

## Penggunaan *Machine Learning* (ML) dan *Natural Language Processing* (NLP) untuk mendeteksi Sentimen Ancaman Siber

Andri Wijaya<sup>1\*</sup>, Steven Adi Putra<sup>2</sup>

<sup>1,2</sup> Jurusan Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas Katolik Misi Charitas

\*Email: andri\_wijaya@ukmc.ac.id

### Abstrak

Keamanan siber telah menjadi isu kritis di era digital, dengan bukti 3 tahun terakhir terdapat 6 kasus cybercrime di Indonesia yang menyerang server salah satunya yang terbaru pencurian data Bank Syariah Indonesia pada bulan Mei yang mengakibatkan server lumpuh selama 5 hari dan dampaknya nasabah tidak dapat mengakses aplikasi mobile banking. Dari berbagai kasus cybercrime yang terjadi di Indonesia perlu kita mengetahui tren sentimen publik saat ini tentang hal tersebut dan salah satu sumber sentimen publik yaitu Twitter. Penggunaan *Machine Learning* (ML) dan *Natural Language Processing* (NLP) telah menjadi fokus utama dalam memahami sentimen publik yang terkandung dalam data Twitter. Penelitian ini mengusulkan sebuah pendekatan yang menggabungkan teknik-teknik ML dan NLP untuk mendeteksi sentimen dalam tweet. Metode ini mencakup tahap pre-processing untuk membersihkan dan mengubah teks tweet menjadi representasi vektor, diikuti oleh penerapan model klasifikasi ML yaitu Naïve Bayes untuk mengidentifikasi sentimen positif, negatif atau netral dari dataset tweet. Penelitian ini memanfaatkan sekumpulan data tweet yang telah dikumpulkan dan diannotasi dengan menggunakan Python untuk melatih dan menguji model. Hasil eksperimen menunjukkan bahwa pendekatan yang diusulkan berhasil menghasilkan klasifikasi sentimen dengan tingkat akurasi sebesar 62% dapat disimpulkan akurasi model masih memuaskan dengan nilai recall positif sebesar 74% artinya sentimen publik dari tweet masih mengandung kata-kata bersifat positif.

**Kata kunci:** cybercrime, naive bayes, natural language processing

### Abstract

Cybersecurity has become a critical issue in the digital era, with evidence in the last 3 years there have been 6 cybercrime cases in Indonesia that attacked servers, one of which was the latest theft of Bank Syariah Indonesia data in May which resulted in the server being paralyzed for 5 days and the impact was that customers could not access the mobile banking application. From the various cybercrime cases that have occurred in Indonesia, we need to know the current trend of public sentiment about it and one of the sources of public sentiment is Twitter. The use of *Machine Learning* (ML) and *Natural Language Processing* (NLP) has become a major focus in understanding public sentiment contained in Twitter data. This research proposes an approach that combines ML and NLP techniques to detect sentiment in tweets. The method includes a pre-processing stage to clean and transform the tweet text into a vector representation, followed by the application of ML classification model namely Naïve Bayes to identify positive, negative or neutral sentiments from the tweet dataset. This research utilizes a set of collected and annotated tweet data using Python to train and test the model. The experimental results show that the proposed approach successfully produces sentiment classification with an accuracy rate of 62%. It can be concluded that the accuracy of the model is still satisfactory with a positive recall value of 74%, meaning that the public sentiment of the tweets still contains words of a positive nature.

**Keyword :** cybercrime, naive bayes, natural language processing

## PENDAHULUAN

Keamanan siber telah menjadi isu kritis di era digital ini. Salah satunya yang terbaru pencurian data Bank Syariah Indonesia pada bulan Mei server lumpuh selama 5 hari yang menyebabkan para nasabah tidak dapat mengakses aplikasi *mobile banking* (AMT, 2023).

Untuk menghadapi tantangan keamanan siber yang semakin kompleks ini, sangat penting bagi para peneliti, ahli keamanan, dan pihak berwenang untuk memahami pola dan tren terkini mengenai ancaman keamanan siber. Sumber informasi yang berpotensi menjadi penunjuk arah adalah media sosial, khususnya platform populer seperti *Twitter*. Jutaan pengguna aktif di *Twitter* secara terbuka berbagi pendapat, pengalaman, dan persepsi mereka tentang keamanan siber. Oleh karena itu, analisis sentimen *Twitter* dapat menjadi alat yang efektif untuk mendeteksi dan memahami tren ancaman keamanan siber.

