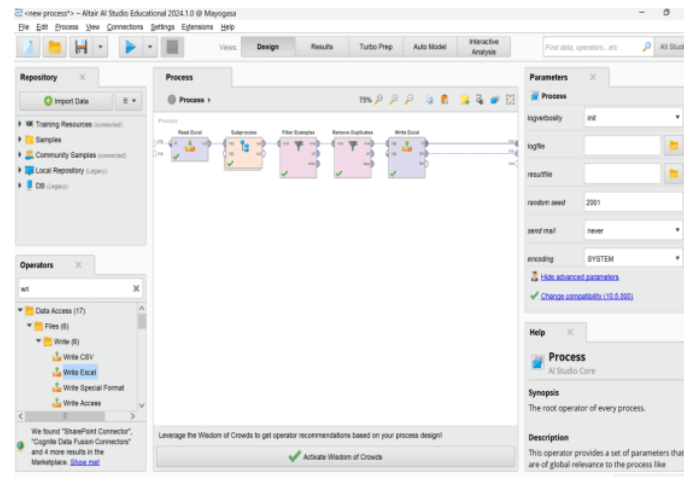
Gambar 6. Hasil labeling menggunakan google colab

3.3 Pembersihan Data dan Pra-Pemrosesan Teks

Pembersihan data dan *preprocessing text* merupakan langkah penting dalam analisis sentimen untuk memastikan data yang digunakan bersih dan relevan. Proses ini meliputi penghapusan elemen yang tidak diperlukan, *case folding*, *tokenizing*, *stopwords*, *stemming*, dan *filtering*. Tahapan ini bertujuan untuk meningkatkan kualitas data agar lebih siap dianalisis.

3.4 Proses *Cleaning*

Pada proses ini menghapus atribut yang tidak dibutuhkan dan tidak memiliki makna, seperti *hashtag*, *mention*, *retweet*, *whitespace*, dll. Berikut gambar 7 adalah contoh *cleaning* pada dataset.

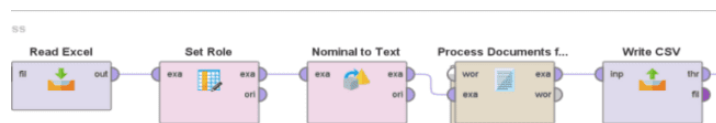


Gambar 7. Proses *cleaning* Data Set

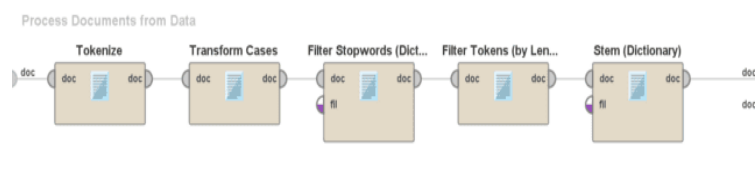
Gambar 7 merupakan proses pembersihan menggunakan beberapa operator *Read excel* memungkinkan pengguna untuk mengimpor data dari file *Excel*, memulai analisis dengan dataset yang ada. Subproses mengelompokkan operator menjadi satu unit yang dapat digunakan kembali, meningkatkan organisasi dan keterbacaan alur kerja. *Filter examples* menyaring dataset berdasarkan kriteria tertentu, sehingga pengguna dapat fokus pada data yang relevan. *Remove duplicates* mengidentifikasi dan menghapus entri yang sama, memastikan analisis dilakukan pada data yang unik. Terakhir, *Write excel* mengeksport hasil ke format *excel*,

3.5 *PreProcessing* Teks

Pada *preprocessing* text meliputi tahapan-tahapan *case folding*, *Tokenizing*, *Stopword removal*, dan *Stemming*, *Filtering*. Pada gambar 8 dan gambar 9



Gambar 8. Proses *Preprocessing* Text



Gambar 9. Proses *Preprocessing* Text

Gambar 8 menunjukkan alur Preprocessing teks dimulai dengan operator *read excel* untuk mengimpor data dari file *excel*. Selanjutnya, *set role* mengatur peran atribut, diikuti oleh *nominal to text* yang mengubah atribut nominal menjadi teks. Selanjutnya pada Gambar 9 operator *process documents* kemudian melakukan analisis lebih mendalam, termasuk *Tokenize* yang memecah teks menjadi token. Proses dilanjutkan dengan *transform cases* yang menyamakan huruf menjadi kecil, dan *filter stopwords (dictionary)* yang menghapus kata-kata umum yang tidak signifikan. *filter tokens* (berdasarkan Panjang) menghilangkan token yang tidak relevan, sementara *stem (dictionary)* mengubah kata-kata menjadi bentuk dasarnya. Setelah semua proses ini, data disimpan menggunakan operator *write csv* untuk mengeksport hasil ke dalam format csv

