

SENTIMEN ANALISIS APLIKASI FERIZY MENGUNAKAN ALGORITMA NAIVE BAYES

Beni Rahmatullah¹; Pungkas Budiyo²; Suwanda Aditya Saputra³
Universitas BSI

beni.brh@bsi.ac.id¹, pungkas.pby@bsi.ac.id²,
suwanda.sdz@bsi.ac.id³

ABSTRAK : Transportasi merupakan kebutuhan yang paling digunakan dalam aktifitas sehari-hari dalam bekerja, dan kegiatan yang lainnya. Transportasi laut salah satunya menjadi yang diminati oleh masyarakat terlebih lagi Indonesia negara kepulauan. Penggunaan aplikasi Ferizy menjadi salah satu penunjang dan mempermudah dalam membeli tiket secara online, Saran dan kritik dari pelanggan guna memperbaiki sistem dan pelayanan yang diberikan. Dalam hal ini penulis memperoleh data dari komentar di playstore dan menggunakan algoritma Naive Bayes. Hasil akurasi yang didapat membuktikan komentar negative tertinggi dengan dengan Akurasi 98 % dan AUC 0.988.

Kata Kunci: Naive Bayes, Transportasi, Ferizy

1. PENDAHULUAN

Sosial media telah menjadi informasi yang dibutuhkan seiring dengan berkembangnya teknologi dan internet (Joshi, Prajapati, & Vala, 2017). Media sosial seperti blog, situs jejaring social, komunitas konten, dan dunia virtual menjadi salah satu sumber paling kuat untuk berita, berbelanja, industri, dan banyak lagi (Duwairi, Marji, Sha'Ban, & Rushaidat, 2014). Analisis sentimen juga telah diterapkan pada berbagai rekayasa perangkat lunak (Lin, Zampetti, Bavota, Di Penta, Lanza, & Oliveto, 2018). Analisis sentimen adalah penelitian dalam bidang pemrosesan teks (Medhat, Hassan, & Korashy, 2014)..

Pemrosesan sentimen terhadap teks adalah salah satu bagian penting dalam proses komputasi (Ren & Wang, 2017). Pelabelan teks menjadi positive dan negative (Stine, 2019). Data yang sudah didapat dilabeli positif dan negatif. Setelah itu dilakukan *preprocessing* baik itu mengubah kata tidak baku menjadi baku atau biasa disebut normalisasi menggunakan kamus dan mencari akar kata yaitu *stemming* dengan bantuan aplikasi Sastrawi Master. Selanjutnya dilakukan juga tokenisasi dan filter terhadap kalimat, kemudian menghilangkan kata-kata yang umum digunakan dan tidak mempunyai Informasi yang berharga pada suatu konteks atau biasa disebut *stopword removal*. Algoritma klasifikasi yang digunakan yaitu *Naive Bayes*.

2. Metode Penelitian

Dalam riset ini dilakukan secara kualitatif dan dimulai dengan pengumpulan data atau penarikan data

ulasan sampai di tahap terakhir didapatkan hasil pemrosesan.



Gambar 1, Kerangka Kerja

2.1. Data Mining

Data mining merupakan proses yang memakai kecerdasan buatan, matematika, statistik, serta metode pendidikan mesin buat mengenali serta mengekstrak data yang bermanfaat dari bermacam database skala besar (Ipmawati, Kusriani & Luthfi, 2017). Data mining pula didefinisikan selaku proses otomatis dari informasi yang sangat besar serta bertujuan buat memperoleh ikatan ataupun pola yang menguntungkan. Data mining bisa dikatakan

selaku proses pendukung pengambilan keputusan yang mencari pola data dalam informasi. Pencarian ini bisa dicoba oleh pengguna (Sari, 2013)

2.2. Klasifikasi

Klasifikasi ialah sesuatu proses yang bertujuan buat memastikan sesuatu obyek kedalam sesuatu kelas ataupun jenis yang telah ditetapkan lebih dahulu. Klasifikasi merupakan proses dari pembangunan terhadap sesuatu model yang mengklasifikasikan sesuatu objek cocok dengan atribut- atributnya. Klasifikasi informasi maupun dokumen pula bisa diawali dari membangun ketentuan klasifikasi tertentu yang memakai data training yang kerap diucap selaku tahapan pembelajaran serta pengujian digunakan selaku data testing (Raharjo, Suwanto & Edi Winarko, 2014). Dalam riset ini memakai satu algoritma saja, ialah *Naïve Bayes*.

