

Analisis Sentimen Komentar Youtube Tentang Relawan Patwal Ambulance Menggunakan Algoritma Naïve Bayes dan Decision Tree

Abd Wahid, Galuh Saputri*

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹abdwahid8813@gmail.com, ^{2,*}dosen02693@unpam.ac.id

Email Penulis Korespondensi: dosen02693@unpam.ac.id

Submitted: 05/10/2022; Accepted: 23/12/2022; Published: 31/12/2022

Abstrak—Kehadiran patwal ambulance terkadang dianggap meresahkan oleh masyarakat, banyak pendapat masyarakat yang masih pro dan kontra tentang aksi yang dilakukan oleh relawan patwal ambulance sipil karena dianggap illegal dan terkadang juga bersifat arogan. Pada penelitian ini peneliti ingin mengetahui opini dan tanggapan masyarakat tentang aksi yang dilakukan oleh para relawan patwal ambulance sipil. Metode yang digunakan dalam penelitian ini yaitu dengan melakukan analisis sentimen dengan teknik data mining untuk menemukan polaritas kalimat dalam suatu dokumen dengan menggunakan algoritma Naïve Bayes dan Decision Tree. Tahapan awal yang dilakukan pada penelitian ini yaitu dengan cara mengumpulkan data komentar yang dilakukan dengan cara scraping komentar youtube menggunakan API Key Youtube V3, kemudian dilanjutkan dengan pelabelan data secara manual, cleaning data, dan preprocessing data dan pembobotan kata menggunakan TF-IDF. Dari Hasil keseluruhan pengujian dengan 600 data latih, algoritma Naïve Bayes memiliki nilai accuracy lebih tinggi yaitu sebesar 66,72% sedangkan nilai Recall nya sebesar 64,98%. Pengujian dengan algoritma Decision Tree pada penelitian ini memiliki Recall yang lebih tinggi dibandingkan dengan Naïve Bayes. %. Dari hasil dataset komentar youtube yang digunakan dalam penelitian ini, bisa disimpulkan algoritma Naïve Bayes memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma Decision Tree, sehingga bisa ditarik kesimpulan bahwa Naïve Bayes memiliki akurasi terbaik pada dataset komentar youtube yang digunakan dalam penelitian ini.

Kata Kunci: Analisis Sentimen; Naïve Bayes; Decision Tree; Rapid Miner; Data Mining

Abstract—The presence of the ambulance patrol is sometimes considered disturbing by the community, many people's opinions are still pro and contra about the actions carried out by civilian ambulance patrol volunteers because they are considered illegal and sometimes also arrogant. In this study, the researchers wanted to know the opinions and responses of the public about the actions taken by civilian ambulance patrol volunteers. The method used in this study is to perform sentiment analysis with data mining techniques to find the polarity of sentences in a document using the Naïve Bayes and Decision Tree algorithms. The initial steps in this research were collecting comment data by scraping YouTube comments using API Key Youtube V3, followed by manual data labeling, data cleaning, data preprocessing, and word weighting using TF-IDF. From the overall results of testing with 600 training data, the Naïve Bayes algorithm has a higher accuracy value of 66.72% while the Recall value is 64.98%. Testing with the Decision Tree algorithm in this study has a higher Recall compared to Naïve Bayes. %. From the results of the YouTube comment dataset used in this study, it can be concluded that the Naïve Bayes algorithm has a higher accuracy value than the Decision Tree algorithm, so it can be concluded that Naïve Bayes has the best accuracy in the YouTube comment dataset used in this study.

Keywords: Analisis Sentimen; Naïve Bayes; Decision Tree; Rapid Miner; Data Mining

1. PENDAHULUAN

Ambulans merupakan kendaraan yang diberi kelengkapan medis untuk membawa korban kecelakaan, memindahkan pasien ke hospital lain, atau membawa orang sakit. Ambulans dilengkapi lampu rotator darurat dan sirene supaya bisa melewati kemacetan lalu lintas.