Untuk mengeksplorasi dan menganalisis sentimen yang terkandung dalam cuitan-cuitan di *Twitter* yang terkait dengan topik keamanan siber, penelitian ini akan menggunakan pendekatan data *crawling* dan analisis berbasis klasifikasi dengan metode *Naive Bayes*. Analisis sentimen menggunakan *Naive Bayes* telah dipelajari secara ekstensif di berbagai domain. Penelitian sebelumnya telah mempelajari analisis sentimen dalam konteks yang berbeda, seperti memeriksa persepsi publik terhadap pemilihan (Juanita, 2020), mengevaluasi respons publik terhadap vaksinasi COVID-19 (Septiana, Susanto and Tukiya, 2021), dan menilai sentimen terhadap layanan pemerintah (Salim and Solichin, 2022). Pemilihan *Naive Bayes* sebagai metode klasifikasi telah dijustifikasi oleh kinerjanya yang mapan dalam tugas-tugas analisis sentiment (Fathonah and Herliana, 2021). Selain itu, *Naive Bayes* telah dibandingkan dengan metode klasifikasi lain seperti *Support Vector Machine* (SVM) dan *Decision Tree*, yang menunjukkan keampuannya dalam klasifikasi sentiment (Fitriana, Setifani and Yusuf, 2020). Selain itu, *Naive Bayes* telah diterapkan dalam analisis sentimen di platform media sosial seperti *Twitter* untuk mengukur sentimen publik tentang topik-topik seperti tindakan lockdown (Dwison Alizah *et al.*, 2020), pengangkatan menteri (Pebrianto *et al.*, 2022), dan isu-isu terkait pandemi (Nurhazizah,

Nur Ichsan and Widiyanesti, 2022). Penelitian-penelitian ini telah menunjukkan kemampuan beradaptasi dan ketergantungan *Naive Bayes* dalam tugas analisis sentimen di berbagai domain dan platform. Maka dari itu untuk *novelty* (keterbaharuan) dari penelitian ini adalah mengkolaborasikan *Algrithm Machine Learning* yaitu *Naive Bayes* dan pengelolaan *Natural Language Processing* (NLP) dalam dataset tidak terstruktur.

## TINJAUAN PUSTAKA

### *CyberCrime*

*Cybercrime* merupakan bentuk-bentuk kejahatan yang timbul karena pemanfaatan teknologi internet. Menurut Hamzah (Hamzah and Marsita, 2012) dalam tulisannya mengartikan kejahatan komputer sebagai: "Kejahatan di bidang komputer secara umum dapat diartikan sebagai penggunaan komputer secara ilegal". Dari pengertian di atas, secara ringkas dapat dikatakan bahwa *cybercrime* dapat didefinisikan sebagai perbuatan melawan hukum yang dilakukan dengan menggunakan internet yang berbasis pada kecanggihan teknologi komputer dan telekomunikasi.

### Analisis Sentimen

Sentimen memiliki definisi pemikiran, pandangan, perasaan, emosi, pendapat, atau rasa umum dan analisis sentimen juga dikenal sebagai penambangan opini dengan penggunaan pemrosesan bahasa alami, analisis teks, dan linguistik komputasi untuk mengidentifikasi dan mengekstrak informasi subjektif. Analisis sentimen berusaha untuk mengidentifikasi sudut pandang yang mendasari suatu rentang teks contoh aplikasinya adalah mengklasifikasikan ulasan film apakah bagus atau tidak (Pang and Lee, 2004). Untuk melakukan analisis sentimen tentang suatu peristiwa, kita perlu mengajarkan komputer apa itu sentimen bagaimana mendefinisikan "positif" atau "negatif" dan "baik" atau "buruk". Di sinilah pembelajaran mesin berperan kita harus mengajari komputer arti positif dan negatif. Langkah pertama dalam proses ini adalah membangun model dari sekumpulan data pelatihan. Setelah model dibangun, kita akan menggunakannya untuk menganalisis data baru.