2.3. Naive Bayes

Metode ini memakai *teorema Bayes*, yang ditemukan oleh Thomas Bayes abad ke- 18 (Suyanto, 2017). *Naïve Bayes Classification (NB)* merupakan klasifikasi statistik yang bisa digunakan memprediksi probabilitas keanggotaan dalam sesuatu kelas. Bagi Wu serta Kumar, algoritma *Naïve Bayes* ialah metode klasifikasi yang terkenal serta tercantum dalam 10 algoritma terbaik dalam data *mining*. *Naïve Bayes* memakai cabang matematika yang diketahui dengan teori probabilitas guna menciptakan kesempatan terbanyak dari mungkin klasifikasi dengan memandang frekuensi tiap klasifikasi pada data latih (Mukminin & Riana, 2017)

2.4. Analisis Sentimen

Analisis sentimen bertujuan untuk menentukan bagian mana saja penulis menentukan sikap keterkaitan dengan barang, produk, merek, atau layanan tertentu. Dalam analisis sentimen, fokusnya adalah pada kalimat subyektif dan untuk menemukan dan mengekstrak informasi subyektif dari teks yang diberikan (Vinodhini & Chandrasekaran, 2012) .

3. Hasil dan Pembahasan

Penelitian ini menggunakan data yang telah diambil dari komentar di playstore seperti yang telah disebutkan pada bagian *data understanding* diatas. Adapun data yang diambil secara keseluruhan adalah berjumlah 100 data komentar. Kemudian data tersebut pada tahap awal akan dilakukan *data cleansing* dan dilakukan pemberian label dengan label positif atau negatif pada setiap komentar. Dari tahap awal tersebut didapatkan data sebanyak 100 komentar yang

telah berlabel maka, data tersebutlah yang akan menjadi dataset pada penelitian ini.

3.1. Pre-processing

Pembahasan pada tahapan ini adalah proses awal pengolahan dataset sebelum dapat diproses untuk klasifikasi dengan algoritma *Naïve Bayes (NB)*. Penelitian ini menggunakan beberapa tahapan *pre-processing* untuk dataset teks komentar, berikut ini adalah tahapan tersebut:

a. Case Folding

Tahap ini adalah proses untuk mengkonversi teks yang tidak beraturan dalam penggunaan huruf di penulisan teks komentar, sehingga teks komentar yang ditulis tidak konsisten. Proses *case folding* ini berfungsi untuk mengubah huruf pada teks komentar kedalam bentuk standar yaitu, menjadi huruf kecil seluruhnya. Hasil dari proses *case folding* tersebut dapat dilihat dibawah ini.

Tabel 1. Perbandingan Teks Komentar Sebelum dan Sesudah Proses *Case Folding*

Sebelum	Sesudah
Beli tiket kapal dijamin cepat lemotnyaðŸ¥£	beli tiket kapal dijamin cepat lemotnyaðŸ¥£
Aplikasi lemot	aplikasi lemot
Kenapa tidak bisa buka aplikasinya	kenapa tidak bisa buka aplikasinya
knapa gak bisa dibuka sih	knapa gak bisa dibuka sih
Aplikasi tidak bisa di gunakan	aplikasi tidak bisa di gunakan
Lemot amat woi!!! mau login susah.	lemot amat woi!!! mau login susah.

Sumber : Rahmatullah (2023)

b. Tokenizing

Tahap *Tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Proses ini menghasilkan kalimat yang memiliki simbol khusus akan di hilangkan seperti pada kalimat nomor 5. Hasil dari proses tersebut dapat dilihat pada tabel dibawah ini.

