DOI: https://doi.org/10.51903/elkom.v17i1.1593

JURNAL ILMIAH ELEKTRONIKA DAN KOMPUTER, Vol.17, No.1, Juli 2024, pp.103-110
p-ISSN : 1907-0012 (print)
e-ISSN : 2714-5417 (online)
http://journal.stekom.ac.id/index.php/elkom

Analisis Sentimen Twitter Terhadap Kualitas Udara Jakarta Menggunakan Metode NBC

Dhani Wahyu Wicaksono[1], Budi Hartono[2]
[1]Teknik Informatika-Unisbank Semarang, dhaniwahyu2121@gmail.com
[2]Teknik Informatika-Unisbank Semarang, budihartono@edu.unisbank.ac.id
Jalan Tri Lomba Juang Semarang, Telp. (024) 8451976

ARTICLE INFO

Article history:

Received Octoberber 29, 2023
Received in form 2 November 2023
Accepted 14 Desember 2023
Available online 1 Juli 2024

ABSTRACT

According to the Jakarta Air Quality Index (AQI US) 12 July 2023, 200 indicates unhealthy air quality with an index value between 151 and 200. This figure even shows that Jakarta is currently the second most polluted city in Southeast Asia. (CNN Indonesia., 2023). This incident gave rise to responses from the public which were expressed via social media Twitter. From this incident, sentiment analysis was carried out regarding Jakarta's air quality. The amount of data used for this research was 500 tweet data. The results of the positive and negative sentiment analysis show that negative sentiment appears more frequently than positive sentiment with a percentage of 7% positive sentiment and 14% negative sentiment, by using the Rstudio application. This method uses the naïve Bayes classifier. Data division in the dataset with training data 1:499 and test data 1:476. It was found that the results of the Accuracy, Precision, Recall, and F1-Score values were Accuracy 87.50%, Precision 87.50 Recall 93.33%, and F1-Score 82.35%.
Keywords: *Sentiment Analysis, Jakarta Air Quality, Twitter, Naïve Bayes Classifier*

1. Introduction

Analisis sentimen adalah proses untuk menentukan kelompok sentimen yang menganalisis opini, sentimen, penilaian, sikap, dan emosi terdapat karakter seperti entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, dan topik . (Hafiz Irsyad dkk., 2020).

Twitter adalah aplikasi media sosial dengan layanan mikroblog yang memungkinkan pengguna mengirim pesan secara real time. Pesan di Twitter disebut tweet. Pengguna Twitter dapat mengirimkan 140 karakter untuk diposting di Twitter, 140 karakter tersebut dapat berupa pujian atau komentar terhadap sesuatu yang sedang dibicarakan dan juga karena terbatasnya jumlah karakter, khusus yang dapat ditulis 140 karakter, sehingga tweet sering kali mengandung singkatan dan bahasa gaul. (Delima Ayu Wulandari., dkk., 2021)

Indeks Kualitas Udara Jakarta (AQI US) 12 Juli 2023 pada pukul 13.00 WIB adalah 166. Menurut indeks referensi IQAir, nilai indeks antara 151 dan 200 menunjukkan kualitas udara yang tidak sehat. Angka ini bahkan menunjukkan bahwa Jakarta saat ini merupakan kota terpolusi kedua di Asia Tenggara.(CNN Indonesia., 2023). Dilansir dari salah satu akun pemantau kualitas udara IQAir indeks kualitas udara di Jakarta pada tanggal 30 september 2023 sekitar pukul 06.24 mencapai di angka 163. Artinya Jakarta masuk dalam kategori tidak sehat dengan PM 2,5. Badan Pengawasan Kualitas Udara (BMKG) mencatat kota Jakarta sebagai kota dengan kualitas udara terburuk (tvOnenews.2023)

Received October 29, 2023; Received November 2, 2023; Accepted December 14, 2023

Penelitian ini menggunakan metode naïve bayes classifer (NBC) yang dimana Naïve Bayes Classifier memakai konsep yang diklaim adalah sebagai probabilitas yang dipergunakan dalam proses penjabaran untuk analisis sentimen. Penelitian ini dilakukan dengan menerapkan metode Naive Bayes untuk klasifikasi data kualitas udara jakarta. Metode klasifikasi Naïve Bayes digunakan dalam analisis sentimen dan berpotensi baik dalam klasifikasi data dalam hal presisi dan komputasi. (Samsir dkk., 2021)

Pada penelitian ini menggunakan teknik Naive Bayes Classiffier (NBC) dan perhitungan confusion matrix data dari komentar Twitter untuk mengoptimalkan analisis sentimen positif dan negatif. Dengan menggunakan aplikasi Rstudio kode.