### Naive Bayes

*Naive Bayes* yang disebut *idiot's bayes*, *simple bayes*, dan *independence bayes* adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk data set berukuran besar. Estimasi Bayesian digunakan ketika memiliki beberapa informasi sebelumnya mengenai parameter. Misalnya sebelum melihat sampel untuk memperkirakan  $\mu$  rata-rata distribusi, mungkin memiliki beberapa keyakinan sebelumnya bahwa itu adalah dekat dengan 2 antara 1 dan 3. Keyakinan sebelumnya tersebut sangat penting ketika memiliki sebuah contoh kecil. Dalam kasus seperti itu, untuk menggabungkan data yaitu nilai yang dihitung dari sampel dan informasi yang didapat sebelumnya. Klasifikasi *Bayes* juga dikenal dengan *Naive Bayes*, memiliki kemampuan sebanding dengan pohon keputusan dan *neural network*. *Naive Bayes* juga merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema *bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas dan Defenisi lain yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Ghani and Sulaiman, 2023). Persamaan dari teorema *Naive Bayes* adalah :

$$P(H | X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

### Natural Language Processing (NLP)

Pemrosesan bahasa alami (*Natural Language Processing/NLP*) telah mengalami

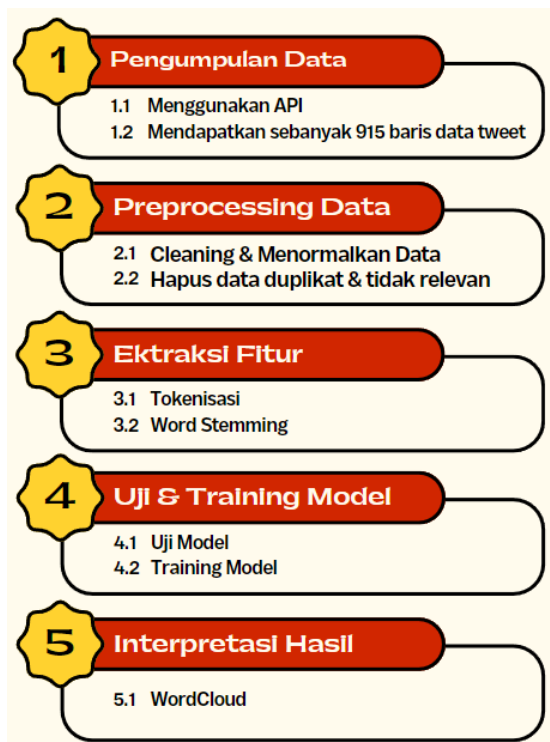
kemajuan dan aplikasi yang signifikan dalam berbagai domain. Penelitian di bidang NLP telah mengeksplorasi berbagai bidang seperti NLP sumber daya rendah (*Proceedings of The 3rd Workshop on Deep Learning Approaches for Low-Resource NLP*, 2022), pertimbangan etika dalam literatur NLP (Santy, Rani and Choudhury, 2021), grafik pengetahuan untuk COVID-19 (Chatterjee *et al.*, 2021), deteksi email phishing (Salloum *et al.*, 2022), penulisan ulang kueri percakapan (Liu *et al.*, 2021), dan peran NLP selama pandemi COVID-19 (Al-Garadi *et al.*, 2022).

### Literatur Review

Pada literatur review ini akan disampaikan hasil-hasil penelitian terdahulu yang menjadi rujukan dalam penelitian ini seperti terkait dengan analisis sentiment dengan menggunakan algoritma *naïve bayes* dan menggunakan dataset yang di crawling dari media social. Analisis sentimen menggunakan algoritma *Naive Bayes* telah mendapatkan perhatian yang signifikan dalam berbagai domain penelitian. Berbagai penelitian telah mengeksplorasi analisis sentimen dalam konteks yang berbeda, seperti menganalisis sentimen publik terhadap layanan pemerintah di platform media sosial seperti Twitter (Salim and Solichin, 2022), mengevaluasi sentimen terhadap pembelajaran online selama pandemi COVID-19 (Samsir *et al.*, 2021), dan menilai sentimen terhadap peristiwa tertentu seperti Sirkuit Mandalika Internasional (Mujahidin *et al.*, 2022). Pengklasifikasi *Naive Bayes* telah disukai karena kinerjanya dalam tugas analisis sentimen, sebagaimana dibuktikan dengan penerapannya dalam menganalisis persepsi publik terhadap pemilihan umum (Juanita, 2020), kebijakan pemerintah (Isnain, Marga and Alita, 2021), dan bahkan pemanfaatan teknologi oleh pemerintah daerah (Nanti Pikir *et al.*, 2021). Kesederhanaan dan akurasi yang tinggi dari algoritma *Naive Bayes* telah disorot dalam penelitian yang berfokus pada analisis sentimen dari berbagai dataset, termasuk ulasan film (Hamzah, 2021), evaluasi kinerja dosen (Sasmita, Pradnyana and Divayana, 2022), dan tanggapan publik terhadap layanan tertentu seperti Google Meet (Khoirunnisa and Februariyanti, 2022).

