



Penerapan BERT untuk Klasifikasi Aspek dalam Media Massa Otomotif Indonesia

Zahra Tifani Apriliana¹, Taufik Edy Sutanto²

^{1,2}Matematika, Universitas Islam Negeri Syarif Hidayatullah Jakarta

¹zahratifani05@gmail.com

Corresponding author: ²taufik.sutanto@uinjkt.ac.id

Abstract: The huge market opportunity for the automotive industry in Indonesia is reflected in the 5.7% increase in vehicle sales by March 2024. In this digital era, mass media is one of the main tools for automotive companies to market vehicles and optimize sales. To optimize sales, research on popular automotive aspects of both our company and competitors is needed. This research uses data from several mass media from January to July 2023 with a total of 14,654 data. The research method used is Bidirectional Encoder Representations from Transformers with accuracy results reaching 77%. In this study, we classified six aspects of the automotive industry. Based on the results of the study, the most frequently appearing aspects are environmentally friendly and safety. The top five automotive brands show different frequencies of aspect occurrence.

Keywords: Automotive Aspects, Automotive Industry, BERT, Mass Media

Abstrak: Peluang besar pasar industri otomotif di Indonesia tercermin dari peningkatan penjualan kendaraan sebesar 5,7% pada Maret 2024. Dalam era digital ini, media massa menjadi salah satu alat utama bagi perusahaan otomotif untuk memasarkan kendaraan dan mengoptimalkan penjualan. Untuk mengoptimalkan penjualan, penelitian mengenai aspek-aspek otomotif yang populer baik di perusahaan kita maupun kompetitor sangat diperlukan. Penelitian ini menggunakan data dari beberapa media massa dari Januari hingga Juli 2023 dengan total sebanyak 14.654 data. Metode penelitian yang digunakan adalah *Bidirectional Encoder Representations from Transformers* dengan akurasi sebesar 77%. Dalam penelitian ini, peneliti mengklasifikasikan enam aspek dalam industri otomotif. Berdasarkan hasil penelitian, aspek yang paling sering muncul adalah aspek ramah lingkungan dan keselamatan. Lima merek otomotif teratas menunjukkan frekuensi kemunculan aspek yang berbeda-beda.

Kata kunci: Aspek Otomotif, BERT, Industri Otomotif, Media Massa

I. PENDAHULUAN

Di tengah perkembangan pesat industri otomotif di Indonesia, klasifikasi aspek-aspek dalam media massa otomotif menjadi kunci untuk memahami dinamika pasar dan tren yang mempengaruhi konsumen. Kemajuan teknologi informasi dan komunikasi saat ini memudahkan akses terhadap informasi dari berbagai sumber, termasuk artikel mengenai industri otomotif yang tersebar luas. Namun, tantangan kompleks terletak pada pemilahan dan pengklasifikasian informasi ini. Penelitian mendalam mengenai aspek-aspek otomotif yang populer sangat penting untuk mendapatkan wawasan strategis yang mendukung peningkatan daya saing dan inovasi dalam menghadapi dinamika pasar yang terus berkembang.

Di era saat ini, kendaraan telah menjadi kebutuhan sekunder yang mendukung produktivitas sehari-hari manusia. Dengan keberadaan kendaraan, mobilitas menjadi lebih efektif dan efisien. Peluang pasar industri otomotif di Indonesia, khususnya untuk kendaraan roda empat, semakin terbuka lebar dengan peningkatan penjualan setiap tahunnya [1]. Permintaan yang meningkat terhadap kendaraan memberikan keuntungan sekaligus tantangan bagi perusahaan otomotif. Persaingan untuk menyediakan kendaraan yang memenuhi kebutuhan masyarakat Indonesia menjadi fokus utama bagi setiap perusahaan otomotif yang ingin memenangkan pangsa pasar. Perkembangan industri otomotif yang terus berubah dipacu oleh permintaan, kebutuhan konsumen, dan persaingan antar pelaku industri [2]. Perusahaan-perusahaan otomotif harus senantiasa memiliki ide dan inovasi baru dalam menghasilkan kendaraan terbaik, baik dari segi teknologi, fitur, keselamatan, maupun harga yang terjangkau, untuk menarik minat masyarakat.



Pemanfaatan media massa sebagai alat pemasaran produk otomotif sangat penting untuk meningkatkan penjualan di era digital saat ini. Dalam dunia yang semakin terhubung, konsumen memperoleh informasi dan membuat keputusan pembelian melalui berbagai platform digital. Oleh karena itu, perusahaan otomotif harus mengoptimalkan media massa untuk mencapai audiens yang lebih luas dan efektif mempromosikan produk mereka. Solusi konvensional untuk memahami konteks dan popularitas topik tertentu pada suatu media melibatkan pengelompokan dataset dan pemilihan kluster yang paling relevan dengan topiknya, serta penerapan teknik menyimpulkan dan mengvisualisasikan kluster guna mengekstraksi konteks dan informasi yang bermakna [3]. Strategi pemasaran yang tepat dapat meningkatkan eksposur merek dan mendorong pertumbuhan penjualan. Dengan pendekatan yang sesuai, media massa dapat menjadi alat yang kuat dalam meraih keunggulan kompetitif dan memperkuat hubungan dengan konsumen. Oleh karena itu, penting bagi perusahaan untuk memahami tren dan preferensi pasar yang sedang populer dalam media massa otomotif guna merancang strategi pemasaran yang efektif dan relevan.

Banyaknya artikel mengenai otomotif dalam media massa menjadikan analisis menjadi sulit. Untuk mengatasi hal ini, diperlukan pengklasifikasian artikel tersebut menjadi enam aspek utama, yaitu biaya, keselamatan, ramah lingkungan, fitur, reliabilitas, dan lainnya. Oleh karena itu, penelitian ini akan menggunakan metode *Bidirectional Encoder Representations Transformers* (BERT) untuk mengklasifikasikan artikel media massa tentang otomotif.

Model deep learning yang sering digunakan dalam pemrosesan teks adalah *Bidirectional Encoder Representations Transformers* (BERT). BERT merupakan sebuah model representasi kata yang dipelajari sebelumnya menggunakan *Masked Language Model* (MLM) dengan menggunakan pendekatan *Transformers* dua arah [4]. Diperkenalkan pada tahun 2018, BERT telah terbukti memberikan kinerja yang sangat baik dalam berbagai penelitian NLP [5]. Berbeda dengan model bahasa lainnya, BERT dirancang sebagai model pra-latih yang dilatih secara *bidirectional* dari data teks tanpa label, yang menggabungkan konteks dari kedua arah pada setiap layer. Pendekatan ini memungkinkan BERT untuk di-*fine-tuning* dengan menambahkan satu layer tambahan [6]. Pada tahun 2020, terjadi kemajuan signifikan dengan pengembangan model BERT yang disesuaikan untuk bahasa Indonesia, dikenal sebagai IndoBERT [7].