Selain pemadam kebakaran, Ambulans ini memiliki hak istimewa seperti melawan arah, menerobos lampu merah, dan melalui lajur bahu jalan. Undang-undang perlalulintasan menjelaskan bahwa kendaraan seperti ambulans serta kendaraan gawat darurat yang lainnya mesti diberi jalan di jalan raya buat menyelamatkan nyawa.

Untuk membelah kemacetan lalu lintas ambulance terkadang membutuhkan patwal (patrol dan pengawalan), ada banyak komunitas pengawal ambulans di Indonesia salah satunya komunitas Indonesia Escorting Ambulance (IEA) komunitas ini sudah memiliki banyak anggota di berbagai daerah. seringkali terlihat pengendara motor yang mengawal ambulans untuk membelah kemacetan dikota-kota besar, namun aksi relawan ini justru dianggap meresahkan. Seperti halnya yang disampaikan oleh Kepala Satuan Lalu Lintas (Kasatlantas) Porles Metro Depok AKBP Andi Indra Waspada, seperti yang dikutip dari (kompas.com 07/05/2021) mengatakan, ada laporan masyarakat yang merasa resah dengan pengawal sipil. Pengawal sipil ini dirasa terlalu arogan bahkan melebihi Patwal dari kepolisian.

Berdasarkan dari berita dari channel youtube Redaksi Trans 7 Official tentang komunitas pengawal ambulans yang pro kontra di kalangan masyarakat, masyarakat yang pro terhadap patwal ambulance beralasan bahwa kehadiran patwal sangat membantu karena terkadang banyak pengendara yang belum menyadari tentang prioritas mobil ambulans di jalan raya, sedangkan pendapat yang kontra mereka beralasan bahwa kehadiran patwal ini adalah illegal dan terkadang meresahkan.

Maka dari itu perlu diketahui langsung opini masyarakat terhadap aksi relawan saat melakukan pengawalan pada ambulance, masyarakat sekarang cenderung mengemukakan pendapatnya langsung melalui media sosial, salah satu platform yang cukup banyak digunakan yaitu youtube. Banyak video-video aksi patwal ambulans yang viral di platform youtube, banyak komentar-komentar masyarakat di youtube yang menilai tentang aksi patwal sipil ambulans, dalam platform youtube masyarakat dapat memberikan opini langsung atau penilaian terhadap berita atau konten yang mereka tonton. Sehingga kita dapat mengambil kesimpulan dari opini-opini tersebut apakah masyarakat setuju atau tidak dengan aksi patwal tersebut.

Berdasarkan studi literatur yang peneliti amati mengenai Sentimen analisis yang dilakukan oleh Syarifuddin tentang: “Analisis Sentimen Opini Publik Terhadap Efek PSBB Pada Twitter Menggunakan Metode Decision Tree-KNN-Naive Bayes” [1]. Dalam penelitian ini peneliti menggunakan dataset dari twitter dengan menggunakan tools rapid miner dengan hasil penelitian nilai akurasi tertinggi terdapat pada metode naive bayes.

Penelitian lain yang dilakukan oleh Harun & ananda tentang : “Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naive Bayes dan Decision Tree”[2]. Dalam penelitian menggunakan dataset dari komentar fanspage facebook Kementerian Kesehatan, metode yang digunakan yaitu membandingkan algoritma Naive Bayes dan Decision Tree akurasi tertinggi terdapat pada metode Naive Bayes.

Berdasarkan Peneliti lainnya yang dilakukan oleh Puspita & Widodo tentang “ Perbandingan Metode KNN, Decision Tree, naive Bayes pada analisis sentimen penggunaan Layanan BPJS”[3]. Dari penelitian ini data yang digunakan berupa tweet dari komentar di twitter tentang layanan BPJS, Hasil yang didapat dalam penelitian ini metode Naive Bayes memiliki hasil presentase tertinggi yaitu denan nilai akurasi sebesar 96.13%

Perlu diketahui Opinion mining atau sentiment analysis adalah suatu bidang studi untuk menganalisis pendapat, sentimen, evaluasi penilaian, sikap bahkan emosi seseorang terhadap suatu entitas, peristiwa, topik, serta atribut lainnya[4]. Analisis sentiment dapat digunakan untuk mengekstraksi opini-opini atau pendapat dari sebuah dokumen, media sosial, review blog, serta data-data lainnya.

