

PENGUNAAN TEXT MINING UNTUK KLASIFIKASI TANGGAPAN PESERTA PELATIHAN TERHADAP PERFORMA TRAINER MENGGUNAKAN METODE ALGORITMA K-NEAREST NEIGHBOR

Faizal Riza,

*Program Studi Teknik Informatika, Institut Teknologi Budi Utomo Jakarta,
faizalriza@itbu.ac.id*

Abstrak

Analisis dan klasifikasi data komentar dilakukan dengan cara membaca dan mengklasifikasikan satu per satu komentar yang bersifat negatif dengan menggunakan lembar kerja (*spreadsheet*) akan menjadi kurang efektif apabila data yang akan diproses berjumlah sangat banyak. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan teknik *analisis sentimen* pada data komentar menggunakan metode K-Nearest Neighbor (KNN). Data komentar yang digunakan yaitu komentar peserta pelatihan terhadap performa trainer Educato.Id yang diisi oleh setiap peserta setelah mengikuti pelatihan. Selanjutnya data komentar tersebut dianalisis dengan melakukan langkah pre-processing, pembobotan kata menggunakan *Term Frequence-Invers Document Frequence* (TFIDF), menghitung tingkat kemiripan antara data latih dan data uji dengan algoritma *cosine similarity*. Proses implementasi *analisis sentimen* dilakukan untuk menentukan apakah komentar tersebut termasuk kategori komentar positif atau negatif. Selanjutnya komentar-komentar tersebut akan ditentukan dalam empat kategori yaitu : trainer, materi, sarana dan prasarana. Hasil penelitian ini menghasilkan sebuah sistem yang dapat mengklasifikasi data komentar secara otomatis dengan tingkat akurasi sebesar 94,23 %.

Kata Kunci : *Analisis sentimen; Komentar; TF-IDF; K-Nearest Neighbor; KNN*

1. PENDAHULUAN

Menurut Kamus Besar Bahasa Indonesia (KBBI), komentar didefinisikan sebagai ulasan atau tanggapan atas berita, pidato, dan sebagainya dengan tujuan menerangkan atau menjelaskan (KBBI, 2018). Komentar merupakan serangkaian penuturan fakta, pendapat atau hasil pengamatan pada suatu objek yang mengandung kritik atau tanpa kritik. Komentar itu sendiri dapat berupa komentar yang bersifat positif maupun negatif yang dapat digunakan sebagai bahan evaluasi. Sebagai penyelenggara pelatihan, Educato.Id memerlukan masukan berupa komentar dari para peserta pelatihan sebagai bahan untuk evaluasi proses pembelajaran yang telah berjalan. Peserta pelatihan dapat memberikan komentar yang positif atau negatif terhadap proses pembelajaran serta hal-hal yang terkait dengan proses pembelajaran yang telah diikutinya, yang pada akhirnya menggambarkan tingkat kepuasan khususnya peserta pelatihan terhadap proses pelatihan di Educato.Id.

Salah satu tantangan dalam pengolahan komentar yang diberikan yaitu proses analisisnya. Sistem yang menampung

komentar yang digunakan oleh Educato.Id hanya sebatas untuk menampung komentar peserta pelatihan. Namun cara pengolahan data komentarnya dilakukan dengan melihat dan merekap data komentar satu per satu komentar yang termasuk komentar negatif dan komentar yang bersifat positif. Kemudian komentar yang bersifat negatif atau dapat dikatakan sebagai komentar yang menyatakan ketidakpuasan terhadap Educato.Id akan diklasifikasikan satu per satu kedalam empat kategori (kategori trainer, materi, sarana dan prasarana) dengan menggunakan Ms.Excel. Melakukan klasifikasi komentar dengan cara tersebut kurang efektif apabila data yang akan diproses dalam jumlah yang banyak. Untuk mempermudah dan mempercepat proses pengolahan data komentar maka diterapkan analisis sentimen untuk proses analisa apakah komentar tersebut bersifat positif atau negatif dan untuk proses pengklasifikasian komentar digunakan algoritma k-nearest neighbor. dengan menerapkan Analisis sentimen menggunakan metode K-Nearest Neighbor (KNN) dapat mempermudah dalam proses pengolahannya. Komentar akan diekstrak dan dianalisis secara otomatis untuk kemudian

ditentukan dan diklasifikasi informasi opini yang terkandung di dalamnya (Sipayung et al., 2016).

Menurut (Riza et al., 2020), dinyatakan bahwa *"the data of matching method will count the number of keywords that appear in the document then return the order of documents with the highest number of occurrences to the user. This can result in the performance of information retrieval becoming less good because they do not pay attention to the information semantically"*.

