

Sentimen Analisis Opini Masyarakat Terhadap UMKM Pada Media Sosial Twitter Dengan Metode Naïve Bayes Classifier

Angga Aditya Permana^{1)*}, Wahyu Aldhi Noviyanto²⁾, Dinar Ajeng Kristiyanti³⁾

^{1)*}Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Banten, Indonesia

²⁾Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang, Banten, Indonesia

³⁾Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Banten, Indonesia

^{1)*} angga.permana@umn.ac.id, ²⁾ wahyu.alzhi@ft-umt.ac.id, ³⁾ dinar.kristiyanti@umn.ac.id

Abstrak :

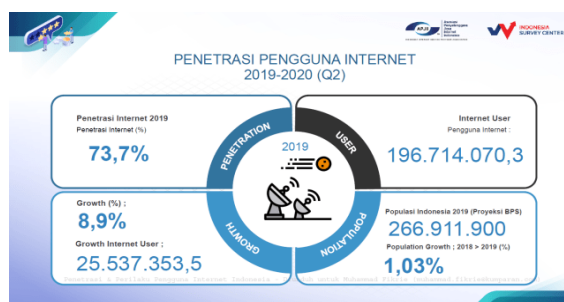
Pertumbuhan *Twitter* selalu meningkat seiring berjalannya waktu, namun masih banyak kekurangan yang ditemukan, seperti sumber daya manusia yang masih belum berkembang. Metode yang digunakan dalam penelitian ini adalah metode *Naïve Bayes Classifier (NBC)* untuk dilakukan analisis sentimen. Untuk proses penentuan kalimat opini positif, netral atau negatif diambil dari *Twitter*. Penentuan ini sebagai proses pengklasifikasian. Dengan menggunakan Aplikasi Python 3.77, langkah awal adalah pengambilan data atau disebut *crawling* lalu di proses menggunakan Bahasa alami yang biasa disebut *Natural Language Processing (NLP)* menggunakan library *textblob* dilanjutkan tahap *preprocessing* lalu menghitung Probabilitas kata. Sentimen ini untuk mengklasifikasi data *tweet* menggunakan library *TextBlob* sebanyak 100 data menghasilkan 58% positif, 37% netral, dan 4% negatif.

Kata kunci :

Sentimen Analisis; UMKM; *Twitter*; Naïve Bayes; Media Sosial

PENDAHULUAN

Perkembangan teknologi internet saat ini sangat cepat sekali dan sangat mudah memperoleh informasi dan membagikannya. Di era yang maju sekarang ini pertumbuhan internet dan media social selalu berkembang sangat pesat sehingga menyediakan berbagai informasi opini dari orang lain. Menurut survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) hingga kuartal II tahun 2020, jumlah pengguna internet di Indonesia ada 196,7 juta orang atau 73,7 persen dari total populasi Indonesia 266,91 juta (Herdhianto et al. 2020). Menunjukkan bahwa jumlah pengguna internet di Indonesia mengalami kenaikan beberapa tahun terakhir (Zuhdi, Utami, and Raharjo 2019).



Gambar 1. Pengguna Internet

Media sosial *Twitter* adalah salah satu sebagai media penghubung yang diminati oleh masyarakat seluruh dunia (Andrianto and Adinugroho 2018). Hal ini bisa dilihat pada peningkatan pada pengguna media sosial *Twitter* yang tercatat pada seluruh dunia termasuk Indonesia. Sehingga hal ini dapat

