

Perbandingan Algoritma *Naïve Bayes* dan *Support Vector Machine* untuk Analisis Sentimen terhadap Review Produk Aster Kosmetik Malang Marketplace Shopee

Dhiva Mustikananda¹, Dian Eka Ratnawati², Bayu Rahayudi³

Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹dhivamustika@student.ub.ac.id, ²dian_ilkom@ub.ac.id, ³ubay1@ub.ac.id

Abstrak

Aster Kosmetik Malang merupakan toko kecantikan populer yang berada di Kota Malang. Aster Kosmetik menggunakan platform belanja *online* Shopee sebagai media promosi produk. Belanja online memberi banyak kemudahan ditengah pandemi covid-19. Pada platform belanja *online* terdapat fitur untuk menambahkan *review* produk, dimana *review* produk dapat digunakan sebagai pertimbangan bagi calon pembeli dalam mengambil keputusan. Shopee digunakan sebagai sumber data untuk analisis sentimen terhadap *review* produk Aster Kosmetik Malang. Data *review* diklasifikasikan kedalam kategori positif, negatif, dan netral. Pengklasifikasian dilakukan dengan membandingkan algoritma *Naïve Bayes* dan *Support Vector Machine*. Pada penelitian menggunakan data sebanyak 300 data untuk klasifikasi 3 sentimen yaitu positif, negatif, dan netral kemudian 200 data untuk klasifikasi 2 sentimen yaitu positif dan negatif. Pengujian dengan membandingkan kedua metode menggunakan parameter uji yang sama yaitu presentase data latih dan data uji 80%:20% dan $cv=10$. Hasil klasifikasi mendapatkan nilai akurasi tertinggi metode *Support Vector Machine* pada klasifikasi 2 sentimen dengan nilai akurasi 87.56%, *precision* 88.81%, *recall* 86.76% dan *f-measure* 86.85%, hasil akurasi tertinggi pada klasifikasi 3 sentimen metode *Support Vector Machine* dengan nilai akurasi 79.71%, *precision* 81.56%, *recall* 79.56% dan *f-measure* 79.36%.

Kata kunci: Aster Kosmetik Malang, klasifikasi, review, *Naïve Bayes*, *Support Vector Machine*

Abstract

*Aster Cosmetics Malang is a popular beauty shop in Malang City. Aster Cosmetics uses the Shopee online shopping platform as a product promotion medium. Online shopping provides many conveniences in the midst of the COVID-19 pandemic. On the online shopping platform there is a feature to add product reviews, where product reviews can be used as a consideration for potential buyers in making decisions. Shopee is used as a data source for sentiment analysis on Aster Cosmetics Malang product reviews. The review data are classified into positive, negative, and neutral categories. Classification is done by comparing the Naïve Bayes algorithm and Support Vector Machine. This study uses 300 data for the classification of 3 sentiments, namely positive, negative, and neutral, then 200 data for the classification of 2 sentiments, namely positive and negative. Testing by comparing the two methods using the same test parameters, namely the percentage of training data and test data 80%:20% and $cv=10$. The classification results get the highest accuracy value of the Support Vector Machine method in the classification of 2 sentiments with an accuracy value of 87.56%, precision 88.81%, recall 86.76% and *f-measure* 86.85%, the highest accuracy results in the classification of 3 sentiments using the Support Vector Machine method with an accuracy value of 79.71%, precision 81.56%, recall 79.56% and *f-measure* 79.36%.*

Keywords: Aster Kosmetik Malang, classification, review, *Naïve Bayes*, *Support Vector Machine*

1. PENDAHULUAN

Aster Kosmetik adalah salah satu toko di Kota Malang yang menyediakan kebutuhan

salon dan kecantikan dengan berbagai macam produk yang terbilang cukup lengkap. Awal pandemi covid-19 cukup memberikan dampak pada omzet penjualan yang menurun pada Aster

Kosmetik. Belanja *online* memberikan banyak kemudahan bagi penjual dan pelanggan dalam transaksi jual beli produk. Salah satu platform belanja *online* yang ada di Indonesia adalah Shopee. Aster Kosmetik memanfaatkan platform Shopee sebagai media promosi produk untuk meningkatkan penjualan. Shopee terdapat fitur untuk menambahkan komentar atau *review* bagi pelanggan atas produk dan jasa yang diberikan oleh penjual. *Review* produk dapat membantu calon pelanggan lainnya dalam pengambilan keputusan bertransaksi.

