

Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier

by Joins 5883

Submission date: 31-Mar-2022 11:24AM (UTC+0700)

Submission ID: 1797586055

File name: 5883-17790-2-BR.docx (434.29K)

Word count: 3040

Character count: 18342

Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier

22

Abstrak

Twitter adalah salah satu media sosial dan fasilitas microblogging yang menjadi tempat bagi penggunaannya berbagi pengalamannya secara bebas, realtime, dan bersifat publik. Hal ini dapat menjadikan twitter sebagai sumber informasi yang dapat berupa opini, ataupun komentar yang bersifat positif maupun negatif. Dari opini masyarakat tersebut dapat diimplementasikan sebagai tolak ukur, karena memiliki nilai bagi suatu perusahaan agar dapat menjadi bahan evaluasi untuk menentukan langkah dalam meningkatkan layanannya. Oleh karena itu untuk mengolah opini tersebut dibutuhkan teknik analisis sentimen untuk dapat mengidentifikasi opini baik positif maupun negatif. Pada penelitian ini akan menganalisis tweet berbahasa Indonesia dengan topik layanan yang ada pada E-commerce shopee yaitu layanan ShopeeFood yang sedang populer dikalangan masyarakat saat ini. Metode yang akan digunakan untuk analisa sentimen pada penelitian ini yaitu Naïve Bayes Classifier untuk proses mengklasifikasi. Berdasarkan hasil pengujian klasifikasi tweet pada penelitian ini dibuktikan keakuratan yang didapatkan melalui confusion matrix dengan nilai accuracy sebesar 90,62%, precision sebesar 88,23%, dan recall sebesar 93,75%.

4

Kata kunci: Analisis Sentimen, Teks Mining, Naïve Bayes Classifier, Twitter, ShopeeFood

Abstract

Twitter is a social media and microblogging facility that is a place for users to share their experiences freely, in real-time, and publicly. This can make Twitter a source of information that can be in the form of opinions, or comments that are positive or negative. From the public opinion, it can be implemented as a benchmark, because it has value for a company so that it can be used as an evaluation material to determine steps to improve its services. Therefore, to process these opinions, sentiment analysis techniques are needed to be able to identify both positive and negative opinions. In this study, we will analyze Indonesian-language tweets with the topic of services that exist in Shopee E-commerce, namely the ShopeeFood service which is currently popular among the public. The method that will be used for sentiment analysis in this research is the Naïve Bayes Classifier for the classification process. Based on the results of the tweet classification test in this study, it was proven that the accuracy was obtained through the confusion matrix with an accuracy value of 90.62%, precision of 88.23%, and recall of 93.75%.

Keywords: Sentiment Analysis, Text Mining, Naïve Bayes Classifier, Twitter, ShopeeFood

1. PENDAHULUAN

Di era teknologi digital yang serba *modern* membuat kemajuan akan pada konsep *bisnis* berkembang sangat pesat. Salah satunya yaitu transaksi jual beli sekarang dapat dilakukan *secara online* dengan menggunakan perangkat elektronik yang biasa disebut dengan *E-commerce* yang membuat segala proses transaksi menjadi lebih praktis dan sederhana. Selain itu memberikan dampak yang positif bagi masyarakat dalam membantu aktifitas yang dilakukan dan juga dapat memenuhi segala kebutuhan manusia di seluruh dunia [1].

Shopee merupakan platform *E-commerce* yang mendukung perkembangan ekonomi digital dan memiliki pangsa pasar terbesar di Indonesia. Salah satu layanan yang ditawarkan *Shopee* adalah *ShopeeFood* yang merupakan layanan *food delivery*. *ShopeeFood* launching perdana di bulan April 2020, pada saat itu hanya untuk pembelian makanan beku, minuman ringan, kue, dan aneka olahan makanan. Setelah itu *ShopeeFood* memingkatkan layanannya dengan memperluas cakupannya dan *merchant* yang terdaftar di *ShopeeFood* juga semakin beragam pada awal tahun 2021 [2]. *ShopeeFood* sekarang ini sangat diminati oleh masyarakat selain layanannya baru, banyak keuntungan juga yang didapatkan [3]. Namun tidak dapat menutup kemungkinan bahwa *ShopeeFood* ini pasti juga memiliki banyaknya kekurangan dalam berbagai faktor pelayanannya. Jika kekurangan terhadap layanan tersebut tidak diatasi maka dampak yang akan muncul yaitu turunya minat masyarakat untuk menggunakan layanan ini. Salah satu upaya untuk meningkatkan kualitas layanan yaitu dengan menganalisa respon para pengguna yang mengandung berbagai faktor layanan seperti pengiriman, penawaran diskon, ketersediaan atau kelengkapan dan faktor-faktor lainnya yang ada pada media sosial seperti *Twitter*.

