

Analisis Sentimen Mahasiswa Terhadap Perkuliahan dalam Jaringan Menggunakan Metode Naïve Bayes Classifier

Bagus Aziz Rahmatullah*, Imam Sujarwo, Erna Herawati

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

bagusazra11@gmail.com*, imamsoejarwo@gmail.com, mahyafarida666@gmail.com

Abstrak

Sejak terjadinya pandemi virus Corona di Indonesia, pemerintah mengedarkan surat edaran Kementerian Pendidikan dan Kebudayaan (Kemendikbud) Direktorat Pendidikan Tinggi No.1 tahun 2020 tentang pencegahan penyebaran *Corona Virus Disease* (Covid-19) di perguruan tinggi. Melalui surat edaran tersebut, pihak kemendikbud memberikan instruksi kepada perguruan tinggi untuk menyelenggarakan pembelajaran jarak jauh dan menyarankan mahasiswa untuk belajar di rumah masing-masing. Perubahan mendadak dari perkuliahan tatap muka menjadi perkuliahan daring menyebabkan berbagai macam respon dari para mahasiswa. Tujuan dari penelitian ini adalah untuk menganalisis sentimen atau respon mahasiswa terhadap perkuliahan daring di masa pandemi Covid-19 di Indonesia dengan menggunakan data yang dikumpulkan lewat kuisioner dan diolah dengan menggunakan metode *naïve bayes classifier*. Penelitian ini adalah penelitian studi kasus deskriptif kuantitatif. Penelitian ini dilakukan dengan mengumpulkan data terlebih dahulu. Data dikumpulkan lewat kuisioner dengan pertanyaan mengenai tanggapan mereka terhadap perkuliahan dalam jaringan pada masa pandemi Covid-19. Hasil menunjukkan bahwa pembelajaran daring memiliki sentimen negatif lebih banyak daripada sentimen positif. Tingginya sentimen negatif disebabkan karena ketidaknyamanan mahasiswa dalam pembelajaran daring. Kata yang sering muncul adalah tidak efektif, susah, dan tugas. Hasil akurasi dari model ini adalah 75% di mana hasil akurasi ini merupakan hasil yang cukup baik dalam pengklasifikasian

Kata kunci: kuliah daring; Naïve Bayes; analisis sentiment; Covid-19; akurasi

Abstract

Since the pandemic of Covid-19 was happened in Indonesia, the government shared the letter of The Ministry of Education and Higher Culture Education Directorate No.1, year 2020 about prevention of the spread of Corona Virus Disease (Covid-19) in higher education. Through the letter, The Ministry of Education and Culture gave an instruction for college to organize online learning and suggested students to study at their home. Online learning which was considered as a strategy then became a controversy because it needed adaption. This sudden change from normal learning to online learning caused many responses from students. The aim of this research was to analyze student sentiment or responses on online learning in this pandemic era of Covid-19 in Indonesia by using data which had been collected using questionnaire and processed using naïve bayes classifier method. This research was case study descriptive quantitative research. The research was done by collecting the data first. The data was collected through questionnaire with the question about their opinion on online learning in this pandemic Covid-19 era. The data was 157 student's data opinion on online learning. After the data was collected, the data was cleaned first from question mark and the words which didn't give an effect in sentiment analysis. After the data was cleaned, then the result of the classification will be showed as well as the accuracy which the model earned. The result showed that online learning had negative sentiment more than positive sentiment. The height of negative sentiment was caused by discomfort of student in online learning. The word which frequently showed was 'tidak efektif', 'susah', and 'tugas'. The accuracy of this model was 75% when the result of this accuracy was good result in classification.

Keywords: online learning; *naïve bayes*; sentiment analysis; covid-19; accuracy

PENDAHULUAN

Peraturan Pemerintahan Republik Indonesia No. 30 tahun 1990 tentang perguruan tinggi mengatakan bahwa pendidikan tinggi adalah pendidikan jenjang yang lebih tinggi daripada pendidikan menengah di jalur pendidikan sekolah. Perguruan tinggi merupakan suatu pendidikan yang menjadi terminal akhir bagi seorang yang berpeluang belajar setinggi-tingginya melalui jalur pendidikan sekolah [1].

Ketika menuntut ilmu di perguruan tinggi, biasanya dosen dan mahasiswa melakukan kegiatan belajar mengajar di dalam kelas. Khusus mahasiswa yang melakukan praktikum, dosen dan mahasiswa melakukan praktikum di laboratorium. Sejak terjadinya pandemi virus Corona di Indonesia, pemerintah mengedarkan surat edaran Kementerian Pendidikan dan Kebudayaan (Kemendikbud) Direktorat Pendidikan Tinggi No.1 Tahun 2020 tentang pencegahan penyebaran Corona Virus Disease (Covid-19) di perguruan tinggi. Melalui surat edaran tersebut pihak Kemendikbud memberikan instruksi kepada perguruan tinggi untuk menyelenggarakan pembelajaran jarak jauh dan menyarankan mahasiswa untuk belajar di rumah masing-masing atau yang dikenal sekarang dengan sebutan perkuliahan dalam jaringan (daring) [2].