Penelitian sebelumnya yang telah dilakukan oleh Binti Najibah (2021) tentang analisis sentimen *review* produk kecantikan menggunakan metode *Naïve Bayes*. Penelitian lain oleh Edgar Maulana (2021) mengenai analisis sentimen opini publik terhadap vaksin covid-19 menggunakan algoritma *Support Vector Machine* dan TF-IDF pada media sosial twitter.

Penelitian ini dilakukan untuk membandingkan klasifikasi terbaik antara dua metode yakni *Naïve Bayes* dengan *Support Vector Machine* dan klasifikasi *review* terbaik antara 2 sentimen dan 3 sentimen. *Naïve Bayes* merupakan klasifikasi menggunakan probabilitas dan statistik berdasarkan *Teorema Bayes*. *Support Vector Machine* merupakan klasifikasi yang menggunakan *hyperlane* untuk menentukan batas antar garis luar dan dalamnya (Samsudiney, 2019). Pengujian menggunakan *Naïve Bayes* terbaik mendapatkan nilai akurasi 82.39% pada penelitian analisis sentimen ulasan produk toko *online* rbylicious untuk peningkatan layanan (Humairah, Darmawan, & Pratiwi, 2020). Metode *Support Vector Machine* nilai akurasi terbaik sebesar 97.00% pada penelitian analisis sentimen *review* produk kosmetik algoritma *Support Vector Machine* dan *Particle Swarm Optimization* sebagai metode seleksi fitur (Dinar Ajeng, 2015).

Berdasarkan penelitian-penelitian yang sudah ada, metode *Naïve Bayes* dan *Support Vector Machine* memberikan nilai evaluasi terbaik, sehingga untuk penelitian ini dapat memberi manfaat sebagai saran dan masukan untuk membantu meningkatkan penjualan Aster Kosmetik Malang.

2. TINJAUAN PUSTAKAN

2.1 Text Preprocessing

Text preprocessing adalah proses untuk mengolah data menjadi data yang siap

digunakan (Mustaqhfi, 2011). *Text Preprocessing* merupakan tahapan awal yang dilakukan yang bertujuan untuk membersihkan data agar tidak terjadi noise pada data. Tahapan *text preprocessing* yakni *case folding*, *tokenizing*, *normalization*, *filtering*, dan *stemming* (Hermawan & Ismiati, 2020).

2.2 Pembobotan Kata (TF-IDF)

Pembobotan kata adalah proses setelah *text preprocessing* untuk memberi bobot pada kata terhadap dokumen (Mulyana, 2012). Tahap awal pada proses pembobotan kata adalah menghitung nilai *term frequency* (TF) yaitu jumlah kata yang sering muncul dalam dokumen. Kata dalam dokumen yang semakin banyak muncul, maka nilai TF dalam dokumen semakin besar. Tahapan kedua adalah mencari *inverse document frequency* (IDF) dengan melakukan log pada hasil pembagian jumlah dokumen dengan frekuensi dokumennya. Tahap ketiga adalah menghitung *term frequency-inverse document* (TF-IDF) yang dapat dilakukan dengan cara mengalikan nilai TF dengan nilai IDF-nya. Pembobotan kata ini akan digunakan dan berperan penting dalam proses klasifikasi dokumen.