1. Proses *Case Folding* : Pada proses ini semua huruf akan diubah menjadi huruf kecil untuk mencegah kasus yang sensitif. Hal ini dapat meningkatkan kinerja algoritma dalam mengklasifikasikan teks yang tidak konsisten.
2. Proses *Tokenizing* : Pada proses *tokenizing*, setiap spasi dianggap sebagai pemisah dari sebuah token. Pengubahan kalimat menjadi token.
3. Proses *Stemming* : Proses *stemming* akan mengubah kata-kata yang ditemukan dalam teks menjadi bentuk dasarnya.
4. Proses *Filtering* : *NaturalLanguage Processing* (NLP) yang dikenal sebagai *filtering* memungkinkan penghapusan elemen tertentu dari teks yang dianggap tidak relevan atau tidak penting untuk analisis.

3.6 Feature Extraction Using TF-IDF

Pada proses ini dilakukan menghitung banyaknya *term* atau kata yang muncul pada *tweet* (tf), menghitung banyaknya *tweet* yang mengandung *term* tersebut (df), menghitung inverse dokumen *frequency* (idf), dan mengalikan tf dengan idf sebagai bobot dari *term* pada setiap *tweet*. Hasil dari proses ekstraksi fitur menggunakan TFIDF diperlihatkan pada gambar 47, gambar 48, gambar 49, gambar 50, gambar 51 menunjukkan hasil dari TF-IDF merek mobil Daihatsu, Honda, Mitsubishi, Toyota, Suzuki. di mana setiap baris mencantumkan kata atau token yang telah diproses. Setiap kolom-kolom yang menunjukkan jumlah atau frekuensi kemunculan setiap kata dalam dataset yang dianalisis.

abangkuhhh Number	abdul Number	abis Number	acara Number	aceh Number	acer Number	adeptusartia Number	adil Number	adjeliu Number
0	0	0	0	0	0.233	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Gambar 10. Hasil TF-IDF merek mobil Daihatsu

Analisis Pola Pembelian Mobil di Indonesia Menggunakan Metode ID3

abibmisbah Number	abis Number	abroad Number	acara Number	accord Number	aceh Number	acura Number	afeela Number	agunghan Number
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Gambar 11. Hasil TF-IDF merek mobil Mitsubishi

abalkan Number	adaptasi Number	adventure Number	aerodinamis... Number	airbag Number	ajak Number	ajang Number	akibat Number	aksen Number
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0.301	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Gambar 12. Hasil TF-IDF merek mobil Suzuki

ganti Number	gardan Number	gaya Number	gegap Number	gelaran Number	gelombang Number	gempita Number	generasi Number	gigi Number
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0.277	0.277	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.216	0
0	0	0	0	0	0	0	0	0

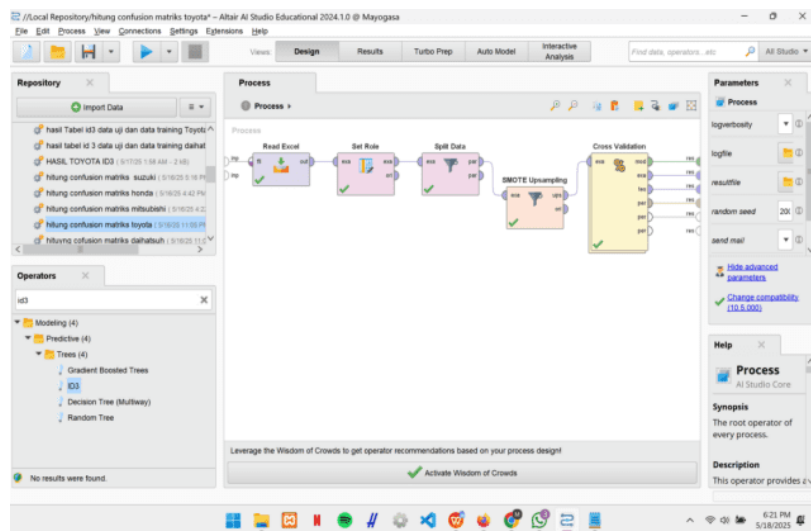
Gambar 13. Hasil TF-IDF merek mobil Toyota

abibmisbah Number	abis Number	abroad Number	acara Number	accord Number	aceh Number	acura Number	afeela Number	agunghan Number
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Gambar 14. Hasil TF-IDF merek mobil Honda

