

Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, KNN, dan Decision Tree

Tobby Wiratama Putra ¹, Agung Triayudi ^{2*}, Andrianingsih ³

^{1,2,3} Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

article info

Article history:

Received 2 June 2021

Received in revised form

13 July 2021

Accepted 25 August 2021

Available *online* January 2022

DOI:

<https://doi.org/10.35870/jtik.v6i1.368>

Keywords:

Corona Virus; Twitter; Naïve Bayes; KNN; Decision Tree.

Kata Kunci:

Virus Corona; Twitter; Naïve Bayes; KNN; Decision Tree.

abstract

Corona Virus Disease or better known as Corona Virus has attacked all corners of the world, including Indonesia. It has been almost +1 year that this virus is still attacking Indonesia and has an impact on various sectors, one of which is education. To contain the spread of the virus, the government has set an online education system. Many complaints were experienced by students, especially students who returned home, many of them experienced difficulties in accessing the internet, so many tasks to complete, and many more. The purpose of this study is to obtain the accuracy of the classification results based on the level of accuracy of the perceptions of students who use Twitter regarding the implementation of online lectures. In this research, the method used is Naive Bayes, KNN, and Decision Tree. The data used in this study is Twitter data by crawling data. From the results of this study, the Decision Tree method has a high value among other methods with an accuracy of 61.92%, precision of 73.63%, and recall of 11.42%.

abstrak

Corona Virus Disease atau lebih dikenal Virus Corona telah menyerang ke seluruh penjuru dunia tidak terkecuali Indonesia. Sudah hampir +1 tahun virus ini masih menyerang Indonesia dan berdampak di berbagai sektor bidang, salah satunya bidang pendidikan. Demi menanggulangi penyebaran virus, pemerintah menetapkan sistem pendidikan dilaksanakan secara online. Banyak keluhan kesah yang dialami mahasiswa terutama mahasiswa yang pulang kampung, banyak yang dialami mereka mulai kesulitan akses internet, tugas yang harus diselesaikan begitu banyak dan masih banyak lagi. Adapun tujuan dari penelitian ini yaitu untuk mendapatkan ketepatan hasil klasifikasi berdasarkan tingkat akurasi dari persepsi mahasiswa yang menggunakan twitter mengenai pemberlakuan kuliah secara online. Dalam penelitian ini ada tiga metode yang akan dipakai untuk proses klasifikasi. Metodenya tersebut yaitu Naive Bayes, KNN dan Decision Tree. Data yang dipakai dalam penelitian ini merupakan data twitter yang diambil dengan cara crawling data. Dari hasil pengujian diketahui metode Decision Tree memiliki nilai yang lebih tinggi diantara metode lainnya dengan akurasi sebesar 61.92%, presisi 73.63%, dan recall 11.42%.

*Corresponding author. Email: agungtriayudi@civitas.unas.ac.id ².

© E-ISSN: 2580-1643.

Copyright © 2022. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET) (<http://creativecommons.org/licenses/by/4.0/>).

1. Latar Belakang

Virus Corona sudah menyerang diberbagai penjuru dunia tidak terkecuali di Indonesia. Hampir satu tahun sudah Indonesia terpapar Virus Corona. Sebagian orang meyakini berasal dari Wuhan, China kemudian menyerang ke beberapa negara bahkan ke penjuru dunia. Viirus ini mengakibatkan sejumlah negara melakukan kebijakan seperti memakai masker, membatasi jarak dan lain sebagainya.

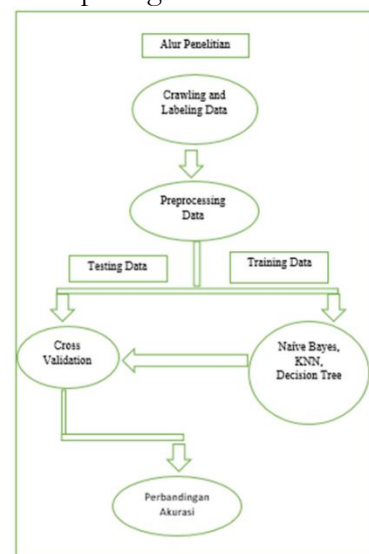
Demikian juga di Indonesia, banyak kebijakan pemerintah guna mencegah penyebaran virus. Tanpa disadari virus ini juga memiliki dampak yang cukup merugikan bagi kalangan masyarakat. Sementara itu, ada aspek yang penting menyangkut keselamatan dari pandemi yang sedang terjadi yang menjadi fokus utama [1]. Pemerintah melakukan kebijakan demi mengurangi penularan virus dengan melakukan kebijakan seperti *Work from Home* (WFH), PSBB dan juga *New Normal*. Dengan diberlakukannya kebijakan ini tentu memiliki dampak diberbagai bidang salah satunya bidang pendidikan [2]. Dengan diberlakukannya kebijakan ini mengakibatkan dunia pendidikan belum kembali sepenuhnya, sehingga mengharuskan mahasiswa melakukan pembelajaran secara *online*/daring. Banyak pendapat mahasiswa yang positif dan negatif terhadap pemberlakuan kuliah *online*.

Selain itu jejaring sosial merupakan tempat untuk menyampaikan segala aspirasi, kritik, dan pendapat. Media sosial yang cukup sering digunakan adalah Twitter [3]. Twitter merupakan aplikasi media sosial yang sering dipakai oleh penggunaanya dalam menyampaikan opininya [4]. Dalam *tweet* tersebut banyak pendapat yang tidak selaras atau berbeda oleh sebab itu akan dilakukan analisis sentimen dengan klasifikasi positif, dan negatif menggunakan metode Naïve Bayes, KNN, dan Decision Three. Pada penelitian Muhammad Riefky dan Aliffia Rahma Anandayani (2021) tentang klasifikasi persepsi pengguna twitter terhadap keringanan UKT 2020, Penelitian tersebut bertujuan untuk mendapatkan hasil ketepatan klasifikasi persepsi pengguna twitter terhadap tuntutan keringan biaya UKT menggunakan metode k-nearest neighbor, penelitian tersebut menggunakan kata “mendikbuddicarimasiswa”, “kampus”, “dan “ukt” dan mendapat nilai akurasi sebesar 83,25% [2].

Penelitian tentang analisis sentimen mengenai transportasi KRL *Commuterline* juga dilakukan oleh Nova Tri Romadloni dkk (2019), penelitian tersebut untuk mengklasifikasikan opini masyarakat terhadap KRL *Commuterline* menggunakan metode naïve bayes, k-NN, dan decision three yang mendapatkan hasil 80% untuk akurasi, 66,67% untuk presisi, dan 100% untuk sensitifitas, dan 66,67% untuk *specivicity* dengan menggunakan metode naïve bayes. Kemudian hasil yang didapat menggunakan metode k-NN adalah sebagai berikut nilai akurasi sebesar 80%, 100% untuk *precision*, *sensitivity* 50%, dan *specificity* 100%. Kemudian menggunakan metode decision three didapatkan hasil akurasi sebesar 100%, 100% untuk *precision*, kemudian *sensitivity* 100%, dan *specificity* 100% [4]. Pada penelitian yang dilakukan oleh Sigit Kurniawan dkk (2019), tentang analisis sentimen tokoh politik. Tujuan penelitian adalah untuk mendapatkan metode yang terbaik dalam klasifikasi analisis sentimen. Dalam penelitian yang dilakukan Sigit (2019) ternyata metode *Support vector machines* (SVMs) yang berbasis *Particle Swarm Optimization* mendapat nilai yang baik sebesar 78.40% akurasi dan AUC 0.850 [5]. Hasil dari penelitian ini akan memberikan gambaran apakah Pemberlakuan Perkuliahan Secara *Online*/Daring mendapat opini positif atau negatif, berdasarkan tingkat akurasi dari ketiga metode tersebut dan membandingkannya.

2. Metode Penelitian

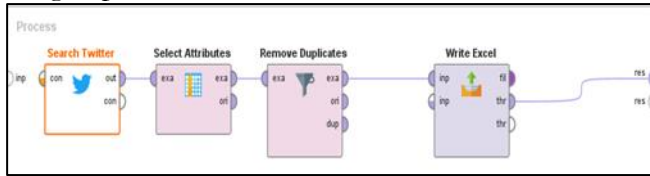
Dalam penelitian ini, ada tahapan-tahapan atau langkah-langkah yang harus dilakukan. Untuk lebih *detail*-nya terlihat pada gambar 1.



Gambar 1. Alur Penelitian

Crawling data

Langkah pertama yang dilakukan dalam proses analisis sentimen adalah pengumpulan data. Proses pengumpulan data dalam penelitian ini diambil dari media sosial twitter menggunakan API Twitter dengan pencarian kata kuliah *online*.



Gambar 2. Proses *Crawling data*

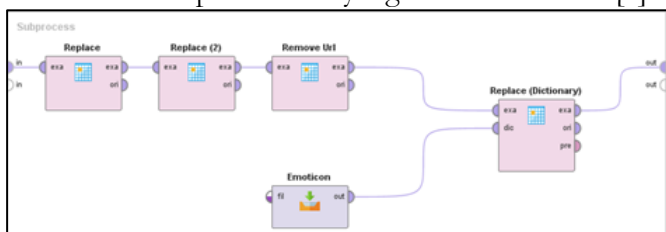
Pada gambar 2 menunjukkan proses *crawling data*. Ada beberapa *operator* yang digunakan untuk melakukan *crawling data* seperti "search twitter", "select atribut", "remove duplicate", dan "write excel". Search twitter digunakan untuk mengambil data dari Twitter menggunakan API Twitter, select atribut digunakan untuk memilih atribut yang ingin digunakan seperti text, id, username dan lain-lain. Operator remove duplicates dipakai untuk menghapus jika ada data yang sama, kemudian operator write excel dipakai untuk menyimpan data dalam format excel. Kemudian proses berikutnya melakukan proses *labeling data*. Pada proses ini data akan dibagi menjadi kelas negatif, positif. Hal ini dilakukan untuk menentukan *tweet* yang diambil apakah masuk kedalam kelas negatif atau positif, berdasarkan *data training* dan *testing*. Berikut adalah contoh pemberian label pada *dataset* twitter.

Tabel 1. Labelling Data

Sebelum	Sesudah
:(((, :(((, :-@, :- , :-p, :s, :x	Sedih
(^.), (^_.), (:, :), (-, :))))) :3, :9	Senang

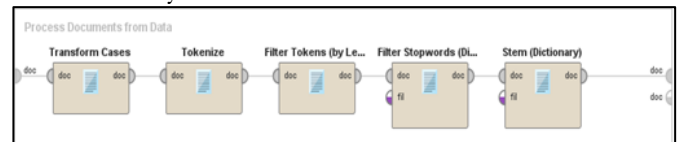
Preprocessing Data

Setelah melakukan *labeling data*, langkah selanjutnya adalah melakukan pra-pemrosesan data. Dalam proses ini yang tadinya data sudah siap kemudian akan dilakukan proses data yang lebih terstruktur [6].



Gambar 3. Operator dalam *Subprocess*

Gambar diatas menunjukkan isi *operator* yang ada pada *operator subprocess*. Dalam penelitian ini digunakan *operator replace*, untuk menghilangkan huruf @ dan RT. Kemudian *operator remove* Url, untuk menghilangkan halaman *web* yang ada pada *tweet*. *Operator Read* Excel (*Emoticon*), digunakan untuk mengkonversi kombinasi huruf menjadi kalimat tekstual. *Operator Replace* Atribut digunakan untuk mengganti atribut satu ke atribut satunya.



Gambar 4. Operator dalam *Precess Document Data*

Gambar diatas menampilkan beberapa *operator* yang ada didalam *operator process documents from data*. Berikut ini adalah penjelasannya.

1) *Cleansing Data*

Merupakan pembersihan data terhadap data yang tidak diperlukan [2,7]. Hal ini dilakukan untuk mengurangi kesalahan ketika memproses data. Dalam penelitian ini adapun yang dihilangkan yaitu huruf @, RT, dan URL.

2) *Emoticon*

Merupakan bentuk ekspresi digambarkan dalam bentuk wajah yang digunakan penggunaanya untuk mengekspresikan suasana hatinya [7,8]. Dalam hal ini dilakukan convert yang dimana ekspresi tadi dalam bentuk kombinasi huruf diubah ke kalimat yang lebih jelas. Contohnya ada pada tabel 2.

Tabel 2. *Convert Emoticon*

Sentiment	Text
Negatif	Hawa ngantuk semakin kuat dikuliah <i>online</i>
Negatif	Asli dah capean kuliah <i>online</i> . Cape mata cape otak
Positif	Good morning~~ semangat kuliah <i>online</i> nya <33
Positif	Semangat buat kamu yang hari ini sekolah daring,kuliah <i>online</i> ,kerja offline dan yg kerja <i>online</i> !??

3) *Case Folding*

Dalam *tweet* terdapat penulisan huruf yang berbeda, oleh karena itu dilakukan proses pengubahan huruf menjadi huruf besar maupun huruf kecil [6,7,8]. Contoh "Mahasiswa Kesulitan

Mengikuti Kuliah Daring” menjadi “mahasiswa kesulitan mengikuti kuliah daring”.

4) *Tokenize*

Merupakan proses pemecahan kata menjadi beberapa bagian kata [6,5]. Proses ini juga digunakan untuk menghilangkan tanda baca yang tidak diperlukan. Contoh ”Mahasiswa kesulitan mengikuti kuliah *online*” menjadi ” Mahasiswa, kesulitan, mengikuti, kuliah, *online*”.

5) *Filtering*

Merupakan proses penyaringan kata, yang dimana kata tersebut tidak memiliki arti. Contoh ”Mahasiswa kesulitan untuk mengikuti kuliah daring” menjadi “Mahasiswa kesulitan mengikuti kuliah daring” [9].

6) *Stemming*

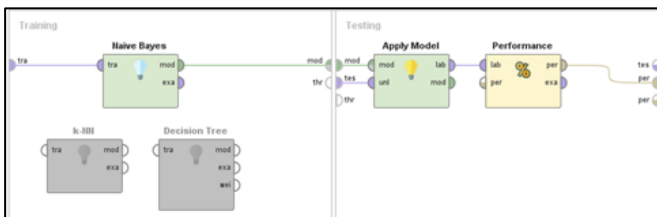
Merupakan pemotongan kata pada kata yang berimbuhan menjadi kata dasar. Contoh ”Mahasiswa kesulitan mengikuti kuliah daring” menjadi ”Mahasiswa sulit ikuti kuliah daring” [7].

Tabel 3. Contoh *Stemming*

Sebelum	Sesudah
Akademiknya	Akademik
Akunya	Aku
Biasanya	Biasa
Caranya	Cara
Selebihnya	Lebih

Klasifikasi Sentimen Analisis

Setelah melakukan proses Preprocessing Data Tahap selanjutnya melakukan analisis sentimen dengan tiga metode klasifikasi. Berikut adalah isi *operator* dari *Cross Validation*.



Gambar 5. *Operator* dalam *Cross Validation*

Untuk lebih jelasnya, akan dijelaskan dibawah ini:

Naive Bayes

Naive Bayes adalah algoritma klasifikasi yang berfungsi untuk mencari nilai probabilitas yang baik [10]. Rumusnya dapat dilihat dibawah ini.

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

K-NN

K-nearest neighbor adalah algoritma klasifikasi data yang berdasarkan *data training* dengan jarak yang paling dekat dengan objek. Rumus yang digunakan dalam perhitungan algoritma KNN adalah rumus Enclidean Distance [4].

$$d(x, y) = \sqrt{(\sum_{i=1}^n (xi - yi)^2)} \quad (2)$$

Decision Tree

Decision Tree adalah algoritma klasifikasi data yang menggunakan pohon keputusan. Algoritma Decision Tree memiliki dua input yaitu sample latih dan sample. Training samples merupakan contoh data dipakai dalam membangun *tree* yang telah diproses keasliannya, kemudian sample digunakan untuk klasifikasi data [11].

$$Entropy(A) = \sum_{i=1}^n -pi \log_2 pi \quad (3)$$

Confusion Matrix

Setelah melakukan klasifikasi analisis sentimen langkah berikutnya melakukan evaluasi dari hasil yang dicoba. Pada tahap ini akan dilakukan perhitungan dengan paramater Accuracy, *Precision*, dan Recall [2]. Berikut adalah tabel *Confusion Matrix* dan rumus perhitungannya dapat dilihat berikut ini

Tabel 4. *Confusion Matrix*

Prediction Class	Actual Class	
	Predic Negative	Predic Positive
Negative	TN	FP
Positive	FN	TP

Accuracy merupakan perhitungan semua data yang diklasifikasikan baik itu positif maupun negatif. Adapun cara menghitungnya dapat menggunakan rumus dibawah ini [8].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)} \times 100\% \quad (4)$$

Precision merupakan prediksi *rasio true positif* dibandingkan dengan perhitungan semua data hasil prediksi positif [8]. Adapun cara menghitungnya dapat menggunakan rumus dibawah ini [10].

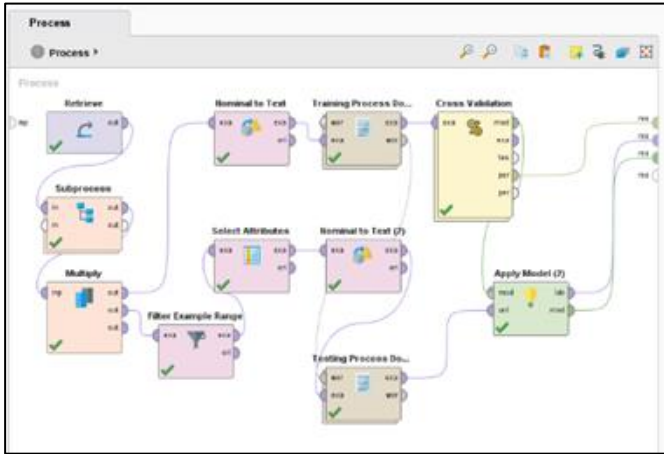
$$Precision = \frac{TP}{(TP + FN)} \times 100\% \quad (5)$$

Kemudian *Recall* merupakan prediksi rasio true positif dibandingkan perhitungan semua data yang true Positif [8]. Adapun cara menghitungnya dapat menggunakan rumus dibawah ini [10].

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \tag{6}$$

3. Hasil dan Pembahasan

Tahap ini merupakan tahap hasil dari uji analisis menggunakan software Rapid Miner.



Gambar 6. Proses Analisis Sentimen

Gambar 6 menampilkan *operator* proses analisis sentimen. Terdapat *operator retrieve* untuk membaca import data yang dilakukan sebelumnya, *operator Process Document* untuk melakukan *Preprocessing* Data, kemudian *operator Cross Validation* untuk melakukan klasifikasi.

Row No.	Sentimen	prediction(S...	confidence(...	confidenc...	text
211	Positive	Positive	0	1	pendidikan nasio...
212	Negative	Negative	1	0	kuliah online non...
213	Negative	Negative	1	0	semenjak kuliah ...
214	Negative	Negative	0.897	0.103	kaget kalo sampe...
215	Negative	Negative	1	0	kuntoajiw pengen...
216	Positive	Positive	0	1	prstyo godd mn t...
217	Positive	Positive	0	1	kuliah online ema...
218	Negative	Negative	1	0	tertipu udah sun...

Gambar 7. Hasil Prediksi

Gambar 7 menampilkan hasil prediksi menggunakan Rapid Miner dari label Sentimen dan Prediction Sentiment. Berikut ini menampilkan hasil tabel *Confusion Matrix* menggunakan Rapid Miner.

Tabel 5. Hasil Confusion Matrix

Metode	TP	FP	TN	FN
Naive Bayes	160	185	210	112
KNN	109	94	301	163
Decision Tree	31	13	382	241

Tabel diatas menampilkan hasil proses Confusion Matrix dari setiap metode. Untuk hasil *accuracy*, *precision* dan *recall*. Hasilnya dapat dilihat pada tabel 6.

Tabel 6. Hasil Accuracy, Precision dan Recall dengan Perhitungan Manual

Metode	Accuracy	Precision	Recall
Naive Bayes	55.47 %	46.37 %	58.82 %
KNN	61.47 %	53.69 %	40.07 %
Decision Tree	61.92 %	70.45 %	11.40 %

Tabel diatas menampilkan hasil perhitungan manual untuk menghitung *Accuracy*, *Precision*, dan *Recall*. Ada sedikit perbedaan nilai dari perhitungan manual dan menggunakan Rapid Miner, berikut hasil perhitungan menggunakan Rapid Miner.

Tabel 7. Hasil Accuracy, Precision dan Recall menggunakan Rapid Miner

Metode	Accuracy	Precision	Recall
Naive Bayes	55.49 %	46.45 %	58.91 %
KNN	61.47 %	53.90 %	40.07 %
Decision Tree	61.92 %	73.63 %	11.42 %

Hasil *accuracy* menggunakan ketiga metode tersebut masing-masing 55.49%, 61.47%, 61.92%. Kemudian hasil *precision* dari ketiga metode tersebut adalah 46.45%, 53.90%, 73.63%. Dan hasil *Recall* dari ketiga metode tersebut adalah 58.91%, 40.07%, 11.42%.

4. Kesimpulan

Penelitian ini melakukan klasifikasi analisis sentimen menggunakan tiga metode yaitu Naive Bayes, KNN, dan Decision Tree dalam menentukan persepsi mahasiswa tentang pemberlakuan kuliah secara daring dari data twitter. Dalam proses yang telah diuji didapat hasil yang berbeda dari ketiga metode tersebut. Hasil

Accuracy dari tiga metode tersebut adalah 55.49%, 61.47%, dan 61.92%. Hasil *Precision* dari tiga metode tersebut adalah 46.45%, 53.90%, dan 73.63%. Dan hasil *Recall* dari tiga metode tersebut adalah 58.91%, 40.07%, dan 11.42%. Dapat disimpulkan dari hasil tersebut bahwa metode Decision Tree memiliki tingkat *accuracy* yang paling tinggi dari kedua metode lainnya sebesar 61.92% *accuracy*, 73.63% *precision* dan 11.42% *recall*.

5. Daftar Pustaka

- [1] Fauziyyah, A.K., 2020. Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python. *Jurnal Ilmiah SINUS*, 18(2), pp.31-42, doi: 10.30646/sinus.v18i2.491.
- [2] Riefky, M. and Anandayani, A.R., 2020. Klasifikasi Persepsi Pengguna Twitter Terhadap Tuntutan Keringanan Pembayaran Uang Kuliah Tunggal (Ukt) Pada Masa Pandemi Covid-19 Menggunakan K-Nearest Neighbor. In *Seminar Nasional Official Statistics* (Vol. 2020, No. 1, pp. 247-257), doi: 10.34123/semnasoffstat.v2020i1.443.
- [3] Syarifuddin, M., 2020. Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes dan KNN. *INTI Nusa Mandiri*, 15(1), pp.23-28.
- [4] Romadloni, N.T., Santoso, I. and Budilaksono, S., 2019. Perbandingan Metode Naïve Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika*, 3(2), pp.1-9.
- [5] Kurniawan, S., Gata, W., Puspitawati, D.A., Tabrani, M. and Novel, K., 2019. Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), pp.176-183, doi: 10.29207/resti.v3i2.935.
- [6] 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.
- [7] Ruhjana, N., 2019. Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes. *IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika*, 3(1), pp.94-99.
- [8] Bayhaqy, A., Sfenrianto, S., Nainggolan, K. and Kaburuan, E.R., 2018, October. Sentiment analysis about E-commerce from tweets using decision tree, K-nearest neighbor, and naïve bayes. In *2018 international conference on orange technologies (ICOT)* (pp. 1-6). IEEE, doi: 10.1109/ICOT.2018.8705798.
- [9] Puspita, R. and Widodo, A., 2021. Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), pp.646-654, doi: 10.32493/informatika.v5i4.7622.
- [10] A. Muhammad., 2018. A Comparative Analysis of Classification Algorithms on Diverse Datasets, *Engineering, Technology & Appl. Sci. Res.*, vol. 8, no. 2, pp. 2790-2795, 2018, doi: 10.48084/etasr.1952.
- [11] Cynthia, E.P. and Ismanto, E., 2018. Metode Decision Tree Algoritma C. 45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 3, pp.1-13, doi: 10.30645/jurasik.v3i0.60.
- [12] Triayudi, A., Hidayatullah, D. and Rosalina, V., 2019. K-Nearest Neighbor Algorithm for Determination of High School Programs. *Journal Of Software Engineering & Intelligent Systems (JSEIS)*, 4.

- [13] Tineges, R., Triayudi, A. and Sholihati, I.D., 2020. Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), pp.650-658, doi: 10.30865/mib.v4i3.2181.
- [14] Samsir, S., Ambiyar, A., Verawardina, U., Edi, F. and Watrianthos, R., 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(1), pp.157-163, doi: 10.30865/mib.v5i1.2604.
- [15] Daeli, N.O.F. and Adiwijaya, A., 2020. Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. *Journal of Data Science and Its Applications*, 3(1), pp.1-7, doi: 10.34818/JDSA.2020.3.22.
- [16] Dharmawan, L.R., Arwani, I. and Ratnawati, D.E., 2020. Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, p.964X.