

**ANALISIS SENTIMEN KEPUASAN MAHASISWA TERHADAP
PEMBELAJARAN *ONLINE* SELAMA PANDEMI COVID-19 PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN PERBANDINGAN
ALGORITMA KLASIFIKASI**

***ANALYSIS OF STUDENT SATISFACTION
SENTIMENT TOWARDS ONLINE LEARNING
DURING THE COVID-19 PANDEMIC ON
SOCIAL MEDIA TWITTER USING
COMPARISON OF CLASSIFICATION
ALGORITHMS***

Ezza Ramadhanta Machmud Razaq¹, Deden Witarsyah Jacob², Faqih Hamami³

^{1,2,3} Universitas Telkom. Bandung

¹ezzaramadhanta@student.telkomuniversity.ac.id, ²dedenw@telkomuniversity.ac.id,

³faqihhamami@telkomuniversity.ac.id

Abstrak

Twitter merupakan platform media sosial yang digunakan banyak orang untuk mengekspresikan pikiran, minat dan pendapat terhadap banyak hal. Pendapat-pendapat tersebut tentunya memiliki nilai yang bervariasi, ada yang bernada positif, negatif, atau netral. Pandangan dan pendapat tersebut dapat dimanfaatkan sebagai sumber data untuk melakukan penilaian sentimen terhadap suatu topik pada Twitter. Saat ini dunia sedang dilanda pandemi *Covid-19*, suatu penyakit yang disebabkan dengan penyebaran virus dari manusia ke manusia sehingga pergerakan dan aktivitas manusia dibatasi, dengan tujuan untuk memutuskan rantai penyebaran virus *Covid-19* diantara manusia. Pembatasan aktivitas manusia mempengaruhi semua aspek kehidupan masyarakat termasuk dalam hal pendidikan. Kegiatan belajar mengajar berubah dari *offline* menjadi pembelajaran *online*, dimana interaksi dilakukan melalui layar dengan memanfaatkan teknologi. Regulasi pembelajaran *online* yang ditetapkan pemerintah merupakan hal baru yang dirasakan mayoritas pelajar di Indonesia, sehingga memunculkan beraneka opini dan pandangan yang diungkapkan para pelajar (mahasiswa) melalui media sosial twitter. Pelajar khususnya mahasiswa membagikan komentar mereka pada media sosial twitter tentang keberlangsungan pembelajaran *online* selama pandemi. Oleh sebab itu peneliti ingin melakukan analisis sentimen untuk mengetahui kecenderungan komentar para pelajar terhadap regulasi baru dalam pembelajaran selama pandemi di Indonesia, apakah positif, negatif, atau netral. Dalam melakukan analisis sentimen untuk mengetahui tanggapan pengguna twitter terhadap pembelajaran *online* selama pandemi, peneliti menggunakan algoritma klasifikasi yang memiliki tingkat akurasi tinggi, penulis akan membandingkan 3 algoritma klasifikasi yaitu, algoritma klasifikasi *Naïve Bayes*, *K-Nearest Neighbor* (K-NN) dan C4.5. Adapun tahapan yang dilakukan pada penelitian ini adalah dengan cara *preprocessing* data, *processing* data, klasifikasi, dan evaluasi. Model yang dihasilkan akan diuji dan dievaluasi dengan melihat nilai akurasi, presisi, *recall*, dan *F1-Measure*. Setelah membandingkan ketiga algoritma tersebut didapatkan akurasi tertinggi yaitu 60.80% menggunakan algoritma klasifikasi *K-Nearest Neighbor* (K-NN). Hasil akurasi yang didapat menggunakan ekstraksi fitur TF-IDF sebesar 60.80%, presisi (*specificity*) 40.73%, *recall* (*sensitifty*) 72%, dan *F1-Measure* 38.43%. Analisis sentimen yang didapat pada penelitian ini menunjukkan bahwa pengguna twitter di Indonesia lebih banyak memberikan komentar netral.