Metode atau teknik dalam data mining sangat banyak, oleh karena itu dalam pemilihan teknik atau algoritma yang tepat dalam melakukan analisis akan bergantung pada tujuan yang diinginkan.

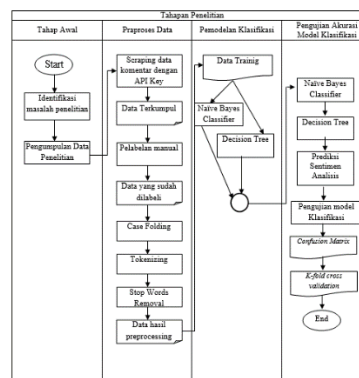
Dengan begitu peneliti tertarik untuk menganalisis sentimen publik pada dataset komentar youtube dengan menggunakan metode naive bayes dan decision tree untuk membandingkan tingkat akurasi yang dihasilkan dari dataset komentar youtube dengan Metode Naive Bayes dan Decision Tree.

Adapun masalah dalam penelitian ini adalah bagaimana peneliti mengumpulkan data dari komentar youtube dan berapa tingkat akurasi dari metode Naive Bayes dan Decision Tree dalam analisis sentimen pada dataset komentar youtube.

2. METODOLOGI PENELITIAN.

2.1 Tahapan Penelitian

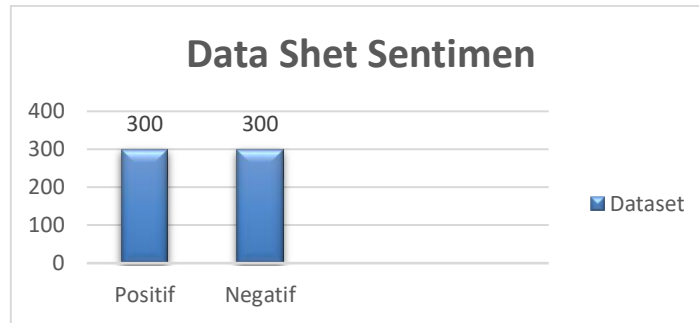
Tahapan penelitian dibagi menjadi 4 step, tahap awal, praproses data, pemodelan algoritma klasifikasi, dan tahapan terakhir yaitu melakukan pengujian akurasi pada metode klasifikasi, data yang digunakan dalam penelitian ini menggunakan data latih sebanyak 600 data, diantaranya 300 sentimen negatif dan 300 data sentimen positif, untuk lebih jelasnya peneliti sajikan dalam bentuk diagram alur yang terdapat pada gambar berikut:



Gambar 1. Rancangan Metode Penelitian

a. Pengumpulan Data

Pada tahapan awal pada penelitian ini yaitu dengan cara melakukan pengumpulan data komentar yang dilakukan dengan menggunakan teknik scraping. Scrapin merupakan teknik untuk mengambil data online secara otomatis, atau dengan kata lain scraping merupakan teknik yang digunakan untuk mengekstrak informasi yang terdapat dalam dari sebuah werbsite secara otomatis[5]. Dari proses scraping dilakukan proses pelabelan sentimen secara manual dan kemudian pembuatan data latih, dalam peneletian ini peneliti menggunakan 600 data komentar yang dijadikan sebagai data latih, dimana 300 data dengan sentimen positif dan 300 data dengan sentimen negatif.



Gambar 2. Dataset

b. Preprocessing

Tahapn selanjtunya preprocessing dilakukan untuk mengolah data mentah menjadi data collection yang siap dipakai sehingga menghasilkan data yang telah terstruktur dengan jelas, adapun tahapan-tahapan yang dilakukan dalam preprocessing data berikut beberapa tahapan yang dilakukan dalam penelitian [6]:

1. Cleaning, merupakan tahapan yang bertujuan untuk membersihkan data yang ganda dan karakter-karakter yang tidak diperlukan yang terdapat pada komentar[7].