Twitter adalah fasilitas *microblogging* yang menjadi tempat berbagi pengalaman tanpa ada penghalang dan *realtime* [4]. Untuk mengakses *twitter* mudah karena *tweet* pada akun pengguna yang terekam bersifat publik. Hal ini dapat menjadikan *twitter* sebagai sumber data opini masyarakat terkait penilaian kepuasan, kualitas produk, kualitas pelayanan dan lain sebagainya yang memiliki nilai bagi suatu perusahaan untuk mengetahui persepsi pelanggan dan karakter pelanggan karena dapat menjadikan evaluasi agar dapat menentukan langkah yang strategis untuk meningkatkan layanan dan mengerti kebutuhan pelanggan.

Untuk mengolah data opini maka dilakukan analisis sentimen yang merupakan teknik *text mining* yang sering dihubungkan dengan proses analisis data media sosial sehingga mendapatkan hasil positif dan negatif [5]. Proses ini termasuk dalam bidang *Natural Language Processing (NLP)* yang bertugas dalam mengidentifikasi dan mengekstaksi makna bahasa yang berbentuk teks [6]. Untuk mendukung proses ini maka diperlukan metode klasifikasi *text* pada penelitian yang akan dilakukan agar mendapatkan hasil terbaik dan akurat. Banyak metode yang sudah pernah diterapkan dalam penelitian analisis sentimen seperti *Naïve Bayes Classifier*, *K-Nearest Neighbor (K-NN)*, *Support Vector Machine (SVM)*, *Decision Tree* dan lain sebagainya. Salah satu metodenya yaitu *Naïve Bayes Classifier* yang termasuk dalam *supervised learning*, telah dibuktikan bahwa metode ini efektif karena sederhana, cepat dan memiliki akurasi yang tinggi untuk kategorisasi teks [7] [8] [9].

2. METODE PENELITIAN

2.1 Pengumpulan Data

Pengumpulan data didapatkan melalui media sosial *Twitter* dengan menggunakan *twitter* API untuk proses *crawling* data. Data yang diambil adalah data *tweets* berbahasa Indonesia yang didapatkan dengan memakai *keyword* “*shopeefood*” dengan periode *tweet* pada bulan Oktober 2021. Data yang digunakan ialah *tweet* yang mengandung opini layanan *ShopeeFood*, seperti kecepatan pengiriman, penawaran diskon, ketersediaan atau kelengkapan, driver, dan lain sebagainya. Variabel yang diambil yaitu waktu, ID *user*, *username*, dan teks *tweet*. Data *tweet*

yang diambil untuk proses analisa sentimen yakni sebanyak 1020 data *tweet*. Data yang sudah diperoleh, selanjutnya dilakukan pemberian label secara manual yaitu positif dan negatif.

2.2 Tahap Preprocessing

Penelitian ini diperlukan proses *Preprocessing* data yang merupakan proses mengubah suatu dokumen yang tidak terstruktur menjadi data yang terstruktur untuk kebutuhan analisis lebih lanjut agar tidak mengurangi performanya sendiri karena pola data yang belum sesuai dengan sistem. Pada tahap *preprocessing* hanya menggunakan variabel data *tweet* opini. Berikut tahapan *Preprocessing* yang dilakukan penulis pada penelitian ini :

- 1) Pada tahapan *Cleaning*, dilakukan beberapa tahap yaitu :
 - a. Melakukan tahapan *case folding* guna menyamaratakan huruf pada data *tweet* menjadi huruf kecil (*lowercase*).
 - b. Menghilangkan url, *mention*, *punctuation* dan angka.
 - c. Melakukan tokenisasi, yang dimana proses memisahkan setiap kata pada suatu data *tweet*.
- 2) Pada tahapan *Stopword Removal*, menghapus kata yang dianggap tidak memberikan dampak terhadap informasi yang ada pada suatu data *tweet*.
- 3) Pada tahapan *Stemming*, dilakukan agar pada suatu data *tweet* yang memiliki imbuhan menjadi kata dasar dan menjadi suatu pola untuk dapat melakukan klasifikasi kata lain yang memiliki kata dasar maupun yang memiliki arti serupa namun berbeda karena memiliki imbuhan yang beda.