Perkuliahan daring adalah pembelajaran interaktif. Dalam perkuliahan daring di masa pandemi membuat mahasiswa tidak diperkenankan untuk pergi ke kampus. Dosen dan mahasiswa dapat berkomunikasi dengan menggunakan media internet [3].

Perkuliahan daring yang dianggap sebagai strategi kemudian menjadi suatu kontroversi karena diperlukannya adaptasi. Perbedaan infrastruktur yang dimiliki, kualitas koneksi, perangkat yang digunakan, dan masih mahalanya kuota internet menjadi hambatan utama perkuliahan daring ini. Perubahan mendadak dari kuliah tatap muka menjadi perkuliahan daring pada skala besar menyebabkan berbagai respons dari para mahasiswa [4]. Penelitian ini bertujuan untuk menganalisis sentimen terhadap perkuliahan daring atau dalam jaringan pada masa pandemi Covid-19 di Indonesia dengan menggunakan data dari kuesioner dan diolah dengan menggunakan *naïve bayes* untuk menganalisis sentimen.

Pada penelitian yang ditulis oleh Tika dkk (2020) tentang Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Menggunakan Naïve Bayes mencoba untuk menganalisis opini masyarakat mengenai aturan larangan mudik dengan mengklasifikasikan opini menjadi positif dan negatif. Pada penelitian ini juga menunjukkan hasil akurasi yang tinggi yaitu sebesar 83,38%. Penelitian yang ditulis oleh Fitri dkk (2020) tentang Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naïve Bayes, Random Forest dan Support Vector Machine. Penelitian ini menunjukkan jumlah komentar positif lebih banyak diberikan dibandingkan dengan komentar negatif dari para pengguna aplikasi Ruangguru. Penelitian ini mendapatkan akurasi sebesar 94,16% untuk metode naïve bayes. Pada penelitian yang ditulis oleh Fanissa dkk (2018) yang meneliti tentang Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naïve Bayes Classifier dan seleksi Fitur Query Expansion Ranking mengungkapkan bahwa klasifikasi menggunakan Naïve Bayes Classifier mendapatkan hasil akurasi sebesar 86,6%.

METODE

Pendekatan Penelitian

Pada penelitian ini, pendekatan yang digunakan adalah pendekatan studi kasus deskriptif kuantitatif. Pada studi kasus yaitu dengan pengumpulan data menggunakan kuesioner yang disebarkan kepada responden. Pada penelitian ini penulis menggunakan Google Form. Sedangkan pendekatan deskriptif kuantitatif menganalisis data dan menyusun data yang telah ada agar sesuai dengan kebutuhan peneliti.

Variabel Penelitian

Pada penelitian ini, variabel penelitian dibagi menjadi dua, yaitu variabel bebas dan variabel terikat. Pada variabel bebas yaitu beberapa opini mahasiswa tentang pembelajaran daring, sedangkan untuk variabel terikatnya adalah label positif dan negatif yang diberikan sendiri oleh para responden.

Jenis dan Sumber Data

Pada penelitian ini data yang digunakan adalah yang berjenis teks opini mahasiswa terhadap perkuliahan dan responden juga yang menentukan label dari opini tersebut, apakah opini tersebut termasuk opini positif atau opini negatif

Data bersumber dari responden menggunakan instrumen berupa kuesioner. Pada instrumen tersebut responden akan menjawab pertanyaan tentang bagaimana pendapat mereka tentang perkuliahan daring. Data yang digunakan hanya sebagai simulasi pada program saja. Responden pada penelitian ini adalah mahasiswa dari beberapa universitas.

Pra-pemrosesan Data

Sebelum melakukan klasifikasi, data disiapkan dulu sehingga mendapatkan data yang diinginkan. Tahap ini dinamakan pra-pemrosesan data. Tahap untuk melakukan pra-pemrosesan data adalah sebagai berikut.

1. Case Folding

Pada tahap ini semua dokumen akan diubah menjadi huruf kecil. Hal ini dilakukan agar tidak terjadi redundansi atau pengulangan data yang hanya berbeda pada hurufnya saja.

2. Cleansing

Pada tahap ini semua kata dibersihkan dari kata-kata yang tidak berpengaruh sama sekali dengan analisis sentimen. Komponen pada dokumen memiliki berbagai macam atribut seperti karakter simbol atau tanda baca. Atribut yang tidak berhubungan ini akan dihilangkan dari dokumen dan diganti dengan karakter spasi.