2. Research Method

2.1 Analisis Masalah

Dalam penelitian ini, sebelum melakukan analisis sentiment yaitu pengambilan data tweet yang dikumpulkan dari aplikasi media sosial Twitter menggunakan program Python dengan menggunakan kata kunci “kualitas udara”.Penelitian ini menggunakan metode klasifikasi Naive Bayes.

2.2 Alur Penelitian

Dalam alur penelitian terdapat sebuah gambaran umum tahapan alur penelitian tersebut, berikut langkah langkahnya yaitu studi literatur, pengumpulan data, pelabelan manual, pengumpulan data, preprocessing data, pengklasifikasian, lalu hasil dan analisis. Penelitian ini menggunakan aplikasi Rstudio code untuk pembagian data, menghitung klasifikasi naïve bayes dan confusion matrix.

2.3 Pengambilan Data Tweet

Pengambilan data tweet melalui program python di google colabs dengan kata kunci “kualitas udara” lalu filter tanggal bulan dan tahun setelah itu run all . Tunggu beberapa menit lalu masukan token dari halaman tweet lalu enter. Setelah mendapatkan dataset dari tweet disimpan dengan format csv.

2.4 Pelabelan Manual

Selanjutnya setelah mendapatkan data dari tweet lalu dilakukan pelabelan manual untuk diketahui sentiment dari masing masing tweet apakah bernilai positif atau negatif. Untuk pelabelan manual dalam penelitian ini menggunakan Microsoft excel. Contoh cara menentukan kata kata positif dapat diungkapkan seperti kata “lebih baik”, “bagus”,” luar biasa”, dan contoh cara menentukan kata kata negatif dapat diungkapakan seperti kata “buruk”, “memalukan”,”membahayakan”.

Tabel 2.1 Hasil Sentimen Tweet Data.

Text Tweet	Kategori Sentimen
@NarasiNewsroom Riset setiap jam pengukuran polusi udara penting agar kita tau pergerakan naik dan turunnya polusi udara dan juga penyebab nya.. semoga upaya saya bisa sedikit membantu untuk menjadi masukan pengendalian polusi udara di jakarta.. saya sungguh2	Positif

#TempoFoto Hari ini langit di tepi ibukota kembali terlihat sedikit biru, setelah hampir 2 minggu terakhir seperti berkabut. Mujarab. Jokowi tunjuk luhut untuk Permasalahan Polusi Udara Jakarta https://t.co/ZjoQwpFz3c	Positif
#TempoFoto Kualitas udara jakarta terburuk di dunia, Begini pemandangan langitnya https://t.co/MQn89pXhYa	Negatif
#EditorialMediaIndonesia hari Jumat (23/6) LIVE pukul 06.30 WIB di Metro TV akan membahas tentang kualitas udara buruk yang jadi kado sekaligus PR dari ulang tahun kota Jakarta. #mediaindonesia #bedaheditorialmi @mediaindonesia@samosirleonard https://t.co/iRctGuiPTc	Negatif

2.5 Naïve Bayes Classiffer

Penelitian ini menggunakan metode Naive Bayes Classifier (NBC) yang digunakan dalam analisis sentimen untuk mengklasifikasi data tweet. Untuk proses klasifikasi, dataset yang digunakan berasal dari tahap preprocessing, yang di mana data sudah dibersihkan. Berikut rumus Naïve Bayes Classifier.

Persamaan dari teorema bayes:

$$P(A|B)= \frac{P\,(B|A) \times P\,(A)}{P\,(B)}$$

Keterangan:

- P(A|B) : peluang hipotesis A jika diberikan data B
- P(B|A) : peluang data B benar jika hipotesis A benar
- P(A) : peluang hipotesis A benar (terlepas dari datanya)
- P(B) : peluang data (terlepas dari hipotesis).

2.6 Confusion Matrix

Confusion Matrix adalah perhitungan keakuratan konsep data mining. Akurasi adalah perbandingan prediksi kasus positif dengan data sebenarnya positif. Di bawah ini adalah tabel model matriks konfusi.

