

Implementasi Algoritma *k-Nearest Neighbor* (k-NN) pada Data Ulasan Pelaksanaan Pembelajaran Daring

Angge Firizkiansah¹, Ali Muhammad², Dita Setiawan³
^{1,2,3}Universitas Sains Indonesia, Kabupaten Bekasi

E-mail Korespondensi: angge.firizkiansah@lecturer.sains.ac.id

Abstrak

Perkembangan teknologi informasi memberikan dampak dalam proses pembelajaran di perguruan tinggi, dimana pelaksanaannya dapat dilakukan secara daring. Pembelajaran daring yang dilakukan harus dievaluasi untuk menjaga mutu pembelajarannya. Di salah satu perguruan tinggi swasta di Indonesia, telah melakukan proses evaluasi terhadap pelaksanaan pembelajaran daring, yaitu melalui ulasan pelaksanaan pembelajaran daring oleh mahasiswa. Berdasarkan data ulasan dari mahasiswa, dilakukan analisis untuk menggambarkan setiap ulasan merupakan saran atau masukan terhadap pelaksanaan pembelajaran daring yang diklasifikasikan dari sisi akademik maupun nonakademik. Pengklasifikasian ulasan tersebut dapat dilakukan menggunakan pendekatan *machine learning*. Pada penelitian ini, dilakukan pengukuran tingkat akurasi terhadap pengklasifikasian ulasan pelaksanaan pembelajaran daring menggunakan algoritma *k-Nearest Neighbor* (k-NN). data yang digunakan berupa data teks ulasan pelaksanaan pembelajaran daring dari suatu perguruan tinggi dengan jumlah data sebanyak 4320 ulasan. Hasil uji menggunakan algoritma k-NN untuk pengkategorian data teks ulasan pelaksanaan pembelajaran daring diperoleh nilai tertinggi pada $k=2$ dengan *accuracy* 87,39%, *precision* 84,17%, *recall* 94,07%, dan AUC 0,843. Hal tersebut dapat disimpulkan bahwa algoritma k-NN cukup baik dalam mengkategorikan data teks ulasan pelaksanaan pembelajaran daring.

Kata kunci: *k-nearest neighbor*; klasifikasi; *machine learning*; pembelajaran daring.

Abstract

Information technology in this era has an impact on the learning process in higher education, where its implementation can be done online. One private university in Indonesia has conducted an evaluation of the implementation of online learning through student reviews. An analysis was carried out to describe each review as a suggestion or input for the implementation of online learning, which was classified from the academic and non-academic sides. The classification of these reviews can be done using a machine learning approach. In this study, the accuracy level of the classification of online learning implementation reviews was measured using the k-Nearest Neighbor algorithm. The data used is in the form of text data on reviews of the implementation of online learning from a university with a total of 4320 reviews. The test results using the k-NN algorithm for the categorization of text data on reviews of the implementation of online learning obtained the highest value at $k=2$ with an accuracy of 87.39%, precision of 84.17%, recall of 94.07%, and AUC of 0.843. It can be concluded that the k-NN algorithm is quite good at categorizing text data on reviews of the implementation of online learning.

Keywords: *k-nearest neighbor*; classification; *machine learning*; online learning.

1. PENDAHULUAN

Teknologi informasi semakin berkembang dengan pesat. Dampak perkembangan tersebut dirasakan di berbagai sektor, tidak terkecuali pada sektor pendidikan. Pelaksanaan proses pembelajaran saat ini dapat dilakukan secara daring. Pembelajaran daring merupakan proses pembelajaran yang dilakukan dengan memanfaatkan jaringan internet, antara dosen dan mahasiswa secara tidak langsung [1]. Pelaksanaan pembelajaran daring sangat dirasakan dampaknya pada masa pandemi COVID-19. Kebutuhan pelaksanaan pembelajaran yang didukung dengan teknologi informasi menjadi semakin tinggi. Hal tersebut terus berlanjut hingga saat ini, meskipun pandemi COVID-19 sudah berakhir.

