

Paper

Klasifikasi Sentimen Tweet Pada Twitter Terhadap Pembelajaran E-Learning Menggunakan Metode k-Nearest Neighbor

Author: Nurcholis Geofany, Tommy, Risiko Liza

Klasifikasi Sentimen *Tweet* Pada Twitter Terhadap Pembelajaran *E-Learning* Menggunakan Metode *k-Nearest Neighbor*

Nurcholis Geofany¹, Tommy², Risiko Liza³

^{1,2,3}Universitas Harapan Medan, Medan, Indonesia

¹nurcholis.geofany@gmail.com, ²tommy.shirakawa@gmail.com, ³risko.liza@gmail.com

Abstrak-Indonesia merupakan salah satu pengguna Twitter yang paling aktif, sehingga Twitter dapat digunakan sebagai media untuk menganalisis topik e-learning. Belajar di sekolah menjadi salah satu hal yang terkena dampak dari pandemi Covid-19 yang berlangsung selama ini. Pada 4 Maret 2020, UNESCO (United Nations Educational, Scientific and Educational Organization) merekomendasikan agar sekolah menggunakan sistem pendidikan jarak jauh yang memungkinkan guru menjangkau siswa dari jarak jauh dan membatasi gangguan pendidikan di jejaring sosial. Penelitian ini bertujuan untuk menjawab opini publik yang berfokus pada media sosial Twitter terhadap kebijakan pembelajaran online/jarak jauh ini. Analisis sentimen dilakukan untuk mengetahui apakah opini terhadap suatu isu memiliki nilai tren positif, netral, atau negatif dan dapat digunakan sebagai tolak ukur untuk meningkatkan suatu pelayanan. Analisis sentimen menggunakan metode *k-Nearest Neighbor* dengan data Twitter sebanyak 7800 data tweet yang telah diolah sebelumnya dan mencapai nilai akurasi terbaik sebesar 53,03% dengan perbandingan data latih dan data uji 80:20.

Kata Kunci: *Analisis Sentimen, k-Nearest Neighbor, Twitter*

Abstract-Indonesia is one of the most active Twitter users, so Twitter can be used as a medium to analyze e-learning topics. Studying at school is one of the things that have been affected by the current Covid-19 pandemic. On March 4, 2020, UNESCO (United Nations Educational, Scientific and Educational Organization) recommended that schools adopt a distance education system that allows teachers to reach students remotely and limits educational disruption on social networks. This study aims to answer public opinion that focuses on social media Twitter on this online/distance learning policy. Sentiment analysis is conducted to determine whether opinions on an issue have a positive, neutral, or negative trend value and can be used as a benchmark to improve a service. Sentiment analysis using the *k-Nearest Neighbor* method with Twitter data as much as 7800 tweet data that has been previously processed and achieves the best accuracy value of 53.03% with a comparison of training data and test data of 80:20.

Keywords: *Sentiment Analysis, k-Nearest Neighbor, Twitter*

1. PENDAHULUAN

Perkembangan teknologi informasi telah berkembang dengan sangat pesat dalam beberapa tahun terakhir, sehingga dengan perkembangan ini paradigma masyarakat telah berubah dalam pencarian dan perolehan informasi yang tidak lagi terbatas pada informasi jurnalistik, audiovisual dan elektronik, tetapi juga sumber informasi lain yang tergabung melalui *internet* [1]. Salah satu aplikasi Internet yang mendukung konten buatan pengguna adalah *microblogging*.

Saat ini *microblogging* atau yang lebih dikenal dengan Twitter sedang populer sebagai alat komunikasi di kalangan pengguna internet. Hal ini terlihat dari peningkatan jumlah pengguna Twitter yang terdaftar di seluruh dunia. Pada Q4 2018, Twitter rata-rata memiliki 126 juta pengguna aktif harian [2]. Indonesia adalah negara dengan jumlah pengguna Twitter terbesar ke-3 di dunia menurut data yang dirilis oleh Twitter Indonesia pada akhir tahun 2016, dilaporkan bahwa 77% pengguna Twitter di Indonesia adalah pengguna aktif yang produktif dalam menulis *tweet*. Hal ini terlihat dari jumlah *tweet* yang dihasilkan pada tahun 2016 yang mencapai 4,1 miliar *tweet*[3].