## METODE

Metode yang digunakan dalam penelitian ini terdiri dari 5 tahapan atau metode yang dijalankan oleh peneliti diantaranya yaitu Pengumpulan data, Preprocessing data, ekstraksi fitur, uji dan training model dan interpretasi hasil. Berikut metode penelitian dapat dilihat pada gambar 1 berikut ini :



Gambar 1. Metode Penelitian.

## HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan hasil dan pembahasan dari penelitian ini sesuai dengan metode-metode yang dipergunakan :

### Pengumpulan Data

Pengumpulan data yang dilakukan melalui metode data *scraping* dari media social *twitter* dengan menggunakan *Application Programming Interface* (API). Dari *twitter* data yang dikumpulkan dalam bentuk format *comma separated values* (CSV) yang berisikan komentar-komentar publik yang biasa disebut dengan *tweet* terkait dengan ancaman siber yang terjadi di dunia. Data yang dikumpulkan dari media social mengandung kata kunci “*cyber attack*”, “*cyber security*” dan “*malware*” serta dengan waktu penarikan data tahun 2022. Data yang terkumpul sebanyak 915

baris data. Selanjutnya data dijadikan satu untuk diolah dalam proses *processing data*.

### Praprocessing Data

Pada praproses data dilakukan proses *cleaning data* yaitu dengan melakukan pembersihan dataset dari kolom-kolom yang tidak terpakai, karakter-karakter khusus, dan data yang redundansi. Berikut hasil dari praprocessing data :

```

id_str,full_text
0 1697036281123832033,hurry up and contact https...
1 1697036227122282692,the 2023 ciso survival gui...
2 1697036044321853605,yuram3ki cause like someon...
3 1697036037640335809,tomilahren are you kidding...
4 1697035991364563396,uk cybersecurity agency wa...
..
910 1696931658757451979,questionsaglore ravisutanj...
911 1696931546844790799,ransomware this type of ma...
912 1696931508404027743,uh oh ⚠ this android malw...
913 1696931098939261236,scareware this is a type o...
914 1696931053716341090,get the cloud access secur...

[915 rows x 1 columns]
  
```

Gambar 2. Hasil Hapus Kolom

```

id_str full_text
0 1697036281123832033 hurry contact instagram cyber security expert ...
1 1697036227122282692 ciso survival guide great see kosmos recognize...
2 1697036044321853605 cause like someone like entry level coding hac...
3 1697036037640335809 kidding incited insurrection based lie fired h...
4 1697035991364563396 uk cybersecurity agency warns chatbot prompt i...
..
783 1696931658757451979 time educate u ur deficiencies malware amit ma...
784 1696931546844790799 type malware encrypts user files entire system...
785 1696931508404027743 uh oh android malware drain bank accounts
786 1696931098939261236 type malware uses fear tactics deceptive messa...
787 1696931053716341090 get cloud access security broker casb buyer gu...

[788 rows x 2 columns]
  
```

Gambar 3. Hasil Penghapusan Duplikasi Data

### Ekstraksi Fitur

Metode ekstraksi fitur dilakukan dengan menggunakan bahasa python untuk melakukan perintah tokenisasi dan word stemming. Berikut hasilnya :

```

id_str full_text \
0 1697036281123832033 hurry contact instagram cyber security expert ...
1 1697036227122282692 ciso survival guide great see kosmos recognize...
2 1697036044321853605 cause like someone like entry level coding hac...
3 1697036037640335809 kidding incited insurrection based lie fired h...
4 1697035991364563396 uk cybersecurity agency warns chatbot prompt i...
..
783 1696931658757451979 time educate u ur deficiencies malware amit ma...
784 1696931546844790799 type malware encrypts user files entire system...
785 1696931508404027743 uh oh android malware drain bank accounts
786 1696931098939261236 type malware uses fear tactics deceptive messa...
787 1696931053716341090 get cloud access security broker casb buyer gu...

lemmatized_text
0 hurry contact instagram cyber security expert ...
1 ciso survival guide great see kosmos recognize...
2 cause like someone like entry level code hack ...
3 kid incite insurrection base lie fire head cyb...
4 uk cybersecurity agency warn chatbot prompt in...
..
783 time educate u ur deficiencies malware amit ma...
784 type malware encrypt user file entire system r...
785 uh oh android malware drain bank account
786 type malware use fear tactic deceptive message...
787 get cloud access security broker casb buyer gu...

[788 rows x 3 columns]
  