2.3 Naïve Bayes

Naïve Bayes merupakan metode klasifikasi berdasarkan *Teorema Bayes* yang dapat dijelaskan pada persamaan 1 berikut ini.

$$P(c|w) = \frac{p(w*c)*P(c)}{P(w)} \quad (1)$$

Keterangan:

$P(c|w)$: probabilitas kemunculan c terhadap w (*posteriori*)

$P(w|c)$: probabilitas kata w terhadap kelas c (*likelihood*)

$P(c)$: probabilitas munculnya kelas c (*prior*)

$P(w)$: probabilitas kemunculan kata w (*evidence*)

Persamaan 1 diatas dapat disederhanakan pada persamaan 2 dengan menghilangkan *evidence* karena tidak berpengaruh terhadap hasil klasifikasi setiap kelas (Rahman and Doewes, 2017).

$$P(c|w_i) = P(c) \times P(w_1|c) \times P(w_2|c) \times P(w_3|c) \times \dots \times P(w_n|c) \quad (2)$$

Perhitungan *prior* dijabarkan pada Persamaan 3 berikut ini.

$$P(C_j) = \frac{N_c}{N} \tag{3}$$

Keterangan:

N_c : total dokumen pada data latih dengan kelas c

N : total dokumen pada data latih

Perhitungan *likelihood* menggunakan model multinomial dengan cara menghitung jumlah kemunculan kata setiap dokumen yang dapat ditunjukkan pada persamaan 4.

$$P(w|c) = \frac{\text{count}(w,c)}{\text{count}(c)} \tag{4}$$

Keterangan:

$\text{count}(w|c)$: jumlah kemunculan w pada kelas c

$\text{count}(c)$: jumlah kemunculan semua kata pada kata c

Perhitungan model multinomial terdapat permasalahan yakni ketika kemunculan kata tidak ada sehingga perhitungan bernilai nol (Kikuchi *et al.*, 2015). Diberikan solusi untuk permasalahan tersebut dengan menggunakan metode *laplace smoothing*. Perhitungan *laplace smoothing* dapat dilihat pada persamaan 5.

$$P(w|c) = \frac{\text{count}(w,c)+1}{\text{count}(c)+|v|} \tag{5}$$

Keterangan:

$|v|$: jumlah kata unik

2.4 Support Vector Machine

Support Vector Machine adalah klasifikasi yang menerapkan algoritma *supervised* yang bekerja dengan menggunakan *hyperlane* sebagai jarak antar kelas yang didapatkan secara maksimal (Samsudiney, 2019). Persamaan *Support Vector Machine* dijabarkan pada Persamaan 6.

$$y = wx + b \tag{6}$$

Keterangan:

y : label data

w : garis *hyperplane* dan *support vector*

b : nilai bias

Penelitian ini menggunakan kernel linear yang dijelaskan kedalam Persamaan 7.

$$\sum_{i=1, j=1}^1 x_i x_j^T, (i, j = 1, 2, \dots, n) \tag{7}$$

Keterangan:

x_i : data input

x_j^T : nilai *transpose* dari nilai x_i

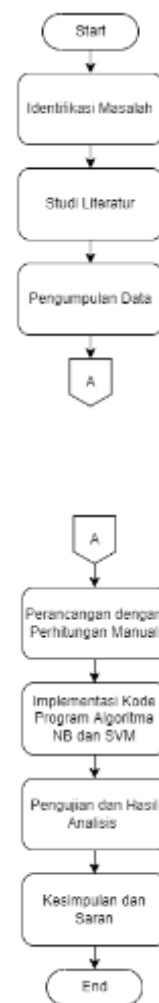
2.5 K-Fold Validation

K-fold Validation merupakan pengujian dengan menggunakan parameter yang sama sebanyak k kali (Khalimi, 2020). *K-fold validation* biasa disebut dengan nama *cross validation*. *Cross validation* ini merupakan pengujian untuk mendapatkan nilai akurasi terbaik.

