

Analisis Sentimen Ekspedisi Sicepat Dari Ulasan *Google Play* Menggunakan *Algoritma Naïve Bayes*

Ayu Kusuma Dewi¹, Sulastri*²

^{1,2} Universitas Stikubank Semarang; Jl Tri Lomba Juang Mugas Sari Semarang

^{1,2} Sistem Informatika, Fakultas Teknologi dan Industri, Universitas Stikubank

e-mail: ¹ayudemak4@gmail.com, *²sulastri@edu.unisbank.ac.id

Abstrak

SiCepat adalah layanan jasa ekspedisi pengiriman barang di Indonesia yang terletak dititik pusat Jakarta. SiCepat memberikan tempat kritik dan saran untuk pengguna aplikasi maupun pengguna layanan ekspedisi yang berada di google play dengan tujuan sebagai peningkatan layanan lebih baik kedepannya. Dengan perkembangan zaman yang semakin canggih pengguna layanan SiCepat bebas memberikan opini dan anggapan apapun. Pada penelitian ini dibutuhkan pengambilan data dari aplikasi SiCepat di google play yang bertujuan untuk mengetahui analisis sentimen. Proses menjalankan klasifikasi diperlukan dua kelas sentimen yaitu sentimen positif dan sentimen negatif. Pengumpulan data yang dilakukan adalah melakukan teknik scrapping menggunakan AppFollow di google play dengan data yang didapatkan sebanyak 457 ulasan positif dan negatif. Selanjutnya pembagian data dibagi dua yaitu data latih dan data uji, dengan melakukan pembagian data sebesar 80% : 20%. Penerapan yang dilakukan dalam pengujian teks mining adalah untuk mengetahui hasil probabilitas dalam pengklasifikasian Naive Bayes. Algoritma Naive Bayes yaitu proses melakukan klasifikasi dengan menggunakan metode probabilitas. dari hasil melakukan penelitian didapatkan hasil tingkat akurasi sebanyak 80%, ulasan negatif sebanyak 87% dan ulasan positif sebanyak 57%.

Kata kunci— *Naïve Bayes, Google Play, Analisis Sentimen, Klasifikasi, SiCepat Ekspres*

Abstract

SiCepat is a freight forwarding service in Indonesia which is located at the central point of Jakarta. SiCepat provides a place for criticism and suggestions for application users and expedition service users who are on Google Play with the aim of improving services for the better in the future. With the development of an increasingly sophisticated era, SiCepat service users are free to provide any opinions and assumptions. In this study, it is necessary to collect data from the SiCepat application on Google Play which aims to determine sentiment analysis. The process of carrying out the classification requires two classes of sentiment, namely positive sentiment and negative sentiment. The data collection was carried out by using a scrapping technique using AppFollow on Google Play with the data obtained as many as 457 positive and negative reviews. Furthermore, the distribution of data is divided into two, namely training data and test data, by dividing the data by 80%: 20%. The application carried out in text mining testing is to determine the probability results in the Naive Bayes classification. The Naive Bayes algorithm is the process of classifying using the probability method. from the results of conducting research, the results obtained an accuracy rate of 80%, negative reviews as much as 87% and positive reviews as much as 57%.

Keywords— *Naïve Bayes, Google Play, Sentiment Analysis, Classification, SiCepat Ekspres*



1. PENDAHULUAN

Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII) yaitu Lembaga yang rutin melakukan survey internet tiap tahunnya. Internet merupakan salah satu kebutuhan pokok bagi kehidupan masyarakat di Indonesia, dengan adanya internet membuat orang mempermudah dalam melakukan apapun sehingga pengaksesan internet semakin naik tiap tahun. dari hasil survey APJII tahun 2019 indonesia mengalami kenaikan sebanyak 73,7% dari tahun sebelumnya yang berjumlah 64,8%. Kenaikan akan terus meningkat sesuai perkembangan infrastruktur internet di Indonesia [1].

