

KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Zuhdiyyah Ulfah Siregar¹, Riki Ruli A. Siregar², Rakhmat Arianto³

¹Departemen Sistem Informasi Indonesia Power Jakarta

^{2,3}Departemen Teknik Informatika Sekolah Tinggi Teknik PLN Jakarta

E-mail : zuhdiyyah.ulfah@gmail.com

ABSTRACT

The process of analyzing and classifying comment data done by reading and sorting one by one negative comments and classifying them one by one using Ms. Excel not effective if the data to be processed in large quantities. Therefore, this study aims to apply sentiment analysis on comment data using K-Nearest Neighbor (KNN) method. The comment data used is the comments of the participants of the training on Udiklat Jakarta filled by each participant who followed the training. Furthermore, the comment data is processed by pre-processing, weighting the word using Term Frequency-Invers Document Frequency, calculating the similarity level between the training data and test data with cosine similarity. The process of applying sentiment analysis is done to determine whether the comment is positive or negative. Furthermore, these comments will be classified into four categories, namely: instructors, materials, facilities and infrastructure. The results of this study resulted in a system that can classify comment data automatically with an accuracy of 94.23%

Keywords : *Sentiment Analysis; Comments; TF-IDF; K-Nearest Neighbor; KNN*

ABSTRAK

Proses analisis dan pengklasifikasian data komentar yang dilakukan dengan cara membaca dan memilah satu per satu komentar yang bersifat negatif dan mengklasifikasikannya satu per satu dengan menggunakan Ms.Excel kurang efektif apabila data yang akan diproses dalam jumlah yang banyak. Oleh karena itu, penelitian ini bertujuan untuk menerapkan sentiment analysis pada data komentar menggunakan metode K-Nearest Neighbor (KNN). Data komentar yang digunakan yaitu komentar peserta diklat terhadap Udiklat Jakarta yang diisi oleh setiap peserta yang mengikuti diklat. Selanjutnya data komentar tersebut diproses dengan melakukan pre-processing, pembobotan kata menggunakan Term Frequence-Invers Document Frequence, menghitung tingkat kemiripan antara data latih dan data uji dengan cosine similarity. Proses penerapan sentiment analysis dilakukan untuk menentukan apakah komentar tersebut bersifat positif atau negatif. Selanjutnya komentar-komentar tersesebut akan diklasifikasikan kedalam empat kategori yaitu : instruktur, 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 : *Sentiment Analysis; Komentar; TF-IDF; K-Nearest Neighbor; KNN*

1. PENDAHULUAN

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 diklat, Udiklat Jakarta memerlukan masukan berupa komentar dari para peserta diklat sebagai bahan untuk evaluasi proses pembelajaran yang telah berjalan. Peserta diklat 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 peserta diklat terhadap proses diklat di Udiklat Jakarta.

Salah satu tantangan dalam pengolahan komentar yang diberikan yaitu proses analisisnya. Sistem yang menampung komentar yang digunakan oleh Udiklat Jakarta hanya sebatas untuk menampung komentar peserta diklat. 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 komentar yang menyatakan ketiaktepatan terhadap Udiklat Jakarta diklasifikasikan satu per satu kedalam empat kategori (kategori instruktur, 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 *sentiment analysis* untuk proses analisa apakah komentar tersebut bersifat positif atau negatif dan untuk proses pengklasifikasian komentar digunakan algoritma *k-nearest neighbor*. dengan menerapkan *Sentiment Analysis* menggunakan metode *K-Nearest Neighbor (KNN)* dapat mempermudah dalam proses pengolahannya. Komentar akan diekstrak dan dianalisis secara otomatis untuk kemudian ditentukan dan diklasifikasikan informasi opini yang terkandung di dalamnya.

1.1 *Sentiment Analysis*

Sentiment Analysis adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Jika diberikan satu set dokumen teks D yang berisi opini (atau sentimen) mengenai suatu objek, maka *opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen dan untuk menentukan apakah komentar tersebut positif atau negatif [3]

1.2 *Text Preprocessing*

Proses perubahan bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam *data mining*, yang biasanya akan menjadi nilai-nilai numerik. Proses ini sering disebut *Text Preprocessing* [2]. Setelah data menjadi data terstruktur dan berupa nilai numerik maka data dapat dijadikan sebagai sumber data yang dapat diolah lebih lanjut. Beberapa proses yang dilakukan adalah sebagai berikut:

a. *Case Folding*

Case Folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter [2].

b. *Tokenizing*

Tokenizing yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), tanda kutip ("), tanda kurung (()), spasi dan karakter angka yang ada pada kata tersebut [7].

c. *Stopword removal* atau *Filtering*

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang. Contoh *stopword* adalah "yang", "dan", "di", dan lain-lain. [2]

d. *Stemming*

Setelah melalui proses stopword removal tindakan selanjutnya yaitu proses stemming. Stemming adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (stem) [6].