3. METODOLOGI

3.1 Alur Penelitian

Pada bagian ini akan dijelaskan alur penelitian yang dimulai dengan fase identifikasi masalah hingga fase terakhir yaitu kesimpulan dan saran. Gambar 1 menjelaskan alur penelitian.



Gambar 1. Alur Metodologi

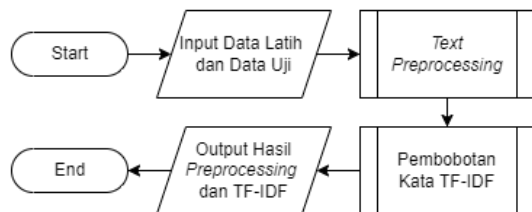
3.2 Pengumpulan Data

Pengumpulan data pada Shopee sebagai media sumber data. Data *review* diambil pada toko online Aster Kosmetik Malang dari tahun 2016 hingga maret 2022 dengan *web scraping*. Data diklasifikasi menjadi 3 kategori kelas yaitu positif, negatif, dan netral. Data yang berhasil diperoleh sebanyak 300 data dengan rincian 110 data positif, 100 data netral dan 90 data negatif.

4. PERANCANGAN DAN IMPLEMENTASI

4.1 Diagram Alir Sistem

Proses yang dilakukan selanjutnya setelah pengumpulan data adalah *text preprocessing*. Tahap pertama *text preprocessing* adalah *case folding* dengan mengubah semua kalimat menjadi *lowercase* dan menghapus semua karakter selain huruf, proses selanjutnya *tokenizing* yaitu proses memisahkan kalimat menjadi token, proses *normalization* mengubah kata yang tidak baku seperti kata singkatan menjadi kata yang baku, proses *filtering* dengan menghapus kata yang tidak diperlukan, dan proses terakhir adalah *stemming* yaitu mengubah kata menjadi bentuk dasarnya. Setelah *text preprocessing* selesai dilakukan kemudian masuk ke tahap selanjutnya yakni pembobotan kata. Pembobotan kata dimulai dari pencarian kata yang sering muncul pada dokumen atau *term frequency* (TF). Kemudian *inverse document frequency* (IDF) dengan menggunakan perhitungan yang didapatkan dari hasil log dari pembagian jumlah dokumen keseluruhan dengan *document frequency* (DF) yang didapatkan dari hasil penjumlahan TF tiap dokumen. Tahap terakhir dari pembobotan kata adalah perkalian TF-IDF. Gambar 2 menunjukkan diagram alir sistem pada penelitian.



Gambar 2. Diagram Alir Sistem

4.2 Diagram Alir Naïve Bayes

Klasifikasi metode *Naïve Bayes* dimulai

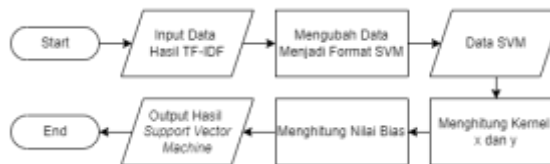
dengan membagi data antara data *training* dan data uji. Tahapan pada *Naïve Bayes* dimulai dari menghitung *prior* pada setiap kelas, kemudian perhitungan *likelihood* dan *posterior*. Hasil dari perhitungan ini akan dipilih yang memiliki probabilitas tertinggi. Gambar 3 merupakan diagram alir *Naïve Bayes*.



Gambar 3. Diagram Alir Naïve Bayes

4.3 Diagram Alir Support Vector Machine

Klasifikasi pada *Support Vector Machine* dimulai dengan membagi data menjadi data latih dan data uji hasil pembobotan kata. Proses selanjutnya mengubah data menjadi bentuk format *Support Vector Machine* dan melakukan perhitungan untuk mendapatkan nilai kernel x dan y dengan matriks untuk mendapatkan nilai *support* dan *vector*, lalu menambahkan nilai bias. Gambar 4 merupakan diagram alir *Support Vector Machine*.