3. Stopword Removal

Tahap ini merupakan tahap di mana semua kata yang tidak sesuai dengan topik dihilangkan jika ada kata yang tidak memengaruhi akurasi klasifikasi maka kata akan dihilangkan. Kata-kata yang termasuk pada stopwords di antaranya “dengan”, “ada”, “harus”, “kepada”, dan lain-lain. Pada penelitian ini, peneliti menghapus kata ‘tidak’ pada kamus *stopwords* karena kata ‘tidak’ merupakan kata yang mengandung makna negatif.

4. Stemming

Pada tahap ini, semua kata diubah ke dalam bentuk dasarnya dengan menghapus imbuhan pada depan kata maupun imbuhan pada belakang kata.

5. Tokenisasi

Pada tahap ini kalimat dipecah menjadi beberapa bagian yang disebut token. Sebuah token dianggap sebagai bentuk kata atau frasa yang memiliki arti.

Analisis Data

Data yang digunakan yaitu data teks yang dikumpulkan dari tanggal 17 Oktober 2020 sampai 6 Desember 2020. Jumlah data yang dikumpulkan adalah sebanyak 100 data teks opini tentang perkuliahan daring. Data terdiri dari nama mahasiswa, jenis kelamin, universitas, fakultas/Program Studi/angkatan, komentar, dan jenis komentar apakah positif atau negatif. Pada data tersebut responden yang menentukan apakah opini yang diberikan positif atau negatif.

Adapun langkah-langkah yang digunakan untuk menganalisis sentimen mahasiswa terhadap perkuliahan daring yaitu sebagai berikut:

1. Menguji model naïve bayes classifier pada data uji.
2. Prediksi hasil tes.
3. Membuat *confusion matrix*.
4. Menghitung nilai akurasi.

HASIL DAN PEMBAHASAN

Sebelum data digunakan, peneliti melakukan pra-pemrosesan data terlebih dahulu. Untuk menghilangkan *noise* dan memperjelas fitur, proses tokenisasi dan konversi data perlu dilakukan

dalam pra-pemrosesan data. Tahap ini bertujuan untuk membersihkan data dan menghapus atribut-atribut yang tidak diperlukan seperti tanda baca. Proses ini akan mengolah teks menjadi data yang mudah untuk diterima oleh sistem.

Case Folding

Case folding dilakukan untuk menyeragamkan huruf yang digunakan menjadi huruf kecil. Berikut merupakan dokumen yang sudah diterapkan *case folding*.

Tabel 2. Case Folding

Sebelum	Sesudah
menurut saya, selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar di rumah. Tapi, dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan.	menurut saya, selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar di rumah. tapi , dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan.

Cleansing

Selanjutnya adalah *cleansing*. Tahap ini dilakukan untuk menghilangkan tanda baca, simbol, angka atau karakter yang bukan huruf. Berikut merupakan dokumen yang sudah diterapkan *cleansing*.

Stopword Removal

Pada tahap ini, kata-kata yang kurang bermakna atau tidak memiliki arti akan dilakukan penghapusan. Hal ini dilakukan karena kata pada *stopword* tidak memengaruhi akurasi dalam klasifikasi sentimen. Jika ada kata pada dokumen yang sama dengan kata dalam *stopword*, maka kata tersebut akan dihilangkan dan diganti dengan spasi. Berikut adalah dokumen yang sudah diterapkan tahap *stopword removal*.

Tabel 3. Stopword Removal

Sebelum	Sesudah
menurut saya selama kuliah online sistem belajar jadi teratur baik kuliah maupun kegiatan belajar dirumah tapi dengan kuliah online juga kadang mengalami kendala jaringan yang tadinya baik jadi tidak ada jaringan untuk daerah yang susah mendapat jaringan	kuliah online sistem belajar teratur kuliah kegiatan belajar dirumah kuliah online kadang mengalami kendala jaringan tidak jaringan daerah susah jaringan

Stemming

Setelah tahap *stopword removal* dilakukan, tahap selanjutnya adalah melakukan *stemming*. Pada tahap *stemming*, semua kata diubah ke dalam bentuk dasarnya dengan menghapus kata imbuhan di depan maupun di belakang kata. Berikut dokumen yang sudah diterapkan *stemming*.

Tabel 4. Proses Stemming

Sebelum	Sesudah
kuliah online sistem belajar teratur kuliah kegiatan belajar dirumah kuliah online kadang mengalami kendala jaringan tidak jaringan daerah susah jaringan	kuliah online sistem ajar atur kuliah giat ajar rumah kuliah online kadang alami kendala jaring tidak jaring daerah susah jaring

38% untuk kelas positif dan 88% untuk kelas negatif. Sebagai kesimpulan, kemampuan suatu model dalam menemukan kembali suatu informasi sangat rendah.