1.3 TF-IDF (Term Frequency-Inverse Document Frequency)

Salah satu fungsi penting yang disediakan oleh tahapan proses ini adalah untuk dapat memilih *term* atau kata apa saja yang dapat dijadikan wakil penting untuk kumpulan dokumen yang akan dianalisis dengan kata lain kita melakukan pembobotan terhadap setiap *term*.

Metode TF-IDF ini merupakan metode pembobotan dalam bentuk sebuah metode yang merupakan integrasi antar term frequency (TF), dan *inverse document frequency* (IDF). Metode TF-IDF dapat dirumuskan sebagai berikut :

$$w(t,d) = tf(t,d) * idf \tag{1}$$

$$idf = \log(N/df) \tag{2}$$

Dimana, $tf(t,d)$ adalah kemunculan kata t pada dokumen d , N adalah jumlah dokumen pada kumpulan dokumen. Dan df = jumlah dokumen yang mengandung term t .

1.4 Cosine Similarity

Cosine Similarity Metode Cosine Similarity adalah metode untuk menghitung kesamaan antara dua dokumen. Penentuan kesesuaian dokumen dengan *query* dipandang sebagai pengukuran (*similarity measure*) antara *vector* dokumen (D) dengan *vector query* (Q). Semakin sama suatu *vector* dokumen dengan *vector query* maka dokumen dapat dipandang semakin sesuai dengan *query* [8]. Rumus yang digunakan untuk menghitung *cosine similarity* adalah sebagai berikut:

$$cosSim(A,B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{(\sum_{i=1}^n A_i^2)} \sqrt{(\sum_{i=1}^n B_i^2)}} \tag{3}$$

Dimana, A adalah dokumen Uji, B adalah dokumen Latih. A_i dan B_i adalah nilai bobot yang diberikan pada setiap term pada dokumen. Kedekatan *query* dan dokumen diindikasikan dengan sudut yang dibentuk. Nilai cosinus yang cenderung besar mengindikasikan bahwa dokumen cenderung sesuai *query*. Dalam proses membandingkan dokumen yang sesuai dengan dokumen yang telah ada atau dokumen lainnya, maka digunakan perhitungan dengan rumus pada persamaan (3) untuk mengetahui angka similaritas dari dokumen tersebut [4], jika angka similaritas semakin tinggi, maka tingkat kemiripan antara dokumen test dan dokumen training semakin besar [5].

1.5 K-Nearest Neighbor (KNN)

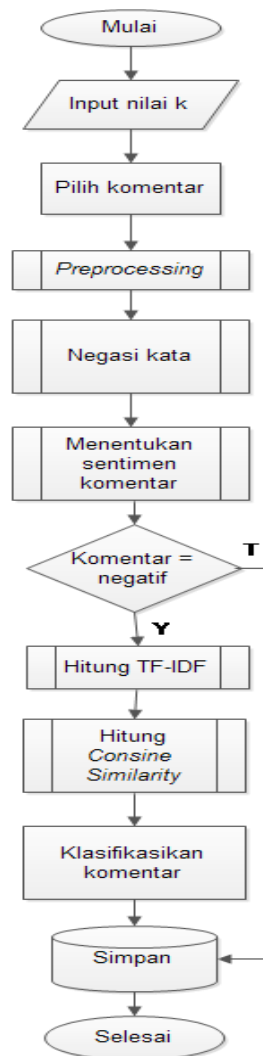
K-Nearest Neighbor merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap vektor data pembelajaran dihitung, dan diambil sejumlah K yang paling mendekati. Titik yang baru klasifikasinya diprediksi termasuk pada klasifikasi terbanyak dari titik-titik tersebut [2].