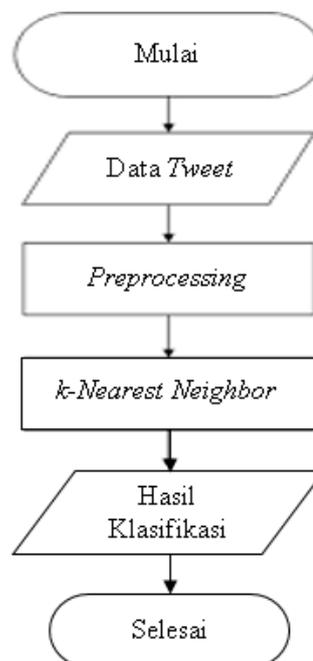
Analisis sentimen juga dapat disebut sebagai jajak pendapat yang merupakan bidang studi yang menganalisis pendapat, perasaan, penilaian, peringkat, sikap, dan emosi seseorang dalam kaitannya dengan subjek, layanan, produk, individu, suatu organisasi atau kegiatan. masalah, memiliki kecenderungan positif, netral atau negatif dan dapat dijadikan sebagai tolak ukur untuk meningkatkan suatu pelayanan atau meningkatkan kualitas produk [4]. Penelitian terkait sebelumnya didapat nilai akurasinya dari pengklasifikasian

sebesar 58,10% [5]. Kemudian penelitian dilakukan [6] mendapatkan hasil akurasi yang cukup tinggi dengan tingkat akurasi sebesar 79,21%.

Dalam penelitian [7] banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi, dalam pemanfaatannya untuk membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Metode yang digunakan adalah *k-Nearest Neighbor* yang akan mengklasifikasikan perasaan positif, netral atau negatif. Keuntungan *k-NN* adalah bekerja berdasarkan jarak terpendek antara *instance query* dan sampel pelatihan untuk menentukan *k-NN* [8]. Dari kesimpulan ini, data Twitter akan dipindai dan diproses terlebih dahulu untuk mendapatkan nilai presisi yang lebih tepat.

2. METODOLOGI PENELITIAN

Dalam implementasi *k-Nearest Neighbor* pada kasus sentimen analisis dengan topik *e-learning*, dataset yang digunakan diambil dari hasil *crawling* dengan topik *e-learning*. Pada data latih aspek sudah ditentukan bersama sentimennya pada setiap kalimat. Sebelum masuk pada pelatihan *k-Nearest Neighbor* untuk klasifikasi sentimen, perlu dilakukan proses *preprocessing* [9]. Tahap *preprocessing* ini meliputi tahap *case folding*, *filtering*, *word normalization*, *tokenization*, *stopword removal* [10]. Selain kalimat pada data latih dilakukan *preprocessing* sebelum masuk tahap pengujian.



Gambar 1. Gambaran Umum Sistem yang Akan Dibangun

Berikut merupakan penjelasan dari Gambar 1:

1. Data masukan (data latih dan data uji) diambil dari hasil *crawling*. Pada data latih aspek sudah ditentukan bersama sentimennya.
2. Pada proses pelatihan akan dilakukan *preprocessing* terhadap kalimat yang menjadi data latih, dimana dalam *preprocessing* dilakukan *filtering*, *case folding*, *tokenization*.
 - a. *Filtering*
Filtering mengambil kata-kata penting dari hasil *tokenization* atau biasa disebut dengan *rule-based word deletion*. Algoritma *stop word removal* inilah yang digunakan untuk melakukan langkah penyaringan.
 - b. *Case Folding*
Case Folding biasa disebut dengan keseragaman kata dengan mengubah semua kata menjadi huruf kecil (*lowercase*). Hanya huruf a sampai z yang dapat diterima oleh karakter selain huruf yang dihilangkan. Ada juga kata-kata tertentu yang harus menghormati aturan yang tidak baku, seperti kata lembaga atau lembaga yang selalu diawali dengan huruf kapital dan juga nama gelar seperti ST, M.Psi dll. Itu tergantung pada sumber data yang digunakan untuk pemrosesan.