*Angga Aditya Permana



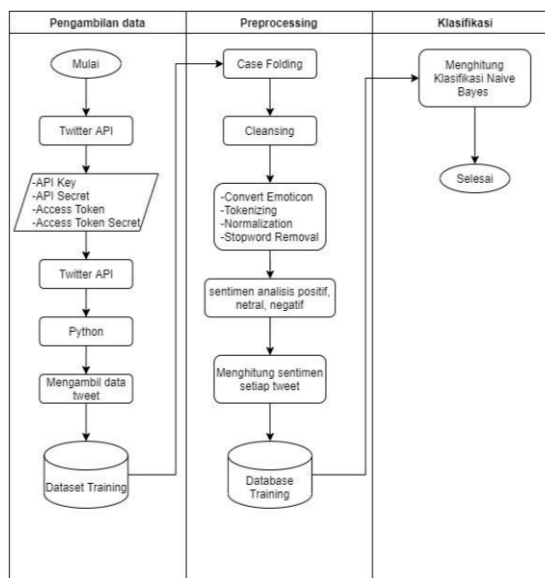
dimanfaatkan bagi para pengguna *Twitter* untuk menyampaikan informasi berupa komentar kritik dan saran terhadap UMKM dapat dilakukan dengan mudah. Belakangan ini persaingan penjualan produk lokal dan impor di *e-commerce* menjadi *trending topic*. Tagar #SellerAsingBunuhUMKM menjadi viral hingga mencapai sepuluh ribu cuitan. Masyarakat meminta para penjual di *e-commerce* seperti Shopee, Tokopedia ataupun Bukalapak untuk mengutamakan Usaha Mikro, Kecil, dan Menengah (UMKM) dan menjual produk lokal dibanding produk impor. Semakin banyak masukan pendapat dari masyarakat dalam bentuk opini yang disampaikan tentunya dapat dijadikan masukan terhadap suatu penilaian pelayanan pada UMKM di Indonesia agar menjadi lebih baik.

Analisis sentimen ini membantu untuk mendapatkan gambaran umum opini masarakat dengan cara mengelompokkan suatu opini menjadi kategori positif, netral, atau negatif. Klasifikasi ini sangat membantu sekali bagi semua pengguna untuk melihat kritik dan saran (Dwiraswati and Siregar 2019).

Tujuan penelitian ini adalah untuk mencari komentar positif, netral dan negative dengan review data *Twitter* UMKM di Indonesia menggunakan *library* TextBlob lalu melakukan perhitungan data positif, netral dan negatif menggunakan algoritma *naïve bayes* secara manual. Penelitian yang dilakukan fokus terhadap kritik dan saran dalam pelayanan UMKM yang diambil pada keyword UMKM di cuitan *Twitter* dengan menggunakan data realtime pada tanggal 2 bulan juli tahun 2021 sebanyak 100 data. Proses pengambilan data dan klasifikasi sentiment analisis ini menggunakan Bahasa Python 3.77.

METODE PENELITIAN

Pada penelitian analisis sentiment yang dilakukan diimplementasikan dengan Bahasa python. Berikut merupakan gambaran tentang kerangka penelitian yang dibuat dari pengambilan data sampai terjadinya klasifikasi *naïve bayes*. (PERMANA and NOVIYANTO 2023)(Putra and Permana 2023)



Gambar 2. Tahapan Penelitian

penulis mengambil data dari API Twitter yang berkaitan dengan topik UMKM khususnya tweet yang diambil menggunakan bahasa indonesia yang terdapat dalam media sosial Twiter. Pengambilan data ini dilakukan pada tanggal 2 Juli 2021. Tweet yang digunakan adalah tweet-tweet yang mengandung emosi cinta, senang, marah, sedih, takut. Pemilihan data secara otomatis dengan menggunakan bahasa Python yaitu memilih kalimat-kalimat tweet yang berbahasa indonesia yang tidak mengandung gambar. Penulis membuat 29 sebuah tabel yang bernama corpus_tweets untuk menyimpan data tweet dan yang nantinya akan dipakai untuk proses stemming dan stopword removal.