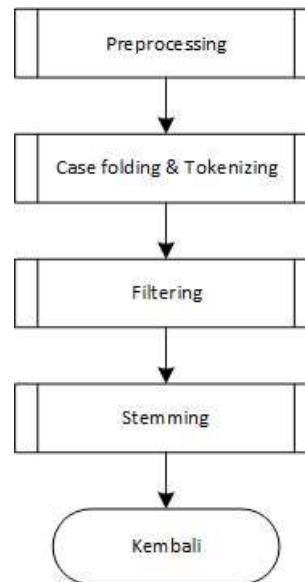
Beberapa metode telah dikembangkan dalam pengklasifikasian teks. Salah satu metode tersebut adalah K-Nearest Neighbor (K-NN). Algoritme ini memiliki kelemahan yaitu nilai k ditetapkan sama pada semua kategori. Namun hal tersebut dapat diatasi dengan menggunakan algoritme Improved K-Nearest Neighbor. Penelitian oleh Baoli, Shiwen, dan Qin (2003) menunjukkan bahwa dengan algoritme Improved K-Nearest Neighbor, mendapatkan hasil kestabilan pada proses kategorisasi dengan k-values yang bervariasi. Dalam metode ini, nilai k-values pada setiap kategori disesuaikan dengan besar-kecilnya jumlah dokumen latih yang dimiliki kategori, sehingga nilai k-values tiap kategori berbeda-beda (Putri, 2013).

2. METODOLOGI

Analisis sentimen merupakan teks yang berupa informasi tekstual yang bersifat opini. Maka dari itu sifat dari analisis sentimen adalah subjektif terhadap suatu hal. Yang dimaksud subjektif disini adalah dapat berupa positif atau negatif. Informasi tekstual yang dikelompokkan menjadi positif atau negatif akan mengandung suatu nilai. Nilai tersebut yang selanjutnya akan digunakan sebagai parameter dalam menentukan suatu keputusan terhadap suatu dokumen (Indriati et al., 2016).

Untuk proses pengolahan datanya akan diterapkan analisis sentimen untuk menganalisa data apakah komentar bersifat negatif atau positif, kemudian hasil dari hasil analisa maka komentar yang bersifat negatif akan diklasifikasikan dengan menerapkan algoritma k-nearest neighbor untuk pengklasifikasian datanya.

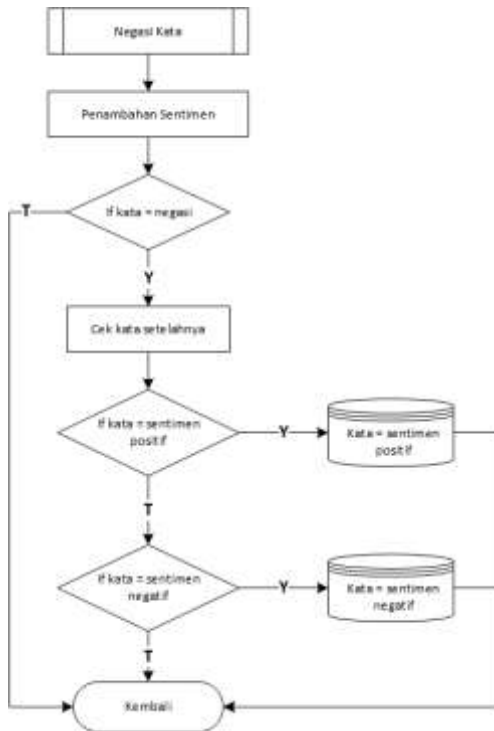
Tahap pertama adalah tahap analisis preprocessing. Pada tahap ini akan dilakukan proses case folding, tokenizing, stopword removal / filtering dan stemming pada semua dokumen (Atina et al., 2017) seperti alur pada gambar 1.



Gambar 1.
Proses Preprocessing
Sumber: Penelitian Mandiri

Tahap kedua adalah tahap analisis sentimen. Pada tahap ini akan dilakukan analisa sentimen pada data uji. Langkah pertama dalam menganalisa sentimen adalah melakukan proses negasi teks. Proses negasi teks ini diperlukan proses load dictionary, dimana dibutuhkan kamus kata kunci sentimen positif, dan kamus kata kunci sentimen negatif.

Berdasarkan kamus kata kunci positif dan negatif maka dibutuhkan pula kata negasi untuk mengubah sentimen pada kata seperti "tidak bersih". Kata "bersih" termasuk kata dengan sentimen positif, tetapi terdapat kata "tidak" yang mengubah makna positif pada kata "bersih" menjadi negatif. Adapun kata negasi yang digunakan yaitu : tidak, bukan, kurang, dan belum. Alur analisis sentimen dapat dilihat pada gambar 2.