Kata kunci : twitter, covid-19, pembelajaran Online, analisis sentimen, klasifikasi

Abstract

Twitter is a social media platform that many people use to express thoughts, interests and opinions on many things. These opinions certainly have varying values, some are positive, negative, or neutral. These views and opinions can be used as a data source to assess sentiment on a topic on Twitter. Currently the world is being hit by the Covid-19 pandemic, a disease caused by the spread of viruses from human to human so that human movement and activities are limited, with the aim of breaking the chain of spread of the covid-19 virus among humans. Restrictions on human activities affect all aspects of people's lives including education. Teaching and learning activities change from offline to online learning, where interaction is done through the screen by utilizing technology. Online learning regulations set by the government is a new thing that is felt by the majority of students in Indonesia, thus giving rise to a variety of opinions and views expressed by students through social media tweeters.

Students in particular shared their comments on social media Twitter about the continuity of online learning during the pandemic. Therefore, researchers want to conduct sentiment analysis to find out the tendency of students' comments on new regulations in learning during the pandemic in Indonesia, whether positive, negative, or neutral.

In conducting sentiment analysis to determine twitter users' responses to online learning during the pandemic, researchers used classification algorithms that have a high level of accuracy, researchers compared 3 classification algorithms namely, classification algorithms Naïve Bayes, K-Nearest Neighbor (K-NN) and C4.5. The stages carried out in this study are by means of preprocessing data, data processing, classification, and evaluation. The resulting model is tested and evaluated by looking at accuracy, precision, recall, and F1-Measure values. After comparing the three algorithms obtained the highest accuracy that is 60.80% by using the classification algorithm K-Nearest Neighbor (K-NN). Accuracy results obtained using TF-IDF feature extraction of 60.80%, specificity of 40.73%, recall (sensitivity) of 72%, and F1-Measure of 38.43% Sentiment analysis obtained in this study shows that twitter users in Indonesia give more neutral comments.

Keywords: *twitter, covid-19, online learning, sentiment analysis, classification*

1. Pendahuluan

Twitter merupakan platform media sosial yang digunakan banyak orang untuk mengekspresikan pikiran, minat dan pendapat terhadap banyak hal. Pendapat-pendapat tersebut tentunya memiliki nilai yang bervariasi, ada yang bernada positif, negatif, atau netral. Saat ini dunia sedang dilanda pandemi *Covid-19*, suatu penyakit yang disebabkan dengan penyebaran virus dari manusia ke manusia sehingga pergerakan dan aktivitas manusia dibatasi, dengan tujuan untuk memutuskan rantai penyebaran virus *Covid-19* diantara manusia. Kegiatan belajar mengajar berubah dari *offline* menjadi pembelajaran *online*, dimana interaksi dilakukan melalui layar dengan memanfaatkan teknologi. Regulasi pembelajaran *online* yang ditetapkan pemerintah merupakan hal baru yang dirasakan mayoritas pelajar di Indonesia, sehingga memunculkan beraneka opini dan pandangan yang diungkapkan para pelajar (mahasiswa) melalui media sosial twitter. Oleh sebab itu peneliti ingin melakukan analisis sentimen untuk mengetahui kecenderungan komentar para pelajar terhadap regulasi baru dalam pembelajaran selama pandemi di Indonesia. Pada penelitian ini analisis sentimen yang dilakukan bertujuan untuk mendapatkan sebuah informasi mengenai respon mahasiswa terhadap pembelajaran online, yang mana informasi tersebut mungkin dapat membantu institusi pendidikan seperti universitas ataupun yang lainnya dalam menentukan sistem pembelajaran online seperti apa yang tepat di Indonesia. Penelitian ini juga dilakukan untuk mendapatkan keterkaitan data yang terekam di media sosial twitter untuk memprediksi ketepatan akurasi dari algoritma yang dipilih. Tahapan yang dilakukan pada penelitian ini adalah dengan cara *preprocessing* data, *processing* data, klasifikasi, dan evaluasi. Hasil akurasi dari algoritma yang dipilih akan digunakan untuk menyimpulkan algoritma klasifikasi yang memiliki tingkat akurasi tertinggi.

2. Dasar Teori /Material dan Metodologi/perancangan

2.1 Twitter

Twitter adalah salah satu platform yang memungkinkan penggunanya mengirim dan membaca pesan yang tidak lebih dari 280 karakter atau disebut sebagai tweet. Tetapi sebelum 7 November 2017 pesan twitter hanya bisa sampai 140 karakter saja hingga akhirnya ditambah menjadi 280 karakter [1].

2.2 Pembelajaran Online

Pembelajaran *online* adalah sistem belajar mengajar antara dosen dan mahasiswa yang dilakukan dengan tidak dengan bertatap muka, melainkan dengan jarak jauh, yang memungkinkan dosen dan mahasiswa berada pada tempat yang berbeda [2].

