

ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK DAN GRAB PADA MEDIA SOSIAL TWITTER MENGGUNAKAN RANDOM FOREST

Melia ¹, Bambang Irawan ², Odi Nurdiawan ³

^{1,2} Teknik Informatika, STMIK IKMI Cirebon

³ Manajemen Informatika STMIK IKMI Cirebon

Jalan Perjuangan 10B Majasem Kota Cirebon, Indonesia

meliamega9a@gmail.com

ABSTRAK

Teknologi informasi dan komunikasi telah berkembang dan mengubah pola hidup masyarakat secara global. Perkembangan teknologi informasi telah menyebabkan dunia menjadi tanpa batas dan menyebabkan perubahan sosial, budaya, dan pola penegakan hukum yang secara signifikan berlangsung cepat. Alat transportasi juga telah berkembang dengan adanya transportasi online. Transportasi merupakan sarana yang sangat penting dalam mendukung aktifitas atau mobilitas manusia setiap harinya. Banyaknya pengguna jasa transportasi online yang sudah menjadi kebutuhan, maka perlu melakukan analisis sentimen terhadap pengguna aplikasi transportasi online. Jasa transportasi online tersebut adalah Gojek dan Grab. Media yang digunakan untuk mengambil data merupakan salah satu platform media sosial yaitu Twitter. Masalah ini, mengklasifikasikan pengguna Twitter kedalam respon positif dan negatif melalui aplikasi transportasi online tersebut. Proses pengambilan data dilakukan dengan cara Crawling data Twitter menggunakan Rapid Miner, kemudian dilakukan proses text mining dengan menggunakan algoritma random forest. Berdasarkan studi pustaka, algoritma random forest dapat menghasilkan ketepatan yang tinggi. Menurut hasil analisis, terdapat 132 komentar yang mendapat penilaian positif, sementara 68 komentar dianggap negatif. Berdasarkan eksperimen yang telah dilakukan, algoritma random forest mencapai tingkat akurasi sekitar 76.25%. Algoritma random forest dianggap efektif dalam melakukan klasifikasi sentimen dengan cepat.

Kata kunci : Analisis sentimen, Data Mining, Random Forest, Twitter, Transportasi online

1. PENDAHULUAN

Dalam kehidupan yang bermasyarakat, setiap harinya kita melakukan aktivitas mengungkapkan sentimen atau pendapat terhadap orang lain. Analisis Sentimen merupakan salah satu teknik untuk menilai suatu komentar atau suatu tulisan kalimat bernada positif dan negatif. Berkembangnya teknologi yang semakin pesat, masyarakat dapat mengungkapkan sentimen atau pendapat atau pandangan melalui media sosial seperti Twitter, Facebook, dan Instagram. Media sosial sendiri merupakan suatu media online untuk komunikasi dan interaksi jarak jauh. Salah satu media sosial terbesar dengan pengguna terbanyak di dunia, yaitu Twitter. Twitter merupakan salah satu media yang banyak digunakan oleh masyarakat Indonesia untuk memberikan sentimen atau pendapat antar sesama pengguna media sosial. Menurut laporan dari website We Are Social, pada tahun 2022 jumlah pengguna Twitter di Indonesia mencapai 18,45 juta. Dimana Indonesia masuk kedalam negara pengguna Twitter peringkat kelima terbesar di dunia. [1]

Dalam kehidupan yang bermasyarakat, setiap harinya kita melakukan aktivitas mengungkapkan sentimen atau pendapat terhadap orang lain. Analisis Sentimen merupakan salah satu teknik untuk menilai suatu komentar atau suatu tulisan kalimat bernada positif dan negatif. Berkembangnya teknologi yang semakin pesat, masyarakat dapat mengungkapkan sentimen atau pendapat atau pandangan melalui media sosial seperti Twitter, Facebook, dan Instagram. Media sosial sendiri merupakan suatu media online untuk

komunikasi dan interaksi jarak jauh. Salah satu media sosial terbesar dengan pengguna terbanyak di dunia, yaitu Twitter. Twitter merupakan salah satu media yang banyak digunakan oleh masyarakat Indonesia untuk memberikan sentimen atau pendapat antar sesama pengguna media sosial. Menurut laporan dari website We Are Social, pada tahun 2022 jumlah pengguna Twitter di Indonesia mencapai 18,45 juta. Dimana Indonesia masuk kedalam negara pengguna Twitter peringkat kelima terbesar di dunia [1]

Pada penelitian [2] menunjukkan hasil bahwa metode *Support Vector Machine* memiliki nilai akurasi yang lebih tinggi (90,20%) dibandingkan dengan *Decision Tree* (89,80%). Penelitian ini menggunakan teknik *Random Forest Classifier*, dimana penggunaan metode ini bertujuan untuk menyelesaikan masalah yang berhubungan dengan klasifikasi dan regresi. Terdapat banyak penelitian yang menyebutkan bahwa *Random Forest* memiliki prediksi yang baik dalam klasifikasi dan regresi di berbagai bidang. Penelitian ini bertujuan untuk mengetahui kesesuaian yang lebih baik mengenai opini/sentimen metode *Random Forest*

