

Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree Pada Analisis Sentimen Terhadap Ulasan Aplikasi KitaLulus

Rodyah Mulyani Harun¹, Faisal Fahmi^{2,*}

¹Sekolah Pascasarjana, Magister Pengembangan Sumber Daya Manusia, Universitas Airlangga, Surabaya, Indonesia

²Fakultas Ilmu Sosial dan Ilmu Politik, Departemen Ilmu Informasi dan Perpustakaan, Universitas Airlangga, Surabaya, Indonesia

Email: ¹rodyah.mulyani.harun-2023@pasca.unair.ac.id, ^{2,*}faisalfahmi@fisip.unair.ac.id

Penulis Koresponden: faisalfahmi@fisip.unair.ac.id

Submitted: 29/11/2024; Accepted: 29/12/2024; Published: 30/12/2024

Abstrak—Kemunculan teknologi digital telah mengubah dinamika interaksi antara perusahaan dan calon karyawan, salah satunya dengan terciptanya *platform* pencarian kerja. KitaLulus, salah satu *platform* ternama di Indonesia, memudahkan proses pencarian kerja dengan menyediakan berbagai lowongan dari berbagai perusahaan dalam satu *platform*. Namun, masih terdapat beberapa keluhan dari para pengguna, seperti proses melamar kerja yang kompleks, penyimpanan *file* yang kurang efisien, serta *user interface* (UI) dan *user experience* (UX) yang kurang baik. Disisi lain, Twitter menjadi salah satu tempat yang memuat ulasan pengguna, baik berupa kepuasan ataupun kekecewaan, sehingga dapat dimanfaatkan untuk mengidentifikasi sentimen publik terhadap aplikasi KitaLulus. Karena penting bagi generasi saat ini, maka perlu ada aplikasi pencarian kerja yang berkualitas, di mana rekomendasi untuk peningkatan kualitas aplikasi tersebut dapat diperoleh dari analisis sentimen. Maka dari itu, dilakukan analisis sentimen untuk mengidentifikasi sentimen publik terhadap aplikasi KitaLulus. Analisis dalam penelitian ini menggunakan 600 data ulasan dari Twitter yang kemudian diklasifikasikan sentimennya berdasarkan algoritma Naïve Bayes, KNN, dan Decision Tree. Penelitian ini terdiri dari enam tahap, yaitu pengumpulan data, *cleaning* data, pelabelan data, *preprocessing* data mulai dari SMOTE, *split data*, *transform cases*, *tokenize*, *filter stopwords*, dan *filter tokens (by length)*, klasifikasi sentimen, dan terakhir hasil dan evaluasi. Hasil penelitian, setelah dilakukan SMOTE pada tahap *preprocessing*, menunjukkan bahwa KNN merupakan algoritma terbaik dengan *accuracy* 83.33%, *precision* 80.36%, dan *recall* 71.09%, diikuti oleh Naïve Bayes dan Decision Tree secara berturut-turut.

Kata Kunci: Analisis Sentimen; Naïve Bayes; KNN; Decision Tree; KitaLulus

Abstract—The advent of digital technologies has transformed interaction dynamics between companies and potential employees by creating job search platforms. KitaLulus, one of the leading platforms in Indonesia, facilitates the job search process by providing various vacancies from various companies on one platform. However, there are several complaints from users, such as a complex job application process, inefficient file storage, and poor user interface (UI) and user experience (UX). On the other hand, Twitter is one of the places that contains user reviews, both in the form of satisfaction or disappointment, so that it can be used to identify public sentiment towards the KitaLulus application. Since it is important for the current generation, it is necessary to have a quality job search application, where recommendations for improving the quality of the application can be obtained from sentiment analysis. Therefore, sentiment analysis was conducted to identify public sentiment towards the KitaLulus application. The analysis in this study used 600 review data from Twitter which were then classified by sentiment based on Naïve Bayes, KNN, and Decision Tree algorithms. This research consists of six stages, namely data collection, data cleaning, data labelling, data preprocessing starting from SMOTE, split data, transform cases, tokenize, filter stopwords, and filter tokens (by length), sentiment classification, and finally results and evaluation. The results, after SMOTE was applied at the preprocessing stage, showed that KNN was the best algorithm with accuracy of 83.33%, precision of 80.36%, and recall of 71.09%, followed by Naïve Bayes and Decision Tree respectively.

Keywords: Sentiment Analysis; Naïve Bayes; KNN; Decision Tree; KitaLulus

1. PENDAHULUAN

Kemunculan teknologi digital seperti internet dan media sosial telah mengubah dinamika interaksi antara perusahaan dan calon karyawan atau pencari kerja, salah satunya dengan terciptanya *platform* pencarian kerja. *Platform* pencarian kerja telah merevolusi kapasitas perusahaan dalam mengidentifikasi dan mengevaluasi kandidat karyawan yang potensial dan kapasitas pencari kerja untuk menemukan pekerjaan dengan lebih efisien. Selain itu, *platform* ini juga memberdayakan perusahaan untuk mempromosikan dirinya kepada calon karyawan serta berbagi informasi internal melalui data ulasan karyawan [1].

Salah satu *platform* pencarian kerja adalah KitaLulus. KitaLulus merupakan *platform* pencarian kerja ternama di Indonesia yang memiliki pengguna lebih dari 7 juta orang berusia minimal 18 tahun dan dengan lebih dari 100 ribu perusahaan terdaftar yang tersebar di lebih dari 30 kota. KitaLulus dapat diakses melalui *website* maupun aplikasi *mobile*. Aplikasi KitaLulus terpilih sebagai aplikasi terbaik Google Play pada tahun 2022 dan memiliki lebih dari 5 juta pengunduh [2]. Namun meskipun begitu, masih terdapat beberapa masalah yang dikeluhkan oleh para pengguna, seperti proses pelamaran kerja yang kompleks dan proses penyimpanan *file* yang kurang efisien. Selain itu juga, terdapat masalah terkait *user interface* (UI) dan *user experience* (UX) yang kurang baik ketika mengakses portal KitaLulus sehingga menurunkan ketertarikan pengguna [3]. *Platform* pencarian kerja seperti aplikasi KitaLulus sangat dibutuhkan karena dapat memudahkan pencari kerja dalam mencari informasi mengenai berbagai lowongan pekerjaan dari berbagai perusahaan dalam satu *platform*, yang semuanya dapat dilakukan secara *online* sehingga dapat diakses dari mana saja serta dapat menghemat waktu dan biaya [4]. *Platform* pencarian kerja tidak hanya memudahkan akses

informasi lowongan pekerjaan, namun juga jelas dan terpercaya sumbernya sehingga terhindar dari kejahatan lowongan pekerjaan palsu [5]. Di sisi lain, Twitter merupakan media sosial yang menjadi salah satu tempat yang memuat ulasan pengguna, baik berupa kepuasan ataupun kekecewaan terhadap aplikasi tersebut sehingga dapat dimanfaatkan untuk mencari tahu sentimen publik terhadap aplikasi KitaLulus [6].