3. 7 Analisis Sentimen dengan Algoritma ID3

Data yang sudah melalui *preprocessing* kemudian melakukan pelabelan menggunakan *googke collabs* dengan *phyton* Selanjutnya akan dilakukan pengujian analisis sentimen dengan metode ID3 dan menambahkan atribut Merek,Jenis Tipe,Kapasitas Penumpang.



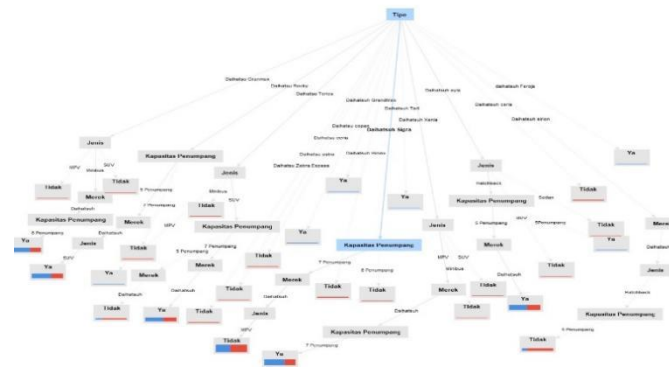
Gambar.15 Proses Implementasi ID3

Gambar 15 menunjukkan alur proses implementasi ID3 yang melibatkan beberapa operator untuk menganalisis data. Proses dimulai dengan *read excel* untuk mengimpor dataset, diikuti oleh *set role* untuk menentukan peran atribut dalam dataset. Selanjutnya, *split data* membagi dataset menjadi bagian pelatihan dan pengujian, memastikan model diuji dengan data yang belum pernah dilihat. Operator *smote* digunakan untuk menangani ketelitian kelas dengan menciptakan salinan sintesis dari kelas minoritas. Pada gambar 21 Setelah keputusan pohon model dibangun, operator *apply model* menerapkan model pada uji data, diikuti oleh operator *performance* untuk mengukur akurasi, presisi, dan recall, sehingga memberikan gambaran tentang kinerja model dalam memprediksi data yang tidak diketahui dan memastikan evaluasi yang menyeluruh.

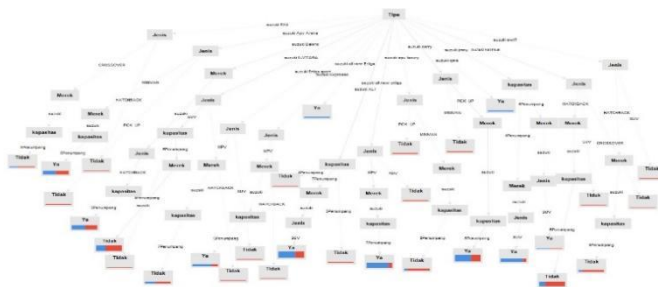


Gambar 16. Hasil Implementasi ID3 Pada merek mobil Toyota

Analisis Pola Pembelian Mobil di Indonesia Menggunakan Metode ID3



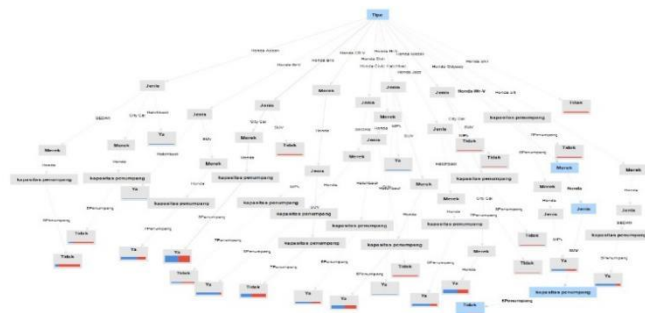
Gambar 17. Hasil Implementasi ID3 Pada merek mobil Daihatsu



