



# Komparasi Metode K-Nearest Neighbor Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Produk Skincare Skintific

Kiki Setiawan<sup>1\*</sup>, Beatrice Yrain<sup>2</sup>

<sup>1,2,3</sup>Sistem Informasi, STIKOM Cipta Karya Informatika, Jakarta, Indonesia  
Email Penulis Korespondensi: <sup>2</sup>beatriceyrain@gmail.com

**Abstrak**—Seringkali banyak kesalahan saat memilih skincare karena belum mengenali jenis wajah kita maupun hanya tergiur karena produk tersebut sedang viral. Dari pembahasan ini banyak pro dan kontra yang disampaikan melalui sosial media mengenai produk skincare. Melalui penelitian ini akan dilakukan Analisis Sentimen terhadap produk skincare menggunakan metode K-Nearest Neighbour dan Support Vector Machine yang akan memberikan hasil uji dan hasilnya akan di optimalkan dengan fitur selection yaitu Particle Swarm Optimization. Adapun tahapan yang dilakukan dalam penelitian ini adalah pengumpulan data, pelabelan data, pre-processing data, pembobotan, penerapan metode, pengujian dan diakhiri hasil/evaluasi. Tujuan dari penelitian agar hasil uji akan menentukan skincare yang baik supaya para konsumen tidak salah lagi memilih jenis skincare dan bertujuan untuk menentukan apakah penggunaan skincare bersentimen positif atau negatif dan diharapkan bisa menghasilkan nilai akurasi yang baik yang sudah dilakukan oleh 2 metode tersebut, dan didapatkan hasilnya dari metode Support Vector Machine 91,37% sedangkan K-Nearest Neighbour 70,78%. Setelah ditingkatkan nilai akurasinya dengan fitur seleksi maka hasil akurasi yang dihasilkan dari metode Support Vector Machine 94,90% sedangkan K-Nearest Neighbour 72,75%. Lalu di dapatkan 667 sentimen positif dan 76 sentimen negatif.

**Kata Kunci:** Skincare, Sosial Media, Analisis Sentimen, K-Nearest Neighbour, Support Vector Machine, Particle Swarm Optimization

**Abstract**—Often there are many mistakes when choosing skincare because we don't recognize our face type or are only tempted because the product is currently viral. From this discussion, many pros and cons were conveyed via social media regarding skincare products. Through this research, Sentiment Analysis will be carried out for skincare products using the K-Nearest Neighbor method and Support Vector Machine which will provide test results and the results will be optimized with the selection feature, namely Particle Swarm Optimization. The stages carried out in this study were data collection, data labeling, data pre-processing, weighting, method application, testing and ending with results/evaluations. The purpose of the research is that the test results will determine good skincare so that consumers will no longer choose the wrong type of skincare and aim to determine whether the use of skincare has positive or negative sentiments and is expected to produce good accuracy values that have been carried out by the 2 methods, and the results obtained from the Support Vector Machine method are 91.37% while K-Nearest Neighbor is 70.78%. After increasing the accuracy value with the selection feature, the accuracy results generated from the Support Vector Machine method are 94.90% while the K-Nearest Neighbor is 72.75%. Then, 667 positive sentiments and 76 negative sentiments were obtained.

**Keywords:** Skincare, Sosial Media, Analisis Sentimen, K-Nearest Neighbour, Support Vector Machine, Particle Swarm Optimization

## 1. PENDAHULUAN

Perawatan kulit adalah rangkaian praktik yang mendukung integritas kulit, meningkatkan penampilan, dan meredakan kondisi kulit. Itu bisa termasuk nutrisi untuk menghindari paparan sinar matahari yang berlebihan. Praktik yang meningkatkan penampilan yaitu dengan penggunaan skincare dan reaksi dari penggunaan skincare bisa pengelupasan kulit atau pun yang lainnya. Beberapa cara yang digunakan dalam penggunaan skincare adalah terapi retinol dan perawatan kulit ultrasonik. Perawatan kulit bisa untuk semua wajah, tidak mengenal jenis kelamin dan yang lainnya. Perawatan kulit adalah prosedur rutin harian di banyak tempat, seperti kulit yang terlalu kering atau terlalu lembap atau juga terlalu berminyak. Perawatan kulit adalah bagian dari perawatan penyembuhan luka yang ada di wajah. US Federal Food, Drug, and Cosmetic Act mendefinisikan skincare sebagai produk yang dimaksudkan untuk membersihkan (misalnya toner).[1]