```

Gambar 4. Hasil Tokenisasi

```

..      _str      full_text \
0  1697836281123832833  hurry contact instagram cyber security expert ...
1  1697836227122282692  ciso survival guide great see kosmos recognize...
2  1697836844321853685  cause like someone like entry level coding hac...
3  1697836837648353889  kidding incited insurrection based lie fired h...
4  1697835991364563396  uk cybersecurity agency warns chatbot prompt 1...
..      ...
783  1696931658757451979  time educate u ur deficiencies malware amit ma...
784  1696931546844798799  type malware encrypts user files entire system...
785  16969315084804827743  uh oh android malware drain bank accounts
786  1696931898939261236  type malware uses fear tactics deceptive messa...
787  1696931853716341098  get cloud access security broker cash buyer gu...

..      _str      lemmatized_text      score_polarity \
0  hurry contact instagram cyber security expert ...      -0.3612
1  ciso survival guide great see kosmos recognize...      0.7579
2  cause like someone like entry level code hack ...      0.7803
3  kid incite insurrection base lie fire head cyb...      -0.8368
4  uk cybersecurity agency warn chatbot prompt in...      -0.5423
..      ...
783  time educate u ur deficiencies malware amit ma...      0.8772
784  type malware encrypt user file entire system r...      -0.7579
785  uh oh android malware drain bank account      0.8808
786  type malware use fear tactic deceptive message...      -0.7269
787  get cloud access security broker cash buyer gu...      0.7269

..      _str      sentiment
0  Negatif
1  Positif
2  Positif
3  Negatif
4  Negatif
..      ...
783  Positif
784  Negatif
785  Netral
786  Negatif
787  Positif

[788 rows x 5 columns]
```

### Gambar 5. Hasil Steaming

## Uji dan Traning Model

Tahap ini algoritma *machine learning* yang digunakan untuk klasifikasi yaitu *naive bayes*. Berikut hasil uji dan traning model.

Accuracy: 0.620253164556962				
Classification Report:				
	precision	recall	f1-score	support
Negatif	0.67	0.67	0.67	54
Netral	0.48	0.31	0.38	35
Positif	0.63	0.74	0.68	69
accuracy			0.62	158
macro avg	0.59	0.57	0.58	158
weighted avg	0.61	0.62	0.61	158

```
1 new_text = ["Ini adalah contoh teks baru untuk diprediksi sentimennya."]
2 new_text_transformed = vectorizer.transform(new_text)
3 prediction = clf.predict(new_text_transformed)
4 print("Predicted Sentiment:", prediction)
```

Predicted Sentiment: ['Positif']

Gambar 6. Hasil Uji dan Training Model

## Interpretasi Hasil

Pada tahap ini akan dilakukan visualisasi dan interpretasi hasil dengan model Word Cloud dari kata-kata yang sering muncul pada *tweet-tweet* publik. Dari kata-kata tersebut dapat kita simpulkan bahwa untuk dan selama tahun 2022 kata-kata *tweet* yang terdapat didalam kalimat-kalimat *tweet* publik masih positif dan rata-rata publik menggunakan kata-kata attack, cyberattack, system, dan security untuk menggambarkan ancaman siber yang terjadi.



Gambar 7. WordCloud Model

## SIMPULAN

Berdasarkan dari hasil pembahasan yang dilakukan dapat diambil simpulan dari penelitian ini bahwa analisis sentiment terhadap tren ancaman keamanan siber sudah berhasil dilakukan dengan berdasarkan *tweet* dari publik melalui media social twitter dengan data tidak terstruktur yang menghasilkan model analisis sentiment yang baik dengan dibuktikan nilai F1-Score mendekati nilai 1 dan juga dibuktikan hasil nilai recall yang tinggi 74%, untuk accuracy yang didapat sebesar 62%. Selanjutnya untuk saran penelitian selanjutnya dapat dilakukan komparasi *algorithm machine learning classification* yang sudah dikolaborasikan dengan NLP dan penambahan jumlah record dalam dataset tidak terstruktur.

