



Analisis Sentimen Pada Pembelajaran Daring Menggunakan Metode K-Nearest Neighbour (Studi Kasus: SMA Negeri 3 Cikampek)

Alfina Noviyanti¹, Yuyun Umaidah², Rini Mayasari³

^{1,2,33}Universitas Singaperbangsa Karawang

Received: 11 Juli 2022
Revised: 17 Juli 2022
Accepted: 22 Juli 2022

Abstract

Abstract: Corona Virus Disease-2019 has been rampant throughout the world, including in Indonesia. Covid-19 has greatly affected several sectors, one of which is the education sector. The Indonesian government itself implements a policy of courageous learning or distance learning which is carried out from their respective homes. SMA Negeri 3 Cikampek is one of the schools that implements bold learning, this bold learning affects the achievement of learning outcomes. Various students from this bold learning, there are those who agree that this bold learning has an effect on the achievement of learning outcomes and some even give a response that does not agree because it has no effect. For this reason, data mining is applied, especially text mining with the K-Nearest Neighbor algorithm to analyze various student responses to bold learning. The data used is a questionnaire data as much as 592 data. Before the data mining stage, the data is divided into 80% of the training data and 20% of the testing data. The classification with K-Nearest Neighbor quality is 85.35% accuracy, 81.19% precision, 92.42% recall and Auc is 0.902. Based on the quantity of negative classes which are more than positive classes, it is known that students will not agree to bold learning because it affects the achievement of learning outcomes.

Keywords: Text mining, Online Learning, Covid-19, K-Nearest Neighbour.

(*) Corresponding Author: alfina.noviyanti18035@student.unsika.ac.id

How to Cite: Yanti, A., Umaidah, Y., & Mayasari, R. (2022). Analisis Sentimen Pada Pembelajaran Daring Menggunakan Metode K-Nearest Neighbour. *Jurnal Ilmiah Wahana Pendidikan*, 8(12), 12-19. <https://doi.org/10.5281/zenodo.6943200>

PENDAHULUAN

Corona Virus Disease-2019 telah merajalela di seluruh dunia, termasuk di Indonesia. Dunia sedang mengalami krisis di bidang kesehatan. Awalnya, wabah Corona Virus Disease-2019 berdampak besar pada lesunya aktivitas sektor ekonomi. Kompas (2020). Setiati & Azwar (2020) menyatakan bahwa pandemi Corona Virus Disease 2019 (Covid-19) merupakan masalah berkelanjutan di dunia. Covid-19 adalah penyebab wabah penyakit menular di Wuhan, China. Hingga 31 Maret 2020, ada 719.758 kasus yang dikonfirmasi di seluruh dunia. Jumlah kematian terkait Covid-19 juga mencapai 33.673 di seluruh dunia. Pandemi ini telah menghasilkan penelitian yang cepat dalam menangani kondisi tersebut.

Perserikatan Bangsa-Bangsa menyatakan bahwa salah satu sektor yang menciptakan wabah ini adalah dunia pendidikan (Purwanto dkk, 2020:1). Hal ini menyebabkan segala aktivitas yang dilakukan di luar rumah menjadi terhenti, salah satunya adalah pendidikan. Pendidikan pada dasarnya merupakan proses pematangan diri terhadap kualitas hidup peserta didik. Dalam pengertian dasarnya, kegiatan pendidikan merupakan bagian dari proses menjadikan seseorang membentuk dirinya sehingga dapat berkembang (Fahrizqi, dkk., 2021).