Pembelajaran secara daring dilakukan dengan memanfaatkan berbagai platform dan aplikasi seperti *WhatsApp group*, *video call*, aplikasi *zoom conference*, dan lainnya [1][2]. Penggunaan aplikasi pembelajaran daring ini menjadi komponen utama dalam kegiatan pembelajaran daring, yang bertujuan agar proses belajar mengajar dapat terlaksana dengan baik [2]. Dalam implementasinya, pelaksanaan pembelajaran daring mengalami beberapa tantangan, seperti rendahnya kualitas koneksi internet, terbatasnya perangkat teknologi yang dimiliki oleh peserta didik, serta kurangnya kemampuan pengajar dalam memanfaatkan teknologi informasi [3]. Dalam rangka perbaikan dan pengembangan pembelajaran daring, diperlukan evaluasi dan analisis terhadap pelaksanaan pembelajaran daring tersebut. Evaluasi pelaksanaan pembelajaran daring, salah satunya dilakukan melalui penyampaian ulasan proses pembelajaran daring dari mahasiswa.

Salah satu perguruan tinggi swasta di Indonesia telah menerapkan pembelajaran daring sebagai bagian dari implementasi proses pembelajaran yang berlangsung sejak pandemi COVID-19. Pelaksanaan pembelajaran daring yang telah dilakukan perlu dilakukan evaluasi secara berkala untuk mengetahui Dalam rangka mengetahui efektivitas proses pembelajaran secara daring, dilakukan evaluasi dengan

mengumpulkan data berupa ulasan atau *review* dari mahasiswa terkait dengan pelaksanaan pembelajaran daring yang dilakukan. Data ulasan tersebut dibagi dalam dua kategori, yaitu akademik dan nonakademik. Hal tersebut guna mengetahui ulasan yang dimaksud untuk perbaikan dari sisi akademik atau nonakademik. Ulasan pada kategori akademik merupakan ulasan yang berkaitan dengan materi pembelajaran, metode pembelajaran, model pembelajaran, bahan ajar, hingga pelaksanaan proses pembelajaran itu sendiri. Sedangkan ulasan pada kategori nonakademik berkaitan dengan jaringan internet, paket data/kuota internet, media pembelajaran, sarana dan prasarana yang mendukung proses pembelajaran secara daring.

Berdasarkan data tersebut, dapat dilakukan klasifikasi kategori terhadap ulasan pelaksanaan pembelajaran daring dengan bantuan komputer melalui penerapan *machine learning*. *Machine learning* merupakan ilmu dan seni dari pemrograman komputer yang membuat program tersebut dapat melakukan pembelajaran dari data tanpa harus diprogram secara eksplisit. Pendekatan *machine learning* dipilih karena memiliki kelebihan diantaranya *machine learning* dapat digunakan untuk menemukan pengetahuan/solusi yang sulit diketahui dengan pendekatan konvensional, sistem *machine learning* juga dapat menangani data yang besar dan dapat beradaptasi terhadap data baru [4]. Penggunaan model *machine learning* yang tepat diharapkan dapat mempermudah pengguna dalam mengkategorisasikan ulasan pelaksanaan pembelajaran daring secara otomatis. Model *machine learning* dalam penelitian ini menggunakan algoritma *k-Nearest Neighbor* (k-NN).

Berdasarkan permasalahan tersebut, Penelitian ini dilakukan dengan tujuan untuk mengetahui seberapa baik tingkat keakuratan model *machine learning* menggunakan algoritma k-NN dalam mengklasifikasikan ulasan pelaksanaan pembelajaran daring kedalam kategori akademik atau nonakademik.