Adapun penelitian-penelitian sebelumnya yaitu mengklasifikasikan berita olahraga menggunakan pendekatan *Naïve Bayes* dengan *Enhanced Confix Striping Stemmer* mampu mengklasifikasi berita olahraga sesuai dengan kategori masing-masing dengan keakuratan sebesar 77% [8]. Peneliti mengklasifikasikan berita menggunakan *naïve bayes* dan dua fase *feature selection* yang mendapatkan akurasi sebesar 86% [9]. Membandingkan model AraBERT yang telah dilatih sebelumnya dan di-*finetuning*, serta digunakan sebagai model ekstraksi fitur dalam kombinasi dengan beberapa *neural network classifier* dan SVM untuk tujuan klasifikasi. Hasil penelitian menunjukkan bahwa model AraBERT yang di-*fine-tune* secara optimal mencapai kinerja yang sangat tinggi, dengan akurasi hingga 99% [10]. Penggunaan IndoBERT *Embedding* dan CNN untuk mendeteksi aspek-aspek dari ulasan pada aplikasi belanja Bukalapak dan mendapatkan akurasi sebesar 94,86% [11].

II. METODE PENELITIAN

2.1. Analisis Pengumpulan Data

Pada penelitian ini, peneliti memperkenalkan pendekatan Analitik Berbasis Relevansi, yang dirancang untuk memahami persepsi merek dan dinamika pasar dalam konteks regulasi ketat terhadap data media sosial. Metodologi ini menggabungkan dua komponen utama: Seleksi Sampel Relevan dan



Pembelajaran Mesin Berbasis Seleksi Sampel Relevan. Setiap elemen dari metode ini telah dirancang secara hati-hati untuk mengatasi tantangan etis dalam pengumpulan data sambil memastikan analisis yang mendalam terhadap konten yang dihasilkan oleh pengguna.

Untuk menghadapi tantangan seputar kekhawatiran privasi dan kendala regulasi, penelitian ini menggunakan strategi Seleksi Sampel Relevan. Pendekatan ini melibatkan penggunaan data yang terindeks oleh mesin pencari untuk membuat sampel yang mewakili konten media sosial terkait kehadiran merek otomotif internasional di Indonesia. Fokus penelitian ini adalah pada data yang tersedia melalui indeks mesin pencari publik, dengan memastikan kepatuhan terhadap kebijakan perlindungan data serta mempertahankan kekayaan dan keberagaman dataset.

Untuk memastikan kepatuhan terhadap etika penelitian dan meningkatkan kehandalan pengumpulan data, penelitian ini menyertakan API protokol hukum dalam proses Seleksi Sampel Relevan. API ini memungkinkan untuk mengekstraksi data sesuai dengan pedoman hukum yang telah ditetapkan, menambahkan jaminan tambahan bahwa metodologi penelitian ini sesuai dengan kerangka kerja regulasi yang berlaku. Integrasi ini menjamin bahwa proses pengumpulan data penelitian ini tidak hanya memenuhi standar keamanan tetapi juga beroperasi dalam batas-batas protokol hukum yang ditetapkan.

Dalam strategi pendekatan penelitian ini, istilah “Relevan” menandakan pendekatan taktis dalam proses pengambilan sampel. Penelitian ini menegaskan kriteria relevansi sebelumnya, seperti konten terkait merek otomotif, dari pengguna di Indonesia, dan berdasarkan jangka waktu tertentu. Dengan fokus pada seleksi sampel yang terarah ini, peneliti dapat memusatkan perhatian pada data yang tidak hanya mematuhi regulasi keamanan tetapi juga sesuai dengan tujuan penelitian untuk memastikan analisis yang terfokus dan signifikan.

Berikut adalah beberapa kata kunci yang digunakan, yaitu “toyota”, “daihatsu”, “honda”, “hyundai”, “suzuki”, “mitsubishi”, dan “wuling”.

2.2. Data

Penelitian ini menggunakan data dalam format .csv yang berisi judul dan kutipan artikel media massa otomotif untuk melakukan klasifikasi terhadap beberapa aspek seperti biaya, keselamatan, ramah lingkungan, fitur, reliabilitas, dan lainnya. Namun, penelitian ini menghadapi tantangan dengan data yang tidak merata, di mana beberapa aspek terlihat lebih dominan dibandingkan yang lain. Untuk mengatasi ini, metode *balancing* data atau *oversampling* diterapkan. Tujuan dari *oversampling* adalah untuk memastikan setiap aspek dalam dataset memiliki representasi yang cukup untuk meningkatkan keadilan dan keakuratan analisis. Dengan melakukan *oversampling*, dataset diperluas dengan menambahkan duplikat atau variasi dari sampel-sampel yang kurang mewakili secara proporsional. Hal ini membantu mengurangi bias yang mungkin muncul akibat ketidakseimbangan dalam distribusi data sehingga hasil analisis dapat lebih akurat dan lebih mewakili keadaan sebenarnya. Penelitian ini menggunakan total 14.654 entri data yang telah dibagi menjadi data *training* dan *testing* dengan perbandingan 80% untuk *training* dan 20% untuk *testing*.

2.3. Preprocessing

Setelah mendapatkan data, langkah awal dilakukan adalah melakukan *preprocessing* untuk membersihkan data dari gangguan atau *noise*. Tahapan *preprocessing* yang dilakukan mencakup *lowercasing*, *unicode normalization*, *punctuation removal*, *tokenization*, dan *stopwords removal*.

Pada tahap *preprocessing*, langkah pertama adalah melakukan *lowercasing*. Proses *lowercasing* dilakukan dengan menggunakan metode “lower()” pada teks. Tujuan dari *lowercasing* adalah untuk



mengubah semua karakter teks menjadi huruf kecil. Hal ini penting untuk memastikan konsistensi dalam analisis teks sehingga kata-kata yang sama dengan kapitalisasi yang berbeda dianggap sama.

Selanjutnya, dilakukan proses *unicode normalization* dengan menggunakan fungsi *unidecode*. Normalisasi ini membantu mengubah karakter-karakter khusus ke dalam bentuk yang lebih umum sehingga dapat diproses dengan lebih baik oleh komputer. Setelah itu, dilakukan *punctuation removal* untuk menghapus semua tanda baca dan karakter non-kata dari teks. Tanda baca sering kali tidak memberikan informasi tambahan yang relevan dalam analisis teks sehingga penghapusan ini dapat membantu membersihkan teks dari *noise* yang tidak diinginkan.

Tahap berikutnya adalah *tokenization*. Tokenisasi dilakukan menggunakan fungsi “word_tokenize” dari NLTK (*Natural Language Toolkit*). Proses tokenisasi ini memecah teks menjadi token atau kata-kata individual berdasarkan spasi dan tanda baca. Misalnya, kalimat “Daftar Harga Toyota” akan dipecah menjadi token “Daftar”, “Harga”, dan “Toyota”. Proses ini adalah hal yang penting agar model dapat memproses teks lebih terstruktur dengan memperlakukan setiap kata sebagai unit yang terpisah.

