

Implementasi Metode *Naive Bayes Classifier* (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)

*Naomi Chatrina Siregar*¹; *Riki Ruli A. Siregar*²; *M. Yoga Distra Sudirman*³

^{1,2,3}Program Studi Teknik Informatika, Institut Teknologi PLN Jakarta

¹ naaomicsiregar7@gmail.com

² riki.ruli@itpln.ac.id

³yogadistra@itpln.ac.id

ABSTRACT

Distance learning systems are learning systems that are widely implemented in different countries. One of them is Indonesia and has been applied at various levels of education ranging from the basic level to the top level. The learning is different from the regular learning system that is a face-to-face learning system that focuses on meeting. In its application, this learning system reaped many comments such as the constraints of lack of understanding of the material submitted and so on. Of the many comments will certainly take a long time in managing the data. Therefore, there is a system that can manage these comments, the data used is comment data about the distance learning system. Then the data will be processed through the preprocessing text stage and classified using the Naïve Bayes Classifier method to select comments that fall into positive, negative, or neutral categories.

Keywords: *Text Mining; TF-IDF Weighting; Naïve Bayes Classifier*

ABSTRAK

Pelaksanaan pembelajaran jarak jauh adalah pelaksanaan pembelajaran yang banyak diterapkan di berbagai negara. Salah satunya adalah Indonesia dan telah diterapkan di berbagai tingkatan pendidikan mulai dari tingkatan dasar hingga tingkatan atas. Pembelajaran tersebut berbeda dengan pelaksanaan pembelajaran reguler yaitu pelaksanaan pembelajaran tatap muka yang berfokus pada adanya pertemuan. Dalam penerapannya pelaksanaan pembelajaran ini menuai banyak komentar seperti kendala kurangnya pemahaman terhadap materi yang disampaikan dan sebagainya. Dari banyaknya komentar tentunya akan membutuhkan waktu yang lama dalam mengelola data. Untuk itu, diperlukan adanya sistem yang dapat mengelola komentar tersebut, data yang digunakan adalah data komentar tentang pelaksanaan pembelajaran jarak jauh. Kemudian data akan diproses melalui tahapan preprocessing text dan akan diklasifikasi menggunakan metode Naïve Bayes Classifier untuk menentukan komentar termasuk dalam kategori positif, negatif, atau netral.

Kata kunci: *Text Mining, Pembobotan TF-IDF, Naïve Bayes Classifier*

1. PENDAHULUAN

Sejak meluasnya pandemi yang disebabkan oleh virus Corona di Indonesia, beragam usaha yang telah dilakukan oleh pemerintah untuk mencegah penyebaran covid-19. Salah satunya adalah dengan menerbitkan surat edaran mengenai arahan tentang kegiatan proses belajar mengajar di rumah dengan tujuan untuk melindungi warga sekolah dari paparan Covid-19. Sekolah Dasar Negeri 060969 adalah salah satu institusi pendidikan yang juga menerapkan kebijakan pelaksanaan pembelajaran jarak jauh kerap kali menerima banyak komentar. Komentar yang diberikan terdiri dari komentar positif, negatif, dan netral. Dari komentar tersebut maka dapat digunakan untuk perbaikan terhadap proses pembelajaran dan mengetahui tingkat kepuasan yang dinyatakan dalam bentuk kategori positif, negatif, dan netral.

Namun, komentar-komentar yang diberikan sebelumnya belum terkelola dengan baik. cara pengolahan data komentar masih dilakukan dengan cara manual yaitu dengan memilah komentar satu demi satu dan menyusunnya berdasarkan kelasnya, lalu mengkategorikannya apakah komentar termasuk dalam kategori komentar negatif yang berarti menyatakan ketidakpuasan atau komentar yang bersifat positif yang menyatakan kepuasan terhadap pelaksanaan pembelajaran. hal tersebut kurang efektif apabila data dalam jumlah banyak sehingga akan sulit dalam mengklasifikasikannya Untuk memudahkan proses pengolahan data komentar maka diterapkan proses pengklasifikasian komentar menggunakan algoritma *Naïve Bayes Classifier* untuk mengetahui apakah komentar tersebut bersifat positif atau negatif.

