

Pemodelan Klasifikasi Popularitas Produk *Skincare* Menggunakan *Support Vector Machine (SVM)*: Studi Komparatif Kinerja Kernel.

Nilia Kamilatutsaniya^{1a)}, Erna Daniati^{2a)}, M.Najibulloh Muzaki^{3a)}

^{a)}*Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri
Jl.K.H.Ahmad Dahlan No.76, Mojoroto, Kediri 64112*

Author Emails

^{a)} *Corresponding author: ernadaniati@unpkediri.ac.id*

^{b)} *nilakamila23@gmail.com*

^{c)} *m.n.muzaki@gmail.com*

Abstract. *The growth of the digital market has increased the variety of skincare products on platforms such as Sephora. This condition ultimately encourages consumers to face challenges in determining popular products. Therefore, this research was conducted to help identify the popularity of skincare products through the development of a classification model based on machine learning algorithms. The purpose of this research is to build a classification model to identify product popularity based on product characteristics and apply the Support Vector Machine (SVM) algorithm with three types of kernels: linear, RBF, and polynomial. Secondary data is obtained from Kaggle which contains skincare product information at Sephora, and processed through the CRISP-DM stages, starting from business understanding, data cleaning, popularity labeling based on the threshold of the number of "loves" and "reviews", balancing data with the SMOTE Technique, to modeling using the Support Vector Machine (SVM) algorithm with three types of kernels: linear, RBF, Polynomial and evaluation. Test results show that the linear kernel provides the highest accuracy of 98.52%. Based on feature selection results, the main factors that affect product popularity are the number of reviews (\log_n of reviews), the number of likes (\log_n of loves), as well as user interaction ratios such as reviews to loves ratio and return on reviews. This research contributes to the development of a machine learning-based prediction model to support decision-making in skincare product marketing.*

Keywords : *Classification, Kaggle, Machine Learning, Skincare, Support Vector Machine(SVM)*

Abstrak. *Pertumbuhan pasar digital telah meningkatkan variasi produk skincare di platform seperti Sephora. Kondisi ini yang pada akhirnya mendorong konsumen menghadapi tantangan dalam penentuan produk yang populer. Oleh karena itu, penelitian ini dilakukan untuk membantu mengidentifikasi popularitas produk skincare melalui pengembangan model klasifikasi berbasis algoritma machine learning. Tujuan penelitian ini adalah membangun model klasifikasi untuk mengidentifikasi popularitas produk berdasarkan karakteristik produk serta menerapkan algoritma *Support Vector Machine (SVM)* dengan tiga jenis kernel: *linear*, *RBF*, dan *polynomial*. Data sekunder diperoleh dari Kaggle yang memuat informasi produk skincare di Sephora, dan diolah melalui tahapan CRISP-DM, mulai dari pemahaman bisnis, pembersihan data, labeling popularitas berdasarkan threshold jumlah "loves" dan "reviews", penyeimbang data dengan Teknik SMOTE, hingga pemodelan menggunakan algoritma Support Vector Machine (SVM) dengan tiga jenis kernel serta evaluasi. Hasil pengujian menunjukkan bahwa kernel linear memberikan akurasi tertinggi sebesar 98,52%. Berdasarkan hasil seleksi fitur, faktor utama yang memengaruhi popularitas produk adalah jumlah ulasan (\log_n of reviews), jumlah suka (\log_n of loves), serta rasio interaksi pengguna seperti *reviews to loves ratio* dan *return on reviews*. Penelitian ini memberikan kontribusi dalam pengembangan model prediksi berbasis *machine learning* untuk mendukung pengambilan keputusan dalam pemasaran produk *skincare*.*

Kata Kunci : *Klasifikasi, Kaggle, Machine Learning, Skincare, Support Vector Machine (SVM)*

PENDAHULUAN

Industri kecantikan menunjukkan pertumbuhan pesat dalam beberapa tahun terakhir. Hal ini ditandai dengan semakin beragamnya produk *skincare* di pasaran. Perkembangan ini turut mengubah preferensi konsumen, terutama Wanita. Saat ini produk kosmetik tidak hanya sebagai alat mempercantik diri, tetapi juga sebagai bagian dari perawatan kulit [1]. Kesadaran ini mendorong konsumen menjadi lebih selektif dalam memilih produk, misalnya produk yang bebas paraben, vegan, dan ramah lingkungan. Tren *global clean beauty* mendukung perubahan ini. Berdasarkan laporan L.E.K. Consulting, sebagian besar konsumen di Amerika Serikat bahkan bersedia membayar 35–40% lebih mahal untuk produk kecantikan ramah lingkungan karena penggunaan bahan alami dan kemasan berkelanjutan [2].

