

Analisis Sentimen Media Sosial *Twitter* Terhadap Tragedi Kanjuruhan

Virgelius Hendrawan Taralandu^{1*}, Magdalena A. Ineke Pakereng²
 Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia
 *e-mail Corresponding Author: 672019220@student.uksw.edu

Abstract

This study aims to analyze public sentiment towards the Kanjuruhan tragedy using the Naive Bayes method. Naive Bayes is a machine learning method that can classify large amounts of data and can automatically label sentiment data from test data or training data based on data models that have been studied by Naive Bayes itself. With this machine learning has obtained results from its own learning. Based on the visualization results, the evaluation values obtained were: Negative prediction (sentiment) = 1,404 while Positive prediction = 1,129, with an accuracy evaluation of 98.20%. For this reason, it can be concluded that from this tweet, more Twitter users have negative sentiments about the Kanjuruhan tragedy. The number of negative sentiments related to this tragedy can be used as a benchmark or reference for those in charge of these activities to be able to find out the public's opinion on the activities that have been held, so that it becomes a lesson in the future in organizing similar activities.

Keywords: Public Sentiment; Twitter Media; Naive Bayes; Visualization

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap tragedi Kanjuruhan menggunakan metode *Naive bayes*. *Naive bayes* adalah metode pembelajaran mesin (*machine learning*) yang dapat melakukan klasifikasi dalam jumlah data yang besar dan dapat melakukan *labelling* otomatis pada sebuah data sentimen dari data uji atau data training berdasarkan model data yang sudah dipelajari oleh *Naive bayes* itu sendiri. Dengan ini *machine learning* telah memperoleh hasil dari pembelajarannya sendiri. Berdasarkan hasil visualisasi nilai evaluasi yang diperoleh adalah: *prediction* (sentimen) Negatif= 1.404 sedangkan *prediction* (sentimen) Positif= 1.129, dengan hasil evaluasi *accuracy* mencapai 98.20%. Untuk itu dapat disimpulkan bahwa dari *tweet* ini lebih banyak pengguna *twitter* yang bersentimen negatif terhadap tragedi Kanjuruhan. Banyaknya sentimen negatif terkait tragedi ini dapat dijadikan sebagai tolok ukur atau acuan bagi penanggung jawab kegiatan tersebut untuk dapat mengetahui pendapat masyarakat terhadap kegiatan yang telah diselenggarakan, agar menjadi pembelajaran di masa mendatang dalam menyelenggarakan kegiatan yang serupa.

Kata kunci: Sentimen Masyarakat; Media Twitter; Naive Bayes; Visualisasi

1. Pendahuluan

Penelitian ini dengan judul analisis sentimen media sosial twitter terhadap tragedi kanjuruhan merupakan suatu tragedi yang penting untuk diteliti fakta narasari berdasarkan data tweet pada media sosial twitter, dikarenakan ini merupakan tragedi yang besar serta mengakibatkan trauma yang mendalam bagi bangsa Indonesia yang dimana pada tanggal 1 oktober 2022 telah menjadi perbincangan di beberapa media sosial terkhususnya twitter yaitu sebuah tragedi di stadion Kanjuruhan. Usai pertandingan antara Arema FC dan Persebaya Surabaya di Stadion Kanjuruhan, Kabupaten Malang, Jawa Timur, yang menimbulkan korban sebanyak 712 orang dengan 132 orang meninggal dunia, 96 orang luka berat, dan 484 orang luka ringan. Kasus ini merupakan sebuah kasus yang cukup membuat pro kontra di kalangan masyarakat Indonesia, serta melahirkan komentar yang positif maupun yang negatif dari pandangan masyarakat Indonesia. Komentar atau pendapat masyarakat pun berbeda-beda, di kalangan usia muda sampai usia lanjut mempunyai pandangan serta pendapatnya masing-masing, terkhususnya terhadap penanganan kerucuhan di tempat kejadian.

Dilihat dari data tweet perbedaan pendapat pro dan kontra masyarakat pada media sosial di twitter yang saling menyalahkan berbagai pihak, terkhususnya pihak penyelenggara serta pihak keamanan, ini merupakan masalah baru bagi pihak penyelenggara yaitu PSSI (persatuan sepak bola seluruh indonesia) dan PT.LIB (PT Liga Indonesia Baru) sebagai badan teknis yang bertanggung jawab mengurus liga ini tidak dapat mengetahui beberapa keseluruhan komentar/sentimen masyarakat terhadap kegiatan sepak bola yang telah mereka adakan.

Pada penelitian ini, data yang diambil berdasarkan data tweet menggunakan API (antarmuka pemrograman aplikasi) Twitter dengan melakukan request ke twitter development untuk mendapatkan data informasi. Kemudian dilakukan pelabelan (labelling) data untuk data latih, kemudian dilakukan processing data, dan dibuat modeling data untuk labelling otomatis dengan metode Naïve Bayes, Metode ini memberikan hasil akhir dari berbagai sentimen negatif dan positif dari labeling otomatis pada tools rapidminer. Penelitian ini diharapkan memberikan hasil analisis sentimen tweet di twitter yang dapat membantu pihak penyelenggara untuk dapat mengetahui sentimen terhadap tragedi Kanjuruhan ini.