1.1 *Text Mining*

Text Mining adalah penambangan data berupa teks, yang memiliki tahapan proses yang lebih banyak karena data teks. Teks merupakan data yang memiliki informasi yang tidak terstruktur sehingga tahapan prosesnya lebih banyak dengan tujuan agar data menjadi lebih terstruktur yang dapat diperoleh melalui pola-pola yang ada pada teks[3]. Tujuan dari *text mining* adalah untuk menganalisa pandangan individu, evaluasi, sikap, dan emosi terhadap orang, dan lain sebagainya agar dapat mengetahui apakah berkenaan dengan suatu topik, layanan, ataupun bidang lainnya[4].

1.2 *Text Preprocessing*

Text Preprocessing adalah proses menggali informasi yang terdapat dalam suatu teks dan mengubahnya menjadi data terstruktur dalam bentuk nilai numerik dengan melewati berbagai tahapan sebagai berikut[2]:

a. *Case Folding*

Case Folding adalah tahap perubahan kata dalam dokumen dari yang memiliki huruf kapital menjadi huruf kecil dan yang akan diproses hanya huruf „a“ sampai dengan „z“, selain dari itu maka akan dihilangkan[2].

b. *Tokenizing*

Tokenizing adalah tahap pemecahan kalimat dan menghilangkan tanda baca seperti tanda titik, koma, dsb, kalimat yang akan dipecah adalah kalimat dari hasil *case folding* [2]. c. *Filtering* *filtering* adalah tahap dalam mengambil kata-kata penting dari hasil *tokenizing* yang dibandingkan dengan kamus data yang digunakan. Contohnya seperti “di”, “ke”, dsb [2]. d. *Stemming*

stemming adalah tahap perubahan kata yang berimbuhan baik itu awalalan, sisipan ataupun akhiran menjadi kata dasar yang didapatkan dari hasil *filtering* [2].

1.3 Pembobotan TF-IDF

Pembobotan TF-IDF adalah salah satu model yang digunakan untuk menghitung bobot kata pada suatu dokumen dengan menggunakan model *term frequency* untuk menghitung jumlah keseluruhan kata yang ada pada dokumen dan *inverse document frequency (idf)* untuk menghitung kata yang muncul di setiap dokumen yang dapat dirumuskan sebagai berikut[2] :

$$W(t,d) = tf(t,d) * idf \quad (1)$$

$$idf = \log (N / df) \quad (2)$$

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

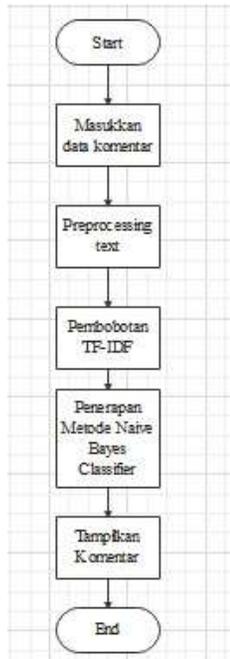
1.4 Naïve Bayes Classifier

Naïve Bayes adalah salah satu metode pengklasifikasian sederhana yang sering digunakan karena mudah diterapkan dan memiliki hasil yang baik saat diterapkan pada banyak kasus. Sedangkan kelemahan dari metode ini sendiri yaitu adanya asumsi atau dengan kata lain kondisi kelas saling bebas , sehingga kurang akurat. Adapun pada prakteknya , kebergantungan ada diantara variabel. Misalnya rumah sakit: pasien, umur, keluarga ,dsb. Kebergantungan diantara variabel ini tidak dapat dimodelkan dengan menggunakan Naïve Bayesian Classifier [6]. Rumus dari teorema bayes akan disajikan pada persamaan berikut [1]