Seiring dengan perkembangan pasar digital memainkan peran krusial dalam membentuk pola konsumsi. Kemajuan teknologi juga telah mengubah pola konsumsi, di mana *platform e-commerce* seperti Sephora menghadirkan berbagai pilihan produk *skincare* yang semakin beragam [3]. Sephora merupakan salah satu pelaku industri ritel kecantikan global yang kini hadir di banyak negara, termasuk Indonesia [4]. Dengan konsep toko yang terbuka dan nyaman, Sephora menyediakan beragam produk kecantikan dari berbagai merek, mulai dari *skincare* hingga perawatan tubuh [4]. Menurut "*2015 PinkReport: The Shopper*", Sephora berhasil menarik minat milenial sebagai kelompok konsumen terbesar yang berbelanja baik secara langsung maupun online [5]. Keberagaman produk ini mempermudah akses konsumen, tetapi juga menimbulkan tantangan dalam memilih produk yang sesuai kebutuhan dan preferensi. Menentukan popularitas produk secara manual menjadi kurang efisien. Menerapkan model algoritma berbasis data pada penelitian ini dapat membantu konsumen menemukan produk terpopuler dengan lebih efisien.

Perkembangan *data science* dan *machine learning* telah banyak dimanfaatkan dalam lingkungan bisnis sebagai alat pendukung dalam proses pengambilan keputusan, promosi, hingga evaluasi produk [6]. *Machine learning* memungkinkan system untuk mempelajari data dan mengidentifikasi pola secara langsung tanpa intervensi langsung dari manusia [6]. Penelitian ini menerapkan algoritma *Support Vector Machine (SVM)*, penggunaan *Support Vector Machine (SVM)* sebagai algoritma klasifikasi telah terbukti menghasilkan akurasi tinggi dalam berbagai penelitian [7], sebagai salah satu metode *machine learning*, SVM dilatih dengan memanfaatkan kumpulan data yang tersedia, SVM unggul dalam menemukan hyperlane yang paling efektif dalam memisahkan data [8].

Penelitian ini menekankan pada pengembangan model klasifikasi popularitas produk, menggunakan data sekunder dari platform Kaggle. Menggunakan kernel linear dan non-linier, kernel linier diterapkan untuk memisahkan data yang dapat dipisahkan secara linier, sebaliknya kernel RBF dan polynomial digunakan ketika data tidak dapat dipisahkan secara linear [9]. Tujuan dilakukannya penelitian ini untuk mengklasifikasikan produk *skincare* berdasarkan Tingkat popularitas dan mengidentifikasi faktor-faktor yang memengaruhi popularitas produk, serta mengevaluasi efektivitas dan perbandingan kinerja ketiga jenis kernel.

TINJAUAN PUSTAKA

Machine Learning

Machine learning merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang menitikberatkan pada pengembangan sistem komputer agar mampu mempelajari pola dari data yang tersedia untuk melakukan prediksi, klasifikasi, maupun pengambilan keputusan tanpa instruksi pemrograman secara langsung [10]. Penelitian terkait klasifikasi popularitas produk berbasis *Machine Learning* semakin marak digunakan, oleh penelitian Pratama (2023) membuat model *Machine Learning* untuk memprediksi churn pelanggan di sektor asuransi. Dengan menggunakan data numerik mereka dapat menemukan pola yang berkontribusi terhadap kemungkinan pelanggan meninggalkan layanan dengan menggunakan data pelanggan [11]

Dataset Skincare

Dataset *skincare* adalah kumpulan data terstruktur yang berisi kumpulan informasi berbagai produk perawatan kulit. Memanfaatkan data sekunder yang diperoleh dari platform *Kaggle* www.kaggle.com/datasets/thedevastator/skincare-products-that-perform-the-best-at-sepho Data tersebut mencakup atribut seperti merek, nama produk, harga, jumlah ulasan, jumlah suka, skor penilaian, ukuran produk, dan lainnya.