Pada beberapa penelitian sebelumnya yang sejenis, diperoleh hasil diantaranya, Mara A, dkk menjelaskan bahwa penggunaan algoritma k-NN pada data teks untuk melakukan analisis sentimen terhadap opini metode pembelajaran daring sebanyak 200 dokumen diperoleh tingkat akurasi 87% [5]. A Onan menjelaskan bahwa penggunaan algoritma k-NN dalam melakukan analisis sentimen terhadap *review* evaluasi penggunaan MOOC menggunakan data sebanyak 93000 *review* diperoleh tingkat akurasi sebesar 76,11% [6]. Penelitian berikutnya yang dilakukan oleh Muhammad Akmal Hakim, dkk diketahui bahwa implemenasi algoritma k-NN pada data pendapat pakar AI tentang kemajuan kecerdasan buatan menghasilkan tingkat akurasi sebesar 70,82% [7]. Penelitian yang dilakukan oleh Egi Puji Sutrisno dan Safrina Amini dalam melakukan analisis sentimen menggunakan algoritma k-NN pada data ulasan pengguna aplikasi Digital Korlantas POLRI sebanyak 600 data, diperoleh tingkat akurasi sebesar 93,26% [8]. Dewi Sartika mengemukakan hasil implementasi algoritma k-NN dalam analisis sentimen terhadap tanggapan masyarakat tentang program Merdeka Belajar Kampus Merdeka (MBKM) di twitter diperoleh tingkat akurasi sebesar 95% [9]. Selanjutnya penelitian yang disampaikan oleh Wiyli Yustanti dan Naim Rochmawati dalam menggunakan algoritma k-NN untuk mengklasifikasikan karakteristik mahasiswa pada pembelajaran daring dengan menggunakan seleksi fitur diperoleh tingkat ketepatan sebesar 84,8% (sebelum diseleksi fitur) dan 88,8% (setelah diseleksi fitur) [10]. Berdasarkan beberapa penelitian sebelumnya, dapat diketahui bahwa algoritma k-NN merupakan algoritma *machine learning* yang cukup baik dalam melakukan klasifikasi pada data teks. Diharapkan implementasi algoritma k-NN pada data teks berupa ulasan pelaksanaan pembelajaran daring diperoleh tingkat ketepatan yang tinggi, sehingga dapat digunakan dalam membantu mengkategorikan ulasan pelaksanaan pembelajaran daring kedalam kategori akademik maupun nonakademik secara otomatis.

2. METODE

Penelitian ini dilakukan dengan menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) dan menggunakan alat bantu analisis dan pemodelan *machine learning*, yaitu AI Studio 2024. KDD memiliki tahapan yang terdiri dari *data selection*, *data preprocessing*, *data transformation*, *data mining*, *interpretation/evaluation*, dan *knowledge* [11]. Sedangkan AI Studio 2024 merupakan salah satu perangkat lunak yang digunakan dalam proses *data mining*. Fungsi yang dapat dilakukan menggunakan perangkat lunak ini diantaranya pemrosesan data, visualisasi data, analisis prediktif, pemodelan statistik, evaluasi, dan penyebaran.

Berdasarkan hal tersebut, tahapan yang dilakukan dalam penelitian ini dimulai dari tahap persiapan dan seleksi data. Tahap ini merupakan langkah awal yang dilakukan guna mengetahui data yang akan digunakan dalam pemodelan *machine learning*. Pada penelitian ini, digunakan data ulasan pelaksanaan pembelajaran daring selama satu tahun dari suatu perguruan tinggi swasta di Indonesia. Data ulasan tersebut berjumlah 4320 ulasan dari mahasiswa. Data tersebut memiliki atribut No (nomor urut data), Fakultas (nama pengelola program studi tempat mahasiswa melakukan proses pembelajaran), Prodi (Nama Program Studi tempat mahasiswa melakukan proses pembelajaran), NPM (Nomor Pokok Mahasiswa), Saran dan Masukan (data teks yang menerangkan kalimat ulasan dari mahasiswa), dan Kategori (klasifikasi data ulasan termasuk dalam kategori akademik/nonakademik). Kemudian data tersebut dilakukan seleksi untuk menentukan atribut yang akan digunakan sebagai data latih dan data *testing* untuk dimodelkan dengan *machine learning* menggunakan algoritma k-NN.