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (3)$$

2. METODE PENELITIAN

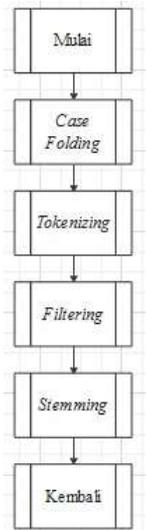
Tahap ini adalah tahap yang menggambarkan keseluruhan proses yang dilakukan dengan tujuan dapat menghasilkan suatu sistem berbasis web yang dapat mengklasifikasikan komentar ke dalam kategori positif,negati, atau netral. Langkah pertama saat komentar dimasukkan ke dalam sistem, kemudian data tersebut akan melalui rangkaian proses *processing Text* yang dimulai dari tahap *case folding, tokenizing, filtering, dan stemming*. Setelah data telah melewati tahap *processing text* , lalu data hasil *processing text* yaitu *hasil stemming* akan dihitung kembali bobot masing-masing kata secara keseluruhan melalui proses pembobotan *tf*, kemudian dari data tersebut diterapkan algoritma *Naïve Bayes Classifier (NBC)* untuk mengklasifikasikan komentar, setelah proses perhitungan data selesai, maka hasil akhir perhitungan akan ditampilkan pada sistem yaitu hasil klasifikasi data komentar termasuk dalam kategori positif, negatif, atau netral yang dapat digambarkan melalui gambar berikut :



Gambar 1 Alur Analisa Keseluruhan Sistem Usulan

2.1 Tahap *Text Preprocessing*

Tahap ini merupakan proses pengolahan teks, teks yang akan diolah dalam penelitian ini berupa komentar yang akan diproses melalui tahap seperti *case folding*, *tokenizing*, *filtering*, dan *stemming* yang digambarkan pada gambar berikut.



Gambar 2 Tahapan *Text Preprocessing*

2.2 Pembobotan TF-IDF

Data yang sudah diproses sebelumnya pada proses *Text Preprocessing* akan digunakan untuk diolah kembali. Data yang digunakan adalah hasil dari proses *stemming*, data tersebut nantinya akan dilakukan perhitungan pembobotan pada setiap kata yang terdapat dalam dokumen. Data yang dihasilkan dari proses ini berupa bobot dari setiap kata dan akan digunakan kembali untuk

melakukan proses klasifikasi menggunakan metode *Naïve Bayes Classifier* ke dalam kategori positif, negatif, dan netral .

2.3 *Naïve Bayes Classifier*

Algoritma *Naïve Bayes Classifier* adalah algoritma yang sering digunakan untuk mengklasifikasikan data dengan menggunakan perhitungan pada probabilitas. Dari perhitungan probabilitas pada setiap kata dalam dokumen akan dihasilkan klasifikasi kategori positif, negatif, dan netral yang diperoleh dari proses perhitungan sebelumnya yaitu pembobotan tf-idf.

3. HASIL DAN PEMBAHASAN

Dari hasil perancangan yang telah dilakukan, maka dihasilkan suatu sistem yang dapat digunakan oleh dua pengguna yaitu dari sisi pengguna yang dalam hal ini adalah orang tua siswa, siswa dan guru. Lalu sisi pengguna lainnya adalah admin yang dapat menjalankan proses dalam sistem seperti menambah, mengedit, mencetak data, dan lain sebagainya.

3.1 Hasil Rancangan

Dalam proses klasifikasi data komentar pada sistem yang telah dibuat diawali dengan tampilan halaman *login* yang dapat diakses oleh *admin* dan pengguna dengan memasukkan *username* dan *password*. Apabila *username* dan *password* terverifikasi dengan benar maka dapat masuk ke dalam tampilan berikutnya pada sistem.



Gambar 3 Tampilan Halaman *Login*

1. Halaman *Data User*

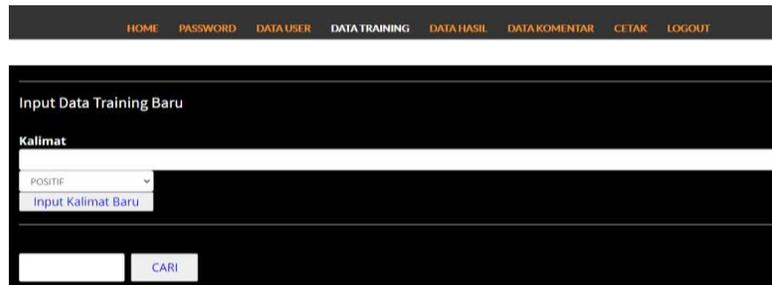
Pada Halaman ini saat *user* berhasil memasukkan *username* dan *password* yang benar maka akan tampilan menu awal seperti data *training*, data komentar, dan data hasil. Halaman ini dibuat dengan tujuan agar dapat memudahkan *administrator* dalam mengelola data pengguna dalam melakukan pencarian data pengguna yang ditampilkan sebagai berikut:



Gambar 4 Tampilan Halaman Data *User*

2. Halaman *Data Training*

Pada halaman ini *admin* dapat menambahkan data *training* baru dengan memasukkan komentar ke dalam kolom komentar dan melakukan pencarian data *training* yang dapat digambarkan sebagai berikut:



Gambar 5 Tampilan Halaman Data *Training*

3. Halaman *Data Komentar*

Halaman ini merupakan halaman untuk menguji data dan menampilkan hasil dari klasifikasi komentar apakah termasuk ke dalam kategori positif, negatif dan netral.



Gambar 6 Menu Halaman Data *Komentar*

4. Halaman *Data Hasil*

Halaman ini adalah halaman yang akan menampilkan hasil klasifikasi dari data komentar yang terdapat pada sistem.



Gambar 7 Menu Halaman Data Hasil

5. Halaman Cetak

Halaman cetak adalah halaman yang akan menampilkan halaman berupa laporan dari komentarkomentar dalam sistem dalam bentuk *pdf*.

model pembelajaran disertai contoh memberikan pemahaman tentang topik yang dibahas
sangat efisien jadi tidak perlu antar jemput sekolah sehingga biaya jadi berkurang
Proses belajar mengajar terasa menyenangkan karena banyak menggunakan media sehingga lebih mudah mengerti
Sistem belajarnya hanya membuat anak-anak semakin bosan di rumah
belajarnya kok jadi semakin sulit
murid selalu antusias ketika belajar melalui tayangan video
pembelajarannya membosankan, anak-anak merasa jenuh belajar
semoga belajarnya jadi lebih baik
arahannya bagus dan jelas jadi mudah mengerti

Gambar 8 Menu Halaman Cetak

3.1 Hasil Pengujian

Pengujian akurasi dilakukan dengan menguji data komentar sebanyak 50 data komentar dengan menggunakan metode *confusion matrix* yang menggunakan rumus sebagai berikut :

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

Keterangan :

- Akurasi = Nilai akurasi aplikasi
- Nilai benar = Jumlah data uji yang bernilai benar
- Jumlah data = Jumlah data yang di uji

Setelah dilakukan pengujian yang dilakukan dengan menguji data *training* sebanyak 120 data dan digunakan data komentar sebagai data *testing* sebanyak 50. Berdasarkan rumusan diatas diperoleh tingkat akurasi sebagai berikut :

$$\begin{aligned} \text{Akurasi} &= \frac{34}{50} \times 100\% \\ &= 68 \%. \end{aligned}$$

3.2 Pembahasan

Dari hasil pengujian yang telah dilakukan dengan menentukan kategori komentar terhadap pelaksanaan pembelajaran di SDN 060969, didapatkan data sebanyak 170 data yang terdiri dari 120 data untuk digunakan sebagai data *training* dan 50 data digunakan sebagai data uji. Namun, data jumlah data uji berbeda setiap kategorinya, sehingga sistem akan memproses menjadi 3 kategori yaitu positif, negatif, dan netral.

Pada saat sistem memproses data uji terdapat perbedaan hasil yang diperoleh dengan klasifikasi yang dilakukan secara manual. Hal tersebut disebabkan oleh data yang digunakan. Dimana pada penelitian ini data yang digunakan berisikan komentar dengan jumlah kata yang

terlalu sedikit dalam satu kalimat komentar. Hal tersebut membuat sistem tidak dapat melakukan analisa dengan benar karena penggunaan bahasa dalam kalimat komentar cenderung tidak sesuai dengan kaidah bahasa indonesia yang baik dan benar dan juga terdapat kesalahan dalam proses mengetikkan komentar untuk dimasukkan ke dalam sistem yang sangat berpengaruh terhadap proses analisa dan proses klasifikasi dalam sistem. Komentar yang tidak sesuai dengan kaidah bahasa indonesia yang baik dan benar atau dengan kata lain tidak sesuai dengan tata bahasa yang ada akan dianalisa oleh sistem sebagai *noise*, hal tersebut terjadi karena kamus yang digunakan untuk proses *pre-processing* adalah kamus bahasa indonesia. Kata yang tidak terdapat dalam kamus yang tersimpan dalam *database* akan dihilangkan sehingga mengakibatkan kata dalam dokumen yang akan diproses berubah. Perubahan tersebut berdampak pada proses pembobotan kata dalam dokumen yang mengakibatkan hasil klasifikasi tidak sesuai.