Informasi ini digunakan untuk mengidentifikasi karakteristik produk yang memengaruhi tingkat popularitasnya dan menjadi dasar dalam proses klasifikasi popularitas serta pengembangan model menggunakan algoritma *machine learning* [6]

Klasifikasi

Teknik klasifikasi memegang peran krusial dalam proses penambangan data [12]. Klasifikasi merupakan proses pengelompokan data berdasarkan kelas label atau target yang telah ditentukan [13]. Dalam tugas ini, program komputer berfungsi untuk memprediksi atau menetapkan kategori, kelompok, atau kelas dari suatu data. Sebagai metode pembelajaran terarah (*supervised learning*),

klasifikasi bertujuan untuk mengidentifikasi hubungan antara atribut masukan dan atribut target. Akurasi klasifikasi dihitung dari seberapa besar persentase data yang berhasil diprediksi dengan label kelas yang benar. Tujuan fundamental dari klasifikasi adalah untuk meningkatkan keakuratan dan keandalan hasil yang diperoleh dari data.

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma dalam pembelajaran terawasi (*supervised learning*) yang dirancang khusus untuk tugas klasifikasi. Penerapan SVM dalam sistem klasifikasi telah menunjukkan efektivitas yang signifikan di berbagai bidang [14]. Meskipun secara fundamental SVM adalah pengklasifikasi *linear*, kemampuannya telah diperluas untuk menangani masalah klasifikasi *non-linier* dengan mengimplementasikan konsep fungsi kernel. Fungsi kernel ini secara efektif memetakan data ke ruang berdimensi tinggi, memungkinkan pemisahan yang sebelumnya tidak mungkin dilakukan dalam ruang dimensi asli [15]. Berikut beberapa metode kernel dalam SVM dijelaskan dalam rumus persamaan seperti berikut [16]

- a. Kernel Linear

$$K(x_i, x) = x_i^T \cdot x \tag{1}$$

- b. Kernel RBF

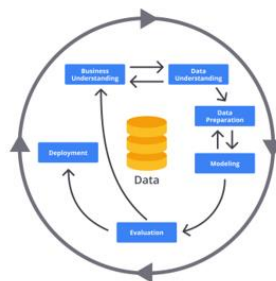
$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \tag{2}$$

- c. Kernel Polynomial

$$K(x_i, x) = (x \cdot x_i + 1)^P \tag{3}$$

METODE PENELITIAN

Penelitian ini mengadopsi metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai pendekatan terstruktur untuk klasifikasi popularitas produk *skincare*. Untuk tujuan ini, algoritma *machine learning Support Vector Machine (SVM)* dimanfaatkan. Metode CRISP-DM terdiri dari enam tahap utama, meliputi (1) *business understanding* (2) *data understanding* (3) *data preparation* (4) *modelling*, (5) *evaluation*, (6) *deployment* [17].



GAMBAR 1. Metode CRISP-DM

Bussines Understanding

Tahap awal dari metodologi CRISP-DM dalam penelitian ini adalah memahami tujuan dan kebutuhan bisnis. Penelitian ini bertujuan untuk mengklasifikasikan tingkat popularitas produk skincare yang tersedia di *platform Sephora*, berdasarkan karakteristik produk seperti harga, jumlah ulasan, skor ulasan, dan atribut lainnya. Informasi ini penting bagi pemangku kepentingan utama, seperti pemilik brand, manajer pemasaran, dan tim pengembangan produk, dalam membuat keputusan strategis terkait promosi, pengembangan produk, dan pengelolaan stok (sitasi).

Data Understanding

Setelah tahap pemahaman bisnis, langkah berikutnya dalam proses penelitian adalah pengumpulan data historis produk skincare yang relevan sebagai dasar analisis. Data yang digunakan bersumber dari platform *Kaggle*, yang menyediakan dataset publik mengenai berbagai produk skincare dan makeup di *Sephora*, mencakup informasi seperti nama produk, merek, harga, jumlah ulasan, skor ulasan, jumlah suka (*loves*), ukuran produk, serta status *clean beauty*. Pada tahap ini juga dilakukan eksplorasi awal terhadap data menggunakan pendekatan *Exploratory Data Analysis (EDA)* guna menilai kualitas data, mengidentifikasi nilai yang hilang dan duplikat, serta memahami distribusi awal dari fitur-fitur utama (sitasi). Selain itu, analisis korelasi diterapkan untuk mengevaluasi hubungan antar fitur dengan variabel target, sehingga dapat memberikan indikasi awal mengenai faktor-faktor yang berpotensi memengaruhi popularitas produk *skincare*.

Data Preparation

Tahap data preparation dilakukan untuk memastikan data siap digunakan dalam pemodelan. Proses ini mencakup pembersihan data dengan menghapus kolom tidak relevan, duplikat, dan menangani nilai hilang. Transformasi logaritma diterapkan pada fitur numerik seperti jumlah ulasan dan jumlah suka guna mengurangi *skewness*. Fitur kategorikal dikonversi menggunakan *one-hot encoding*. Label target ditentukan berdasarkan kuartil ke-75, memisahkan produk menjadi kelas 'Populer' dan 'Belum Populer'. Karena data tidak seimbang, dilakukan Teknik SMOTE yang merupakan metode over sampling dan digunakan untuk mengatasi masalah ketidakseimbangan dalam kumpulan data [18].