Tabel 1. Proses Case Folding

| Input | Output |
|---|---|
| Logika aja kalo para pengemudi kendaraan itu PUNYA OTAK ya gk bakalan ada PENGAWAL AMBULANCE simple kan | logika aja kalo para pengemudi kendaraan itu punya otak ya gk bakalan ada pengawal ambulance simple kan |

2. Pada tabel 1. Proses Case Folding diperlukan untuk mengubah semua text yang terdapat dalam komentar menjadi lowercase.

Tabel 2. Proses Tokenizing

| Input | Output |
|---|---|
| Logika aja kalo para pengemudi kendaraan itu PUNYA OTAK ya gk bakalan ada PENGAWAL AMBULANCE simple kan | ["logika", "aja", "kalo", "para", "pengemudi", "kendaraan", "itu", "punya", "otak", "ya", "gk", "bakalan", "ada", "pengawal", "ambulance", "simple", "kan"] |

3. Pada tabel 2. Proses Tokenizing tahapan ini bertujuan untuk memecah teks yang berupa kalimat menjadi sebuah token atau membaginya menjadi kata perkata.

Tabel 3. Proses Stopword Removal

| Input | Output |
|---|---|
| Logika aja kalo para pengemudi kendaraan itu PUNYA OTAK ya gk bakalan ada PENGAWAL AMBULANCE simple kan | logika pengemudi kendaraan otak pengawal ambulance simple |

4. Pada Tabel 3. Merupakan contoh Stopword Removal, hal ini diperlukan dalam proses pemilihan kata-kata penting dari sebuah dokumen, pada tahapan ini dilakukan penghapusan kata-kata yang dianggap stopwords[8].

Tabel 4. Proses Stemming

| Input | Output |
|---|---|
| Logika aja kalo para pengemudi kendaraan itu PUNYA OTAK ya gk bakalan ada PENGAWAL AMBULANCE simple kan | logika kemudi kendara punya otak kawal ambulance simple |

Tabel 4. Proses Stemming, bertujuan untuk menemukan base word dengan menghilangkan semua imbuhan yang menyatu pada tiap-tiap kata. Selanjutnya dilakukan penghitungan frekuensi kata dengan tujuan untuk menghitung berapa banyak kata yang sama dalam setiap sentimen yang menjadi ciri masing-masing dari kata sentimen. Tahapan Probabilitas kata dilakukan untuk mendapatkan term dengan nilai yang lebih penting dan dianggap relevan untuk dijadikan kata kunci, adapun persamaannya sebagai berikut:

$$P(W_k|V_j) = \frac{n_{k+1}}{n+|kosakata|} \tag{1}$$

Diketahui : n Positif : 300
n Negatif : 300

Kata Positif

$$P(\text{membantu}|\text{Positif}) = \frac{201+1}{300+|600|} = \frac{202}{600} = 0,33$$

$$P(\text{membantu}|\text{Negatif}) = \frac{235+1}{300+|600|} = \frac{236}{600} = 0,39$$

Kata Negatif

$$P(\text{arogan}|\text{Positif}) = \frac{231+1}{300+|600|} = \frac{234}{600} = 0,38$$

$$P(\text{arogan}|\text{Negatif}) = \frac{234+1}{300+|600|} = \frac{235}{600} = 0,39$$

Tabel 5. Contoh Kata Dasar Sentimen

| No | Kata | Positif | Negatif | Probab Positif | Probab Negatif |
|----|---------------|---------|---------|----------------|----------------|
| 1 | Arogan | 231 | 234 | 0,38 | 0,39 |
| 2 | Kemanusiaan | 242 | 247 | 0,40 | 0,41 |
| 3 | Legal | 227 | 240 | 0,38 | 0,40 |
| 4 | Ilegal | 242 | 238 | 0,40 | 0,39 |
| 5 | Melanggar | 248 | 250 | 0,41 | 0,41 |
| 6 | Apresiasi | 249 | 250 | 0,41 | 0,41 |
| 7 | Bahaya | 250 | 0 | 0,41 | 0,001 |
| 8 | Strobo | 0 | 250 | 0,001 | 0,41 |
| 9 | Menolong | 245 | 249 | 0,41 | 0,41 |
| 10 | Kesadaran | 241 | 249 | 0,40 | 0,41 |
| 11 | Miris | 0 | 250 | 0,001 | 0,41 |
| 12 | Menyelamatkan | 248 | 249 | 0,41 | 0,41 |
| 13 | Vlog | 0 | 250 | 0,001 | 0,41 |
| 14 | Hukum | 240 | 246 | 0,40 | 0,41 |
| 15 | Gerombolan | 0 | 250 | 0,001 | 0,41 |
| 16 | Ganggu | 0 | 250 | 0,001 | 0,41 |
| 17 | Tilang | 0 | 249 | 0,001 | 0,41 |

Pada tabel 5 merupakan contoh kata dasar sentimen yang dihasilkan dari setiap data komentar, dan dijadikan kata kunci untuk melakukan analisis sentimen pada dataset komentar.

2.2 Naïve Bayes

Pengklasifikasian statistic yang bisa dipakai buat memprediksi probabilitas keanggotaan suatu class adalah Naïve Bayes[9]. Konsep dasar yang dipakai pada Naïve Bayes merupakan Teorema Bayes dengan asumsi ketidakbergantungan yang kuat. Dalam ilmu statistic, teorema dipakai untuk menghitung peluang satu class dari masing-masing kelompok attribute yang ada dan memastikan class mana yang paling optimal[10]. Metode ini merupakan salah satu metode data mining dalam metode machine learning yang memiliki akurasi yang tinggi serta merapkan algoritma yang sederhana. Adapun tahapan metode Naïve Bayes sebagai berikut[11]:

a. Menghitung nilai probabilitas bersyarat

$$P(B_i|A)=P(B_1, B_2, \dots, B_n|A) \tag{2}$$

Keterangan :

A : Kelas

B : Vektor dari nilai atribut n

P(B_i|A) : Nilai atribut A_i yang ada pada kelas B

b. Menghitung nilai probabilitas prior tiap kelas

$$P(A) = \frac{X_j}{X} \tag{3}$$

Keterangan :

N_j : Jumlah dokumen pada satu kelas

N : Jumlah total dokumen

c. Menghitung nilai probabilitas posterior

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \tag{4}$$

2.3 Decision Tree

Pohon keputusan atau decision tree adalah suatu metode klasifikasi yang lumayan terkenal, metode Decision Tree ini mengganti fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan[12]. Pohon keputusan merupakan struktur flowchart yang memiliki tree, setiap cabang merepresentasikan hasil tes yang dimana masing-

masing simpul internal menerangkan suatu tes attribute, dan simpul daun merepresentasikan class atau class distribution[13]. Alur dari decision tree bisa kita telusuri dari simpul ke akar ke simpul daun yang memangu prediksi. Konsep dari algoritma Decision Tree adalah mengganti data menjadi aturan-aturan keputusan, dan konsep pohon keputusan dapat digambarkan sebagai berikut [14]. Metode Decision Tree memiliki beberapa tahapan yaitu:

1. Data Training disiapkan
2. Pemilihan attribut yang akan dijadikan akar dengan menggunakan rumus seperti berikut[11]

$$\text{Entropy}(S) = \sum_{i=1}^n p_i * \log_2 p_i \tag{5}$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \tag{6}$$

Keterangan :

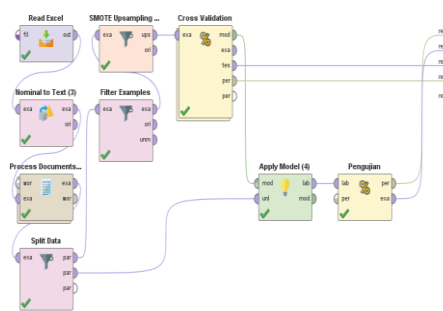
- S : Himpunan Kasus
- n : Jumlah Partisi
- pi : Porporasi dari Si terhadap S
- |Si| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah Kasus dalam S

3. HASIL DAN PEMBAHASAN

Pada penelitian ini data komentar youtube yang telah dikumpulkan kemudian di jadikan sebagai dataset dalam bentuk file csv kemudian diolah dengan menggunakan software rapidminer, untuk melakukan pengujian dengan menggunakan metode yang digunakan, kemudian data akan dilakukan pengujian dengan menggunakan operator Cross Validation dengan membagi data menjadi dua yaitu data training dan data uji, operator Cross Validation mempunyai dua subproses yaitu subproses training yang digunakan untuk melatih model dan subproses testing untuk metode yang digunakan serta untuk mengukur kinerja model klasifikasi yang dibuat.