Tabel 2.2 Model Confusion Matrix

Predicted Class	Actual Class	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

- TP = True Positif
- FN = False Negatif
- FP = False Positif
- TN = True Negatif

Berikut rumus persamaan Akurasi, Presisi, Recall dan F1-score.

1. $Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$
2. $Presisi = \frac{TP}{TP+FP}$
3. $Recall = \frac{TP}{TP+FN}$
4. $f1-score = \frac{2 \times Presisi \times Recall}{Presisi \times Recall}$

2.7 Uji Model

Pengujian ini menggunakan confusion matrix yang digunakan untuk penyempurnaan perhitungan dalam nilai akurasi berdasarkan metode Naive Bayes Classifier. Metode klasifikasi Naive Bayes digunakan dalam analisis sentimen dan berpotensi cukup baik dalam klasifikasi data dalam hal presisi dan komputasi.

2.8 Evaluasi Model

Evaluasi Model yang digunakan untuk melihat kinerja model metode berdasarkan tampilan hasil,.Confusion Matrix adalah cara untuk menghitung keakuratan konsep pemanggilan data. Akurasi adalah perbandingan prediksi kasus positif dengan data sebenarnya positif.

3. Results and Analysis

3.1 Pengumpulan Data

Langkah awal dalam melakukan studi analisis sentimen adalah mengumpulkan data dari platform Twitter. Data dikumpulkan menggunakan program Google Colabs Python.Untuk pengumpulan data dalam sebuah eksperimen, terdapat batasan hanya 50 dataset yang dapat diambil. Oleh karena itu, diperlukan beberapa percobaan untuk mendapatkan kumpulan data yang lebih banyak. Dalam penelitian ini, peneliti memperoleh sekitar 500 dataset dalam 10 percobaan. Pada bagian bawah gambar Anda akan melihat tampilan program dimana dapat memasukkan kata kunci dan melakukan proses pencarian data.

Gambar 3.1 Proses Pengambilan Data.

3.2 Preprocessing Data

Preprocessing adalah fase dimana kata-kata yang tidak penting dan tidak berarti dihilangkan. Gambar berikut menunjukkan kode program di Rstudio dan tabel hasil tahap preprocessing data.

```
df$class <- as.factor(df$KATEGORI)
corpus <- VCorpus(VectorSource(df$TEXT))
inspect(corpus[1:5])

stopwordID <- "D:/danu/stoplist.txt"
cstopwordID<-readLines(stopwordID,warn=FALSE)

corpus.clean <- tm_map(corpus, content_transformer(tolower))
corpus.clean <- tm_map(corpus.clean, removePunctuation)
corpus.clean <- tm_map(corpus.clean, removeNumbers)
corpus.clean <- tm_map(corpus.clean, removeWords,cStopwordID)
corpus.clean <- tm_map(corpus.clean, stripwhitespace)

dtm <- DocumentTermMatrix(corpus.clean)
inspect(dtm[1:5,1:4])
```

Gambar 3.2 Kode Preprocessing Data

3.3 Hasil Klasifikasi Naïve Bayes

Tahap klasifikasi menggunakan metode Naive Bayes Classifier(NBC). Untuk mendapatkan hasil probabilitas apakah termasuk dalam kelas positif atau negatif. Pengklasifikasi Naïve Bayes digunakan untuk melakukan prediksi terhadap kasus berdasarkan hasil klasifikasi yang diperoleh. Perhitungan klasifikasi Naive Bayes dilakukan untuk menentukan nilai probabilitas prior positif dan negatif. Sebelum melakukan klasifikasi naïve bayes, yaitu pisahkan data pelatihan dan data pengujian.Pembagian datanya adalah 1:.499 untuk data latih dan 1:.476 untuk data uji.Di bawah ini adalah perhitungan klasifikasi Naive Bayes, dan diagram kode yang memisahkan data pelatihan dan pengujian ditunjukkan pada Gambar 3.3.

$$Probabilitas\ Prior\ Positif = \frac{Jumlah\ Kelas\ Positif}{Jumlah\ Keseluruhan\ Data} = \frac{7}{499} = 0{,}0140$$

$$Probabilitas\ Prior\ Negatif = \frac{Jumlah\ Kelas\ Negatif}{Jumlah\ Keseluruhan\ Data} = \frac{14}{499} = 0{,}2857$$

```
show(dtm)
df.train <- df[1:475,]
df.test <- df[476:499,]

dtm.train <- dtm[1:475,]
dtm.test <- dtm[476:499,]

dim(dtm.train)
corpus.clean.train <- corpus.clean[1:475]
corpus.clean.test <- corpus.clean[476:499]

fivefreq <- findFreqTerms(dtm.train, 5)
length((fivefreq))
```

Gambar 3.3 Kode Program Pembagian Data.

