

SENTIMEN ANALISIS PENGGUNA PRODUK PONSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES

Nurfaizah^{1)*}, Salafudin Ridwan Hidayat²⁾

^{1),2)} Sistem Informasi Universitas Amikom Purwokerto

email : nurfaizah@amikompurwokerto.ac.id¹⁾, salafudinridwan.ok@gmail.com²⁾

Abstraksi

Penggunaan ponsel di Indonesia sudah tidak menjadi barang yang asing lagi hampir seluruh kalangan menggunakan ponsel untuk berbagai kebutuhan dari pendidikan, bisnis dan *lifestyle*. Berkembangnya kebutuhan ponsel sejalan dengan berkembangnya berbagai merk ponsel yang menawarkan berbagai keunggulan dengan harga yang terjangkau. Penelitian ini akan melihat bagaimana sentimen dari pengguna produk ponsel dimana dasar penentuan sentimen analisis terhadap produk diambil dari komentar konsumen di platform X dengan menggunakan algoritma naïve bayes. Dimana algoritma naïve bayes adalah algoritma klasifikasi yang berdasarkan probabilitas sederhana dengan keunggulan memiliki tingkat nilai error yang rendah dan memiliki akurasi yang tinggi. Adapun dataset yang digunakan dalam penelitian ini berjumlah 244 data dengan tingkat akurasi yang dihasilkan dari algoritma naïve bayes sebesar 88.93%.

Kata Kunci:

Sentimen analisis, ponsel, naïve bayes

Abstract

The use of mobile phone in Indonesia is no longer a strange thing, almost all people use mobile phone for various needs from education, business, and lifestyle. The growing need for mobile phone is in line with the development of various mobile phone brands that offer various advantages at affordable prices. This research will look at the sentiment of mobile phone product users where the basis for determining product sentiment analysis is taken from consumer comments on X platform using the Naïve Bayes algorithm. Whereas the naïve Bayes algorithm is a classification algorithm based on simple probability with the advantage of having a low error rate and high accuracy. The dataset used in this research consisted of 24 data with an accuracy level resulting from the Naïve Bayes algorithm of 88.93%.

Keywords :

Sentiment analysis, mobile phone, naïve bayes

Pendahuluan

Peningkatan penggunaan ponsel di Indonesia hampir 190 juta pengguna dan akan meningkat setiap tahunnya seperti dikutip dari data statistika [1]. Menurut data laporan statistik juga menunjukkan pengguna ponsel sudah mencapai 73.58 % dari total penduduk perkotaan dan 60.18% dari seluruh penduduk perdesaan [2].

Peningkatan tersebut tidak lepas dari pengaruh dari semakin banyaknya produk yang menawarkan berbagai kemudahan dan kelebihan mulai dari penawaran keunggulan fitur yang setiap saat diperbaharui dan penawaran harga antar produk.

Semakin banyaknya pengguna ponsel dan juga meningkatnya berbagai jenis produk ponsel dari berbagai produk mengharuskan para produsen terus meningkatkan berbagai promo untuk produk-produk terbarunya dimana pelanggan juga dapat dengan mudah memberikan penilaian terhadap produk yang dibelinya melalui platform social. Adanya komentar pengguna dapat mempengaruhi kepercayaan calon konsumen atas suatu produk.

Banyak pengguna X memberikan ulasan atas produk ponsel yang mereka gunakan baik berupa komentar

atau opini baik maupun buruk. Data X memiliki karakteristik tidak terstruktur dan memiliki banyak noise sehingga membutuhkan teks mining untuk menggali informasi dari komentar dan opini tersebut. Sentimen analisis merupakan salah satu cara untuk melihat penilaian konsumen berdasarkan komentar yang dituliskan pada social media. Salah satu algoritma yang dapat digunakan untuk melakukan sentiment analisa pengguna yaitu algoritma naïve bayes.