2.3 Tahap Pemodelan

Pada tahapan ini menerapkan metode *Naïve Bayes Classifier* sebagai klasifikasi pada analisa sentimen. *Naïve bayes classifier* termasuk kedalam teknik data mining *supervised learning* yang merupakan pendekatan dengan guru, data setnya sudah memiliki label/class [10]. Adapun langkah untuk proses klasifikasi yaitu melakukan perhitungan *Posterior Probability* yang didapatkan dari nilai *Prior Probability* dan *Conditional Probability* dengan persamaan sebagai berikut.

- *Posterior Probability*

$$C_{map} = \operatorname{argmax}_{C \in \mathcal{C}} P(C) \prod_{i=1}^n P(w_i | C) \quad (1)$$

- *Prior Probability*

$$P(c) = \frac{N_c}{N} \quad (2)$$

- *Conditional Probability*

Untuk menghindari terjadinya peluang bernilai 0 maka digunakan *Laplace Smoothing* yang cara kerjanya menambahkan angka 1 pada *numerator* dan jumlah kosakata pada *dominator*.

$$P(w_i | C) = \frac{\text{count}(w_i, C) + 1}{\text{count}(C) + |V|} \quad (3)$$

Keterangan :

C	= Class,
N_c	= Jumlah dokumen dalam kelas C ,
N	= Jumlah total dokumen <i>training</i> ,
w_i	= Kata w ke i ,
$\text{count}(w_i, C)$	= Jumlah kata w_i dalam C ,
$\text{count}(C)$	= Jumlah kata di class C ,
$ V $	= Jumlah kosakata.

3. HASIL DAN PEMBAHASAN

Bab ini mengulas uraian hasil penelitian yang sudah dilakukan oleh peneliti untuk mengetahui seberapa baik hasil klasifikasi menggunakan metode *Naïve Bayes* dalam mengklasifikasi data teks.

3.1 Tahapan *Preprocessing*

Sebelum dilakukan pengujian data *tweet*, dilakukan *preprocessing* terlebih dahulu agar mendapatkan data bersih dan sudah terstruktur. Tahapan ini melakukan *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.

Berikut hasil data *training* yang dilakukan *preprocessing* yang tertera pada Tabel 1.

Tabel 1. Data *Training* Sebelum dan Sesudah *Preprocessing*

Dokumen	<i>Original Tweet</i>	Hasil <i>Preprocessing</i>	Kelas
D1	@ShopeeID mau beli makanan enaaak di shopeefood mumpung banyak promo #1010ShopeexITZY	beli makanan enak shopeefood selagi banyak promo	Positif
D2	Pertama kali order di shopeefood dimarahin drivernya	pertama pesan shopeefood marah driver	Negatif
D3	@██████ Pake shopeefood deh banyak gratis ongkir sama voucher cashback nya	pakai shopeefood banyak gratis ongkir voucher cashback	Positif
D4	@ShopeeID kalo driver shopeefood gak jalan-jalan bisa gak sih secara otomatis dibatalin dari shopeenya	driver shopeefood tidak jalan secara otomatis batal shopee	Negatif

3.2 Tahapan *Naïve Bayes Classifier*

Data yang dapat dilanjutkan proses klasifikasi yaitu sebanyak 960 data *tweet* dengan uraian 480 positif dan 480 negatif. Pada data *Testing* dilakukan klasifikasi berdasarkan data *Training* yang sudah ada. Berikut kumpulan *term* pada data *Training* pada Tabel 1 yang disusun kedalam satu kolom dan dihitung kemunculan katanya pada setiap dokumen, yang dapat dilihat pada Tabel 2.