Oleh karena itu penelitian ini dilakukan agar pihak penyelenggara dan pihak penanggung jawab kegiatan dapat mengetahui sentimen masyarakat terhadap tragedi ini, melihat rata-rata sentimen dari tweet positif serta negatif dari masyarakat terhadap kegiatan yang telah diselenggarakan oleh pihak penyelenggara. juga sebagai acuan untuk beberapa organisasi penanggung jawab kegiatan tersebut agar dapat mengetahui pendapat masyarakat terhadap kegiatan yang telah diselenggarakan, agar menjadi pembelajaran dan pengalaman penting ke depannya dalam menyelenggarakan kegiatan yang serupa.

2. Tinjauan Pustaka

penelitian terdahulu yang berjudul “Penerapan Teks Mining untuk mengidentifikasi pengguna twitter terhadap fenomena peran DPR RI”, menjelaskan terkait pengumpulan data yang dilakukan berdasarkan kata kunci serta hastag dalam jumlah tertentu, dimana setelah datanya terkumpul kemudian diolah menggunakan metode text mining dan hasil olahnya diproses dengan algoritma support vector machine (SVM) dalam mencari klasifikasinya. Penelitian ini menghasilkan sebuah sistem otomatis yang dapat mengklasifikasikan/memilih mana komentar yang negatif serta mana komentar yang positif [1].

Pada penelitian sebelumnya juga dengan judul “Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia”, Text mining diterapkan untuk mengelompokkan metode K-means pada data tweet Shopee Indonesia untuk mengetahui jenis konten tweet yang paling banyak di-retweet oleh followers Shopee Indonesia. Hasil yang diperoleh adalah tipe konten pada grup dengan retweet tinggi, antara lain kuis hadiah (cluster 4, cluster 20, dan cluster 28), ulang tahun Shopee Indonesia (cluster 11), dan hobi, kuis, dan gaya hidup (cluster 7), sehingga Shopee Indonesia dapat menggunakan konten tweet jenis ini untuk beriklan ke pengguna Twitter [2].

Pada Penelitian terdahulu sebelumnya “Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial”, Text mining mengacu pada proses penggalian informasi berkualitas tinggi dari teks, salah satu metode klasifikasi yang dapat digunakan adalah algoritma K-nearest neighbor. Berdasarkan teori kepribadian Big Five, hasil dari penelitian menyimpulkan bahwa tingkat akurasi yang diperoleh adalah 92,92%, dari 550 data memiliki ciri kepribadian Openness tertinggi yaitu 239, Conscientiousness 16 data, Extraversion 173 data, untuk Agreeableness sebanyak 50 data, neurotic menjadi 33 data, dan 39 data yang tidak dapat diklasifikasikan [3].

Pada penelitian sebelumnya dengan judul “Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes”, Dalam klasifikasi email, nilai probabilitas dihitung berdasarkan kemunculan kata dalam data email. Pengujian akurasi sistematis ditampilkan sebagai grafik nilai akurasi, positif salah dan negatif salah. Hasil pengujian menunjukkan bahwa aplikasi ini dapat digunakan dan mempunyai nilai akurasi sistem sebesar 89,6% [4].

Pada penelitian sebelumnya dengan judul “Penerapan text mining analisis sentimen mengenai vaksin covid -19 menggunakan metode naive bayes”, pengguna telah memberikan berbagai komentar, baik positif maupun negatif. Untuk mengetahui tanggapan masyarakat terhadap vaksin COVID-19, penelitian ini melakukan jajak pendapat terhadap vaksin COVID-19 dengan menggunakan pendekatan Naive Bayes. Naive Bayes memiliki peluang bagus untuk melakukan klasifikasi. Berdasarkan hasil pengujian dengan 100 data latih, 3 di antaranya

kemudian disampling ulang menggunakan teknik pengindeksan data, analisis opini dari pengguna Twitter tentang vaksin COVID-19 dan didapatkan nilai akurasi sebesar 100% [5].

Pada penelitian sebelumnya dengan judul "Penerapan Text Mining Untuk Pengelompokan Penelitian Dosen", dilakukan dengan menggunakan text mining untuk mengelompokkan judul-judul penelitian dosen. Pengelompokan ini bertujuan untuk memperjelas topik penelitian dosen. Berdasarkan hasil tersebut, dapat membantu program pengajaran dan peneliti lain melihat tren penelitian yang dilakukan oleh fakultas. Selain itu, program akademik dapat menggunakan hasil penelitian ini untuk menentukan pembimbing mahasiswa yang mengambil mata kuliah skripsi. Selain itu, hasil penelitian ini dapat membantu fakultas menemukan relasi/mitra dalam penelitian fakultas selanjutnya [6].

Pada penelitian sebelumnya dengan judul "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan" dilakukan dengan menggunakan text mining, Atribut yang digunakan dalam proses klasifikasi pelanggan adalah jumlah pembelian, periode waktu dan lokasi. Hasil sistem klasifikasi adalah 23 reaksi benar dan 2 reaksi salah. Berdasarkan hasil yang diperoleh dengan metode Confusion Matrix menunjukkan nilai presisi mencapai 92%, nilai presisi 100%, nilai recall 91%.[7].