Data yang telah diseleksi untuk digunakan sebagai data latih dan data testing, kemudian diolah untuk bisa dimodelkan menggunakan algoritma *machine learning*. Pemrosesan data ini dilakukan pembersihan data dan transformasi data menggunakan AI Studio 2024. Tahap

pertama dilakukan pembersihan data melalui penghapusan duplikasi data. Setelah data dibersihkan, selanjutnya dilakukan *tokenizing (regular expression)*, *tokenizing (non letter)*, *transform lower case*, *filter stopwords removal*, *stemming*, dan *filter tokens (by length)*. *Tokenizing (regular expression)* digunakan untuk menghapus tanda baca dan karakter spesial yang terdapat dalam setiap kalimat saran pada dataset. Sedangkan *tokenizing (non letter)* digunakan untuk menghilangkan bagian dari kalimat yang bukan huruf. Kemudian proses tokenisasi tersebut mengubah struktur setiap kalimat ulasan menjadi per kata yang disebut sebagai token. Selanjutnya dilakukan proses *transform lower case* untuk menyeragamkan struktur setiap token menjadi *lower case*. Berikutnya dilakukan proses *filter stopwords removal* yang berfungsi untuk menghilangkan kata-kata yang bersifat umum dan/atau tidak memiliki makna seperti kata hubung.

Selanjutnya dilakukan *stemming*. Proses *stemming* merupakan proses mengubah kata-kata yang berimbuhan menjadi kata dasarnya. Setelah dilakukan *stemming*, proses berikutnya adalah *filter tokens (by length)* yang berfungsi untuk mengambil kata yang memiliki panjang karakter 3 – 25. Sehingga kata-kata yang memiliki panjang karakter diluar dari ketentuan di atas tidak dimasukkan ke dalam *corpus* data latih dan *testing* nantinya.

Tahap selanjutnya setelah *corpus* data diperoleh adalah melakukan proses *data mining*. *Data mining* merupakan salah satu cabang ilmu yang berkaitan dengan bagaimana mengekstraksi suatu informasi/pengetahuan yang bermanfaat dari suatu basisdata yang besar [12]. *Data mining* juga dapat diartikan sebagai proses mengekstrak data yang sebelumnya tidak diketahui manfaatnya menjadi pengetahuan/informasi/pola dari data yang jumlahnya besar [13]. Secara umum *data mining* terbagi menjadi klasifikasi, klastering, asosiasi, estimasi, dan prediksi [14]. Proses *data mining* dilakukan dengan pendekatan algoritma *machine learning*.

Pada penelitian ini algoritma *machine learning* yang digunakan, yaitu algoritma k-NN.

Algoritma k-NN merupakan algoritma sederhana yang disebut sebagai *lazy-learner*. Hal tersebut dikarenakan algoritma k-NN bekerja dengan menyimpan dan mengolah semua data sebagai data latih tanpa dilakukan proses pengolahan ke dalam suatu format tertentu. Algoritma ini bekerja dengan mencari similaritas/kemiripan dari suatu data yang akan diklasifikasikan berdasarkan data latih yang telah disimpan. Similaritas tersebut diperoleh dengan cara menghitung jarak antara data latih dan data *testing* dengan asumsi nilai k yang menggambarkan jumlah tetangga data terdekat dari data yang akan diklasifikasikan [13]. Jarak tersebut secara umum dihitung menggunakan persamaan *euclidian distance*. Berikut persamaan untuk menghitung tetangga terdekat dalam algoritma k-NN.

$$d(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \quad (1)$$

Langkah-langkah yang dilakukan dalam algoritma k-NN, yaitu (1) Menentukan nilai k secara manual; (2) Hitung jarak antara data latih dan data testing; (3) Mengurutkan data latih berdasarkan jarak terdekat; (4) Menetapkan kelas dengan cara memilih kelas yang memiliki jumlah nilai k terbanyak pada data *testing*.