Tabel 2. Daftar *Term Frequency*

<i>Term</i>	<i>Term Frequency</i>			
	D1	D2	D3	D4
beli	1	0	0	0
makan	1	0	0	0
enak	1	0	0	0
shopeefood	1	1	1	1
selagi	1	0	0	0
banyak	1	0	1	0
promo	1	0	0	0
pertama	0	1	0	0
pesan	0	1	0	0
marah	0	1	0	0
driver	0	1	0	1
pakai	0	0	1	0
gratis	0	0	1	0
ongkir	0	0	1	0
voucher	0	0	1	0

Term	Term Frequency			
	D1	D2	D3	D4
cashback	0	0	1	0
tidak	0	0	0	1
jalan	0	0	0	1
secara	0	0	0	1
otomatis	0	0	0	1
batal	0	0	0	1
shopee	0	0	0	1
pada	0	0	0	1

Selanjutnya hitung *Prior Probabilitas* untuk dapat mengetahui banyak jumlah dokumen pada kelas data *training*, hasilnya dapat dilihat pada perhitungan berikut :

1. Probabilitas $P(C_{\text{Positif}}) = \frac{2}{4} = 0,5$
2. Probabilitas $P(C_{\text{Negatif}}) = \frac{2}{4} = 0,5$

Setelah mendapatkan hasil probabilitas setiap kelas, dilanjutkan mengitung nilai *Conditional Probabilitas* dan menentukan kelas berdasarkan probabilitas tertinggi dari masing-masing kelas.

- 1) Data Testing yang digunakan
D5 = mantap shopeefood voucher gratis ongkir makanan cepat sampai
- 2) Perhitungan *Conditional Probabilitas* D5 untuk kelas **Positif** dan **Negatif**.

Tabel 3. Perhitungan *Conditional Probabilitas*

Term	Kelas	
	Positif	Negatif
Mantap	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
Shopeefood	$(2 + 1)/(12 + 23) = \mathbf{0,0857}$	$(2 + 1)/(12 + 23) = \mathbf{0,0857}$
Voucher	$(1 + 1)/(12 + 23) = \mathbf{0,0571}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
gratis	$(1 + 1)/(12 + 23) = \mathbf{0,0571}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
ongkir	$(1 + 1)/(12 + 23) = \mathbf{0,0571}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
makanan	$(1 + 1)/(12 + 23) = \mathbf{0,0571}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
cepat	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$
sampai	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$	$(0 + 1)/(12 + 23) = \mathbf{0,0285}$

- 3) Perhitungan *Posterior Probability* D5 untuk kelas Positif.
 $P(C_{\text{Positif}}|D5) = P(C_{\text{Positif}}) \times P(\text{mantap}|Positif) \times P(\text{shopeefood}|Positif) \times P(\text{voucher}|Positif) \times P(\text{gratis}|Positif) \times P(\text{ongkir}|Positif) \times P(\text{makanan}|Positif) \times P(\text{cepat}|Positif) \times P(\text{sampai}|Positif)$
 $= 0,5 \times 0,0285 \times 0,0857 \times 0,0571 \times 0,0571 \times 0,0571 \times 0,0571 \times 0,0285 \times 0,0285$
 $= 1,0544$
- 4) Perhitungan *Posterior Probability* D5 untuk kelas Negatif.
 $P(C_{\text{Negatif}}|D5) = P(C_{\text{Positif}}) \times P(\text{mantap}|Negatif) \times P(\text{shopeefood}|Negatif) \times P(\text{voucher}|Negatif) \times P(\text{gratis}|Negatif) \times P(\text{ongkir}|Negatif) \times P(\text{makanan}|Negatif) \times P(\text{cepat}|Negatif) \times P(\text{sampai}|Negatif)$
 $= 0,5 \times 0,0285 \times 0,0857 \times 0,0285 \times 0,0285 \times 0,0285 \times 0,0285 \times 0,0285 \times 0,0285$
 $= 0,6544$

Hasil pada perhitungan data *testing* D5, disimpulkan bahwa data *testing* D5 termasuk pada kelas kelas **Positif**, karena pada perhitungan probabilitas kelas positif lebih tinggi dibandingkan probabilitas kelas negatif.