- c. *Tokenezation*
Tokenizing suatu tahapan pemotongan *string* kata berdasarkan penyusunan kata tersebut.
3. Algoritma *k-Nearest Neighbor* bekerja berdasarkan jarak terpendek antara *instance* permintaan dan data pelatihan untuk menentukan *k-NN*-nya. Salah satu cara untuk menghitung jarak dekat atau jauh dari tetangga adalah dengan menggunakan metode jarak *Euclidean*. Dalam algoritma ini, nilai *k* terbaik tergantung pada jumlah data. belum tentu nilai *k* terbaik dan sebaliknya. Langkah-langkah menghitung algoritma *k-NN*:
 - a. Tentukan nilai *k*.
 - b. Menghitung kuadrat jarak *Euclidean* (contoh kueri) dari setiap objek sehubungan dengan data pelatihan data.
 - c. Kemudian urutkan benda-benda tersebut ke dalam kelompok-kelompok yang memiliki jarak *Euclidean* terkecil.
 - d. Koleksi label Kelas Y (klasifikasi tetangga terdekat).
 - e. Dengan menggunakan kategori tetangga terdekat yang mayoritas, nilai *instance query* yang dihitung dapat diprediksi.
4. Setelah itu hasil klasifikasi dilakukan *confussion matrix* untuk menghitung hasil akurasi.

3. HASIL DAN PEMBAHASAN

Bagian ini menjabarkan proses ekstraksi yang digunakan untuk mengklasifikasi kata kata yang bernilai positif, netral atau negatif dan studi kasus yang diangkat pada penelitian ini adalah media sosial Twitter tentang pembelajaran *online e-learning*. Akuisisi data teks tentang *tweet* pembelajaran *onlinee-learning* kemudian diekstraksi teks setelah itu diidentifikasi dan dikelompokan menjadi kata-kata yang mengandung sentimen positif, sentimen netral, dan sentimen negatif. Algoritma *k-Nearest Neighbor* digunakan sebagai metode klasifikasi terhadap vektor fitur teks yang diakuisisi lewat media sosial Twitter. Pada penelitian ini data diambil hanya data *tweet* yang menggunakan teknik *crawling* melalui media sosial Twitter kemudian diklasifikasi berdasarkan kelas kata tersebut.

3.1 Implementasi

Implementasi dilakukan setelah perancangan selesai dilakukan dan selanjutnya akan di implementasikan pada bahasa pemrograman yang akan digunakan pada penelitian ini menggunakan bahasa pemrograman *python*. Tujuan implementasi adalah untuk dapat melakukan klasifikasi kata positif, netral dan negatif.

3.2 Hasil dan Implementasi

Pada bagian ini akan dipaparkan hasil dari implementasi proses pembuatan dan implementasi sistem klasifikasi kata positif, netral dan negatif dengan teknik sentimen analisis menggunakan algoritma *k-Nearest Neighbor*.