2. TINJAUAN PUSTAKA

Pada penelitian [3] menjelaskan *Algoritma Multinomial naive bayes* yang mampu mengetahui sentimen pengguna transportasi online melalui tweet pada jasa layanan ojek online Grab Indonesia ke dalam kelas positif, netral, serta negatif. Hasil klasifikasi yang diambil dari tweet sentimen netral menampilkan

tingkatan mutu layanan *Grab* Indonesia dengan layanan terbaik pada *GrabFood* sebesar 42, 29%, *GrabBike* sebesar 32, 57%, serta *GrabCar* sebesar 25, 14%. Pengujian analisis sentimen pada layanan *Grab* Indonesia dengan proses metode *Multinomial Naive Bayes Classifier* ini menciptakan akurasi sebesar 86, 57%

Menurut penelitian [4] menjelaskan angka persebaran *COVID-19* di Indonesia. Membatasi pertemuan di publik dengan Pembatasan Sosial Berskala Besar atau PSBB untuk mengurangi laju pertumbuhan korban *COVID-19*. Penerapan PSBB ini menerima banyak komentar dari pengguna media sosial khususnya melalui media sosial Twitter. Pada jurnal ini menggunakan 2 metode yaitu Random forest dan Suppor Vector Machine.

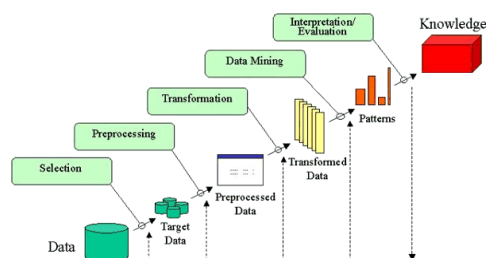
Berdasarkan penelitian [5] dijelaskan mengenai sentimen masyarakat terhadap aspirasi yang disampaikan melalui Twitter, Masukan pada sistem ini berupa tweet yang diperoleh dari Twitter menggunakan keyword seperti #coronavirusindonesia atau #covid-19 dengan jumlah data tidak melebihi 500 data tweet. Sedangkan outputnya berupa pengelompokkan sentimen positif dan negatif dari setiap tweet yang sudah melewati tahap pre processing dengan menggunakan Metode *Naive Bayes Classifier*. Hasil dari penelitian ini diperoleh 75 tweet dari pengukuran akurasi *recall* 32%, *precision* 80%, *F-Measure* 45% serta rata-rata akurasi 36%.

3. METODE PENELITIAN

Metode yang digunakan pada penelitian ini adalah metode algoritma Random Forest, yaitu sebuah metode algoritma yang digunakan dalam machine learning untuk tugas-tugas seperti klasifikasi dan regresi. Metode ini menggabungkan beberapa pohon keputusan (decision trees) menjadi sebuah "hutan" yang memiliki kemampuan untuk mengatasi overfitting (overfit) dan bias yang mungkin muncul dalam model pohon keputusan tunggal. Berikut adalah langkah-langkah utama dalam algoritma Random Forest.

3.1. Analisis Sentimen

Analisis senstimen adalh sentimen dari teks subjektif memproses, menganalisis, meringkas dan proses inferensial. Saat ini analisis sentimen dibagi menjadi penggunaan klasifikasi pembelajaran berdasarkan aturan dan klasifikasi pembelajaran mesin



Gambar 1. Proses KDD

Pada Gambar 1, dijelaskan mengenai Tahapan dari KDD (*Knowledge Discovery in Database*), Berikut adalah Penjelasan dari setiap tahapannya :

3.1.1. Data Selection

Pemilihan data yang relevan dan dapat dilakukan analisis dari data operasional. Data disatukan ke dalam sebuah tabel. Pada tahap ini, data yang digunakan yaitu data pelanggan.

3.1.2. Preprocessing

Pada tahapan ini diterapkan pola *preprocessing* yaitu pembersihan data guna meminimalisir data yang tidak lengkap.

3.1.3. Transformasi

Proses tranformasi data kedalam bentuk format tertentu sehingga data tersebut sesuai untuk proses data mining

3.1.4. Data Mining

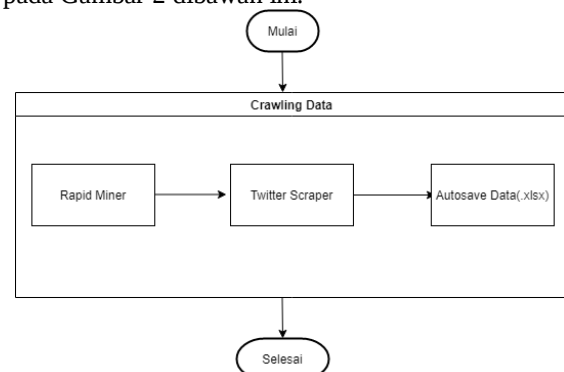
Data Mining adalah suatu proses untuk mengekstraksi informasi atau pengetahuan yang berguna dan berarti dari suatu database yang besar dan kompleks menggunakan teknik-teknik machine learning dan analisis statistik. Proses ini dilakukan secara otomatis dengan tujuan untuk menemukan pola dan hubungan yang tersembunyi dalam data yang ada.[6]