Tahap terakhir adalah melakukan *stopwords removal*. Penghapusan stopwords dilakukan untuk menghilangkan kata-kata umum yang tidak memberikan nilai tambah dalam analisis teks, seperti kata hubung (“dan”, “atau”, “yang”) atau kata-kata lain yang sering muncul namun tidak relevan. Pada tahap ini digunakan daftar *stopwords* Bahasa Indonesia yang telah diunduh dari GitHub repository “taudataanalytics/taudata-Academy” untuk melakukan *filtering* terhadap token-token yang diperoleh dari tokenisasi. Langkah ini membantu meningkatkan relevansi kata-kata yang tersisa dalam teks yang akan diproses oleh model.

2.4. Bidirectional Encoder Representations Transformers

Tahap selanjutnya adalah melakukan pengklasifikasian dengan menggunakan metode *Bidirectional Encoder Representations Transformers* (BERT). Pertama-tama, kita memuat *tokenizer* BERT untuk memproses teks. Selanjutnya, kita menghitung panjang setiap kalimat untuk menentukan panjang maksimum token yang akan digunakan, dengan memperhatikan rata-rata dan median panjang kalimat. Proses tokenisasi dimulai dengan penambahan token khusus, seperti [CLS] pada awal kalimat dan [SEP] pada akhir kalimat, sesuai dengan aturan BERT. Setiap kata dalam kalimat kemudian diubah menjadi token ID. BERT mengonversi setiap token kata ke dalam token ID dan menghasilkan output berupa “input_ids” dan “attention_masks”, yang digunakan sebagai masukan untuk model BERT. Langkah selanjutnya adalah membuat *dataset* tensor. *Dataset* ini dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data validasi. Data ini kemudian dimuat ke dalam *DataLoader*, yang digunakan untuk mengiterasi dataset dalam bentuk *batch*. Setelahnya, model BERT untuk pengklasifikasian artikel media massa siap dibuat. Dilanjutkan dengan proses pelatihan dan validasi menggunakan model tersebut. Setelah model dilatih, model akan digunakan untuk mengklasifikasikan artikel media massa yang baru.

2.5. Teknik Evaluasi

Dalam penelitian ini, digunakan metode *cross validation* dengan 5 *fold* dan 5 *epoch* untuk memastikan keakuratan dan keandalan model prediksi distribusi aspek otomotif. *Cross validation* melibatkan pembagian data menjadi lima bagian (*fold*) menggunakan metode *K-Fold Cross Validation*. Model dilatih pada empat *fold* data dan diuji pada satu *fold* yang tersisa. Proses ini diulang lima kali dengan kombinasi data yang berbeda, sehingga setiap bagian data digunakan baik untuk pelatihan maupun pengujian. Dengan cara ini, hasil penelitian lebih akurat dan dapat dipercaya, karena model diuji secara menyeluruh dan mengurangi kemungkinan kesalahan prediksi.



III. HASIL DAN PEMBAHASAN

3.1 Pengaruh Jumlah Epoch

Penelitian ini mengeksplorasi pengaruh jumlah *epoch* terhadap kinerja model dengan menguji beberapa nilai *epoch* yang berbeda. Hal ini bertujuan untuk memperoleh nilai *epoch* optimal sehingga model dapat memberikan hasil performa yang baik. *Epoch* dalam konteks pelatihan *neural network* adalah saat seluruh dataset diteruskan ke depan untuk membuat prediksi dan ke belakang untuk menyesuaikan parameter berdasarkan hasilnya melalui *neural network* hanya satu kali [12]. Pada penelitian ini, dilakukan pengujian terhadap beberapa nilai *epoch* yaitu, 3, 5, 7, dan 10. Pada saat melakukan pengujian, beberapa parameter seperti *batch size* diatur sebesar 32 dan *learning rate* sebesar $2e-5$. Berikut hasil pengujian *epoch* dapat dilihat pada tabel di bawah.

Tabel 1. Perbandingan Akurasi Terhadap Penggunaan Jumlah *Epoch*

<i>Epoch</i>	Akurasi
3	71.54
5	73.70
10	77.39

Berdasarkan Tabel 1, dapat dilihat bahwa semakin tingginya nilai *epoch* maka akan menghasilkan nilai akurasi yang semakin baik. Peningkatan akurasi ketika menggunakan nilai *epoch* 3 ke *epoch* 10 menunjukkan bahwa artinya dengan tingginya nilai *epoch* maka model terus belajar dari data dan meningkatkan kemampuannya untuk membuat prediksi yang lebih akurat seiring berjalannya waktu. Oleh karena itu, dapat disimpulkan bahwa *epoch* 10 adalah nilai *epoch* yang optimal dibandingkan dengan nilai *epoch* lainnya.

3.2 Pengaruh Nilai Batch Size

Penelitian ini juga mengevaluasi pengaruh nilai *batch size* terhadap kinerja model dengan mencoba beberapa variasi nilai *batch size*. Tujuannya adalah untuk menemukan nilai *batch size* yang paling efektif dalam mencapai akurasi tinggi dan stabilitas selama proses pelatihan. *Batch size* adalah sebuah istilah dalam pembelajaran mesin yang menggambarkan jumlah sampel pelatihan yang diproses dalam satu iterasi. *Batch size* merupakan salah satu *hyperparameter* utama yang harus diatur dengan cermat dalam sistem *deep learning*, *batch size* mempengaruhi efisiensi dan hasil pelatihan model [13]. Pada penelitian ini, dilakukan pengujian terhadap beberapa nilai *batch size* yaitu, 16 dan 32. Pada saat melakukan pengujian, beberapa parameter seperti *epoch* diatur sebesar 10 dan *learning rate* sebesar $2e-5$. Berikut hasil pengujian *batch size* dapat dilihat pada tabel di bawah.

Tabel 2. Perbandingan Akurasi Terhadap Penggunaan Jumlah *Batch Size*

<i>Batch Size</i>	Akurasi
16	76.39
32	77.39

Berdasarkan Tabel 2, dapat dilihat bahwa *batch size* 32 memberikan sedikit peningkatan dalam akurasi dibandingkan dengan *batch size* 16. Hal ini menunjukkan bahwa ukuran *batch* yang lebih besar mungkin lebih sesuai untuk kasus ini dengan mengoptimalkan keseimbangan antara akurasi dan efisiensi komputasi. Pada saat melakukan pelatihan model dengan menggunakan nilai *batch size* 32 maka didapatkan akurasi sebesar 77.39%.