Debat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*. Untuk pengklasifikasian text, penentuan *Euclidean Distance*-nya menggunakan *cosine similarity*. Semakin besar nilai *Euclidean Distance*-nya akan semakin jauh tingkat kemiripan antara data uji dan dokumen training-nya dan sebaliknya jika nilai *Euclidean Distance*-nya semakin kecil maka akan semakin dekat tingkat kemiripan antara data uji dan dokumen trainingnya. Namun dalam pengklasifikasian text, semakin besar nilai *cosine similarity*-nya akan semakin dekat tingkat kemiripan antara data uji dan data latihnya dan sebaliknya jika nilai *cosine similarity*-nya semakin kecil maka akan semakin jauh tingkat kemiripan antara data uji dan data latihnya [5-8].

2. METODOLOGI PENELITIAN

Sistem yang akan dibuat berbasis *web application* yang dibangun dengan bahasa pemrograman PHP dan *database* MYSQL.

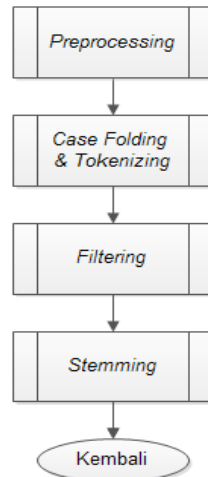
Untuk proses pengolahan datanya akan diterapkan *sentiment analysis* untuk menganalisa data apakah komentar bersifat negatif atau positif, kemudian hasil dari hasil analisa maka komentar yang bersifat negatif akan diklasifikasikan dengan menerapkan algoritma *k-nearest neighbor* untuk pengklasifikasian datanya. 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 *consine similarity* digunakan untuk menghitung jarak kedekatan/kemiripan dokumen latih dengan dokumen uji. Proses secara keseluruhan pada sistem yang diusulkan dapat dilihat pada diagram alir berikut :



Gambar 1 Proses Keseluruhan Sistem Usulan

2.1 Tahap *Preprocessing*

Pada tahap ini akan dilakukan proses *case folding*, *tokenizing*, *stopword removal / filtering* dan *stemming* pada semua dokumen.

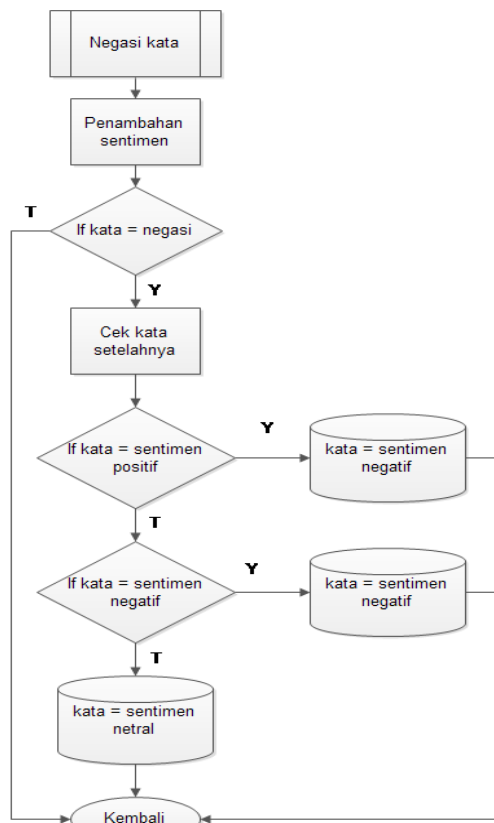


Gambar 2 Proses Preprocessing

2.2 Analisa Sentiment

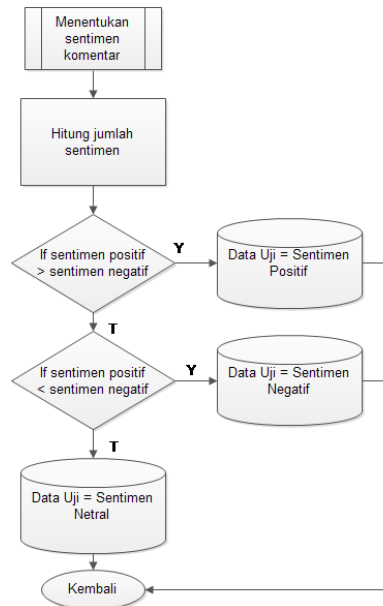
Pada tahap ini akan dilakukan analisa sentimen pada dokumen uji. Langkah pertama dalam menganalisa sentimen adalah melakukan proses negasi teks. Proses negasi teks ini diperlukan proses *load dictionary*, dimana dibutuhkan kamus kata kunci sentimen positif, dan kamus kata kunci sentimen negatif.