Pada penelitian sebelumnya dengan judul "Penerapan Text Mining Dengan Menggunakan Metode TF-IDF Untuk Menentukan Genre Dari Komik" Penelitian ini menggunakan text mining dan TF-IDF untuk mendefinisikan genre komik. Text mining dapat diartikan sebagai pencarian informasi baru yang sebelumnya tidak diketahui oleh komputer dengan cara mengekstrak informasi secara otomatis dari berbagai sumber [8].

Pada penelitian sebelumnya dengan judul "Penerapan Text Mining Pada Sistem Penyeleksian Judulu Skripsi Menggunakan Algoritma Latent Dirichlet Allocation(LDA)", merancang sebuah sistem yang dapat merekomendasikan kelayakan text mining menggunakan judul skripsi yang diajukan menggunakan algoritma LDA. Algoritme LDA mampu mengidentifikasi topik dari koleksi dokumen dan jumlah kemunculan topik tersebut [9].

Pada penelitian sebelumnya dengan judul " Application of Text Mining for Classification of COVID-19 Vaccination Hoax News Using Support Vector Machine Algorithm" dengan menggunakan 1.000 data berita Indonesia, 500 berita non-fraud, 500 berita penipuan, 500 data yang diperoleh dari Twitter Mining digunakan sebagai data pelatihan untuk data penelitian. Data dikumpulkan dari portal berita online Indonesia dengan menggunakan metode text mining. Hasil penelitian menunjukkan bahwa akurasi tertinggi 89,27% dicapai dengan Support Vector Machine 10 kali lipat dan akurasi tertinggi 89,48% dengan Naive Bayes 5 kali lipat[10].

Pada penelitian sebelumnya dengan judul " Penerapan Text Mining untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K- Means Clustering", Cluster text mining menggunakan metode K-means clustering dengan jumlah cluster terbaik diperoleh dari metode koefisien siluet data tweet Twitter @blibliidotcom untuk menentukan jenis konten tweet @blibliidotcom yang paling banyak diikuti retweet. Tweet yang paling banyak diretweet dan difavoritkan adalah promo dan flash sale, sehingga Blibli Indonesia dapat menggunakan tweet tersebut untuk mempromosikan Twitter di media sosial, karena anggota akun Twitter @blibliidotcom menyukai tweet berhadiah[11].

Pada penelitian sebelumnya dengan judul " Text Mining - Analisis Teks Terkait Isu Vaksinasi COVID-19 (Text Mining - Text Analysis Related to COVID-19 Vaccination Issues" pada penelitian ini menggunakan TF-IDF dan LSI/LSA yang dimana adalah metode penambahan aturan yang digunakan dalam penambahan data. Hasil analisis penelitian ini dapat mempengaruhi data lainnya. Misalnya, analisis pendapat pengguna digunakan untuk melihat gambaran luas topik atau topik diskusi penting, serta contact point untuk isu-isu terkait vaksinasi COVID-19[12].

Pada penelitian sebelumnya dengan judul " Penerapan Text Mining Dengan Algoritma Naive Bayes Untuk Mengklasifikasikan Sentimen Rakyat Terhadap Minyak Goreng Subsidi Pemerintah" Penelitian ini menggunakan algoritma Naive Bayesian untuk mengklasifikasikan pendapat masyarakat Indonesia terhadap subsidi minyak goreng pemerintah dan didapatkan akurasi yang dihasilkan adalah 69,4 % dengan presisi 65,3%, sedangkan recall yang dihasilkan 31,6% dan tingkat kesalahan 30,6%[13]

Pada penelitian sebelumnya dengan judul " Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification". Studi ini membandingkan klasifikasi berdasarkan judul artikel. Algoritma yang digunakan adalah Support Vector Machine (SVM). menggunakan empat kernel SVM termasuk kernel linier, kernel polinomial, kernel sigmoid

dan inti RBF. Distribusi data menggunakan distribusi uji latih dibagi menjadi empat skenario yaitu 60:40, 70:30, 80:30 dan 90:10. Hasil survei diukur setelah pengujian model dengan nilai presisi, akurasi, daya ingat, dan ukuran F. Hasil terbaik adalah akurasi 70%, akurasi 75%, 69 persen f-measure 71% dalam skenario referensi 90:10 dan inti linier[14]

Pada penelitian sebelumnya dengan judul "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid – 19 Menggunakan Metode Naive Bayes", Pengguna telah memberikan berbagai komentar, baik positif maupun negatif. Untuk mengetahui reaksi masyarakat terhadap ketersediaan vaksin covid-19, penelitian ini melakukan analisis sentimen terhadap vaksin covid-19 dengan menggunakan metode Naive Bayesian. Berdasarkan hasil pengujian, dari 100 dataset latih yang kemudian diseleksi ulang menggunakan teknik data indexing, ditemukan 34 dataset yang mencapai akurasi 100% dalam analisis opini pengguna Twitter tentang vaksin COVID-19[15]