*Angga Aditya Permana



HASIL PENELITIAN DAN DISKUSI

A. Pengumpulan data tweet

Proses pengambilan data pada twitter atau disebut *crawling* menggunakan *software* Python 3.77. Topik analisis ini berkaitan mengenai *tweet* tentang UMKM dan pengambilan data pada media sosial twitter sejumlah 100 *tweet*. Dengan cara pemanggilan kode token dan di implementasikannya pada python, hasil data *crawling* yang di dapat berbentuk json file dapat dilihat pada gambar 3.

```

1:
pengguna: "ujakmasuraleho"
isi_tweet: "di @kemerekuari: Indonesia bisa jadi pusat produsen hajeil dunia. UMKM punya peran penting untuk mewajarkannya dengan @kneks_id sebagai kula."
sentimen: "positif"

2:
pengguna: "kadesku"
isi_tweet: "RT @KIVAcold: UMKM Jangan Gitu Gitu Aja, mande! Tantang Perusahaan Harus Kreatif! https://t.co/cdaqTTF9v"
sentimen: "positif"

3:
pengguna: "ujakmasuraleho"
isi_tweet: "di @kemerekuari: Komposisi paling bawah pada sektor UMKM banyak diisi oleh pelaku usaha Ultra Mikro (UM). Di masa pandemi ini, usaha UMI me..."
sentimen: "positif"

4:
pengguna: "CandraCeliana"
isi_tweet: "RT @My_Ansara1701: Yuk diadon produk teman-teman UMKM @Boviraakabamuila guys 🙌🏻"
sentimen: "netral"

5:
pengguna: "CandraCeliana"
isi_tweet: "RT @tly31670477: Tahadotabek free ongkir yes! Trinalastih yg sudah dukung UMKM yg sudah order, yg sudah bertanya, yg masih pna salim, seru..."
sentimen: "positif"

6:
pengguna: "bdiw 8144K"
isi_tweet: "RT @Bdrowaborn: @lungky2501 @novi_sbg @tialovra13 @kyra_Triah2 @Pni_Ainegrus @BellaRator @drosi912 @Baka_ri @Blaucans Uraan abwah..."

```

Gambar 3. Hasil Crawling

B. Penggunaan Library Python TextBlob

TextBlob merupakan *library* pada *Python* yang digunakan untuk proses data tekstual. *TextBlob* sendiri menyediakan API yang bisa digunakan untuk proses Bahasa alami yang biasa disebut *Natural Language Processing* (NLP) yang nantinya untuk analisis sentiment, klasifikasi, penerjemahan. Hasil pada objek *TextBlob* dapat digunakan untuk proses pembelajaran Bahasa alami dan *library TextBlob* mempunyai kekurangan yaitu hanya mengenali Bahasa Inggris. Hasil klasifikasi menggunakan *library TextBlob* bisa dilihat pada gambar 4.

```

print("Hasil Sentimen")
print("positif: ", len(tweet_positif), "{:}%".format(100*len(tweet_positif)/len(hasilAnalisis)))
print("netral: ", len(tweet_netral), "{:}%".format(100*len(tweet_netral)/len(hasilAnalisis)))
print("negatif: ", len(tweet_negatif), "{:}%".format(100*len(tweet_negatif)/len(hasilAnalisis)))

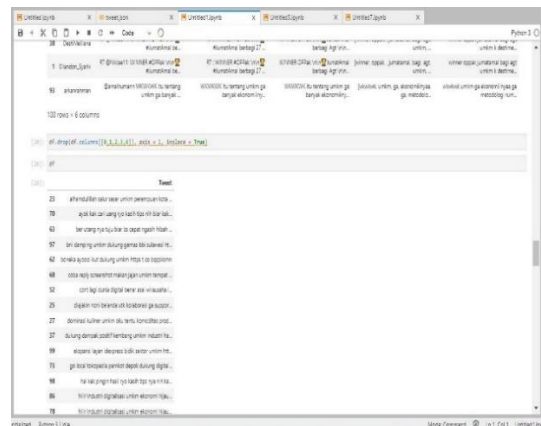
Hasil Sentimen
positif: 58 (58.5058505850585059%)
netral: 37 (37.37373737373738%)
negatif: 4 (4.040404040404041%)

```

Gambar 4. Hasil Sentimen Library TextBlob

C. Preprocessing

Data yang diambil dari hasil *crawling* masih belum terstruktur karna adanya karakter-karakter, tanda baca, kata singkatan, dan penggunaan huruf kapital. Pada tahap *preprocessing* ini guna mendapatkan data bersih agar memudahkan ke tahap berikutnya dan bisa dilihat pada gambar 5.



Gambar 5. Preprocessing

*Angga Aditya Permana



D. Menghitung Probabilitas Kata

Pada perhitungan probabilitas kata ini bertujuan agar mendapatkan term dengan nilai yang lebih penting dan dianggap relevan untuk digunakan sebagai kata kunci. Proses pembobotan dengan menggunakan algoritma naïve bayes dalam proses perhitungan persamaan dapat dilihat pada Tabel 1.