Tabel 2. Perbandingan Teks Komentar Sebelum dan Sesudah Proses *Tokenizing*

Sebelum	Sesudah
---------	---------

Beli tiket kapal dijamin cepat lemotnya Aplikasi lemot	Beli tiket kapal dijamin cepat lemotnya Aplikasi lemot
Kenapa tidak bisa buka aplikasi nya knapa gak bisa dibuka sih	Kenapa tidak bisa buka aplikasi nya knapa gak bisa dibuka sih
Aplikasi tidak bisa di gunakan Lemot amat woi!!! mau login susah.	Aplikasi tidak bisa di gunakan Lemot amat woi mau login susah.

c. Filter Token

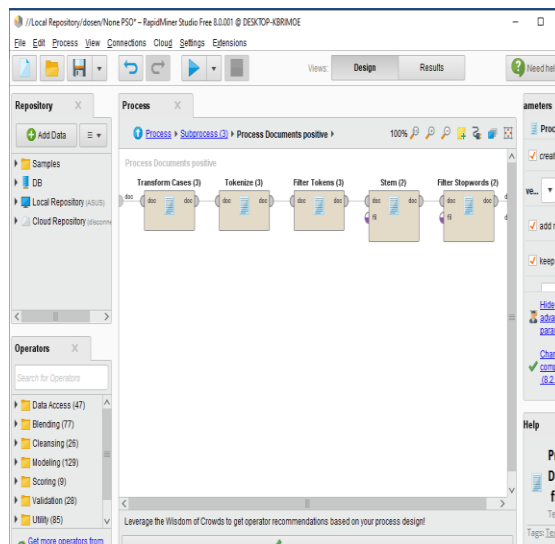
Filter Token untuk membatasi jumlah kata minimal dan maksimal dalam suatu kalimat yang telah diproses. Jika kata yang muncul jumlahnya berada dibawah batas minimal dan maksimal maka tidak akan muncul. Proses dapat dilihat pada tabel dibawah ini.

Tabel 3. Perbandingan Teks Komentar Sebelum dan Sesudah Proses Filter Token

Sebelum	Sesudah
Beli tiket kapal dijamin cepat lemotnya Aplikasi lemot	beli tiket kapal dijamin cepat lemotnya
Kenapa tidak bisa buka aplikasi nya knapa gak bisa dibuka sih	kenapa tidak bisa buka aplikasi knapa bisa dibuka
Aplikasi tidak bisa di gunakan Lemot amat woi!!! mau login susah.	aplikasi tidak bisa digunakan lemot amat login susah

Tabel 1 dan Tabel 2 diolah dengan menggunakan tools yang terdapat pada rapidminer.

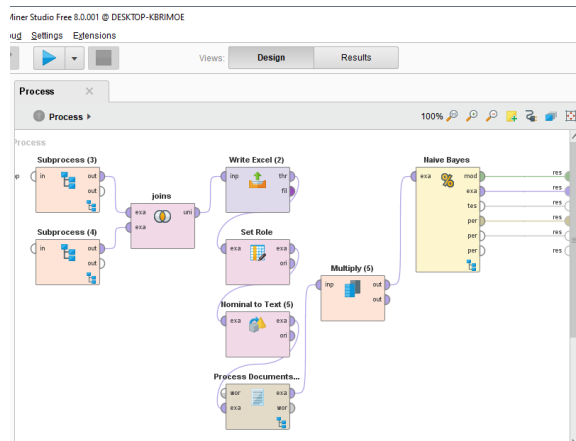
Pada gambar 2 dibawah proses data pertama akan dilakukan proses transform Cases untuk merubah semua kata menjadi huruf kecil setelah di proses tahap selanjutnya melalui Tokenize. Proses Tokenize menghilangkan simbol tanda baca. Filter Token melakukan pembatasan minimum sebuah kata. Stemming digunakan untuk merubah sebuah kata menjadi kata dasar tanpa imbuhan.



Gambar 2. Tahap Preprocessing.

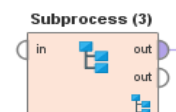
3.2. Pembuatan Model Klasifikasi

Tahap ini pada penelitian adalah untuk membuat model dengan menggunakan algoritma klasifikasi untuk dataset teks komentar yang sudah melalui tahap *pre-processing*. Tahap ini menggunakan algoritma klasifikasi *Naive Bayes* (NB) dan menggunakan *tools* Rapidminer versi 8.0 untuk mengolah dataset teks komentar yang sudah melalui tahap *data preparation* dengan *pre-processing* teks. Tahap pertama dari proses ini adalah data teks komentar akan diunggah ke dalam tools dengan menggunakan file excel yang kemudian akan di proses dengan algoritma *Naive Bayes* (NB) untuk mendapatkan hasil awal dari algoritma tersebut.