Setelah model k-NN diperoleh, hasil pemodelan tersebut dievaluasi untuk mengetahui tingkat ketepatan klasifikasi yang dilakukan pada data ulasan pelaksanaan pembelajaran daring. Tingkat ketepatan model k-NN yang telah diterapkan diukur menggunakan *confusion matrix* dan nilai *Area Under Curve (AUC)* dari kurva ROC yang didukung dengan *10-fold cross validation*. *Confusion matrix* digambarkan dalam bentuk tabulasi yang berisikan nilai kelas prediksi dan kelas aktual [9].

Tabel 1. Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

- a. *True Positive (TP)*: Prediksi klasifikasi pada label positif sesuai aktualnya, yaitu positif.

- b. *True Negative* (TN): Prediksi klasifikasi pada label negatif sesuai aktualnya, yaitu negatif.
- c. *False Positive* (FP): Prediksi klasifikasi pada label bernilai positif tidak sesuai aktualnya, yaitu bernilai negatif.
- d. *False Negative* (FN): Prediksi klasifikasi pada label bernilai negatif tidak sesuai aktualnya.

Confusion matrix digunakan untuk mengetahui nilai *accuracy*, *precision*, dan *recall*. *Accuracy* merupakan jumlah data yang diklasifikasikan benar dibandingkan jumlah keseluruhan data. *Precision* berfungsi untuk mengetahui kemampuan sistem dalam menemukan peringkat yang relevan. Sedangkan *recall* berfungsi untuk mengetahui kemampuan sistem dalam menemukan seluruh item yang relevan dari koleksi dokumen [9][15].

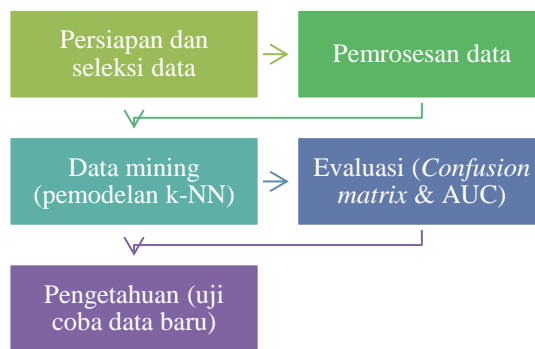
$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Nilai AUC diperoleh dari hasil kurva ROC yang menggambarkan grafik dua dimensi dari nilai *True Positive Rate* (TPR) pada sumbu Y dan *False Positive Rate* (FPR) pada sumbu X [23]. Hasil evaluasi menggunakan nilai AUC terbagi menjadi lima tingkat, yaitu (1) *Excellent Classification* untuk nilai AUC antara 0,90 – 1,00; (2) *Good Classification* untuk nilai AUC antara 0,80 – 0,90; (3) *Fair Classification* untuk nilai AUC antara 0,70 – 0,80; (4) *Poor Classification* untuk nilai AUC antara 0,60 – 0,70; (5) *Failure Classification* untuk nilai AUC di bawah 0,60 [16].

Selanjutnya ditentukan model terbaik yang dapat digunakan untuk melakukan uji coba menggunakan data baru yang belum memiliki kategori. Uji coba dilakukan untuk mengetahui apakah model *machine learning* yang digunakan dapat mengklasifikasikan kategori ulasan pelaksanaan pembelajaran daring dengan baik.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

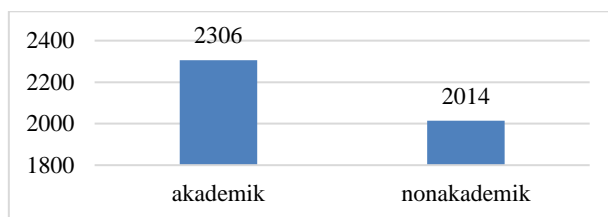
3.1. Persiapan dan Seleksi Data

Berdasarkan dataset ulasan pelaksanaan pembelajaran daring sebanyak 4320 ulasan dari mahasiswa, dilakukan seleksi data yang akan digunakan sebagai data latih dan data *testing* yang akan digunakan oleh model *machine learning* untuk melakukan klasifikasi menggunakan algoritma k-NN. Berdasarkan atribut dataset tersebut, data atribut Saran dan Masukan dijadikan sebagai data input dan atribut Kategori sebagai label atau target klasifikasi berdasarkan data input yang ditetapkan. Atribut Saran dan Masukan dipilih sebagai data input karena data tersebut merupakan data teks yang dapat dikelompokkan kedalam kategori akademik maupun nonakademik.