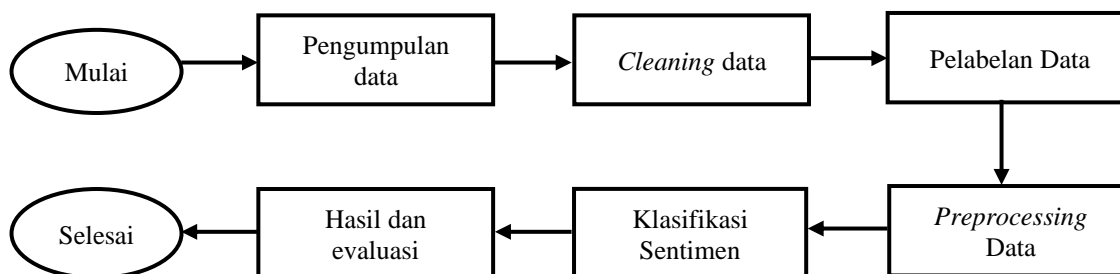
Kualitas aplikasi KitaLulus, baik itu kualitas sistem maupun kualitas informasi terbukti mempengaruhi kepuasan pengguna dalam menggunakan aplikasi tersebut. Artinya apabila kualitas sistem dan kualitas informasi meningkat, maka kepuasan pengguna juga akan meningkat [7]. Oleh karena itu, diperlukan analisis lebih lanjut mengenai kepuasan pelanggan, yaitu melalui analisis sentimen. Analisis sentimen bertujuan untuk memecahkan masalah opini publik, sikap dan emosi individu, apakah mayoritas individu memiliki opini positif, negatif ataupun netral terhadap aplikasi KitaLulus. Hal ini dapat digunakan sebagai bahan peningkatan kualitas aplikasi, yang dapat bermanfaat bagi pengembang aplikasi ataupun manajemen perusahaan dan dapat membantu pengguna baru aplikasi KitaLulus ataupun para pencari kerja untuk memperoleh gambaran mengenai aplikasi tersebut berdasarkan ulasan dari pengguna sebelumnya, sebelum memutuskan untuk menggunakan aplikasi yang sama. Selain itu, peningkatan kualitas layanan aplikasi juga berguna untuk membangun kepercayaan pengguna [5][8].

Analisis sentimen terhadap aplikasi KitaLulus dilakukan dengan menggunakan algoritma Naïve Bayes, K-Nearest Neighbour (KNN) dan Decision Tree. Ketiga algoritma tersebut merupakan algoritma populer dan telah banyak digunakan dalam penelitian-penelitian sebelumnya, seperti penelitian mengenai analisis sentimen aplikasi KitaLulus pada ulasan Google Play Store yang menggunakan Naïve Bayes mendapatkan hasil akurasi sebesar 94% [9], penelitian lain mengenai analisis sentimen terhadap kemacetan di Jakarta yang menggunakan algoritma tiga metode yaitu KNN, Naïve Bayes, dan Decision Tree menunjukkan bahwa hasil algoritma KNN merupakan yang paling baik dengan nilai akurasi tertinggi, yaitu sebesar 86.88%, kemudian Decision Tree dengan akurasi 80.84%, dan terakhir Naïve Bayes dengan akurasi 63.60% [10], sedangkan pada mengenai ulasan aplikasi Threads dan Twitter menunjukkan hasil yang sebaliknya, yaitu algoritma Naïve Bayes merupakan yang paling baik dengan akurasi tertinggi sebesar 65.52% untuk aplikasi Threads dan 77.93% untuk aplikasi Twitter, dilanjut oleh Decision Tree yang memiliki akurasi 65.43% (Threads) dan 72.97% (Twitter), dan terakhir KNN dengan akurasi sebesar 62.46% (Threads) dan 71.57% (Twitter) [11], kemudian penelitian serupa lain yang membahas analisis sentimen *marketplace* Bukalapak dengan membandingkan tiga algoritma yang sama mendapatkan hasil bahwa Decision Tree memiliki akurasi tertinggi yaitu sebesar 89.12%, disusul oleh KNN dengan akurasi 85.03%, dan Naïve Bayes dengan akurasi 73.95% [12], sebaliknya pada penelitian analisis sentimen dari data Twitter, Decision Tree justru mendapatkan akurasi paling rendah yaitu 47.48%, tidak jauh berbeda dengan Naïve Bayes yang mendapatkan akurasi 47.58%, dan yang paling tinggi yaitu KNN dengan akurasi 54.42% meskipun ketiga hasil ini masih terhitung rendah [13]. Perbedaan dari penelitian ini dengan penelitian sebelumnya terletak pada proses *preprocessing data* yaitu SMOTE dan metode pelabelan data yang menggabungkan dua metode sekaligus, yaitu menggunakan leksikon vader dan pelabelan manual.

Berdasarkan permasalahan di atas, penelitian ini melakukan analisis sentimen menggunakan algoritma Naïve Bayes, KNN, dan Decision Tree terhadap ulasan aplikasi KitaLulus yang diambil dari Twitter dengan bantuan RapidMiner untuk membandingkan hasil kinerja klasifikasi dari ketiga algoritma tersebut, yang bertujuan untuk mengidentifikasi sentimen publik terhadap aplikasi KitaLulus.

2. METODOLOGI PENELITIAN

Penelitian ini fokus melakukan analisis sentimen dengan membandingkan tiga algoritma, yaitu Naïve Bayes, KNN, dan Decision Tree yang terdiri dari beberapa tahapan seperti terlihat pada Gambar 1:



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan data

Gambar 1 menunjukkan bahwa tahapan penelitian dimulai dari pengumpulan data. Data dalam penelitian ini dikumpulkan dari Twitter menggunakan python dengan Google Colab. Data yang dikumpulkan adalah berbagai tweet dengan kata kunci, seperti “KitaLulus” dan “Aplikasi Kitalulus”. Hasil pengumpulan data ini kemudian disimpan dalam format csv.

2.2 Cleaning Data

Berdasarkan gambar 1, tahapan penelitian selanjutnya adalah pembersihan atau cleaning data. *Cleaning data* adalah langkah untuk menghapus kata-kata yang tidak perlu, seperti RT atau *retweet*, URL, *mention* atau *username* (@), simbol atau emoji, spasi berlebih dan duplikat kata atau kalimat [14]. Pembersihan data bertujuan untuk menghilangkan *noise* atau kebisingan yang berasal dari kata-kata yang tidak diperlukan, yang dapat mengganggu proses analisis. Proses pembersihan data dilakukan dengan dua cara, yaitu secara manual dan otomatis melalui RapidMiner dengan menggunakan beberapa operator untuk memastikan keoptimalan data sehingga dapat meningkatkan akurasi dan keandalan analisis data [15]. Beberapa operator yang digunakan yaitu seperti *Replace RT*, *Replace URL1*, *Replace URL2*, *Replace Mention*, *Replace Simbol*, *Trim*, dan *Remove Duplicate*.