Ada kategori terpisah untuk skincare, ada yang dimaksudkan untuk menyembuhkan, mengurangi, merawat di wajah. Skincare juga disimpan secara teratur dan beberapa produk harus disimpan ditempat yang sejuk, misalnya krim/pelembab muka adalah produk yang biasanya disimpan ditempat yang sejuk. Ada beberapa kandungan yang terdapat dalam skincare, diantaranya ada AHA/BHA, Salicylic Acid, Vitamin C, Niacinamide dan masih banyak yang lainnya. Dan setiap kandungan yang terdapat dalam skincare memiliki fungsinya masing masing. Banyak dari kita masih salah memilih skincare karena belum mengenali jenis kulit kita dan hanya tergiur memilih karena produk tersebut sedang viral [2] dan tidak melihat kandungan yang terdapat dalam skincare tersebut, maka dari itu memilih skincare pun tidak boleh sembarangan.[3]

Kita perlu memastikan bahwa produk skincare yang kita pilih memiliki kualitas yang baik dan bahan yang aman. Tentunya, jangan lupa juga agar memilih skincare dengan kandungan yang cocok dengan jenis kulit kita agar hasilnya lebih maksimal dan tidak

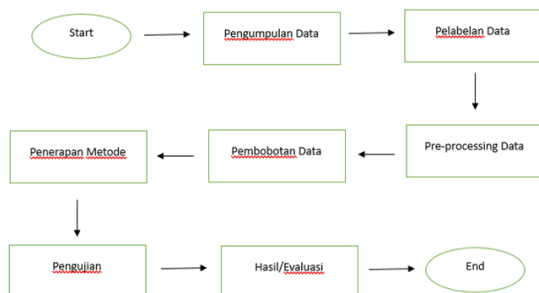


menyebabkan efek samping. Banyak merek skincare yang beredar di sosial media untuk kita coba dan salah satu merek skincare yang sedang banyak diperbincangkan di social media adalah Skintific. Skintific adalah salah satu brand skincare yang memformulasikan produknya di Canada berdasarkan pada scientific research dan teknologi TTE (Triangle Trilogy Effect) yang bekerja secara cepat dan aman untuk barrier kulit. Tahapan basic untuk pemakaian skincare di pagi hari adalah toner, pelembab, dan sunscreen. Lalu untuk pemakaian skincare di malam hari adalah toner, serum, dan ditutup dengan pelembab.[4]

Metode yang mendukung penelitian ini adalah metode K-Nearest Neighbour dan Support Vector Machine menggunakan fitur Particle Swarm Optimization.[5] Merupakan metode yang bisa memberikan hasil uji melalui sebuah dataset menjadi data yang sebelumnya masih berantakan menjadi data yang sudah bersih. Dengan adanya metode K-Nearest Neighbour dan Support Vector Machine Menggunakan Fitur Particle Swarm Optimization dapat mempermudah proses pengelolaan data secara efisien sesuai dengan yang diinginkan.

## 2. METODE PENELITIAN

Metodologi penelitian merupakan tahap-tahap penelitian sistematis untuk membantu penelitian menjadi terarah dengan baik. Berikut adalah metodologi penelitian yang dilakukan penulis.



Gambar 1 Tahapan Penelitian

### A. Pengumpulan Data

Dataset dikumpulkan dengan cara mengetik di mesin pencari pada Youtube sesuai topik studi yang akan diteliti menggunakan Bahasa Indonesia yaitu dengan kata kunci “Skincare” dan “Skintific”

### B. Pelabelan Data

Pelabelan data dilakukan secara manual dengan menentukan klasifikasi sentimen dari dataset tersebut, biasanya bersifat positif atau negatif. Sentimen positif berisi opini dan komentar yang baik tentang produk *skintific*, sebaliknya untuk sentimen negatif berisi opini dan komentar yang tidak baik tentang produk *skintific* tersebut.

### C. Pre-Processing Data

Pre-processing Data merupakan proses untuk mengolah dataset agar membersihkan setiap kata. Data yang berasal dari sosial media biasanya bersifat teks yang masih belum terstruktur, adanya simbol dan emoji, serta banyak kata yang tidak sesuai seperti kata slank atau kata singkatan

### D. Pembobotan Data

Data yang telah melalui tahap Pre-processing dilanjutkan ke tahap pembobotan data untuk mendeteksi nilai keunikan pada teks. Pembobotan data dilakukan dengan menggunakan RapidMiner dengan menerapkan nilai *TF-IDF*.