Algoritma naïve bayes merupakan algoritma yang probabilitas yang memiliki tingkat akurasi yang tinggi [3]. Selain itu algoritma naïve bayes merupakan algoritma berdasarkan pada teorema bayes, dapat dikatakan bahwa naïve bayes bukan algoritma tunggal tetapi satu grup algoritma dimana masing-masing anggota algoritma memiliki prinsip kerja yang mirip, dimana algoritma ini bekerja berdasarkan prinsip probabilitas dari teorema bayes [4]. Adapun ciri utama dari naïve bayes merupakan asumsi yang sangat kuat atau sering disebut sebagai naïf akan independensi dari masing-masing kondisi atau kejadian yang ada [5] atau dapat dikatakan

naïve bayes memiliki nilai kemandirian yang kuat [6].

Penelitian melakukan sentiment analisis pengguna produk ponsel yang diambil komentar dari platform X dengan mengklasifikasikan komentar positif dan komentar negatif serta memberikan tingkat akurasi dan penggunaan algoritma naïve bayes. Adapun dataset yang digunakan dalam penelitian ini sebanyak 244 data yang diambil dari pada tahun 2020.

Tinjauan Pustaka

Pada penelitian sebelumnya penggunaan algoritma naïve bayes sudah berhasil digunakan pada penelitian sentiment analisis pengguna aplikasi mypertamina dengan akurasi 72% [7]. Penelitian serupa juga telah dilakukan untuk menilai produk amazon dengan tingkat akurasi 100% dan berhasil melakukan sentimen atas komentar positif dan negatif [8]. Algoritma naïve bayes juga telah digunakan dalam menganalisa sistem informasi akademik dengan menghasilkan akurasi sebesar 84.55% dan mampu mengklasifikasikan sentiment positif dan negatif [6].

Pada penerapan penelitian sentimen analisis untuk ulasan Instagram dimana dataset diambil dari google play algoritma naïve bayes juga sudah memberikan akurasi tinggi [9] [10] dan mampu mengklasifikasikan komentar dengan cepat [11] naïve bayes algoritma yang yang mampu bekerja secara efektif bahkan dengan dataset dalam skala besar [12].

Penelitian ini menggunakan algoritma naïve bayes untuk melakukan proses klasifikasi terhadap komentar konsumen dari produk ponsel yang bersumber dari ulasan pengguna X atas sentimen positif dan negatif serta mencari hasil akurasi dari algoritma yang digunakan.

Metode Penelitian

Penelitian ini menggunakan algoritma naïve bayes digunakan untuk memperoleh hasil akurasi dari data komentar yang diambil dari komentar postingan twitter tentang produk ponsel dan digunakan untuk mengklasifikasi komentar positif dan negatif review produk dari data komentar yang telah diperoleh. Gambar 1 berikut adalah diagram alir penelitian yang dilakukan.



Gambar 1. Konsep Penelitian

1.Studi Literatur

Tahap ini digunakan penulis melakukan pencarian terhadap berbagai sumber tertulis, baik berupa buku-buku, artikel, dan jurnal, atau dokumen-dokumen yang relevan yang berkaitan dengan analisis

sentiment untuk mengetahui gambaran umum mengenai penelitian yang akan dilakukan.

2.Scrapping Data

Pada penelitian ini, data yang digunakan adalah data primer yaitu data yang diambil dari komentar-komentar pada produk ponsel di X.