2.3 Pelabelan Data

Setelah dilakukan cleaning, kemudian dilanjutkan dengan tahap pelabelan data sesuai dengan gambar 1. Pelabelan data adalah proses untuk mengkategorikan kumpulan data ke dalam beberapa kelas, yaitu positif, negatif, atau netral. Proses pelabelan ini bergantung pada nilai dari susunan kalimat yang ada dalam dataset [15]. Pelabelan data dilakukan melalui dua metode, yaitu pelabelan secara otomatis menggunakan metode leksikon vader dan pelabelan secara manual oleh seorang annotator. Metode leksikon vader mendeteksi sentimen berdasarkan nilai dari polaritas kata-kata menggunakan kamus yang tersedia. Sedangkan metode manual dilakukan dengan cara melabeli setiap ulasan satu per satu tanpa bantuan sistem [16]. Penggabungan dua metode ini dilakukan untuk memastikan bahwa hasil pelabelan menjadi lebih akurat dengan mempertimbangkan berbagai perspektif, baik dari mesin otomatis maupun dari evaluasi manusia [15].

2.4 Preprocessing Data

Sebelum melakukan analisis data, dataset yang telah dikumpulkan diolah terlebih dahulu untuk menghasilkan data yang siap pakai. Berdasarkan gambar 1, tahapan tersebut dijelaskan dalam *Preprocessing data*, yang merupakan metode modifikasi data untuk mengidentifikasi kelengkapan dan keakuratan data sebelum dilakukan analisis lebih lanjut [13]. *Preprocessing data* untuk mengubah data mentah menjadi data yang siap digunakan terdiri dari beberapa tahapan, yaitu:

- SMOTE: menyeimbangkan dataset dengan meningkatkan ukuran kelas minor
- Split data*: membagi data menjadi 2 bagian, yaitu data pelatihan dan data pengujian
- Transform cases*: mengubah karakter data menjadi huruf kecil
- Tokenize*: menyaring token untuk menghapus tanda baca dan memecah kalimat menjadi kata berdasarkan spasi
- Filter stopwords*: menghilangkan kata-kata tertentu atau yang tidak memiliki makna
- Filter tokens (by length)*: mengatur kata berdasarkan banyaknya susunan huruf yang dimiliki.

2.5 Klasifikasi Sentimen

Tahap klasifikasi sentimen dilakukan setelah melalui *preprocessing data* berdasarkan pada gambar 1. Data yang telah diberi label terlebih dahulu dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian dengan perbandingan 80:20. Setelah itu, dilanjutkan dengan proses klasifikasi data secara keseluruhan untuk membangun model dan melihat kinerja dari data ulasan tersebut menggunakan *cross validation*. Klasifikasi ini dilakukan dengan menggunakan tiga algoritma, yaitu Naïve Bayes, KNN, dan Decision Tree.

Algoritma Naïve Bayes adalah algoritma yang sangat sederhana dan efisien. Naïve Bayes merupakan teknik machine learning yang populer untuk klasifikasi teks dan memiliki kinerja yang baik di banyak domain. Namun di sisi lain, Naïve Bayes memiliki kekurangan yaitu sangat sensitif dan pemilihan fitur. Terlalu banyak fitur untuk proses klasifikasi tidak hanya meningkatkan waktu perhitungan tetapi juga menurunkan presisi [17]. Selain itu, Naïve Bayes adalah metode probabilistik sederhana yang diterapkan pada teorema bayes dengan kecepatan pemrosesan yang tinggi dalam data besar [18].

KNN atau K-Nearest Neighbor adalah algoritma klasifikasi berdasarkan pada data pembelajaran tetangga atau data yang paling dekat, di mana nilai ini disebut dengan K [19]. Prinsip kerja KNN adalah menentukan dan mengidentifikasi jarak terpendek ke nilai tetangga terdekat pada data latih dengan data yang akan diuji. Nilai k terbaik bergantung pada nilai data, yang mana biasanya nilai k tertinggi mengurangi kekeliruan atau *noise* pada proses klasifikasi, namun menciptakan batas yang kurang optimal antar klasifikasi [10]. KNN memiliki kelebihan dalam hal kemudahan implementasi dan interpretasi, kemampuan mengatasi kumpulan data tidak terstruktur dengan kompleksitas yang tinggi, ketahanan terhadap *noise* dalam data pelatihan serta data dengan jumlah besar. Namun kinerja KNN dapat kurang optimal pada data berdimensi tinggi, karena pengaruh sensitivitas terhadap data outlier dan perhitungan jarak yang rumit sehingga diperlukan pemilihan yang tepat dan cermat dalam menentukan nilai k [20].

Decision Tree adalah algoritma yang didasarkan pada struktur pohon, yang dapat digunakan untuk membagi dataset menjadi himpunan data yang lebih kecil [21]. Algoritma Decision Tree adalah struktur diagram berbentuk pohon, di mana simpul bagian atas pohon disebut sebagai simpul akar, kemudian setiap simpul internal bertugas untuk menguji sebuah atribut, setiap cabang menggambarkan hasil pengujian tersebut, dan setiap simpul daun atau simpul terminal menunjukkan label kelas [14]. Decision Tree mampu mengelola data yang banyak, memiliki fleksibilitas

yang baik, hasilnya mudah dipahami, dan akurasi yang lebih baik. Tetapi, akan menjadil lebih kompleks jika memiliki kelas dan kriteria yang banyak [22].

2.6 Hasil dan evaluasi

Berdasarkan Gambar 1, dijelaskan bahwa tahapan akhir penelitian adalah hasil dan evaluasi. Pada tahap ini, data yang telah diproses hasil akhirnya dievaluasi untuk menilai kualitas keberhasilan analisis. Evaluasi dihitung menggunakan confusion matrix yang mencakup *accuracy*, *precision* dan *recall*. Evaluasi kinerja model klasifikasi yang dibangun didapatkan dari operator *performance* di RapiMiner.

3 HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

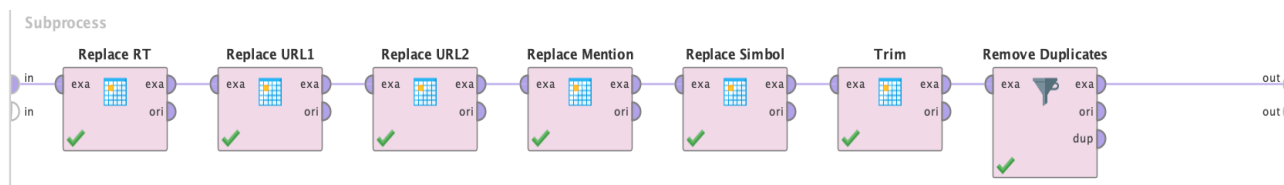
Penelitian ini menggunakan data tweet berbahasa Indonesia yang dikumpulkan dengan cara *crawling* menggunakan Google Colab yang berjumlah 1.306 data pada awal pengumpulan. Data ini merupakan opini-opini dari publik maupun para pengguna KitaLulus. Data-data yang dikumpulkan digabungkan dan disimpan dalam format csv yang kemudian *convert* dalam format excel. Contoh data yang digunakan dalam penelitian ini disajikan pada Tabel 1:

Tabel 1. Contoh Data Tweet KitaLulus

| # | text |
|------|--|
| 1 | Coba pake aplikasi kitalulus Biasa disitu ada info loker kerjaan |
| 2 | @strwbry28 @worksfess Aplikasi kitalulus kak |
| 3 | @jogmfs Hlow sender km cari kerjaan di Jogja kah?memang susah cari kerja di Jogja (temanku cerita) dia udh lamar sana sini jg sm aja. Coba kamu cari yg diluar Jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km & siapin diri. Bismillah lancar ya nder |
| 4 | Hai saya membuat profil di aplikasi KitaLulus. Cek profil saya di sini! https://t.co/9UQLp3BWOG Yuk buat profil profesionalmu sekarang juga di aplikasi KitaLulus |
| 5 | @worksfess Dari aplikasi kitalulus |
| ... | ... |
| ... | ... |
| 1305 | @miawlexa Loker mah banyak di Li Jobstreet Glass door Kaliber Kitalulus IG TELEGRAM WA tergantung fokusmu kemana |
| 1306 | @convomfs dari apk kitalulus dan itu udah kerja sampe skrg |

3.2 Cleaning Data

Cleaning merupakan proses pembersihan data untuk mengurangi *noise* yang banyak terdapat dalam data mentah hasil *crawling*, sehingga data mentah yang awalnya berjumlah 1.306 data berkurang menjadi 1.122 data yang siap dipakai. Berikut merupakan tahapan cleaning data:



Gambar 2. Proses Cleaning

Berdasarkan Gambar 2, proses *cleaning* data yang dilakukan terdiri dari tujuh tahapan, di mana tiap tahapan menggunakan operator tertentu di RapidMiner. Operator pertama yaitu *Replace RT* yang digunakan untuk membersihkan data dari kata “RT” atau retweet di twitter. Operator selanjutnya yaitu *Replace URL1* untuk membersihkan URL atau link di awal kalimat dan *Replace URL2* untuk membersihkan URL atau link di akhir kalimat. Operator *Replace mention* untuk membersihkan *mention* atau tanda yang menyebutkan *username*. Operator simbol untuk membersihkan simbol ataupun emoji. Operator trim untuk menghilangkan spasi berlebih pada data. Dan operator *remove duplicate* untuk menghilangkan data yang sama. Selain itu, proses *cleaning* data juga dilakukan secara manual untuk membersihkan data ulasan yang tidak relevan seperti iklan atau promosi. Hasil dari proses *cleaning* data disajikan pada Tabel 2 sebagai berikut:

Tabel 2. Contoh Hasil Cleaning

| Sebelum <i>cleaning</i> | Setelah <i>cleaning</i> |
|--|--|
| Coba pake aplikasi kitalulus Biasa disitu ada info loker kerjaan | Coba pake aplikasi kitalulus Biasa disitu ada info loker kerjaan |
| @strwbry28 @worksfess Aplikasi kitalulus kak | Aplikasi kitalulus kak |

@jogmfs Hlow sender km cari kerjaan di Jogja kah?memang susah cari kerja di Jogja (temanku cerita) dia udh lamar sana sini jg sm aja. Coba kamu cari yg diluar Jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km & siapin diri. Bismillah lancar ya nder

@worksfess Dari aplikasi kitalulus

@worksfess Glints jobstreet kitalulus. Tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter. Yg aku perhatiin aku lulus di loker yg klop sama isi CV aku

Hlow sender km cari kerjaan di Jogja kahmemang susah cari kerja di Jogja temanku cerita dia udh lamar sana sini jg sm aja Coba kamu cari yg diluar Jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin diri Bismillah lancar ya nder

Dari aplikasi kitalulus

Glints jobstreet kitalulus Tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter Yg aku perhatiin aku lulus di loker yg klop sama isi CV aku

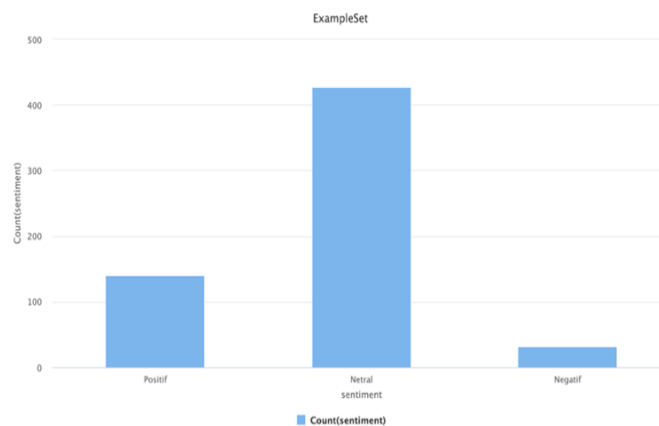
3.3 Pelabelan Data

Setelah melakukan *preprocessing data*, data yang telah dibersihkan selanjutnya dilakukan pelabelan data berupa positif, negatif, atau netral. Pelabelan dilakukan dengan dua cara, yaitu secara otomatis menggunakan leksikon vader dan pelabelan secara manual oleh seorang annotator. Hasil pelabelan akhir adalah penggabungan hasil label yang sama dari kedua metode ini. Jika terdapat perbedaan label maka data ulasan dieliminasi dan tidak digunakan, sehingga jumlah data mengalami pengurangan menjadi 600 data. Berikut disajikan Tabel 3 yang merupakan hasil pelabelan data tweet terkait aplikasi KitaLulus.

Tabel 3. Contoh Hasil Pelabelan Data

| Text | Sentimen |
|--|----------|
| Coba pake aplikasi kitalulus Biasa disitu ada info loker kerjaan | Netral |
| @strwbry28 @worksfess Aplikasi kitalulus kak | Netral |
| @jogmfs Hlow sender km cari kerjaan di Jogja kah?memang susah cari kerja di Jogja (temanku cerita) dia udh lamar sana sini jg sm aja. Coba kamu cari yg diluar Jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km & siapin diri. Bismillah lancar ya nder | Positif |
| @worksfess Dari aplikasi kitalulus | Netral |
| @worksfess Glints jobstreet kitalulus. Tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter. Yg aku perhatiin aku lulus di loker yg klop sama isi CV aku | Netral |

Total data yang diberi label berjumlah 600 data, yang terbagi dalam 141 sentimen positif, 427 sentimen netral, dan 32 sentimen negatif. Berikut Gambar 3 merupakan grafik distribusi data:



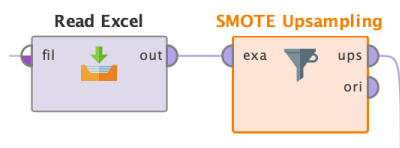
Gambar 3. Jumlah Sentimen Positif, Netral, dan Negatif

3.4 Preprocessing Data

Data ulasan yang diambil dari twitter masih terdapat banyak noise sehingga tidak dapat langsung digunakan. Maka dari itu, diperlukan *preprocessing data* untuk mengubah data mentah menjadi data yang siap digunakan, dengan beberapa tahapan yaitu:

a. SMOTE

Dataset dalam penelitian ini memiliki kelas yang tidak seimbang, di mana terdapat jarak yang jauh antara jumlah kelas netral yang sangat banyak dan jumlah kelas negatif yang sangat sedikit, sehingga berdampak pada penurunan nilai akurasi hasil klasifikasi. Ketidakseimbangan kelas dan jumlah atribut yang banyak pada dataset dapat mempengaruhi hasil kinerja klasifikasi. Maka dari itu, untuk mengatasi ketidakseimbangan kelas dilakukan teknik over-sampling menggunakan SMOTE *upsampling* di RapidMiner [23]. Hasil dari SMOTE *upsampling* adalah dataset yang seimbang, yaitu jumlah data yang sama pada setiap kelas, sehingga dapat meningkatkan kinerja klasifikasi, baik *accuracy*, *precision*, dan *recall* [24]. Proses SMOTE dapat dilihat pada Gambar 4 di bawah:



Gambar 4. Proses SMOTE

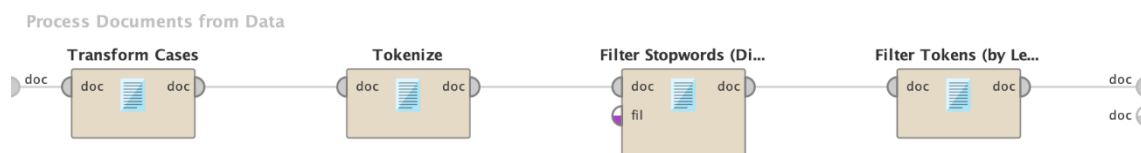
b. Split data

Split data merupakan tahap pembagian data secara otomatis menjadi dua bagian, yaitu data pelatihan dan data pengujian. Pembagian data dalam penelitian ini ditentukan dengan rasio 80% atau 0.8 data pelatihan dan 20% atau 0.2 data pengujian, seperti pada Gambar 5 di bawah:



Gambar 5. Proses Split Data

Berdasarkan metode yang dijelaskan sebelumnya, tahap *preprocessing* data dilanjutkan dengan beberapa operator proses dokumen lainnya dalam RapidMiner, yang terdiri dari operator *transform cases*, *tokenize*, *filter stopwords* dan *filter tokens (by length)*, seperti Gambar 6 di bawah:



Gambar 6. Proses Dokumen

c. Transform cases

Transform cases dalam hal ini yaitu *lower case* digunakan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Tabel 4 berikut merupakan contoh hasil *transform cases*:

Tabel 4. Contoh Hasil Transform Cases

| Sebelum transform cases | Setelah transform cases |
|---|---|
| Coba pake aplikasi kitalulus Biasa disitu ada info loker kerjaan | coba pake aplikasi kitalulus biasa disitu ada info loker kerjaan |
| Aplikasi kitalulus kak | aplikasi kitalulus kak |
| Hlow sender km cari kerjaan di Jogja kahmemang susah cari kerja di Jogja temanku cerita dia udh lamar sana sini jg sm aja Coba kamu cari yg diluar Jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin diri Bismillah lancar ya nder | hlow sender km cari kerjaan di jogja kahmemang susah cari kerja di jogja temanku cerita dia udh lamar sana sini jg sm aja coba kamu cari yg diluar jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin diri bismillah lancar ya nder |
| Dari aplikasi kitalulus | dari aplikasi kitalulus |
| Glints jobstreet kitalulus Tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter Yg aku perhatiin aku lulus di loker yg klop sama isi CV aku | glints jobstreet kitalulus tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter yg aku perhatiin aku lulus di loker yg klop sama isi cv aku |

d. Tokenize

Tokenize diperlukan untuk memecah kalimat menjadi satuan kata terpisah agar dapat dilakukan *filtering* kata untuk tahap selanjutnya. Tabel 5 berikut merupakan contoh kalimat ulasan yang telah dilakukan *tokenize*:

Tabel 5. Contoh Hasil Tokenize

| Sebelum tokenize | Setelah tokenize |
|---|--|
| coba pake aplikasi kitalulus biasa disitu ada info loker kerjaan aplikasi kitalulus kak | 'coba', 'pake', 'aplikasi', 'kitalulus', 'biasa', 'disitu', 'ada', 'info', 'loker', 'kerjaan' 'aplikasi', 'kitalulus', 'kak' |



hlow sender km cari kerjaan di jogja kahmemang susah cari kerja di jogja temanku cerita dia udh lamar sana sini jg sm aja coba kamu cari yg diluar jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin diri bismillah lancar ya nder

dari aplikasi kitalulus glints jobstreet kitalulus tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter yg aku perhatiin aku lulus di loker yg klop sama isi cv aku

'hlow', 'sender', 'km', 'cari', 'kerjaan', 'di', 'jogja', 'kehmemang', 'susah', 'cari', 'kerja', 'di', 'jogja', 'temanku', 'cerita', 'dia', 'udh', 'lamar', 'sana', 'sini', 'jg', 'sm', 'aja', 'coba', 'kamu', 'cari', 'yg', 'diluar', 'jogja', 'nder', 'di', 'beberapa', 'aplikasi', 'like', 'kitalulus', 'glints', 'dst', 'cari', 'yg', 'sesuai', 'passion', 'km', 'amp', 'siapin', 'diri', 'bismillah', 'lancar', 'ya', 'nder', 'dari', 'aplikasi', 'kitalulus', 'glints', 'jobstreet', 'kitalulus', 'tapi', 'sebenarnya', 'bukan', 'faktor', 'aplikasi', 'melainkan', 'kecocokan', 'skill', 'kita', 'sama', 'yg', 'dicari', 'recruiter', 'yg', 'aku', 'perhatiin', 'aku', 'lulus', 'di', 'loker', 'yg', 'klop', 'sama', 'isi', 'cv', 'aku'

e. *Filter stopwords*

Penelitian ini menggunakan *filter stopwords (dictionary)* karena data yang digunakan dimuat dalam bahasa Indonesia dan mengacu pada kamus *stopwords* yang diambil dari situs Kaggle.com, yang memuat sekitar 750 kata-kata berbahasa Indonesia dalam format txt. Tabel 6 berikut merupakan contoh hasil *filter stopwords*:

Tabel 6. Contoh Hasil *Filter Stopwords*

| Sebelum <i>filter stopwords</i> | Seteah <i>filter stopwords</i> |
|--|--|
| coba pake aplikasi kitalulus biasa disitu ada info loker kerjaan aplikasi kitalulus kak hlow sender km cari kerjaan di jogja kahmemang susah cari kerja di jogja temanku cerita dia udh lamar sana sini jg sm aja coba kamu cari yg diluar jogja nder di beberapa aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin diri bismillah lancar ya nder dari aplikasi kitalulus glints jobstreet kitalulus tapi sebenarnya bukan faktor aplikasi melainkan kecocokan skill kita sama yg dicari recruiter yg aku perhatiin aku lulus di loker yg klop sama isi cv aku | coba pake aplikasi kitalulus disitu info loker kerjaan aplikasi kitalulus kak hlow sender km cari kerjaan jogja kahmemang susah cari kerja jogja temanku cerita udh lamar jg sm aja coba cari yg diluar jogja nder aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin bismillah lancar ya nder aplikasi kitalulus glints jobstreet kitalulus sebenarnya faktor aplikasi kecocokan skill yg dicari recruiter yg perhatiin lulus loker yg klop isi cv |

f. *Filter tokens (by length)*

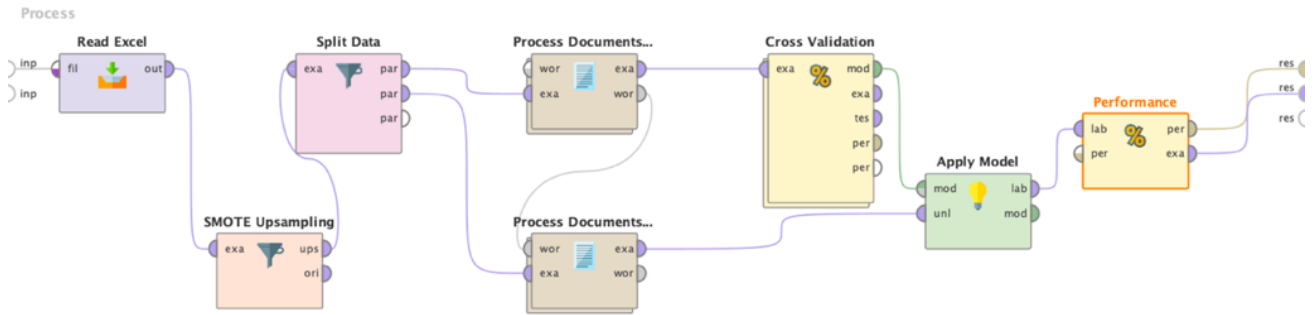
Filter tokens (by length) dalam penelitian diatur dengan ketentuan minimal 4 karakter dan maksimal 10 karakter untuk panjang kata dalam teks untuk lebih memperjelas makna kalimat. Tabel 7 berikut merupakan contoh hasil *filter tokens (by length)*:

Tabel 7. Contoh Hasil *Filter Tokens (by length)*

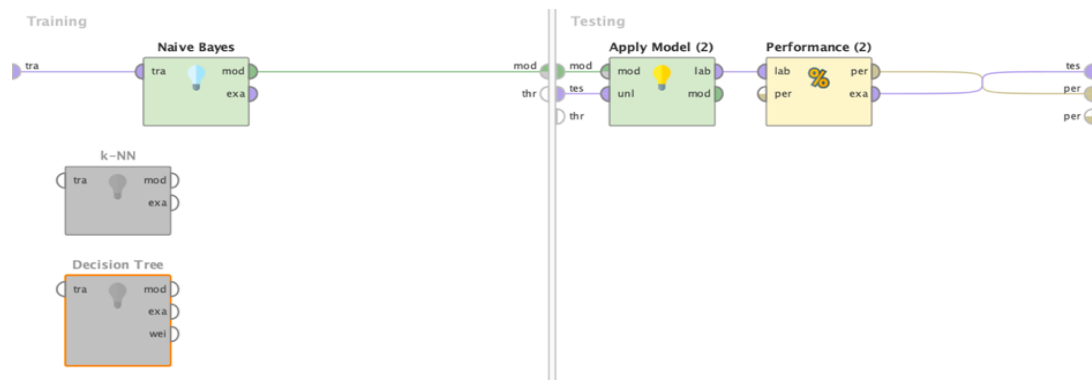
| Sebelum <i>filter tokens (by length)</i> | Setelah <i>filter tokens (by length)</i> |
|--|---|
| coba pake aplikasi kitalulus disitu info loker kerjaan aplikasi kitalulus kak hlow sender km cari kerjaan jogja kahmemang susah cari kerja jogja temanku cerita udh lamar jg sm aja coba cari yg diluar jogja nder aplikasi like kitalulus glints dst cari yg sesuai passion km amp siapin bismillah lancar ya nder aplikasi kitalulus glints jobstreet kitalulus sebenarnya faktor aplikasi kecocokan skill yg dicari recruiter yg perhatiin lulus loker yg klop isi cv | coba pake aplikasi kitalulus disitu info loker kerjaan aplikasi kitalulus hlow sender cari kerjaan jogja kahmemang susah cari kerja jogja temanku cerita lamar coba cari diluar jogja nder aplikasi like kitalulus glints cari sesuai passion siapin bismillah lancar nder aplikasi kitalulus glints jobstreet kitalulus sebenarnya faktor aplikasi kecocokan skill dicari recruiter perhatiin lulus loker klop |

3.5 Klasifikasi Sentimen

Proses klasifikasi sentimen dianalisis menggunakan tiga algoritma secara terpisah, yaitu Naïve Bayes, KNN, dan Decision Tree. Analisis klasifikasi diawali dengan tahap pelatihan data melalui operator *corss validation*, sekaligus melakukan validasi untuk melihat data kinerja awal. Data yang telah dilatih ini kemudian diterapkan pada dokumen baru melalui operator *apply model*. Terakhir dilakukan pengujian kinerja klasifikasi dengan operator *performance*. Berikut merupakan proses klasifikasi sentimen secara keseluruhan pada Gambar 7 dan *cross validation* pada Gambar 8:



Gambar 7. Proses Klasifikasi Sentimen



Gambar 8. Proses Cross Validation

3.6 Hasil dan evaluasi

Pada tahap ini, penelitian melakukan klasifikasi menggunakan tiga algoritma, yaitu Naïve Bayes, KNN, dan Decision Tree. Hasil dari ketiga algoritma tersebut dibandingkan untuk mengidentifikasi algoritma mana yang memiliki kinerja paling bagus dalam menganalisis sentimen terhadap aplikasi KitaLulus.

a. Naïve Bayes

Klasifikasi menggunakan Naïve Bayes menghasilkan *accuracy* 65.15%, rata-rata *precision* 60.38%, dan rata-rata *recall* 62,56%. Berikut merupakan hasil confusion matrix algoritma Naïve Bayes:

accuracy: 65.15%

| | true Netral | true Negatif | true Positif | class precision |
|---------------|-------------|--------------|--------------|-----------------|
| pred. Netral | 29 | 0 | 11 | 72.50% |
| pred. Negatif | 27 | 85 | 2 | 74.56% |
| pred. Positif | 29 | 0 | 15 | 34.09% |
| class recall | 34.12% | 100.00% | 53.57% | |

Gambar 9. Confusion Matrix Naïve Bayes

Berdasarkan Gambar 9, diketahui bahwa nilai *accuracy* algoritma Naïve Bayes merupakan yang paling tinggi, kemudian disusul oleh nilai *recall* dan terakhir yang paling rendah yaitu *precision*.

b. KNN

Klasifikasi menggunakan KNN mendapatkan hasil *accuracy* 83.33%, rata-rata *precision* 80.36%, dan rata-rata *recall* 71.09%. Berikut merupakan hasil confusion matrix algoritma KNN:

accuracy: 83.33%

| | true Netral | true Negatif | true Positif | class precision |
|---------------|-------------|--------------|--------------|-----------------|
| pred. Netral | 76 | 4 | 17 | 78.35% |
| pred. Negatif | 6 | 81 | 3 | 90.00% |
| pred. Positif | 3 | 0 | 8 | 72.73% |
| class recall | 89.41% | 95.29% | 28.57% | |

Gambar 10. Confusion Matrix KNN

Gambar 10 di atas menunjukkan bahwa hasil algoritma KNN yang paling baik terdapat pada accuracynya, dilanjut dengan *precision* dan terakhir *recall*.

c. Decision Tree

Klasifikasi menggunakan Decision Tree mendapatkan hasil *accuracy* 48.48%, rata-rata *precision* 45.91%, dan rata-rata *recall* 47.23%. Berikut merupakan hasil confusion matrix algoritma Decision Tree:

accuracy: 48.48%

| | true Netral | true Negatif | true Positif | class precision |
|---------------|-------------|--------------|--------------|-----------------|
| pred. Netral | 84 | 85 | 16 | 45.41% |
| pred. Negatif | 0 | 0 | 0 | 0.00% |
| pred. Positif | 1 | 0 | 12 | 92.31% |
| class recall | 98.82% | 0.00% | 42.86% | |

Gambar 11. Confusion Matrix Decision Tree

Gambar 11 di atas menunjukkan bahwa hasil algoritma Decision Tree tidak jauh berbeda, paling baik terdapat pada *accuracy*, kemudian *recall* dan *precision*.

Berdasarkan hasil klasifikasi dari tiga algoritma, mulai dari Naïve Bayes, KNN, dan Decision Tree, dapat disimpulkan bahwa KNN memiliki kinerja yang paling baik secara keseluruhan, baik dari tingkat *accuracy*, *precision*, maupun *recall* jika dibandingkan dengan dua algoritma lainnya. Hal ini berarti bahwa algoritma KNN merupakan algoritma yang paling baik digunakan dalam melakukan prediksi sentimen terhadap aplikasi KitaLulus. Hasil perbandingan skor antara ketiga algoritma ini memperlihatkan bahwa skor KNN berada jauh di atas dibandingkan algoritma Naïve Bayes dan Decision Tree, seperti yang dapat dilihat pada Tabel 8.

Tabel 8. Hasil Perbandingan Algoritma

| Algoritma | Accuracy | Precision | Recall |
|---------------|----------|-----------|--------|
| Naïve Bayes | 65.15% | 60.38% | 62.56% |
| KNN | 83.33% | 80.36% | 71.09% |
| Decision Tree | 48.48% | 45.91% | 47.23% |

4. KESIMPULAN

Analisis sentimen terhadap ulasan Aplikasi KitaLulus di Twitter menggunakan algoritma Naïve Bayes, KNN, dan Decision Tree, menemukan bahwa algoritma KNN memiliki kinerja yang paling baik dalam melakukan klasifikasi sentimen. Hal ini dibuktikan dari hasil penelitian setelah dilakukan SMOTE pada tahap preprocessing, menunjukkan bahwa hasil *confusion matrix* KNN lebih tinggi jika dibandingkan dengan algoritma Naïve Bayes dan Decision Tree, di mana KNN mendapatkan *accuracy* sebesar 83.33%, *precision* 80.36%, dan *recall* 71.09%. Kemudian dilanjutkan dengan Naïve Bayes yang mendapatkan *accuracy* sebesar 65.15%, *precision* 60.38%, dan *recall* 62.56%, dan terakhir yaitu Decision Tree dengan *accuracy* sebesar 48.48%, *precision* 45.91%, dan *recall* 47.23%. Dalam hal melakukan prediksi, algoritma KNN juga lebih unggul daripada dua algoritma lainnya untuk klasifikasi sentimen netral dan negatif. KNN berhasil mengklasifikasikan sebanyak 76 data dengan benar dari total 97 data netral. Begitu juga untuk sentimen negatif, KNN berhasil mengklasifikasikan sebanyak 81 data dengan benar dari total 90 data negatif. Sedangkan untuk prediksi sentimen positif, algoritma Decision Tree merupakan yang paling baik, dengan hasil klasifikasi benar sebanyak 12 data dari total 13 data sentimen positif. Hasil analisis yang lebih rinci ini mengindikasikan bahwa meskipun KNN menunjukkan kinerja yang lebih baik secara keseluruhan, baik dari tingkat akurasi maupun kemampuan dalam melakukan prediksi, namun Decision Tree menunjukkan kinerja yang lebih baik dalam memprediksi sentimen positif, sedangkan KNN lebih unggul dalam memprediksi sentimen netral dan negatif. Dalam konteks perbaikan dan peningkatan kualitas aplikasi, maka kelas sentimen negatif yang menjadi fokus utama, sehingga algoritma KNN adalah yang paling tepat untuk digunakan sebagai acuan dalam membangun model analisis sentimen ini. Dari penelitian yang telah dilakukan, ketidakseimbangan kelas data sangat mempengaruhi kinerja model yang berdampak pada hasil akurasi yang dihasilkan. Maka dari itu, untuk memperoleh hasil yang lebih akurat, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih banyak ataupun menggunakan metode *preprocessing* lainnya selain SMOTE untuk mengatasi masalah ketidakseimbangan data.