3.3 Pengaruh Nilai Learning Rate

Penelitian ini meneliti pengaruh nilai *learning rate* terhadap kinerja model dengan menguji berbagai nilai *learning rate*. Tujuannya adalah untuk menemukan *learning rate* yang optimal yang memungkinkan model untuk belajar secara efektif tanpa mengalami masalah seperti konvergensi lambat atau divergensi. *Learning rate* merupakan parameter penting dalam *training* yang berfungsi untuk menentukan besarnya koreksi bobot yang diterapkan selama proses pelatihan model [13]. Pada penelitian ini, dilakukan pengujian terhadap beberapa nilai *learning rate* yaitu, $5e-5$, $3e-5$, dan $2e-5$. Pada saat melakukan pengujian, beberapa parameter seperti *epoch* diatur sebesar 10 dan *batch size* sebesar 32. Berikut hasil pengujian *learning rate* dapat dilihat pada tabel di bawah.

Tabel 3. Perbandingan Akurasi Terhadap Penggunaan Jumlah *Learning Rate*

<i>Learning Rate</i>	Akurasi
$5e-5$	76.79
$3e-5$	77.32
$2e-5$	77.39

Berdasarkan Tabel 3, dapat dilihat bahwa semakin tingginya nilai *learning rate* maka akan menghasilkan nilai akurasi yang semakin baik. penurunan kecil dari $3e-5$ ke $2e-5$ menghasilkan peningkatan akurasi yang signifikan, menunjukkan bahwa penyesuaian kecil dalam *learning rate* dapat memiliki dampak besar pada kinerja akhir model. *Learning rate* dengan akurasi yang lebih baik pada pengujian ini adalah $2e-5$.

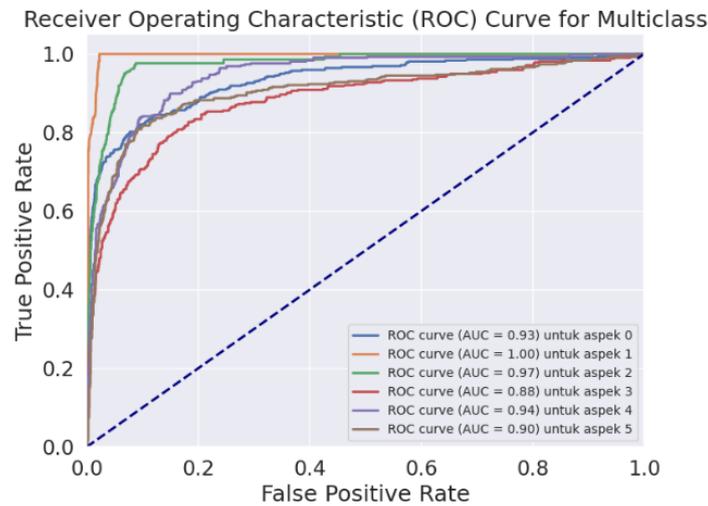
3.4 Metrik Evaluasi Performa Model Setiap Aspek

Hasil pengujian menunjukkan bahwa model BERT yang telah dilatih menggunakan konfigurasi *hyperparameter* yang telah ditentukan berhasil mencapai akurasi 77%. Setelah model digunakan untuk *testing* pada data baru, didapatkan evaluasi performa model klasifikasi menggunakan metrik-metrik seperti *precision*, *recall*, dan *f1-score* untuk setiap aspek yang ada dalam data.

Tabel 4. Metrik Evaluasi Model Setiap Aspek

Aspek	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Biaya	0.82	0.74	0.78
Keselamatan	0.89	1.00	0.94
Ramah Lingkungan	0.78	0.90	0.83
Fitur	0.70	0.63	0.66
Reliabilitas	0.69	0.74	0.72
Lainnya	0.76	0.63	0.69

Pada Tabel 4 di atas, dapat dilihat bahwa Model evaluasi menunjukkan performa yang beragam dalam mengklasifikasikan berbagai aspek otomotif. Aspek Keselamatan memperoleh nilai tertinggi dengan presisi 0.89 dan *recall* 1.00, menandakan bahwa model dapat dengan sangat baik mengidentifikasi dan mengingatkan terhadap contoh-contoh yang termasuk dalam kategori Keselamatan. Sementara itu, aspek Biaya dan Ramah Lingkungan juga menunjukkan hasil yang baik dengan presisi masing-masing 0.82 dan 0.78, serta *recall* 0.74 dan 0.90. Namun, beberapa aspek lain seperti Fitur, Reliabilitas, dan Lainnya menunjukkan tantangan dengan nilai *recall* yang lebih rendah, mengindikasikan bahwa model mungkin mengalami kesulitan dalam mengidentifikasi semua contoh yang sebenarnya masuk dalam kategori-kategori tersebut. Keseluruhan, nilai *F1-score* yang diperoleh mencerminkan keseimbangan antara presisi dan *recall* untuk masing-masing aspek, menunjukkan keefektifan model dalam melakukan klasifikasi aspek-aspek otomotif yang berbeda.

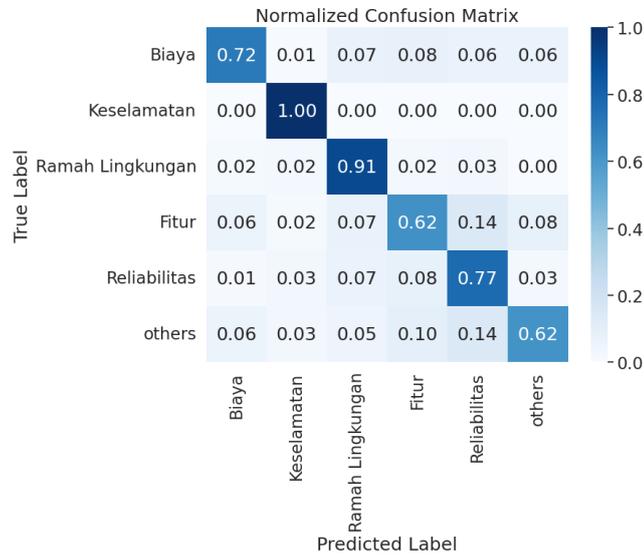


Gambar 1. Kurva ROC

Pada Gambar 1, dapat dilihat bahwa hasil evaluasi AUC (*Area Under Curve*) untuk berbagai aspek otomotif pada kurva ROC di atas menunjukkan kemampuan model dalam membedakan kelas-kelas yang berbeda. Aspek Keselamatan mendapatkan AUC sempurna yaitu 1.00. Hal ini menunjukkan bahwa model secara sempurna memisahkan artikel media massa yang termasuk dalam aspek Keselamatan dari aspek lainnya. Aspek Biaya dan Reliabilitas juga menunjukkan AUC yang tinggi dengan nilai masing-masing 0.93 dan 0.94. Hal ini menunjukkan kemampuan model dalam mengklasifikasikan dengan baik antara artikel-artikel media massa yang relevan dengan aspek-aspek tersebut dan yang tidak. AUC yang lebih rendah pada aspek Fitur dan Lainnya, meskipun masih cukup baik dengan nilai masing-masing 0.88 dan 0.90. Hal ini menunjukkan beberapa tantangan dalam membedakan aspek-aspek ini. Secara keseluruhan, hasil evaluasi AUC memberikan gambaran tentang seberapa baik model dapat memprediksi dan mengklasifikasikan berbagai aspek otomotif dengan akurasi yang baik.

