



Prediksi Perilaku Konsumtif Remaja Menggunakan Algoritma Catboost Berbasis Machine Learning

Sifa Aurahman^{a,1}, Umi Mahmudah^{b,2*}

^{a,b} Sains Data, UIN K.H. Abdurrahman Wahid Pekalongan, Pekalongan, Indonesia

¹ sifa.aurahman24002@mhs.uingusdur.ac.id; ² umi.mahmudah@uingusdur.ac.id*

* corresponding author

ABSTRAK

ARTICLE INFO

Perilaku konsumtif pada remaja menjadi fenomena yang semakin marak dan berpotensi menimbulkan dampak negatif jangka panjang, seperti berkurangnya sikap hemat dan produktif. Ciri perilaku ini antara lain kebiasaan boros, pengeluaran berlebihan untuk memenuhi keinginan, serta mengikuti tren gaya hidup yang sedang berkembang. Penelitian ini bertujuan memprediksi perilaku konsumtif remaja berdasarkan faktor gaya hidup, penggunaan media sosial, dan lingkungan sosial menggunakan pendekatan *machine learning*. Metode yang digunakan adalah *supervised learning* dengan algoritma Categorical Boosting (CatBoost), yang mampu mengelola data kategorikal secara efisien dan merupakan pengembangan dari *Gradient Boosting Decision Tree* (GBDT). Model dibangun dengan mempelajari fitur-fitur yang relevan untuk mengklasifikasikan tingkat perilaku konsumtif. Hasil penelitian menunjukkan bahwa gaya hidup, media sosial, dan lingkungan sosial memiliki pengaruh signifikan terhadap perilaku konsumtif remaja, dengan tingkat akurasi prediksi sebesar 91,8% dan nilai AUC sebesar 0,93. Temuan ini diharapkan dapat menjadi dasar pengembangan strategi pencegahan perilaku konsumtif di kalangan remaja.

Article history

Received: 15 Agustus 2025

Revised: 11 September 2025

Accepted: 24 Oktober 2025

Keywords

perilaku konsumtif

gaya hidup

media sosial

machine learning

CatBoost

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Perilaku konsumtif di kalangan remaja telah berkembang melampaui pemenuhan kebutuhan dasar menjadi medium ekspresi identitas, pencarian status, dan perilaku materialistic [1], [2]. Remaja, yang sedang menjalani proses pencarian jati diri (*identity formation*), menjadi segmentasi konsumen yang sangat rentan terhadap tekanan sosial, tren, serta eksposur digital [3].

Peran media sosial dalam proses sosialisasi konsumsi semakin dominan, terutama dalam memfasilitasi interaksi parasosial dengan influencer yang merepresentasikan gaya hidup ideal dan materialistic [2], [4]. Interaksi semu semacam ini bukan hanya memengaruhi minat beli, tetapi juga memelihara aspirasi, perasaan kurang diri (*inadequacy*), dan legitimisasi konsumsi impulsif [5]. Selain itu, penelitian di Indonesia menunjukkan bahwa media sosial secara signifikan membentuk gaya hidup remaja, yang kemudian menjadi mediator penting dalam menjembatani pengaruh sosial seperti pendidikan ekonomi keluarga dan konformitas teman sebaya terhadap perilaku konsumsi [6]. Penggunaan media sosial yang intensif juga terbukti berdampak buruk pada gaya hidup sehat dan kepercayaan terhadap nilai-nilai sehat. Sebagai contoh, kecanduan media sosial pada remaja



menjelaskan hingga 8,9% varians dalam kepercayaan terhadap gaya hidup sehat—indikator perlunya kontrol serta intervensi di ranah keluarga dan sekolah [7].

Dampak konsumtif tersebut tidak hanya bersifat perilaku, tetapi juga memiliki implikasi psikologis dan sosial [8], [9]. Eksposur berulang terhadap konten "flexing" atau gaya hidup glamor berlebihan, terutama dari influencer dan YouTuber, bisa memicu perasaan rendah diri, kompetisi sosial, hingga masalah mental seperti kecemasan [10]. *Overconsumption* juga berpotensi merusak kesehatan mental dan menggerus kepuasan hidup, yang ironisnya tidak seimbang dengan peningkatan kepemilikan materi [11].

Sebagian besar kajian perilaku konsumtif sebelumnya mengandalkan metode survei dan analisis statistik linier, seperti regresi atau analisis korelasi. Pendekatan ini efektif untuk menguji hubungan antarvariabel secara parsial, namun memiliki keterbatasan ketika berhadapan dengan data kategorikal yang kompleks, interaksi antarvariabel yang bersifat non-linear, dan pola perilaku yang dinamis [12]. Model linier umumnya mengasumsikan hubungan yang stabil dan proporsional, padahal perilaku konsumtif remaja sering kali dipengaruhi oleh interaksi multi-faktor yang tidak bersifat linier, misalnya pengaruh simultan gaya hidup, intensitas penggunaan media sosial, dan tekanan lingkungan sosial. Dalam konteks tersebut, *machine learning* menawarkan pendekatan yang lebih adaptif, mampu menangkap pola kompleks tanpa memerlukan asumsi distribusi data yang ketat, serta memiliki kemampuan *feature interaction learning* yang lebih tinggi [13]–[15]. Salah satu kelompok algoritma *machine learning* yang relevan adalah *boosting*, yang bekerja dengan menggabungkan banyak *weak learners* (umumnya *decision trees*) secara iteratif untuk meningkatkan kinerja prediksi [15]–[18].