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.



Gambar 3 Flowchart Negasi Text

Setelah setiap kata diketahui sentimennya maka selanjutnya yaitu menghitung jumlah setimen positif dan sentimen negatif pada data uji. Jika sentimen positif > sentimen negatif maka dokumen uji termasuk komentar dengan sentimen positif, dan sebaliknya apabila sentimen positif < sentimen negatif maka dokumen uji termasuk komentar dengan sentimen negatif.



Gambar 4 Flowchart Penentuan Sentimen Data Uji

2.3 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.

2.4 K-Nearest Neighbor

Algoritma *k-nearest neighbor* merupakan algoritma yang menggunakan kedekatan/kemiripan antar dokumen dalam proses pengklasifikasiannya. Dalam penelitian ini digunakan perhitungan *cosine similarity* untuk menghitung jarak kedekatan antar dokumennya.

3. HASIL DAN PEMBAHASAN

Berdasarkan perancangan aplikasi Pengklasifikasian Data Komentar, secara garis besar aplikasi ini hanya mempunyai dua user, yaitu Admin (*staff* bagian pengembangan, evaluasi dan mutu (PEM)) dan Peserta Diklat. 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 Diklat hanya bisa mengisi kolom komentar dan tidak memiliki hak akses untuk mengakses menu lainnya.

3.1 Hasil Rancangan

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



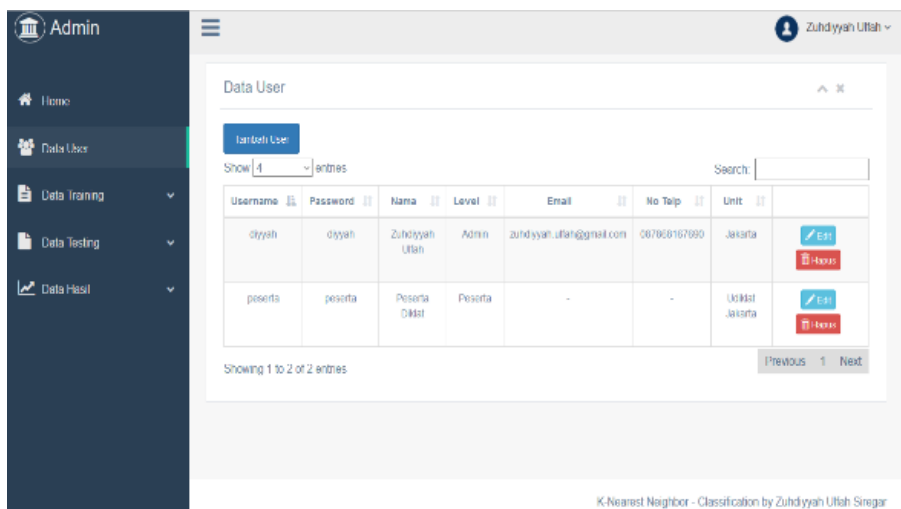
Login

Level Pengguna

Gambar 5 Interface Form Login

1. Form Data User

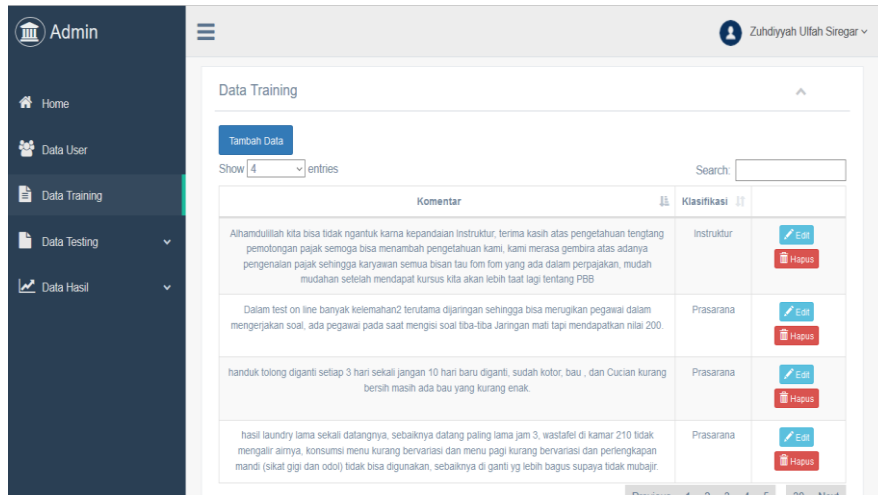
Halaman data *user* hanya dapat diakses oleh pengguna yang login ke dalam sistem aplikasi sebagai *admin*. Halaman data *user* ini menampilkan semua data *user* yang tersimpan di dalam *database* dibuat dengan maksud untuk memudahkan *administrator sistem* dalam mengelola data pengguna, baik membuat pengguna baru, mengedit pengguna lama, menghapus pengguna maupun melakukan pencarian data pengguna sebagaimana ditampilkan berikut ini:



Gambar 6 Interface Form Data Pengguna

2. Form Data Training

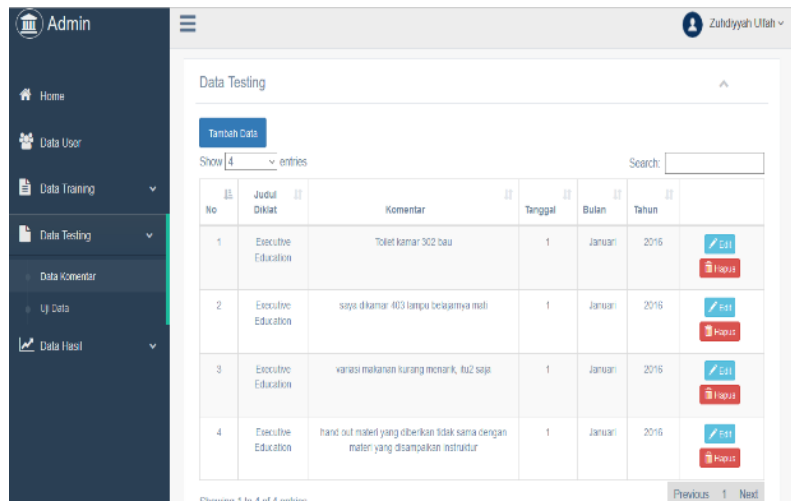
Halaman data *training* menampilkan data-data komentar yang telah diklasifikasikan sebelumnya. Pada halaman ini *admin* dapat menambahkan data *training* baru, mengedit data *training* dan menghapus data *training* yang sudah ada.



Gambar 7 Interface Form Data Training

3. Form Data Testing

Halaman data *testing* merupakan halaman yang menampung semua komentar yang telah di input oleh peserta diklat. 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.

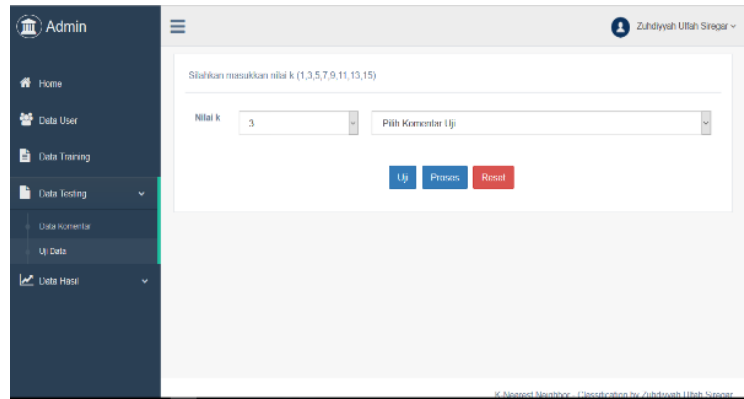


Gambar 8 Interface Form Data Testing

4. Form Uji Data

Halaman uji data merupakan halaman digunakan untuk memilih dan memproses data uji. Pada halaman ini *admin* diharuskan untuk menginput nilai k dan memilih komentar yang akan di uji. Terdapat tiga tombol yaitu tombol uji, tombol proses dan tombol reset.