Tujuan utama dalam penelitian ini yaitu dapat mengetahui dari tingkat akurasi metode Decision Tree dan Naïve Bayes dalam mengklasifikasikan sentimen negative dan positif pada komentar youtube. Pada pengujian ini peneliti menggunakan RapidMiner Versi 9.10 yang digunakan untuk melakukan klasifikasi sentimen data training dan data testing pada dataset komentar youtube.

Setelah semua tahapan cleaning, pre processing data telah selesai dilakukan, dilanjut dengan mengolah data menggunakan aplikasi Rapidminer berikut adalah proses dari pembuatan model dari analisis sentimen pada penelitian ini.



Gambar 3. Proses Analisis Sentimen

Pada Gambar 3 Operator proses read excel digunakan untuk membaca file data komentar yang telah melewati tahapan cleaning data dan labeling manual. Pada Rapid miner versi 9.10 penggunaan operator Nomintal To Text dibutuhkan untuk mengubah data nominal menjadi text, kemudian dilanjutkan dengan preprocessing document dimana di dalamnya terdapat proses (tokenizing, transform case, filter token by length, filter stopwords).



Gambar 4. Pembagian Data Menggunakan Operator Split Data

Pada gambar 4. penggunaan operator Split Data untuk membagi dataset menjadi dua bagian yaitu data uji dan data latih, dalam penggunaan operator split data ini peneliti menggunakan rasio dataset pengujian dengan 0.7 dan 0.3, dimana 0.7 merupakan 70% dari data yang akan digunakan sebagai data latih, 0.3 adalah 30% dari dataset yang akan dijadikan sebagai data testing.

Dalam pengujian menggunakan Cross Validation dilakukan 10 kali eksperimen pengujian klasifikasi dengan bantuan operator cross validation guna mengetahui hasil analisis serta hasil klasifikasi dari model yang telah dibuat[15], kemudian menggunakan operator Apply Model untuk menerapkan model yang sudah dilatih sebelumnya dengan data training, pada unlabeled data (data testing). Tujuan dari apply model yaitu mendapatkan prediksi untuk unlabeled data (data testing) yang belum memiliki label sentimen.

3.1 Hasil Klasifikasi Naïve Bayes

accuracy: 66.67%

| | true negatif | true positif | class precision |
|---------------|--------------|--------------|-----------------|
| pred. negatif | 82 | 39 | 67.77% |
| pred. positif | 40 | 76 | 65.52% |
| class recall | 67.21% | 66.09% | |

Gambar 5. Performace Vector Naive Bayes (Data Uji)

Hasil pengujian pada model klasifikasi yang digunakan, performa polaritas data pada klasifikasi sentiment dari data training menghasilkan akurasi sebesar 70,62% dengan nilai sentimen negatif memiliki prediksi yang tinggi.

Pada Gambar 5 merupakan hasil dari keseluruhan pengujian data uji yang digunakan dengan model naïve bayes di dapat hasil klasifikasi dengan akurasi sebesar 66.67% dengan hasil prediksi negatif memiliki nilai tertinggi.

3.2 Hasil Klasifikasi Decision Tree

accuracy: 59.69% +/- 5.13% (micro average: 59.68%)

| | true negatif | true positif | class precision |
|---------------|--------------|--------------|-----------------|
| pred. negatif | 96 | 41 | 70.07% |
| pred. positif | 188 | 243 | 56.38% |
| class recall | 33.80% | 85.56% | |

Gambar 4. Performace Vector Decision Tree (Data Uji)

Pada Gambar 6 merupakan hasil Percobaan menggunakan metode Decision tree dengan pembagian data menggunakan operator Split Data dengan rasio pembagian data 70% data latih dan 30% dari 1200 data latih yang digunakan pada algoritma Decision Tree akurasi yang dihasilkan dari data latih sebesar 59.69%. dan untuk keseluruhan hasil data uji berdasarkan data latih yang telah di buat menggunakan model Decision Tree mendapatkan akurasi 64.98% terjadi peningkatan akurasi pada data uji yang digunakan dalam penelitian ini.