Di antara berbagai algoritma *boosting*, Categorical Boosting (CatBoost) menonjol karena dirancang khusus untuk menangani variabel kategorikal tanpa memerlukan proses *one-hot encoding* yang memakan sumber daya [19], [20]. CatBoost mengimplementasikan ordered boosting untuk mengurangi *prediction shift* dan pohon simetris untuk mempercepat pelatihan serta menjaga konsistensi struktur model [21], [22]. Keunggulan ini menjadikannya lebih efisien dibandingkan algoritma *boosting* lain seperti XGBoost atau LightGBM, terutama dalam dataset dengan proporsi besar variabel kategorikal dan ukuran data yang masif.

Secara empiris, kinerja CatBoost telah diuji dalam berbagai konteks. Misalnya, sebuah penelitian melaporkan bahwa CatBoost mampu mencapai skor prediksi yang sangat tinggi (ROC AUC $\approx 0,985$) dalam klasifikasi perilaku pembelian konsumen, setara atau bahkan melampaui XGBoost dan *Support Vector Machine* [23]. Dalam studi lain pada bidang ekonomi pembangunan di Filipina, penelitian menunjukkan bahwa CatBoost mengungguli algoritma pembandingan dengan skor F1 mencapai 91% dan efisiensi pengujian yang sangat tinggi, menjadikannya pilihan ideal untuk aplikasi prediktif berskala besar [16]. Selain akurasi, CatBoost juga memiliki keunggulan dalam interpretabilitas model. Melalui fitur *model visualization* seperti *feature importance* dan *SHAP values*, peneliti dapat mengidentifikasi faktor-faktor paling berpengaruh terhadap prediksi, yang relevan untuk menjawab pertanyaan penelitian secara substantif [24], [25]. Kemampuan ini sangat penting dalam studi perilaku konsumtif remaja, karena tidak hanya menyediakan model prediktif yang akurat, tetapi juga memberikan wawasan mengenai variabel-variabel kunci yang dapat menjadi sasaran intervensi.

Meskipun potensinya besar, penerapan CatBoost untuk memprediksi perilaku konsumtif remaja—terutama dengan mengintegrasikan variabel gaya hidup, media sosial, dan lingkungan sosial secara simultan—masih jarang dilakukan. Sebagian besar studi terdahulu hanya berfokus pada salah satu faktor atau menggunakan metode prediksi konvensional yang terbatas dalam menangkap kompleksitas hubungan antarvariabel. Oleh karena itu, penelitian ini tidak hanya bertujuan membangun model prediksi berbasis CatBoost, tetapi juga mengisi kekosongan literatur dengan menyediakan analisis komprehensif yang dapat memandu kebijakan dan strategi pengendalian perilaku konsumtif di kalangan remaja.

Sebagian besar kajian perilaku konsumtif sebelumnya mengandalkan survei konvensional dengan analisis linier. Metode tersebut memiliki keterbatasan ketika berhadapan dengan data kategorikal yang kompleks dan interaksi non-linear antarvariabel. Penelitian ini juga menggunakan survei, namun dengan pendekatan berbeda. Pertama, desain kuesioner disusun secara lebih terstruktur dengan fokus pada variabel kategorikal yang dapat langsung diolah oleh CatBoost. Kedua, variabel dioperasionalisasi ke dalam indikator-indikator terukur, sehingga data yang dihasilkan lebih sesuai untuk pemodelan *machine learning*. Ketiga, analisis menggunakan CatBoost memungkinkan mengatasi kelemahan survei konvensional dengan menangkap pola kompleks tanpa asumsi distribusi

data yang ketat. Dengan demikian, penelitian ini mengombinasikan metode survei dengan analisis prediktif modern untuk memperoleh hasil yang lebih komprehensif.

2. Method

Penelitian ini merupakan penelitian kuantitatif berbasis analisis prediksi yang dilaksanakan melalui serangkaian tahapan sistematis untuk memastikan hasil yang diperoleh akurat, dapat diinterpretasikan dengan baik, dan memiliki tingkat keandalan tinggi. Tahapan meliputi: pengumpulan data, pemrosesan data, pembagian data, pemodelan menggunakan algoritma CatBoost, dan evaluasi model. Pendekatan ini digunakan untuk memahami pola perilaku konsumtif remaja berdasarkan faktor gaya hidup, penggunaan media sosial, dan lingkungan sosial.