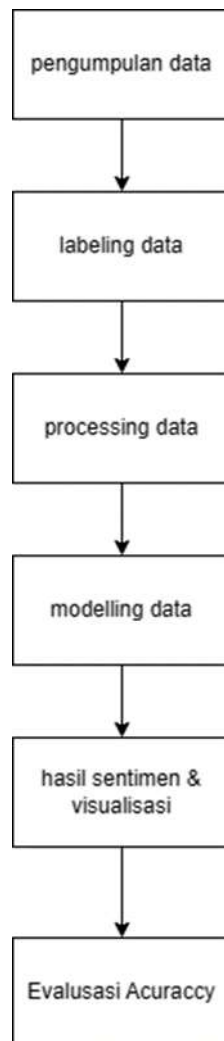
Adapun Fokus dari penelitian ini yaitu untuk dapat mengetahui berapa banyak sentimen/nilai sentimen rata-rata dari pandangan masyarakat di media sosial twitter terhadap tragedi kanjuruhan, fokus penelitian ini penulis melakukan penerapan *text mining* menggunakan metode naive bayes sebagai modelnya, dan datanya diambil dari API twitter, serta berfokus untuk mencari hasil sentimen negatif dan sentimen positif dari 2.533 data *tweet* serta hasil pengujian yang di tentukan dengan accuracy model.

3. Metodologi

Jenis penelitian yang digunakan merupakan penelitian sentimen positif dan negatif berdasarkan data tweet yang diselesaikan dengan metode Naive Bayes. Penelitian ini bertujuan untuk menggambarkan secara umum menggunakan suatu fakta pendapat pengguna twitter saat ini untuk dapat mengatasi sebuah masalah yang akan disajikan ataupun dieksplorasi. Objek penelitian adalah percakapan di media sosial twitter terhadap tragedi di Kanjuruhan. Peneliti melakukan pengumpulan data menggunakan pendekatan media sosial yaitu twitter. Tools penarikan data (API) twitter yang digunakan yaitu netlytic. API atau biasa disebut antarmuka pemrograman aplikasi dari media sosial twitter digunakan dalam software netlytic melalui pengumpulan data serta menemukan pola komunikasi. Netlytic sendiri adalah sebuah platform software online yang berbasis cloud serta mempunyai fungsi untuk melakukan penarikan data, serta dapat digunakan juga untuk menganalisis teks dan pola jaringan sosial dalam sosial media. Tools analisis sentimen yang digunakan yaitu rapidminer, rapidminer sendiri merupakan perangkat lunak yang bersifat terbuka (open source). Rapidminer merupakan solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang baik, dengan mengandalkan kurang lebih 500 operator data mining, termasuk operator yang digunakan untuk melakukan penginputan, output, melakukan preprocessing data, serta melakukan visualisasi hasil data. Rapidminer sendiri dapat bekerja di semua sistem operasi, karena rapidminer sendiri ditulis dengan menggunakan bahasa pemrograman Java.

Pada saat pengumpulan data menggunakan tools netlytic, dapat diuraikan dengan cara mereduksi data, dengan memilih data yang sesuai dengan topik yang akan dikaji, kemudian penyajian data. Penyajian data sangat diperlukan untuk dapat mudah dipahami, berbagai bentuk data dapat digambarkan dalam bentuk skema, tabel, dan gambar. Pada penelitian ini, digunakan tweet sebagai objek data latih. Kemudian dari objek data latih ini digunakan untuk menjadi patokan dari data uji yang belum ada label sentimen artinya data latih ini digunakan sebagai data latihan untuk Naive Bayes melakukan klasifikasinya. Kemudian dibuatlah labelling otomatis menggunakan metode Naive Bayes.

Kemudian yang terakhir adalah penarikan kesimpulan atau verifikasi. Data yang didapat akan dianalisis, serta proses penarikan kesimpulan pada tahap awal masih lemah, masih terbuka serta skeptis. Setelah pengumpulan data dan processing serta labelling sentimen sudah selesai, maka penarikan kesimpulan akan tercapai. Penelitian ini dilakukan melalui beberapa tahapan penelitian, di antaranya yaitu tahap pengumpulan data, tahap labelling data latih, tahap Processing, tahap modelling dan tahap analisis hasil sentimen. Untuk gambarannya dapat dilihat pada Gambar 1.



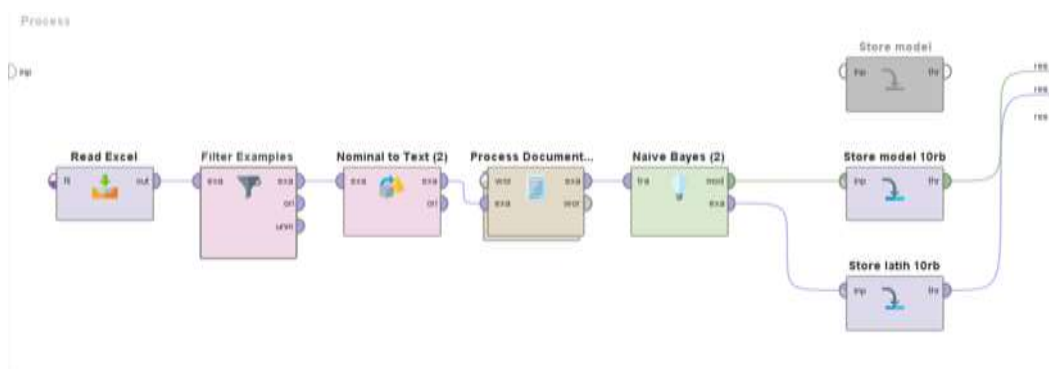
Gambar 1 Tahapan penelitian

Pada saat analisis, penelitian ini dilakukan dengan beberapa tahap. Dimulai dari tahapan registrasi, kemudian ke tahap editing untuk membuat project data baru. Di tahap ini, dilakukan request data baru dengan memilih button "new dataset" kemudian masukan keyword atau kalau di twitter itu hashtag Kanjuruhan, kemudian tahap berikutnya yaitu preview, pada tahap ini netlytic menampilkan data hasil pencarian yang digambarkan dalam bentuk tabel pada format excel dan csv, file preview ini dapat di-download. Tahap berikutnya adalah melakukan labelling data. Pada bagian ini, excel untuk merapikan data text yang masih berantakan dan juga dilakukan labeling data latih sebagai data latihan untuk metode Naïve Bayes, kemudian pada tahap selanjutnya digunakan rapidminer untuk melakukan processing data, pada tahap ini dilakukan filter examples berdasarkan data sentimen yang tidak kosong (is not missing), kemudian mengubah data nominal menjadi data teks (nominal to text) setelah itu melakukan proses dokumen dengan melakukan ekstraksi kata-kata penting data tweet atau bisa diartikan sebagai suatu teknik untuk menghitung kemunculan frekuensi suatu kata-kata dari data tweet, kemudian setelah itu melakukan set role dari sentimen dengan target role adalah label dan nama atribut adalah text, setelah itu data yang sudah siap dilanjutkan pembelajaran menggunakan metode Naïve Bayes agar metode ini tahu data teks seperti apa yang sentimennya positif atau negatif, kemudian data hasil dari pembelajaran Naïve Bayes tadi disimpan dalam store model untuk data modelnya, dan untuk data latihnya disimpan dalam store latih, kedua data ini disimpan untuk melakukan penelitian tahap berikutnya yaitu tahap penggabungan data sentimen yang sudah dipelajari oleh Naïve Bayes dan data yang masih kosong sentimennya.

Kemudian pada tahap berikutnya adalah tahap modelling data dimana tahap ini adalah tahap data latih dan data yang masih kosong sentimen akan diisi secara otomatis menggunakan model dari Naïve Bayes yang sudah dipelajari dari data latih. Hasilnya adalah data yang sudah digabungkan sudah berisi sentimen negatif dan positif. Berdasarkan data ini dapat disimpulkan dengan menghitung jumlah data positif dan negatif, kemudian dari data ini dilakukan analisis terkait sentimen media sosial twitter terkait tragedi Kanjuruhan.

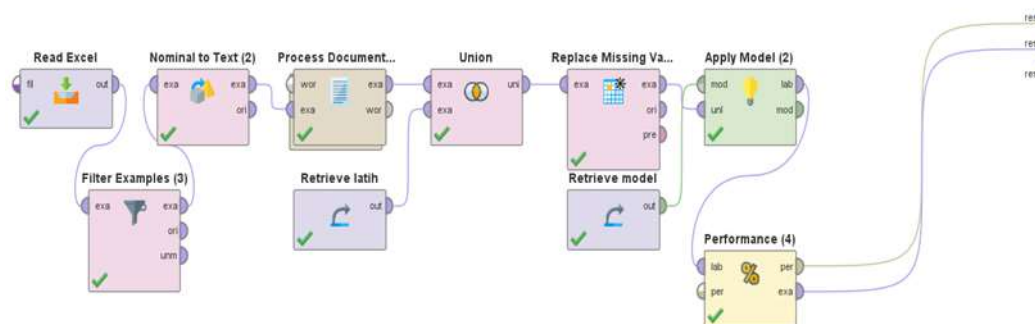
4. Hasil dan Pembahasan

Pada tahap awal penelitian yaitu mempersiapkan beberapa kebutuhan tahapan penelitian, dijelaskan sebagai berikut. Tahap pengumpulan data yaitu tahap awal penarikan data yang dilakukan untuk mempersiapkan sumber daya data yang berasal dari data tweet pada twitter, dengan menggunakan data tweet 4 hari setelah hari kejadian tragedi Kkanjuruhan, data diambil menggunakan tools netlytic, data yang berhasil diambil yaitu berjumlah 10.000 data, kemudian dilakukan pembersihan data yang duplikat sehingga menjadi 2.533 data; Tahapan labelling data latih, pada tahap ini dilakukan labeling data dengan tahapan awal labelling data secara manual, digunakan 500 data yang akan di-labelling sebagai data latih, data latih dibuat menggunakan rapidminer; Tahap processing data dan tahap pembelajaran mesin naive bayes, pada tahap ini digunakan operator read excel untuk memasukkan dataset, kemudian dataset tadi dilakukan filter examples yang berisi parameter filter "sentimen is not missing" atau isi sentimen yang tidak kosong, ini dilakukan untuk memfilter dataset agar data yang masuk hanya data yang tidak kosong pada value sentimen, selanjutnya menggunakan operator nominal to text yang berfungsi untuk mengubah data yang atributnya adalah nominal agar menjadi teks. Kemudian dilanjutkan menggunakan operator Process Documents from Data yang berfungsi sebagai operator yang melakukan processing pada dataset, Process Documents from Data mempunyai sub operator di dalamnya yaitu yang pertama tokenize dengan mode non letters yang berfungsi untuk pemisahan kata-kata dalam suatu kalimat, kemudian operator tranform cases dengan transform to lower case digunakan untuk mengecilkan semua huruf menjadi huruf kecil, kemudian sub operator Filter Stopwords (Dictionary) yang berisikan file dataset dengan encoding system, kemudian operator Filter Tokens (by Length) dengan parameter filter min chars 4, dan max chars 25, dimana jika data yang memiliki kurang dari 4 dan lebih dari 25 akan dihapus. Kemudian tahap selanjutnya menggunakan operator Naive Bayes yang digunakan untuk klasifikasi dataset berdasarkan data latih atau data labelling (500 data yang telah dilakukan labelling) untuk mendapatkan sebuah data model yang sudah dipelajari oleh operator naive bayes dan digunakan sebagai dictionary pada tahap labeling otomatis untuk mengetahui sentimen yang positif maupun negatif pada data yang belum ada nilai sentimennya. Kemudian menggunakan operator store untuk menyimpan data model pada file processes di rapidminer dan satu operator Store lagi sebagai tempat penyimpanan data latih dan disimpan dalam file data pada rapidminer. Setelah semuanya tersimpan maka tahap pembelajaran dari naive bayes selesai dan dilanjutkan ke proses labelling data otomatis. Untuk proses ini dapat dilihat pada Gambar 2.



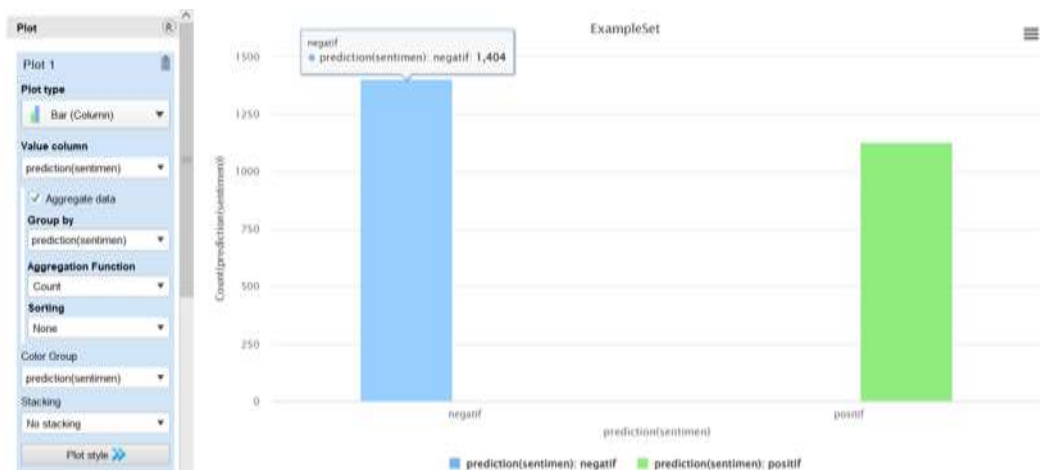
Gambar 2 Proses Tahapan Labeling Data (Source: Rapidminer)

Modelling data uji merupakan tahap dimana dilakukan labelling otomatis data yang belum memiliki label sentimen positif maupun negatif pada dataset, dilakukan dengan langkah awal menggunakan operator read excel untuk memasukkan dataset, kemudian dataset tadi dilakukan filter examples yang berisi parameter filter "sentimen is missing" atau isi sentimen yang kosong, ini dilakukan untuk memfilter dataset agar data yang masuk hanya data yang kosong pada value sentimen, kemudian memfilter pada bagian description dengan parameter is not missing, kemudian tahap selanjutnya menggunakan operator Nominal to Text, yaitu operator yang berfungsi untuk mengubah data berupa atribut nominal menjadi data teks, selanjutnya menggunakan operator Process Documents From Data yang berfungsi sebagai operator yang melakukan processing pada dataset dengan sub operator yang berjumlah 4 sub operator yaitu yang pertama tokenize dengan mode non letters yang berfungsi untuk pemisahan kata-kata dalam suatu kalimat, kemudian operator transform cases dengan transform to lower case digunakan untuk mengecilkan semua huruf menjadi huruf kecil, kemudian sub operator Filter Stopwords (Dictionary) yang berisikan file dataset dengan encoding system, kemudian operator Filter Tokens (by Length) dengan parameter filter min chars 4, dan max chars 25, dimana jika data yang memiliki kurang dari 4 dan lebih dari 25 akan dihapus. Selanjutnya menggunakan operator Union yang merupakan operator yang berfungsi untuk menggabungkan dua tabel dalam bentuk baris baru ke bawah, dengan cara menyambungkan data yang akan diisi sentimennya dan data latih yang sudah menjadi pembelajaran mesin dari Naive Bayes, setelah tergabung maka dilanjutkan menggunakan operator replace missing values yang digunakan untuk mengganti data missing value menjadi data yang di-set menjadi zero, kemudian operator yang digunakan selanjutnya yaitu operator apply model digunakan untuk menerapkan model yang telah dilatih dengan data training yang masih kosong sentimennya, dengan begitu mesin dapat menyimpulkan data value sentimen dan positif dan negatif, berdasarkan data model yang sudah dipelajari naive bayes, untuk prosesnya dapat dilihat pada Gambar 3.