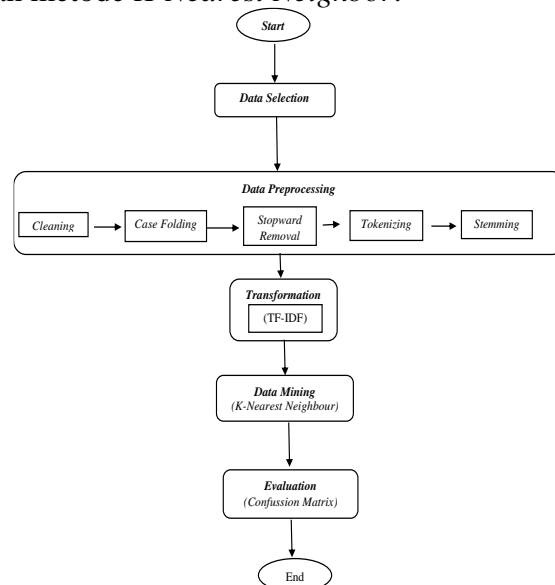


Di masa pandemi, SMA Negeri 3 Cikampek mematuhi kebijakan pemerintah dengan melakukan proses pembelajaran online. Mulai dari terbitnya rekomendasi belajar dari rumah. Sistem pembelajaran online yang berjalan memanfaatkan platform pembelajaran online seperti Google Form, Google Classroom, Zoom, dan media komunikasi seperti WhatsApp. Platform digunakan sebagai alat untuk menyajikan materi pembelajaran dan alat untuk mengumpulkan kembali tugas-tugas yang telah diberikan oleh guru. Banyak siswa merasa dalam proses pembelajaran online. Salah satunya terkait dengan pencapaian hasil belajar. Termasuk apa yang dirasakan oleh pendidik atau guru, guru merasa bahwa selama ini berani belajar sangat berpengaruh terhadap pendidikan karakter dan pencapaian hasil belajar siswa. Dalam proses pembelajaran yang berani, akankah hasil belajar yang diperoleh siswa pada raport mendapatkan hasil yang baik (Pratomo, Cahyo, dkk 2021).

Mengacu pada latar belakang di atas, maka penulis bertujuan akan melakukan penelitian tentang “Analisis Sentimen pada pembelajaran Daring Menggunakan Metode *K-Nearest Neighbour*”. Alasan melakukan penelitian ini bertujuan untuk mengetahui respon dan pandangan siswa terhadap kualitas ketercapaian hasil belajar selama pembelajaran daring. Hal yang akan dilakukan dalam penelitian ini adalah melakukan analisis sentimen dari opini siswa tentang pembelajaran daring. Semoga penelitian ini dapat menjadi bahan evaluasi bagi sekolah dan tenaga pendidik dalam proses pembelajaran daring di Sma Negeri 3 Cikampek.

METODOLOGI PENELITIAN

Metodologi penelitian yang akan digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD). Pada tahap preprocessing data dilakukan *case folding*, *cleaning*, *tokenizing*, *stopwords*, *stemming* dan normalisasi. Setelah proses preprocessing data dilakukan, langkah selanjutnya adalah transformasi data menggunakan TF-IDF dan tahap data mining diklasifikasikan menggunakan metode *K-Nearest Neighbor*.



Gambar 1 Rancangan Penelitian

1. *Data Selection*

Penelitian ini menggunakan dataset angket yang bersumber dari seluruh siswa di Sma Negeri 3 Cikampek. Proses yang dilakukan dalam pemilihan data penulis dan jumlah data yang digunakan untuk proses mining masih mewakili data asli dari kumpulan data. Kemudian dari hasil seleksi tersebut, data tersebut disimpan dan dipisahkan menjadi dataset baru untuk digunakan dalam proses data mining

2. *Data Preprocessing*

Pada tahap Data Preprocessing yaitu proses pembentukan dan pembersihan data sebelum melakukan proses klasifikasi. Berikut adalah proses dari tahap Data Preprocessing

3. *Data Transformation*

Pada tahap ini, skor diberikan pembobotan untuk setiap kata menggunakan TF-IDF. Data sebelumnya dibagi dengan persentase data latih sebesar 80% setelah hasil stemming dan data pengujian pengujian sebesar 20%. Distribusi data dengan rasio 80:20 Merujuk pada penelitian Putri Fitrianti, Risma dkk (2019) menggunakan data perbandingan rasio menghasilkan akurasi analisis sentimen menggunakan metode K-Nearest Neighbor dengan hasil yang cukup baik.