2.1. Pengumpulan Data

Data diperoleh melalui kuesioner yang disebarakan kepada remaja di Desa Sidosari, Kecamatan Kesesi, Kabupaten Pekalongan. Instrumen penelitian dirancang untuk menggali informasi terkait variabel gaya hidup, intensitas penggunaan media sosial, dan pengaruh lingkungan sosial terhadap perilaku konsumtif. Variabel Gaya hidup diukur melalui frekuensi belanja per minggu, jenis produk dominan yang dibeli, dan rata-rata pengeluaran bulanan, variabel media sosial diukur melalui durasi penggunaan per hari, jumlah platform yang aktif digunakan, serta jumlah akun influencer yang diikuti. Sedangkan variabel Lingkungan sosial diukur melalui pengaruh teman sebaya, intensitas diskusi konsumsi dengan keluarga, dan keterlibatan kelompok sosial dalam keputusan pembelian. Variabel-variabel tersebut dikodekan menjadi data kategorikal/numerik agar sesuai dengan input yang dapat diproses CatBoost. Data yang terkumpul direkap, diperiksa kelengkapannya, dan diseleksi untuk memastikan validitasnya sebelum masuk tahap analisis.

2.2. Pemrosesan Data

Data yang telah terkumpul diproses agar siap digunakan dalam pemodelan. Tahapan yang dilakukan meliputi: a) Pembersihan Data, yaitu menghapus data tidak lengkap, duplikat, atau tidak valid; b) Penyeimbangan Kelas, yaitu jika terjadi ketidakseimbangan pada kategori target, digunakan teknik *resampling* untuk menyeimbangkan distribusi; c) Seleksi Fitur, yaitu mengidentifikasi dan memilih variabel paling relevan terhadap perilaku konsumtif melalui analisis awal; d). Pengolahan Data Kategorikal, yaitu mengonversi variabel kategorikal ke format numerik menggunakan teknik internal CatBoost yang menghindari *one-hot encoding*, sehingga efisien dan minim kehilangan informasi.

2.3. Pembagian Data

Data yang telah diproses dibagi menjadi: a). Training Data (80%), yaitu digunakan untuk membangun dan menyesuaikan model; b). Testing Data (20%), yaitu digunakan untuk menguji kemampuan model memprediksi data baru yang belum pernah dilihat sebelumnya.

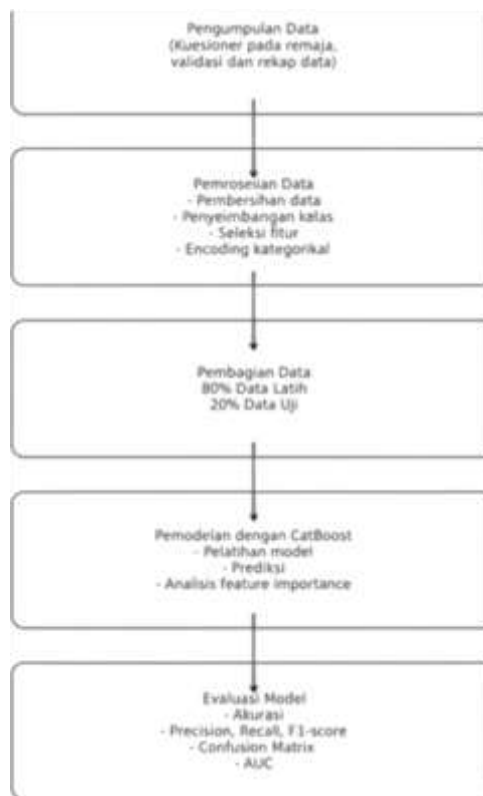
2.4. Pemodelan dengan CatBoost

Algoritma Categorical Boosting (CatBoost) digunakan karena kemampuannya menangani variabel kategorikal secara langsung, efisiensi komputasi tinggi, serta minim risiko *overfitting*. Proses pemodelan meliputi: a). Pelatihan Model, dimana model dilatih menggunakan data latih dengan penyesuaian hiperparameter (*hyperparameter tuning*) untuk mendapatkan kinerja optimal; b). Prediksi, yaitu model memprediksi perilaku konsumtif remaja berdasarkan variabel input yang telah ditentukan; c). Analisis Feature Importance, yaitu mengidentifikasi kontribusi relatif masing-masing variabel terhadap hasil prediksi untuk memberikan interpretasi yang lebih bermakna.

2.5. Evaluasi Model

Kinerja model dievaluasi menggunakan beberapa metrik: a. Akurasi, yaitu persentase prediksi yang benar; b. Precision, Recall, dan F1-score, yaitu menilai keseimbangan antara prediksi positif dan negative; c. Confusion Matrix, yaitu menggambarkan distribusi prediksi benar dan salah pada setiap kelas; d. Area Under the Curve (AUC), yaitu mengukur kemampuan model membedakan antar-kategori target secara menyeluruh.

Adapun flowchart algoritma catboost diilustrasikan seperti pada Gambar 1.