Gambar 3. Design Model Algorithm Naive Bayes

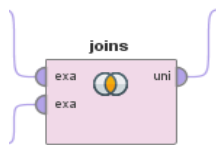
a. Operator subprocess



Gambar 4. Tahap Preprocessing.

Operator ini digunakan sebagai penampung operator lain dalam proses berbeda dan bertujuan untuk memisahkan pemrosesan dan output yang dihasilkan hasil dari tahap preprocessing

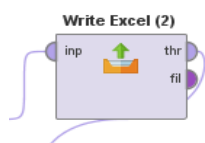
b. Operator join



Gambar 5. Tahap Preprocessing.

Digunakan untuk menggabungkan data ulasan *positive* dan *negative* antar *subprocess* yang sudah dilakukan.

c. Operator Write Excel



Gambar 6. Tahap Preprocessing.

Proses ini untuk menghasilkan *output* dalam bentuk excel dari hasil operator join

3.3. Evaluasi Model Klasifikasi

Tahapan evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil dibuat pada langkah sebelumnya. Untuk evaluasi digunakan *10 fold-cross validation*. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *pre-processing*. Data tersebut diambil dari operator *Read Excel*, hal ini dilakukan karena dataset disimpan dalam bentuk Excel. *Process documents from files* untuk mengkonversi file menjadi dokumen. Proses validasi terdiri dari data pelatihan dan data pengujian. Pada tahap ini juga menggunakan *Set Role* yang berfungsi untuk menentukan *field* pada *class*.

3.4. Hasil nilai Algoritma Naive Bayes

Hasil percobaan yang didapat dari 100 data komentar pengguna di playstore menghasilkan nilai sebagai berikut :

a. Nilai *accuracy*

accuracy: 82.00% +/- 12.49% (mikro: 82.00%)

	true Positive	true Negative	class precision
pred. Positive	23	13	63.89%
pred. Negative	5	59	92.19%
class recall	82.14%	81.94%	

Gambar 7. Tahap Preprocessing.

Terdapat 23 *True Positive (TP)* yang tergolong positif dan *False Positive (FP)* sebanyak 5 *record* yang tergolong Negatif pada nilai akurasi. Selanjutnya, 13 Negatif Benar (TN) diklasifikasikan sebagai Negatif dan 59 catatan Negatif Palsu (FN) diklasifikasikan sebagai Negatif. Berdasarkan tabel diatas terlihat bahwa tingkat akurasi *review* pengguna aplikasi ferizy menggunakan algoritma *Naive Bayes* adalah 82,00%.

b. Nilai *Precision*

precision: 93.25% +/- 10.64% (mikro: 92.19%) (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	23	13	63.89%
pred. Negative	5	59	92.19%
class recall	82.14%	81.94%	

Gambar 8. Tahap Preprocessing.

Terdapat 23 *True Positive (TP)* yang tergolong positif dan *False Positive (FP)* sebanyak 5 *record* yang tergolong Negatif pada nilai akurasi. Selanjutnya, 13 Negatif Benar (TN) diklasifikasikan sebagai Negatif dan 59 catatan Negatif Palsu (FN) diklasifikasikan sebagai Negatif. Berdasarkan tabel diatas terlihat bahwa tingkat akurasi *review* pengguna aplikasi ferizy menggunakan algoritma *Naive Bayes* adalah 93.25%.

c. Nilai *Recall*

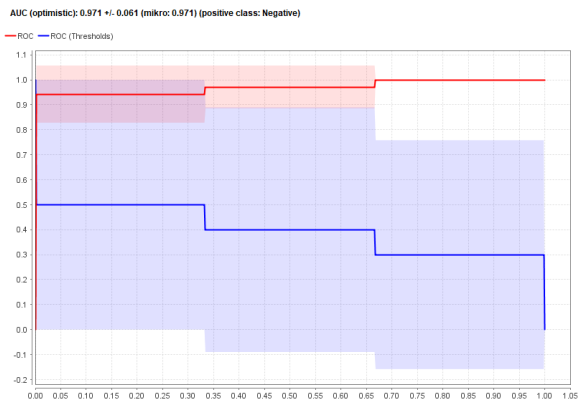
recall: 82.32% +/- 13.28% (mikro: 81.94%) (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	23	13	63.89%
pred. Negative	5	59	92.19%
class recall	82.14%	81.94%	

Gambar 9. Tahap Preprocessing.