Tabel 1. Kemunculan Term Frekuensi

No	Kosa Kata	Tf(Positif)	Tf(Neutral)	Tf(Negatif)
1.	umkm	48	28	2
2.	dukung	28	1	0
3.	pandemi	19	4	0
4.	bantu	14	3	0
5.	usaha	6	8	1
6.	order	13	2	0
7.	jabodetabek	13	0	0
8.	free	13	0	0
9.	ongkir	13	0	0
10.	trimakasih	13	0	0
Jumlah Term		180	46	3

Count positif = 180, Count Netral = 46, Count Negatif = 3 dengan total 10 kata

1. Hitung probabilitas prior setiap kategori yang menjadi kategori ada 3, yaitu kelas positif, netral, dan negatif.

$$P\left(\frac{\text{Positif}}{\text{Netral}}}{\text{Negatif}}\right) = \frac{x(\text{positif/netral/negatif})}{|C|} \dots\dots\dots$$

$$P(\text{Positif}) = \frac{fx(\text{positif})}{|C|} = \frac{1}{3} = 0,333333333333$$

$$P(\text{Netral}) = \frac{fx(\text{netral})}{|C|} = \frac{1}{3} = 0,333333333333$$

$$P(\text{Negatif}) = \frac{fx(\text{negatif})}{|C|} = \frac{1}{3} = 0,333333333333$$

2. Menghitung probabilitas likelihood setiap term dari semua dokumen. Terdapat 10 jumlah kata, yang dibagi antara lain 180 term kelas positif, 46 dari kelas netral, dan 3 term dari kelas negatif. Banyaknya jumlah dari term bergantung pada hasil praproses data.

$$P(w|\text{positif/netral/negatif}) = \frac{(n.k(\text{positif/netral/negatif}) + 1)}{(n.\text{positif/netral/negatif}) + |T|}$$