### E. Penerapan Metode

Peneliti menggunakan bantuan GataFramework untuk melakukan pre-processing data. GataFramework merupakan framework data mining khususnya untuk pre-processing text mining [6] yang dapat digunakan untuk pengolahan kata. Untuk mengolah kata bisa secara manual satu persatu atau sekaligus banyak dengan mengupload dokumen berformat Excel. Beberapa fitur GataFramework yang peneliti gunakan, diantaranya sebagai berikut;

1. Indonesian Stop Removal
2. Indonesian Stemming
3. Transformation: Not Negative
4. Transformation: Remove URL
5. Regexp
6. @Anotation Removal
7. Indonesian Slang
8. Normalalization : Emoticon

### F. Pengujian

Pengujian dilakukan untuk mengetahui performansi algoritma *K-Nearest Neighbor* dan *Support Vector Machine* terhadap sentimen analisis dan untuk menentukan kelas dari tanggapan netizen terhadap produk *skincare* dalam media sosial youtube. Pengujian dilakukan menggunakan RapidMiner Studio. Dataset yang digunakan diperoleh dari komentar pengguna youtube menggunakan teknik crawling [7] dengan keyword “produk *skintific*”. Kemudian data tersebut akan melalui beberapa proses yaitu pelabelan, cleaning data, pre-processing, pembobotan, pemodelan dan kemudian dilakukan pengujian. Pada tahap pelabelan dilakukan secara manual untuk menentukan kelas sentimennya, sentimen yang di tetapkan menjadi dua kelas yaitu kelas positif dan kelas negatif.

### G. Hasil/Evaluasi

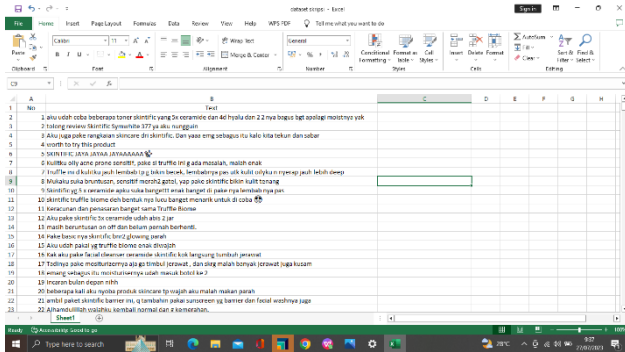
Hasilnya dari metode Support Vector Machine 91,37% sedangkan K-Nearest Neighbour 70,78%. Setelah ditingkatkan nilai akurasi dengan fitur seleksi maka hasil akurasi yang dihasilkan dari metode Support Vector Machine 94,90% sedangkan K-Nearest Neighbour 72,75%. Lalu di dapatkan 666 sentimen positif dan 76 sentimen negatif.

## 3. HASIL DAN PEMBAHASAN



**A. Pengumpulan Data**

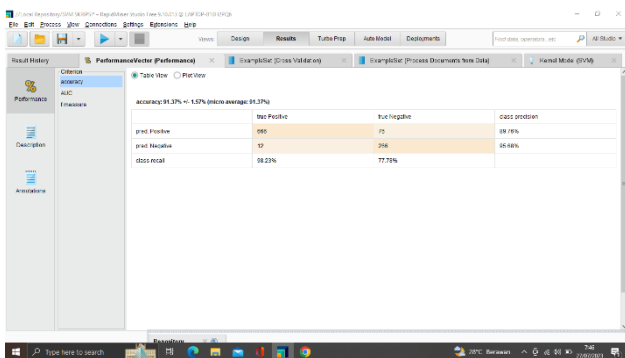
Dataset dikumpulkan dengan cara menyetik di mesin pencari pada Youtube sesuai topik studi yang akan diteliti menggunakan Bahasa Indonesia yaitu dengan kata kunci “Skincare”[8] dan “skintific”. Setelah ditemukan dilakukan pembersihan seperti kata yang duplikat atau tidak berkaitan. Hasil data yang ditemukan adalah sebanyak 1020 data.[9]



Gambar 1. Hasil Data

**B. Pelabelan Data**

Pelabelan data dilakukan secara manual dengan menentukan klasifikasi sentimen dari dataset tersebut, biasanya bersifat positif atau negatif.[10] Sentimen positif berisi opini dan komentar yang baik tentang produk skintific, sebaliknya untuk sentimen negatif berisi opini dan komentar yang tidak baik tentang produk skintific tersebut.

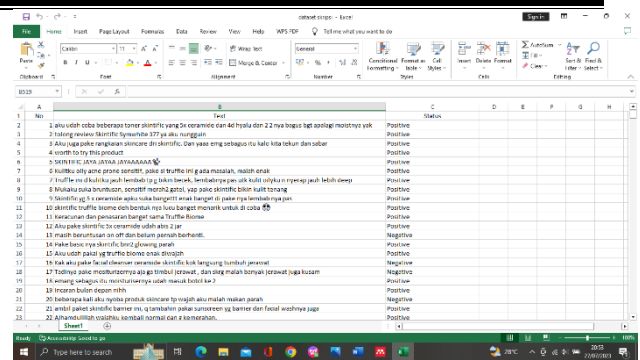


Gambar 2. Pelabelan Data

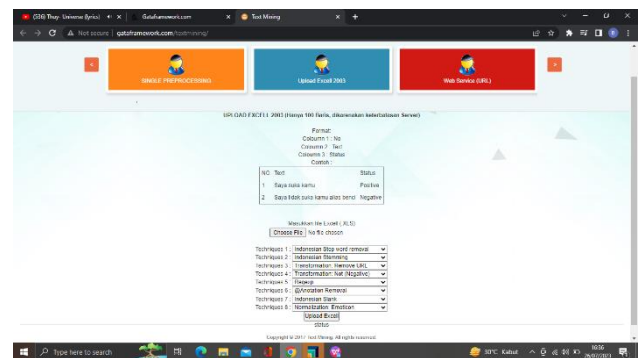
**C. PRE-PROCESSING DATA**

Peneliti menggunakan bantuan GataFramework untuk melakukan pre-processing data. GataFramework merupakan framework data mining khususnya untuk pre-processing text mining[7] yang dapat digunakan untuk pengolahan kata. Untuk mengolah kata bisa secara manual satu persatu atau sekaligus banyak dengan mengupload dokumen berformat Excel. Beberapa fitur GataFramework yang peneliti gunakan, diantaranya sebagai berikut;

1. Indonesian Stop Removal
2. Indonesian Stemming
3. Transformation: Not Negative
4. Transformation: Remove URL
5. Regexp
6. @Anotation Removal



7. Indonesian Slang
8. Normalization : Emoticon

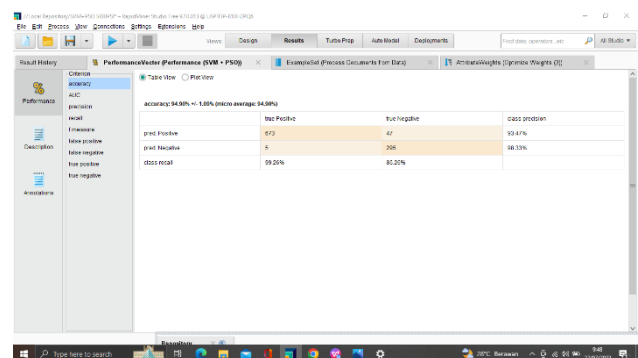


Gambar 3. Tampilan GataFramework

**D. CROSS VALIDATION**

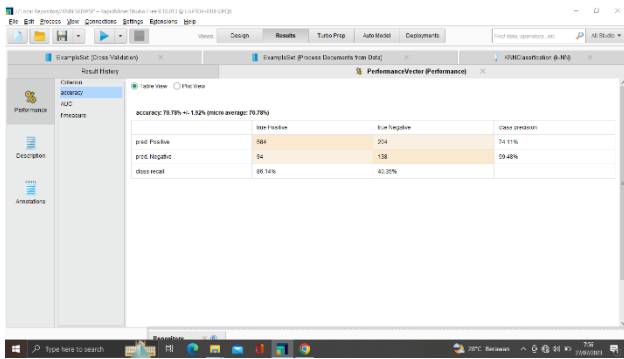
Cross validation merupakan tahap akhir di RapidMiner untuk mendapatkan nilai akurasi dari metode yang peneliti gunakan. Berikut hasil akurasi dari cross validation:

Gambar 4. Hasil Support Vector Machine  
Setelah melakukan cross validation, didapatkan hasil nilai akurasi dari metode Support Vector Machine [11]91,37%+-1,57%



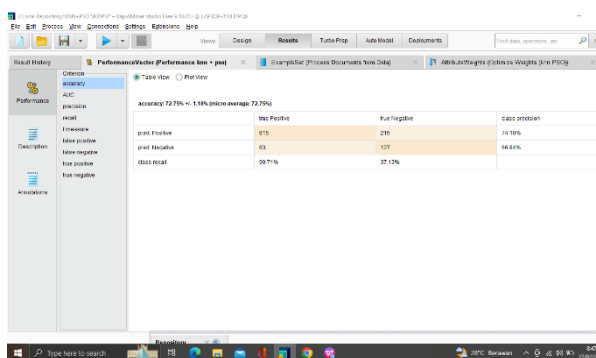
Gambar 5. Hasil Support Vector Machine+Particle Swarm Optimization