Terdapat 23 *True Positive (TP)* yang tergolong positif dan *False Positive (FP)* sebanyak 5 *record* yang tergolong Negatif pada nilai akurasi. Selanjutnya, 13 Negatif Benar (TN) diklasifikasikan sebagai Negatif dan 59 catatan Negatif Palsu (FN) diklasifikasikan sebagai Negatif. Berdasarkan tabel diatas terlihat bahwa tingkat akurasi *review* pengguna aplikasi ferizy menggunakan algoritma *Naive Bayes* adalah 82.32%.

3.5. Nilai *AUC* Aplikasi Ferizy



Gambar 10. AUC Algoritma Naive Bayes

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Kurva ROC Naive Bayes dengan nilai AUC (Area Under Curve) yang dihasilkan dari gambar 5 diatas sebesar 0.971 dimana diagnosa hasilnya adalah *excellent classification*

4. Kesimpulan

Dari hasil penelitian yang telah dilakukan terkait dengan sentimen analisis aplikasi ferizy pada komentar playstore, penggunaan Algoritma Naive Bayes dengan tingkat akurasi 82%, nilai recall 82.32%, nilai precision 93.25%. Hasil dari proses yang dilakukan dapat disimpulkan bahwa sentimen *negative* lebih banyak dari sentimen positif.

Ucapan Terima Kasih

Peneliti mengucapkan banyak terimakasih kepada Suwanda Aditya dan Pungkas Budiono yang telah memberikan pembiayaan penelitian ini dan telah mendukung selama berjalannya kegiatan penelitian.

5. DAFTAR PUSTAKA

- [1] Joshi, M., Prajapati, P., Shaikh, A., & Vala, "A Survey on Sentiment Analysis," International Journal of Computer Applications., 163(6), 34–38. doi:10.5120/ijca2017913552, 2017.
- [2] Duwairi, R. M., Marji, R., Sha'Ban, N., & Rushaidat, S. (2014). Sentiment analysis in arabic tweets. In 2014 5th International Conference on Information and Communication Systems, ICICS 2014. doi:10.1109/IACS.2014.6841964.
- [3] Lin, B., Zampetti, F., Bavota, G., Di Penta, M., Lanza, M., & Oliveto, R. (2018). Sentiment

analysis for software engineering (pp. 94–104). doi:10.1145/3180155.3180195.

- [4] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. doi:10.1016/j.asej.2014.04.011.
- [5] Ren, F., & Wang, L. (2017). Sentiment nalysis of text based on three-way decisions. Journal of Intelligent and Fuzzy Systems, 33(1), 245–254. doi:10.3233/JIFS-161522.
- [6] Stine, R. A. (2019). Sentiment Analysis. Annual Review of Statistics and Its Application, 6(1), 287–308. doi:10.1146/annurev-statistics-030718-105242
- [7] J. Ipmawati, Kusriani, and E. Taufiq Luthfi, (2017) "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," Indones. J. Netw. Secur.
- [8] E. N. Sari, (2013) "Analisa Algoritma Apriori untuk Menentukan Merek Pakaian yang Paling Diminat i pada Mode Fashion Group Medan," Pelita Inform. Budi Darma.
- [9] Susilowati, Elly, et al (2015) "Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter." E-Proceeding of Engineering, vol. 2, no. 1, pp. 1–7.
- [10] Raharjo, Suwanto, and Edi Winarko (2014) "Klasterisasi, Klasifikasi Dan Peringkasan Teks Berbahasa Indonesia.", vol. 8, no. Kommit, 2014, pp. 391–401.
- [11] Suyanto (2017) "Data Mining untuk Klasifikasi dan Klasterisasi Data." Bandung: Informatika Bandung.
- [12] A. Mukminin and D. Riana (2017) "Komparasi Algorit ma C4 . 5 , Naive Bayes Dan Neural Network Untuk Klasifikasi Tanah," J. Inform.
- [13] G. Vinodhini dan R. M. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey," Int. J. of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, hal. 282-292, 2012