Modeling

Tahap pemodelan pada penelitian ini, algoritma Support Vector Machine (SVM) digunakan untuk membuat model klasifikasi popularitas produk skincare. Tiga jenis kernel diuji, yaitu linear, *Radial Basis Function (RBF)*, dan *polynomial*. Data yang telah diproses kemudian dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Pemilihan fitur berdasarkan hasil seleksi mutual information, dari hasil seleksi fitur

Evaluasi

Setelah model SVM dibangun, fase evaluasi dilakukan untuk menilai kinerjanya dalam mengklasifikasikan popularitas produk skincare. Setiap model dievaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* adalah metode yang digunakan untuk mengevaluasi sejauh mana keberhasilan model dalam melakukan prediksi klasifikasi ketika output terdiri dari dua kelas atau lebih [19]. Ditunjukkan dengan tabel 1.

Tabel 1. Confusion Matrix

| Kelas | Prediksi Yes | Prediksi No | Total |
|------------|------------------------|------------------------|-----------------|
| Actual yes | True Positive (TP) | False Negative (PN) | Positive (P) |
| Actual no | False Negative (FN) | True Negative (TN) | Negative (N) |
| Total | P' | N' | P+N |

Berdasarkan tabel 2 dijelaskan bahwa:

- TP (True Positive) : Jumlah data yang diklasifikasikan dengan benar sebagai positif.
- FP (False Positive) : Jumlah data yang salah diklasifikasikan sebagai positif.
- FN (False Negative) : Jumlah data yang salah diklasifikasikan sebagai negatif.

HASIL DAN PEMBAHASAN

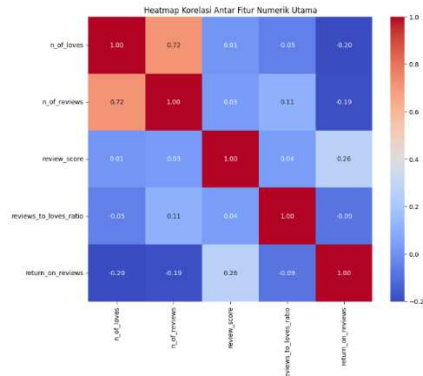
Hasil dan Pembahasan pada penelitian memuat temuan yang didasari melalui tahapan CRISP-DM meliputi:

Bussines Understanding

Penelitian ini mengacu pada tujuan bisnis untuk membantu pihak manajemen memahami pola popularitas produk skincare di platform Sephora. Identifikasi ini dilakukan dengan menganalisis karakteristik produk seperti jumlah ulasan, jumlah suka, harga, dan skor penilaian. Informasi ini diharapkan dapat digunakan oleh pemangku kepentingan, seperti manajer pemasaran dan pengembang produk, dalam menyusun strategi promosi, penentuan stok, serta penyesuaian produk dengan preferensi pasar. Dengan demikian, penerapan model klasifikasi berbasis machine learning bertujuan untuk mendukung pengambilan keputusan yang lebih tepat sasaran dalam meningkatkan penjualan dan kepuasan pelanggan.

Data Understanding

Pada tahap ini, data yang digunakan dalam penelitian merupakan data produk skincare dan kosmetik milik brand Sephora. Data tersebut merupakan data sekunder yang bersifat publik dan diperoleh melalui platform Kaggle. Dari data yang di gunakan pada penelitian ini sebanyak 1689 data berisi informasi seputar produk skincare. Tahapan dalam pemahaman data pada penelitian menggunakan *Exploratory Data Analysis (EDA)* merupakan analisis dan visualisasi data yang dilakukan untuk memperoleh pemahaman yang lebih mendalam serta mengungkap insight yang bermakna dari data tersebut [20]. Pada proses EDA analisis korelasi dilakukan untuk melihat hubungan fitur dengan popularitas produk.