Gambar 18. Hasil Implementasi ID3 Pada merek mobil Suzuki



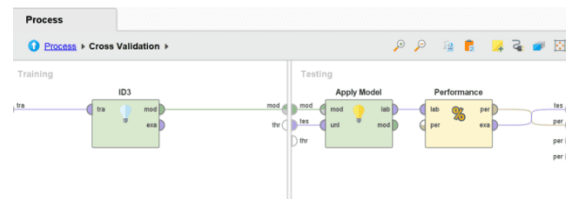
Gambar 19. Hasil Implementasi ID3 Pada merek mobil Mitsubishi



Gambar 20. Hasil Implementasi ID3 Pada merek mobil Honda

3.8 Evaluasi

Hasil evaluasi yang digunakan pada kedua metode ini berupa nilai *confusion matrix* berisi nilai akurasi, presisi, dan recall yang diambil dari data test. Dalam menentukan nilai *confusion matrix*, penelitian ini akan menggunakan *k-fold cross validation* dengan nilai $k=10$ agar nilai yang dihasilkan maksimal.



Gambar 21. Tahap *Cross Validation*

accuracy: 75.85% +/- 7.39% (micro average: 75.85%)

	true Ya	true Tidak	class precision
pred. Ya	89	28	76.07%
pred. Tidak	29	90	75.63%
class recall	75.42%	76.27%	

Gambar 22. Hasil *Confusion Matrix* merek Mitsubishi

Nilai akurasi 75,85%, yang berarti 75,85% data dapat diklasifikasikan dengan benar, dengan margin error sebesar $\pm 7,39\%$ yang menunjukkan kemungkinan kesalahan dalam prediksi. Model memberikan rekomendasi seimbang, dengan 118 untuk "Ya" dan 118 untuk "Tidak." Model presisi adalah 76,07% untuk prediksi "Ya" dan 75,63% untuk "Tidak," menunjukkan tingkat akurasi yang tinggi dalam rekomendasi. Nilai recall menunjukkan keberhasilan model dalam menemukan informasi, dengan hasil 75,42% untuk data "Ya" dan 76,27% untuk data "Tidak," yang mengindikasikan bahwa model dapat diandalkan dalam memberikan rekomendasi yang tepat kepada calon pembeli.

accuracy: 60.45% +/- 9.10% (micro average: 60.45%)

	true Ya	true Tidak	class precision
pred. Ya	67	64	57.62%
pred. Tidak	23	46	66.67%
class recall	79.09%	41.82%	

Gambar 23. Hasil *Confusion Matrix* merek Honda

Nilai akurasi 60,45%, yang menunjukkan bahwa 60,45% data dapat diklasifikasikan dengan benar, dengan margin error $\pm 9,10\%$ yang mengindikasikan kemungkinan kesalahan dalam prediksi. Model memberikan rekomendasi seimbang dengan 110 untuk "Ya" dan 110 untuk "Tidak", namun akurasi yang rendah berpotensi mempengaruhi kualitas rekomendasi. Model presisi adalah 57,62% untuk prediksi "Ya" dan 66,67% untuk "Tidak," menunjukkan bahwa terdapat kemungkinan besar ketidakakuratan pada rekomendasi "Ya." Nilai recall menunjukkan 79,09% untuk data

“Ya,” menandakan kemampuan model yang baik dalam menemukan data yang benar, sementara 41,82% untuk data “Tidak” menunjukkan kesulitan dalam mengidentifikasi data yang tidak direkomendasikan.

● Table View ○ Plot View

accuracy: 64.81% +/- 6.66% (micro average: 64.81%)

	true Ya	true Tidak	class precision
pred. Ya	103	71	59.20%
pred. Tidak	5	37	88.10%
class recall	95.37%	34.26%	