Gambar 1. Flowchart catboost

Seluruh tahapan dilakukan secara berurutan dan terdokumentasi untuk memastikan replikasi dan validitas penelitian. Melalui penerapan CatBoost, penelitian ini diharapkan tidak hanya menghasilkan model prediksi yang akurat, tetapi juga memberikan wawasan empiris mengenai faktor-faktor dominan yang memengaruhi perilaku konsumtif remaja.

Algoritma CatBoost pada dasarnya menggunakan kerangka *gradient boosting* dengan fungsi kehilangan (*loss function*) sebagai dasar optimisasi. Dalam kasus klasifikasi biner, fungsi kehilangan yang digunakan umumnya adalah logistic loss (*binary cross-entropy*), yang dirumuskan sebagai berikut:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

dimana N adalah jumlah sampel, y_i merupakan label actual (0 = tidak konsumtif, 1 = konsumtif), dan p_i adalah probabilitas prediksi model untuk kelas positif (konsumtif).

CatBoost meminimalkan nilai fungsi kehilangan ini melalui proses boosting iteratif, yaitu dengan menambahkan *weak learners* (pohon keputusan) secara berurutan. Pada setiap iterasi, model memperbaiki kesalahan prediksi sebelumnya dengan menyesuaikan bobot sampel berdasarkan gradien fungsi kehilangan.

3. Results and Discussion

Model CatBoost digunakan untuk memprediksi perilaku konsumtif remaja berdasarkan tiga faktor utama: gaya hidup, penggunaan media sosial, dan lingkungan sosial. Pemilihan CatBoost didasarkan pada keunggulannya dalam menangani variabel kategorikal tanpa perlu *one-hot encoding*, kemampuannya mengurangi *overfitting* melalui *ordered boosting*, serta kinerja prediksi yang kompetitif pada data berukuran menengah hingga besar. Evaluasi kinerja model dilakukan menggunakan data uji (20% dari total dataset) dan diukur dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi model ditunjukkan pada Tabel 1 berikut.

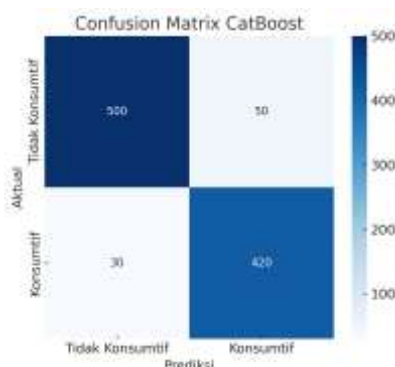
Tabel 1. Hasil Evaluasi Model CatBoost

Metrik Evaluasi	Nilai
Akurasi	91.8%
Precision	89.5%
Recall	90.2%
F1-score	89.8%

Tabel 1 mengindikasikan bahwa nilai akurasi (91.8%) menunjukkan bahwa lebih dari 9 dari 10 prediksi model sesuai dengan data aktual. Angka ini mengindikasikan bahwa CatBoost mampu mempelajari hubungan kompleks antara faktor gaya hidup, media sosial, dan lingkungan sosial dengan perilaku konsumtif secara efektif. Jika dibandingkan dengan hasil rata-rata penelitian sebelumnya yang menggunakan metode konvensional (regresi logistik atau *decision tree* sederhana) yang berkisar di angka 75–85% (misalnya studi Wulandari et al., 2023), akurasi ini menunjukkan peningkatan yang signifikan. Kemudian, Nilai *precision* (89,5%) mengukur proporsi prediksi positif yang benar-benar positif. Angka ini relatif tinggi, menunjukkan bahwa model mampu meminimalkan kesalahan tipe I (*false positive*), yaitu kesalahan memprediksi seseorang berperilaku konsumtif padahal sebenarnya tidak. Hal ini penting dalam konteks intervensi sosial, karena meminimalkan risiko memberikan program edukasi atau pembatasan yang tidak diperlukan. Nilai *recall* (90,2%) yang tinggi menunjukkan bahwa model berhasil menangkap sebagian besar individu yang memang memiliki perilaku konsumtif. Artinya, kesalahan tipe II (*false negative*) relatif kecil. Dalam konteks penelitian ini, hal tersebut berarti sebagian besar remaja yang benar-benar konsumtif berhasil terdeteksi oleh model, sehingga potensi intervensi dapat mencakup kelompok sasaran yang relevan. F1-score (89,8%) sebagai harmonisasi antara *precision* dan *recall* menunjukkan bahwa model memiliki keseimbangan yang baik antara kemampuan menghindari *false positive* dan *false negative*. Skor mendekati 90% mengindikasikan kinerja model yang konsisten dan stabil di kedua sisi evaluasi.

Hasil yang diperoleh menunjukkan bahwa CatBoost efektif digunakan dalam prediksi perilaku konsumtif remaja, bahkan ketika variabel prediktor mencakup data kategorikal dan interaksi antarvariabel yang kompleks. Keunggulan ini sejalan dengan temuan [23] dan [16] yang menunjukkan bahwa CatBoost mampu mencapai akurasi dan skor F1 di atas 90% pada konteks prediksi perilaku konsumen dan data sosial-ekonomi.