3.5 Confusion Matrix

Pada Gambar 2, *Confusion matrix* menunjukkan kinerja model dalam mengklasifikasikan enam aspek yaitu, Biaya, Keselamatan, Ramah Lingkungan, Fitur, Reliabilitas, dan Lainnya. Setiap sel menunjukkan persentase prediksi yang benar atau salah dari model untuk aspek tertentu, dengan nilai yang ter-normalisasi. Diagonal dari kiri atas ke kanan bawah mewakili prediksi yang benar. Pada aspek Biaya, model memprediksi dengan benar 72%, sementara untuk Keselamatan, akurasi prediksinya mencapai 100%. Namun, ada juga beberapa kesalahan, seperti Fitur yang sering salah diprediksi sebagai Ramah Lingkungan dan sebaliknya. Secara keseluruhan, model memiliki kinerja yang baik di beberapa aspek seperti Keselamatan dan Reliabilitas.



Gambar 2. Confusion Matrix

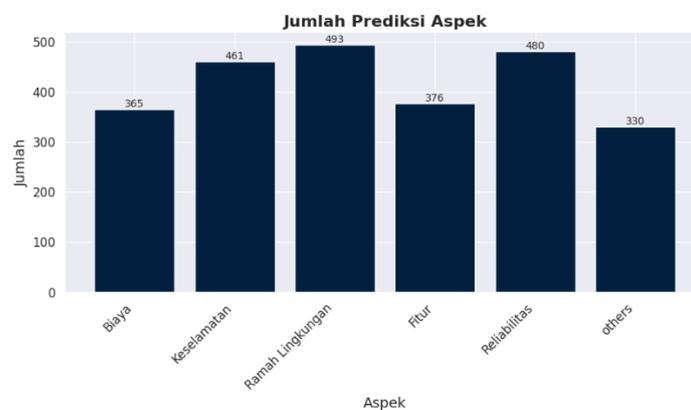
3.6 Cross Validation

Tabel 5. Cross Validation

Fold	Akurasi
1	72%
2	70%
3	69%
4	69%
5	68%
average	69%

Pada tabel 5 di atas, dapat dilihat bahwa hasil cross validation menunjukkan bahwa model memiliki variasi akurasi di setiap fold. Akurasi pada fold pertama adalah 72%, fold kedua 70%, fold ketiga 69%, fold keempat 69%, dan fold kelima 68%. Rata-rata akurasi dari kelima fold ini adalah 69% yang menunjukkan bahwa model memiliki performa yang konsisten dan cukup baik dalam memprediksi distribusi aspek otomotif di seluruh dataset yang digunakan.

3.7 Distribusi Prediksi Klasifikasi Aspek



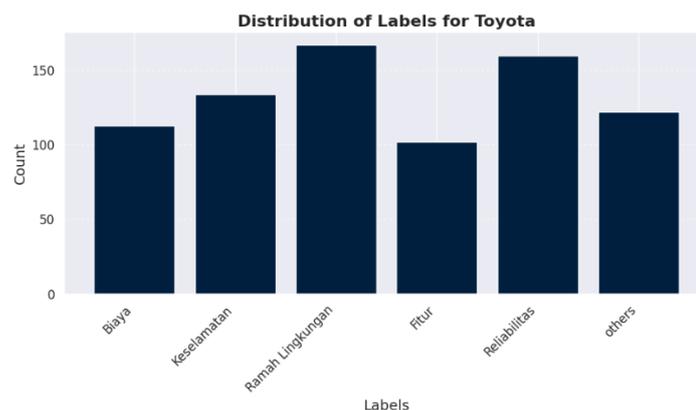
Gambar 3. Distribusi Setiap Aspek



Dari gambar 3 di atas, dapat disimpulkan bahwa hasil distribusi aspek otomotif pada beberapa media massa otomotif Indonesia menunjukkan bahwa aspek Ramah Lingkungan memiliki jumlah prediksi tertinggi dengan 493, diikuti oleh Reliabilitas dengan 480, Keselamatan dengan 461, Fitur dengan 376, Biaya dengan 365, dan Lainnya dengan 330.

Distribusi ini mencerminkan tren pasar dan preferensi konsumen saat ini. Tingginya perhatian pada aspek ramah lingkungan mencerminkan kesadaran yang semakin meningkat terhadap isu-isu lingkungan dan permintaan akan kendaraan yang lebih bersih dan hemat energi. Konsumen kini lebih peduli terhadap dampak lingkungan dari kendaraan mereka, mendorong industri otomotif untuk berinovasi dalam teknologi hijau. Reliabilitas yang juga tinggi menunjukkan bahwa konsumen menginginkan kendaraan yang dapat diandalkan dan memiliki umur panjang, yang penting untuk mengurangi biaya perawatan dan menjamin keamanan. Aspek keselamatan yang berada di posisi ketiga menggarisbawahi pentingnya fitur keselamatan bagi konsumen, yang mungkin didorong oleh meningkatnya kesadaran akan keselamatan berkendara. Sementara itu, aspek fitur menyoroti keinginan konsumen akan teknologi dan kenyamanan tambahan dalam kendaraan mereka. Biaya yang berada di urutan kelima menunjukkan bahwa meskipun penting, konsumen lebih cenderung memprioritaskan faktor lingkungan, reliabilitas, dan keselamatan terlebih dahulu. Akhirnya, kategori lainnya menunjukkan bahwa terdapat beberapa aspek tambahan yang juga menjadi perhatian, meskipun tidak sepopuler lima aspek utama tersebut.

Secara keseluruhan, distribusi ini mencerminkan dinamika pasar otomotif di Indonesia, di mana isu lingkungan, keandalan, dan keselamatan kendaraan menjadi prioritas utama bagi konsumen, diikuti oleh ketertarikan terhadap fitur-fitur modern dan pertimbangan biaya yang efektif. Tren ini menunjukkan bahwa produsen kendaraan harus terus berinovasi dalam teknologi ramah lingkungan dan fitur keselamatan, sambil tetap menjaga biaya yang kompetitif untuk memenuhi harapan konsumen.



Gambar 4. Distribusi Setiap Aspek pada Toyota

Pada gambar 4 di atas, dapat dilihat bahwa distribusi aspek otomotif untuk merek Toyota pada beberapa media massa otomotif di Indonesia menunjukkan bahwa Ramah Lingkungan adalah aspek yang paling sering dibahas. Diikuti oleh Reliabilitas, Keselamatan, Biaya, Fitur, dan Lainnya.