Saran dan Masukan	Kategori
Dapat subsidi kuota yang telah dibilangin	nonakademik
jangan dadakan kalo share link zoom atau google meet	nonakademik
Semoga ada solusi dari permasalahan koneksi internet yang tidak baik	nonakademik
Tugas lebih di minimalisir dan memberikan waktu yang lebih panjang untuk pengerjaan tugas	akademik

Gambar 2. Atribut Data yang digunakan

Berdasarkan data yang telah diseleksi, diketahui bahwa jumlah ulasan pada kategori akademik sebanyak 2306 ulasan dan kategori nonakademik sebanyak 2014 ulasan.



Gambar 3. Jumlah Data per Kategori

3.2. Pemrosesan Data

Langkah selanjutnya dilakukan pemrosesan data menggunakan AI Studio 2024. Tahap pertama dilakukan penghapusan duplikasi data untuk mengurangi redundansi dan menjaga integritas dari data itu sendiri. Data ulasan pelaksanaan pembelajaran daring yang semula sebanyak 4320 ulasan, setelah dibersihkan menjadi sebanyak 4297 ulasan. Selanjutnya setiap data kalimat ulasan dilakukan transformasi *lower case* dengan tujuan untuk menyeragamkan struktur kalimat menjadi huruf kecil. Hal tersebut dilakukan menggunakan fitur *transform lower case* seperti yang ditampilkan pada tabel 2.

Tabel 2. Transform Lower Case

Data Awal	Hasil Proses
Dapat subsidi kuota yang telah dibilangin	dapat subsidi kuota yang telah dibilangin
Semoga ada solusi dari permasalahan koneksi internet yang tidak baik	semoga ada solusi dari permasalahan koneksi internet yang tidak baik
Untuk meet jangan terlalu sering, karena sinyal kurang bagus	untuk meet jangan terlalu sering, karena sinyal kurang bagus

Setelah kalimat diubah dengan struktur seluruh kata menggunakan huruf kecil, dilakukan pemecahan data teks menjadi token menggunakan fitur *tokenizing regular expression* untuk menghapus tanda baca dan karakter spesial pada setiap kalimat dan *tokenizing non-letter* untuk menghilangkan bagian dari kalimat yang bukan huruf. Hasil dari proses ini seperti pada tabel 3.

Tabel 3. Tokenizing (Regular Expression & Non-Letter)

Data Awal	Hasil Proses
Dapat subsidi kuota yang telah dibilangin	dapat subsidi kuota yang telah dibilangin

Langkah pemrosesan data selanjutnya adalah menggunakan fitur *filter stopwords removal*. Fitur tersebut berfungsi untuk menghilangkan kata-kata yang bersifat umum dan/atau tidak memiliki makna seperti kata hubung. Hasil proses tersebut dapat dilihat seperti tabel 4.

Tabel 4. Filter Stopwords Removal

Data Awal	Hasil Proses
dapat subsidi kuota yang telah dibilangin	subsidi kuota dibilangin
semoga ada solusi dari permasalahan koneksi internet yang tidak baik	semoga solusi permasalahan koneksi internet

Selanjutnya dilakukan *stemming* untuk mengubah setiap kata kembali ke kata dasarnya. Hasil proses *stemming* seperti dicontohkan pada tabel 5.