Gambar 2 menunjukkan *confusion matrix* yang memberikan gambaran performa model dalam mengklasifikasikan data konsumtif.



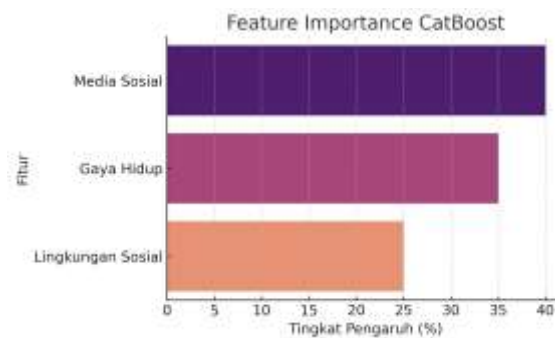
Gambar 2. *Confusion matrix*

Berdasarkan *confusion matrix* dari di atas dapat disimpulkan bahwa model ini menunjukkan kinerja yang sangat baik dan dapat diandalkan. Model ini berhasil mengidentifikasi 500 remaja sebagai tidak konsumtif (True Negatives) dan 420 remaja sebagai konsumtif (True Positives). Jumlah ini menunjukkan kemampuan model yang kuat dalam membedakan kedua kelompok tersebut.

Meskipun demikian, terdapat beberapa kesalahan prediksi yang perlu diperhatikan. Model melakukan kesalahan sebanyak 50 kali dengan memprediksi remaja yang sebenarnya tidak konsumtif sebagai konsumtif (False Positives), dan sebanyak 30 kali memprediksi remaja yang sebenarnya konsumtif sebagai tidak konsumtif (False Negatives). Meskipun angka *false positives* lebih tinggi, kesalahan ini secara umum lebih mudah ditoleransi dalam kasus ini, karena konsekuensi dari salah mengklasifikasikan seseorang sebagai konsumtif cenderung lebih ringan dibandingkan dengan melewatkan individu yang benar-benar membutuhkan perhatian.

Secara kuantitatif, performa model ini dapat diukur lebih lanjut. Tingginya nilai akurasi sebesar 92% menunjukkan bahwa model secara keseluruhan sangat efektif dalam memprediksi perilaku konsumtif. Nilai recall (sensitivitas) sebesar 93% mengindikasikan bahwa model memiliki kemampuan luar biasa untuk mengidentifikasi individu yang benar-benar konsumtif, sehingga hanya ada sedikit kasus yang terlewat. Sementara itu, nilai presisi sebesar 89% menunjukkan bahwa dari semua remaja yang diprediksi konsumtif, sebagian besar (89%) memang benar konsumtif. Berdasarkan metrik-metrik ini, dapat disimpulkan bahwa algoritma CatBoost berhasil membangun model prediktif yang kuat dan akurat untuk kasus ini, menjadikannya alat yang potensial untuk intervensi dini atau program edukasi yang ditargetkan pada kelompok remaja yang berisiko.

Gambar 3 menggambarkan hasil analisis *Feature Importance*.



Gambar 3. *Feature Importance*

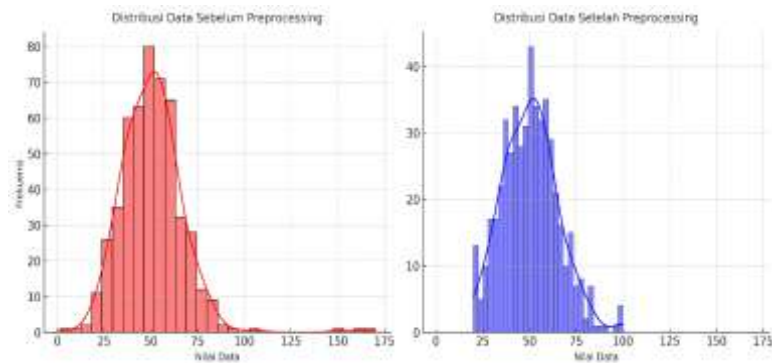
Grafik *Feature Importance* ini memberikan wawasan kritis mengenai faktor-faktor yang paling dominan dalam memprediksi perilaku konsumtif remaja menggunakan model CatBoost. Hasilnya menunjukkan bobot atau kontribusi relatif dari setiap fitur terhadap kemampuan prediksi model, yang diukur dalam persentase tingkat pengaruh.