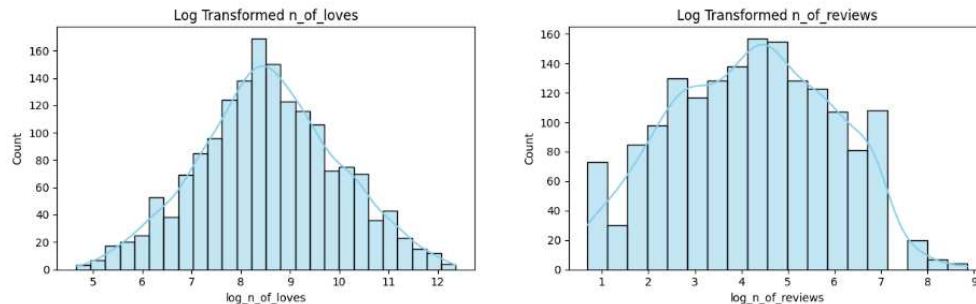


GAMBAR 2. Heatmap Korelasi Fitur

Gambar 2. memperlihatkan heatmap korelasi yang menunjukkan bahwa jumlah ulasan ($\log_n_of_reviews$) dan jumlah suka ($\log_n_of_loves$) memiliki korelasi positif paling kuat terhadap popularitas. Selain itu, rasio interaksi pengguna seperti $reviews_to_loves_ratio$ dan $return_on_reviews$ juga relevan meski korelasinya sedang. Temuan ini mendukung pemilihan fitur utama dalam pemodelan klasifikasi.

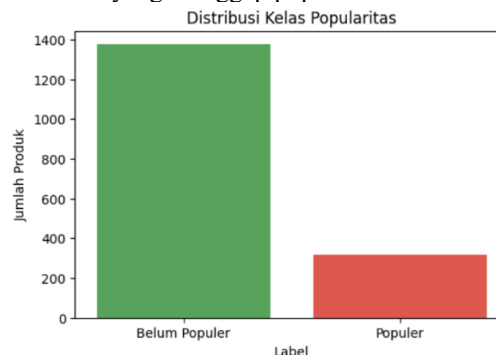
Data Preparation

Data yang telah diperoleh nantinya melalui beberapa langkah penting meliputi persiapan data agar sesuai dengan kebutuhan algoritma *machine learning*, dengan tujuh tahapan sebagai berikut: 1) Drop data yang berarti menghapus kolom atau baris yang tidak relevan atau mengganggu analisis. Total 38 kolom tidak relevan dalam analisis yang dihapus. 2) Standarisasi Teks pada kolom category yaitu dengan tujuan pembersihan spasi berlebih atau menyamaratakan kapitalisasi huruf pada teks agar teks lebih rapi dan konsisten. 3) Data Transformation, dilakukan untuk mengatasi masalah skewness pada fitur numerik agar distribusinya lebih mendekati normal.



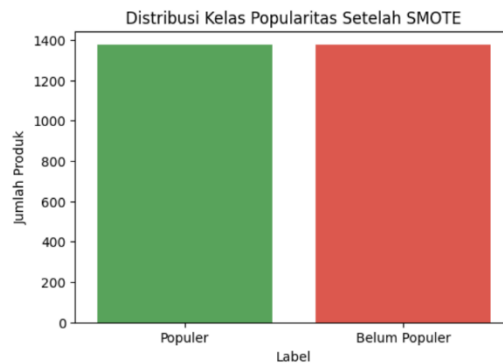
GAMBAR 2. Transformasi Logaritma

Selanjutnya 4) Imputasi, pada tahap ini melakukan imputasi / pengisian pada kolom yang bernilai NaN dengan median. 5) Encoding, merupakan peran penting yang perlu dilakukan untuk mengubah fitur kategorikal menjadi numerik agar bisa digunakan oleh algoritma machine learning. 6) *labeling* pada tahap ini produk diklasifikasikan ke dalam dua kelas: ‘Populer’ dan ‘Belum Populer’. Dilakukan berdasarkan analisis deskriptif dengan pemilihan nilai (Q3) atau kuartil ke-75. Penentuan label menggunakan kuartil ke-75 (Q3) dari distribusi nilai (*log_n_of_loves*) dan (*log_n_of_reviews*). Pendekatan ini dipilih karena Q3 secara statistik merepresentasikan batas atas (upper bound) dari sebaran data, sehingga hanya 25% produk teratas yang dianggap populer



GAMBAR 4. Distribusi Kelas Popularitas

Pada gambar 4 diperoleh hasil distribusi dua kelas dengan data sebesar 1.374 berlabel “Belum Populer” dan 315 data “Populer”. Selanjutnya 7) *Balancing Data* berdasarkan hasil yang diperoleh, bahwa distribusi kelas tidak seimbang (imbalance). Ketidakseimbangan kelas berisiko menurunkan performa model klasifikasi karena model cenderung bias pada kelas mayoritas. Oleh karena itu, diterapkan teknik penyeimbangan data dengan metode SMOTE untuk memperbesar jumlah data pada kelas minoritas sehingga distribusi kelas menjadi lebih seimbang.