3.3 Performa Naïve Bayes Classifier

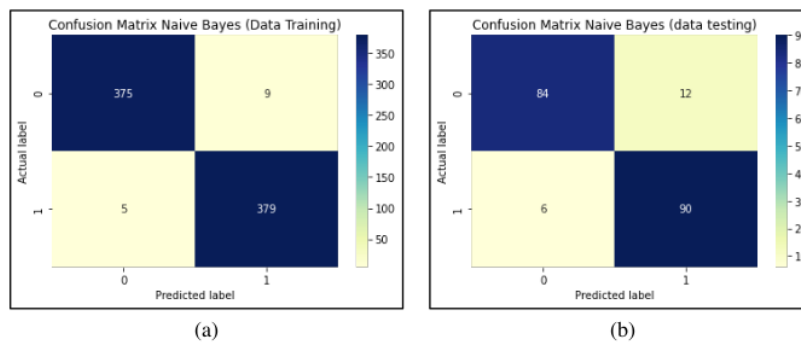
Evaluasi mengenai *actual* dan performa hasil klasifikasi didapatkan dengan *Confusion Matrix*. Pengukuran performa dilakukan dengan menghitung *statistic*, yaitu *True Positives (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* dengan persamaan berikut :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

Berikut hasil *Confusion Matrix* data *Training* dan data *Testing* dengan pengujian 8:2 yang terbagi 768 *tweet data training*, dan 192 *tweet data testing*.



Gambar 1. *Confusion Matrix*
(a) *Data Training*, (b) *Data Testing*

Data Training

$$\text{Accuracy} = (379 + 375)/(379 + 375 + 9 + 5) \times 100\% = 98,17\%$$

$$\text{Precision} = (379)/(379 + 9) \times 100\% = 97,68\%$$

$$\text{Recall} = (379)/(379 + 5) \times 100\% = 98,69\%$$

Data Testing

$$\text{Accuracy} = (90 + 84)/(90 + 84 + 12 + 6) \times 100\% = 90,62\%$$

$$\text{Precision} = (90)/(90 + 102) \times 100\% = 88,23\%$$

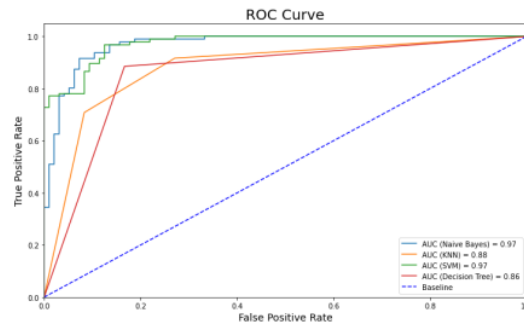
$$\text{Recall} = (90)/(90 + 6) \times 100\% = 93,75\%$$

Dari hasil penelitian ini menunjukkan data *testing* bersifat *overfitting* karena nilai *accuracy* pada data *training* lebih tinggi dibandingkan data *testing* dengan menggunakan metode *naïve bayes*.

3.4 Komparasi Model

Berdasarkan hasil performa yang sudah didapatkan dari proses sebelumnya, maka dilanjutkan untuk memvisualisasikan hasil klasifikasi berdasarkan metode yang digunakan yaitu metode utama *naïve bayes* dan metode pembandingan KNN, SVM, *decision tree* dalam bentuk grafik ROC (Receiver Operating Characteristic). Dengan grafik ROC akan menunjukkan apakah hasil klasifikasi yang didapatkan bernilai baik atau buruk berdasarkan nilai AUC yang didapatkan.

Berikut grafik ROC yang tertera pada Gambar 2 yang akan dihitung nilai AUC untuk metode utama *naïve bayes* dan metode pembandingan KNN, SVM, *decision tree* yang dikomparasikan pada penelitian ini :



Gambar 2. Grafik ROC

Grafik pada Gambar 2 menunjukkan bahwa metode utama *Naive Bayes* dan metode pembanding SVM yang digunakan pada penelitian ini menghasilkan performa ‘sangat baik’ dengan nilai AUC sebesar 0.97, sedangkan dengan menggunakan metode pembanding KNN menghasilkan nilai AUC 0.88 yang membuktikan bahwa metode pembanding KNN termasuk metode yang menghasilkan performa yang ‘baik’, begitu pula dengan metode pembanding *Decision Tree* yang menghasilkan nilai AUC 0.86 yang memiliki performa ‘baik’.

3.5 Implementasi Prototype

Testing model yang digunakan yaitu *Naive Bayes Classifier*. Langkah yang dilakukan pada proses klasifikasi data yaitu menginput data teks yang kemudian dilanjutkan proses *preprocessing* sehingga menjadi data teks yang bersih agar mendapatkan hasil klasifikasinya.