Setelah melakukan cross validation, didapatkan hasil nilai akurasi dari metode Support Vector Machine + Particle Swarm Optimization[12] 94,90%+-1,09%



Gambar 6. Hasil K-Nearest Neighbour

Setelah melakukan cross validation, didapatkan hasil nilai akurasi dari metode K-Nearest Neighbour [13] 70,78% +1,92%



Gambar 7. Hasil K-Nearest Neighbour+Particle Swarm Optimization

Setelah melakukan cross validation, didapatkan hasil nilai akurasi dari metode K-Nearest Neighbour [14] + Particle Swarm Optimization 72,75% +1,18%

#### 4. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan, dapat diambil kesimpulan bahwa nilai akurasi tentang skincare [15] dari metode Support Vector Machine 91,37% sedangkan K-Nearest Neighbour 70,78%. Setelah ditingkatkan nilai akurasinya dengan fitur seleksi maka hasil akurasi yang dihasilkan dari metode Support Vector Machine 94,90% sedangkan K-Nearest Neighbour 72,75%. Lalu di dapatkan 667 sentimen positif dan 76 sentimen negatif. Jadi bisa disimpulkan bahwa hasil sentiment [16] dari opini Masyarakat melalui komentar Youtube adalah produk *Skincare Skintific* bersifat positif,

#### UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

#### 5. REFERENSI

[1] Didik Gunawan, Aiga Dwi Pratiwi, Yenni Arfah, and Bobby Hartanto, *Keputusan Pembelian Skincare Safi Berbasis Media Marketing*. Indonesia: PT Inovasi Pratama Internasional, 2022.

[2] J. M. Teknologi Informatika dan Komputer Thamrin, A. Fisty Setyaningsih, D. Septiyani, and S. Rahma Widiyari, 'Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia', doi: 10.37012/jtik.v9i1.1409.

[3] '4. Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)'.

[4] 'PENGARUH MEDIA SOSIAL ONLINE DAN MEDIA PROMOSI OFFLINE TERHADAP PEMILIHAN MEREK PRODUK SKINCARE DAN KLINIK KECANTIKAN'.

[5] T. Astuti and Y. Astuti, 'Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO)', *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 1806, Oct. 2022, doi: 10.30865/mib.v6i4.4119.

[6] 'ANALISIS SENTIMEN TWITTER BAHASA INDONESIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK'.

[7] G. Tokoh, M. Dur, N. Metode, and D. Bayes, 'SKRIPSI ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP'.

[8] C. A. Agustin and G. Reveria Hellianto, 'PENGARUH REPUTASI TERHADAP MINAT BELI PRODUK SKINCARE DI SHOPEE'.

[9] L. Retno Hariatiningsih, 'Penggunaan Skincare Dan Penerapan konsep Beauty 4.0 Pada Media Sosial (Studi Netnografi Wanita Pengguna Instagram)', *Journal Komunikasi*, vol. 11, no. 2, 2020, doi: 10.31294/jkom.

[10] M. Hamka, N. Alfahari, and D. Ratna Sari, 'Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier', *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 64, Sep. 2022, doi: 10.30865/json.v4i1.4740.

[11] N. Fadila Putri, S. Al Faraby, and M. Dwifabri, 'Analisis Sentimen pada Produk Kecantikan dari Ulasan Female Daily Menggunakan Information Gain dan SVM Classifier'.

[12] T. B. Sasongko, 'Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)', 2016.

[13] M. Rangga, A. Nasution, and M. Hayaty, 'Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter', *JURNAL INFORMATIKA*, vol. 6, no. 2, pp. 212–218, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>

[14] M. Kafil, 'PENERAPAN METODE K-NEAREST NEIGHBORS UNTUK PREDIKSI PENJUALAN BERBASIS WEB PADA BOUTIQ DEALOVE BONDOWOSO', 2019.

[15] V. Maarif and H. M. Nur, 'SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN SKINCARE YANG SESUAI DENGAN JENIS



---

KULIT WAJAH MENGGUNAKAN LOGIKA FUZZY', *Jurnal Sains dan Manajemen*, vol. 7, no. 2, 2019.

- [16] R. Febrilia, T. Wulandari, and D. Anubhakti, 'IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MEMPREDIKSI HARGA SAHAM PT. GARUDA INDONESIA TBK', 2021. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexRatihFebriliaTriWulandari>|<http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>