Gambar 3. Diagram Alir Support Vector Machine

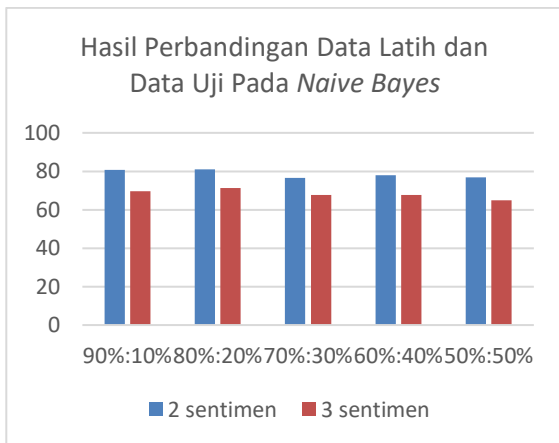
4.4 Perangkat Lunak

Pada penelitian ini proses implementasi kode program menggunakan bahasa pemrograman Python 3.9.7 dan aplikasi *text editor Jupyter Notebook*.

5. ANALISA DAN PENGUJIAN

Pada pengujian *Naïve Bayes* dan *Support Vector Machine* dilakukan dengan membandingkan jumlah data latih dan data uji secara acak serta pengujian parameter nilai C (kompleksitas) dimana pengujian ini akan dilakukan sebanyak 10 kali dan hasil dirata-rata. Pengujian pada penelitian ini dibagi menjadi pengujian klasifikasi 2 sentimen dan 3 sentimen. Pengujian *Naïve Bayes* didapatkan hasil yakni presentase data latih dan data uji 80%:20% untuk

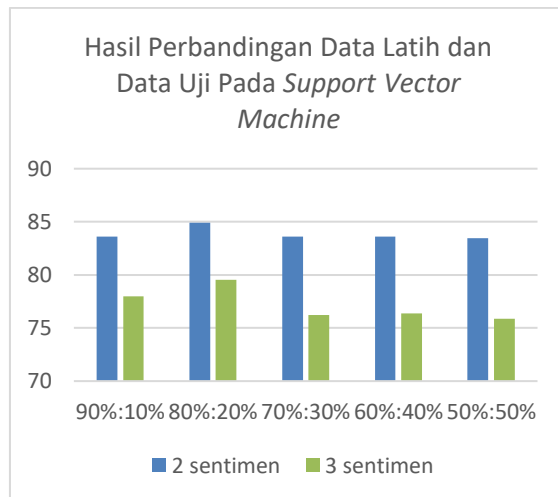
klasifikasi 2 sentimen dan klasifikasi 3 sentimen. Gambar 4 merupakan diagram batang hasil perbandingan data latih dan data uji *Naïve Bayes*.



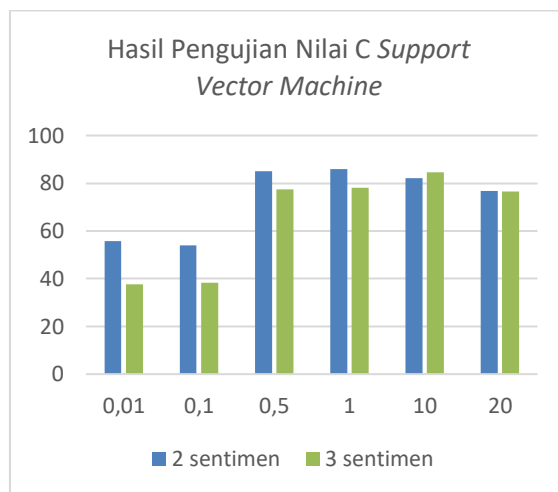
Gambar 4. Diagram Batang Perbandingan Data Latih dan Data Uji *Naïve Bayes*

Gambar 4 dapat dijelaskan bahwa hasil pengujian dengan cara membandingkan data latih dan data uji untuk klasifikasi 2 sentimen didapatkan hasil akurasi terbaik 81.13% dan klasifikasi 3 sentimen mendapat hasil akurasi terbaik 71.38%.