4. *Data Mining*

Pada tahap ini dilakukan pengolahan data dan pengetahuan dengan menggunakan dua proses klasifikasi. yaitu klasifikasi manual secara manual yang dilakukan oleh ahli bahasa dan menggunakan metode *K - Nearest Neighbor*.

5. *Evaluation*

Pada tahap ini akan dilakukan pengujian dengan menggunakan Confusion Matrix. Hasil evaluasi berupa nilai akurasi, presisi, dan recall. Berdasarkan nilai ini, seberapa baik dan akurat proses klasifikasi akan.

HASIL DAN PEMBAHASAN

1. *Data Selection*

Tahapan ini diawali dengan pengumpulan data berupa angket untuk seluruh siswa di SMA Negeri 3 Cikampek. Rentang waktu pengisian angket adalah 15 April 2022 – 25 April 2022. Angket pembelajaran yang mempengaruhi pencapaian hasil belajar menghasilkan 592 data sentimen.

2. *Data Preprocessing*

a. *Cleaning*

Menghapus atribut yang ada pada data seperti simbol, angka, tagar, url, dan emoji. Gambar di bawah ini adalah datanya mentang yang telah melalui tahap pembersihan jadi simbol, url, tagar, angka, dan emoji sudah hilang.

tapi kendalanya saat tidak ada koreksi tidak bisa mengikuti pembelajaran
 tapi karena udah dua tahun ini daring terus jadi mulai terbiasa
 sulit dipahami
 soal
 seperti tugas sekolah
 seperti orang tua siswa dan pemerintah
 semangat pembelajaran daring
 semangat daring nilai itu menurun
 sekolah kepada pemangku kepentingan pendidikan
 sehingga ketercapaian hasil belajar di rapor pun berpengaruh sangat besar
 sebenarnya saya sendiri tidak tau berpengaruh atau tidak soalnya kan masih kelas 10 jadi tidak bisa memba...
 sebelum daring di ahmdulillah nilai saya di atas 80m semua di rapor tapi pas daring udah lumayan anj...
 saya merasa kurang efektif saat belajar daring terkadang saya kurang mengerti materi pembelajaran ya...
 sangat berpengaruh pada nilai yang menjadi kurang maksimal
 sangat berpengaruh

Gambar 2 Cleaning

b. Case folding

proses yang mengubah huruf dalam data, huruf itu diubah dari huruf besar ke huruf kecil huruf kecil (lowercase) dari "a" ke "z". Di bawah ini adalah data mentah yang telah dikonversi ke huruf kecil di panggungkasus lipat.

ada beberapa materi yang memang paham ketika pembelajaran daring, jadi berpengaruh ke nilai rapor yang sedikit turun
 ada yang berpengaruh, ada yang tidak, dan masih banyak guru yang memberi nilai rapor di sama kelas, tidak melaku s...
 agak kurang nilai rapornya
 agak susah di mengerti saat pembelajaran daring
 tapi saya hasil rapor masih kurang dengan pembelajaran daring ini
 banyak materi yang kurang paham
 banyak pelajaran yg tidak dapat dipahami
 berpengaruh karena saat pembelajaran daring minat untuk belajar kadang dikurangi yang membuat nilai rapor kita menurun...
 belajar dengan guru dan teman semangat agar nilai rapor lebih bagus lagi
 belajar bisa menjadi maksimal
 belum bisa mencapai target lagi disebabkan sementara ini hasil rapor saya bagus
 berhasil karena selalu tidak lupa waktu
 berpengaruh apabila banyak faktor seperti kurang nya konsentrasi, materi kurang bisa dipahami
 berpengaruh atau kurangnya nilai karena dengan daring ada banyak yg tidak paham dan harus ditany siapa yang membantu...
 berpengaruh sangat kuat aku karena semangat sendiri belajar daring dituntut nilai aku sendiri turun karena ketidakl...