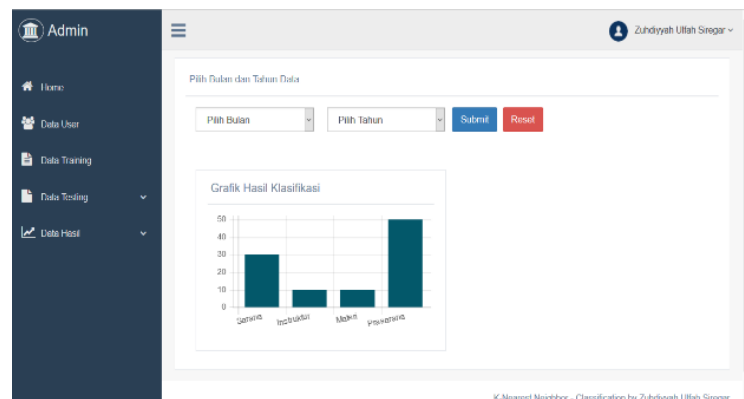
Setelah *admin* menginput nilai k dan memilih komentar yang akan di uji maka *admin* menekan tombol uji. Tombol uji ini berfungsi untuk mendeklarasikan nilai k dan id komentar uji kedalam *query* untuk diproses. Selanjutnya *admin* menekan tombol proses yang berfungsi untuk memproses komentar untuk di klasifikasikan. Tombol reset digunakan untuk mengembalikan ke kondisi awal.



Gambar 9 Halaman Uji Data

5. Form Grafik

Pada halaman ini berisi grafik yang menampilkan persentase hasil klasifikasi pada tiap kategorinya (sarana, intruktur, materi dan pelayanan). Data pada grafik merupakan data yang diklasifikasikan pada setiap bulan.



Gambar 10 Interface Form Grafik

6. Form Hasil Klasifikasi

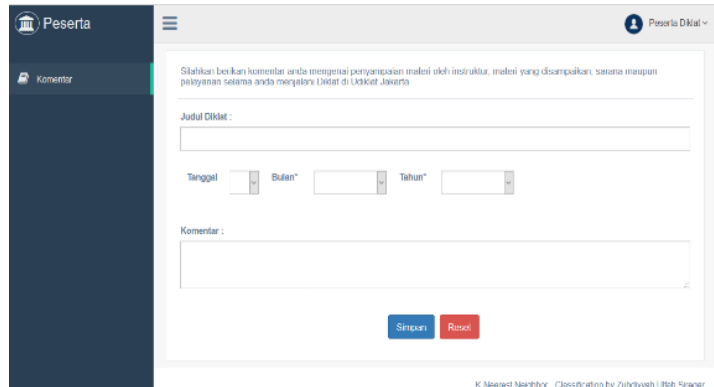
Halaman form hasil klasifikasi merupakan halaman yang menampilkan hasil klasifikasi data yang di uji. Hasil klasifikasi yang ditampilkan adalah hasil klasifikasi data uji dengan sentimen negatif.

Komentar	Kategori	Bulan	Tahun	Sentimen
materi yang disampaikan sudah pernah dibahas sewaktu pembelajaran, mohon materinya di update (?)	Materi	Januari	2016	Negatif
Pendingin kamar (AC) tidak dingin dan berisara berisik, perlu penggantian AC yg lebih efisien	Pasarsana	Januari	2016	Negatif

Gambar 11 Halaman Hasil Klasifikasi

7. Form Pengisian Komentar

Halaman pengisian komentar ini dikhususkan untuk peserta diklat. Pada halaman ini peserta diklat dapat mengisi komentar mengenai Udiklat Jakarta selama menjalani pendidikan dan pelatihan di Udiklat Jakarta. Peserta diklat wajib mengisi judul diklat yang di ikuti beserta tanggal, bulan dan tahun pada saat mengisi komentar.



Gambar 11 Interface Pengisian Komentar

3.2 Hasil Pengujian

Pengujian metode yang dilakukan pada system klasifikasi komentar ini, dilakukan dengan menguji 64 data komentar pada bulan Agustus. Tingkat akurasi pengujian sistem dapat dilakukan dengan menggunakan rumus akurasi pada *confusion matrix* sebagai berikut :

$$Akurasi = \frac{jumlah\ nilai\ yang\ benar}{jumlah\ data\ keseluruhan} \times 100\% \tag{4}$$