2.3 Text Mining

Text mining merupakan salah satu cara untuk mencari pola dalam data teks tidak terstruktur yang dilakukan secara otomatis oleh komputer untuk menyarikan informasi yang bermanfaat untuk tujuan tertentu [3].

2.4 Analisis Sentimen

Analisis sentimen adalah cabang penelitian dari data mining dan *text mining*, atau sebuah metode yang diterapkan untuk mengolah berbagai opini yang dipublikasikan oleh konsumen atau para pakar melalui berbagai media, mengenai suatu hal tertentu agar mendapatkan sebuah sentiment yang terkandung dalam sebuah opini [4].

2.5 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah bagian dari aplikasi *Artificial Intelligence* (AI) yang mempunyai kemampuan dalam memahami Bahasa manusia [5].

2.6 Google Colaboratory

Google Colaboratory adalah *tools* yang dirancang oleh *Google Internal Research* untuk membantu para peneliti mengolah data dalam keperluan belajar ataupun bereksperimen khususnya dalam bidang *Machine Learning* dengan menggunakan bahasa pemrograman *Python* [6].

2.7 Algoritma Naïve Bayes

Naïve Bayes merupakan metode klasifikasi berdasarkan *teorema bayes* [7]. *Teorema bayes* adalah perhitungan statistik dengan menghitung probabilitas kemiripan kasus lama yang ada dibasis kasus dengan kasus baru [8]. Persamaan dari *Naïve Bayes* sebagai berikut [9]:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

2.8 Algoritma K-Nearest Neighbor

K-Nearest Neighbor adalah algoritma yang diperuntukan untuk melakukan klasifikasi terhadap suatu objek [10]. Pada prinsip kerjanya K-NN melakukan klasifikasi terhadap objek berlandaskan data training yang menerapkan jarak terdekat atau kesamaan dari objek tersebut [11]. Rumus K-NN adalah sebagai berikut [12]:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (2)$$

2.9 Algoritma Decision tree (C.45)

C.45 merupakan salah satu algoritma yang digunakan untuk menghasilkan pohon keputusan (*decision tree*) [13]. *Decision tree* adalah metode klasifikasi dan prediksi yang terkenal, pada prinsipnya *decision tree* mencari solusi permasalahan dengan menjadikan kriteria sebagai *node* yang saling berhubungan membentuk seperti struktur pohon [14]. Ada beberapa algoritma dalam pembentukan pohon keputusan antara lain ID3, CART, dan C4.5 yang merupakan pengembangan dari algoritma ID3 [15].

2.10 Ekstrasi Fitur TF-IDF (Term Frequency – Inverse Document Frequency)

Pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*) merupakan metode yang diperuntukan untuk menentukan seberapa dalam keterhubungan kata (term) pada dokumen dengan memberikan nilai setiap kata [16]. Persamaan TF-IDF adalah sebagai berikut [17]:

$$\begin{aligned} idf_t &= \log \left(\frac{D}{df_t} \right) \\ W_{d,t} &= tf_{d,t} \times idf_{d,t} \end{aligned} \quad (3)$$

2.11 Ekstrasi Fitur TF (Term Frequency)

TF (*Term Frequency*) adalah cara pembobotan term (kata) yang paling sederhana, semakin bobot suatu term (TF tinggi) dalam dokumen, maka akan memberikan nilai kesesuaian yang semakin besar [18]. Berikut adalah rumus TF (*Term Frequency*) [19]:

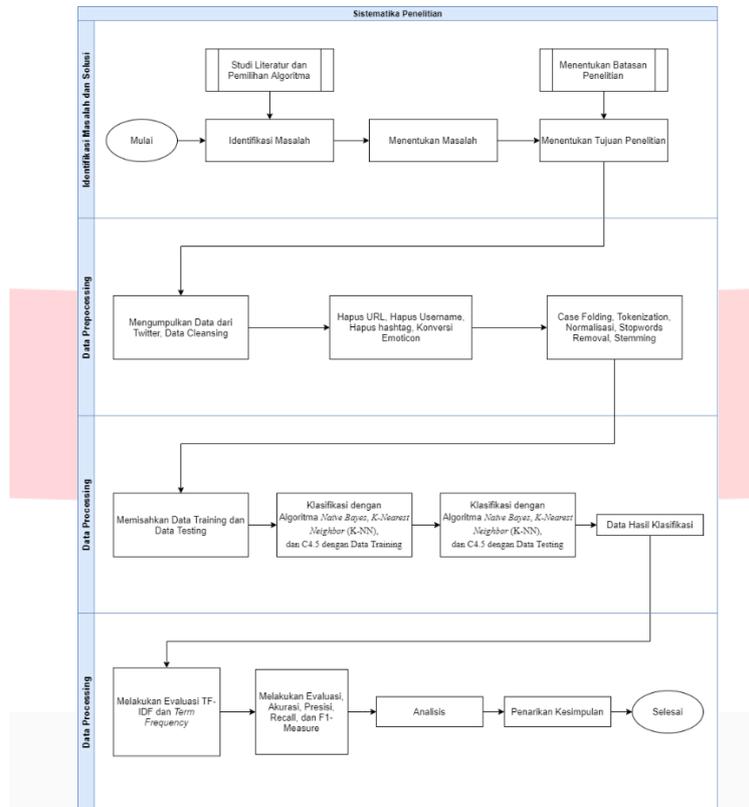
$$TF = \begin{cases} 1 + \log_{10}(f_{t,d}), f_{t,d} > 0 \\ 0, otherwise \end{cases} \quad (4)$$

3. Pembahasan

3.1. Cara Pengajuan dan Pengutipan

Pada penelitian ini, peneliti akan menggunakan metode klasifikasi. Metode ini akan menghasilkan

akurasi algoritma dari sentimen data relevan pada twitter terkait pembelajaran *online* selama *Covid-19* menggunakan tools *Google Colaboratory*. Pada tahap sistematika penelitian dilakukan beberapa tahap yaitu inialisasi, metode, dan evaluasi atau kesimpulan. Diagram sistematika yang telah dibuat adalah sebagai berikut.



Gambar 1. Sistematika Penelitian

3.2. Deskripsi Ruang Lingkup

Dalam penelitian ini akan dilakukannya analisis sentimen pada media sosial twitter. Objek yang ditentukan peneliti adalah pembelajaran *online* selama *Covid-19*. Pembelajaran *online* yang ditetapkan pemerintah merupakan hal baru yang dirasakan mayoritas pelajar di Indonesia, sehingga memunculkan beraneka opini dan pandangan yang diungkapkan para pelajar (mahasiswa) melalui media sosial twitter. Peneliti melakukan analisis sentimen pada twitter untuk mengetahui sentimen mahasiswa terhadap pembelajaran online selama *Covid-19* di Indonesia. Dalam analisis sentimen pada penelitian ini peneliti menggunakan tools *Google Colaboratory* dan membandingkan tiga algoritma klasifikasi yaitu *Naïve Bayes*, *K-Nearest Neighbor* (K-NN) dan C4.5. Penarikan data yang dilakukan tidak dapat memproses kata-kata tidak batu seperti singkatan atau bahasa gaul.

3.3. Dataset

Pada penelitian ini data yang dipakai adalah komentar para pengguna media sosial twitter di Indonesia yang diambil menggunakan program *crawling* dalam bahasa pemograman *python* dengan *library tweepy* yang memanfaatkan *API Twitter*. Data yang diambil menggunakan kata kunci yang terkait dengan pembelajaran *online*. Sebanyak 8850 komentar berhasil didapatkan. Selanjutnya adalah melakukan *preprocessing* data, untuk mendapatkan format data yang sesuai untuk tahap klasifikasi. Berikut adalah tahapan pertama dalam *preprocessing* data, yaitu hapus URL

Sebelum	Sesudah
Lulus S2 dari Harvard, Mikha Tambayong Ceritakan Susahnya Kuliah Online https://t.co/1HmE4chVXl https://t.co/plpQF40AI3	Lulus S2 dari Harvard, Mikha Tambayong Ceritakan Susahnya Kuliah Online

Selanjutnya dilakukan proses hapus username.

Sebelum	Sesudah
@moonareas terbebas dari beban kuliah online	terbebas dari beban kuliah online

Selanjutnya merupakan tahap penghapusan hashtags.

Sebelum	Sesudah
Sekolah / kuliah online kayak gini masih ada biaya uang gedung ga ya? #seriustanya	Sekolah / kuliah online kayak gini masih ada biaya uang gedung ga ya?

Tahapan berikutnya adalah case folding yang merupakan tahapan untuk mengubah semua huruf dalam suatu kalimat menjadi huruf kecil semua.

Sebelum	Sesudah
Walaupun aku lebih suka online karena hemat duit, tapi tetep suasana kuliah offline ga bisa digantikan dengan apapun	walaupun aku lebih suka online karena hemat duit, tapi tetep suasana kuliah offline ga bisa digantikan dengan apapun

Kemudian tahap normalisasi kata merupakan proses mengubah kata yang tidak baku menjadi kata-kata yang sesuai dengan kamus besar bahasa Indonesia (KBBI).

Sebelum	Sesudah
mau bgt kuliah lagi tp ga mau online	Mau banget kuliah lagi tetapi tidak mau online

Selanjutnya tokenization merupakan proses mengubah dokumen menjadi bagian-bagian seperti token

Sebelum	Sesudah
Aku benci kuliah online tetapi kalau ujian online aku suka	aku benci kuliah online tetapi kalau ujian online aku suka

Berikutnya adalah tahapan stopword removal merupakan tahap untuk menghilangkan kata-kata yang tidak mengandung makna atau tidak diperlukan.

Sebelum	Sesudah
aku benci kuliah online tetapi kalau ujian online aku suka	benci kuliah online ujian online suka

Proses yang terakhir adalah stemming yang mana proses ini untuk merubah kata-kata ke dalam bentuk dasarnya dengan cara menghapus imbuhan awal, akhir, atau keduanya.

Sebelum	Sesudah
demi kayanya gue	kayak butuh tutor

butuh	kuliah
tutor	tidak
kuliah	erti
deh	banget
tidak	online
ngerti	dosen
banget	tidak
online	erti
gini	
mana	
dosennya	
tidak	
pengertian	

Setelah itu data tersebut dipilih secara manual dan terkumpul 1548 data relevan, data tersebut akan dibagi menjadi data *training* dan data *testing* dengan perbandingan rasio 60:40 dimana 60% data *training* dan 40% data *testing*. Berikut adalah perbandingan dataset dengan rasio 60:40.

Tabel 1 Pembagian Data *Training* dan Data *Testing*

Jumlah Data	Data <i>Training</i>	Data <i>Testing</i>	Total
	928	620	1548

Kemudian setelah pembagian data, label yang diberikan pada sentimen dilakukan secara manual, dan akan dibandingkan dengan labeling dari algoritma klasifikasi. Hasil dari data *testing* ini akan dibandingkan dengan data *training*, untuk label yang berbeda antara sentimen dengan *prediction* nanti akan dihitung oleh sistem untuk menghasilkan nilai evaluasi performansi.

3.4. Evaluasi Performansi

Evaluasi performansi dilakukan untuk mendapatkan hasil dan klasifikasi algoritma yang sudah dipilih dengan ekstraksi fitur TF-IDF dan *Term Frequency*. Berikut adalah hasil perbandingan akurasi dari ketiga algoritma.

Algoritma	TF-IDF	TF
<i>Naïve Bayes</i>	59.35%	58.87%
K-NN	60.80%	59.67%
<i>Decision Tree (C4.5)</i>	57.90%	57.74%

Dapat diketahui algoritma K-NN memiliki tingkat akurasi tertinggi dibandingkan dengan algoritma *Naïve Bayes* dan *Decision Tree (C4.5)*. Berdasarkan tabel di atas, akurasi yang didapatkan oleh algoritma K-NN menggunakan ekstraksi fitur TF-IDF sebesar 60.80%.

Setelah itu ekstraksi fitur akan menghasilkan confusion matrix yang memuat hasil akurasi, presisi, *recall*, dan *F1-Measure*. Berikut merupakan hasil dari pengukuran presisi, *recall*, dan *F1-Measure* untuk ekstraksi fitur TF-IDf pada algoritma K-NN.