3.4 Hasil Uji Model

Setelah melakukan proses klasifikasi, langkah berikutnya adalah menguji model untuk mengetahui seberapa baik teknik tersebut bekerja. Hasil klasifikasi akan divisualisasikan dalam tabel confusion matrix. Tabel ini digunakan untuk mengukur model klasifikasi untuk menentukan perhitungan prediksi benar dan salah. Dalam pengujian, model klasifikasi data yang digunakan berasal dari 476

data tweet dari 499 data tweet yang diuji. Hasil confusion matrix dan kode confusion matrix pada Gambar 3.4 dan 3.5.

14	2
1	7

Gambar 3.4 Hasil Confusion Matrix

```
# Prepare the confusion matrix
conf.mat <- confusionMatrix(pred, df.testSclass)
```

Gambar 3.5 kode Confusion Matrix

3.5 Hasil Evaluasi Model

Setelah dilakukan pengujian model maka selanjutnya dilakukan tahap evaluasi model yang bertujuan untuk menghasilkan matrix dengan ukuran 2x2. Confusion matrix dapat memberikan informasi untuk perbandingan hasil klasifikasi. Berikut pada tabel merupakan hasil evaluasi model dengan confusion matrix.

Tabel 3.1 Hasil Confusion Matrix

Predicted Class	Actual Class	
	Positif	Negatif
Positif	TP(7)	FP(1)
Negatif	FN(2)	TN(14)

Berdasarkan hasil confusion matrix dengan perhitungan rumus maka didapatkan nilai Akurasi, Presisi, Recall, dan F1-Score dalam kelas sentimen positif dan negatif.

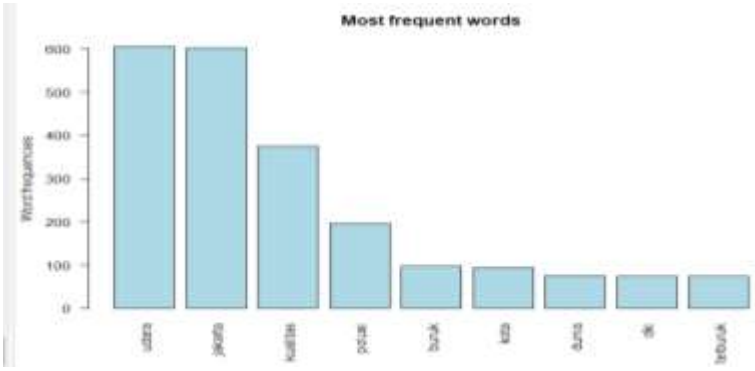
- Tabel Nilai Rata Rata Evalusasi Model

Akurasi	Presisi	Recall	F1-Skor
87.50%	87.50%	93.33%	82.35%

Dari data perbandingan dari pengujian eksperimen menggunakan Naïve Bayes berikut Diketahui bahwa hasil nilai Akurasi, Presisi, Recall, dan F1-Skor yaitu Akurasi 87.50%, Presisi 87,50 Recall 93.33%, dan F1-Score 82,35%.

3.6 Visualisasi

Setelah fase selesai, langkah selanjutnya adalah memvisualisasikan kata yang paling sering digunakan dalam bentuk histogram dan world cloud. Tampilan histogram dan wordcloud adalah format yang menampilkan kata yang sering muncul. Dibawah ini merupakan gambar visualisasi Most Frequent Words dan Word Cloud.