Tingginya perhatian terhadap aspek Ramah Lingkungan mencerminkan upaya Toyota dalam mengembangkan teknologi *hybrid* dan mobil listrik yang ramah lingkungan, sejalan dengan tren global yang semakin mengedepankan isu keberlanjutan dan dampak lingkungan. Konsumen semakin menyadari pentingnya kendaraan yang tidak hanya efisien dalam penggunaan bahan bakar, tetapi juga



memiliki emisi yang rendah sesuai dengan kebijakan pemerintah yang mendukung kendaraan ramah lingkungan.

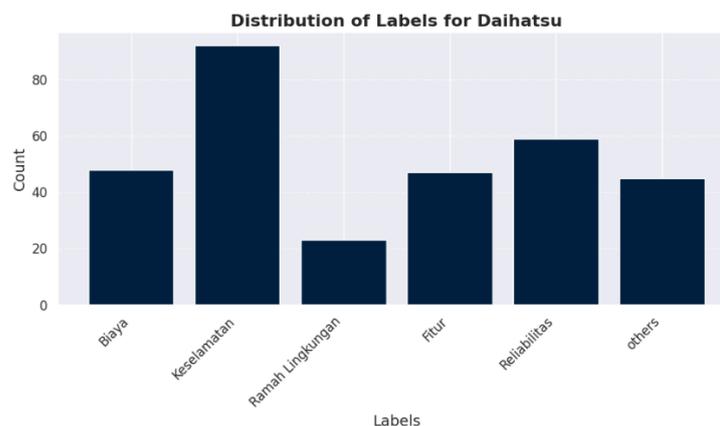
Reliabilitas yang menduduki posisi kedua menunjukkan bahwa konsumen sangat menghargai keandalan mobil Toyota. Produk-produk otomotif Toyota telah dipercaya sebagai produsen kendaraan yang tahan lama dan minim masalah teknis menjadikannya pilihan utama bagi banyak konsumen yang mengutamakan keamanan jangka panjang.

Aspek Keselamatan juga mendapat perhatian besar, menandakan bahwa konsumen sangat peduli dengan fitur keselamatan yang ditawarkan oleh Toyota. Peningkatan fitur keselamatan seperti sistem pengereman canggih, *airbag*, dan teknologi bantuan pengemudi menjadi nilai tambah yang baik.

Aspek Biaya yang cukup tinggi menunjukkan bahwa konsumen juga sangat mempertimbangkan harga pembelian, biaya perawatan, dan efisiensi bahan bakar. Kendaraan Toyota dikenal dengan efisiensi bahan bakarnya yang baik yang menjadi faktor penting bagi konsumen di pasar yang sensitif terhadap harga.

Sementara aspek Fitur yang mendapat perhatian sedikit lebih rendah mencerminkan bahwa meskipun fitur teknologi dan kenyamanan penting, mereka mungkin tidak seutama aspek lain seperti lingkungan, keandalan, dan keselamatan dalam keputusan pembelian mobil Toyota. Aspek Lainnya yang cukup banyak dibahas menunjukkan bahwa terdapat faktor-faktor tambahan yang juga mempengaruhi persepsi konsumen terhadap merek Toyota.

Secara keseluruhan, distribusi ini mencerminkan bahwa Toyota berhasil menarik perhatian konsumen melalui inovasi teknologi ramah lingkungan, keandalan, dan keselamatan dengan tetap menjaga efisiensi biaya yang kompetitif. Tren ini menunjukkan bahwa untuk tetap kompetitif, Toyota harus terus berinovasi dalam teknologi hijau dan keselamatan, sambil menjaga reputasi mereka sebagai produsen mobil yang andal dan hemat biaya.



Gambar 5. Distribusi Setiap Aspek pada Daihatsu

Gambar 5 di atas menunjukkan bahwa distribusi aspek otomotif untuk merek Daihatsu pada beberapa media massa otomotif di Indonesia menunjukkan bahwa Keselamatan adalah aspek yang paling sering dibahas. Diikuti oleh Reliabilitas, Biaya, Fitur, Lainnya, dan Ramah Lingkungan.

Tingginya perhatian terhadap aspek Keselamatan mencerminkan fokus konsumen pada keamanan berkendara yang menjadi prioritas utama saat memilih mobil Daihatsu. Hal ini bisa disebabkan oleh meningkatnya kesadaran akan pentingnya fitur keselamatan.

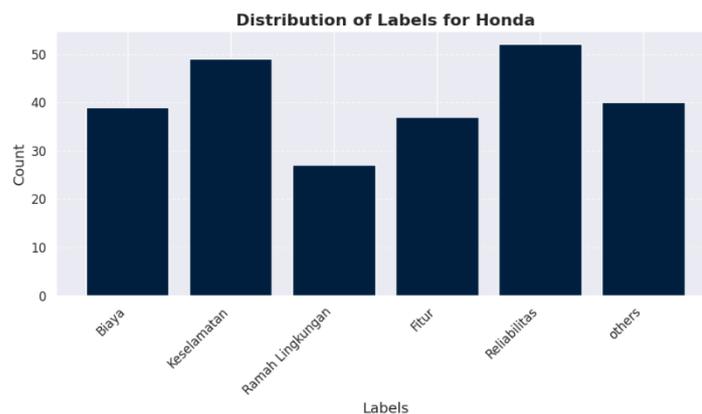
Aspek Reliabilitas yang berada di posisi kedua menunjukkan bahwa konsumen sangat menghargai keandalan mobil Daihatsu. Daihatsu dikenal sebagai produsen kendaraan yang andal dan

minim masalah teknis membuatnya menarik bagi konsumen yang mencari mobil dengan performa yang stabil.

Aspek Biaya dan Fitur yang hampir seimbang menunjukkan bahwa meskipun harga dan fitur teknologi penting, mereka mungkin tidak seutama aspek keselamatan dan keandalan dalam keputusan pembelian mobil Daihatsu. Kendaraan Daihatsu dikenal dengan harga yang terjangkau dan efisiensi biaya yang baik dimana ini menjadi faktor yang penting bagi konsumen.

Rendahnya perhatian terhadap aspek Ramah Lingkungan menunjukkan bahwa saat ini, konsumen Daihatsu mungkin kurang memperhatikan atau kurang melihat inovasi signifikan dari Daihatsu dalam hal teknologi hijau dan ramah lingkungan. Ini bisa menjadi area potensial bagi Daihatsu untuk mengembangkan dan mempromosikan lebih lanjut kendaraan yang lebih efisien dan rendah emisi.

Secara keseluruhan, distribusi ini mencerminkan bahwa Daihatsu berhasil menarik perhatian konsumen melalui keandalan dan keselamatan, sambil tetap menjaga efisiensi biaya yang kompetitif. Tren ini menunjukkan bahwa untuk tetap kompetitif, Daihatsu harus terus berinovasi dalam teknologi keselamatan dan keandalan, sambil mengeksplorasi peluang dalam teknologi ramah lingkungan untuk memenuhi harapan konsumen yang semakin peduli terhadap isu lingkungan.