3.Tahap Pre-processing

Tahap pre-processing adalah tahap yang bertujuan untuk menghasilkan data yang berkualitas baik sehingga menghasilkan keputusan yang baik. Adapun tahapan pre-processing yang dilakukan diantaranya pertama seleksi komentar yang mengandung hashtag(#) pada komentar X. Kedua proses Cleansing menghilangkan noise pada kalimat seperti kesalahan acak. Ketiga Punctuation Removal bertujuan untuk menghapus tanda baca yang ada pada dataset seperti tanda tanya (?), seru (!), titik (.), koma (,) dan tanda baca lainnya. Keempat Normalisasi Kalimat yang bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal, cek tanda baca (punctuation) dan simbol selain alphabet, mengubah semua huruf menjadi huruf kecil semua, menghilangkan huruf yang berulang, menghilangkan emoticon, normalisasi kata tidak baku menjadi kata baku. Kelima Tokenisasi Setelah normalisasi kalimat, selanjutnya kalimat tersebut di pecah kedalam token-token menggunakan pembatas atau delimiter spasi. Keenam Filtering adalah tahap menghilangkan kata-kata yang muncul dalam jumlah besar namun dianggap tidak memiliki arti (stopwords). Contohnya: “dan”, “di”, “yang”, “ke” dan seterusnya. Ketujuh Stemming adalah tahap untuk membuat akhiran kata menjadi kata dasar dan akan dicari akar kata kemudain menghilangkan imbuhan pada kata.

4. Klasifikasi Algoritma Naïve bayes

Penggunaan algoritma naïve bayes, naïve bayes membuat asumsi yang sangat kuat akan independensi dari masing-masing kelas. Naïve bayes digunakan untuk memperoleh hasil akurasi dari data komentar yang diambil dari komentar postingan twitter tentang produk ponsel dan digunakan untuk mengklasifikasi komentar positif dan negatif review produk dari data komentar yang telah diperoleh.

Pada proses evaluasi yang dilakukan berdasarkan pada kategori komentar positif dan negatif review produk menggunakan bantuan confusion matrix untuk menentukan nilai accuracy, nilai precision dan nilai f-measure, nilai recall.

5.Evaluasi

Tahap selanjutnya evaluasi berupa penarikan kesimpulan hasil akhir. Dalam tahap ini penulis melakukan penyimpulan dari hasil penelitian, penarikan kesimpulan berdasarkan hasil dari kelas positif dan kelas negatif dari data komentar review produk, perhitungan akhir dengan menggunakan confusion matrix meliputi nilai accuracy, nilai precision, nilai f-measure, dan nilai recall yang diperoleh menggunakan algoritma Naïve Bayes

Hasil dan Pembahasan

1. Scrapping Data

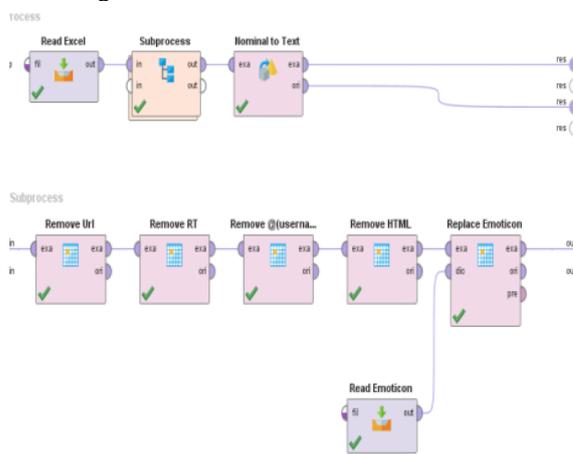
Merupakan proses pengambilan data dari platform social media X, dimana diperoleh dataset sebanyak 244 data.

2. Pre Processing

a. Cleansing

Tahapan ini bertujuan untuk membersihkan noise. Noise merupakan suatu bentuk data yang nantinya akan mengganggu proses pengolahan data tersebut. Noise tersebut diantaranya adalah url, RT, @username, #(hashtag), HTML dan Emoticon. Untuk emoticon tidak dihapus tetapi diganti dengan arti yang sesungguhnya misalnya emoticon “ :(“ emoticon disamping menunjukkan arti “sedih”.

