ANALISIS SENTIMEN PADA PENGGUNA TWITTER TERHADAP PROGRAM KAMPUS MERDEKA MENGGUNAKAN NAÏVE BAYES

Ahmad Nursidik Dinar, Agung Susilo Yuda Irawan, Yuyun Umaidah
Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170157@student.unsika.ac.id

ABSTRAK

Pembuatan berbagai jenis program Kampus Merdeka ramai saat ini, hal ini menyebabkan munculnya banyak masalah yang menyertainya. dalam tiga tahun pelaksanaannya, banyak kesalahan yang terus berulang, seperti keterlambatan pencairan uang saku untuk biaya hidup mahasiswa selama menjalankan program. Masalah ini bahkan menyebabkan mahasiswa membuat petisi untuk menuntut pencairan yang lebih cepat, namun hal tersebut tidak memberikan dampak yang signifikan karena masalah keterlambatan pencairan uang saku terulang kembali. Namun, tweet yang diposting oleh peserta di Twitter dapat dijadikan sumber informasi yang penting dan bermanfaat. Oleh karena itu, analisis sentimen dapat digunakan sebagai solusi untuk mengolah suara tersebut dengan menggunakan algoritma Naïve Bayes Classifier. Tujuan dari penelitian ini adalah untuk mengevaluasi program Kampus Merdeka agar dapat menghindari mengulangi kesalahan yang sama di masa depan, dengan mengklasifikasikan pendapat tentang program Kampus Merdeka berdasarkan kelas positif dan negatif. Pengujian dilakukan dengan menggunakan empat skenario, yang berarti data dibagi menjadi dua bagian dengan perbandingan 90:10, 80:20, 70:30 dan 60:40 untuk data training dan data testing. Hasil evaluasi menunjukkan bahwa pembagian dataset 90:10 adalah yang terbaik dengan menghasilkan akurasi yang terbaik baik yaitu 97,92%.

Kata kunci: *Analisis Sentimen, Naïve Bayes, Kampus Merdeka, Twitter*

1. PENDAHULUAN

Program Kampus Merdeka, yang merupakan inisiatif baru dari Kementerian Pendidikan dan Kebudayaan, menjadi salah satu topik yang sedang ramai diperbincangkan oleh masyarakat saat ini. Program ini bertujuan untuk mendorong mahasiswa untuk menguasai berbagai ilmu yang dapat membantu mereka memasuki dunia kerja. Namun, meskipun program ini memiliki tujuan yang baik, rencana pelaksanaannya yang dikeluarkan oleh Menteri Pendidikan dan Kebudayaan telah membawa berbagai komentar di kalangan masyarakat Indonesia.

Oleh karena itu, diperlukan analisis untuk mengevaluasi program Kampus Merdeka agar dapat menghindari mengulangi kesalahan yang sama di masa depan. Analisis ini dapat dilakukan melalui berbagai media, dan salah satu media yang dapat digunakan sebagai alat untuk melakukan analisis adalah Twitter.

Sebelumnya, sebuah penelitian telah menggunakan komentar terhadap calon presiden Indonesia sebagai objek penelitiannya. Dalam penelitian tersebut, digunakan algoritma Naive Bayes Classifier dan berhasil mencapai tingkat akurasi sebesar 60% [1]. Namun, dalam studi lain tentang analisis sentimen yang dilakukan dengan algoritma Naive Bayes Classifier juga digunakan objek penelitiannya adalah maskapai penerbangan berhasil mencapai tingkat akurasi sebesar 81% dalam melakukan analisis sentimen pada objek penelitiannya [2].