Tabel 5. Stemming

Data Awal	Hasil Proses
subsidi kuota dibilangin	subsidi kuota bilang
semoga permasalahan internet	semoga solusi masalah koneksi internet

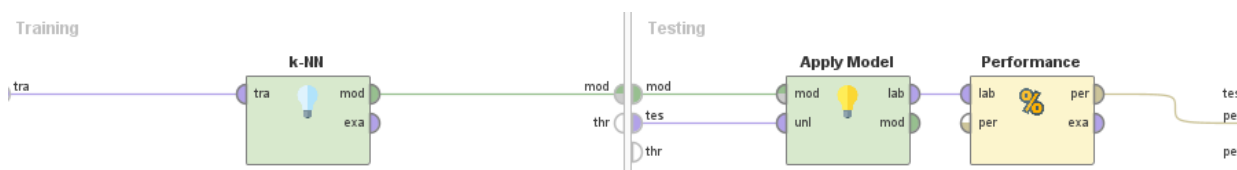
3.3. Data Mining

Data yang telah selesai ditransformasikan, selanjutnya dilakukan pembobotan menggunakan metode *Term Frequency – Invers Document Frequency* (TF-IDF). Metode tersebut merupakan metode untuk memberikan bobot pada suatu kata (*term*) pada setiap dokumen yang menjadi data latih dan data *testing*. Hal tersebut guna mengetahui tingkat kepentingan dari suatu kata dalam dokumen. Contoh hasil pembobotan kata menggunakan metode TF-IDF dapat dilihat pada tabel 6.

Tabel 6. Pembobotan TF-IDF

Dokumen	Term	Skor TF-IDF
Subsidi kuota bilang	bilang	0,929
	kuota	0,205
	subsidi	0,307
tugas minimal kerja tugas	kerja	0,581
	minimal	0,590
	Tugas	0,561

Berdasarkan hasil pembobotan yang telah dilakukan terhadap seluruh data ulasan pelaksanaan pembelajaran daring, dilakukan pemodelan *machine learning* menggunakan algoritma k-NN dengan bantuan AI Studio 2024. Implementasi algoritma k-NN digunakan pada data latih dan data *testing*. Pembagian data latih dan data *testing* menggunakan *10-Fold Cross Validation*.



Gambar 4. Model *Machine Learning* Algoritma k-NN pada AI Studio 2024

Metode ini membagi data menjadi 10 subset yang sama rata, dimana 9 subset dijadikan sebagai data latih dan 1 subset dijadikan sebagai data *testing*. Selanjutnya data latih dan data *testing* tersebut diuji menggunakan model *machine learning* dengan algoritma k-NN untuk nilai $k=1$ sampai dengan $k=10$ dalam mengklasifikasikan kategori ulasan pelaksanaan pembelajaran daring.

3.4. Interpretasi/Evaluasi

Berdasarkan pengujian menggunakan model *machine learning* dengan algoritma k-NN, dilakukan evaluasi tingkat ketepatan model tersebut menggunakan *confussion matrix* dan kurva ROC. Berdasarkan hasil evaluasi diperoleh nilai *accuracy*, *precision*, *recall*, dan AUC seperti pada tabel 7.

Tabel 7. Hasil Evaluasi Pengujian Model k-NN

Nilai k	Akurasi	Precision	Recall	AUC
1	87,36%	84,02%	94,29%	0,500
2	87,39%	84,17%	94,07%	0,843
3	85,41%	80,71%	95,51%	0,839
4	85,55%	80,69%	95,90%	0,890
5	83,50%	78,10%	96,03%	0,886
6	83,76%	78,27%	96,34%	0,875
7	82,01%	76,18%	96,51%	0,858
8	82,13%	76,20%	96,78%	0,886
9	80,36%	74,36%	96,51%	0,867
10	80,43%	74,35%	96,73%	0,914

Tabel 7 menjelaskan tingkat ketepatan pengklasifikasian kategori ulasan pelaksanaan pembelajaran daring. Berdasarkan hasil evaluasi tersebut, dapat diketahui bahwa nilai *accuracy* terbaik diperoleh pada model k-NN dengan $k=2$ sebesar 87,39%, nilai *precision* pada $k=2$ sebesar 84,17%, nilai *recall* pada $k=8$ sebesar 96,78%, dan nilai AUC pada $k=10$ sebesar 0,914.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa tingkat

ketepatan model k-NN untuk melakukan klasifikasi pada data teks berupa ulasan pelaksanaan pembelajaran daring dapat dinyatakan cukup baik menggunakan model k-NN dengan $k=2$, mengingat nilai *accuracy*, *precision*, dan *recall* rata-rata di atas 80% dan nilai AUC 0,843 yang masuk ke dalam *Good Classification*.