Gambar 2.
Flowchart Negasi Text
Sumber: Penelitian Mandiri

Pada tahap analisis sentimen, setelah setiap kata diketahui sentimennya maka selanjutnya yaitu menghitung jumlah kata dengan sentimen positif dan sentimen negatif pada data uji. Jika kata sentimen positif lebih banyak dari kata sentimen negatif maka dokumen uji termasuk komentar dengan sentimen positif, dan sebaliknya apabila kata sentimen positif lebih sedikit dari kata sentimen negatif maka dokumen uji termasuk komentar dengan sentimen negatif. Alur penentuan sentimen data uji dapat dilihat pada gambar 3.



Gambar 3.
Penentuan Sentimen Data Uji
Sumber: Penelitian Mandiri

Tahap ketiga adalah tahap pembobotan TF-IDF. Setelah proses analisa sentimen dilakukan maka selanjutnya yaitu proses perhitungan TF-IDF. Dokumen uji yang akan dilakukan pembobotan TF-IDF adalah komentar yang termasuk sentimen negatif sedangkan komentar dengan sentimen positif tidak akan diproses untuk diklasifikasikan.

Tahap keempat adalah implementasi algoritma K-Nearest Neighbor untuk menghitung klasifikasi data uji. Algoritma k-nearest neighbor adalah algoritma yang proses pengklasifikasian datanya dilakukan dengan menghitung jarak kedekatan antar dokumen latih dan dokumen uji. Oleh karena itu penerapan algoritma k-nearest neighbor pada sistem yang akan dibuat didukung dengan metode tf-idf untuk pembobotan term dari setiap dokumen dan metode cosine similarity digunakan untuk menghitung jarak kedekatan/kemiripan dokumen latih dengan dokumen uji. Proses secara keseluruhan pada sistem yang diusulkan dapat dilihat pada Gambar 3.



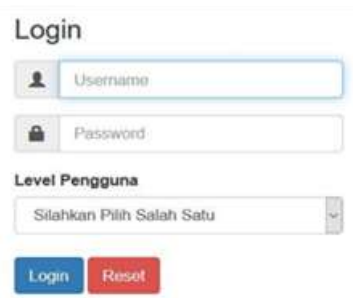
Gambar 4.
Proses Keseluruhan Teknik Analisis Sentimen
Sumber: Penelitian Mandiri

3. ANALISIS DAN PEMBAHASAN

Berdasarkan perancangan aplikasi Pengklasifikasian Data Komentar, secara garis besar aplikasi ini hanya mempunyai dua user, yaitu Admin (staff administasi Educato.Id) dan User (peserta pelatihan). Pada aplikasi ini admin dapat menjalankan atau melakukan

proses menginput, menghapus, mengedit data training (data komentar), melakukan testing data dan mencetak hasil pengklasifikasian. Sedangkan peserta pelatihan hanya bisa mengisi kolom komentar dan tidak memiliki hak akses untuk mengakses menu lainnya.

Tampilan awal dari sistem aplikasi ini adalah menu login, proses login dilakukan oleh peserta pelatihan dan staff administrasi yang berperan sebagai admin. Proses login dilakukan dengan memasukkan user name, password dan memilih level pengguna.



Gambar 5.
Antarmuka Form Login
Sumber: Hasil Olah Data

Form Login User. Halaman data user hanya dapat diakses oleh pengguna yang login ke dalam sistem aplikasi sebagai admin. Halaman data user ini menampilkan semua data user yang tersimpan di dalam database sehingga memudahkan administrator sistem dalam mengelola data pengguna. Admin dapat membuat pengguna baru, mengedit pengguna lama, menghapus pengguna dan pencarian data pengguna seperti terlihat pada gambar 6.



Gambar 6.
Antarmuka Form Data Pengguna
Sumber: Hasil Olah Data

Form Data Training. Halaman data training menampilkan data-data komentar yang telah diklasifikasikan sebelumnya. Pada halaman ini admin dapat menambahkan data

training baru, mengedit data training dan menghapus data training yang sudah ada.



Gambar 7.
Antarmuka Form Data Pelatihan
Sumber: Hasil Olah Data

Form Data Uji. Halaman data uji merupakan halaman yang menampung semua komentar yang telah di input oleh peserta pelatihan. Data komentar yang ditampilkan merupakan data komentar yang belum diolah dan belum diklasifikasikan. Seperti halaman data training, pada halaman ini admin juga dapat menambahkan komentar baru, mengedit komentar, dan juga menghapus komentar yang sudah ada seperti terlihat pada gambar 8.