GAMBAR 5. Distribusi Popularitas Setelah Imbalance

Dan yang terakhir dari preparation adalah 8) *Feature Selection* dilakukan dengan tujuan memilih atribut atau fitur yang paling relevan secara otomatis. Metode yang digunakan adalah Mutual Information, yaitu teknik yang mengukur seberapa besar informasi yang dibagikan antara setiap fitur dengan variabel target. Dengan hasil yang diperoleh dari 4 pemilihan fitur terbaik, Fitur yang terpilih: `['log_n_of_reviews', 'log_n_of_loves', 'reviews_to_loves_ratio', 'return_on_reviews']`

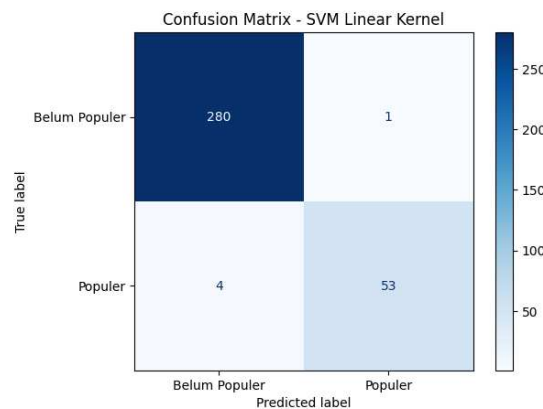
Modeling

Proses pemodelan dalam penelitian ini melibatkan penerapan algoritma *Support Vector Machine* (SVM). SVM dikenal efektif untuk klasifikasi data berdimensi tinggi dan mampu membentuk batas pemisah optimal, sehingga sesuai untuk memetakan popularitas produk skincare. dengan tiga jenis kernel, yaitu *linear*, *RBF*, dan *polynomial*, untuk membangun model klasifikasi popularitas produk *skincare*. Data dibagi menjadi dua bagian, yaitu 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Pemilihan rasio ini bertujuan untuk menjaga keseimbangan antara proses pelatihan dan evaluasi, agar model dapat mempelajari pola dari karakteristik data.

Evaluasi

Evaluasi performa model dilakukan untuk mengukur seberapa baik model yang telah dibangun dalam melakukan klasifikasi. Pengukuran performa dilakukan dengan *confusion matrix* untuk mengukur Tingkat *accuracy*, *precision*, *recall*, dan *F1-score*. Selanjutnya dilakukan prediksi dengan hasil yang ditunjukkan pada *Confusion Matrix* dan *Classification Report* dari masing-masing kernel seperti berikut:

- a. Confusion Matrix pada algoritma SVM Kernel Linear



GAMBAR 6. Confusion Matrix Kernel Linear

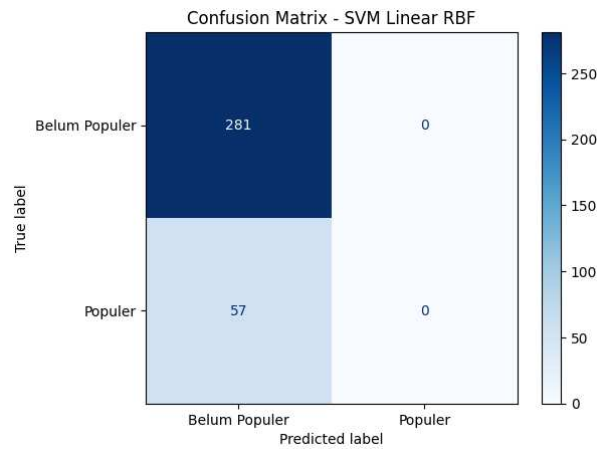
b. Classification Report Kernel Linear

Classification Report (Kernel Linear):

| | precision | recall | f1-score | support |
|---------------|-----------|--------|----------|---------|
| Belum Populer | 0.99 | 1.00 | 0.99 | 281 |
| Populer | 0.98 | 0.93 | 0.95 | 57 |
| accuracy | | | 0.99 | 338 |
| macro avg | 0.98 | 0.96 | 0.97 | 338 |
| weighted avg | 0.99 | 0.99 | 0.99 | 338 |