Dari uraian latar belakang yang telah dijabarkan, maka akan dilakukan sebuah penelitian yang bertujuan untuk melakukan analisis sentimen pada pengguna Twitter terkait dengan program Kampus Merdeka. Penelitian ini bertujuan untuk mengumpulkan opini dan pendapat dari para pengguna Twitter terkait dengan pelaksanaan program Kampus Merdeka. Metode yang akan digunakan dalam penelitian ini adalah algoritma Naïve Bayes yang dikenal sebagai salah satu algoritma yang terkenal dan masuk dalam 10 besar algoritma pada data mining [3]. Diharapkan bahwa penelitian ini akan memberikan hasil yang baik dan dapat dijadikan acuan dalam pengembangan program Kampus Merdeka di masa depan.

2. TINJAUAN PUSTAKA

2.1. Kampus Merdeka

Kampus Merdeka adalah kebijakan Menteri Pendidikan dan Kebudayaan, Nadiem Makarim, yang memberikan kesempatan bagi mahasiswa untuk mengambil mata kuliah di luar program studi selama tiga semester. Mahasiswa dapat mengeksplorasi dan memilih mata kuliah yang diinginkan di universitas [4].

2.2. Twitter

Twitter ialah platform media sosial gaya microblogging yang didirikan oleh Jack Dorsey dan diluncurkan pada Juli 2006. Twitter populer di kalangan pendidik untuk keperluan profesional dan penelitian [5].

2.3. Analisis Sentimen

Analisis sentimen ialah sebuah disiplin ilmu yang mempelajari opini, perasaan, penilaian, review, dan sikap seseorang terhadap suatu objek, seperti

organisasi, layanan, produk, isu, individu, dan peristiwa. Analisis ini bertujuan untuk memahami pandangan individu yang menunjukkan emosi positif atau negatif terhadap objek tertentu, khususnya pada media sosial. Saat ini, analisis sentimen menjadi inti dari penelitian media sosial [6].

2.4. Text Mining

Text mining adalah ilmu yang mengolah teks menjadi informasi melalui pola statistik untuk memprediksi pola dan tren. Tujuannya adalah untuk menganalisis sikap, perasaan, pendapat, penilaian, evaluasi, dan emosi individu terhadap topik, organisasi, layanan, orang, atau aktivitas tertentu menurut Liu [7].

2.5. Knowledge Discovery in Database (KDD)

KDD adalah proses untuk menemukan informasi baru yang lebih berharga dan mudah dipahami dari gudang data yang besar dan kompleks [8]. Metode KDD menggabungkan dan menginterpretasikan hasil yang diperoleh dari gugusan data dengan menggunakan ilmu lain. Proses KDD dimulai dengan menetapkan tujuan dan diakhiri dengan evaluasi.

2.6. Data Selection

Sebelum memulai fase ekstraksi informasi, KDD harus melakukan Data Selection. Data yang terpilih akan digunakan dalam proses data mining dan disimpan dalam file terpisah dari database operasional. Dalam pengumpulan data menggunakan teknik sampling, teknik sampling adalah proses memilih sampel data yang representatif yang dapat menjelaskan semua data dalam kumpulan data tertentu dengan menguji sebagian data [9].

2.7. Preprocessing

Preprocessing adalah proses persiapan data mentah, biasanya dengan menghilangkan data yang tidak relevan atau tidak sesuai, atau mengubah data ke dalam format yang lebih mudah diproses oleh sistem. Ini adalah langkah penting dalam analisis sentimen, terutama untuk data media sosial yang tidak terstruktur, karena berisi kata-kata dan kalimat informal yang menimbulkan banyak noise [10]. Sebelum melakukan data mining, data harus dibersihkan agar sesuai dengan KDD. Proses pembersihan meliputi perbaikan kesalahan data seperti kesalahan penulisan, pemeriksaan ketidakkonsistenan data, dan penghapusan data duplikat.

2.8. Text Transformation

Text transformation dilakukan agar dapat menggunakan algoritma dalam pembelajaran dengan mudah, seperti pencarian data numerik [11]. Proses ini melibatkan pemecahan kalimat menjadi kata-kata, kemudian mengubah himpunan kata menjadi term vector menggunakan algoritma TF-IDF. Hal ini diperlukan karena himpunan kata tidak dapat langsung digunakan dalam algoritma pembelajaran.

