

**TEXT MINING DALAM ANALISIS SENTIMEN PEMBELAJARAN DARING DI MASA PANDEMI
COVID 19 MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR**

Nengah Widya Utami^{1*}, Made Artana²
^{1,2}Sistem Informasi Akuntansi, STMIK Primakara
email: widya@primakara.ac.id

Abstrak: Sistem pembelajaran di masa pandemi Covid19 ini telah beralih menjadi pembelajaran dalam jaringan (daring). Berbagai komentar baik positif maupun negatif diungkapkan khususnya oleh peserta didik. Mahasiswa STMIK Primakara menyampaikan kritik/saran/pendapat terkait metode atau strategi mengajar yang dilakukan oleh dosen dalam pembelajaran daring melalui sistem SISKAPrimakara. pengisian data EDOM oleh mahasiswa dilakukan setiap 6 (enam) bulan sekali. Hingga saat ini data komentar edom berjumlah sangat besar dan belum dimanfaatkan dengan baik. Selain itu, saat ini belum ada strategi yang tepat mengenai metode pembelajaran yang harus diimplementasikan dalam situasi pembelajaran daring. Untuk itu dilakukan penerapan *text mining* dalam analisis sentiment pembelajaran daring di STMIK Primakara menggunakan algoritma *K-Nearest Neighbor* (KNN). Hasil penelitian ini diharapkan dapat menjadi acuan dalam penentuan metode dan strategi pembelajaran yang tepat guna tercapainya peningkatan kepuasan mahasiswa dalam pembelajaran daring. Tahapan penelitian meliputi identifikasi masalah, studi literatur, pengumpulan data, *text preprocessing* (*filtering, cleansing, case folding, tokenizing, stemming, stopword, dan cross validation*), analisis dan kesimpulan, pembuatan laporan, serta publikasi luaran. Hasil penelitian menunjukkan komentar mahasiswa lebih cenderung sentimen negatif. Persepsi negatif dihasilkan karena ketidakpuasan mahasiswa terhadap pembelajaran daring. Hasil *confusion matrix* menunjukkan prediksi komentar negatif yakni *True Negatif* (TN) sebesar 127 dan 10 *False Negatif* (FN), sedangkan data prediksi komentar positif *True Positif* (TP) sebesar 14 dan 20 *False Positif* (FP). Berdasarkan uji *cross validation* menunjukkan bahwa nilai *Precision* menggunakan algoritma KNN sebesar 86.39% positif dan 58.33% negatif. Sedangkan nilai *recall* sebesar 92.70% negatif dan 41.18% positif. Nilai akurasi KNN sebesar 82.48% yang berarti KNN dapat dikatakan akurat dalam melakukan analisis sentimen komentar pembelajaran daring. Selain itu dilakukan perbandingan dengan SVM dan Naïve Bayes, dimana hasil menunjukkan KNN memiliki nilai akurasi yang lebih tinggi sedangkan algoritma Naïve Bayes memiliki akurasi yang paling rendah. Hasil penelitian ini membuktikan bahwa pembelajaran daring belum dapat diterapkan secara maksimal di STMIK Primakara, berbagai kendala baik dari dosen maupun mahasiswa menyebabkan ketidakpuasan mahasiswa dalam mengikuti perkuliahan.

Kata Kunci : text mining; analisis sentimen; pembelajaran; covid19; knn

Abstract: *The learning system during the Covid-19 pandemic has turned into online learning. Various comments, both positive and negative, were expressed especially by students. STMIK Primakara students submit criticisms/suggestions/opinions related to teaching methods or strategies carried out by lecturers in online learning through the SISKAPrimakara system. EDOM data filling by students is done every 6 (six) months. Until now, the data for Edom comments is very large and has not been used properly. In addition, there is currently no appropriate strategy regarding learning methods that must be implemented in online learning situations. For this reason, the application of text mining in online learning sentiment analysis at STMIK Primakara uses the K-Nearest Neighbor (KNN) algorithm. The results of this study are expected to be a reference in determining appropriate learning methods and strategies in order to achieve increased student satisfaction in online learning. The research stages include problem identification, literature study, data collection, text preprocessing (filtering, cleansing, case folding, tokenizing, stemming, stopword, and cross validation), analysis and conclusions, report generation, and output publication. The results showed that student comments were more likely to have negative sentiments. Negative perceptions are generated due to student dissatisfaction with online learning. The results of the confusion matrix show the predictions for negative comments, namely True Negative (TN) of 127 and 10 False Negative (FN), while the prediction data for positive comments of True Positive (TP) of 14 and 20 False Positive (FP). Based on the cross validation test, it shows that the Precision value using the KNN algorithm is 86.39% positive and 58.33% negative. While the recall value is 92.70% negative and 41.18% positive. The KNN accuracy value is 82.48%, which means that KNN can be said to be accurate in analyzing online learning comment sentiments. In addition, a comparison is made with SVM and Nave Bayes, where the results show that KNN has a higher accuracy value, while the Naïve Bayes algorithm has the lowest accuracy. The results of this study prove that online learning cannot be applied optimally at STMIK Primakara, various obstacles from both lecturers and students cause student dissatisfaction in attending lectures.*

Keywords : text mining; analysis sentiment; learning; covid19; knn

PENDAHULUAN

Pandemi Covid-19 yang masih berlangsung hingga saat ini menimbulkan dampak dan perubahan besar di berbagai bidang khususnya pendidikan. Sistem pembelajaran kini telah beralih menjadi sistem pembelajaran daring. Pembelajaran daring merupakan pembelajaran yang memanfaatkan media

interaktif dengan dasar internet yang digunakan antara pendidik dan peserta didik [1]. Regulasi pembelajaran daring yang telah ditetapkan pemerintah merupakan hal yang telah dirasakan pelajar hampir selama 2 (tahun) belakangan ini, sehingga memunculkan berbagai opini dan pandangan yang bersifat pro maupun kontra.