1. Probabilitas kata “umkm”

$$P(\text{umkm}|\text{positif}) = \frac{48 + 1}{180 + 10} = 0.2578947368$$

$$P(\text{umkm}|\text{netral}) = \frac{28 + 1}{46 + 10} = 0.5178571429$$

$$P(\text{umkm}|\text{negatif}) = \frac{2 + 1}{3 + 10} = 0.2307692308$$

2. Probabilitas kata “dukung”

$$P(\text{dukung}|\text{positif}) = \frac{28 + 1}{180 + 10} = 0.1526315789$$

$$P(\text{dukung}|\text{netral}) = \frac{1 + 1}{46 + 10} = 0.03571428571$$

$$P(\text{dukung}|\text{negatif}) = \frac{0 + 1}{3 + 10} = 0.07692307692$$

*Angga Aditya Permana



3. Probabilitas kata “pandemi”

$$P(\text{pandemi}|\text{positif}) = \frac{19+1}{180+10} = 0.1052631579$$

$$P(\text{pandemi}|\text{netral}) = \frac{4+1}{46+10} = 0.08928571429$$

$$P(\text{pandemi}|\text{negatif}) = \frac{0+1}{3+10} = 0.07692307692$$

4. Probabilitas kata “bantu”

$$P(\text{bantu}|\text{positif}) = \frac{14+1}{180+10} = 0.07894736842$$

$$P(\text{bantu}|\text{netral}) = \frac{3+1}{46+10} = 0.07142857143$$

$$P(\text{bantu}|\text{negatif}) = \frac{0+1}{3+10} = 0.07692307692$$

5. Probabilitas kata “usaha”

$$P(\text{usaha}|\text{positif}) = \frac{6+1}{180+10} = 0.03684210526$$

$$P(\text{usaha}|\text{netral}) = \frac{8+1}{46+10} = 0.1607142857$$

$$P(\text{usaha}|\text{negatif}) = \frac{1+1}{3+10} = 0.1538461538$$

6. Probabilitas kata “order”

$$P(\text{order}|\text{positif}) = \frac{13+1}{180+10} = 0.07368421053$$

$$P(\text{order}|\text{netral}) = \frac{2+1}{46+10} = 0.05357142857$$

$$P(\text{order}|\text{negatif}) = \frac{0+1}{3+10} = 0.07692307692$$

7. Probabilitas kata “jabodetabek”

$$P(\text{jabodetabek}|\text{positif}) = \frac{13+1}{180+10} = 0.07368421053$$

$$P(\text{jabodetabek}|\text{netral}) = \frac{0+1}{46+10} = 0.01785714286$$

$$P(\text{jabodetabek}|\text{negatif}) = \frac{0+1}{3+10} = 0.076912307692$$

8. Probabilitas kata “free”

$$P(\text{free}|\text{positif}) = \frac{13+1}{180+10} = 0.07368421053$$

$$P(\text{free}|\text{netral}) = \frac{0+1}{46+10} = 0.01785714286$$

$$P(\text{free}|\text{negatif}) = \frac{0+1}{3+10} = 0.076912307692$$

9. Probabilitas kata “ongkir”

$$P(\text{ongkir}|\text{positif}) = \frac{13+1}{180+10} = 0.07368421053$$

$$P(\text{ongkir}|\text{netral}) = \frac{0+1}{46+10} = 0.01785714286$$

$$P(\text{ongkir}|\text{negatif}) = \frac{0+1}{3+10} = 0.076912307692$$

10. Probabilitas kata “trimakasih”

$$P(\text{trimakasih}|\text{positif}) = \frac{13+1}{180+10} = 0.07368421053$$

$$P(\text{trimakasih}|\text{netral}) = \frac{0+1}{46+10} = 0.01785714286$$

$$P(\text{trimakasih}|\text{negatif}) = \frac{0+1}{3+10} = 0.076912307692$$

*Angga Aditya Permana



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E. Klasifikasi *NBC* secara manual

Berikut adalah hasil rekapan data yang ingin di uji menggunakan perhitungan *naive bayes* dan mencocokkan hasil sentiment dari library *TextBlob*.

Tabel 2. Data Uji

No	Komentator	Komentar	Kelas
1.	@Indopolitika	Peduli umkm, gubernur hd diganjar penghargaan kepala daerah inovatif 2020	Positif
2.	@bisnisjatengdiy	keaktivitas anak muda di balik geliat umkm salatiga	Negatif
3.	@imwinulost	20.833 UMKM di Riau Dapat Bantuan Modal Usaha dari Pemprov Riau	Netral
Data Uji			
4.	@PemkotaBlitar	dukung eksistensi umkm batik, kim asabri gedog bersama wira desa edukasikan digital marketplace	?
5.	@maduuntoday	bahas ekonomi pandemi wali kota kan strategi daya umkm wali kota sapa madiun program	?

Proses dari pengklasifikasian data uji dengan mengalikan semua nilai peluang, Nilai yang lebih tinggi merupakan kelas baru dari data tersebut.

$$P\left(\frac{\text{Positif}}{\text{netral}}{\text{negatif}} \mid d\right) = P\left(\frac{\text{Positif}}{\text{netral}}{\text{negatif}}\right) \times \pi p (w|\text{positif/netral/negatif})$$

1. Pada data uji “dukung eksistensi umkm batik, kim asabri gedog bersama wira desa edukasikan digital marketplace” yang termasuk ke dalam kategori data training adalah kata “umkm” dan “dukung”

a. *P(uji/positif)*

$$= P(\text{positif}) \times P(\text{umkm}|\text{positif}) \times P(\text{dukung}|\text{positif})$$

$$= 0,333333333333 \times 0.2578947368 \times 0.1526315789$$

$$= 0.0131209603$$

b. *P(uji/netral)*

$$= P(\text{netral}) \times P(\text{umkm}|\text{netral}) \times P(\text{dukung}|\text{netral})$$

$$= 0,333333333333 \times 0.5178571429 \times 0.03571428571$$

$$= 0.006164965986$$

c. *P(uji/negatif)*

$$= P(\text{negatif}) \times P(\text{umkm}|\text{negatif}) \times P(\text{dukung}|\text{negatif})$$

$$= 0,333333333333 \times 0.2307692308 \times 0.07692307692$$

$$= 0.005917159764$$

Kesimpulan : Nilai probabilitas tertinggi yang didapatkan sebesar 0.0131209603 pada *P(uji/positif)* sehingga komentar tersebut dapat digolongkan ke dalam kelas “**Positif**” dan sesuai dengan hasil klasifikasi pada *library TextBlob*.

```

pengguna: "PemkotaBlitar"
isi_tweet: "Dukung Eksistensi UMKM Batik, KIM Asabri Gedog Bersama Wira Desa Edukasikan Digital Marketplace"
sentimen: "positif"
    
```

Gambar 6. Hasil library Textblob

2. Pada data uji “bahas ekonomi pandemi wali kota kan strategi daya umkm wali kota sapa madiun program” yang termasuk ke dalam data training adalah kata “pandemi” dan “umkm”

*Angga Aditya Permana



dan seluruh data tweet yang berjumlah 100 data yang diklasifikasikan menggunakan *library TextBlob* mendapatkan hasil dimana 58% tweet positif, 37% tweet netral, dan 4% tweet negatif.