2.9. Data Mining

Data mining adalah proses menggali informasi atau pola dari data yang sebelumnya dianggap tidak berguna atau tersembunyi. Data yang dianggap sampah karena tidak terstruktur dan tidak berguna dapat diubah menjadi informasi atau pengetahuan baru yang berguna melalui data mining [12].

2.10. Evaluation

Tahap evaluasi adalah tahap terakhir dari proses KDD yang mencakup kesimpulan yang diambil dari data mining. Untuk mengukur akurasi, penelitian ini menggunakan confusion matrix. Hasil dari data mining harus disajikan dalam format yang dapat dipahami oleh para pemangku kepentingan.

2.11. Naïve Bayes Classifier (NBC)

Algoritma Naive Bayes Classifier (NBC) adalah metode klasifikasi yang menggunakan teorema Bayesian dari ilmu statistik untuk memprediksi probabilitas keanggotaan kelas. Algoritma ini sangat efektif dalam mengklasifikasikan data nominal. Teorema Bayes yang menjadi dasar dari algoritma ini dapat dirumuskan secara sederhana [12].

2.12. Confusion Matrix

Confusion matrix adalah sebuah matriks yang digunakan untuk mengukur kualitas klasifikasi suatu algoritma pada kelas tertentu [13]. Matriks ini penting untuk mengevaluasi kualitas model statistik atau algoritma data mining. Ada beberapa jenis confusion matrix yang digunakan untuk menguji model data mining, seperti Accuracy dan lainnya. Pada tabel 1 berikut dapat dilihat bentuk confusion matrix.

Tabel 1. Confusion matrix [14]

Actual Class	**Predicted Class**	
	+	-
+	TP	TN
-	FP	FN

2.13. Accuracy

Banyak peneliti setuju bahwa accuracy (rasio jumlah sampel yang diklasifikasikan dengan benar dengan jumlah total sampel) adalah metrik yang paling masuk akal. Namun, ketika jumlah data di setiap kategori bervariasi, akurasi tidak lagi dianggap sebagai ukuran yang andal karena memberikan perkiraan yang terlalu optimis tentang kemampuan mengklasifikasikan kelas mayoritas [15]. Perhitungan accuracy dapat dirumuskan sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2.14. Tools

Peneliti akan menggunakan bahasa pemrograman dan coding environment untuk menunjang penelitian. Bahasa pemrograman adalah bahasa buatan yang digunakan untuk mengontrol

perilaku mesin seperti komputer. Sementara itu, coding environment adalah alat yang digunakan untuk menjalankan bahasa pemrograman dari barisan kata-kata menjadi sebuah aksi. Dalam penelitian ini, peneliti akan menggunakan beberapa bahasa pemrograman dan coding environment untuk mendukung penelitian yang dilakukan.

2.15. Python

Python adalah bahasa pemrograman yang banyak digunakan di berbagai sistem operasi. Python termasuk dalam kategori bahasa pemrograman tipe interpreted language yang artinya kode program tidak perlu diterjemahkan ke dalam kode komputer yang dapat dipahami sebelum dieksekusi, melainkan diterjemahkan saat runtime. Bahasa pemrograman ini dikembangkan oleh Guido Van Rossum dan memiliki banyak aplikasi, seperti dalam bidang ilmiah dan numerik, dengan menggunakan library seperti twint, pandas, dan scikit-learn [16].

2.16. Google Colab

Google Collaboratory atau Google Interactive Notebook adalah sebuah layanan cloud gratis dari Google yang bekerja seperti Notebook Jupyter. Pengguna tidak perlu menginstalnya terlebih dahulu. Layanan ini dapat dengan mudah terhubung ke Notebook Jupyter di komputer pengguna sendiri (runtime lokal), Google Drive, atau Github. Google Colab mendukung pekerjaan kolaboratif, seperti membagikan kode secara online [17].

3. METODE PENELITIAN

Pada penelitian ini, digunakan metodologi KDD (Knowledge Discovery in Database) sebagai metode untuk mengidentifikasi pola dalam data yang relevan, bermanfaat, dan mudah dipahami. Berdasarkan proses KDD (Knowledge Discovery in Database) yang telah dipaparkan dalam tinjauan pustaka diatas, dapat dibuat alur penelitian dengan tahapan yaitu:

Gambar 1. Rancangan penelitian

4. HASIL DAN PEMBAHASAN

Langkah pertama dalam penelitian ini adalah merumuskan masalah untuk menjawab mengapa penelitian perlu dilakukan. Kampus Merdeka adalah kebijakan Menteri Pendidikan dan Kebudayaan Nadiem Makarim, namun program ini mengalami masalah dalam pelaksanaannya, seperti keterlambatan pencairan uang saku yang sering terjadi. Mahasiswa yang mengikuti program sangat terdampak oleh masalah ini, dan beberapa bahkan harus meminjam uang dari pinjaman online untuk memenuhi kebutuhan mereka.

Peneliti menemukan beberapa alasan mengapa penelitian ini perlu dilakukan. Pertama, analisis sentimen pengguna media sosial terhadap program Kampus Merdeka dapat memberikan informasi penting tentang respons masyarakat terhadap program ini. Dengan memahami sentimen pengguna media sosial, pemerintah dan stakeholder terkait dapat meningkatkan kualitas program ini. Kedua, metode Naive Bayes diharapkan dapat menghasilkan hasil analisis sentimen yang lebih efektif dan efisien dibandingkan dengan metode analisis lainnya. Terakhir, dengan menguji performa metode Naive Bayes dalam melakukan analisis sentimen dari pengguna media sosial Twitter terhadap program Kampus Merdeka, peneliti dapat mengevaluasi keakuratan dan kinerja metode tersebut dan memberikan kontribusi pada pengembangan ilmu pengetahuan dan teknologi.

Setelah menentukan tujuan dan merumuskan masalah, peneliti melakukan pencarian literatur yang cermat dan teliti, termasuk referensi terbaru dalam bidang yang akan diteliti. Kemudian, peneliti membaca dan menganalisis setiap sumber yang relevan dan mencari gagasan baru yang dapat diterapkan dalam penelitian ini. Peneliti dapat mengembangkan kerangka teoretis untuk mengorganisir konsep-konsep kunci dan memperjelas hubungan antara variabel yang akan diteliti. Setelah itu, peneliti merancang metode penelitian yang sesuai dengan tujuan dan hipotesis yang telah ditetapkan. Dengan melakukan studi literatur secara komprehensif, peneliti dapat membangun landasan teoritis yang kuat dan meningkatkan validitas dan reliabilitas penelitian.

Data tweet diambil dari media sosial Twitter yang terkait dengan kata kunci "@KampusMerdeka" atau "#KampusMerdeka" dan menggunakan bahasa Indonesia mulai bulan Agustus 2022 hingga Desember 2022. Awalnya, terdapat 6.474 data yang berhasil diambil dengan menggunakan library Python bernama twint.

Setelah crawling data selesai, langkah berikutnya adalah memilih data yang sesuai dengan batasan penelitian. Setelah dipilih, jumlah data awal 6.474 berkurang menjadi 1.911 data. Hanya atribut teks yang digunakan untuk pembersihan data dan klasifikasi selanjutnya.

Dalam tahap pre-processing pada pengolahan data tweet, terdapat beberapa langkah yang dilakukan. Pertama-tama, setiap kata pada dataset akan diubah menjadi huruf kecil atau lower case, serta dilakukan penghapusan karakter yang tidak berpengaruh seperti tanda baca, simbol, link/url, dan angka pada tahap case folding. Selanjutnya, tweet akan dibagi menjadi kata-kata pada tahap tokenisasi untuk memudahkan proses transformasi selanjutnya. Setelah tokenizing, dilakukan normalisasi untuk memperbaiki kata-kata dengan kesalahan pengejaan atau kata-kata yang disingkat. Kemudian, dilakukan filtering atau penghilangan stopword pada kata-kata yang dianggap tidak relevan untuk klasifikasi selanjutnya, seperti kata-kata umum dalam bahasa Indonesia diantaranya "dan", "atau", dan "namun". Untuk melakukan tahap ini, digunakan library nltk pada Python, dan dilakukan juga filtering manual untuk memperbaiki hasil. Tahap terakhir dalam pre-processing adalah stemming, yang berfungsi untuk mengurangi frekuensi kata turunan dengan menghapus imbuhan pada kata. Untuk proses stemming pada data tweet, digunakan library Sastrawi di Python.

Setelah tahap pre-processing selesai, dilakukan pembobotan sentimen untuk pelabelan data dengan menggunakan metode TF-IDF (Term Frequency – Inverse Document Frequency) yang menggunakan fungsi TfidfVectorizer pada library sklearn. Selanjutnya, dilakukan tahap klasifikasi menggunakan algoritma Multinomial Naive Bayes dengan menggunakan library sklearn. naive_bayes pada bahasa pemrograman Python, yang menyediakan kelas MultinomialNB untuk melatih dan membuat model. Sebelumnya, perlu disiapkan data training dan data testing yang akan digunakan dalam proses pelatihan dan pengujian model. Setelah tahap klasifikasi, dilakukan evaluasi performa model dengan mengukur akurasi menggunakan confusion matrix.

Berikut empat skenario pembagian data yang dilakukan dan hasil akurasi yang didapatkan dari setiap scenario.

4.1. Skenario 1

Berikut dapat dilihat pada Gambar 6 pembagian data training dan testing dengan perbandingan 90:10.

```
[13] #Modeling Naive Bayes
     X = df_vektorizer
     y = steming['label']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.1, random_state = 0)
     mnb = MultinomialNB().fit(X_train, y_train)
     y_pred = mnb.predict(X_test)
     y_pred_train = mnb.predict(X_train)
     X_train.shape, X_test.shape

     ((1719, 3166), (192, 3166))
```

Gambar 6. Pembagian Data Skenario 1

```
[23] #evaluasi
     cm = confusion_matrix(y_test, y_pred)
     cm_matrix = pd.DataFrame(data=cm, columns=['Actual : Positive', 'Actual : Negative'],
                              index=['Predict : Positive', 'Predict : Negative'])
     print(cm_matrix)
     print('Accuracy:', accuracy_score(y_test,y_pred))

                    Actual : Positive  Actual : Negative
     Predict : Positive                 1                 4
     Predict : Negative                 0               187
     Accuracy: 0.9791666666666666
```

Gambar 7. Evaluasi Skenario 1

Berdasarkan Gambar 7, hasil dari klasifikasi menggunakan Naive Bayes dengan seleksi fitur TF-IDF pada perbandingan 90:10 didapatkan hasil dengan nilai akurasi 97.92%.

4.2. Skenario 2

Berikut dapat dilihat pada Gambar 8 pembagian data training dan testing dengan perbandingan 80:20.

```
[14] #Modeling Naive Bayes
     X = df_vektorizer
     y = steming['label']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)
     mnb = MultinomialNB().fit(X_train, y_train)
     y_pred = mnb.predict(X_test)
     y_pred_train = mnb.predict(X_train)
     X_train.shape, X_test.shape

     ((1528, 3166), (383, 3166))
```

Gambar 8. Pembagian Data Skenario 2

```
[25] #evaluasi
     cm = confusion_matrix(y_test, y_pred)
     cm_matrix = pd.DataFrame(data=cm, columns=['Actual : Positive', 'Actual : Negative'],
                              index=['Predict : Positive', 'Predict : Negative'])
     print(cm_matrix)
     print('Accuracy:', accuracy_score(y_test,y_pred))

                    Actual : Positive  Actual : Negative
     Predict : Positive                 1                 9
     Predict : Negative                 0               373
     Accuracy: 0.9765013054830287
```

Gambar 9. Evaluasi Skenario 2

Berdasarkan Gambar 9, hasil dari klasifikasi menggunakan Naive Bayes dengan seleksi fitur TF-IDF pada perbandingan 80:20 didapatkan hasil dengan nilai akurasi 97.66%.

4.3. Skenario 3

Berikut dapat dilihat pada Gambar 10 pembagian data training dan testing dengan perbandingan 70:30.

```
[15] #Modeling Naive Bayes
     X = df_vektorizer
     y = steming['label']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 0)
     mnb = MultinomialNB().fit(X_train, y_train)
     y_pred = mnb.predict(X_test)
     y_pred_train = mnb.predict(X_train)
     X_train.shape, X_test.shape

     ((1337, 3166), (574, 3166))
```

Gambar 10. Pembagian Data Skenario 3

```
[27] #evaluasi
     cm = confusion_matrix(y_test, y_pred)
     cm_matrix = pd.DataFrame(data=cm, columns=['Actual : Positive', 'Actual : Negative'],
                              index=['Predict : Positive', 'Predict : Negative'])
     print(cm_matrix)
     print('Accuracy:', accuracy_score(y_test,y_pred))

                    Actual : Positive  Actual : Negative
     Predict : Positive                 0                18
     Predict : Negative                 0               556
     Accuracy: 0.9686411149825784
```

Gambar 11. Skenario 3

Berdasarkan Gambar 11, hasil dari klasifikasi menggunakan Naive Bayes dengan seleksi fitur TF-IDF pada perbandingan 70:30 didapatkan hasil dengan nilai akurasi 96.86%.

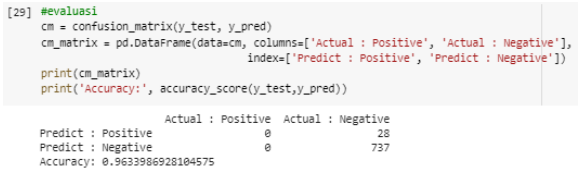
4.4. Skenario 4

Berikut dapat dilihat pada Gambar 12 pembagian data training dan testing dengan perbandingan 60:40.

```
[16] #Modeling Naive Bayes
     X = df_vektorizer
     y = steming['label']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4, random_state = 0)
     mnb = MultinomialNB().fit(X_train, y_train)
     y_pred = mnb.predict(X_test)
     y_pred_train = mnb.predict(X_train)
     X_train.shape, X_test.shape

     ((1146, 3166), (765, 3166))
```

Gambar 12. Pembagian Data Skenario 4



```
[29] #evaluasi
     cm = confusion_matrix(y_test, y_pred)
     cm_matrix = pd.DataFrame(data=cm, columns=['Actual : Positive', 'Actual : Negative'],
                              index=['Predict : Positive', 'Predict : Negative'])
     print(cm_matrix)
     print('Accuracy:', accuracy_score(y_test,y_pred))

                    Actual : Positive  Actual : Negative
     Predict : Positive                 0                 28
     Predict : Negative                 0                737
     Accuracy: 0.9633986928104575
```

Gambar 13. Evaluasi Skenario 4

Berdasarkan Gambar 13, hasil dari klasifikasi menggunakan Naive Bayes dengan seleksi fitur TF-IDF pada perbandingan 60:30 didapatkan hasil dengan nilai akurasi 96.33%.

5. KESIMPULAN DAN SARAN

Penelitian ini melakukan analisis sentiment terhadap program Kampus Merdeka menggunakan algoritma Naïve Bayes Classifier melalui beberapa tahap, yaitu crawling data, preprocessing, transformasi data, dan penerapan algoritma. Setelah validasi oleh ahli bahasa Indonesia, diperoleh data sebanyak 1911 dokumen yang kemudian dilakukan preprocessing dan transformasi dengan menggunakan teknik TF-IDF. Data kemudian dibagi menjadi data training dan data testing untuk melatih dan menguji model.