Pembelajaran daring dapat mengubah gaya mengajar konvensional dan memberi peluang lebih bagi dosen untuk menilai dan mengevaluasi progress pembelajaran setiap mahasiswanya secara lebih efisien [2]. Namun, tak jarang kuliah secara daring dianggap membawa beragam kendala baru dalam pembelajaran [3]. Berbagai masalah muncul seperti ketidaksiapan teknologi, media yang digunakan, dan faktor psikologis dari peserta didik [2].

STMIK Primakara telah menerapkan pembelajaran daring semenjak berlakunya peraturan pemerintah dan gubernur terkait pemberlakuan kebijakan *social distancing*. Dengan diberlakukannya kebijakan ini berbagai komentar mahasiswa baik positif dan negatif terhadap metode pembelajaran yang digunakan dosen selama pembelajaran daring diungkapkan melalui angket kepuasan mahasiswa yang disebut dengan EDOM (Evaluasi Dosen Mahasiswa). Penilaian EDOM ini dilakukan oleh mahasiswa pada setiap akhir semester melalui Sistem Informasi Kampus Primakara (SISKA). *Dataset* komentar pada EDOM di SISKA STMIK Primakara jumlahnya sangat banyak yaitu mencapai ribuan komentar. Namun data tersebut belum dimanfaatkan dengan baik, padahal apabila data tersebut diolah tentunya akan menghasilkan berbagai pengetahuan penting untuk bisa digunakan sebagai pedoman dalam pengambilan strategi/kebijakan. Selain itu hingga saat ini belum diterapkannya metode pembelajaran yang tepat dan sesuai dengan keinginan mahasiswa khususnya dalam proses pembelajaran daring. Adapun teknologi yang dapat dimanfaatkan untuk menggali pengetahuan di dalam kumpulan data berupa teks yang jumlahnya besar yaitu *text mining*.

Text mining merupakan salah satu cara untuk mencari pola dalam data teks tidak terstruktur yang dilakukan secara otomatis oleh komputer untuk menemukan informasi yang bermanfaat untuk tujuan tertentu [4]. *Sentiment analysis* merupakan cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence* yang dilakukan untuk memperoleh informasi yang bermanfaat atau pengetahuan baru dengan cara mengekstrak, memahami, dan mengolah data teks secara otomatis [3]. Melalui proses *text mining* akan terlihat bagaimana kecenderungan pendapat atau opini seseorang terhadap suatu topik atau masalah dengan menentukan klasifikasi sentimen ke dalam dua kelas atau lebih. Salah satu algoritma yang dapat digunakan dalam klasifikasi sentimen adalah *K-Nearest Neighbor* (KNN). KNN adalah algoritma klasifikasi yang berdasarkan *data training* dengan jarak yang paling dekat dengan objek [5].

Dengan demikian, penelitian ini bertujuan untuk klasifikasi komentar pembelajaran daring di STMIK Primakara sehingga diharapkan dapat menjadi bahan evaluasi pelaksanaan kebijakan mengenai pembelajaran daring dan dapat dijadikan sebagai acuan dalam menerapkan metode

pembelajaran yang tepat bagi mahasiswa di STMIK Primakara.

TINJAUAN PUSTAKA

Text mining merupakan sebuah proses untuk melakukan pencarian pengetahuan yang berfokus pada data yang berbentuk dokumen atau teks, dengan tujuan untuk mengekstrak informasi yang berguna dan mengidentifikasinya [6]. Analisis sentimen adalah suatu bidang yang termasuk dalam analisis *text mining*. Analisis sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas tersebut dapat mewakili individu. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif [7]. Langkah-langkah analisis sentimen klasifikasi terhadap data *text mining* atau data teks adalah sebagai berikut [8].

1. Tahap awal yaitu mengumpulkan *dataset* seperti pendapat masyarakat, penilaian terhadap restoran atau produk dan lain-lain.
2. *Pre-processing* yaitu tahapan yang mencakup *Tokenization*, *Stopwords Removal*, dan *Stemming*.
3. *Transformation* yaitu melakukan pembobotan dari data teks
4. *Feature Selection* yaitu tahap mengurangi data yang tidak diperlukan
5. *Classification* yaitu tahap klasifikasi teks yang biasanya menggunakan metode seperti: *Naive Bayes*, *KNN*, *SVM*, dan lain-lain.
6. *Interpretation/Evaluation* yaitu tahap evaluasi untuk menghitung nilai *accuracy* dan nilai *Area Under the Curve*

Salah satu algoritma yang dapat digunakan untuk klasifikasi adalah *K-Nearest Neighbor* (KNN). Menurut Tanggu Mara, dkk (2021) KNN merupakan algoritma yang dapat digunakan untuk melakukan proses klasifikasi opini hasil analisis sentimen. Kelebihan dari algoritma ini terbukti terbukti dapat mencapai akurasi yang tinggi dan sesuai dengan perhitungan yang diterapkan dalam aplikasi-aplikasi [8]. Dengan demikian dalam penelitian ini algoritma KNN digunakan untuk klasifikasi hasil *text mining* dan analisis sentimen yang bertujuan untuk mengklasifikasikan komentar mahasiswa baik berupa opini positif maupun opini negatif dan selanjutnya akan dihitung tingkat akurasi dari algoritma KNN. *State of the art* pada penelitian ini mencakup penelitian-penelitian yang telah dilakukan sebelumnya seperti yang ditunjukkan pada Tabel 1. berikut.

Tabel 1. Tabel State of The Art

No	Judul	Peneliti
1	Analisa Sentimen Terhadap Belajar <i>Online</i> Pada Masa Covid19 Menggunakan Algoritma <i>Support Vector Machine</i> Berbasis <i>Particle Sarm Optimization</i> [6]	Hermanto dan Astrid Noviriandini, 2021
2	Penerapan Algoritma <i>K-Nearest Neighbors</i> Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba [8]	Mara, dkk, 2021
3	Analisis Sentimen dan <i>Topic Modelling</i> Pada Pembelajaran <i>Online</i> di Indonesia Melalui <i>Twitter</i> [9]	Indra Kusumajati Susanto, 2021
4	Analisis Sentimen Pembelajaran Daring Pada <i>Twitter</i> di Masa Pandemi COVID-19 Menggunakan Metode <i>Naïve Bayes</i> [2]	Samsir, dkk, 2021
5	Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di <i>Twitter</i> Menggunakan Algoritma <i>Naïve Bayes</i> [7]	Siti Yana Nursyifa, dkk, 2021
6	Implementasi Metode <i>Naïve Bayes Classifier</i> (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ) [10]	Naomi Chatrina Siregar, dkk, 2020
7	Implementasi metode <i>Naïve Bayes Classifier</i> pada Evaluasi Kepuasan Mahasiswa terhadap Pembelajaran Daring [11]	Bister Purba, Rian Syahputra, 2021
8	Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran <i>Online</i> Selama Pandemi Covid-19 Pada Media Sosial <i>Twitter</i> Menggunakan Perbandingan Algoritma Klasifikasi	Ezza Ramadhanta Machmud Razaq, dkk, 2021
9	Seleksi Fitur <i>Support Vector Machine</i> pada Analisis Sentimen Keberlanjutan Pembelajaran Daring [12]	Amar P. Natasuwarna, 2020
10	Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari <i>Twitter</i> Dataset Menggunakan <i>InSet Lexicon</i> [3]	Desi Musfiroh, dkk, 2021

Beberapa peneliti telah melakukan penelitian terkait analisis sentimen komentar pada social media *Twitter* menggunakan algoritma klasifikasi *Support Vector Machine* seperti yang dilakukan oleh Hermanto dan Astrid Noviriandini (2021) berjudul “Analisa Sentimen Terhadap Belajar *Online* Pada Masa Covid-19 Menggunakan Algoritma *Support Vector Machine* Berbasis *Particle Sarm Optimization*”. Hasil penelitian menunjukkan algoritma ini memiliki tingkat akurasi sebesar 71.39% serta diperoleh hasil analisis sentiment berupa komentar positif dan negatif [6]. Selain itu SVM juga digunakan oleh peneliti Amar P. Natasuwarna (2020) yang menunjukkan bahwa algoritma SVM dengan perbandingan dua seleksi fitur dan variabel *k-Fold* pada *Cross Validation* menghasilkan evaluasi yang cukup baik digunakan untuk analisis sentimen dengan perolehan nilai tertinggi *accuracy* dan *recall* pada kedua seleksi bernilai sama.

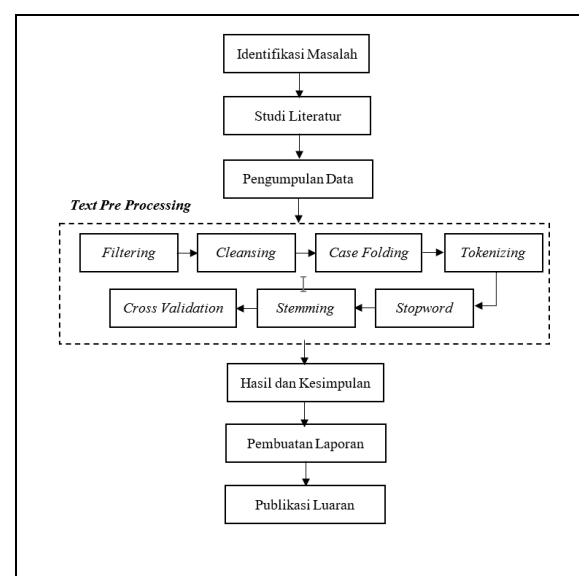
Selain SVM algoritma yang paling banyak digunakan adalah *Naïve Bayes*. Algoritma ini bersifat sederhana dan mudah diimplementasikan. Beberapa penelitian telah dilakukan khususnya dalam klasifikasi hasil

analisis sentiment dimana sumber datanya berasal dari *Twitter*. Hasil *crawling* dan analisis sentimen menghasilkan persentase klasifikasi komentar positif dan negatif. Berdasarkan penelitian oleh Ezza Ramadhanta Machmud Razaq, dkk (2021) yang berjudul “Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran *Online* Selama Pandemi Covid-19 Pada Media Sosial *Twitter* Menggunakan Perbandingan Algoritma Klasifikasi” dimana dalam penelitiannya membanding 3 metode yakni *Naïve Bayes*, *K-Nearest Neighbor* (K- NN) dan C4.5 menunjukkan bahwa KNN memiliki tingkat akurasi tertinggi sebesar 60.80%. penelitian ini juga didukung oleh penelitian Mara, dkk (2021) yang membuktikan bahwa algoritma KNN menghasilkan tingkat akurasi sebesar 87.00% [8].

Berdasarkan studi pustaka yang telah dilakukan maka peneliti menggunakan algoritma KNN dalam proses analisis sentimen. Penelitian terdahulu Sebagian besar menggunakan data pada social media seperti *twitter* dan *facebook*, namun di penelitian ini data yang digunakan adalah data komentar mahasiswa pada EDOM yang bersumber dari *database* STMIK Primakara. hal ini bertujuan untuk mengetahui hasil klasifikasi komentar oleh mahasiswa untuk selanjutnya dapat dimanfaatkan dalam hal penentuan strategi maupun metode pembelajaran yang tepat pada proses belajar mengajar di STMIK Primakara.

METODE

Metode penelitian yang akan digunakan pada penelitian ini sebagaimana ditunjukkan pada Gambar 1 berikut.



Gambar 1. Tahapan Metode Penelitian

Tahapan metode penelitian yang dilakukan dalam penelitian ini yaitu sebagai berikut.

1. Identifikasi Masalah

Dalam penelitian ini, permasalahan yang terjadi adalah belum diterapkannya metode atau strategi pembelajaran yang tepat dalam pembelajaran daring di STMIK Primakara. Selain itu besarnya jumlah data komentar EDOM yang belum dimanfaatkan dengan baik, sehingga dilakukan analisis sentiment menggunakan algoritma *K-Nearest Neighbor* (KNN).

2. Studi Literatur

Studi literatur dilakukan untuk mengkaji dan menggali pengetahuan mengenai *text mining*, analisis sentiment, dan algoritma KNN melalui buku, jurnal, artikel, *paper*, maupun *internet*.

3. Pengumpulan Data

Sumber data utama yang akan digunakan dalam penelitian ini dari *database* komentar hasil EDOM yang bersumber dari bagian PPTI STMIK Primakara.

4. *Text Preprocessing*

Text Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain [7]. Tahapan *text preprocessing* adalah sebagai berikut.

a) *Filtering dan Cleansing*

Pada tahapan ini dilakukan proses *filtering* dan *cleansing* yaitu menggunakan *select* atribut dan *remove duplicate* untuk mengambil kata-kata penting, membersihkan dokumen dari kata yang tidak diperlukan, dan membersihkan kata yang *double* atau duplikat. selanjutnya *dataset* akan meliputi tahap pelabelan manual ke dalam dua kelas sentimen yakni positif dan negatif.

b) *Case Folding*

Case folding dengan menggunakan *transform cases* untuk mengubah teks menjadi huruf kecil.

c) *Tokenizing*

Tahap dimana dilakukannya penghapusan terhadap semua karakter non alphabet, misalnya simbol, spasi, angka menggunakan *tokenize*.

d) *Stemming*

Proses dilanjutkan dengan menghilangkan imbuhan pada masing-masing kata sehingga menjadi kata dasar, dan dalam tahap ini bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat.

e) *Stopword*

Stopword diproses pada sebuah kalimat yang mengandung kata-kata yang sering keluar dan dianggap tidak penting seperti waktu atau penghubung. Untuk itu perlu penghapusan.

f) *Cross Validation*

Proses *cross validation* menggunakan algoritma *KNN* yang meliputi proses *training* dan *testing (apply model dan performance)*. Setelah itu dilakukan proses pemodelan dengan menggunakan algoritma *KNN* untuk mendapatkan nilai *accuracy*.

HASIL DAN PEMBAHASAN

Pelaksanaan penelitian *text mining* dalam analisis sentimen pembelajaran daring di masa pandemi covid19 menggunakan algoritma *K-Nearest Neighbor* Studi Kasus: STMIK Primakara ini dimulai pada bulan Januari hingga September 2022. Adapun hasil penelitian berdasarkan tahapan penelitian yang dilakukan dijabarkan sebagai berikut.

1. Pengumpulan Data

Data utama yang digunakan dalam penelitian ini adalah *database* komentar hasil Evaluasi Dosen (EDOM) yang bersumber dari bagian PPTI STMIK Primakara pada tahun 2020-2021. Hasil pengumpulan *dataset* diperoleh sejumlah 210 komentar berdasarkan hasil *query text* yang memuat kata “daring” dan “online”. *Dataset* yang digunakan dalam penelitian ini seperti yang ditunjukkan pada Tabel 2 berikut.

Tabel 2. *Dataset* Komentar

No	Komentar
1	Tolong selama pandemi di permudah untuk bimbingan online Kadang susah sekali ngubungin dosennya.
2	bapaknya jarang masuk kelas lebih sering online
3	Terlalu banyak courses bhs. Indonesia yang saya dan teman-teman saya ikuti hanya untuk melakukan absen secara online saat melakukan kelas online. Kami jadi bingung untuk mengikuti aturan tersebut, jadinya ada yang tidak melakukan absen.
4	sudah bagus, mungkin perlu meningkatkan komunikasi lagi antara dosen dan mahasiswa, mengingat pandemi seperti ini rentan terjadinya kesalahpahaman via daring ketimbang bertemu langsung. Tidak lupa utk memberikan informasi entah mengenai jadwal perkuliahan atau hal lainnya tidak secara dadakan, karena meskipun stay at home dan study from home, belum menjamin bahwa mahasiswa siap utk diberikan informasi secara mendadak
5	Jika mengajar online pastikan bisa berjalan lancar tanpa hambatan.
6	Mengajar hanya memberi tugas. Tidak mengadakan meet secara online. Padahal waktu bimbingan beliau sangat menguasai mata kuliah ini. Sayang sekali.
7	- penyampaian grup proposal dan skripsi yang kurang jelas, sehingga kebanyakan mahasiswa yang mengambil matkul proposal bergabung ke skripsi yang menyebabkan ketinggalan informasi mengenai proposal skripsi. - kurang siap dalam memberikan informasi kepada mahasiswa - tidak adanya pertemuan/online meet pertama untuk menjelaskan bagaimana proses dan tata aturan proposal skripsi (hanya mencantumkan slide presentasi) - contoh laporan proposal skripsi dan skripai dijadikan 1 sehingga kami bingung t
8	Efek daring, sya kurng paham dan ilmunya kurang masuk.

No	Komentar
9	selama daring kurang paham dengan materi karena cuma tugas
10	selama daring kurang paham dengan materi yang disampaikan
11	selalu daring, tidak efisien
12	Tidak ada, Kuliah Online sudah bisa berjalan dengan baik
13	Tidak ada, Kuliah Online sudah bisa berjalan dengan baik
14	Tidak ada, Kuliah Online sudah bisa berjalan dengan baik
15	mohon untuk tugas-tugas yang di kumpulkan secara online untuk tidak di kirimkan Via WhatsApp, mungkin bisa di kirimkan via email atau schoology sehingga tidak ada lagi tugas yang di contek begitu saja oleh temen-temen yang tidak mau belajar sama sekali
...	...
210	"Terima kasih saya ucapkan kepada Pak dosen telah memberikan kuliah selama 1 semester ini. Ijin, saya mau menyampaikan kritik saran terkait kuliah. Mungkin lebih diperbanyak pemecahan studi kasus terkait Business Intelligent, mengingat ini ada kaitannya dengan sertifikasi Prodi Sistem Informasi. Dari materi, mungkin bisa dibuatkan modul materi dasarnya atau memberikan referensi berupa artikel atau video untuk mahasiswa lebih menguasai topik bahasan materi. Sulit kalau kelas praktik dilakukan online sinkron. Lebih baik diadakan secara offline, akan lebih bagus interaksinya. Setiap pertemuan kelas sinkron sudah direkam oleh dosen dan dishare link YouTubanya, itu saya sangat setuju, mahasiswa dapat mengulang melihat pertemuan kelas."

2. Text Preprocessing

Setelah pengumpulan data langkah selanjutnya adalah *preprocessing*. *Preprocessing* merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem [14]. Pada proses *text mining* dilakukan proses *text preprocessing* yang memiliki tujuan untuk menghasilkan sebuah *set term index* yang bisa mewakili dokumen [15]. Sehingga *text preprocessing* adalah proses menggali informasi yang terdapat dalam suatu teks dan mengubahnya menjadi data terstruktur dalam bentuk nilai numerik [16]. Adapun tahapan *text preprocessing* yaitu sebagai berikut.

a) Filtering dan Cleansing

Pada tahapan ini dilakukan proses *filtering* dan *cleansing* yaitu menggunakan *select* atribut dan *remove duplicate* untuk mengambil kata-kata penting, membersihkan dokumen dari kata yang tidak diperlukan, dan membersihkan kata yang *double* atau duplikat. selanjutnya *dataset* akan meliputi tahap pelabelan manual ke dalam dua kelas sentimen yakni positif dan negatif. Berdasarkan hasil *filtering* dan *cleansing*, dataset yang diperoleh sejumlah 172 komentar, seperti yang ditunjukkan pada Tabel 3 berikut.

Tabel 3. Dataset setelah *Filtering* dan *Cleansing*

class	komentar
negatif	Tolong selama pandemi di permudah untuk bimbingan online Kadang susah sekali ngubungin dosennya.
negatif	Terlalu banyak courses bhs. Indonesia yang saya dan teman-teman saya ikuti hanya untuk melakukan absen secara online saat melakukan kelas online. Kami jadi bingung untuk mengikuti aturan tersebut, jadinya ada yang tidak melakukan absen.
positif	sudah bagus, mungkin perlu meningkatkan komunikasi lagi antara dosen dan mahasiswa, mengingat pandemi seperti ini rentan terjadinya kesalahpahaman via daring ketimbang bertemu langsung. Tidak lupa utk memberikan informasi entah mengenai jadwal perkuliahan atau hal lainnya tidak secara dadakan, karena meskipun stay at home dan study from home, belum menjamin bahwa mahasiswa siap utk diberikan informasi secara mendadak
negatif	Jika mengajar online pastikan bisa berjalan lancar tanpa hambatan.
negatif	Mengajar hanya memberi tugas. Tidak mengadakan meet secara online. Padahal waktu bimbingan beliau sangat menguasai mata kuliah ini. Sayang sekali.
negatif	- penyampaian grup proposal dan skripsi yang kurang jelas, sehingga kebanyakan mahasiswa yang mengambil matkul proposal bergabung ke skripsi yang menyebabkan ketinggalan informasi mengenai proposal skripsi. - kurang siap dalam memberikan informasi kepada mahasiswa - tidak adanya pertemuan/online meet pertama untuk menjelaskan bagaimana proses dan tata aturan proposal skripsi (hanya mencantumkan slide presentasi) - contoh laporan proposal skripsi dan skripai dijadikan 1 sehingga kami bingung t
negatif	Efek daring, saya kurang paham dan ilmunya kurang masuk.
negatif	selama daring kurang paham dengan materi yang disampaikan
negatif	selalu daring, tidak efisien
positif	Tidak ada, Kuliah Online sudah bisa berjalan dengan baik
...	...
negatif	"Terima kasih saya ucapkan kepada Pak dosen telah memberikan kuliah selama 1 semester ini. Ijin, saya mau menyampaikan kritik saran terkait kuliah. Mungkin lebih diperbanyak pemecahan studi kasus terkait Business Intelligent, mengingat ini ada kaitannya dengan sertifikasi Prodi Sistem Informasi. Dari materi, mungkin bisa dibuatkan modul materi dasarnya atau memberikan referensi berupa artikel atau video untuk mahasiswa lebih menguasai topik bahasan materi. Sulit kalau kelas praktik dilakukan online sinkron. Lebih baik diadakan secara offline, akan lebih bagus interaksinya. Setiap pertemuan kelas sinkron sudah direkam oleh dosen dan dishare link YouTubanya, itu saya sangat setuju, mahasiswa dapat mengulang melihat pertemuan kelas."

b) Case Folding

Case folding dengan menggunakan *transform cases* untuk mengubah teks menjadi huruf kecil. Tabel 4 berikut menunjukkan sampel dataset yang dilakukan *case folding*.

Tabel 4. Hasil *Case Folding* pada sampel dataset

No	Label	Komentar
1	negatif	selalu daring, tidak efisien
2	negatif	audio kadang macet saat daring
3	positif	dari segi mengajar dikelas secara online sudah cukup bagus
4	positif	tidak ada yang perlu dikritik karena dari materi, proses pembelajaran, dan interaksi kelas online sudah cukup bagus

No	Label	Komentar
5	negatif	dalam penyampaian materi secara online kurang jelas

b) *Tokenizing*

Tahap dimana dilakukannya penghapusan terhadap semua karakter non alphabet, misalnya simbol, spasi, angka menggunakan *tokenize*. Hasil dataset setelah di *tokenizing* seperti pada Tabel 5 berikut.

Tabel 5. Hasil *Tokenizing* pada sampel *dataset*

No	Label	Komentar
1	negatif	selalu daring tidak efisien
2	negatif	audio kadang macet saat daring
3	positif	dari segi mengajar dikelas secara online sudah cukup bagus
4	positif	tidak ada yang perlu dikritik karena dari materi proses pembelajaran dan interaksi kelas online sudah cukup bagus
5	negatif	dalam penyampaian materi secara online kurang jelas

c) *Stemming*

Proses dilanjutkan dengan menghilangkan imbuhan pada masing-masing kata sehingga menjadi kata dasar, dan dalam tahap ini bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat. Proses *stemming* yang dilakukan seperti mengganti kata “diskusinya” menjadi “diskusi”, “mengajarkan” menjadi “mengajar”, “interaksinya” menjadi “interaksi”, mengganti “penyampaian” menjadi “mengajar”, dst. Tabel 6 berikut adalah sampel dataset setelah dilakukan proses *stemming*.

Tabel 6. Hasil *Stemming* pada sampel *dataset*

No	Label	Komentar
1	negatif	selalu daring tidak efisien
2	negatif	audio kadang macet saat daring
3	positif	dari segi ajar kelas secara online sudah cukup bagus
4	positif	tidak ada yang perlu kritik karena dari materi proses belajar dan interaksi kelas online sudah cukup bagus
5	negatif	dalam mengajar materi secara online kurang jelas

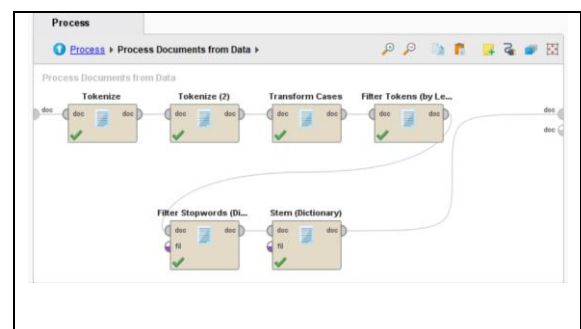
d) *Stopword*

Stopword diproses pada sebuah kalimat yang mengandung kata-kata yang sering keluar dan dianggap tidak penting seperti waktu atau penghubung. Daftar kata yang menjadi *stopword* seperti “aduhh, hehe, yang, via, tolong, terkait, terimakasih, terkadang, terima, kasih, temen, teman, online, kuliah, pandemi, covid, covid19, mahasiswa, dst”. Adapun hasil sampel dataset setelah dilakukan proses *stopword* seperti pada Tabel 7 berikut.

Tabel 7. Hasil *Stopword* pada sampel *dataset*

No	Label	Komentar
1	negatif	tidak efisien
2	negatif	audio macet
3	positif	cukup bagus
4	positif	materi belajar interaksi cukup bagus
5	negatif	materi kurang jelas

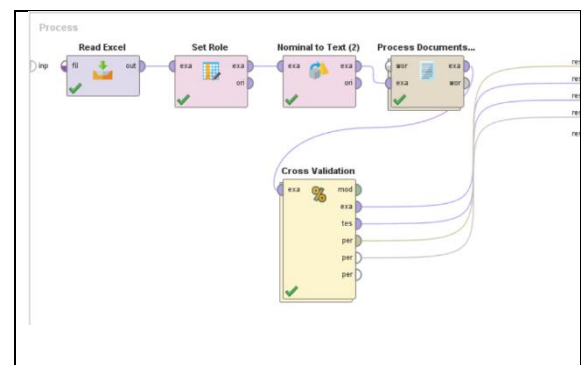
Proses *text processing* secara keseluruhan menggunakan *tools Data Mining* yaitu RapidMiner. Berikut adalah model yang digunakan dalam tahapan *text preprocessing*.



Gambar 2. Model *Text Preprocessing*

3. Penerapan Algoritma KNN

Penerapan algoritma KNN dilakukan dengan RapidMiner, dimana model yang dibangun seperti yang ditunjukkan pada Gambar 3 berikut.



Gambar 3. Model Penerapan Algoritma KNN dalam Sentimen Analisis

Data komentar yang sudah melalui *preprocessing* selanjutnya dilakukan penerapan algoritma KNN dan juga melakukan analisis sentimen terhadap data komentar. Model analisis sentimen dibangun berdasarkan data latihan yang telah disiapkan yang diujicobakan pada pengujian data sejumlah 172 data untuk mengetahui sentimen terhadap pembelajaran daring. Teks yang dianalisis dalam penelitian ini adalah teks dalam berbahasa Indonesia dan menggunakan kosakata dalam bahasa Indonesia. KNN adalah algoritma klasifikasi data yang berdasarkan *data training* dengan jarak yang paling dekat dengan objek. Rumus yang digunakan

dalam perhitungan algoritma KNN adalah rumus *Eucledan Distance*[17]. KNN terbukti mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan dalam sebuah aplikasi [18].

$$d(x, y) = \sqrt{(\sum_{i=1}^n (x_i - y_i)^2)}$$

Dimana:

- d : jarak
- x : data uji
- y : sampel data
- i : variabel data
- n : dimensi data

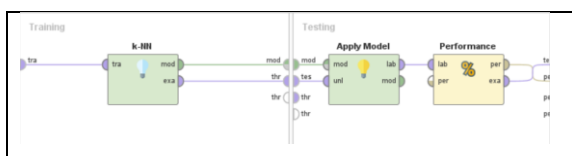
Hasil penerapan algoritma KNN pada data komentar untuk melakukan sentimen analisis seperti yang ditunjukkan pada gambar berikut.

Row No.	class	prediction(cl...	confidence(...	confidence(...
1	negatif	negatif	1	0
2	negatif	negatif	1	0
3	negatif	negatif	1	0
4	negatif	negatif	1	0
5	positif	positif	0	1
6	positif	positif	0	1
7	negatif	negatif	1	0
8	negatif	negatif	1	0
9	negatif	negatif	1	0
10	negatif	negatif	1	0
11	negatif	negatif	1	0
12	negatif	negatif	1	0

ExampleSet (171 examples, 4 special attributes, 458 regular attributes)

Gambar 4. Hasil Analisis Sentimen Positif dan Negatif

Hasil data analisis yang diuji untuk mendapatkan sentimen terhadap persepsi pembelajaran daring di masa pandemi adalah 138 komentar negatif dan 34 komentar positif. Sentimen negatif lebih mendominasi dengan 80.23% dan 19.77% positif. Selanjutnya, peneliti melakukan uji *confusion matrix* dan *cross validation* dalam proses *text mining* menggunakan algoritma KNN dengan model seperti pada gambar berikut.



Gambar 5. Model Uji Performance

Dari 172 data sebagai data uji, hasil data prediksi komentar negatif yakni *True Negatif* (TN) sebesar 127 dan 10 *False Negatif* (FN), sedangkan data prediksi komentar positif *True Positif* (TP) sebesar 14

dan 20 *False Positif* (FP) seperti yang ditunjukkan pada Gambar 6 berikut.

accuracy: 82.48% +/- 7.23% (micro average: 82.46%)			
	true negatif	true positif	class precision
pred. negatif	127	20	86.39%
pred. positif	10	14	58.33%
class recall	92.70%	41.18%	

Gambar 6. Hasil Analisis Confusion Matrix

Berdasarkan data tersebut menunjukkan bahwa komentar mahasiswa terhadap pembelajaran daring pada masa pandemi lebih cenderung negatif. Selanjutnya nilai accracy yang didapatkan algoritma KNN dalam analisis sentimen adalah 82.48% yang membuktikan bahwa KNN memiliki tingkat akurasi yang tinggi dalam melakukan analisis sentimen pembelajaran daring. Berdasarkan uji *cross validation* menunjukkan bahwa nilai *Precision* menggunakan algoritma KNN sebesar 86.39% positif dan 58.33% negatif. Sedangkan nilai *recall* sebesar 92.70% negatif dan 41.18% positif. Selain KNN, peneliti mencoba membandingkan KNN dengan algoritma lain yang sejenis yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes*.

Tabel 8. Perbandingan Tingkat Akurasi

Algoritma	Akurasi
KNN	82.48%
Support Vector Machine	80.69%
Naïve Bayes	68.95%

Berdasarkan Tabel 8 di atas, hasil menunjukkan bahwa SVM memiliki tingkat akurasi sebesar 80.69% dan *Naïve Bayes* sebesar 68.95%. Hal ini menunjukkan bahwa algoritma KNN memiliki nilai akurasi yang paling tinggi meskipun tidak signifikan dari SVM, sedangkan algoritma *Naïve Bayes* memiliki akurasi yang paling rendah dibandingkan KNN dan SVM.

KESIMPULAN

Penelitian ini menerapkan algoritma KNN untuk analisis sentimen terhadap pembelajaran daring. Analisis sentimen menggunakan *database* evaluasi dosen STMIK Primakara dengan kata kunci “daring” dan “online” dalam bahasa Indonesia. Penelitian ini menunjukkan komentar mahasiswa lebih cenderung sentimen negatif. Persepsi negatif dihasilkan karena ketidakpuasan mahasiswa terhadap pembelajaran daring. Hasil *confusion matrix* menunjukkan prediksi komentar negatif yakni *True Negatif* (TN) sebesar 127 dan 10 *False Negatif* (FN), sedangkan data prediksi komentar positif *True Positif* (TP) sebesar 14 dan 20 *False Positif* (FP). Berdasarkan uji *cross validation* menunjukkan bahwa nilai *Precision* menggunakan algoritma KNN sebesar 86.39% positif dan 58.33% negatif. Sedangkan nilai *recall* sebesar

92.70% negatif dan 41.18% positif. Nilai akurasi KNN sebesar 82.48% yang berarti KNN dapat dikatakan akurat dalam melakukan analisis sentimen komentar pembelajaran daring. Selain itu dilakukan perbandingan dengan SVM dan Naïve Bayes, dimana hasil menunjukkan KNN memiliki nilai akurasi yang lebih tinggi meskipun tidak signifikan dari SVM, sedangkan algoritma Naïve Bayes memiliki akurasi yang paling rendah. Hasil penelitian ini membuktikan bahwa pembelajaran daring belum dapat diterapkan secara maksimal di STMIK Primakara, berbagai kendala baik dari dosen maupun mahasiswa menyebabkan ketidakpuasan mahasiswa dalam mengikuti perkuliahan.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang telah mendukung penelitian ini melalui program hibah pendanaan nasional skema Penelitian Dosen Pemula tahun 2022.

DAFTAR PUSTAKA

[1] W. S. A. H. Abdul Bashith, "ANALISIS SENTIMEN PUBLIK TERHADAP PEMBELAJARAN DARING Dr.," vol. 08, no. 01, pp. 190–196, 2020.

[2] R. W. Samsir, Ambiyar, Unung Verawardina, Firman Edi, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.

[3] D. Musfiroh *et al.*, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021.

[4] E. Ramadhanta Machmud Razaq, J. Deden Witasryah, and F. Hamami, "Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9000–9006, 2021.

[5] T. W. Putra, A. Triayudi, and Andrianingsih, "Analisis Sentimen Pembelajaran Daring menggunakan Metode Naïve Bayes, KNN, dan Decision Tree," vol. 6, no. 1, 2021.

[6] Hs. Hermanto and A. Noviriandini, "Analisa Sentimen Terhadap Belajar Online Pada Masa Covid-19 Menggunakan Algoritma Support Vector Machine Berbasis Particle Sarm Optimization," *J. Inform. Kaputama*, vol. 5, no. 1, pp. 129–136, 2021, [Online].

Available:

<https://jurnal.kaputama.ac.id/index.php/JIK/article/view/451>.

- [7] S. Y. Nursyi'ah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, pp. 117–123, 2021.
- [8] A. Tanggu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.
- [9] I. K. Susanto, "Analisis Sentimen dan Topic Modelling Pada Pembelajaran Online di Indonesia Melalui Twitter," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 6, no. 2, p. 85, 2021, doi: 10.31328/jointecs.v6i2.2350.
- [10] N. Chatrina, Siregar, R. Ruli, A. Siregar, and M. Yoga, Distra, Sudirman, "Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)," *J. Teknol.*, vol. 34, no. 1, pp. 102–110, 2020, [Online]. Available: <https://aperti.e-journal.id/teknologia/article/view/67>.
- [11] B. Purba and R. Syahputra, "Implementasi metode Naive Bayes Classifier pada Evaluasi Kepuasan Mahasiswa terhadap Pembelajaran Daring," *InfoTekJar J. Nas. Inform. ...*, vol. 1, 2021, [Online]. Available: <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/4352>.
- [12] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring," *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/te.v19i4.4044.
- [13] P. Puntoriza and C. Fibriani, "Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [14] S. Y. Nursyi'ah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, pp. 117–123, 2021.
- [15] Nengah Widya Utami, "TEXT MINING CLUSTERING UNTUK PENGELOMPOKAN TOPIK DOKUMEN PENELITIAN MENGGUNAKAN ALGORITMA K-MEANS DENGAN COSINE SIMILARITY," *JINTEKS (Jurnal*

- Inform. Teknol. dan Sains*), vol. 4, no. 3, pp. 255–259, 222AD.
- [16] N. Chatrina, Siregar, R. Ruli, A. Siregar, and M. Yoga, Distra, Sudirman, “Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ),” *J. Teknol.*, vol. 34, no. 1, pp. 102–110, 2020, [Online]. Available: <https://aperti.e-journal.id/teknologia/article/view/67>.
- [17] T. W. Putra, A. Triayudi, and Andrianingsih, “Analisis Sentimen Pembelajaran Daring menggunakan Metode Naive Bayes , KNN , dan Decision Tree,” vol. 6, no. 1, 2021.
- [18] A. Tanggu Mara, E. Sedyono, and H. Purnomo, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba,” *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.