SiCepat yaitu salah satu perusahaan logistik yang menawarkan kecepatan pengiriman di Indonesia berdiri pada tahun 2014 didirikan oleh The Kim Hai dan Rudy Darwin Swigo yang berpusat di Jakarta. Dengan banyaknya perkembangan hal ini menjadikan SiCepat menjadi salah satu organisasi koordinasi paling luar biasa di Indonesia dengan sekitar 1.500 perwakilan. Bantuan yang sangat cepat dan aman telah membuat banyak organisasi bisnis internet yang menawarkan kerjasama, membuat SiCepat dikenal luas dan terpercaya [2]. SiCepat memiliki layanan aplikasi di google play yang bernama SiCepat Ekspres yang berfungsi untuk pengecekan nomer resi dan status lacak, pengguna aplikasi SiCepat bebas memberikan komentar yang tersedia di google play sesuai pelayanan SiCepat [3].

Analisis sentimen merupakan sebuah metode pemrosesan memperoleh pengklasifikasian yang kuat serta mengidentifikasi fitur yang sesuai sebagai penghilangan teks opini. Analisis sentimen banyak digunakan pemantauan sebagai gambaran umum media sosial untuk mengetahui masalah pengetahuan perasaan lebih cenderung opini positif atau opini negatif. Analisis sentimen adalah proses mengaplikasikan kata-kata dari review penulis kemudian ditentukan nada emosional yang berasal dari media sosial, e-commerce, dan website. Keunggulan dari analisis sentimen ialah penghematan waktu, tenaga dan biaya karena sangat baik dapat dilakukan secara konsekuen dan memadai [4].

Penentuan klasifikasi menggunakan naïve bayes merupakan pembelajaran mechine learning untuk pengelompokan (klasifikasi) statistik dalam perhitungan probabilitas.

Tabel 1. Gap Analisis Penelitian Terdahulu

Aspek yang ditinjau	Judul	Metode	Gap Hasil penelitian
Aspek tinjauan 1	analisa sentimen twitter pada pilpres 2019 menggunakan algoritma naïve bayes [5].	Naïve Bayes	Data twitter sebanyak 300 tweets menghasilkan akurasi sebesar 73%. Precision kelas negatif sebesar 78% dan precision kelas positif sebesar 66%.
Aspek tinjauan 2	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan	Naïve Bayes	Analisis menggunakan data sebanyak 3780 tweet. hasil analisis positif (60,3 %), netral (34,4 %) menentang (5,4

	Algoritma Naïve Bayes Classifier [6].		%). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).
Aspek tinjauan 3	Analisis sentiment pelanggan toko online JD.ID menggunakan metode naïve bayes classifier berbasis konversi ikon emosi [7]	Naïve Bayes	Dari hasil penelitian menggunakan metode naïve bayes dengan tf-idf pada toko JD.id. dalam perhitungan naïve bayes mendapatkan akurasi sebesar 96,44%, jika digabungkan dengan tf-idf mendapatkan akurasi sebanyak 98%.

Naive Bayes diusulkan oleh peneliti Inggris Thomas Bayes merupakan memanfaatkan prediksi probabilitas untuk perhitungan masa silam hingga masa sekarang. Kalsifikasi naïve bayes lebih memprioritaskan perhitungan probabilitas karena naïve bayes merupakan algoritma yang paling efektif untuk melakukan oerhitungan dibandingkan algoritma yang lain.

Kemudian untuk menyelesaikan persoalan ini yaitu bagaimana cara melakukan klasifikasi naïve bayes ulasan SiCepat di google play, untuk menentukan seberapa akurat hasil akurasi dari tatanan nave Bayes dalam menyampaikan hasil opini positif dan perasaan negatif.

2. METODE PENELITIAN

2.1 Metode Sumber Data

Informasi yang didapat adalah informasi penting yang dikumpulkan pada periode 17 Agustus 2021 hingga 17 September 2021 dari survei google play.