Hasil pengujian menunjukkan bahwa model yang dibangun cukup akurat dalam melakukan sentiment analysis terhadap program Kampus Merdeka dengan nilai akurasi terbesar 0.979166666 atau setara dengan 97,92% dari perbandingan data 90:10.

Saran untuk penelitian selanjutnya adalah menggali metode pelabelan data yang lebih baik untuk meningkatkan nilai akurasi. Selain itu, peneliti dapat menggunakan metode pembobotan dan seleksi fitur seperti N-Gram, Particle Swarm Optimization, Chi Square, dan Stable Mutation Jump Strategy untuk meningkatkan performa model.

DAFTAR PUSTAKA

[1] Yana, A.A. and Santoso, T., 2020, November. Sentiment analysis of facebook comments on indonesian presidential candidates using the naïve bayes method. In Journal of Physics: Conference Series (Vol. 1641, No. 1, p. 012012). IOP Publishing.

[2] Negara, A.B.P., Muhardi, H. and Putri, I.M., 2020. Analisis sentimen maskapai penerbangan menggunakan metode naive bayes dan seleksi fitur information gain. J. Teknol. Inf. dan Ilmu Komput, 7(3).

[3] Brunton, S.L. and Kutz, J.N., 2022. Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press.

[4] Leuwol, N.V., Wula, P., Purba, B., Marzuki, I., Brata, D.P.N., Efendi, M.Y., Masrul, M., Sahri, S., Ahdiyat, M., Sari, I.N. and Gusty, S., 2020. Pengembangan Sumber Daya Manusia Perguruan Tinggi: Sebuah Konsep, Fakta dan Gagasan. Yayasan Kita Menulis.

[5] Carpenter, J., Tani, T., Morrison, S. and Keane, J., 2022. Exploring the landscape of educator professional activity on Twitter: An analysis of 16 education-related Twitter hashtags. Professional Development in Education, 48(5), pp.784-805.

[6] Septian, J.A., Fachrudin, T.M. and Nugroho, A., 2019. Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. INSYST: Journal of Intelligent System and Computation, 1(1), pp.43-49.

[7] Luqyana, W.A., Cholissodin, I. and Perdana, R.S., 2018. Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2(11), pp.4704-4713.

[8] Widaningsih, S., 2019. Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm. Jurnal Tekno Insentif, 13(1), pp.16-25.

[9] Asroni, A., Fitri, H. and Prasetyo, E., 2018. Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). Semesta Teknika, 21(1), pp.60-64.

[10] Syakuro, A., 2017. Analisis sentimen masyarakat terhadap e-commerce pada media sosial menggunakan metode Naïve Bayes Classifier (NBC) dengan seleksi fitur Information Gain (IG) (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

[11] Kalra, V. and Aggarwal, R., 2017. Importance of Text Data Preprocessing & Implementation in RapidMiner. ICITKM, 14, pp.71-75.

[12] Suntoro, J., 2019. DATA MINING: Algoritma dan Implementasi dengan Pemrograman php. Elex Media Komputindo.

[13] Arifin, O. and Sasongko, T.B., 2018. Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier untuk klasifikasi jalur minat SMA. SEMNASTEKNOMEDIA ONLINE, 6(1), pp.1-2.

[14] Aldean, M.Y., Paradise, P. and Nugraha, N.A.S., 2022. Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). Journal of Informatics Information System Software Engineering and Applications (INISTA), 4(2), pp.64-72.

[15] Tanggraeni, A.I. and Sitokdana, M.N., 2022. Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve

Bayes. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 9(2), pp.785-795.

[16] Applications for python (no date) Python.org. Available at: https://www.python.org/about/apps/#scientific-and-numeric (Accessed: October 20, 2022).

[17] Applications for python (no date) Python.org. Available at: https://www.python.org/about/apps/#scientific-and-numeric (Accessed: October 20, 2022).