Pengujian metode *Support Vector Machine* dilakukan dengan membandingkan data latih dan data uji dan parameter nilai C. Hasil pengujian yang didapatkan dengan presentase data latih dan data uji 80%:20% untuk klasifikasi 2 sentimen dan klasifikasi 3 sentimen. Klasifikasi 2 sentimen mendapat nilai akurasi sebesar 84.90% dan nilai C=1 dengan nilai akurasi terbaik 85.94%. Hasil klasifikasi 3 sentimen didapatkan akurasi 79.56% dan nilai C=1 dengan nilai akurasi terbaik 78.00%. Gambar 5 merupakan diagram batang pengujian data latih dan data uji. Gambar 6 merupakan diagram batang hasil pengujian parameter kompleksitas.



Gambar 5. Diagram Batang Perbandingan Data Latih dan Data Uji *Support Vector Machine*



Gambar 6. Diagram Batang Parameter Nilai C pada *Support Vector Machine*

Pengujian *cross validation* menggunakan nilai k=10. Hasil *cross validation* pada klasifikasi 2 sentimen metode *Naïve Bayes* mendapatkan nilai akurasi sebesar 82.13%, *precision* 86.19%, *recall* 80.91% dan *f-measure* 80.12%. Tabel 1 merupakan hasil *cross validation* klasifikasi 2 sentimen *Naïve Bayes*.

Tabel 1. Hasil *Cross Validation Naïve Bayes* Klasifikasi 2 Sentimen

Fold ke-	Accuracy	Precision	Recall	F-measure
1	84.38%	88.85%	83.64%	83.39%
2	81.88%	87.95%	79.29%	79.79%
3	85.00%	77.50%	85.00%	80.00%
4	81.25%	86.91%	80.00%	79.28%
5	80.63%	86.95%	78.84%	78.84%
6	80.00%	85.91%	77.95%	77.73%

7	81.25%	86.40%	79.64%	78.80%
8	81.88%	86.09%	81.16%	80.94%
9	84.38%	88.63%	83.48%	83.19%
10	80.63%	86.75%	80.09%	79.27%
Rata-Rata	82.13%	86.19%	80.91%	80.12%

Hasil *cross validation* pada klasifikasi 3 sentimen metode *Naïve Bayes* diperoleh akurasi 72.96%, *precision* 77.60%, *recall* 72.35% dan *f-measure* 72.04%. Tabel 2 menunjukkan hasil *cross validation* klasifikasi 3 sentimen *Naïve Bayes*.

Tabel 2. Hasil *Cross Validation Naïve Bayes* Klasifikasi 3 Sentimen

Fold ke-	Accuracy	Precision	Recall	F-measure
1	71.25%	77.91%	70.05%	69.26%
2	70.83%	75.11%	70.65%	70.49%
3	70.42%	74.11%	69.17%	68.54%
4	81.11%	83.10%	81.00%	80.85%
5	70.83%	77.27%	70.42%	69.44%
6	71.67%	76.94%	71.22%	71.14%
7	70.00%	75.83%	69.11%	68.71%
8	74.58%	77.94%	74.53%	74.41%
9	77.22%	82.49%	76.49%	76.73%
10	71.67%	75.34%	70.88%	70.82%
Rata-Rata	72.96%	77.60%	72.35%	72.04%

Hasil pengujian *Support Vector Machine* untuk klasifikasi 2 sentimen mendapatkan akurasi terbaik sebesar 87.56%, *precision* 88.81%, *recall* 86.76% dan *f-measure* 86.85%. Tabel 3 merupakan hasil *cross validation* klasifikasi 2 sentimen *Support Vector Machine*.