2.2 Metode Teknik Pengumpulan Data

Data pengumpulan pada penelitian menggunakan metode sebagai berikut:

2.1.1 Studi Pustaka

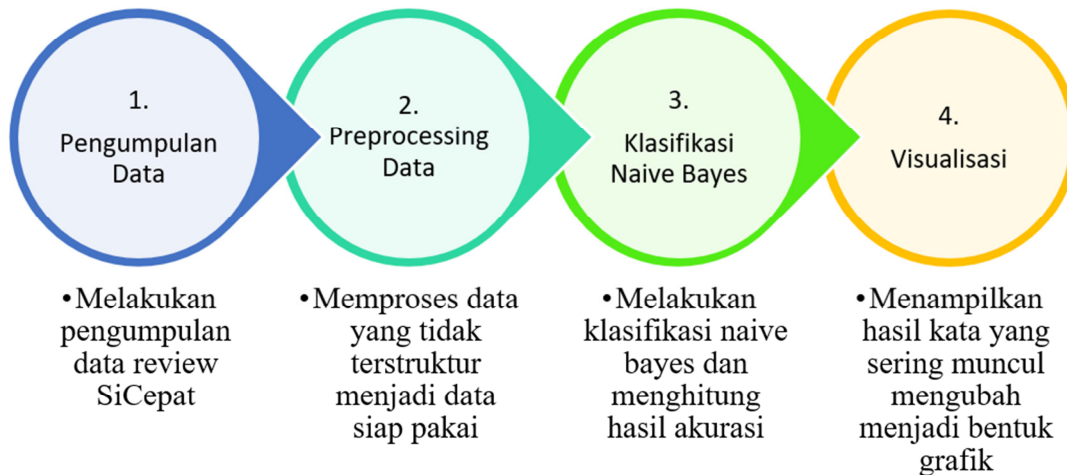
Metode yang dilakukan untuk studi kepustakaan dari jurnal, buku, dan artikel untuk pengumpulan berbagai sumber teori yang meliputi algoritma Naïve Bayes.

2.1.2 Goggle Play

Melakukan pengambilan data yang akurat dari ulasan (review) SiCepat ekspres dari google play.

2.3 Metode Analisis Data

Analisis data yang digunakan metode dalam menyelesaikan analisis sentimen untuk mengetahui kelas sentimen yaitu sebagai berikut :



Gambar 1. Alur Sistem Analisis Sentimen

1. Pengumpulan Data

Pengumpulan data dari hasil scrapping aplikasi AppFollow di google play review SiCepat mendapatkan data sebanyak 457 data, dilakukan pengujian analisis sentimen untuk menghasilkan sentimen positif dan sentimen negatif, kemudian pengolahan penanganan peredaran informasi penyiapan dan informasi pengujian yang diambil secara sewenang-wenang dengan metode 80/20 untuk mengurangi tingkat kesalahan pembagian data latih dan data uji dilakukan pengambilan secara random dengan teknik 80/20 untuk mengurangi tingkat kesalahan.

2. Preprocessing Data

Informasi yang dikumpulkan adalah informasi yang tidak terstruktur oleh karen itu teks preprocessing sangat penting menjadikan data yang terstruktur. Tahapan yang dilakukan preprocessing seperti cleaning, tokenizing, stemming, filtering [8].

3. Klasifikasi Naive Bayes

Algoritma Naive Bayes lebih memprioritaskan perhitungan probabilitas karena Naive Bayes merupakan algoritma yang paling efektif untuk melakukan perhitungan dibandingkan algoritma yang lain.

4. Visualisasi

Fase dari siklus persepsi ini adalah membuat bagan efek samping dari urutan menggunakan strategi Naive Bayes Classifier tergantung pada klasifikasi perasaan dan ekstremitas setiap kata.

2.4 Metode Klasifikasi Naive Bayes

Proses yang dilakukan dalam melakukan klasifikasi ini dibagi beberapa tahap yaitu:

1. Data uji akan dilakukan preprocessing. Tahap teks preprocessing terdiri dari cleaning, tokenizing, stemming, filtering.
2. Setelah melakukan preprocessing kemudian dilakukan perhitungan frekuensi term.
3. Perhitungan nilai Vmap untuk untuk setiap kelas dengan rumus [9]:

$$V_{a} = \underset{V \in_v}{\operatorname{argmax}} \prod_{i=1}^n p(x_i|V)p(V) \quad (1)$$

Keterangan :

Klasifikasi Tweet $j = 1, 2, \dots, n$. Dimana dalam pemeriksaan ini

j_1 = klasifikasi tweet opini positif,

j_2 = klasifikasi tweet opini negatif

$P(x_{il})$ Peluang x_i dalam klasifikasi V_j

$P(V_j)$ Kelas Probabilitas a masih mengudara tergantung pada nilai V_{map} terbesar. Dan hitung akurasi hasilnya confusion matrix dengan rumus sebagai berikut [10]:

$$\frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

3. HASIL DAN PEMBAHASAN

Informasi tersebut didapat dari hasil penolakan menggunakan APPfollow yang diambil melalui google play kemudian, pada saat itu, penolakan pada Aplikasi SiCepat Ekspres pada periode 17 Agustus 2021 hingga 17 September 2021 akibat informasi yang tergores efektif ditambahkan hingga 457 dan disingkirkan. ke dalam pengaturan yang mendominasi untuk meningkatkan proses penyelidikan perasaan. Mempersiapkan informasi dan menguji informasi dikumpulkan secara serampangan untuk mengurangi tingkat kesalahan dengan mempartisi informasi 80/20.

3.1 Pembacaan Data

Rating	Review
0	1 Alamat udah lengkap no hp juga ada
1	1 Perbaiki lagi sistemnya katanya 1 hari sampe t...
2	2 nama ny si cepat tp pengiriman ga ada cepat2 n...
3	1 Ini aplikasi apa sih? Udah saya install
4	1 Semoga kedepan nya lebih baik lagi dalam hal p...
...	...
452	1 Barang kecil cuma stiker kok padahal cuma keci...
453	1 paket diterima dalam keadaan sobek
454	1 Kurang puas
455	1 Belanja di akulaku
456	5 Berhubung barang saya sudah sampai dan sudah s...

457 rows × 2 columns

Gambar 2. Data

3.2 Preprocessing Data

Tahap selanjutnya yaitu melakukan preprocessing supaya data bisa diproses dengan benar, tahap preprocessing yaitu meliputi mengubah menjadi huruf kecil, menghapus tanda baca, menghapus angka dan menghapus kata – kata (stopword removal), tokenizing, stemming. Berikut merupakan script preprocessing pada gambar 3.

```

def remove_data(text):
    text = text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\', "")
    text = text.encode('ascii', 'replace').decode('ascii')
    text = ' '.join(re.sub("([@#][A-Za-z0-9]+) | (\w+:\w+\/\w+)", " ", text).split())
    return text.replace("http://", " ").replace("https://", " ")

data['Review'] = data['Review'].apply(remove_data)

def remove_number(text):
    return re.sub(r"\d+", "", text)
data['Review'] = data['Review'].apply(remove_number)

def remove_punctuation(text):
    return text.translate(str.maketrans("", "", string.punctuation))
data['Review'] = data['Review'].apply(remove_punctuation)

def remove_whitespace_LT(text):
    return text.strip()
data['Review'] = data['Review'].apply(remove_whitespace_LT)

def remove_whitespace_multiple(text):
    return re.sub('\s+', ' ', text)
data['Review'] = data['Review'].apply(remove_whitespace_multiple)

def remove_singl_char(text):
    return re.sub(r"\b[a-zA-Z]\b", "", text)
data['Review'] = data['Review'].apply(remove_singl_char)

```

Gambar 3. Scirpt Preprocessing Teks

3.3 Klasifikasi Naïve Bayes

Model persiapan yang digunakan untuk memainkan karya Naive Bayes dengan Sklearn adalah Gaussian Naïve Bayes. Model Nave Bayes adalah model yang dapat diakses di perpustakaan Python, memiliki tempat dengan model Nave Bayes yang sangat mendasar dan memiliki anggapan bahwa informasi untuk setiap nama diambil dari penyebaran Gaussian dasar.

```

from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer

Ulasan = data['Review']

Ulasan.isnull().sum()

Ulasan = Ulasan.fillna('tidak ada komentar')

cv = CountVectorizer()
term_fit = cv.fit(Ulasan)

print (len(term_fit.vocabulary_))

term_fit.vocabulary_ #mengurutkan berdasarkan urutab abjad kata

#kolom pertama ini berarti jumlah dokumen
#kolom kedua berarti Letak katanya
#kolom ketiga hasil dari tf

term_frequency_all = term_fit.transform(Ulasan)
print (term_frequency_all)

ulasan_tf = Ulasan[1] #memanggil kata pada index ke 1
print (ulasan_tf)

```

Gambar 4. Model Klasifikasi Naïve Bayes

```
term_frequency = term_fit.transform([ulasan_tf]) #hanya menampilkan hasil document 1
print (term_frequency)

(0, 20)      1
(0, 321)     1
(0, 478)     1
(0, 600)     1
(0, 924)     1
```

Gambar 5. Hasil Klasifikasi

```
y_pred_nb = nb.predict(X_test)

from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report

print('----- confusion matrix -----')
print(confusion_matrix(y_test, y_pred_nb))
print('----- classification report -----')
print(classification_report(y_test, y_pred_nb))

----- confusion matrix -----
[[31  7]
 [ 2  6]]
```

Gambar 6. Hasil Confussion Matrix

Dari gambar 6 mengeluarkan confusion matrix dari hasil klasifikasi Naïve Bayes yaitu:

1. True positif (TP) 6
2. True negatif (TN) 31
3. False positif (FP) 2
4. False negatif (FN) 7

```
----- classification report -----
              precision    recall  f1-score   support

negatif      0.94      0.82      0.87         38
positif      0.46      0.75      0.57          8

accuracy          0.80         46
macro avg      0.70      0.78      0.72         46
weighted avg   0.86      0.80      0.82         46
```

Gambar 7. Hasil Akurasi Klasifikasi Naive Bayes

Dari gambar diatas dijelaskan sebagai berikut:

1. Dari gambar 5 mendapatkan hasil akurasi sebanyak:

$$\frac{TP+TN}{TP+FP+TN+FN} = \frac{6+31}{6+2+31+7} = \frac{37}{46} = 0,80 \quad (2)$$

Hasil dari perhitungan rumus akurasi dalam mengklasifikasikan analisis sentimen dari review SiCepat ke dalam sentimen positif dan sentimen negatif menghasilkan akurasi sebanyak 0,80 atau 80%.

2. Sentimen negatif

Hasil dari sentimen negatif dapat diuraikan sebagai berikut:

Pada gambar 6 hasil precision, dan recall, menggunakan GaussianNB mendapatkan hasil percision sebesar 94%, recall sebesar 82%. f1-score digunakan untuk menentukan nilai precision dan recall yang baik hasil f1 score pada klasifikasi mendapatkan hasil sebesar 87%.

3. Sentimen Positif

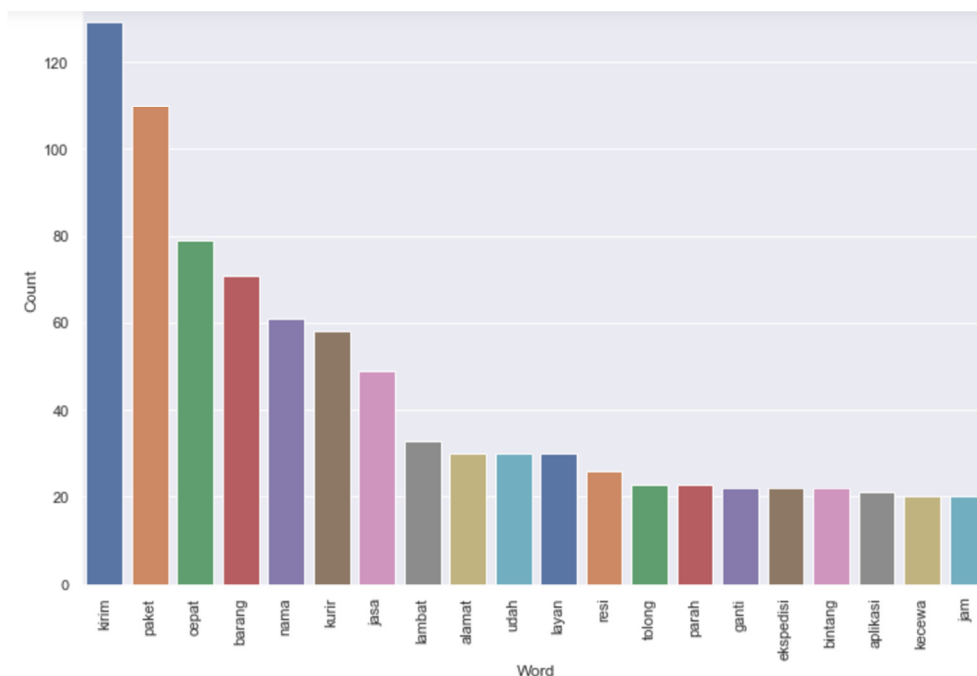
Hasil dari sentimen positif dapat diuraikan sebagai berikut:

Pada gambar 6 hasil precision, dan recall, menggunakan GaussianNB mendapatkan hasil precision sebesar 46%, recall sebesar 75%. f1-score digunakan untuk menentukan nilai precision dan recall yang baik hasil f1 score pada klasifikasi mendapatkan hasil sebesar 57%.

3.4 Visualisasi

Visualisasi pada tahap ini merupakan beberapa kata dari hasil term kemudian dibentuk sebagai berikut terdapat gambar 8 dan 9:

1. Diagram Batang



Gambar 8. Visualisasi Diagram Batang

Gambar 8 merupakan diagram batang hasil term (kata yang sering muncul), dari gambar tersebut dapat dilihat bahwa lima kata terbesar yang sering muncul adalah kirim, paket, cepat, barang, nama.

-
- [2] S. Y. KURNIAWATI, “*Tanggung Jawab Hukum Antara Sicepat Ekspres dan Pengirim Dalam Pengiriman Barang,*” p. 15, 2020.
- [3] P. P. S. Santoso;, “*The Influence of Strategic Agility on Sustainability Competitive,*” Vol. 6, No. 1, 2021.
- [4] R. Wahyudi and G. Kusumawardana, “*Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,*” *J. Inform.*, Vol. 8, No. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [5] A. Imron, “*Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,*” pp. 10–13, 2019.
- [6] W. Y.-J. D. M. dan S. Informasi and undefined 2021, “*Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier,*” *Ejurnal.Teknokrat.Ac.Id*, Vol. 2, No. 2, pp. 1–9, 2021.
- [7] A. Sari, F. V., & Wibowo, “*Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,*” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, Vol. 2, No. 2, pp. 681–686, 2019.
- [8] S. Fanissa, M. A. Fauzi, and S. Adinugroho, “*Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking,*” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 2, No. 8, pp. 2766–2770, 2018.
- [9] Y. S. Mahardika and E. Zuliarso, “*Analisis Sentimen Terhadap Pemerintahan Joko Widodo pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes,*” *Pros. SINTAK 2018*, no. 2015, pp. 409–413, 2018.
- [10] L. D. Mahbubah *et al.*, “*Analisa Sentimen Twitter pada Pilpres 2019 Menggunakan,*” *Sintak*, pp. 194–195, 2019.