UCAPAN TERIMAKASIH

Penelitian ini dapat dilaksanakan dengan baik berkat bantuan dari berbagai pihak, untuk itu peneliti mengucapkan terima kasih kepada Universitas Multimedia Nusantara dan Universitas Muhammadiyah Tangerang atas dukungan dan bantuan yang telah diberikan selama proses penulisan artikel ini.

REFERENSI

- Andrianto, B., & Adinugroho, S. (2018). Analisis Sentimen Konten Radikal Melalui Dokumen Twitter Menggunakan Metode Backpropagation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(12), 7380–7385.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGGER: Journal of Information Technology*, 1(1), 32–41.
- Dwiraswati, O., & Siregar, K. N. (2019). Analisis Sentimen Pada Twitter Terhadap Penggunaan Antibiotik Di Indonesia Dengan Naive Bayes Classifier. *Media Informasi*, 15(1), 1–9. <https://doi.org/10.37160/bmi.v15i1.232>
- Herdhianto, A., Studi, P., Informatika, T., Sains, F., Teknologi, D. A. N., Islam, U., & Syarif, N. (2020). *Sentiment Analysis Menggunakan Naive Bayes Classifier (Nbc) Pada Tweet Tentang Zakat Sentiment Analysis Menggunakan Naive Bayes Classifier (Nbc) Pada Tweet Tentang*.
- Karomi, M. A. Al, & Rusli, C. Y. (2018). Aplikasi Pendukung Keputusan Persetujuan Kredit berbasis WEB dengan Pemanfaatan Algoritma Data Mining. *IC-Tech*, 13(1), 28–32.
- Kilat, J. (2019). *KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR*. 8(1), 81–92.
- Murnawan, M. (2017). Pemanfaatan Analisis Sentimen Untuk Pemingkatan Popularitas Tujuan Wisata. *Jurnal Penelitian Pos Dan Informatika*, 7(2), 109. <https://doi.org/10.17933/jppi.2017.070203>
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *E-Proceeding of Engineering*, 2(1), 1–7.
- PERMANA, ANGGA ADITYA, and WAHYU ALDHI NOVIYANTO. 2023. "Comparison of the Accuracy of the Lexicon-Based and Naive Bayes Classifier Methods To Public Opinions About Removing Masks on Social Media Twitter." *Journal of Theoretical and Applied Information Technology* 101(3):1174–83.
- Putra, Permana Perdana, and Angga Aditya Permana. 2023. "Sentimen Analisis Opini Masyarakat Terhadap Kasus Pelecehan Seksual Di Kpi Pada Social Media Twitter Dengan Metode Naive Bayes Classifier (Nbc) 1." 12(01):122–33.
- Rahutomo, F., Pratiwi, I. Y. R., & Ramadhani, D. M. (2019). Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia. *Jurnal Penelitian Komunikasi Dan Opini Publik*, 23(1). <https://doi.org/10.33299/jpkop.23.1.1805>
- Rosadi, A., Gustiana, D., Informatika, M., Sti, S. J., No, J. B. R. I., Dalam, R., Baru, K., Selatan, J., Kata, P., Baku, T., & Bernegasi, M. K. (2021). *Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naive Bayes Classifier Twitter Text Mining*. 20, 39–52.
- Sartika, D., & Sensuse, D. I. (2017). *Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian* (Vol. 1, Issue 2).
- Syakuro, A. (2017). Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro. *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*, 1–89.
- Zuhdi, A. M., Utami, E., & Raharjo, S. (2019). Analisis sentiment twitter terhadap capres Indonesia 2019 dengan metode K-NN. *Jurnal Informa Politeknik Indonusa Surakarta*, 5, 1–7.

*Angga Aditya Permana



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.