3.3 Evaluasi Hasil Analisis Sentimen

Setelah proses analisis sentimen tahapan terakhir yaitu memasuki tahapan pengujian, butuh 1 tahapan lagi proses yang telah dilakukan dalam menentukan kualitas, yaitu evaluasi hasil. 3 Parameter yang dilakukan evaluasi hasil yaitu, precision; accuracy; dan recall; dari suatu perhitungannya yang telah dilakukan[1]. Akurasi (A) merupakan klasifikasi jumlah dokumen dengan benar (Benar Positif atau Benar Negatif)[16]. Untuk menghitung nilai akurasi bisa memakai persamaan:

$$\text{Accuracy (A)} = \frac{TP+TN}{(TP+TN+FP+FN)} \times 100\% \tag{7}$$

Presisi (P) merupakan informasi yang ingin dicari untuk banyaknya hasil pemrosesan yang relevan. Presisi merupakan klasifikasi True Positive dan semua data di prediksi sebagai positive class. Untuk menghitung nilai presisi bisa memakai persamaan:

$$\text{Precision} = \frac{TP}{(TP+FP)} \times 100\% \tag{8}$$

Recall (R) merupakan koleksi dihasilkan oleh sistem dari banyaknya dokumen yang relevan. Jadi, recall merupakan jumlah dokumen yang memiliki klasifikasi benar-benar positive (termasuk False Negative). Untuk menghitung nilai recall bisa memakai persamaan:

$$\text{Recall} = \frac{TP}{(TP+FN)} \times 100\% \tag{9}$$

Variabel TN (True Negative), TP (True Positive), FN (False Negative) dan FP (False Positive) berasal dari Confusion Matrix yaitu data negatif yang diklasifikasikan menjadi positif, berikut penjelasan lebih detail tentang confusion matrix pada tabel 5 [17]

Tabel 5. Confusion Matrix

| | Prediction | |
|----------|------------|----|
| | Yes | No |
| True Yes | TP | FN |
| True No | FP | TN |

Confusion Matrix adalah pengukuran performa untuk masalah klasifikasi pada machine learning dimana keluaran dapat berupa dua kelas atau lebih. Confusion matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual[18]. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu, True Positif, True Negatif, False Positif dan False Negatif seperti yang terdapat pada Tabel 5. Dari hasil pengujian dengan algoritman Naïve Bayes dan Decision Tree di dapatkan hasil prediksi dengan menggunakan RapidMiner, berikut hasil confusion matrix dari masing-masing algoritman yang digunakan pada analisis sentimen dengan RapidMiner.

Tabel 6. Nilai Accuracy, Precision, dan Recall Hasil Pengujian

| Metode | Accuracy | Precision | Recall |
|---------------|----------|-----------|--------|
| Naïve Bayes | 66,67% | 66,09% | 65,52% |
| Decision Tree | 64,98% | 58,79% | 93,04% |

Pada tabel 8 Accuracy dari Naïve Bayes sebesar 66,67% dengan Precision 66,09% dan Recall 65,52%. Sedangkan hasil pengujian dengan Decision Tree menunjukkan hasil Accuracy sebesar 64,98%, Precision 58,79% dengan Recall 93,04%. Dari hasil dataset komentar youtube yang digunakan dalam penelitian ini, bisa disimpulkan nilai akurasi lebih tinggi adalah Naïve Bayes dibandingkan dengan Decision Tree, sehingga bisa ditarik kesimpulan bahwa Naïve Bayes memiliki akurasi terbaik pada dataset media sosial youtube yang digunakan, Seperti halnya penelitian yang telah dilakukan oleh (Harun, Ahmad, 2021)[2]. Menggunakan algoritma yang sama yaitu Naïve Bayes dan decision Tree dengan dataset media sosial twitter. Akurasi yang lebih baik adalah algoritma Naïve Bayes dibandingkan Decision Tree, hasil akhir akurasi Naïve Bayes sebesar 100,00% dan nilai akurasi dari Decision Tree sebesar 50,39%.