Gambar 3 Case folding

c. Tokenizing

proses pemisahan kalimat yang diubah menjadi per-kat yang dibuat dari kalimat

A
 Ada
 Agak
 Akan
 Akhir
 Aku
 Ahmdulillah
 B
 BANGET
 BLENG
 BRNGT
 BURUK
 Bagi
 Baik
 Banyak

Gambar 4 Tokenizing

d. *Stemming*

proses yang dilakukan untuk pemilihan kata dalam dokumen teks hapus kata-kata yang tidak membantu dalam penelitian.



Gambar 5 *Stemming*

e. *Data Transformation*

Teknik selanjutnya yang dilakukan adalah menghitung term dari setiap dokumen dan dilanjutkan dengan pembobotan kata dengan TF-IDF.

Tabel 1 Nilai Accuracy

Nilai K	Accuracy	Recall	Presisi	AUC
3	85.35%	92.42%	81.19%	0.902

The image shows a matrix of TF-IDF values. The rows represent words and the columns represent documents. The values are numerical, indicating the importance of each word in each document.

Gambar 6 TF-IDF

f. *Data Mining*

Pada tahap ini dilakukan proses klasifikasi sentimen untuk bold learning menggunakan algoritma *K-Nearest Neighbor*. Hasil pengujian variasi nilai k pada sistem klasifikasi sentimen ditunjukkan pada Tabel 4.2.

The image shows a table with columns for 'Nilai k', 'Akurasi', and 'Presisi'. The rows correspond to different values of k, showing the resulting accuracy and precision for each.

Gambar 7 Hasil akurasi



Gambar 8 hasil Kurva ROC

Gambar 8 menjelaskan hasil ROC. Melengkung algoritma svm yang nilai AUC nya adalah mengembalikan 0,902.

g. *Evaluation*

Hasil analisis sentimen menggunakan matriks konfusi dengan algoritma *K-Nearest Neighbors* adalah sebagai berikut:

Perhitungan confusion matrix algoritma *K-Nearest Neighbors*:

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FN} \times 100 \% \\ &= \frac{278+328}{278+328+27+77} \times 100\% = 85.35\% \end{aligned}$$

KESIMPULAN

Hasil tes yang telah dilakukan terhadap proses pembelajaran di SMA Negeri 3 Cikampek. Data yang terkumpul dari kuesioner sebanyak 592 data dengan menggunakan algoritma *K-Nearest Neighbors*. Pada penelitian ini menghasilkan akurasi sebesar 85,35%, recall sebesar 92,42%, presisi sebesar 81,19% dan Auc sebesar 0,902.

DAFTAR PUSTAKA

- Andika, L. A. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode *Naive Bayes Classifier*. *Indonesian Journal of Applied Statistics*, 34.
- Andreas, D., Arieshanti, I., & Zainal Arifin, A. (2020). IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOUR*. 1-7.
- Arviana, G. N. (2022, Februari Senin). *Tableu*. Retrieved from Glints: <https://glints.com/id/lowongan/tableau-> adalah
- Azwar, M. K., & Setiati, S. (2020). COVID-19 and Indonesia. 22-25.
- B, L. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* , University of Illinois at Chicago. *Cambridge University Press*.
- Bersiap tameng ekonomi untuk dampak wabah corona. (2022, Februari Senin). Retrieved from Kompas.com: <https://jeo.kompas.com/bersiap-tameng-ekonomi-untuk-dampak-wabah-corona>
- Bramer, M. (2007). *Principles of Data Mining*. Springer Science.
- dkk, P. (2020). Studi Eksploratif Dampak Pandemi COVID-19 Terhadap . *Journal of Education* , Volume 2 No. 1.
- Fahrizqi, E. B. (2021). *The Learning Motivation and Physical Fitness of University Student During the Implementations of Covid-19 Pandemic* . JUARA: *Jurnal Olahraga*, 88-100.
- Fajrin, A. A. (2018). Penerapan *Data Mining* untuk Analisis Pola Pembelian Konsumen dengan Algoritma Fp- Growth pada Data . *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 27-36.
- Fatyanosa, T. N. (2017). *Classification method comparison on Indonesian social media sentiment analysis*. *International Conference on Sustainable Information Engineering and Technology (SIET)*, 24-25.