Tabel 3. Hasil *Cross Validation Support Vector Machine* Klasifikasi 2 Sentimen

Fold ke-	Accuracy	Precision	Recall	F-measure
1	90.00%	90.27%	89.99%	89.92%
2	86.88%	88.40%	86.49%	86.44%
3	86.88%	88.55%	86.04%	86.23%
4	89.38%	89.97%	89.25%	89.28%
5	86.25%	88.47%	85.28%	85.47%
6	86.25%	87.08%	85.75%	85.79%
7	85.63%	87.06%	84.29%	84.89%
8	85.00%	86.41%	81.45%	81.57%
9	88.13%	89.57%	88.16%	87.88%
10	91.25%	92.27%	90.90%	91.05%
Rata-Rata	87.56%	88.81%	86.76%	86.85%

Hasil pengujian *Support Vector Machine* untuk klasifikasi 3 sentimen diperoleh akurasi terbaik sebesar 79.71%, *precision* 81.56%, *recall* 79.56% dan *f-measure* 79.36%. Tabel 4 merupakan hasil *cross validation* klasifikasi 1 sentimen *Support Vector Machine*.

Tabel 4. Hasil *Cross Validation Support Vector Machine* Klasifikasi 3 Sentimen

Fold ke-	Accuracy	Precision	Recall	F-measure
1	81.67%	82.46%	81.39%	81.17%
2	78.75%	80.87%	78.74%	78.54%
3	79.17%	82.01%	79.26%	79.30%
4	80.00%	82.02%	79.66%	79.49%
5	78.75%	79.94%	78.82%	78.73%
6	79.17%	81.28%	78.88%	78.52%
7	79.58%	81.69%	79.55%	79.00%
8	80.83%	83.38%	80.45%	80.15%
9	79.17%	81.07%	78.82%	78.97%
10	80.00%	80.90%	80.01%	79.71%
Rata-Rata	79.71%	81.56%	79.56%	79.36%

Berdasarkan hasil pengujian *cross validation* kemudian dapat dilakukan perbandingan metode untuk mendapatkan akurasi terbaik pada masing-masing klasifikasi data dengan menggunakan parameter uji presentase data latih dan data uji 80%:20%, nilai $c=1$ untuk metode *Support Vector Machine* dan $cv=10$ sebagai parameter uji. Didapatkan hasil bahwa untuk klasifikasi 2 sentimen metode *Support Vektor Machine* lebih baik dibandingkan *Naïve Bayes* dengan nilai akurasi sebesar 87.56%, hal ini disebabkan karena proses pelabelan hanya membutuhkan 2 kategori kelas saja yaitu positif dan negatif, sehingga pemrosesan data tidak butuh proses yang kompleks dan mendapatkan hasil yang lebih baik dibandingkan klasifikasi 3 sentimen. Klasifikasi 3 sentimen *Support Vector Machine* lebih baik dengan akurasi sebesar 79.71%.

Dapat disimpulkan bahwa metode *Support Vector Machine* mendapatkan nilai akurasi terbaik untuk klasifikasi 2 sentimen dan 3 sentimen, dikarenakan *Support Vector Machine* bekerja dengan menemukan batas antar kelas dan mengelaskan data berdasarkan batas tersebut. Tabel 5 merupakan hasil perbandingan kedua metode pada klasifikasi 2 sentimen dan Tabel 6 merupakan hasil perbandingan kedua metode pada klasifikasi 3 sentimen.

Tabel 5. Hasil Perbandingan Metode Klasifikasi 2 Sentimen

Hasil Evaluasi	Naïve Bayes	Support Vector Machine
Accuracy	82.13%	87.56%
Precision	86.19%	88.81%
Recall	80.91%	86.76%
F-measure	80.12%	86.85%

Tabel 6. Hasil Perbandingan Metode Klasifikasi 3 Sentimen

Hasil Evaluasi	Naïve Bayes	Support Vector Machine
Accuracy	72.96%	79.71%
Precision	77.60%	81.56%
Recall	72.35%	79.56%
F-measure	72.04%	79.36%