GAMBAR 7. Classification Report Kernel Linear

c. Confusion Matrix pada algoritma SVM Kernel RBF



GAMBAR 8. Confusion Matrix Kernel RBF

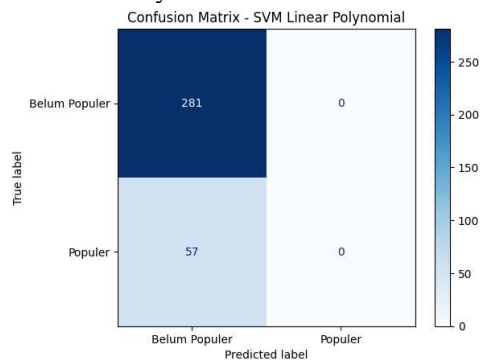
d. Classification Report Kernel RBF

Classification Report (Kernel RBF):

| | precision | recall | f1-score | support |
|---------------|-----------|--------|----------|---------|
| Belum Populer | 0.83 | 1.00 | 0.91 | 281 |
| Populer | 0.00 | 0.00 | 0.00 | 57 |
| accuracy | | | 0.83 | 338 |
| macro avg | 0.42 | 0.50 | 0.45 | 338 |
| weighted avg | 0.69 | 0.83 | 0.75 | 338 |

GAMBAR 9. Classification Report Kernel RBF

e. Confusion Matrix algoritma SVM Kernel Polynomial



GAMBAR 10. Confusion Matrix Kernel Polynomial

f. Classification Report Kernel Polynomial

| Classification | Report (Kernel Polynomial): | | | |
|----------------|-----------------------------|--------|----------|---------|
| | precision | recall | f1-score | support |
| Belum Populer | 0.83 | 1.00 | 0.91 | 281 |
| Populer | 0.00 | 0.00 | 0.00 | 57 |
| accuracy | | | 0.83 | 338 |
| macro avg | 0.42 | 0.50 | 0.45 | 338 |
| weighted avg | 0.69 | 0.83 | 0.75 | 338 |

GAMBAR 11. Classification Report Kernel Polynomial

Hasil dari evaluasi performa model menggunakan algoritma SVM dengan tiga jenis kernel, diketahui bahwa kernel Linear memberikan performa klasifikasi unggul karena pola data popularitas bersifat linier, sehingga pemisahan kelas lebih optimal tanpa perlu transformasi non-linear. Menunjukkan bahwa model mampu mengklasifikasikan kedua kelas dengan sangat baik. Sebaliknya, kernel RBF dan Polynomial menunjukkan performa lebih rendah, hal tersebut dikarenakan kedua kernel tersebut lebih mampu digunakan dalam analisis ketika data tidak terpisah secara linear.

KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi untuk memprediksi popularitas produk skincare, sehingga dapat mendukung pengambilan keputusan pemasaran di platform Sephora. Model tersebut dibangun menggunakan algoritma *Support Vector Machine (SVM)* dengan tiga jenis kernel, yaitu *linear*, *RBF*, dan *polynomial*. Hasil pengujian menunjukkan bahwa kernel linear memberikan akurasi tertinggi sebesar 98,52%, yang menandakan bahwa karakteristik data lebih sesuai dengan pendekatan linier. Kelebihan utama model ini terletak pada kemampuannya mengenali pola popularitas dengan akurasi tinggi, sedangkan kekurangannya adalah ketergantungan pada keseimbangan data dan pemilihan parameter kernel yang tepat. Penelitian ini memberikan kontribusi berupa model klasifikasi popularitas produk skincare dan temuan faktor-faktor dominan, yang dapat dijadikan acuan pengambilan keputusan pemasaran serta pengembangan produk di industri kecantikan.

Saran kepada Penelitian selanjutnya disarankan untuk menggunakan eksplorasi algoritma lain seperti *Random Forest*, *XGBoost*, atau *deep learning* untuk dibandingkan dengan performa SVM. Fitur tambahan seperti analisis sentimen dari ulasan juga dapat diteliti lebih lanjut untuk memperkaya informasi prediktif.