Proses mengklasifikasi tertera pada Gambar 3.

```

tweet = '@hopeefood_id Ini pesanan saya sudah di selesaikan tapi makanan bla sampe ke saya :('

test = cleaning_text(tweet)
print("tahap prepro pertama :")
print(test)

tahap prepro pertama :
Ini pesanan saya sudah di selesaikan tapi makanan bla sampe ke saya

test2 = tokenization(test)
print("tahap prepro kedua :")
print(test2)

tahap prepro kedua :
['ini', 'pesanan', 'saya', 'sudah', 'di', 'selesaikan', 'tapi', 'makanan', 'bla', 'sampe', 'ke', 'saya']

test3 = remove_stopwords(test2)
print("tahap prepro ketiga :")
print(test3)

tahap prepro ketiga :
['pesanan', 'selesaikan', 'makanan', 'bla', 'sampe']

test4 = stemming(test3)
print("tahap prepro keempat :")
print(test4)

tahap prepro keempat :
['pesan', 'selesai', 'makan', 'bla', 'sampe']

hasil = fit_stopwords(test4)
print("hasil Preprocessing :")
print(hasil)

Hasil Preprocessing :
pesan selesai makan bla sampe

print("text tweet termasuk polaritas :", gaussian_predict(preprocess_data(hasil)))

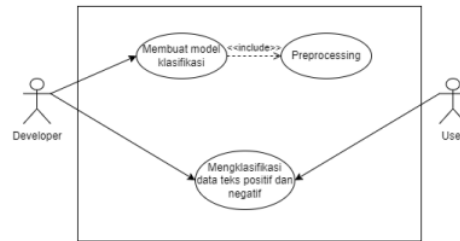
Text tweet termasuk polaritas : [0]
    
```

Gambar 3. Mengklasifikasi Data

A. Perancangan UML

1. Use Case Diagram

Use case mendeskripsikan alur sistem yang hendak diusulkan dengan menunjukkan aktor yang akan berinteraksi langsung dengan sistem. Berikut ini adalah gambaran mengenai *Use Case Diagram* Sistem *Text Mining* pada Gambar 4 :



Gambar 4. Use Case Diagram

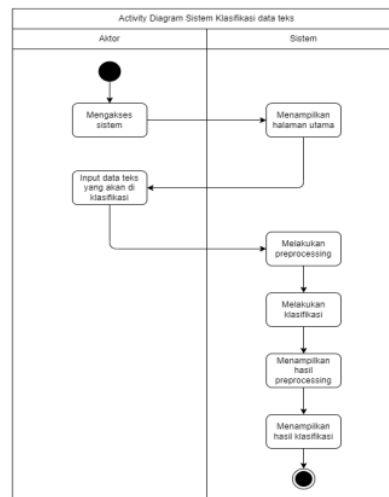
Deskripsi *Use Case Diagram* :

- Aktor *developer* (pengembang) akan melakukan pembuatan model klasifikasi yang sudah dilakukan *preprocessing* untuk diterapkan kedalam sistem. Proses *preprocessing* yang dilakukan yaitu *cleaning* teks, *tokenizing*, *stopword removal*, dan *stemming*.
- Aktor *developer* (pengembang) dan user dapat melakukan klasifikasi positif dan negatif pada sistem secara otomatis.

2. Activity Diagram

23

Activity diagram berisi tindakan yang terjadi pada sistem yang dijalankan. Berikut adalah *activity diagram* untuk sistem klasifikasi data text positif dan negatif pada Gambar 5 :



Gambar 5. Activity Diagram

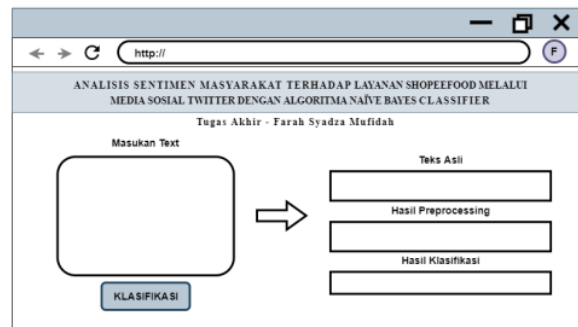
Activity diagram pada sistem untuk klasifikasi data text positif dan negatif dapat dilakukan oleh *developer* dan *user*. Untuk melakukan proses klasifikasi maka aktor harus mengakses sistem terlebih dahulu, jika sistem sudah menampilkan halaman utama maka aktor dapat langsung menginputkan data teks kedalam sistem yang selepas itu sistem akan menjalani proses *preprocessing* terlebih dahulu sebelum melakukan proses klasifikasi. Setelah sistem selesai memproses maka sistem akan menampilkan hasil dari *preprocessing* yang outputnya berupa data teks bersih serta hasil klasifikasi yang didapatkan.