6. KESIMPULAN DAN SARAN

1. Pengklasifikasin *review* produk kecantikan Aster Kosmetik Malang menggunakan metode *Naïve Bayes* dan *Support Vector Machine* pada klasifikasi 2 sentimen dan 3 sentimen. Pengujian menggunakan presentase data latih dan data uji sebesar 80%:20%, nilai $c=1$ dan $cv=10$.
1. Metode *Support Vector Machine* memberikan nilai akurasi terbaik dalam proses pengklasifikasian klasifikasi 2 sentimen dengan akurasi 87.56%, *precision* 88.81%, *recall* 86.76% dan *f-measure* 86.85%. Nilai akurasi terbaik untuk pengklasifikasian 3 sentimen metode *Support Vector Machine* didapatkan nilai akurasi sebesar 79.71%, *precision* 81.56%, *recall* 79.56% dan *f-measure* 79.36%.

Berdasarkan kesimpulan yang sudah dijelaskan peneliti, peneliti memberikan saran yang bisa digunakan sebagai masukan untuk penelitian selanjutnya yaitu:

1. Menambahkan jumlah data penelitian yang digunakan untuk meningkatkan akurasi.
2. Menambahkan parameter pada pengujian *Support Vector Machine* seperti parameter iterasi, *gamma*, dan *lambda* pada kernel.

7. DAFTAR PUSTAKA

Binti Najibah Agus Ratri, Y. A. (2021). Analisis Sentimen *Review* Produk Kecantikan menggunakan Metode *Naïve Bayes*.

Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 5, No. 12.

Edgar Maulana Thoriq, D. E. (2021). Analisis Sentimen Opini Publik pada Media Sosial Twitter terhadap Vaksin Covid-19 menggunakan Algoritma *Support Vector Machine* dan *Term*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 5, No. 12.*

Humairah, I. D. (2020). Analisis Sentimen Ulasan Produk Toko *Online Rubylicious* Untuk Peningkatan Layanan Menggunakan Algoritma *Naive Bayes*. *e-Proceeding of Engineering : Vol.7, No.2 .*

Indrayuni, E. (2018). Komparasi Algoritma *Naive Bayes* Dan *Support Vector Machine* Untuk Analisa Sentimen *Review Film*. *Jurnal Pilar Nusa Mandiri, 14(2), 175.*

Navlani, A. (2018). *Naive Bayes Classification Tutorial using Scikit-learn*. [online] <<https://www.datacamp.com/tutorial/naive-bayes-scikit-learn>> [Diakses 27 Desember 2021]

Sulton Nur Hakim, A. J. (2021). Analisis Sentimen Persepsi Penggunaan *MyIndiHome* Menggunakan Metode *Support Vector Machine (SVM)* dan *Naive Bayes Classifier (NBC)*. *International Journal of Industrial Optimization Vol. 2, No. 2.*

Syukuran, G. (2016). Implementasi Metode *Support Vector machine* Dan Metode *Maximum Marginal Relevance* Untuk Menghasilkan Rangkuman Dari Kumpulan Dokumen Berita Dengan Topik Sejenis. *JBPTUNIKOMPP*

Utami, P. D. (2018). Analisis Sentimen *Review Kosmetik Bahasa Indonesia* Menggunakan Algoritma *Naïve Bayes Classifier*.

Yunus, M. (2020). *Text Preprocessing* menggunakan *Pandas, NLTK* dan *Sastrawi* untuk *Large Dataset*. [online] <<https://yunusmuhammad007.medium.com/text-preprocessing-menggunakan-pandas-nltk-dan-sastrawi-untuk-large-dataset-5fb3c0a88571>> [Diakses 17 November 2021]

Yunus, M. (2020). *TF-IDF (Term Frequency-*

Inverse Document Frequency) :
Representasi *Vector Data Text*. [online] <
<https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>> [Diakses 22
Desember 2021]