Gambar 2 berikut proses pre processing cleansing



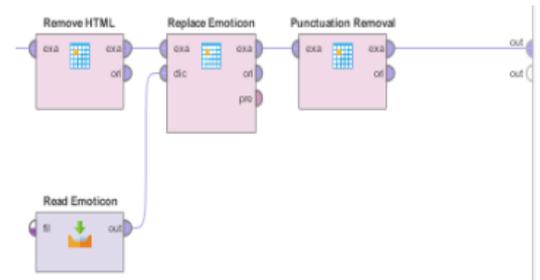
Gambar 2. Proses Cleansing

Pada aplikasi rapidminer Operator Read Excel merupakan operator yang digunakan untuk membaca data dalam file excel, operator subprocess digunakan untuk pra-pengolahan dan operator nominal to text digunakan untuk menyaring attribut apa yang akan digunakan.

Selanjutnya dari operator subprocess. Didalamnya terdapat operator replace yang didalamnya dimasukan perintah untuk menghapus noise seperti menghapus url, rt,@username, #(hashtag), HTML dan emoticon.

b. Punctual Removal

Punctual Removal bertujuan untuk menghapus tanda baca yang ada pada dataset seperti tanda tanya (?), seru (!), titik(.), koma(,) dan tanda baca lainnya. Gambar 3 adalah hasil tahapan Punctual Removal.



Gambar 3 Proses Punctual Removal

c. Normalisasi kalimat

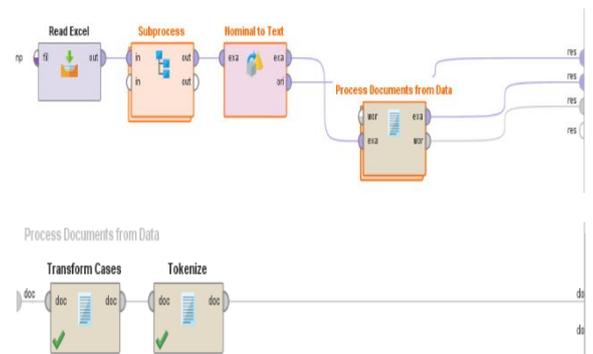
Keterbatasan karakter yang diberikan oleh X mengakibatkan terdapat beberapa kata yang sengaja disingkat oleh pengguna X agar tweet mencangkup opini pengguna. Selain itu juga terdapat beberapa kata yang memiliki arti sama, sehingga kata dengan arti sama diseragamkan. Begitu pula dengan bahasa asing dan bahasa dialek maupun bahasa gaul atau alay diubah ke dalam bahasa indonesia. Oleh karena itu proses normalisasi kalimat dilakukan agar meminimalisir pengulangan kata yang sama memiliki arti sama. Tabel 1 menunjukkan beberapa kata singkatan yang memiliki arti sama. Dalam proses ini peneliti menggunakan cara manual.

Tabel 1. kata yang di normalisasi

Sebelum Normalisasi	Setelah Normalisasi
Ad	Ada
Bgs	Bagus
Gk	Tidak
Gt	Begitu
Gw	Saya
Jgn	Jangan
Jlk	Jelek
Mhn	Mohon
Kpn	Kapan
Krn	Karena

d. Tokenisasi

Sebuah proses melakukan pemotongan kalimat menjadi bagian-bagian atau kata-kata berdasarkan tiap kata yang menyusunnya. Memecahkan atau memotong teks tersebut disebut dengan token.



Gambar 4. Proses Tokenisasi

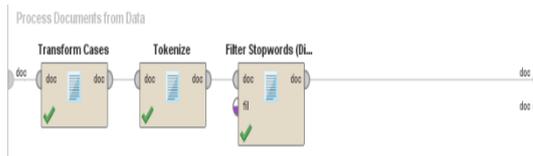
Tabel 2 berikut merupakan contoh hasil kata tokenisasi dari dataset

Tabel 2. Data Sesudah Proses Tokenisasi

Word	Attribute Name	Total Occurrences
abadikan	abadikan	1
ada	ada	4
andalan	andalan	3
apapun	apapun	1
awet	awet	8
bagi	bagi	1
bagus	bagus	113
banding	banding	1
banget	banget	4

e. Filtering

Stopword merupakan proses menghilangkan kata yang sering muncul tetapi tidak berpengaruh terhadap analisis sentimen. Contoh stopwords adalah “dan”, “di”, “yang”, “ke” dan seterusnya.