3. Desain Interface

Berikut ini merupakan penggambaran desain *interface* sistem *text mining* yang diusulkan :



Gambar 6. Halaman Utama



Gambar 7. Halaman Proses Klasifikasi

4. DISKUSI

Berdasarkan hasil pada penelitian ini didapatkan beberapa temuan, yaitu yang pertama pada data *tweet* setelah dilakukan klasifikasi didapatkan banyaknya opini positif dibandingkan sentimen negatif. Temuan kedua yaitu data bersifat *overfitting* karena nilai *accuracy* pada data *training* lebih tinggi dibandingkan data *testing* dengan menggunakan metode *naive bayes*. Temuan yang lainnya yaitu pada komparasi metode *naive bayes*, KNN, SVM, dan *Decision Tree* dengan dataset yang sama menunjukkan bahwa metode *naive bayes* dan SVM menghasilkan performa yang ‘Sangat Baik’, sedangkan dengan menggunakan metode pembanding KNN menghasilkan performa yang ‘Baik’, begitu pula dengan metode pembanding *Decision Tree* yang menghasilkan performa ‘Baik’ pula.

Hal yang penting pada metode klasifikasi yaitu data pada setiap kelas harus *balance* agar performa yang dihasilkan baik, selain itu jika data *training* banyak dan bervariasi dapat menghindari terjadinya *overfitting* maupun *underfitting* karena semakin banyak data *training* membuat sistem dapat mengenali data *testing* dengan tepat. Pada penelitian ini memiliki kekurangan yang perlu dilakukan penyempurnaan untuk penelitian selanjutnya, yaitu tahapan *preprocessing*. Tahapan *preprocessing* terutama *stopword removal* pada penelitian ini hanya mengandalkan korpus yang tersedia pada *library NLTK* saja, sehingga pada penelitian ini tahapan *preprocessing* yang dilakukan kurang maksimal.

5. KESIMPULAN

Berdasarkan hasil pembahasan, pengujian yang sudah dilakukan, serta temuan dalam penelitian ini, maka didapatkan kesimpulan bahwa metode *naive bayes* mampu mengklasifikasikan sentimen positif dan negatif berupa data teks dengan baik dibuktikan dengan hasil performa yang didapatkan pada data *training* yaitu memiliki tingkat *accuracy* sebesar 98,17%, *precision* 97,68%, dan *recall* 98,69%. Sedangkan pada data *testing* memiliki performa dengan tingkat *accuracy* 90,62%, *precision* 88,23%, dan *recall* 93,75%. Hasil dari perhitungan AUC menghasilkan nilai sebesar 0.97 yang dimana nilai tersebut menunjukkan performa “Sangat Baik”. Suatu model dikatakan sempurna jika menghasilkan nilai AUC sebesar 1 yang artinya 100% area dibawah kurva. Maka kemampuan metode *naive bayes* yang digunakan sebagai model pada penelitian ini dinyatakan memiliki 97% area dibawah kurva.