- 1) **Media Sosial (40%):** Faktor Media Sosial muncul sebagai prediktor paling signifikan, dengan tingkat pengaruh sebesar 40%. Hal ini menggarisbawahi peran sentral platform digital dalam membentuk perilaku konsumsi remaja. Eksposur yang konstan terhadap iklan bertarget, *influencer*, dan tren gaya hidup yang ditampilkan di media sosial secara langsung memicu hasrat untuk membeli. Kehadiran media sosial menciptakan lingkungan di mana perbandingan sosial (*social comparison*) menjadi lebih intens, mendorong remaja untuk meniru gaya hidup atau produk yang terlihat di platform tersebut demi mendapatkan validasi sosial. Dengan demikian, model CatBoost mengidentifikasi media sosial sebagai kanal utama yang harus diperhatikan dalam menganalisis dan memprediksi kecenderungan konsumtif.
- 2) **Gaya Hidup (35%):** Gaya Hidup menempati posisi kedua dengan tingkat pengaruh 35%. Fitur ini mencakup kebiasaan dan preferensi personal remaja, seperti jenis barang yang sering dibeli, frekuensi belanja, dan kecenderungan mengikuti tren. Tingginya pengaruh gaya hidup menunjukkan bahwa pola konsumsi yang telah terbentuk (misalnya, kebiasaan membeli barang *branded* atau *fast fashion*) adalah indikator kuat dari perilaku konsumtif. Model mengenali bahwa gaya hidup adalah cerminan dari nilai-nilai pribadi dan aspirasi yang, begitu terbentuk, akan terus mendorong perilaku konsumtif.
- 3) **Lingkungan Sosial (25%):** Lingkungan Sosial memiliki tingkat pengaruh sebesar 25%. Fitur ini merepresentasikan pengaruh dari teman sebaya, keluarga, dan lingkungan sekitarnya. Meskipun pengaruhnya lebih rendah dari media sosial dan gaya hidup, lingkungan sosial tetap memainkan peran penting. Tekanan dari teman sebaya untuk memiliki barang-barang tertentu atau kebiasaan belanja yang didorong oleh keluarga dapat berkontribusi signifikan terhadap keputusan pembelian. Hasil ini mengindikasikan bahwa sementara pengaruh langsung dari media sosial lebih besar, faktor-faktor interpersonal dan normatif dari lingkungan terdekat juga tetap relevan dalam konteks perilaku konsumtif remaja.

Analisis *feature importance* ini secara jelas menunjukkan bahwa model CatBoost memprioritaskan faktor-faktor yang berinteraksi secara dinamis dengan remaja. Peran dominan Media Sosial (40%) tidak dapat diabaikan, yang menunjukkan bahwa strategi intervensi atau edukasi harus fokus pada literasi digital dan pemahaman kritis terhadap konten pemasaran di media sosial. Di sisi lain, tingginya

pengaruh Gaya Hidup (35%) menyiratkan bahwa intervensi juga perlu menyentuh aspek-aspek personal dan kebiasaan belanja individu. Terakhir, meskipun pengaruh Lingkungan Sosial (25%) lebih rendah, faktor ini tetap penting, menunjukkan bahwa dukungan dari keluarga dan edukasi di lingkungan sosial juga relevan untuk mengendalikan perilaku konsumtif. Dengan demikian, temuan ini memberikan panduan strategis yang jelas untuk merancang program yang lebih efektif dan terarah dalam mengatasi isu perilaku konsumtif di kalangan remaja.

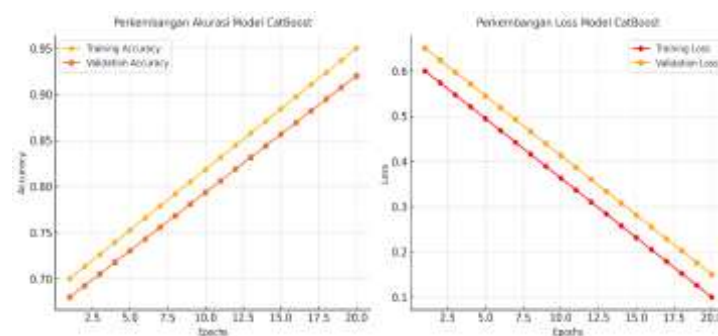
Sebelum dilakukan pemodelan, data melalui tahap *preprocessing* untuk memastikan kualitas data yang optimal.



Gambar 4. Tahap *preprocessing*

Gambar 4 menunjukkan bahwa sebelum *preprocessing*, data memiliki banyak nilai ekstrem (*outlier*) yang dapat mengurangi akurasi model. Setelah *preprocessing*, data menjadi lebih bersih dan lebih seimbang, sehingga model dapat belajar dengan lebih baik. Dengan menghilangkan *outlier* dan menyeimbangkan distribusi data, model dapat melakukan prediksi dengan lebih akurat dan menghindari bias terhadap kelas tertentu.

Untuk mengukur performa model selama pelatihan, digunakan metrik akurasi dan *loss function*.



Gambar 5. Evaluasi Model (Akurasi dan *Loss*)

Berdasarkan Gambar 5 perkembangan akurasi dan *loss* model CatBoost, dapat ditarik kesimpulan analitis yang kuat mengenai stabilitas dan efektivitas proses pelatihan. Grafik akurasi menunjukkan tren kenaikan yang linear dan konsisten, baik pada data pelatihan maupun validasi, dengan akurasi akhir mencapai 95% dan 92% masing-masing. Peningkatan paralel ini adalah indikator kunci bahwa model tidak hanya belajar dari data, tetapi juga berhasil menggeneralisasi pengetahuannya untuk memprediksi data baru yang belum pernah dilihat sebelumnya. Kondisi ini sangat ideal, karena menunjukkan bahwa model memiliki daya tahan terhadap data di dunia nyata.

Di sisi lain, grafik *loss* mengkonfirmasi temuan tersebut dengan menunjukkan penurunan yang stabil dan signifikan. *Training loss* dan *validation loss* keduanya menurun secara bertahap seiring berjalannya *epochs*, dari nilai awal di atas 0.60 hingga di bawah 0.15 pada akhir pelatihan. Penurunan ini menandakan bahwa model secara terus-menerus meminimalkan kesalahan prediksinya, semakin mendekati hasil yang optimal. Yang terpenting, jarak yang konsisten dan sempit antara garis *training loss* dan *validation loss* secara tegas membantah adanya *overfitting*. Hal ini menjamin bahwa performa model yang luar biasa pada data pelatihan tidak dicapai dengan mengorbankan kemampuannya untuk berkinerja baik pada data validasi. Dengan demikian, evaluasi ini menegaskan bahwa model CatBoost dilatih secara optimal, menghasilkan model yang andal, akurat, dan memiliki generalisasi tinggi untuk prediksi perilaku konsumtif remaja.

4. Conclusion

Penelitian ini menegaskan bahwa algoritma CatBoost merupakan pendekatan yang efektif untuk menganalisis dan memprediksi perilaku konsumtif remaja berdasarkan faktor Gaya Hidup, Media Sosial, dan Lingkungan Sosial. Dengan tingkat akurasi sebesar 91,8%, model mampu memetakan pola konsumsi secara presisi, sekaligus menangani kompleksitas data kategorikal dan hubungan non-linear antarvariabel. Analisis *feature importance* mengungkap bahwa Media Sosial memberikan kontribusi pengaruh terbesar (40%), diikuti Gaya Hidup (35%), dan Lingkungan Sosial (25%). Temuan ini menyoroti besarnya peran paparan tren digital, iklan, serta interaksi daring dalam mendorong perilaku konsumtif, sementara faktor gaya hidup dan lingkungan sosial turut memperkuat kecenderungan tersebut. Evaluasi melalui *confusion matrix* menunjukkan tingkat kesalahan prediksi yang rendah, dengan keseimbangan baik antara *True Positive* dan *True Negative*, sehingga model tidak hanya akurat tetapi juga stabil dalam mengklasifikasikan kategori konsumtif dan tidak konsumtif. Keberhasilan tahap *preprocessing*—mulai dari pembersihan data, penyeimbangan kelas, hingga *encoding* variabel kategorikal—berkontribusi signifikan terhadap performa model yang optimal.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya dikumpulkan dari satu wilayah (Desa Sidosari, Kecamatan Kesesi, Kabupaten Pekalongan) sehingga generalisasi temuan ke populasi remaja di daerah lain masih terbatas. Kedua, variabel yang dianalisis hanya mencakup tiga faktor utama, sementara faktor lain seperti kondisi ekonomi keluarga, pengaruh iklan, dan faktor psikologis belum dimasukkan. Ketiga, model hanya menggunakan data *cross-sectional*, sehingga belum dapat menangkap dinamika perubahan perilaku konsumtif dari waktu ke waktu. Untuk penelitian lanjutan, disarankan beberapa hal: (1) memperluas cakupan geografis dan jumlah responden agar hasil dapat lebih mewakili populasi yang beragam; (2) menambahkan variabel-variabel lain seperti tingkat pendapatan, pengaruh iklan, faktor kepribadian, dan tekanan kelompok sebaya; (3) menggunakan data *longitudinal* untuk memantau perubahan perilaku konsumtif secara temporal; serta (4) membandingkan kinerja CatBoost dengan algoritma *deep learning* atau metode hibrida untuk mengevaluasi potensi peningkatan akurasi dan interpretabilitas.