Gambar 5. Proses Stopwords

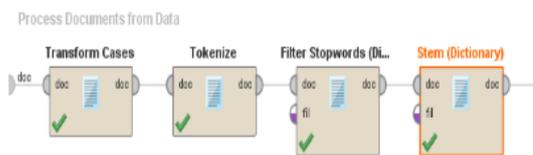
Tabel 3 berikut merupakan hasil dari proses stopwords

Tabel 3. Hasil stopwords

Word	Attribute Name	Total Occurrences
andalan	andalan	3
awet	awet	8
bagus	bagus	113

f. Stemming

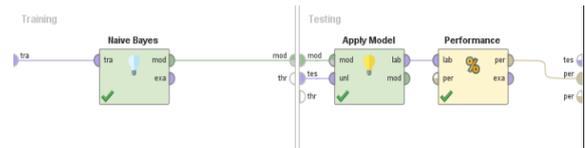
Pada proses stemming, akan dicari akar kata dan menghilangkan imbuhan pada kata. Stemming bertujuan untuk mengurangi variasi kata yang memiliki kata dasar sama.



Gambar 6. Proses Stemming

3. Naïve Bayes

Proses analisis komentar positif dan negatif review produk yang telah dilakukan pada process pre-processing, clustering k-means, tahap selanjutnya adalah proses klasifikasi naïve bayes . Pada tahap ini dilakukan dengan menggunakan bantuan aplikasi rapidminer.



Gambar 7. Pengujian algoritma naïve bayes

Hasil akurasi dari algoritma naïve bayes untuk mendapatkan hasil akurasi, peneliti menggunakan bantuan confusion matrix. Tabel 4 adalah tabel confusion matrix.

Tabel 4 Confusion Matrix

	True Negatif	True Positif
Pred.Negatif	161	3
Pred.Positif	24	56

Berdasarkan tabel 4 dapat dijelaskan True Positive (TP) merupakan kasus dimana tweet diprediksi positif, memang benar (True) tweet tersebut positif . True Negative (TN) merupakan kasus dimana tweet diprediksi negatif dan sebenarnya tweet tersebut memang (True) negatif. False Negative (FN) merupakan kasus dimana tweet yang diprediksi negatif, tetapi ternyata sebenarnya (True) positif. False Positive (FP) merupakan kasus dimana tweet yang diprediksi positif, ternyata negatif prediksinya salah (False)

Tahap selanjutnya melakukan evaluasi dengan mengguakan confusion matrix menghasilkan formulasi perhitungan sebagai berikut :

$$Precision = \frac{TP}{TP + FP} = \frac{161}{161 + 3} = 0,981707 = 98,17\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{161}{161 + 24} = 0,870270 = 87,03\%$$

$$F - Measure = 2 \frac{PRECISION \times RECALL}{PRECISION + RECALL} = 2 \frac{0,981707 \times 0,870270}{0,981707 + 0,870270} = 2 \frac{0,854350}{1,851977} = 0,922635 = 92,27\%$$

Berdasarkan perhitungan diatas antara lain precision recall, dan f-measure. Maka dapat disimpulkan hasil perhitungan akurasi.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{161 + 56}{161 + 3 + 24 + 56} = \frac{217}{244} = 0,889344 = 88,93\%$$

4. Evaluasi

Berdasarkan perhitungan diatas nilai akurasi berdasarkan confusion matrix yang diterapkan pada tabel 5 sebagai berikut:

Tabel 5 Performa Confusion Matrix

Precision	Recall	F-measure	Accuracy
98,17%	87,03%	92,27%	88,93%

Dari hasil perhitungan didapatkan precision sebesar 98,17%, recall 87,03%, f-measure 92,27% dan accuracy 88,93%. Untuk penjelasannya yaitu :

a. Precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil

- yang diprediksi positif. *precision* menjawab pertanyaan “Berapa persen *tweet* yang benar negatif dari keseluruhan *tweet* yang diprediksi negatif”
- b. *Recall* (sensitifitas) merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *recall* menjawab pertanyaan “Berapa persen *tweet* yang diprediksi negatif dibandingkan keseluruhan *tweet* yang sebenarnya negatif”
- c. *F-measure* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan.
- Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. *Accuracy* menjawab pertanyaan “Berapa persen *tweet* yang benar di prediksi negatif dan tidak negatif dari keseluruhan *tweet*”.