DAFTAR PUSTAKA

- [1] G. Pratama, "Analisis Transaksi Jual Beli online Melalui Website Marketplace Shopee Menurut Konsep Bisnis di Masa Pandemi Covid 19," *Ecopreneur: Jurnal Program Studi Ekonomi Syariah*, vol. 1, no. 2, pp. 21-34, 2020.
 - [2] I. Vania and R. Simbolon, "Pengaruh Promo ShopeeFood Terhadap Minat Beli Pengguna Shopee (Di Daerah Tangerang Selatan)," *Jurnal Ekonomis*, vol. 14, no. 2b, 2021.
 - [3] E. S. Buana, "Pengaruh Promo ShopeeFood Terhadap Minat Beli Pengguna Shopee," 2021.
 - [4] N. M. S. Hadna, P. I. Santosa and W. W. Winarno, "Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter," *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*, pp. 57-64, 2016.
 - [5] D. N. Sari, F. Adelia, F. Rosdiana, B. B. Butar and M. Hariyanto, "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier," *JIKA (Jurnal Informatika)*, vol. 4, no. 3, pp. 109-118, 2020.
 - [6] A. Nurzahputra and M. A. Muslim, "Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing," *Seminar Nasional Ilmu Komputer (SNIK 2016)*, pp. 114-118, 2016.
 - [7] Surohman, S. Aji, Rousyati and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *Jurnal Sains dan Manajemen*, vol. 8, no. 1, pp. 93-105, 2020.
 - [8] M. I. Fikri, T. S. Sabrila and Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 2, pp. 71-76, 2020.
 - [9] Z. E. Sholikha, E. Y. Puspaningrum and W. S. JS, "Analisa Sentimen Pengguna E-Money pada Twitter menggunakan Algoritma C4.5 dan Naive Bayes," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, vol. 1, no. 3, pp. 1063-1071, 2020.
 - [10] D. Alita, Y. Fernando and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," *Jurnal Tekno Kompak*, vol. 14, no. 2, pp. 86-91, 2020.
-

Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier

ORIGINALITY REPORT

13%

SIMILARITY INDEX

11%

INTERNET SOURCES

4%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	humanika.penapersada.com Internet Source	1%
2	123dok.com Internet Source	1%
3	jurnal.unai.edu Internet Source	1%
4	pdfs.semanticscholar.org Internet Source	1%
5	Fajar Romadoni, Yuyun Umaidah, Betha Nurina Sari. "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2020 Publication	1%
6	repository.uin-suska.ac.id Internet Source	1%

repository.usd.ac.id

7	Internet Source	1 %
8	p3m.sinus.ac.id Internet Source	<1 %
9	coek.info Internet Source	<1 %
10	jurnal.untan.ac.id Internet Source	<1 %
11	download.atlantis-press.com Internet Source	<1 %
12	repository.trisakti.ac.id Internet Source	<1 %
13	Frizka Fitriana, Ema Utami, Hanif Al Fatta. "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes", Jurnal Komtika (Komputasi dan Informatika), 2021 Publication	<1 %
14	Hermanto, Antonius Yadi Kuntoro, Taufik Asra, Eri Bayu Pratama, Lasman Effendi, Ridatu Ocanitra. "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm And Support Vector Machine Based Smote Technique", Journal of Physics: Conference Series, 2020	<1 %

15

Rifqatul Mukarramah, Dedy Atmajaya, Lutfi Budi Ilmawan. "Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter", ILKOM Jurnal Ilmiah, 2021

Publication

<1 %

16

Rizal Rachman, Rissa Nurfitriana Handayani. "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM", Jurnal Informatika, 2021

Publication

<1 %

17

anzdoc.com

Internet Source

<1 %

18

doku.pub

Internet Source

<1 %

19

ejournal.nusamandiri.ac.id

Internet Source

<1 %

20

journal.binadarma.ac.id

Internet Source

<1 %

21

www.psychology.gatech.edu

Internet Source

<1 %

22

Ikhwanul Hakim, Arifin Nugroho, Sulaeman Hadi Sukmana, Windu Gata. "Sentimen Analisis Stay Home menggunakan metode

<1 %

klasifikasi Naive Bayes, Support Vector Machine, dan k-Nearest Neighbor",
Paradigma - Jurnal Komputer dan Informatika,
2020

Publication

23	core.ac.uk Internet Source	<1 %
24	muaramedia.com Internet Source	<1 %
25	openlibrarypublications.telkomuniversity.ac.id Internet Source	<1 %
26	sakup.pertanian.go.id Internet Source	<1 %
27	www.acarindex.com Internet Source	<1 %
28	www.researchgate.net Internet Source	<1 %
29	www.scribd.com Internet Source	<1 %
30	repository.its.ac.id Internet Source	<1 %
31	Anggit Dwi Hartanto, Ema Utami, Sumarni Adi, Harish Setyo Hudnanto. "Job Seeker Profile Classification of Twitter Data Using the Naïve Bayes Classifier Algorithm Based on the DISC	<1 %

Method", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2019

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography On