

Analisis Sentimen Komentar SIAKAD Menggunakan Metode Naive Bayes Classifier

Dualu Lado Kaka^{1*}, Gergorius Kopong Pati¹, Karolus Wulla Rato¹

¹Program Studi Teknik Informatika, STIMIKOM Stella Maris Sumba, Indonesia

dualuladokaka30@gmail.com*

| Received: 21/09/2023

| Revised: 23/09/2023

| Accepted: 24/09/2023

Copyright©2023 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstrak

Bidang teknologi informasi dan komunikasi berkembang pesat di semua bidang kehidupan. Semua pekerjaan administrasi membutuhkan teknologi. Ketika inovasi menggabungkan teknologi dan informasi, posisi teknologi semakin penting. Jumlah orang yang menggunakan teknologi informasi dan komunikasi dalam kehidupan sehari-hari mereka telah mengalami transformasi yang sangat besar yang tidak pernah terbayangkan sebelumnya. Teknologi informasi telah masuk ke dalam hal-hal terkecil dalam kehidupan manusia, dan salah satu contohnya adalah sistem informasi akademik (SIAKAD), yang berfungsi untuk meningkatkan kualitas layanan akademik. Komentar mahasiswa dikategorikan dan kemudian digunakan untuk melakukan analisis sentimen. Navie Bayes Classifier merupakan teknik utama yang diterapkan dalam penelitian ini. Untuk menilai tingkat keakuratannya, akan dilakukan perbandingan dengan menggunakan pendekatan ini. Emosi positif dan negatif diklasifikasikan sebagai sentimen. Tujuan dari penelitian ini adalah untuk memberitahukan kepada masyarakat tentang SIAKAD dengan memanfaatkan umpan balik mahasiswa dan memahami tingkat keakuratan teknik yang diteliti untuk perbandingan metode. Hasil pengujian akan menjalani pengujian metode dan pengujian akurasi menggunakan alat Rapidminer.

Kata Kunci : Navie Bayes Classifier, SIAKAD

Abstract

In all industries, information and communication technology is expanding very quickly. All administrative tasks have ramifications that genuinely call for technology. When the innovation successfully mixes technology with information, the position of this technology becomes increasingly significant. Information technology has permeated so deeply into even the most insignificant aspects of human life, such as the use of academic information systems (SIAKAD) in raising the caliber of academic services, that there are many users who use it in their daily lives who have witnessed an acceleration of change that was previously unimaginable. Public sentiment was categorized and then used to conduct the analysis. The Navie Bayes Classifier is the main technique applied in this study. To

assess the degree of accuracy, a comparison will be made using this approach. Positive and negative emotions are classified as sentiments. The purpose of this study is to tell the public about SIAKAD by utilizing student feedback and understanding the degree of accuracy of the techniques examined for method comparison. The test results will undergo method testing and accuracy testing using the Rapidminer tool.

Keywords: Navie Bayes Classifier, SIAKAD

1. Pendahuluan

Teknologi informasi kini berkembang dengan pesat. Sistem Pengambilan Keputusan (DSS) adalah satu-satunya metode komputernisasi yang paling banyak digunakan saat ini. Pengambilan keputusan adalah proses memilih antara berbagai pilihan untuk mencapai satu atau lebih tujuan. Setiap perguruan tinggi harus memastikan bahwa kualitas yang diinginkan mahasiswa tersedia, karena ini meningkatkan persaingan perguruan tinggi (Chersoni, 2021; Ernawati & Wati, 2018; Kim et al., 2006).

Untuk memberikan pelayanan kepada Mahasiswa, pegawai dan dosen STIMIKOM Stella Maris Sumba telah menerapkan sistem informasi akademik. Sistem informasi akademik memiliki website yang dapat diakses oleh mahasiswa, pegawai dan dosen untuk memberikan informasi akademik yang mereka perlukan. Salah satu masalah dalam penelitian ini adalah STIMIKOM Stella Maris Sumba belum mengetahui seberapa puas pengguna dengan sistem informasi akademik. Ini karena dengan berkembangnya teknologi, kebutuhan akan informasi yang akurat akan meningkat. STIMIKOM Stella Maris Sumba, telah dilakukan penilaian untuk mengetahui seberapa puas Mahasiswa, pegawai dan dosen dengan layanan sistem informasi akademik setelah dimulainya. Penilaian ini mengevaluasi tingkat kepuasan pengguna sistem dengan memenuhi kebutuhan akademik. Namun, seperti yang ditunjukkan oleh terjadinya kemajuan yang disebutkan di atas sepanjang penelitian ini, sistem yang dimaksud belum melakukan evaluasi yang lebih menyeluruh. Oleh karena itu, penelitian ini berfokus pada kualitas sistem informasi akademik yang memberikan rasa aman bagi penggunanya. Ketika sistem informasi akademik tersedia dan digunakan, pengguna juga akan menemukan bahwa evaluasi informasi layanan akademik masih dilakukan secara manual. Ini berarti bahwa pengguna hanya dapat melihat tanggapan siswa terhadap layanan akademik tanpa mengetahui apakah tanggapan tersebut positif, negatif, atau netral (Liu, 2022; Pang & Lee, 2008).

Analisis sentimen adalah metode untuk memahami, menganalisis, dan secara otomatis mengubah masukan tekstual untuk mengekstrak informasi. Klasifikasi yang penting untuk menyelesaikan permasalahan perekapan angket Mahasiswa. Masih dilakukan penelitian mengenai analisis prior sentience dengan menggunakan algoritma K-Nearest Neighbor. (K-NN) (Salam et al., 2018; Wilianto et al., 2017). Pada penelitian sebelumnya, metode Ensemble Feature dan K-NN digunakan di Twitter untuk menganalisis sentimen industri pariwisata. (Mara et al., 2021; Mentari et al., 2018; Sari, 2020).

Tujuan K-Nearest Neighbor Algorithm (K-NN) adalah untuk mengklasifikasikan objek berdasarkan atribut dan data latih; tidak memiliki model yang dapat dipelajari dari data uji. Oleh karena itu pembelajaran hanya sebatas contoh uji. Deep Belief Network (DBN) mengungguli Nave Bayes dan Support Vector Machine (SVM) dalam analisis tweet berbahasa Indonesia, dan hasilnya lebih unggul (Indini et al., 2022; Winarko, 2017). Tujuan penelitian adalah untuk

memahami hasil analisis tweet akun pornografi homoseksual Indonesia yang dilakukan masing-masing menggunakan Naive Bayes dan DBN .

Tujuan K-Nearest Neighbor Algorithm (K-NN) adalah untuk mengklasifikasikan objek berdasarkan atribut dan data latih; tidak memiliki model yang dapat dipelajari dari data uji. Oleh karena itu pembelajaran hanya sebatas contoh uji. Deep Belief Network (DBN) mengungguli Naive Bayes dan Support Vector Machine (SVM) dalam analisis tweet berbahasa Indonesia, dan hasilnya lebih unggul (Indini et al., 2022; Winarko, 2017). Tujuan penelitian adalah untuk memahami hasil analisis tweet akun pornografi homoseksual Indonesia yang dilakukan menggunakan Naive Bayes dan DBN dengan tingkat akurasi masing-masing sebesar 93,31%, 79,10%, dan 92,18% (Amanda dan Utomo, 2021; Telaumbanua dkk., 2021). Tujuan dari penelitian ini adalah untuk mengidentifikasi sentimen positif atau negatif terhadap data dalam sebuah tweet dan menggunakan informasi tersebut untuk mendorong pengguna Twitter biasa agar menggunakan layanan ini dengan jujur. Hasil analisis sentimen terhadap 500 data.

Penelitian yang menggunakan Naive Bayes Classifier beserta fitur yang dapat mendeteksi data negatif dan pembobotan dengan menggunakan frekuensi frase dan TF-IDF (Mustafa et al., 2018; Saputra, 2018). Hasil percobaan yang dilakukan menggunakan aplikasi yang baru dikembangkan dan alat RapidMiner menunjukkan bahwa amplifikasi dengan frekuensi lebih menguntungkan dibandingkan amplifikasi menggunakan TF-IDF. Support Vector Machine mengungguli Naive Bayes dalam klasifikasi subjek dan kategori. Pada studi di Jawa Barat, metode Naive Bayes Classifier digunakan untuk menganalisis persepsi pengunjung terhadap kawasan resor. Tujuan dari penelitian ini adalah untuk membantu Pemerintah Provinsi Jawa Barat dan Dinas Pariwisata dalam memahami apa saja yang perlu dibicarakan tanpa harus menunggu lama atau berada di lokasi tetap dengan menggunakan Trip Advisor sebagai sumber data dalam pembangunan. dari aplikasi.

Analisis Sentimen dan Kategori Tokoh Publik di Twitter (Ridwan et al., 2013). Dalam penelitian ini, Naive Bayes Classifier digunakan bersama dengan fitur untuk mendeteksi negasi dan pembobotan dengan menggunakan term frekuensi dan TF-IDF (Pudjajana & Manongga, 2018). Hasil pengujian pada aplikasi yang dibangun dan alat RapidMiner menunjukkan bahwa akurasi dengan term frekuensi lebih baik daripada akurasi dengan fitur TF-IDF. Dalam klasifikasi sentimen dan kategori, metode Support Vector Machine mengalahkan metode Naive Bayes. Studi Kasus Jawa Barat: Analisis Sentimen Terhadap Tempat Wisata dari Komentar Pengunjung Dengan Metode Klasifikasi Naive Bayes. Tujuan penelitian ini adalah untuk membantu Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat dalam menentukan tindakan apa yang harus dilakukan tanpa turun langsung dan melihat satu-satu dengan menggunakan Trip Advisor sebagai sumber data. Hasil penelitian menunjukkan bahwa algoritma klasifikasi naive Bayes dapat dengan akurat mengklasifikasikan komentar dan pendapat ke dalam dua kelas, yaitu positif dan negatif (Hidayatullah & Azhari, 2015). Berdasarkan latar belakang yang telah dijelaskan, penulis akan melakukan klasifikasi komentar siswa terhadap kinerja dosen STIMIKOM Stella Maris Sumba dengan menggunakan metode Naive Bayes Classifier.

2. Metode Penelitian

Dalam penelitian ini, metode eksperimen digunakan, dengan tahapan sebagai berikut:

- a) Pengumpulan data: Penulis pertama menggunakan informasi dari situs review TripAdvisor, www.tripadvisor.com. Lima puluh ulasan positif dan lima puluh ulasan

negatif dikumpulkan.

b) Pengolahan Awal Data Preprocessing pada dataset yang dilakukan yaitu:

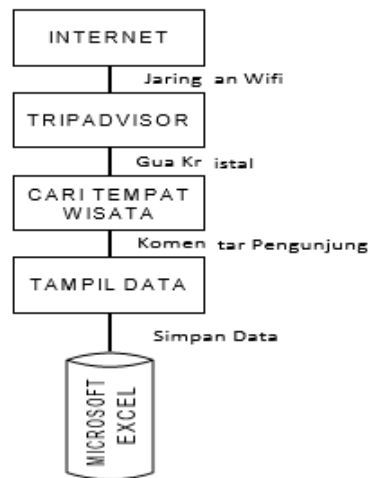
1. Tokenize
2. Filter Stopword
3. Pada tab stopwords, kata sambung, imbuhan, dan akhiran dihasilkan. Termasuk kata “dari”, “yang”, “di”, dan “ke”.
4. N-gram

Sesi ini menggunakan string bi-gram, yang merupakan dua kantong, untuk karakterisasi karakter.

c) Untuk mengukur akurasi yang dihasilkan dalam penelitian ini, metode K-Nearest Neighbor (k-NN) dan Naive Bayes digunakan.

d) Pengujian dan Validasi Hasil divalidasi dengan validasi lipat silang. Untuk menurunkan nilai AUC, digunakan matriks konfusi dan kurva ROC.

Penelitian ini dimulai dengan menentukan tujuan. Setelah itu, data yang relevan untuk penelitian dikumpulkan. Data diambil dari komentar pengunjung. Alat Rapidminer Studio 9.6.0 digunakan untuk memperoleh data dari situs TripAdvisor. Kemudian, data dipindahkan ke Microsoft Excel 2007 untuk digunakan dalam analisis 100 kumpulan data. Proses pengumpulan data digambarkan pada Gambar 1.

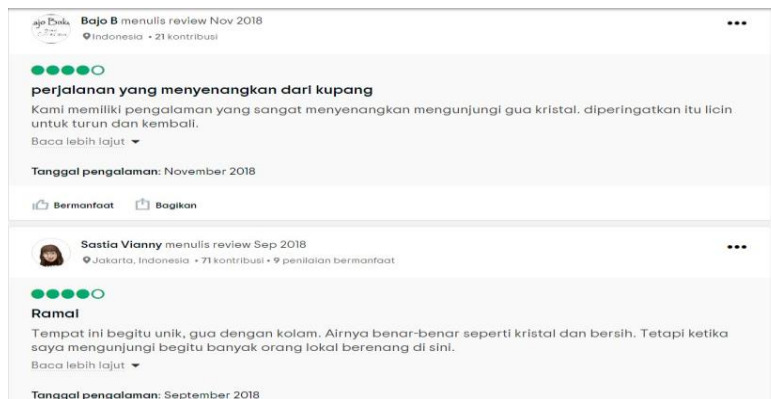


Gambar 1. Proses Pengumpulan Data

Tahapan pengumpulan data yakni:

- a. Terhubung ke internet melalui WiFi
- b. Mengkses Facebook SIAKAD
- c. Mengidentifikasi alamat Facebook yang sesuai dengan analisis
- d. Pengumpulan data berupa komentar mahasiswa masiswa
- e. Langkah terakhir adalah memasukkan data ke dalam editor teks yang menggabungkan komentar pengguna dan sentimen data, baik positif maupun negatif.

Data pengunjung untuk Layanan Sistem Informasi Akademik di STIMIKOM Stella Maris Sumba akan ditampilkan di bawah ini.



Gambar 2. Data Komentar Mahasiswa

2.1 Tahap Preprocessing

Preprocessing dilakukan untuk menghindari data yang tidak konsisten, tidak lengkap, atau gangguan (Fajri et al., 2020). Proses preprocessing teks penelitian termasuk:

- a. Komentar data yang disampaikan pada saat pengumpulan data akan dihilangkan karena mengandung sentimen positif dan negatif sehingga subjek dapat menafsirkan sentimen tersebut.
- b. Mengubah Emoticon: Ini adalah proses mengganti emoticon di komentar dengan kata yang menyerupai emoticon. Daftar konversi emoticon seperti yang ditunjukkan di bawah ini

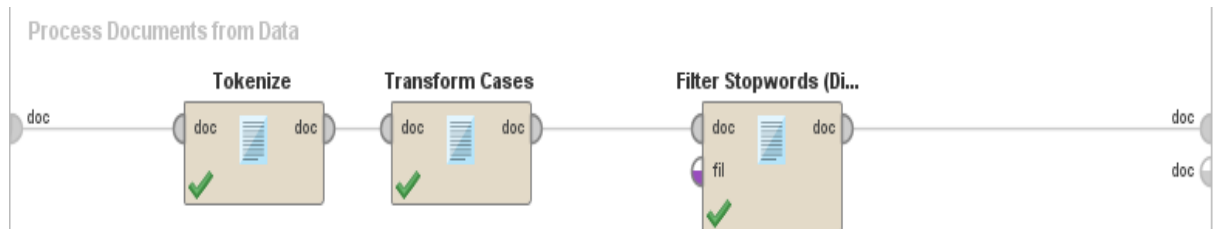
Tabel 1. Konversi Emoticon

Emoticon	Konversi
☺☺☺☺☺)=)	Smile
:D:-D =D	Laugh
☹☹	Sad
:-) :\$ >:[☹	Disappointed
;-) ;)	Blink
:-P :P	Scoffed
:-/;/	Hesitant

- c. Menghapus Simbol: adalah langkah-langkah menghilangkan tanda baca dan simbol dalam saran/ komentar.
- d. Normalisasi Kalimat: adalah proses mengubah data tidak baku menjadi kalimat baku karena banyak komnetar yang menggunakan kalimat tidak baku, yang membuat

pengujian data sulit. Contoh kalimat tidak baku termasuk "bangeetttt", "engga", "tdk", "gak", "tak", "keren", dll. Ini adalah contoh penggunaan tokenizing.

- e. Tokenizing adalah proses membagi teks, seperti paragraf atau kalimat, menjadi bagian-bagian tertentu. Ini adalah langkah akhir dalam pengolahan data menjadi sentimen analisis, di mana peneliti membagi setiap kalimat menjadi masing-masing token biasa. Ini adalah contoh penggunaan tokenizing.



Gambar 3. Tokenizing pada Tolls RapidMiner

Tabel 2. Penerapan Tokenizing

Data Latih	
Input	Output
Tempat ini sangat unik karena memiliki gua dengan air yang bersih seperti kristal. Namun, ketika saya pergi ke sana, saya melihat banyak orang lokal berenang di sini.	[Ini adalah tempat yang sangat unik dengan gua dengan kolam yang airnya seperti kristal dan bersih. Namun, ketika saya mengunjunginya, banyak orang lokal yang berenang di sana]

2.2 Naive Bayes Classifier

Rumus Naive Bayes adalah teknik klasifikasi yang menggunakan probabilitas dan statistik. Algoritma Naive Bayes menggunakan pengalaman masa lalu untuk memprediksi peluang di masa depan, yang dikenal sebagai Teorema Bayes:

$$P(C_j|W_i) = \frac{P(c_j) \times p(w_i|c_j)}{P(w_i)}$$

Keterangan :

x : Data dengan kelas yang belum diketahui

c : Memiliki hipotesis data yang spesifik

$P(c|x)$: Probabilitas hipotesis berdasar kondisi (posteriori probability),

$P(c)$: Probabilitas hipotesis (prior probability)

$P(x/c)$: Probabilitas hipotesis berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

2.3 Algoritma K-Nearest Neighbor (K-NN)

K-NN adalah metode untuk mengklasifikasikan data set berdasarkan klasifikasi training set yang telah dibuat. Rumus k-NN direferensikan dalam percakapan berikut:

$$d_i = \sum_{i=1}^p (x_{2i} - x_{1i})^2$$

Keterangan

x_1 : data latih

x_2 : data uji

i : variabel data

d : jarak

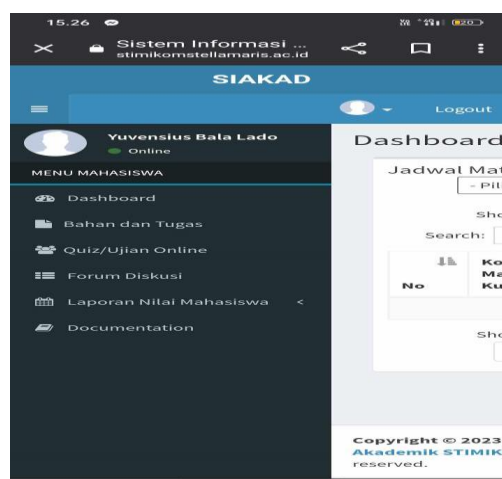
p : dimensi data

3. Hasil dan Pembahasan

3.1 Sistem Informasi Akademik (SIKAD)

SIKAD yang dikembangkan oleh STIMIKOM Stella Maris Sumba, sektor utamanya, memberikan layanan kepada mahasiswa, terutama mereka yang mengambil Prodi Manajemen Informatika dan Teknik Informatika dengan menggunakan SIAKAD.

Data dokumentasi bisa ditemukan di Laporan SIAKAD dengan fasilitas barcode yang digunakan oleh semua stakeholder. Sistem SIAKAD dibuat dengan program PHP dan database MySQL. Google Chrome merupakan framework yang digunakan. Anda dapat mengakses SIAKAD di <http://siak.iainorontalo.ac.id>.



Gambar 4. Siakad

3.2 Sumber Data

Data penelitian dibagi menjadi dua yaitu data latih dan data uji. Pembuatan data latih dimaksudkan untuk membantu menentukan hasil data uji. Gambar berikut menunjukkan data komentar yang disimpan dalam Microsoft Excel:

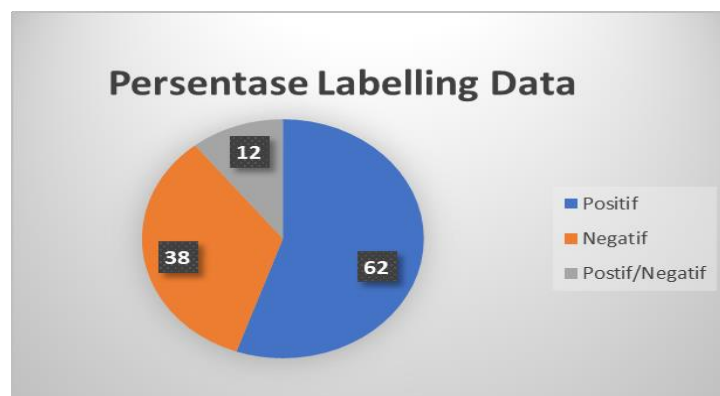
	text	sentimen
1		
2	siakad menyediakan informasi yang di butuhkan, dapat di pahami	positif
3	informasi yang di hasilkan siakad dapat di percaya, sangat sesuai dan tepat waktu	positif
4	sangat di butuhkan, dapat di percaya, tepat waktu, dan relevan	positif
5	informasi sangat di butuhkan dan menghasilkan laporan yang di perlukan serta tepat waktu	positif
6	menghasilkan laporan yang di perlukan dan membantu dalam mengurus adminitrasi	positif
7	menyediakan informasi yang tepat waktu, mudah di gunakan dan terpercaya	positif
8	informasi ter up to date dan tepat waktu dan relevan	positif
9	sangat membantu dan akurat	positif
10	membantu dalam mengurus KRS, mempermudah mahasiswa	positif
11	membantu dalam mengurus adminitrasi, sangat akurat dan dapat di percaya	positif
12	pekerjaan lebih mudah, mudah di gunakan, tepat waktu dan akurat	positif
13	informasi sangat di butuhkan, memeberikan laporan yang sangat jelasa	positif
14	sangat di perlukan dan pekerjaan lebih mudah	positif
15	siakad membantuh mahasiswa mengurus KRS secara tepat waktu dan akurat	positif
16	membentuh mahasiswa dalam pelayanan akademik dan secara tepat waktu	positif
17	sangat mudah digunakan, membantu mahasiswa dalam mengurus adminitrasi	positif
18	tepat waktu, sangat relevan, dan menghasilkan informasi	positif

Gambar 5. Contoh Data Latih

3.3 Persentase Data Hasil Labelling

Pada proses pengambilan data, koisioner yang dibagikan kepada mahasiswa menghasilkan data dengan jumlah 112 komentar. Proses pelabelan dimulai setelah seluruh komentar berhasil disimpan. Salah satu tujuan dari proses pelabelan adalah untuk menentukan apakah komentar yang diterima memiliki nilai positif atau negatif. Namun, beberapa komentar terjadi selama proses pelabelan penelitian ini yang tidak sesuai dengan sentimen komentar yang sebenarnya.

Selanjutnya, hasil pelabelan data selesai menghasilkan data yang terdiri dari 62 komentar bernilai positif, 38 komentar bernilai negatif, dan 9 komentar bernilai positif negatif. Persentase data bernilai positif, negatif, dan negatif, seperti yang ditunjukkan pada gambar 6.



Gambar 6. Persentase Data Hasil Labelling

3.4 Ekstraksi Fitur

Setelah file yang akan digunakan untuk membentuk dataset, data akan diubah menjadi model klasifikasi. Namun, ada beberapa langkah yang harus dilakukan sebelum membentuk model yang baik. Pertama, file xlsx dibaca dan kemudian ditokenisasi. dokumen yang terkandung dalam file tersebut. Penulis ingin mengetahui frekuensi kata yang banyak

diperbincangkan oleh pengunjung wisata berdasarkan hasil tokenisasi yang dilakukan, jadi penulis membuat wordcloud pada gambar 7.



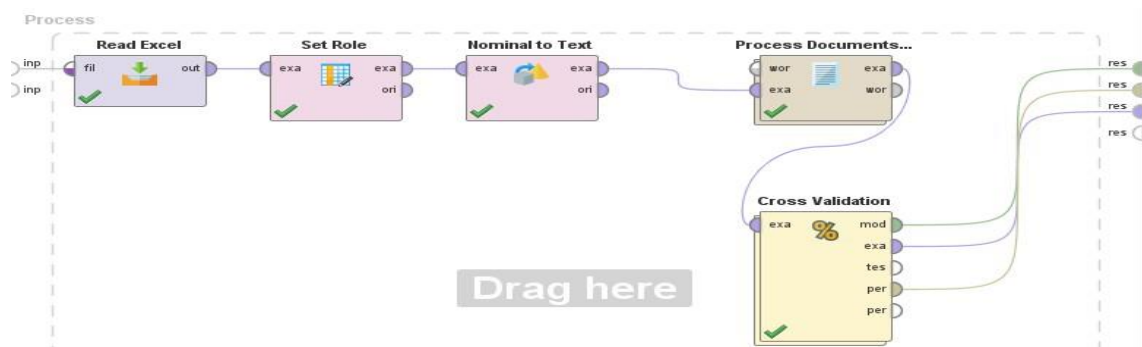
Gambar 7. Persentase Data Hasil Labelling

3.5 Perhitungan Akurasi Dengan RapidMiner

Untuk melakukan penelitian ini, Alat RapidMiner versi 10.1 digunakan untuk melakukan penyelidikan ini. Pemantauan proses dilakukan dengan menggunakan metode Naive Bayes Classifier dan Support Vektor Machine. Tujuan dari percobaan ini adalah untuk mengetahui ambang batas keefektifan kedua metode tersebut.

3.6 Proses Dokument dengan metode Naive Bayes Classifier

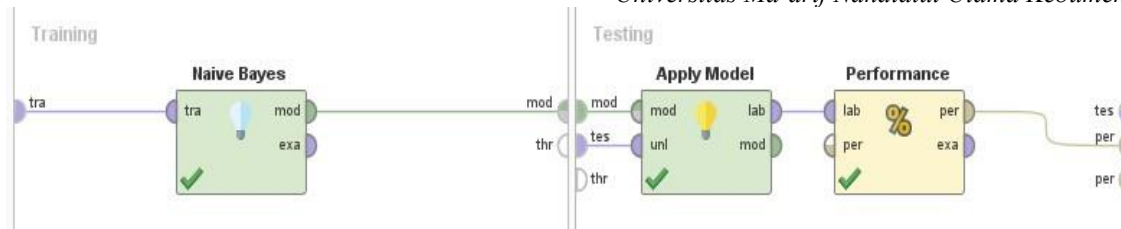
Proses ini diuji dengan data latih Excel, seperti yang ditunjukkan pada gambar berikut:



Gambar 8. Proses Dokumen Data Latih

3.7 Perhitungan Akurasi dengan Naive Bayes Classifier pada RapidMiner

Gambar di bawah ini menunjukkan proses pelatihan dan pengujian pada RapidMiner menggunakan metode Naive Bayes Classifier:



Gambar 9. Proses *Training* dan *Testing* dengan Naive Bayes Classifier pada RapidMiner

Hasil perhitungan klasifikasi menggunakan metode Metode Naive Bayes Classifier diperoleh sebesar 73.47% seperti yang terlihat pada gambar dibawah.

	true positif	true negatif	class precision
pred. positif	62	3	95.38%
pred. negatif	9	38	80.85%
class recall	87.32%	92.68%	

Gambar 10. Akurasi dengan Metode Naive Bayes Classifier pada RapidMiner

Hasil Pengujian Metode K-NN dan Naive Bayes

Metode Naive Bayes Classifier digunakan untuk menguji akurasi tools RapidMiner. Hasil pengujian menunjukkan akurasi 95,38%. Oleh karena itu, tingkat akurasi yang dihasilkan dari metode ini adalah 95,38%.

Tabel 3. Hasil Perhitungan Akurasi

Metode	Tingkat Predikat	True Positif	True Negatif
Naive Bayes Classifier	Pred. Positif	62	3
	Pred. Negatif	9	38
Akurasi pada tools RapidMiner			
Naive Bayes Classifier		89.28%	

4. Kesimpulan

Hasil pengujian akurasi menggunakan metode Naive Bayes Classifier menunjukkan bahwa SIAKAD STIMIKOM Stella Maris Sumba memiliki nilai komentar positif yang sangat baik di antara aplikasi SIAKAD lainnya, dengan tingkat akurasi sebesar 89,28%. Hasil ini juga menunjukkan bahwa SIAKAD memiliki tingkat akurasi yang cukup baik untuk menganalisis sentimen komentar atau tweet dari siswa. Jadi, SIAKAD adalah salah satu aplikasi yang direkomendasikan. Selain itu, peneliti menyarankan agar STIMIKOM lebih mempromosikan aplikasi yang ada sehingga para mahasiswa dapat mengetahuinya dan mengaplikasikannya lebih

mendalam lagi.

Daftar Pustaka

- Amalia, C., & Sibaroni, Y. (2020). Analisis sentimen data tweet menggunakan model jaringan saraf tiruan dengan pembobotan delta tf-idf. *EProceedings of Engineering*, 7(2).
- Amanda, U. R., & Utomo, D. P. (2021). Penerapan Data Mining Algoritma Hash Based Pada Data Pemesanan Buah Impor Cv. Green Uni Fruit. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 5(1).
- Chersoni, E. (2021). *Manfred Stede and Jodi Schneider: Argumentation mining. Synthesis lectures on human language technologies, edited by Graeme Hirst: Morgan & Claypool Publishers, December 2018, 191 pages*. Springer.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, 6(1).
- Fajri, M. S., Septian, N., & Sanjaya, E. (2020). Evaluasi Implementasi Algoritma Machine Learning K-Nearest Neighbors (kNN) pada Data Spektroskopi Gamma Resolusi Rendah. *Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, 3(1), 9–14. <https://garuda.kemdikbud.go.id/documents/detail/1770695>
- Hidayatullah, A. F., & Azhari, A. S. N. (2015). Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter. *Seminar Nasional Informatika (SEMNASIF)*, 1(1).
- Indini, D. P., Siburian, S. R., Nurhasanah, N., & Utomo, D. P. (2022). Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan. *Prosiding Seminar Nasional Sosial, Humaniora, Dan Teknologi*, 325–331.
- Kim, S.-B., Han, K.-S., Rim, H.-C., & Myaeng, S. H. (2006). Some effective techniques for naive bayes text classification. *IEEE Transactions on Knowledge and Data Engineering*, 18(11), 1457–1466.
- Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- Mara, A. T., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. *JOINTER: Journal of Informatics Engineering*, 2(01), 24–31.
- Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2739–2743.
- Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, 4(2), 151–162.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.

- Pudjajana, A. M., & Manongga, D. (2018). Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan Naive Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 313–318.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, 7(1), 59–64.
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). *Analisis Sentimen Data Komentar Sosial Media Facebook Dengan k-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)*.
- Saputra, H. K. (2018). Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Teknologi Informasi Dan Pendidikan*, 11(1), 14–26.
- Sari, R. (2020). Analisis sentimen pada review objek wisata dunia fantasi menggunakan algoritma K-Nearest Neighbor (k-nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1).
- Telaumbanua, F., Purba, J. M., & Utomo, D. P. (2021). Analysis of Online Learning Understanding Patterns at Budi Darma University Using the C5. 0 Algorithm. *The IJICS (International Journal of Informatics and Computer Science)*, 5(2), 118–122.
- Wilianto, L., Pudjiantoro, T. H., & Umbara, F. R. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. *Prosiding SNATIF*, 439–448.
- Winarko, E. (2017). Sentimen analisis tweet berbahasa Indonesia dengan deep belief network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187–198.