Gambar 24. Hasil *Confusion Matrix* merek Daihatsu

Nilai akurasi 64,81%, yang berarti 64,45% data dapat diklasifikasikan dengan benar, dengan margin error $\pm 6,66\%$ menunjukkan kemungkinan kesalahan dalam prediksi. Model memberikan rekomendasi seimbang dengan 108 untuk "Ya" dan 108 untuk "Tidak." Model presisi adalah 59,20% untuk prediksi "Ya," menunjukkan kemungkinan ketidakakuratan dalam rekomendasi pembelian, sementara 88,10% untuk prediksi "Tidak" menunjukkan kehandalan dalam identifikasi mobil yang tidak direkomendasikan. Nilai recall menunjukkan 95,37% untuk data “Ya,” menandakan efektivitas model dalam menemukan data yang benar, tetapi hanya 34,26% untuk data “Tidak,” menunjukkan kesulitan dalam mengidentifikasi mobil yang tidak layak direkomendasikan.

accuracy: 69.37% +/- 8.42% (micro average: 69.35%)

	true Ya	true Tidak	class precision
pred. Ya	100	52	65.79%
pred. Tidak	24	72	75.00%
class recall	80.65%	58.06%	

Gambar 25. Hasil *Confusion Matrix* merek Suzuki

Nilai akurasi 69,35%, menunjukkan bahwa model dapat mengklasifikasikan data dengan benar, dengan margin error $\pm 8,42\%$ yang menunjukkan kualitas hasil. Model memberikan rekomendasi seimbang dengan 124 untuk "Ya" dan 124 untuk "Tidak." Presisi model adalah 66,79% untuk prediksi "Ya," yang berarti ada kemungkinan ketidakakuratan dalam rekomendasi pembelian, sedangkan 75,00% untuk prediksi "Tidak" menunjukkan kehandalan dalam identifikasi mobil yang tidak layak dibeli. Nilai recall menunjukkan 80,65% untuk data “Ya,” menandakan kemampuan model dalam menemukan data relevan, namun hanya 56,06% untuk data “Tidak,” yang menunjukkan kurangnya efektivitas dalam mengidentifikasi mobil yang seharusnya tidak direkomendasikan.

● Table View ○ Plot View

accuracy: 52.00% +/- 9.49% (micro average: 52.00%)

	true Ya	true Tidak	class precision
pred. Ya	57	53	51.82%
pred. Tidak	43	47	52.22%
class recall	57.00%	47.00%	

Gambar 26. Hasil *Confusion Matrix* merek Toyota

Nilai akurasi 52,00%, yang menunjukkan bahwa model dapat mengklasifikasikan data dengan benar, dengan margin error $\pm 9,49\%$ yang. Model memberikan rekomendasi seimbang dengan 100 untuk "Ya" dan 100 untuk "Tidak." Presisi adalah 51,82% untuk prediksi "Ya" dan 52,22% untuk "Tidak," menandakan bahwa model tidak efektif dalam memberikan rekomendasi yang akurat. Nilai recall menunjukkan 57,00% untuk data "Ya" dan 47,00% untuk data "Tidak", menandakan kesulitan model dalam menemukan data relevan untuk kedua kategori, sehingga banyak mobil yang seharusnya direkomendasikan atau tidak direkomendasikan tidak terdeteksi dengan baik.

4. PENUTUP

4.1 Kesimpulan

Berdasarkan metode analisis ID3 terhadap lima merek mobil, Mitsubishi menunjukkan akurasi terbaik sebesar 75,85%, terutama untuk tipe Xpander di kategori MPV dengan 15 rekomendasi "Ya." Suzuki mengikuti dengan akurasi 69,35%, di mana XL7 mendapat 19 rekomendasi dan All New Ertiga 20 rekomendasi terbanyak dalam kategori SUV dan MPV. Daihatsu, dengan akurasi 64,81%, menarik perhatian Xenia yang mendapatkan 21 rekomendasi "Ya" di segmen MPV. Honda, dengan akurasi terendah 60,45%, berhasil menarik minat melalui Brio yang mencatatkan 22 rekomendasi "Ya" dalam kategori City Car. Terakhir, Toyota mencatatkan akurasi 52,00%, dengan Avanza dan Land Cruiser sebagai pilihan populer meskipun tidak meraih jumlah rekomendasi tertinggi.

4.2 saran

Penelitian selanjutnya yaitu membandingkan performa ID3 dengan metode decision tree lainnya menggunakan algoritma, C4.5, *Random Forest*, CART.