4. KESIMPULAN

Tujuan dari penelitian ini yaitu untuk mengetahui tingkat akurasi dari algoritma Naive bayes dan Decision Tree dalam melakukan analisis sentimen pada dataset berupa komentar. Berdasarkan penelitian yang telah dilakukan maka dapat diambil kesimpulan, Teknik untuk mengambil data komentar pada channel youtube bisa dilakukan dengan cara scraping data menggunakan Youtube API Key V3. Hasil pengujian Accuracy dari Naïve Bayes sebesar 66,67% dengan Precision 66,07% dan Recall 65,52%. Sedangkan hasil pengujian dengan Decision Tree menunjukkan hasil Accuracy sebesar 64,98%, Precision 58,79% dengan Recall 93,04%. Dari hasil dataset komentar youtube yang digunakan dalam penelitian ini, bisa disimpulkan algoritma Naïve Bayes memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma Decision Tree, sehingga bisa ditarik kesimpulan bahwa Naïve Bayes memiliki akurasi terbaik pada dataset komentar youtube yang digunakan dalam penelitian ini.

REFERENCES

- [1] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes," INTI Nusa Mandiri, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [2] A. Harun and D. P. Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 1, no. 1, pp. 58–63, 2021.
- [3] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," J. Inform. Univ. Pamulang, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [4] M. F. A. BASHRI and R. KUSUMANINGRUM, "Analisis Sentimen Menggunakan Latent Dirichlet Allocation dan Visualisasi Topic Polarity Wordcloud." Universitas Diponegoro, 2017.
- [5] O. W. Purbo, Text Mining. Bandung: Andi, 2019.
- [6] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," J. Teknol. Inf. dan Ilmu Komput., vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [7] R. P. Sidiq, B. A. Dermawan, and Y. Umidah, "Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," J. Inform. Univ. Pamulang, vol. 5, no. 3, p. 356, 2020, doi: 10.32493/informatika.v5i3.6571.
- [8] E. M. Zamzami, "Aplikasi Edutainment Pendukung Pembelajaran Jarak Jauh TK Merujuk Standar Nasional PAUD," J. Obs. J. Pendidik. Anak Usia Dini, vol. 5, no. 2, pp. 985–995, 2020, doi: 10.31004/obsesi.v5i2.750.
- [9] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," J. Ilm. Ilmu Komput., vol. 2, no. 1, pp. 22–26, 2016.

- [10] A. Wanhari, “Perbandingan Algoritma C4. 5 dan Naive Bayes Untuk Klasifikasi Mustahik.” Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- [11] F. Salmon Pattiha, “Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia,” *J. Ris. Komputer*, vol. 9, no. 2, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i2.4016.
- [12] I. D. Khazari, Sofalul; Marisa, Fitri; Wijaya, “Sistem Rekomendasi Penentuan Judul Skripsi Menggunakan Algoritma Decision Tree,” *J. Teknol. dan Manaj. Inform.*, pp. 123–127, 2017.
- [13] M. R. Fajriansyah and Siswanto, “Analisis Sentimen Pengguna Twitter Terhadap Partai Politik Pendukung Calon Gubernur Di Jakarta Menggunakan Algoritma C4 . 5 Decision Tree Learning,” *Skanika*, vol. 1, no. 2, pp. 697–703, 2018, [Online]. Available: <https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/278>
- [14] A. G. Salman, “Konsep Decision Tree & Random Forest,” 2020. <https://socs.binus.ac.id/2020/05/26/konsep-decision-tree-random-forest/>
- [15] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [16] K. M. Asshiddiqi, M.F, Lhaksamana, “Perbandingan Metode Decision Tree dan Support Vector Machine Untuk Analisis Sentimen Pada Instagram Mengenai Kinerja PSSI,” vol. 1, no. D, pp. 2019–2020, 2020.
- [17] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, no. 2, p. 140, 2020, doi: 10.24002/jbi.v11i2.3769.
- [18] M. A. Muslim et al., *Data Mining Algoritma C4.5*. 2019.