Gambar 6. Distribusi Setiap Aspek pada Honda

Pada gambar 6, dapat dilihat bahwa distribusi aspek otomotif untuk merek Honda pada beberapa media massa otomotif di Indonesia menunjukkan bahwa Reliabilitas adalah aspek yang paling sering dibahas. Diikuti oleh Keselamatan, Lainnya, Fitur, Biaya, dan Ramah Lingkungan.

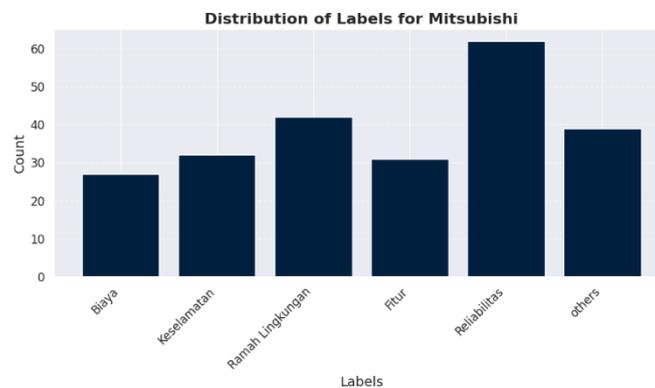
Tingginya perhatian terhadap aspek Reliabilitas menunjukkan bahwa konsumen sangat menghargai keandalan kendaraan Honda. Honda dikenal sebagai produsen kendaraan yang terpercaya membuatnya menarik bagi konsumen yang mencari kendaraan yang bisa diandalkan dalam jangka panjang.

Aspek Keselamatan yang berada di posisi kedua menandakan bahwa konsumen sangat peduli dengan fitur keselamatan yang ditawarkan oleh Honda. Honda dikenal dengan komitmennya terhadap keselamatan melalui berbagai inovasi. Hal ini sejalan dengan peningkatan kesadaran konsumen akan pentingnya keselamatan berkendara. Aspek “Lainnya” yang banyak dibahas menunjukkan bahwa terdapat faktor-faktor tambahan yang juga mempengaruhi persepsi konsumen terhadap merek Honda.

Aspek Fitur dan Biaya yang hampir seimbang menunjukkan bahwa meskipun fitur teknologi dan harga penting, mereka mungkin tidak seutama keandalan dan keselamatan dalam keputusan pembelian mobil Honda. Konsumen Honda mungkin melihat nilai dalam fitur teknologi canggih dan efisiensi biaya, tetapi tetap mengutamakan keamanan dan keandalan.

Rendahnya perhatian terhadap aspek Ramah Lingkungan menunjukkan bahwa saat ini, konsumen Honda mungkin kurang memperhatikan atau kurang melihat inovasi signifikan dari Honda dalam hal teknologi hijau dan ramah lingkungan. Ini bisa menjadi area potensial bagi Honda untuk mengembangkan dan mempromosikan lebih lanjut kendaraan yang lebih efisien dan rendah emisi.

Secara keseluruhan, distribusi ini mencerminkan bahwa Honda berhasil menarik perhatian konsumen melalui keandalan dan keselamatan, sambil tetap menjaga efisiensi biaya dan fitur teknologi yang kompetitif. Tren ini menunjukkan bahwa untuk tetap kompetitif, Honda harus terus berinovasi dalam teknologi keselamatan dan keandalan dengan tetap mengeksplorasi peluang dalam teknologi ramah lingkungan untuk memenuhi harapan konsumen yang semakin peduli terhadap isu lingkungan.



Gambar 7. Distribusi Setiap Aspek pada Mitsubishi

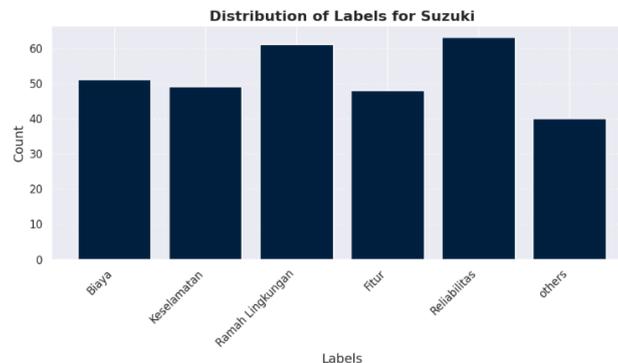
Gambar 7 di atas menampilkan distribusi aspek otomotif untuk merek Mitsubishi pada beberapa media massa otomotif di Indonesia menunjukkan bahwa Reliabilitas adalah aspek yang paling sering dibahas. Diikuti oleh Ramah Lingkungan, Lainnya, Keselamatan, Fitur, dan Biaya.

Tingginya perhatian terhadap aspek Reliabilitas menunjukkan bahwa konsumen sangat menghargai keandalan mobil Mitsubishi. Selanjutnya, aspek Ramah Lingkungan yang berada di posisi kedua menandakan bahwa Mitsubishi mungkin fokus pada pengembangan teknologi ramah lingkungan. Konsumen yang semakin peduli terhadap isu lingkungan cenderung memilih mobil dengan performa ramah lingkungan sehingga Mitsubishi dapat menarik segmen pasar ini dengan strategi ini.

Aspek Keselamatan yang mendapat perhatian lebih rendah mungkin menunjukkan bahwa sementara konsumen menginginkan fitur keselamatan, faktor-faktor lain seperti keandalan dan ramah lingkungan mungkin lebih dominan dalam keputusan pembelian mereka.

Aspek Biaya yang paling sedikit dibahas menunjukkan bahwa Mitsubishi mungkin tidak menonjolkan diri dalam segi harga yang bersaing, atau konsumen Mitsubishi mungkin lebih memilih untuk fokus pada keandalan dan fitur lainnya daripada harga.

Secara keseluruhan, distribusi ini mencerminkan bahwa Mitsubishi berhasil menarik perhatian konsumen dengan fokus pada keandalan dan ramah lingkungan. Tren ini menunjukkan bahwa untuk tetap relevan dan bersaing di pasar otomotif yang semakin kompetitif, Mitsubishi harus terus berinovasi dalam teknologi keandalan dan ramah lingkungan sekaligus mempertahankan nilai tambah dari faktor-faktor lain yang penting bagi konsumen.



Gambar 8. *Distribusi Setiap Aspek pada Suzuki*

Pada gambar 8 di atas, dapat ditarik kesimpulan bahwa distribusi aspek otomotif untuk merek Suzuki pada beberapa media massa otomotif di Indonesia menunjukkan bahwa fokus utama ada pada Reliabilitas. Diikuti oleh Ramah Lingkungan, Biaya, Keselamatan, Fitur, dan Lainnya.