DAFTAR PUSTAKA

- Santosa, B. (2014) 'Algoritma Iterative Dichotomiser 3 (ID3) untuk Pengambilan Keputusan', *Jurnal Teknologi Informasi*, 10(2), pp. 45–52. Available at: <https://media.neliti.com/media/publications/59658-ID-algoritma-iterative-dichotomizer-3-id3-p.pdf>
- Jayapana, R.D. & Yuniarsi, R. (2015) 'Analisis Pola Pembelian Konsumen dengan Algoritma Apriori pada Apotek Rahayu Jepara', *UG Jurnal*, 8, pp. 1–10. Available at: <https://core.ac.uk/download/pdf/35381950.pdf>
- Wianto, P.W.A. (2018) 'Analisis Sentimen Media Sosial untuk Teks Berbahasa Indonesia Menggunakan Algoritma CNN (Convolutional Neural Network) (Studi Kasus: Politik)', Skripsi, Institut Teknologi Sepuluh Nopember. Available at: <https://repository.its.ac.id/52593/>
- Utomo, D.P. & Mesran. (2020) 'Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut pada DataSet Penyakit Jantung', *Jurnal Media Informatika Budidarma*, 4(2), pp. 437–444. Available at: <https://ejurnal.stmik-budidarma.ac.id/mib/article/view/2080/1541>

- Rahim, A. & Jusia, P.A. (2024) 'Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma Decision Tree Classifier', Indonesian Journal of Computer Science, 13(1), pp. 1308–1322. Available at: <https://ijcs.net/ijcs/index.php/ijcs/article/view/3692>
- Kohsasih, R.A. (2022) 'Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor', Jurnal Nasional Informatika dan Teknologi Jaringan, 6(1), pp. 134–139. Available at: <https://repository.horizon.ac.id/items/show/3362>
- Pratama, D.A.P., Nugraha, N.S., Maulana, R., Rahmasari, A. & Fauzi, A. (2021) 'Implementasi Algoritma ID3 untuk Klasifikasi Kualitas Mobil', Proceedings of the Global Digital Communication Symposium (GDSCS). Available at: <https://www.conferences.uinsgd.ac.id/index.php/gdcs/article/view/97>
- Hendra, F. (2017) 'Analisis Tingkat Kematangan Industri Komponen Otomotif di Indonesia', SINTEK JURNAL: Jurnal Ilmiah Teknik Mesin, 11(1), pp. 38–48. Available at: <https://jurnal.umj.ac.id/index.php/sintek/article/view/1473>
- KL, & Kotler, P. (2022). Branding dalam perusahaan B2B. Dalam *Handbook of business-to-business marketing* (hlm. 205-224). Edward Elgar Publishing.
- David, MCG (2004) Tutorial Algoritma Pohon Keputusan ID3 , Fakultas Teknologi Informasi Universitas Monash. Available at https://books.google.co.id/books?id=zLHGEEAAQBAJ&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Han, J. & Kamber, M. (2006) Data Mining: Concepts and Techniques , Morgan Kaufmann Publishers, San Francisco. Available at https://liacs.leidenuniv.nl/~bakkerem2/dbdm2012/03_dbdm2012_Data.pdf
- Hendra, F. (2017) 'Analisis Tingkat Kematangan Industri Komponen Otomotif di Indonesia', SINTEK JURNAL: Jurnal Ilmiah Teknik Mesin, 11(1), pp.38–48. Available at : <https://jurnal.umj.ac.id/index.php/sintek/article/view/1473> .
- Hikmatulloh, RAWD & Ambarsari, DA (2019) 'Penerapan Algoritma Iterative Dichotomiser Three (ID3) Dalam Mendiagnosa Kesehatan Kehamilan', Jurnal Ilmu Komputer, 6(2), pp.116–127. Available at <https://www.academia.edu/download/101003403/pdf.pdf>
- Isnawati, SIPMI & Bangsa, JR (2023) 'Pelatihan Konten Marketing Pada Industri Otomotif Dengan Media Video Marketing Di PT Wahana Investasindo Salatiga', Jurnal Abadimas Adi Buana, 6(2), pp.240–247. Available at <https://jurnal.unipasby.ac.id/index.php/abadimas/article/view>
- <https://industri.kontan.co.id/news/honda-raih-kenaikan-penjualan-retail-58-pada-oktober-2024>)
- <https://otomotif.bisnis.com/read/20241026/46/1810861/gaikindo-pangkas-target-penjualan-mobil-honda-sebut-lebih-realistik>)