Gambar 3. Modeling Data Training (Source: Rapidminer)

Hasil Sentimen Positif dan Negatif, pada tahap ini digunakan jenis visualisasi bar (Column) dengan value column = prediction (sentimen), kemudian aggregate data dicentang, dan untuk group by = prediction (sentimen), dan untuk aggregation function = count, kemudian untuk sorting = value column, direction = descending, kemudian untuk color group = prediction (sentimen) dan untuk stacking = no stacking, yang terakhir untuk plot style-nya default, dari tampilan visualisasi ini terdapat bar berwarna biru yang menandakan sentimen negatif dengan *prediction* (sentimen) : Negatif = 1.404 dan yang berwarna hijau adalah sentimen positif dengan *prediction* (sentimen) : Positif = 1.129, sehingga jumlah total datanya yaitu 2.533 data yang sudah masuk dalam analisis sentimen, untuk hasil prosesnya dapat dilihat pada Gambar 4.



Gambar 4 Hasil Visualiasi Sentimen Positif dan Negatif

Tahap selanjutnya merupakan tahap dimana dilakukanya pencarian accuracy model dari model naive bayes yang sudah mempelajari kata-kata dari setiap data labeling, yang dimana itu merupakan kata-kata dari data *tweet*. hasil *accuracy* nya dapat di lihat pada gambar 5.



Gambar 5. Evaluasi Accuracy

Dari hasil pengujian yang tertera pada gambar 5, evaluasi akhir model dari metode ini (*naive bayes*) dapat menunjukan hasil yang baik dengan mendapat *accuracy* model sampai 98,20%, ketika ingin melakukan evaluasi *performance algoritma* dari machine learning maka diperlukan acuan Confusion Matrix yang merepresentasikan prediksi dan kondisi sebenarnya dari semua data yang dihasilkan oleh algoritma dari machine learning itu sendiri. dengan berdasarkan Confusion Matrix maka dapat di tentukan *accuracy* nya. Evaluasi *Accuracy* merupakan penentu untuk mengukur kinerja pengklasifikasi untuk mencari satu prediksi yang benar dari banyaknya suatu titik data, berdasarkan model dari metode naive bayes dapat dilihat dari 500 data latih, 307 prediksi negatif dengan class precision 100.00%, class recall 97.15% dan 184 prediksi positif dengan 95.34% class precision, class recall 100.00% dirujuk pada gambar 5, kemudian perhitungan untuk mencari nilai accuracy jika secara manual menjadi $(491 * 100 / 500) = 98.2 = 98,20\%$, dan jika dihitung otomatis pada rapidminer maka hasilnya juga sama menjadi 98.20%, dari hasil *accuracy* yang cukup baik ini di dapatkan hasil yang dapat menjawab permasalahan dari penelitian ini yaitu nilai sentimen positif yang berjumlah 1.129 dan sebanyak 1.404 nilai sentimen negatif. Konsep penelitian ini merupakan penelitian yang berdasarkan pada hasil sentimen yang dimana sejalan dengan penelitian sebelumnya Fira Fathonah, Asti Herliana [4] yang dimana telah melakukan jajak pendapat terhadap vaksin COVID-19 dengan menggunakan pendekatan Naive Bayes. Karena Naive Bayes memiliki peluang bagus untuk melakukan klasifikasi, oleh karena itu ada dari metode ini dapat di hasilkan sebuah sentimen positif serta negatif dari data *tweet* jejak pendapat masyarakat.

5. Simpulan

Simpulan Berdasarkan hasil penelitian dan pembahasan, pada skala data yang ditampilkan dengan visualisasi, hasil nilai evaluasi yang didapatkan yaitu sebesar prediction (sentimen) : Negatif = 1.404 sedangkan prediction (sentimen) : Positif = 1.129. Berdasarkan hasil tersebut dapat disimpulkan bahwa dari tweet ini lebih banyak pengguna twitter yang bersentimen negatif terhadap tragedi Kanjuruhan, dengan banyaknya sentimen negatif terkait tragedi ini, tragedi ini dapat dijadikan sebagai tolak ukur atau sebagai acuan untuk beberapa organisasi penanggung jawab kegiatan tersebut agar dapat mengetahui pendapat masyarakat terhadap kegiatan yang telah diselenggarakan, agar menjadi pembelajaran dan pengalaman ke depannya dalam menyelenggarakan kegiatan yang serupa. Saran pengembangan untuk penelitian selanjutnya yaitu dengan memperbanyak lagi data testing dan data latih untuk mendapatkan accuracy model yang lebih baik.