TINJAUAN PUSTAKA

- [1] A. S. Rabiah and F. A. Stefany, "Analisis Preferensi Konsumen Terhadap Produk Perawatan Diri (Skincare) Lokal Yang Mempengaruhi Customer Loyalty," *E-QIEN J. Ekon. dan Bisnis*, vol. 11, no. 3, pp. 726–740, 2022, [Online]. Available: www.janio.com,
- [2] L.E.K. Consulting, "Beautiful Concern: Sustainability Takes Hold in Beauty and Personal Care." Accessed: May 28, 2025. [Online]. Available: <https://www.lek.com/insights/ei/beautiful-concern-sustainability-takes-hold-beauty-and-personal-care>
- [3] F. M. Alja, E. Daniati, and A. Ristyawan, "Perancangan Ui/Ux E-Commerce Menggunakan Metode User Centered Design (Ucd)," *J. Inf. Syst. Manag.*, vol. 6, no. 1, pp. 93–101, 2024, doi: 10.24076/joism.2024v6i1.1669.
- [4] M. Puspitasari and A. Budiarti, "PENGARUH CITRA MEREK DAN KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN PRODUK SEPHORA DENGAN GAYA HIDUP SEBAGAI VARIABEL INTERVENING," *ECOTECHNOPRENEUR J. Econ. Technol. Entrep.*, vol. 3, no. 01, pp. 41–55, Mar. 2024, doi: 10.62668/ECOTECHNOPRENEUR.V3I01.1004.
- [5] W. Wardani *et al.*, "Rahmiawati 1)," *J. Sist. Inf. dan Teknol. Informasi*, vol. 5, no. 3, pp. 287–297, 2023.
- [6] P. R. Sihombing and I. F. Yuliati, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol.

- 20, no. 2, pp. 417–426, May 2021, doi: 10.30812/matrik.v20i2.1174.
- [7] E. Daniati and H. Utama, “Decision Making Framework Based on Sentiment Analysis in Twitter Using SAW and Machine Learning Approach,” *2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020*, pp. 218–222, 2020, doi: 10.1109/ICOIACT50329.2020.9331998.
- [8] D. Wiraharja Athaurrahman and dan Daniel Amadeus Panal Saragih, “Penerapan Support Vector Machine dan Neural Network untuk Memprediksi Nilai Rumah di California pada Dataset ‘California Housing Prices,’” 2024.
- [9] H. Catur Sulistya Ningrum, “Perbandingan Metode Support Vector Machine,” pp. 1–90, 2018.
- [10] T. (Thoyyibah) Thoyyibah, F. (Fajar) Kurniawan, and T. (Taswanda) Taryo, “Dasar-Dasar Machine Learning pada Google Colabs,” Jan. 2024, Accessed: Nov. 20, 2024. [Online]. Available: <https://repository.penerbiteureka.com/uk/publications/567741/>
- [11] R. Pratama, M. I. Herdiansyah, D. Syamsuar, and A. Syazili, “Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 96–104, 2023, doi: 10.32736/sisfokom.v12i1.1507.
- [12] A. Fransisca, K. Wardhani, N. Kamilatutsaniya, A. Alamsyah, E. Daniati, and A. Ristyawan, “Perbandingan Algoritma Naive Bayes Dan Desicion Tree Dalam Pengujian Data Anemia Menggunakan K-Fold Cross Validation,” *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.,* vol. 8, no. 1, pp. 298–309, Jul. 2024, doi: 10.29407/INOTEK.V8I1.4945.
- [13] R. A. Akmal and A. Kurniasih, “Penerapan Algoritma Klasifikasi untuk Menangani Data Tidak Seimbang pada Peningkatan Kualitas Siswa,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, p. 563, 2023.
- [14] T. Andriyanto and E. Daniati, “Classification in the Self Monitoring System for Chronic Kidney Failure Patients on Hemodialysis Therapy with SVM,” *JINAV J. Inf. Vis.*, vol. 3, no. 2, pp. 131–140, 2022, doi: 10.35877/454ri.jinav1410.
- [15] D. W. Athaurrahman, “Penerapan Support Vector Machine dan Neural Network untuk Memprediksi Nilai Rumah di California pada Dataset ‘ California Housing Prices ,”” pp. 1–10.
- [16] A. Z. Praghakusma and N. Charibaldi, “Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi),” *JSTIE (Jurnal Sarj. Tek. Inform.,* vol. 9, no. 2, p. 88, 2021, doi: 10.12928/jstie.v9i2.20181.
- [17] D. Fernando and R. G. Guntara, “Model Klasifikasi Penyebab Turnover Karyawan,” no. 204, pp. 383–392, 2022.
- [18] R. A. Kurniawan, F. I. N. G, and M. R. Pribadi, “Sentimen Analisis Presiden Terpilih Menggunakan Algoritma Support Vector Machine,” *Indones. J. Comput. Sci. Res.*, vol. 3, no. 14, pp. 68–73, 2024.
- [19] T. N. Wijaya, R. Indriati, and M. N. Muzaki, “Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter,” *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjee.v3i2.10885.
- [20] A. R. H. S. H. S. I. Muhammad Radhi, “ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK | Jurnal Sistem Informasi dan Ilmu Komputer.” Accessed: May 30, 2025. [Online]. Available: <https://jurnal.unprimdn.ac.id/index.php/JUSIKOM/article/view/2475>