Namun hasil penelitian tersebut tentunya masih dapat ditingkatkan, seperti melakukan optimasi pemrosesan data untuk menyeimbangkan jumlah data. Hal tersebut dapat dilihat dari kemungkinan jumlah data ulasan pelaksanaan pembelajaran daring per kategori (akademik/nonakademik) bersifat *imbalance*/tidak seimbang. Pengembangan penelitian berikutnya juga bisa menggunakan data ulasan pelaksanaan pembelajaran daring ini untuk dimodelkan menggunakan algoritma *machine learning* lainnya untuk menghasilkan tingkat ketepatan klasifikasi yang lebih baik.

5. DAFTAR PUSTAKA

- [1] I A. Sadikin and A. Hamidah, "Pembelajaran Daring di Tengah Wabah Covid-19," BIODIK, vol. 6, no. 2, pp. 214–224, Jun. 2020, doi: 10.22437/bio.v6i2.9759.
- [2] F. Amri, "Persepsi Siswa tentang Aplikasi Teknologi yang Digunakan dalam Pembelajaran Online," EDUKATIF: JURNAL ILMU PENDIDIKAN, vol. 4, no. 1, pp. 250–258, Dec. 2021, doi: 10.31004/edukatif.v4i1.1216.
- [3] R. Handayani, S. Saharudin, A. M. Z. Mustopa, and S. Trikandi, "Teachers' Experiences in Evaluating Learning Process: A Case Study of Online Learning," *Jurnal Ilmiah Universitas Batanghari Jambi*, vol. 22, no. 3, p. 2061, Oct. 2022, doi: 10.33087/jiubj.v22i3.2802.

- [4] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd Edition., vol. Sebastopol: O'Reilly Media, Inc., 2019.
- [5] A. A. P. T. Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba."
- [6] A. Onan, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 572–589, May 2021, doi: 10.1002/cae.22253.
- [7] M. Akmal Hakim, P. Gunawan, R. Pratama, and F. Kurniawam, "Implementasi Algoritma K-Nearest Neighbors untuk Menganalisis Pendapat Pakar AI tentang Kemajuan Kecerdasan Buatan," 2024. [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [8] E. P. Sutrisno and S. Amini, "3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023- Jakarta," 2023.
- [9] D. Sartika, "Implementasi Algoritma K-Nearest Neighbour dalam Menganalisis Sentiment Terhadap Program ... 69."
- [10] W. Yustanti and N. Rochmawati, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring," 2022.
- [11] A. M. Siregar and A. Puspabhuana, *Data Mining Pengolahan Data Menjadi informasi dengan RapidMiner*. Surakarta: CV KEKATA GROUP, 2017.
- [12] M. Mirmozaffari, A. Alinezhad, and A. Gilanpour, "Data Mining Apriori Algorithm for Heart Disease Prediction," *International Journal of Computing, Communication and Instrumentation Engineering*, vol. 4, no. 1, Jan. 2017, doi: 10.15242/ijccie.dir1116010.
- [13] J. Suntoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT Elex Media Komputindo, 2019.
- [14] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*, Cetakan 1. Yayasan Kita Menulis, 2020.
- [15] A.- Arini, L. K. Wardhani, and D.- Octaviano, "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden," *KILAT*, vol. 9, no. 1, pp. 103–114, Apr. 2020, doi: 10.33322/kilat.v9i1.878.
- [16] J. Indriyanto, *Algoritma K-Nearest Neighbor untuk Prediksi Nasabah Asuransi*. Pekalongan: PT. Nasya Expanding Management, 2021.