Label	Presisi	<i>Recall</i>	<i>F1-Measure</i>
<i>Positive</i>	0.2%	100%	0.3%
<i>Negative</i>	35%	53%	42%
<i>Neutral</i>	87%	63%	73%
Rata - rata	40.73%	72%	38.43%

Berdasarkan tabel di atas, hasil presisi, *recall*, *F1-Measure* pada algoritma *K-Nearest Neighbor* (K-NN) dan ekstraksi fitur TF-IDF diperoleh hasil presisi (*specificity*) sebesar 40.73%, *recall (sensitifity)* sebesar 72%, sehingga menghasilkan *F1-Measure* sebesar 38.43%. Dapat disimpulkan bahwa akurasi, presisi, *recall*, dan *F1-Measure* besarnya antara 38% sampai 72%.

4. Kesimpulan

Berdasarkan hasil yang didapat pada penelitian ini, dapat disimpulkan bahwa K-NN dengan menggunakan ekstraksi fitur TF-IDF memiliki tingkat akurasi paling tinggi sebesar 60.80% dibandingkan algoritma klasifikasi *Naïve Bayes* dan *Decision Tree* (C4.5). Algoritma K-NN mendapatkan hasil persentase dari klasifikasi sentimen positif sebesar 0.16%, sentimen negative sebesar 21,77%, dan sentimen netral sebesar 78.06%. Menunjukkan kepuasan mahasiswa terhadap pembelajaran *online* selama pandemi *Covid-19* terhadap sentimen komentar pengguna twitter adalah netral.

Referensi:

- [1] Hannani, N. (2019). *Pengertian Twitter Beserta Sejarah dan Manfaat Twitter yang Dibahas Secara Lengkap*. Wwww.Nesabamedia.Com.
- [2] Jatmoko, D., & Faizun, M. (2020). Persepsi Mahasiswa Terhadap Pelayanan Pembelajaran Online di Masa Pandemi Covid-19. *Jurnal Pendidikan Surya Edukasi (JPSE)*, 6(1), 104–113. <https://doi.org/10.37729/jpse.v6i1.6495>
- [3] Putra, M. P. R., & Wardani, K. R. N. (2020). Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 5(1), 63–71. <https://doi.org/10.32767/jutim.v5i1.791>
- [4] Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes , Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl. *Jurnal IKRA-ITH Informatika*, 3(2), 1–9.
- [5] Maitri, A. L., & Sutopo, J. (2019). RANCANG BANGUN CHATBOT SEBAGAI PUSAT INFORMASI LEMBAGA KURSUS DAN PELATIHAN MENGGUNAKAN PENDEKATAN NATURAL LANGUAGE PROCESSING. *Eprints.Uty.Ac.Id*, 1–9.
- [6] Digmi, I. (2018). *Google Colab Gratis Untuk Belajar Deep Learning*. <https://Imam.Digmi.Id/>
<https://imam.digmi.id/post/google-colab-gratis-untuk-belajar-deep-learning/>
- [7] Jeremy, A., M, V. C., & Mulyawan, B. (2017). *OPINION MINING UNTUK ULASAN PRODUK DENGAN KLASIFIKASI NAIVE BAYES*. 9–16.
- [8] Novianti, D. (2019). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner. *Paradigma - Jurnal Komputer Dan Informatika*, 21(1), 49–54. <https://doi.org/10.31294/p.v21i1.4979>
- [9] Sartika, D., & Indra, D. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2), 151–161.
- [10] Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31. <https://doi.org/10.21609/jsi.v13i1.500>
- [11] Siregar, R. R. A., Siregar, Z. U., & Arianto, R. (2019). Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. *Kilat*, 8(1), 81–92. <https://doi.org/10.33322/kilat.v8i1.421>
- [12] Asiyah, S. N. (2016). Klasifikasi Berita Online Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor. *Jurnal Sains Dan Seni ITS*, 5(2).
- [13] Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 134–138.
- [14] Sartika, D., & Indra, D. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2), 151–161.
- [15] Mutmainnah, S., Abdurrahman, G., Azizah, H., & Faruq, A. (2018). *Optimasi algoritma c4.5 menggunakan teknik bagging pada data kadar karat emas*.
- [16] Herwijayanti, B. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 306–312.
- [17] Wahyuni, R. T., Prastiyanto, D., & Suprpto, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro*, 9(1), 18–23. <https://doi.org/10.15294/jte.v9i1.10955>
- [18] Wijaya, A. P., & Santoso, H. A. (2016). Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. *Journal of Applied Intelligent System*, 1(1), 48–55.
- [19] INFORMATIKALOGI. (2017). *Pembobotan Kata atau Term Weighting TF-IDF*. <https://Informatikalogi.Com/>.