Gambar 3.6 Most Frequent Words

Gambar 3. 7 Word Cloud

4. Conclusion

Penelitian ini menggunakan klasifikasi naive bayes classifier untuk menganalisis sentiment masyarakat di media sosial twitter mengenai kualitas udara jakarta. Penelitian ini menggunakan data berupa data tweet dari platform media sosial twitter. Pengambilan data tweet menggunakan program python di google colabs. pelabelan manual dengan menggunakan Microsoft excel. kata kuncinya adalah kualitas udara setelah itu dilanjutkan dengan aplikasi Rstudio..Data yang diambil berjumlah 500 data tweet yang terbagi menjadi sentimen positif dan negatif kemudian dilanjutkan pembagian data seperti data uji 1:499 dan data latih 1: 476. Dengan hasil sentimen positif 7% dan sentimen negatif 14% . Hasil dari rumus klasifikasi naive bayes classifier dan disempurnakan dengan model confusion matrix didapatkan hasil Akurasi sebesar 87,50%, presisi 87.50%, recall 93.33%, dan F-measure 82.35%.

References

[1] Andreyestha, A., & Azizah, Q. N. (2022). Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE. Infotek: Jurnal Informatika Dan Teknologi, 5(1), 108-116.

[2] Astuti, A. P., Alam, S., & Jaelani, I. (2022). Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo. Jurnal Bangkit Indonesia, 11(2), 1-6.

[3] Azhar, M., Hafidz, N., Rudianto, B., & Gata, W. (2020). Marketplace Sentiment Analysis Using Naive Bayes And Support Vector Machine. Piksel: Penelitian Ilmu Komputer Sistem Embedded And Logic, 8(2), 91-100.

[4] Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. Jurnal Tekno Kompak, 15(1), 131-145.

[5] Fikri, Mujaddid Izzul, Trifebi Shina Sabrila, and Yufis Azhar. "Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter." SMATIKA Jurnal: STIKI Informatika Jurnal 10.02 (2020): 71-76.
[6] Irsyad, H., & Pribadi, M. R. (2020). Klasifikasi Opini Terhadap Pertanian Sawit (Palm Oil) Indonesia Menggunakan Naïve Bayes. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 6(2), 230-239.
[7] Hidayatulloh, M. Y., Sunanto, A., Armansyah, A., Gevin, M. F. A., & Saputra, D. D. (2023). Optimasi Sentimen Analisis Informatif dan Tidak Informatif dari Tweet di BMKG Menggunakan Algoritma Naive Bayes dan Metode Teknik Pengambilan Sampel Minoritas Sintetis. J-SAKTI (Jurnal Sains Komputer dan Informatika), 7(1), 1-12.
[8] Hant, M. I. P., & Hendry, H. (2022). DATA MINING TECHNIQUE USING NAÏVE BAYES ALGORITHM TO PREDICT SHOPEE CONSUMER SATISFACTION AMONG MILLENNIAL GENERATION. Jurnal Teknik Informatika (Jutif), 3(4), 829-838.
[9] Pribadi, M. R., Purnomo, H. D., Hartomo, K. D., Sembiring, I., & Iriani, A. (2022, October). Improving the accuracy of text classification using the over sampling technique in the case of sinovac vaccine. In 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 106-110). IEEE.
[10] Manalu, D. R., Tobing, M. C. L., & Yohanna, M. (2022). ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE. METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi, 6(2), 149-156.
[11] Mufidah, F. S., Winarno, S., Alzami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat terhadap Layanan Shopeefood Melalui Media Sosial Twitter dengan Algoritma Naïve Bayes Classifier. vol, 7, 14-25.
[12] Pandunata, P., Ananta, C. K., & Nurdiansyah, Y. (2022). Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classififer. INFORMAL: Informatics Journal, 7(2), 146-156.
[13] Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 11(2), 195-200.
[14] Prasetiawan, F., Widiyanesti, S., & Widarmanti, T. (2022). Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Barang Sicepat Di Media Sosial Twitter. eProceedings of Management, 9(2).
[15] Pratiwi, Y., & Yaqin, A. (2022). Klasifikasi Tweet Tidak Senonoh Twitter dengan Naïve Bayes Classifier. E-JURNAL JUSITI: Jurnal Sistem Informasi dan Teknologi Informasi, 11(1), 70-80.
[16] Permana, R. A., & Sahara, S. (2021). Review Analisis Produk Marketplace Online pada Algoritma Support Vector Machine. Jurnal Ilmiah Informatika, 6(1), 50-58.
[17] Verawati, I., & Audit, B. S. (2022). Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By. u. Jurnal Media Informatika Budidarma, 6(3), 1411-1417.
[18] Wati, Risha Ambar, Hafiz Irsyad, and M. Ezar Al Rivan. "Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine." J. Algoritm 1.1 (2020): 21-32.