## DAFTAR PUSTAKA

- Al-Garadi, M.A.; *et al.* (2022) ‘Citation: The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges’, *MDPI Healthcare* [Preprint].
- AMT (2023) *6 Kasus Cybercrime di Indonesia yang Menyerang Server*.
- Chatterjee, A. *et al.* (2021) ‘Knowledge graphs for covid-19: An exploratory review of the current landscape’, *Journal of Personalized Medicine*. MDPI AG.
- Dwison Alizah, M. *et al.* (2020) ‘Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter’, *IJSE-Indonesian Journal on Software Engineering*, 6(2),
- Fathonah, F. and Herliana, A. (2021) ‘Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes’, *Jurnal Sains dan Informatika*, 7(2), pp. 155–164.

- Fitriana, D.N., Setifani, N.A. and Yusuf, A. (2020) 'Perbandingan Algoritma Naïve Bayes, Svm, Dan Decision Tree Untuk Klasifikasi Sms Spam', *JUSIM (Jurnal Sistem Informasi Musirawas)*, 5(2).
- Ghani, M.A. and Sulaiman, H. (2023) *Deteksi Spam Email dengan Metode Naive Bayes dan Particle Swarm Optimization (PSO)*, *Jurnal Informatika dan Teknologi*.
- Hamzah, A. and Marsita, B.D. (2012) *Aspek-Aspek Pidana dibidang Komputer*. Jakarta: Sinar Grafika.
- Hamzah, M.B. (2021) 'Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting', *Journal of Advances in Information Systems and Technology*, 3(1).
- Isnain, A.R., Marga, N.S. and Alita, D. (2021) 'Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(1), p. 55.
- Juanita, S. (2020) 'Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes', *Jurnal Media Informatika Budidarma*, 4(3).
- Khoirunnisa, F. and Februariyanti, H. (2022) 'Analisis Sentimen Kualitas Layanan Google Meet Menggunakan Naïve Bayes Classifiers Dan Association', *semanTIK*, 8(1), p. 35.
- Liu, H. *et al.* (2021) 'Conversational query rewriting with self-supervised learning', in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Institute of Electrical and Electronics Engineers Inc.,
- Mujahidin, S. *et al.* (2022) 'Bianglala Informatika Implementasi Analisis Sentimen Opini Publik Mengenai Sirkuit Internasional Mandalika Pada Twitter Menggunakan Metode Multinomial Naïve Bayes Classifier', *Jurnal Bianglala Informatika*, 10(2).
- Nanti Pikir, B. *et al.* (2021) 'Sentiment Analysis of Technology Utilization by Pekanbaru City Government Based on Community Interaction in Social Media', *Journal Of Artificial Intelligence and Application (JAIA)*, 2(1).
- Nurhazizah, E., Nur Ichsan, R. and Widiyanesti, S. (2022) 'Analisis Sentimen Dan Jaringan Sosial Pada Penyebaran Informasi Vaksinasi di Twitter', *Jurnal Swabumi*, 10(1), p. 2022.
- Pang, B. and Lee, L. (2004) 'A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts'.
- Pebrianto, R. *et al.* (2022) 'Analisis Sentimen Twitter Terhadap Menteri Indonesia Dengan Algoritma Support Vector Machine Dan Naive Bayes', *Jurnal Elektro dan Informatika*, 17(1).
- Proceedings of The 3rd Workshop on Deep Learning Approaches for Low-Resource NLP* (2022).
- Salim, E. and Solichin, A. (2022) 'Analisis Sentimen Pada Media Sosial Twitter Terhadap Pelayanan Dinas Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes', *Indonesia Journal Information System (IDEALIS)*, 5(2).
- Salloum, S. *et al.* (2022) 'A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques', *IEEE Access*. Institute of Electrical and Electronics Engineers Inc
- Samsir *et al.* (2021) 'Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes', *Jurnal Media Informatika Budidarma*, 5(1).
- Santy, S., Rani, A. and Choudhury, M. (2021) 'Use of Formal Ethical Reviews in NLP Literature: Historical Trends and Current Practices', *ACL-IJCNLP*.
- Sasmita, A., Pradnyana, G.A. and Divayana, D.G.H. (2022) 'Pengembangan Sistem Analisis Sentimen Untuk Evaluasi Kinerja Dosen Universitas Pendidikan Ganesha Dengan Metode Naïve Bayes', *JST (Jurnal Sains dan Teknologi)*, 11(2).
- Septiana, R.D., Susanto, A.B. and Tukiyat (2021) 'Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic Dan Particle Swarm Optimization', *Jurnal Sistem Komputer dan Kecerdasan Buatan*, 5(1).