- hafira, N. H. (2020). Persepsi Mahasiswa Terhadap Perkuliahan Daring Sebagai Sarana Pembelajaran. *Jurnal Bisnis Dan Kajian Strategi Manajemen*.
- Hidayat. (2017). ANALISIS HAMBATAN BELAJAR MAHASISWA PADA MATA KULIAH STATISTIKA. *Jurnal JPPM*.
- Indonesia, K. K. (2022, Februari Senin). *Profil Kesehatan* . Retrieved from Kementerian Kesehatan Republik Indonesia.
- J. Iawe. Han, M. K.-H.-F. (2012). *Data Mining in Healthcare and Biomedicine. A Survey of The Literature*, 2431–2448.
- Kristina, E. (2011). Penerapan Metode Statistik Dan Average Energy Untuk Menguji Tingkat Kemiripan Pada Identifikasi Suara. *Jurnal Informatika*, Vol 7, No 1.
- Kusrini. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Larose, D. T. (2005). Discovering Knowledge in Data Mining An Introduction to Data Mining. *Wiley Interscience*.
- Monika, N. &. (2013). Analisis Chi-Square dan Transformasi Data Ordinal ke Data Interval Menggunakan Methods Of Succesive Interval . *Eksponensial*, 85-90.
- Negara, A. B. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode *Naive Bayes* dan Seleksi Fitur *Information Gain*. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 599.
- Nuramilus, S. R. (2017). Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode *K-Nearest* . *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 426-435.
- Paramita, A. (2016). Perancangan Data Warehouse pada Perpustakaan Yayasan Lentera Insan. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 83-88.
- Putra, R. R. (2018). Implementasi *Data Mining* Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Journal of Information Technology and Computer Science (INTECOMS)*, 72-77.
- Putrianti, R. P., Kurniati, A., & Agustin, D. (2019, August). Implementasi Algoritma *K- Nearest Neighbor* Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*.
- R. Akbar, R. O. (2017). Implementasi Business Intelligence Un-tuk Menentukan Tingkat. *Jurnal Ilmiah Informati-ka*, 135-138.
- Sani, R. R. (2016). Pengembangan Aplikasi Penentuan Tema Tugas Akhir Berdasarkan Data Abstrak Menggunakan . *Prosiding seminar Nasional Multi Disiplin Ilmu & Call For Papers*, 103-111.
- Santoso, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis (1 ed)*. Yogyakarta: Graha Ilmu.
- Syakuro, A. (2017). Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode *Naive Bayes Classifier* (NBC) Dengan Seleksi Fitur *Information Gain* (IG). *Skripsi*, 1-89.
- Waryanto, N. (2006). Online learning sebagai salah satu inovasi pembelajaran. *Jurnal Matematika*, 10-23.

- Yunianto, T. K. (2022, Februari Senin). *Survei SMRC: 92% Siswa Memiliki Banyak Masalah dalam Belajar Daring*. Retrieved from katadata.co.id: <https://katadata.co.id/ekarina/berita/5f3bc04617957/survei-smrc-92-siswa-memiliki-banyak-masalah-dalam-belajar-daring>
- Yustanti, W. (2012). Algoritma *K-Nearest Neighbour* untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika Statistika & Computing*, 57-68.
- Zulkifli, A. (2016). Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler. *Journal Of Computer Science*, 65-76.