Setelah dilakukan pengujian, didapatkan hasil analisa sentimen komentar yang benar sebanyak 59 komentar. Berdasarkan rumus diatas maka tingkat akurasi analisa sentimen adalah sebagai berikut :

$$Akurasi = \frac{59}{64} \times 100$$

$$Akurasi = 92.18 \%$$

Setelah dilakukan analisa sentimen pada komentar, didapatkan sebanyak 52 komentar bersifat negatif yang akan di proses untuk diklasifikasikan. Setelah dilakukan proses pengklasifikasian oleh sistem didapatkan hasil sebanyak 4 data komentar yang hasil klasifikasinya tidak sesuai dan sebanyak 48 data komentar diklasifikasikan sesuai kategori. Tingkat akurasi pengklasifikasian dapat diketahui berdasarkan perhitungan sebagai berikut :

$$Akurasi = \frac{49}{52} \times 100$$

$$Akurasi = 94.23 \%$$

Berdasarkan dengan rumus diatas maka didapatkan hasil akurasi sebesar 92,18 % untuk analisa sentimen komentar dan akurasi sebesar 94,23 % untuk proses pengklasifikasian komentar.

Tabel 1 Hasil Pengujian Sistem

Uji ke-	Data Uji	Data Training	Jumlah Analisa Sentimen Benar	Jumlah Komentar Negatif	Jumlah Klasifikasi Benar	Tingkat Akurasi
1	64	11	59	52	38	73,07 %
2	64	20	59	52	41	78,84 %
3	64	120	59	52	49	94,23 %

3.3 Pembahasan

Berdasarkan hasil pengujian klasifikasi komentar yang dilakukan dengan menggunakan data komentar pada bulan Agustus sebanyak 64 komentar. Setelah dilakukan analisa sentimen pada komentar maka didapatkan sebanyak 59 komentar dianalisa dengan benar dan sebanyak 5 komentar hasil analisisnya tidak tepat. Setelah semua komentar dianalisa maka data yang digunakan untuk selanjutnya diklasifikasikan sebanyak 52 data komentar dengan sentimen negatif dan didapatkan data sebanyak 49 komentar diklasifikasikan dengan benar atau sama dengan data hasil klasifikasi yang dilakukan secara konvensional oleh staff bagian PEM dan sebanyak 3 data komentar yang tidak sama dengan data hasil klasifikasi yang dilakukan oleh staff bagian PEM.

Ketidaksamaan hasil klasifikasi oleh sistem dengan data yang ada sebagian besar dipengaruhi oleh jumlah kata yang terlalu sedikit dalam satu kalimat komentar. Jumlah kata yang sedikit tidak dapat dianalisa oleh sistem dengan benar. Penggunaan bahasa dalam kalimat komentar dan kesalahan pengetikan juga sangat berpengaruh dalam proses analisa dan pengklasifikasiannya. Komentar yang tidak menggunakan bahasa indonesia yang baik dan benar akan dianalisa sistem sebagai *noise*, hal ini dikarenakan kamus yang digunakan untuk proses *pre-processing* 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 pengklasifikasian, namun 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 pengklasifikasian yang menghitung kedekatan antara data uji dengan data *training*, oleh sebab itu semakin banyak jumlah data *training* yang digunakan maka semakin tinggi tingkat akurasi pengujiannya. 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 94,23 %. 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

- [1] Dave, K. L. (2003). Mining the Peanut Galery: Opinion Extraction and Semantic Classification of Product Reviews. *Word Wide Web Conference*.
- [2]. Feldman, R. d. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. New York.: Cambridge University Press.
- [3]. Liu, B. (2015). *Sentiment Analysis and Opinion Mining*. Chicago: Morgan & Claypool Publisher.

- [4] Luhulima, Yugo Yudasha, Marji, & Lailil Muflikhah. 2015. Sentiment Analysis Pada review Barang Berbahasa Indonesia Dengan Metode K-Nearest Neighbor (K-NN). Universitas Brawijaya Malang.
- [5] Putri, P. A., Ridok, A., & Indriati. (n.d.). Implementasi Metode Improved K-Nearest Neighbor Pada Analisis Sentimen Twitter Berbahasa Indonesia.
- [6] Tala, F. Z. 2003. A Study Of Stemming Effects On Information Retrieval in Bahasa Indonesia. The Netherlands: Universiteit van Amsterdam
- [7] Weiss, S. M., Indurkha, N., Zhang, T., & Damerau, F. J. 2005. Text Mining: Predictive Methods fo Analyzing Unstructured Information. New York: Springer.
- [8] Siregar, R. R. A., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Computatio: Journal of Computer Science and Information Systems*, 1(2), 171-186.