Gambar 8.
Antarmuka Form Data Uji
Sumber: Hasil Olah Data

Form Grafik. Pada halaman ini berisi grafik yang menampilkan persentase hasil klasifikasi pada tiap kategorinya (sarana, instruktur, materi dan pelayanan). Data pada grafik merupakan data yang diklasifikasikan pada setiap bulan. Tampilan form uji data seperti terlihat pada gambar 9.



Gambar 9.
Antarmuka Form Grafik
Sumber: Hasil Olah Data

Form Grafik. Pada halaman ini berisi grafik yang menampilkan persentase hasil klasifikasi pada tiap kategorinya (sarana, trainer, materi dan pelayanan). Data pada grafik merupakan data yang diklasifikasikan pada setiap bulan. Tampilan form grafik seperti terlihat pada gambar 10.



Gambar 10.
Antarmuka Form Grafik
Sumber: Hasil Olah Data

Form Hasil Klasifikasi. Halaman form hasil klasifikasi merupakan halaman yang menampilkan hasil klasifikasi data yang di uji. Hasil klasifikasi yang ditampilkan adalah hasil klasifikasi data uji dengan sentimen negatif seperti terlihat pada gambar 11.



Gambar 11.
Antarmuka Form Hasil Klasifikasi
Sumber: Hasil Olah Data

Form Pengisian Komentar. Halaman pengisian komentar ini dikhususkan untuk peserta pelatihan. Pada halaman ini peserta pelatihan dapat mengisi komentar mengenai Educato.Id selama menjalani pendidikan dan

pelatihan di Educato.Id. Peserta pelatihan wajib mengisi judul pelatihan yang di ikuti beserta tanggal, bulan dan tahun pada saat mengisi komentar. Tampilan form pengisian komentar seperti terlihat pada gambar 12.



Gambar 12.
Antarmuka Form Pengisian Komentar
Sumber: Hasil Olah Data

Hasil Pengujian yang dilakukan pada sistem analisis sentimen komentar ini, dilakukan dengan menguji 264 data komentar pada bulan Januari sampai dengan Maret tahun 2022. Tingkat akurasi pengujian sistem dapat dilakukan dengan menggunakan rumus akurasi pada *confusion matrix* sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah nilai benar}}{\text{Jumlah data keseluruhan}} \times 100\%$$

Gambar 13.
Rumus Akurasi pada *Confusion Matrix*
Sumber: Hasil Olah Data

Setelah dilakukan pengujian, idapatkan hasil analisa sentimen komentar yang benar sebanyak 251 komentar. Berdasarkan rumus diatas maka tingkat akurasi analisa sentimen adalah sebagai berikut :

$$\text{Akurasi} = \frac{251}{264} \times 100\%$$

$$\text{Akurasi} = 95,08 \%$$

Setelah dilakukan analisa sentimen pada komentar, didapatkan sebanyak 89 komentar bersifat negatif yang akan di proses untuk diklasifikasikan. Setelah dilakukan proses pengklasifikasian oleh sistem didapatkan hasil sebanyak 3 data komentar yang hasil klasifikasinya tidak sesuai dan sebanyak 86 data komentar diklasifikasikan sesuai kategori. Tingkat akurasi pengklasifikasian dapat diketahui berdasarkan perhitungan

sebagai berikut :

$$\text{Akurasi} = \frac{86}{89} \times 100\%$$

$$\text{Akurasi} = 97\%$$

Berdasarkan perhitungan dengan rumus akurasi pada *Confusion Matrix* maka didapatkan hasil akurasi sebesar 95,08 % untuk analisa sentimen komentar dan akurasi sebesar 97 % untuk proses pengklasifikasian komentar.

Tabel 1.
Hasil Pengujian Sistem

Uji No-	Data Uji	Data Training	Jumlah Analisa Sentimen	Jumlah Komentar Negatif	Jumlah Klasifikasi Benar	Tingkat Akurasi
1	264	45	251	89	66	74 %
2	264	83	251	89	71	89 %
3	264	490	251	89	88	97 %

Sumber: Hasil Olah Data

Berdasarkan hasil pengujian klasifikasi komentar yang dilakukan dengan menggunakan data komentar pada bulan Agustus sebanyak 264 komentar. Setelah dilakukan analisa sentimen pada komentar maka didapatkan sebanyak 251 komentar dianalisa dengan benar dan sebanyak 13 komentar hasil analisisnya tidak tepat. Setelah semua komentar dianalisa maka data yang digunakan untuk selanjtnya diklasifikasikan sebanyak 251 data komentar dengan sentimen negatif dan didapatkan data sebanyak 86 komentar diklasifikasikan dengan benar atau sama dengan data hasil klasifikasi yang dilakukan secara konvensional oleh staff administrasi dan sebanyak 3 data komentar yang tidak sama dengan data hasil klasifikasi yang dilakukan oleh staff administrasi.