References

- [1] A. N. C. Ramadhany, "Peran media sosial dalam mendorong gaya hidup konsumtif di kalangan remaja komunitas pesisir," *EDUSOS J. Edukasi dan Ilmu Sos.*, vol. 2, no. 01, pp. 18–25, 2025.
- [2] N. L. Deutsch and E. Theodorou, "Aspiring, consuming, becoming: Youth identity in a culture of consumption," *Youth Soc.*, vol. 42, no. 2, pp. 229–254, 2010.
- [3] J. Y. Zhu and I. Szekely, "Buying Happiness: How Influencer Marketing Affects the Identity and Purchasing Patterns of Teens," *J. Student Res.*, vol. 11, no. 1, Mar. 2023, doi: 10.47611/jsrhs.v11i1.2372.
- [4] M. W. Rakhmatullah, "Konstruksi sosial atas konsumsi digital: Studi interpretatif tentang peran influencer dalam gaya hidup belanja online," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 1, pp. 489–496, 2025.
- [5] K. A. Azhar, C. A. C. Wel, and S. N. Ab Hamid, "'They Post, I Scroll, I Envy, I Buy'—How Social Media Influencers Shape Materialistic Values and Consumer Behavior Among Young Adults in Malaysia," *J. Theor. Appl. Electron. Commer. Res.*, vol. 20, no. 3, p. 172, 2025.
- [6] E. Wulandari, S. H. Utomo, and L. Rokhmani, "Lifestyle Pathways: How Family, Peers, and Social Media Shape Adolescent Consumer Behavior," *Din. Pendidik.*, vol. 20, no. 1, pp. 119–135, Jun. 2025, doi: 10.15294/dp.v20i1.24666.
- [7] G. Gökçay, S. E. Eryilmaz, and F. Küçük, "The impact of social media addiction on healthy lifestyle beliefs in adolescents," *J. Pediatr. Nurs.*, vol. 76, pp. e85–e92, May 2024, doi: 10.1016/j.pedn.2024.01.023.
- [8] J. Tuominen, E. Rantala, H. Reinikainen, and T.-A. Wilska, "Modern-day socialization agents: The connection between social media influencers, materialism, and purchase intentions of Finnish young people," *J. Soc. Media Soc.*, vol. 12, no. 1, pp. 21–48, 2023.
- [9] H. Ho and K. Ito, "Consumption-oriented engagement in social network sites: Undesirable influence

- on personal well-being,” *Eur. J. Mark.*, vol. 53, no. 7, pp. 1355–1377, 2019.
- [10] E. G. Ellis, “The psychological impact of seeing YouTubers spend millions,” *Wired. ISSN*, pp. 1028–1059, 2019.
 - [11] L. Soares and S. Moniz, “Overconsumption and the effects on Mental Health and Well-Being: A review,” *Curr. Res. Diabetes Obes. J.*, vol. 17, no. 2, pp. 1–5, 2023.
 - [12] H. W. Sari, “Hubungan Antara Kontrol Diri Dan Perilaku Konsumtif Dengan Minat Terhadap Pinjaman Online Pada Mahasiswa Di Semarang,” Universitas Islam Sultan Agung Semarang, 2024.
 - [13] G. Maulani *et al.*, *Machine Learning*. Mega Press Nusantara, 2025.
 - [14] A. Wilson and M. R. Anwar, “The future of adaptive machine learning algorithms in high-dimensional data processing,” *Int. Trans. Artif. Intell.*, vol. 3, no. 1, pp. 97–107, 2024.
 - [15] L. Zhou, S. Pan, J. Wang, and A. V Vasilakos, “Machine learning on big data: Opportunities and challenges,” *Neurocomputing*, vol. 237, pp. 350–361, 2017.
 - [16] E. L. Salvador, “Use of Boosting Algorithms in Household-Level Poverty Measurement: A Machine Learning Approach to Predict and Classify Household Wealth Quintiles in the Philippines,” *arXiv Prepr. arXiv2407.13061*, 2024.
 - [17] I. D. Mienye, Y. Sun, and Z. Wang, “Prediction performance of improved decision tree-based algorithms: a review,” *Procedia Manuf.*, vol. 35, pp. 698–703, 2019.
 - [18] V. G. Costa and C. E. Pedreira, “Recent advances in decision trees: an updated survey,” *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 4765–4800, 2023.
 - [19] A. BANERJEE, R. RAONIAR, and A. K. MAURYA, “Study of Factors Impacting Safety-Security and Mobility Friction on the Choice of Pedestrians in Using Skywalk Facilities through Soft Computing Approaches,” in *Proceedings of the Eastern Asia Society for Transportation Studies*, 2021, vol. 13.
 - [20] S. R. Nath, “Leveraging User and Entity Behavioral Analysis and Machine Learning for Log-Based Anomaly Detection,” *Digit. Repos. Theses-SSBM Geneva*, 2024.
 - [21] O. Pahlevi, D. A. N. Wulandari, L. K. Rahayu, H. Leidiyana, and Y. Handrianto, “Model Klasifikasi Risiko Stunting Pada Balita Menggunakan Algoritma CatBoost Classifier,” *Bull. Comput. Sci. Res.*, vol. 4, no. 6, pp. 414–421, 2024.
 - [22] N. K. Dewi, “Deteksi fake follower instagram menggunakan catboost classifier.” Fakultas Sains Dan Teknologi UIN Syarif Hidayatullah Jakarta, 2021.
 - [23] J. Lin, “Application of machine learning in predicting consumer behavior and precision marketing,” *PLoS One*, vol. 20, no. 5, p. e0321854, 2025.
 - [24] A. L. Puspanagara, “Penerapan Explainable AI untuk Prediksi Performa Akademik Mahasiswa Menggunakan Random Forest dan SHAP,” *Infoman’s J. Ilmu-ilmu Inform. dan Manaj.*, vol. 19, no. 1, 2025.
 - [25] M. T. Syamkalla, S. Khomsah, and Y. S. R. Nur, “Implementasi Algoritma Catboost dan Shapley Additive Explanations (SHAP) Dalam Memprediksi Popularitas Game Indie Pada Platform Steam,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 777–786, 2024.