Algoritma *Naïve Bayes* adalah algoritma yang dapat digunakan dalam mengklasifikasi dengan cara memprediksi kemungkinan/ probabilitas yang ada di masa depan berdasarkan pada data yang diperoleh di masa sebelumnya dan memiliki hasil yang baik saat diterapkan pada banyak kasus dengan jumlah data yang banyak sehingga bila semakin banyak jumlah data *training* yang digunakan maka hasil klasifikasi yang didapatkan akan semakin tinggi pula tingkat akurasi.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan sebelumnya maka dapat ditarik beberapa kesimpulan yaitu bahwa sistem yang dibangun dapat membantu dalam menganalisis komentar dengan lebih tepat dibandingkan dilakukan secara manual. Adapun dalam proses klasifikasi komentar mengenai pelaksanaan pembelajaran jarak jauh dimulai dengan melakukan proses *preprocessing text*, proses ini adalah pengolahan teks. Diawali dengan membaca teks yang akan dilatih dan mengubahnya menjadi kata dalam bentuk tunggal. Bentuk kata yang telah diubah kemudian akan diproses untuk mendapatkan kata yang penting dengan cara menghilangkan kata yang tidak terdapat pada kamus data yang tersimpan dalam *database* dan setelah kata didapatkan selanjutnya adalah menghilangkan tanda baca. Selanjutnya dilakukan proses perhitungan pembobotan TF-IDF lalu melakukan perhitungan *Naïve Bayes Classifier*. Untuk menentukan kategori dilakukan pengujian terhadap 50 data komentar dengan menggunakan algoritma *Naïve Bayes Classifier* didapatkan bahwa hasil nilai akurasi yaitu sebesar 68%.

4.2 Saran

Dalam pengembangan selanjutnya, agar dapat melakukan penambahan data *training* (data latih) dalam jumlah yang lebih banyak dengan tujuan menghasilkan tingkat akurasi yang lebih tinggi karena semakin banyak data latih yang digunakan maka akan semakin akurat pula hasil klasifikasi dan dalam pengolahan komentar dapat menggunakan bahasa asing seperti bahasa inggris dan lainnya.

DAFTAR PUSTAKA

- [1] Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *Jurnal Teknik Informatika Musirawas (JUTIM)*, 3(1), 1. <https://doi.org/10.32767/jutim.v3i1.286>
- [2] Feldman, R. &. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- [3] Jiawei, H.(2012). *Data Mining: Concepts and Techniques*. *Morgan Kaufmann*.

- [4] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- [5] Wandathira. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram Dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4704-4713.
- [6] Wilianto, L. P. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. *SNATIF*.
- [7] Budiana ND, Siregar RR, Susanti MN. Penetapan Instruktur Diklat Menggunakan Metode Clustering K-Means dan Topsis Pada PT PLN (Persero) Udiklat Jakarta. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*. 2019 Aug 6;12(2):111-21.
- [8] Prianty AF, Siregar RR, Arianto R. Penanganan Gangguan Listrik Rumah Tangga Menggunakan Algoritma Greedy Untuk Penentuan Jarak Optimal. *JURNAL TEKNOLOGIA*. 2019 Aug 9;2(1).
- [9] Siregar RR, Sinaga FA, Arianto R. Aplikasi penentuan dosen penguji skripsi menggunakan metode TF-IDF dan vector space model. *Computatio: Journal of Computer Science and Information Systems*. 2017 Nov 1;1(2):171-86.
- [10] Siregar RR, Siregar ZU, Arianto R. Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. *KILAT*. 2019 May 17;8(1).
- [11] Prasatya A, Siregar RR, Arianto R. Penerapan Metode K-Means Dan C4. 5 Untuk Prediksi Penderita Diabetes. *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*. 2020 Mar 22;13(1):86-100.