Untuk mengevaluasi suatu model dilakukan pengujian kedua dengan menggunakan *k-fold cross validation*. Di mana jumlah *k* yang digunakan pada penelitian ini adalah 10. Dalam *10-fold cross validation*, data dibagi menjadi 10 bagian yang kira-kira berukuran sama, sehingga memiliki 10 *subset* data untuk mengevaluasi suatu model. Dengan menggunakan *cross validation* nilai akurasi menjadi **0,61**. *Cross validation* dilakukan agar hasil akurasi yang didapat menjadi lebih valid dibanding dengan memisahkan data 80% untuk data latih dan 20% untuk data uji.

KESIMPULAN

Berdasarkan hasil pengujian menggunakan *naïve bayes classifier* yang telah dilakukan, ada beberapa hal yang dihasilkan Hasil klasifikasi pada penelitian ini menunjukkan bahwa sentimen negatif lebih banyak dibandingkan sentimen positif. Ketika data dipisah 20% untuk menguji suatu model, 21 sentimen diklasifikasikan sebagai sentimen yang negatif sedangkan tiga sentimen diklasifikasikan sebagai sentimen yang positif dan pada penelitian ini, metode *naïve bayes classifier* mendapatkan hasil akurasi sebesar 75%. Hasil ini bisa dikatakan cukup baik. Untuk memastikan hasil penelitian ini dilakukan juga pengujian menggunakan *k-fold cross validation* dengan *k* yaitu 10 mendapatkan nilai akurasi sebesar 61% sehingga lebih akurasi lebih bagus menggunakan pemisahan 80% untuk data latih dan 20% untuk data uji.

DAFTAR PUSTAKA

- [1] Dardjowidjojo, S. (1991). *Pedoman Pendidikan Tinggi*. PT. Gramedia Widiasarana Indonesia.
- [2] R. M. McLeod, K. Ranson and L. Biehl, The generalized Riemann integral, JSTOR, 1980.
- [3] C. Godsil and G. F. Royle, Algebraic graph theory, vol. 207, Springer Science & Business Media, 2013.
- [4] A. Gara, M. A. Blumrich, D. Chen, G.-T. Chiu, P. Coteus, M. E. Giampapa, R. A. Haring, P. Heidelberger, D. Hoenicke, G. V. Kopcsay and others, "Overview of the Blue Gene/L system architecture," *IBM Journal of Research and Development*, vol. 49, no. 2, pp. 195-212, 2005.
- [5] J. France, J. H. Thornley and others, *Mathematical models in agriculture.*, Butterworths, 1984.
- [6] F. E. Browder, "Nonexpansive nonlinear operators in a Banach space," *Proceedings of the National Academy of Sciences*, vol. 54, no. 4, pp. 1041-1044, 1965.
- [7] M. J. Berger and J. Olinger, "Adaptive mesh refinement for hyperbolic partial differential equations," *Journal of computational Physics*, vol. 53, no. 3, pp. 484-512, 1984.
- [8] Hermanto, H., Kuryanti, S. J., & Khasanah, S. N. (2019). Comparison of Naïve Bayes Algorithm, C4.5 and Random Forest for Classification in Determining Sentiment for Ojek Online Service. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 3(2), 266–274. <https://doi.org/10.33395/sinkron.v3i2.235>
- [9] Inyang, I., Ozuomba, S., & Ezenkwu, C. (2017b). Comparative analysis of Mechanisms for Categorization and Moderation of User Generated Text Contents on a Social E-Governance Forum. *Mathematical and Software Engineering*, 3(1), 78–86.
- [10] Budiharto, W., & Suhartono, D. (2014). *Artificial Intelligence: Konsep dan Penerapannya*. Andi Publisher. <https://research.binus.ac.id/publication/95BD64D6-904F-4D97-BE58-E2BA778F5422/artificial-intelligence-konsep-dan-penerapannya/>
- [11] Rodiyansyah, S. F., & Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 6(1). <https://doi.org/10.22146/ijccs.2144>

- [12] Hidayatillah, R., Mirwan, M., Hakam, M., & Nugroho, A. (2019). Levels of Political Participation Based on Naive Bayes Classifier. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), 73. <https://doi.org/10.22146/ijccs.42531>
- [13] Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- [14] Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data. *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–8. <https://doi.org/10.1109/KSE.2019.8919368>
- [15] Permana, A. Y., Ismasari, & Effendi, M. M. (2018). Optimasi Stemming Porter KBBI dan Cross Validation Naïve Bayes untuk Klasifikasi Topik Soal UN Bahasa Indonesia. *Jurnal Ilmiah Komputasi*, 17(4). <https://doi.org/10.32409/jikstik.17.4.2492>