Perbedaan hasil klasifikasi yang diperoleh dari pengolahan sistem, dengan data yang diolah secara manual, dominan dipengaruhi oleh jumlah kata yang terlalu sedikit dalam satu kalimat komentar. Jumlah kata yang sedikit tidak dapat dianalisa oleh sistem dengan benar. Pengaruh lain dalam analisis sentimen dan klasifikasi adalah bahasa yang digunakan untuk berkomentar serta kesalahan pengetikan. Komentar yang menggunakan bahasa indonesia dengan kurang baik dan benar akan dianalisa sistem sebagai galat. Kesalahan dalam proses preprocessing dan tokenizing dikarenakan kamus yang digunakan untuk proses analisisnya adalah kamus bahasa indonesia. Kata yang tidak tercantum di dalam kamus akan dibuang sehingga sangat berpengaruh

pada proses selanjutnya yang mengakibatkan hasil klasifikasi tidak sesuai.

Penentuan nilai k yang tidak tepat juga mempengaruhi tingkat akurasi klasifikasi yang dilakukan sistem. Saat ini belum ada aturan yang menetapkan cara untuk menentukan nilai k yang baik. Sehingga salah satu kelemahan dari metode k-nearest neighbor adalah adanya penentuan nilai k yang dapat mempengaruhi tingkat akurasi. Algoritma K-Nearest Neighbor adalah algoritma klasifikasi yang bekerja dengan menghitung kedekatan antara data uji dengan data training, oleh sebab itu semakin banyak jumlah data training yang digunakan maka semakin tinggi tingkat akurasi pengujian. Hasil pengujian berdasarkan jumlah data training yang berbeda-beda dapat di lihat pada tabel 1.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan dapat diambil beberapa kesimpulan, antara lain: Proses analisa sentimen pada komentar dilakukan dengan menghitung jumlah bobot sentimen positif dan negatif yang terdapat dalam setiap komentar. Kemudian Proses pengklasifikasian komentar dapat dilakukan dengan melalui tahap *pre-processing*, analisa sentimen, perhitungan bobot kata dengan TF-IDF dan menghitung jarak kedekatan antar data uji dengan data latih menggunakan *cosine similarity*. Pada Proses analisa komentar dengan sentiment analysis dapat menganalisa sentimen / opini yang terkandung di dalam setiap komentar dan dapat menentukan komentar yang bersifat positif dan negatif. Hasil Klasifikasi analisa sentimen pada komentar peserta diklat dengan menggunakan metode k-nearest neighbor didapatkan hasil dengan tingkat akurasi sebesar 97%. Pengujian Aplikasi untuk pengklasifikasian komentar melibatkan user secara langsung, sehingga proses perubahan dalam pengembangan aplikasi dapat dilakukan dengan cepat dan mendapatkan hasil yang sesuai dengan keinginan user.

DAFTAR PUSTAKA

Atina, V., Sedyono, E., & Rizal, R. (2017). Information Retrieval System for Indonesian Manuscript using Semantic Web. *International Journal of Computer Applications*. <https://doi.org/10.5120/ijca2017914>

930

- Indriati, I., and, A. R.-J. of E. E., & 2016, undefined. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwknn). *Jeest.Ub.Ac.Id*, 03(01), 23–32. <https://jeest.ub.ac.id/index.php/jeest/article/view/45>
- KBBI. (2018). *Pada KBBI Daring. Diambil 05 Mei 2022, dari https://kbbi.kemdikbud.go.id/entri/komentar*.
- Putri, P. (2013). *Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter Berbahasa Indonesia*. <http://repository.ub.ac.id/145872/>
- Riza, F., Rifai, S., Dirgantara, A., Sfenrianto, Rasenda, & Herdyansyah, S. (2020). Information Retrieval Technique for Indonesian PDF Document with Modified Stemming Porter Method Using PHP. *Journal of Physics: Conference Series*, 1477(3). <https://doi.org/10.1088/1742-6596/1477/3/032016>
- Sipayung, E., ... H. M.-J. J. S., & 2016, undefined. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Ejournal.Unsri.Ac.Id*, 8(1), 2355–4614. <https://ejournal.unsri.ac.id/index.php/jsi/article/view/3250>