REFERENSI

- [1] H. Jang, S. Kim, J. Jeon, and J. Oh, "Voice of Employee: Impact of Online Reviews on Company and Job Seeker Matching Performance," in *2023 IEEE International Conference on Big Data (Big Data)*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 6172–6174. doi: 10.1109/BigData59044.2023.10386598.
- [2] KitaLulus, "Platform Pencarian Kerja Berorientasi Komunitas Terbesar dan Teraman di Indonesia." [Online]. Available: <https://www.kitalulus.com/tentang-kitalulus>



- [3] G. Y. Rahayu and A. Indrati, “Perancangan Ulang Antarmuka Portal KitaLulus dengan Menggunakan Metode user Centered Design (UCD),” *JURNAL JUIT*, vol. 3, no. 1, pp. 66–73, 2024, doi: <https://doi.org/10.56127/juit.v3i1.1163>.
- [4] V. Pavani, N. M. Pujitha, P. V. Vaishnavi, K. Neha, and D. S. Sahithi, “Feature Extraction based Online Job Portal,” in *Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1676–1683. doi: 10.1109/ICEARS53579.2022.9752295.
- [5] G. S. Al-Husna, D. Asmarajati, I. A. Ihsanuddin, and R. Mahmudati, “Perbandingan Metode Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen pada Ulasan Pengguna Aplikasi LinkedIn,” *STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123.
- [6] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [7] A. S. Teguh, A. Gandhi, and A. P. Kurniati, “Analisis Kualitas Sistem dan Kualitas Informasi Terhadap Kepuasan Pelanggan (Studi Kasus KitaLulus),” *e-Proceeding of Engineering*, vol. 10, no. 3, pp. 3704–3714, Jun. 2023.
- [8] S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, “Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm,” in *2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 163–167. doi: 10.1109/ISMODE53584.2022.9742906.
- [9] D. Fristikasari, S. Alam, and I. Kurniawan, “Analisis Sentimen Pengguna Aplikasi Kitalulus pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes,” *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 2, pp. 458–473, Sep. 2024, doi: 10.37012/jtik.v10i2.2244.
- [10] S. Rahayu, B. R. Asmoro, and E. Rinaldi, “Classification of Congestion in Jakarta Using KNN, Naïve Bayes and Decision Tree Method,” *Jurnal Syntax Admiration*, vol. 4, no. 7, pp. 928–952, Jul. 2023, doi: 10.46799/jsa.v4i7.654.
- [11] M. Iqbal, A. D. Wiranata, R. Suwito, and R. F. Ananda, “Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.
- [12] E. N. Halim, B. Huda, and A. Elanda, “Perbandingan KNN, Decision Tree Dan Naïve Bayes Untuk Analisis Sentimen Marketplace Bukalapak,” *CESS (Journal of Computing Engineering, System and Science)*, vol. 8, no. 1, pp. 71–79, 2023, doi: <https://doi.org/10.24114/cess.v8i1.41385>.
- [13] E. Apriliyanto and Y. S. Rahayu, “Comparison of Sentiment Analysis from Twitter Data Collection with Naïve Bayes, Decision Tree, and k-Nearest Neighbor Methods,” *Jurnal Ilmiah SINUS*, vol. 22, no. 2, p. 1, Jul. 2024, doi: 10.30646/sinus.v22i2.833.
- [14] N. N. Wilim and R. S. Oetama, “Sentiment Analysis about Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, and Decision Tree,” *IJNMT (International Journal of New Media Technology)*, vol. 8, no. 1, p. 50, 2021, doi: <https://doi.org/10.31937/ijnmt.v8i1.1965>.
- [15] M. B. Husna and W. Gata, “Analisis Sentimen Terhadap Layanan Aplikasi Jenius di Media Sosial Menggunakan Algoritma Long Short-Term Memory,” *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 2, pp. 793–805, 2024, doi: 10.35889/progresif.v20i2.1957.
- [16] D. S. Putri, N. Sulistiyowati, and A. Voutama, “Analisis Sentimen dan Pemodelan Ulasan Aplikasi AdaKami Menggunakan Algoritma SVM dan KNN,” *Journal Sensi*, vol. 9, no. 2, pp. 209–225, 2023, doi: <https://doi.org/10.33050/sensi.v9i2.2914>.
- [17] D. A. Kristiyanti, D. A. Putri, E. Indrayuni, A. Nurhadi, and A. H. Umam, “E-Wallet Sentiment Analysis Using Naïve Bayes and Support Vector Machine Algorithm,” *J Phys Conf Ser*, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012079.
- [18] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 646, Dec. 2021, doi: 10.32493/informatika.v5i4.7622.
- [19] M. F. Fakhrezi, A. F. Rochim, and D. M. K. Nugraheni, “Comparison of Sentiment Analysis Methods Based on Accuracy Value Case Study: Twitter Mentions of Academic Article,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, pp. 161–167, Feb. 2023, doi: 10.29207/resti.v7i1.4767.
- [20] P. A. R. Devi, “Klasifikasi Penyakit Gagal Ginjal Kronis Dengan Metode KNN (Studi Kasus RS Di Kab Gresik),” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1739–1748, Sep. 2024, doi: 10.29100/jupi.v9i3.6226.
- [21] A. Fatkhudin, F. A. Artanto, N. A. Safli, and D. Wibowo, “Decision Tree Berbasis SMOTE dalam Analisis Sentimen Penggunaan Artificial Intelligence untuk Skripsi,” *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 8, no. 2, pp. 495–505, 2024, doi: 10.33395/remik.v8i2.13531.
- [22] M. Maharani and F. Fathoni, “Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode Decision Tree,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 18, no. 1, 2024, doi: <https://doi.org/10.32815/jitika.v18i1.1002>.
- [23] S. G. Barus, “Klasifikasi Sentimen Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor Pada Ulasan Pengguna Aplikasi PeduliLindungi,” *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 3, no. 2, pp. 162–173, 2022.
- [24] E. Eviyanti, B. Irawan, and A. Bahtiar, “Penggunaan Algoritma Naïve Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi AdaKami Di Google Play Store,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, pp. 3879–3885, 2023, doi: <https://doi.org/10.36040/jati.v7i6.8272>.