Kesimpulan dan Saran

Hasil dari tingkat akurasi algoritma naïve bayes untuk menganalisa sentimen produk ponsel diperoleh hasil 88.93%. Dimana jumlah komentar tertinggi adalah komentar positif sebanyak 185 data komentar positif. maka dapat disimpulkan bahwa metode klasifikasi Naïve Bayes yang digunakan cenderung stabil karena probabilitas kemunculan kata dalam nilai. Nilai akurasi adalah salah satu parameter penilaian dari metode yang digunakan, nilai akurasi didapat dari jumlah banyaknya data yang berhasil di klasifikasikan dengan benar sesuai kelas sentimennya dari seluruh jumlah data yang diklasifikasi.

Daftar Pustaka

- [1] Statista.com, “Number of smartphone users in Indonesia from 2018 to 2028,” 2023. [Online]. Available: <https://www.statista.com/>
- [2] Badan Pusat Statistik, “Statistik Telekomunikasi Indonesia 2022,” 2022.
- [3] M. Qamal *et al.*, “ANALISIS SENTIMEN TOKO ONLINE MENGGUNAKAN dilakukan oleh Mehdi Mursalat Ismail dan Kemas Muslim Lhaksamana dengan judul ‘ Sentimen Analisis Pada Media Online Mengenai Pemilihan Presiden 2019 dengan Menggunakan Metode Naive Bayes ’,” no. 1.
- [4] D. Berrar, “Bayes ’ Theorem and Naive Bayes Classifier Bayes ’ Theorem and Naive Bayes Classifier,” no. January 2018, pp. 0–18, 2019, doi: 10.1016/B978-0-12-809633-8.20473-1.
- [5] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” vol. 2, no. 4, pp. 148–155, 2022.
- [6] B. Irawan, A. Bahtiar, P. Studi, T. Informatika, K. Cirebon, and A. Adakami, “PENGUNAAN ALGORITMA NAÏVE

- BAYES DALAM MENGANALISIS,” vol. 7, no. 6, pp. 3879–3885, 2023.
- [7] I. G. Bgs, D. Putra, C. Rai, and A. Pramarthar, “Analisis Sentimen Pengguna Aplikasi MyPertamina Dengan Menggunakan Algoritma Naïve Bayes,” vol. 12, no. 3, pp. 633–638, 2024.
- [8] J. Yolanda, B. Sibarani, A. Junaidi, M. A. Azis, and Y. T. Arifin, “Analisis Sentimen Produk Amazon Menggunakan Algoritma Naïve Bayes Pada Data Review Pelanggan,” vol. 10, no. 1, pp. 13–19, 2024, doi: 10.31294/jtk.v4i2.
- [9] A. Syahjuddin and T. S. Rizky, “ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES TERHADAP PRODUK PT . IMIN TECHNOLOGY BERDASARKAN SENTIMENT ANALYSIS USING THE NAÏVE BAYES METHOD ON PT . IMIN TECHNOLOGY PRODUCTS BASED ON REVIEWS FROM,” vol. 2, no. September, pp. 445–452, 2023.
- [10] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [11] D. S. Sayogo *et al.*, “ANALISIS SENTIMEN ULASAN INSTAGRAM DI GOOGLE PLAY STORE,” vol. 7, no. 6, pp. 3314–3319, 2023.
- [12] A. F. Setyaningsih, D. Septiyani, and S. R. Widiyari, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia,” vol. 9, no. 1, pp. 224–235, 2023.