3.2.1 Proses Akuisisi

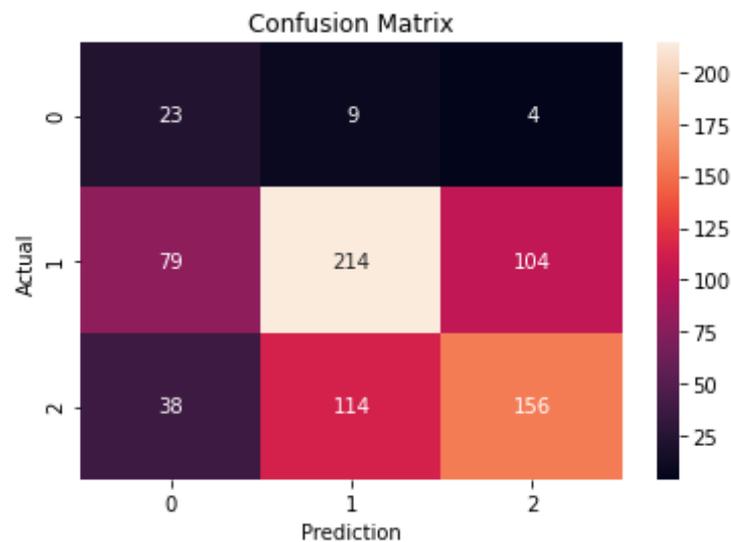
Data Berikut ini tabel dari hasil *crawling* terhadap data *tweet* yang mengandung kata yang berkaitan dengan topik yang dianalisis:

Tabel 1. Data *Tweet* Pembelajaran *e-learning*

No	Data <i>Tweet</i>
1	F@aku fenomena yang cukup sering ditemui. Apalagi yang kelas 12. Apalagi di masa sekolah daring yang katanya lebih banyak tugasnya. ☹(
2	6, "Pemerintah mengumumkan pembelajaran tatap muka secara terbatas di sekolah bisa dimulai pada Juli 2021 mendatang. Apakah langkah itu tepat di tengah kasus COVID-19 yang masih tinggi? @kasd #daring
3	8, "@adekumala, Aku cuma berdoa semoga sekolah anakku gak jadi tatap muka per 1 April ini. Kalau maksa, aku akan buat surat pernyataan keberatan karena kasus di sini masih sangat tinggi tapi ditutupi.
4	Sudah siap belajar tatap muka di sekolah atau masih betah belajar daring? :*(
5	25, "PENCEGAHAN CYBERBULLYING SAAT SEKOLAH DARING
6	26, "Pembelajaran melalui daring banyak menemui kendala dan banyak dikeluhkan org tua mulai dr tdk mempunyai hp android, sinyal, kuota, d kebingungan org tua dlm membimbing anaknya di rumah @dkfee

3.5 Evaluasi

Confusion Matrix tabel dengan 6 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada enam istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu *True Positif*, *True Netral*, *True Negatif*, *False Positif*, *False Netral* dan *False Negatif*.



Gambar 4. *Confusion Matrix*

Keterangan Gambar 4 akan menjelaskan bahwa gambar diatas menunjukkan sebaran prediksi data TP (*True Positive*), TNe (*True Netral*), TN (*True Negatif*), FP (*False Positif*), FNe (*False Netral*), FN (*False Negatif*). *Class 0* adalah *class positif*, *class 1* adalah *class netral*, dan *class 2* adalah *class negatif*. Berdasarkan Gambar 3.3 *class 0* telah terprediksi positif benar sebesar 23 dan 13 salah. Kemudian untuk *class 1*, telah terprediksi netral benar sebesar 214 dan 183 salah. Pada *class 2*, 156 data terprediksi negatif benar dan 153 data terprediksi salah. Sehingga didapatkan untuk hasil akurasi dari analisis sentimen berikut:

Keterangan:

Kelas Positif (0) = 36

Kelas Netral (1) = 397

Kelas Negatif (2) = 308

Setelah itu dilakukan perhitungan untuk melihat akurasinya. Dengan cara dibawah ini:

$$TP + TNe + TN / TP + TNe + TN + FP + FNe + FN * 100 =$$

$$23 + 214 + 156 = 393 / 23 + 214 + 156 + 13 + 183 + 153 = 741$$

$$393 / 741 = 0,5303 * 100 = 53,03\%$$

Dan akurasi yang didapat pada klasifikasi terhadap topik pembelajaran *e-learning* yang di analisis, yaitu sebesar 53,03%.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh bahwa penelitian tentang klasifikasi sentimen pengguna Twitter terhadap topik pembelajaran *e-learning* dengan menggunakan metode *k-Nearest Neighbor* dengan kesimpulan sebagai berikut:

1. Penelitian dilakukan dengan menggunakan 741 data dan diproses menggunakan metode *k-Nearest Neighbor* dan mencapai akurasi keseluruhan sebesar 53,03%.
2. Hasil analisis sentimen menunjukan nilai opini netral menempati peringkat pertama dengan jumlah sentimen sebanyak 397. Hasil ini menunjukan bahwa antusias masyarakat tidak terlalu tinggi dalam menanggapi pembelajaran *online e-learning* selama pandemi ini.
3. Hasil analisis sentimen menunjukan nilai opini negatif menempati posisi kedua dengan jumlah sentimen sebanyak 308 dan masih lebih besar daripada opini positif yang menempati posisi ketiga dengan jumlah

- sentimen sebanyak 36. Hasil ini menunjukkan bahwa pembelajaran *online e-learning* ini selama pandemi masih dianggap masyarakat kurang efektif.
4. Faktor data yang tidak seimbang, kumpulan data yang tidak sesuai, dan kurangnya pemahaman tentang klasifikasi sentimen menyebabkan proses klasifikasi tidak berjalan maksimal dan hanya mencapai nilai akurasi sebesar 53,03 %. Pada penelitian lain, kualitas kumpulan data perlu ditingkatkan untuk hasil yang optimal.
 5. Kosakata tidak baku mempengaruhi hasil klasifikasi terhadap suatu kelas data uji apabila suatu kelas data latih memiliki data jumlah kata tidak baku lebih banyak dibandingkan kelas data latih lainnya.

DAFTAR PUSTAKA

- [1] Pratomo, Y. 2019. "Untuk Pertama Kali, Twitter Ungkap Jumlah Pengguna Harian." *Kompas.Com*, pp.1-6. Retrieved from
- [2] Herman dan Mononimbar, D.A. 2017. "Indonesia Fifth-Largest Country in Terms of Twitter Users. JakartaGlobe, pp.4-7. Retrieved from
- [3] Nurjanah, W. E, Perdana, R. S, dan Fauzi. 2017. "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter Menggunakan Metode *k-Nearest Neighbor* dan Pembobotan Jumlah *ReTweet*."
- [4] Yuliana, Tanti. 2020. "Analisis Sentimen Publik Terhadap Kebijakan Pemerintah Dalam Penanganan Covid-19 Di Indonesia Menggunakan Naïve Bayes Classifier Dan K-Nearest Neighbor."
- [5] M Syarifuddin. 2020. "Analisis Sentimen Opini Publik Mengenai *Covid-19* Pada Twitter Menggunakan Metode *k-NN*."
- [6] A Salam, et, al. 2018. "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan *k-Nearest Neighbor* (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)."
- [7] Mahardika, Y. S, dan Zuliarso, E. 2018. "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes."
- [8] Kumar, Akash, Aniket Verma, Gandhali Shinde, Yash Sukhdeve, And Nidhi Lal. 2020. "Crime Prediction Using K-Nearest Neighboring Algorithm." *International Conference On Emerging Trends In Information Technology And Engineering, Ic-Etite 2020* 4–7. Doi: 10.1109/Ic-Etite47903.2020.155.
- [9] Primantoro, Faizal Asri Primantoro. 2018. "Aplikasi Klasifikasi Film Berdasarkan Analisis Sentimen Berbasis Web."
- [10] Wijaya, Arnandia Raka, dan Arief Jananto. 2018. "Mencari Pola Pembelian Konsumen Menggunakan Algoritma Fp-Growth."