Daftar Referensi

- [1] Nuryanti Musatari, Wahdania Suardi, Ulfiah Syukri, "Analisis Sentimen Media Sosial : penerapan E- Katalog Dalam Pengadaan Barang Dan Jasa Di Indonesia.", Jurnal Universitas Muhammadiyah Makassar, 193-200, Vol.10, No.3, 03 oktober 2022, DOI: 10.0803/klj.f0u7.987
- [2] Nurtriana Hidayati, Joko Suntoro, Galet Guntoro Setiaji, "Perbandingan Algoritma Klasifikasi Untuk Prediksi Cacat Software Dengan Pendekatan Crisp-Dm.", Jurnal Sains dan Teknik Informatika, Vol.7, No.2, pp.117-121, 2021, DOI: 10.34128/jsi.v7i2.313
- [3] Herianto, "Penerapan Text-Mining Untuk Mengidentifikasi Pengguna Twitter Terhadap Fenomena Peran DPR RI.", Jurnal Universitas Darma Persada, Vol.8, No.2, pp.36-44, 15 september 2022, DOI: 10.98478/Ysj.S9vi2.762
- [4] Fira Fathonah, Asti Herliana, "Penerapan Text Mining Analisa Sentimen Mengenai Vaksin Covid – 19 Menggunakan Metode Naive Bayes.", Jurnal Sains Dan Informatika, Vol.7, No.2, pp.155-164, 4 december 2021. [Online]. Available: DOI:<https://doi.org/10.34128/jsi.v7i2.331>
- [5] Dedi Darwis, Nery Siskawati, Zaenal Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional.", Jurnal Teknokompak, Vol.15, No.1, pp.131-145, 4 december 2021. [Online]. Available: DOI: <https://doi.org/10.33365/jtk.v15i1.744>
- [6] M Pramadani Riyanis Putra, Kiki Rizky Nova Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial.", Jurnal Teknik Informatika Msirawas, Vol.5, No.1, pp.63-71, 01 juni 2021, DOI: 10.98865/mining.v4i1.9609
- [7] Hakam Febtadianrano Putro, Retno Tri Vlandari, Wawan Laksito Yuly Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan.", Teknologi Informasi dan Komunikasi Sinar Nusantara, Vol.8, No.2, pp.19-24, 2 oktober 2020, [Online]. Available: DOI: <https://doi.org/10.30646/tikomsin.v8i2.500>
- [8] Eka Sabna, "Penerapan Text Mining Untuk pengelompokan Penelitian Dosen.", Jurnal Ilmu Komputer, Vol.9, No.2, pp.16-164, 2 oktober 2020, [Online]. Available: <https://doi.org/10.33060/JIK/2020/Vol9.Iss2.183>
- [9] Windy Sri Utami Saragih, Nelly Astuti Hasibuan, Rivalri Kristianto Hondroo, "Penerapan Text Mining Dengan Menggunakan Metode TF-IDF Untuk Menentukan Genre Dari Komik.", Konferensi Nasional Teknologi Informasi dan Komputer, Vol.4, No.1, pp.191-199, 1 oktober 2020, DOI: 10.30865/komik.v4i1.2679
- [10] Rakhmat Kurniawan R, ilka Zufria, "Penerapan Text Mining Pada Sistem Penyeleksian Judulu Skripsi Menggunakan Algoritma Latent Dirichlet Allocation(LDA).", Indonesian Journal of Computer Science, Vol.11, No.3, pp.1035-1052, 31 desember 2022, [Online]. Available: DOI: <https://doi.org/10.33022/ijcs.v11i3.3120>
- [11] Aditya Wijaya, Chaerur Rozikin, Betha Nurina Sari, "Application of Text Mining for Classification of COVID-19 Vaccination Hoax News Using Support Vector Machine Algorithm.", Jurnal Ilmiah Wahana Pendidikan, Vol.8, No.16, pp.67-77, 25 November 2022, DOI: 10.1109/ICTC55196.2022.9952630.
- [12] Syiva Multi Fani, Rukun Santoso, Suparti Suparti, "Penerapan Text Mining untuk Melakukan Clustering Data Tweet Akun Bliibli Pada Media Sosial Twitter Menggunakan K- Means Clustering.", Jurnal Ilmiah Wahana Pendidikan, Vol.10, No.4, pp.583-593, 1 oktober 2021, [Online]. Available: <https://doi.org/10.14710/j.gauss.10.4.583-593>

-
- [13] Novita Anggraini, Edi Surya Negara Harahap, Tri Basuki Kurniawan, "Text Mining - Analisis Teks Terkait Isu Vaksinasi COVID-19 (Text Mining - Text Analysis Related to COVID-19 Vaccination Issues.", *Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi*, Vol.23, No.2, pp.141-153, 2 desember 2021, DOI: 10.1109/YITC55196.2022.9952690
- [14] Rion Aztin, Krisna Adiyarta Musodo, "Penerapan Text Mining Dengan Algoritma Naïve Bayes Untuk Mengklasifikasikan Sentimen Rakyat Terhadap Minyak Goreng Subsidi Pemerintah.", *Prosiding SENAFI (Buku 2 - Artificial Intelligence)*, Vol.1, No.1, pp.645-652, 30 oktober 2022, DOI: 10.34128/ysi.v74i.390
- [15] Nur Arifin, Ultach Enri, Nina Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification.", *Satuan Tulisan Riset dan Inovasi Teknologi*, Vol.6, No.6, pp.129-136, 2 Desember 2021, [Online]. Available: DOI: <http://dx.doi.org/10.30998/string.v6i2.10133>