Tingginya perhatian terhadap aspek Reliabilitas mencerminkan reputasi Suzuki sebagai produsen kendaraan yang handal menjadi faktor utama bagi konsumen dalam mencari kendaraan yang dapat diandalkan dalam jangka panjang.

Aspek Ramah Lingkungan yang mendapat perhatian cukup tinggi menunjukkan bahwa Suzuki mungkin telah fokus pada pengembangan teknologi ramah lingkungan. Konsumen yang semakin peduli terhadap lingkungan cenderung memilih kendaraan yang lebih ramah lingkungan, sehingga Suzuki bisa mendapatkan keunggulan kompetitif di pasar ini.

Perhatian yang signifikan terhadap aspek Biaya menunjukkan bahwa Suzuki mungkin menawarkan harga yang kompetitif atau nilai tambah yang baik dalam segi biaya operasional kendaraan. Hal ini menarik bagi konsumen yang mempertimbangkan efisiensi biaya dalam keputusan pembelian kendaraan.

Meskipun Keselamatan dan Fitur juga penting, tingkat perhatian yang sedikit lebih rendah mungkin menunjukkan bahwa konsumen Suzuki mungkin melihat aspek-aspek lain seperti keandalan dan ramah lingkungan lebih dominan dalam keputusan pembelian mereka.

Secara keseluruhan, distribusi ini mencerminkan strategi Suzuki dalam memenuhi preferensi konsumen yang mencari keandalan, efisiensi biaya, dan semakin peduli terhadap lingkungan. Untuk tetap bersaing di pasar otomotif yang dinamis, Suzuki perlu terus berinovasi dalam teknologi keandalan dan ramah lingkungan.

IV. KESIMPULAN

Distribusi aspek otomotif pada media massa Indonesia menunjukkan bahwa setiap merek kendaraan memiliki fokus dan prioritas yang berbeda berdasarkan preferensi konsumen dan tren pasar. Untuk keseluruhan industri, aspek Ramah Lingkungan menjadi perhatian utama, mencerminkan meningkatnya kesadaran konsumen terhadap isu lingkungan dan permintaan akan kendaraan yang lebih bersih. Reliabilitas dan Keselamatan juga sangat penting, menunjukkan bahwa konsumen menghargai kendaraan yang dapat diandalkan dan aman.

Toyota lebih banyak dibicarakan dalam hal Ramah Lingkungan, mencerminkan inovasi mereka dalam teknologi hijau seperti mobil *hybrid* dan listrik. Konsumen Toyota juga menghargai keandalan dan keselamatan kendaraan mereka. Sebaliknya, Daihatsu lebih fokus pada aspek Keselamatan dan Reliabilitas, dengan perhatian lebih rendah pada Ramah Lingkungan, mungkin karena fokus mereka



Honda dan Mitsubishi mendominasi pada aspek Reliabilitas, dengan Honda juga memiliki fokus kuat pada Keselamatan, sementara Mitsubishi menunjukkan perhatian besar pada teknologi Ramah Lingkungan. Suzuki dikenal karena keandalannya dan efisiensi biaya, serta mulai mendominasi dalam aspek Ramah Lingkungan.

Perbedaan ini mencerminkan bagaimana setiap merek merespons tren pasar dan preferensi konsumen yang berbeda. Sebagai contoh, peningkatan kesadaran akan isu lingkungan mendorong merek-merek seperti Toyota dan Mitsubishi untuk fokus pada teknologi ramah lingkungan. Sementara itu, konsumen yang mencari kendaraan andal dengan harga terjangkau cenderung memilih Daihatsu atau Suzuki. Inovasi dalam fitur keselamatan juga menjadi prioritas penting seiring dengan meningkatnya kesadaran akan pentingnya keselamatan berkendara.

Temuan ini memberikan wawasan berharga bagi produsen otomotif untuk lebih menekankan aspek-aspek yang penting bagi konsumen dalam strategi pengembangan dan pemasaran produk mereka. Dengan memahami prioritas konsumen maka produsen dapat fokus mengembangkan kendaraan yang sesuai dengan tren pasar saat ini. Ini tidak hanya membantu dalam memenuhi kebutuhan dan preferensi konsumen, tetapi juga memungkinkan produsen untuk bersaing lebih efektif di pasar yang semakin kompetitif.

REFERENSI

1. K. Emir and S. Nuno, “Peningkatan Daya Saing Dan Pangsa Pasar Dalam Bisnis,” vol. 23, no. 1, pp. 65–74, 2021.
2. Rachel Nasywa Aurelia, Centurion Chandratama Priyatna, and Hanny Hafiar, “Tinjauan Brand Monitoring terhadap Pemberitaan Motor Listrik Brand Alva,” *Harmon. J. Ilmu Komun. dan Sos.*, vol. 2, no. 2, pp. 70–83, 2024, doi: 10.59581/harmoni-widyakarya.v2i2.3051.
3. R. N. T. Sutanto, “Fast Knowledge Discovery in Social Media Data using Clustering via Ranking,” *Int. Conf. Cyber IT Serv. Manag.*, pp. 1–8, 2021, doi: 10.1109/CITSM52892.2021.9588866.
4. Y. Peng, S. Yan, and Z. Lu, “Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets,” *BioNLP 2019 - SIGBioMed Work. Biomed. Nat. Lang. Process. Proc. 18th BioNLP Work. Shar. Task*, no. iv, pp. 58–65, 2019, doi: 10.18653/v1/w19-5006.
5. R. Merdiansah, S. Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.
6. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 4171–4186, 2019.
7. T. Iskandar Zulkarnain Maulana Putra, A. Farhan Bukhori, dan Ilmu Pengetahuan Alam, and U. Gadjah Mada, “Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long ,” *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022, [Online]. Available: <https://doi.org/10.35912/jisted.v1i1.1509>
8. Y. D. Pramudita, S. S. Putro, and N. Makhmud, “Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 269–276, 2018, doi: 10.25126/jtiik.201853810.
9. M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, “Indonesian news classification using naïve bayes and two-phase feature selection model,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 3, pp. 610–615, 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.
10. F. zahra El-Alami, S. Ouatik El Alaoui, and N. En Nahnahi, “Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8422–8428, 2022, doi: 10.1016/j.jksuci.2021.02.005.



SENADA
Seminar Nasional Sains Data

Seminar Nasional Sains Data 2024 (SENADA 2024)
UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

11. S. Imron, E. I. Setiawan, and J. Santoso, “Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN,” *J. Intell. Syst. Comput.*, vol. 5, no. 1, pp. 10–16, 2023, doi: 10.52985/insyst.v5i1.267.
12. S. Sharma, “Epoch vs Batch Size vs Iterations,” Medium. Accessed: Jul. 06, 2024. [Online]. Available: <https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9>
13. N. Rochmawati, H. B. Hidayati, Y. Yamasari, H. P. A. Tjahyaningtjas, W. Yustanti, and A. Prihanto, “Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 44–48, 2021, doi: 10.26740/jieet.v5n2.p44-48.