3.1.5. Interpretation/Evaluation

Tahap interpretasi merupakan simpulan hasil dari *data mining*. Berbagai hipotesa yang muncul dari *data mining* dilebur menjadi satu dan ditemukan kesimpulan akhirnya.

3.2. Sumber data

Sumber data untuk penelitian ini yaitu data sekunder. Data sekunder ialah sumber data yang tidak diperoleh langsung oleh peneliti. Sumber data yang didapat dalam penelitian bersifat data publik yang diperoleh dari proses *Crawling* pada media sosial twitter menggunakan aplikasi Rapid Miner versi 10.0 dengan operator *search twitter*. Teknik pengumpulan data yang dilakukan oleh peneliti seperti yang tertera pada Gambar 2 dibawah ini.



Gambar 2. Proses Crawling[7]

4. HASIL DAN PEMBAHASAN

Memuat hasil, Pengujian dan pembahasan tentang sentiment analisis yang telah dilakukan

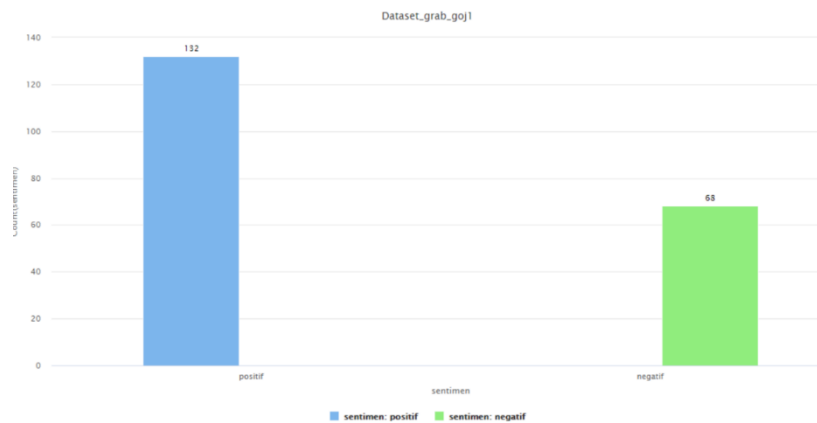
4.1. Data Selection

Data ini didapat dari proses crawling pada twitter menggunakan RapidMiner versi 10.1 dengan operator search twitter dengan kata kunci Gojek, setelah crawling Gojek dilakukan crawling dengan kata kunci Grab. Pengambilan data tanggal 11 Februari 2023. Data yang didapat sebanyak 361 data tweet, setelah proses pembersihan, data menjadi 200 tweet seperti yang terlihat pada Gambar 3. Pada Tabel 1 ialah beberapa contoh Sentimen yang terdiri dari sentimen positif dan negatif.

Tabel 1. Sentimen positif dan negatif

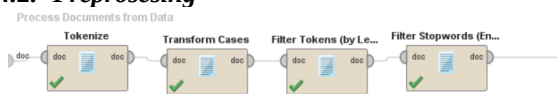
Sentimen	Text
Positif	Jalani hidup itu emang capek makanya ada aplikasi gojek tuk membantumu
Negative	Sekali2nya itu bang Gua beneran jarang banget beli lewat aplikasi Tapi yang gua bingung beberapa tempat ready aplikasi gojek tapi pembayarannya gabisa pakai gopay
Negative	Pagi min kenapa ya sy gabisa pake kode promo terus apakah kena soft banned Tolong bantuannya ya supaya bisa enjoy lg pake aplikasi gojek
Negative	Buat masuk ke aplikasi sekarang gojek grab sama shopee rata2 layanannya ada biaya aplikasinya dan itu belum termasuk komisi 20 yg diambil dari aplikasi ke driver Kalo di total aplikasi ambil komisinya 35
Positif	Baik Kak Setelah kami lakukan pengecekan untuk transaksi pembayaran perjalanan Commuterline dengan Aplikasi Gojek pada menu Gotransit di Stasiun Tambun saat ini sudah kembali normal ya Kak Terima kasih
Positif	Tapi kalo pecah pun ngga perlu panik kalau udah ada proteksinya Buat kamu yang belum punya cari proteksi yang paling pas buat kamu dari PasarPolis di aplikasi Shopee DANA dan Gojek ke menu GoSure
Positif	Bukan username tapi dn sering dipanggil Remy kok Sama ibu kos kadang garagara nama aplikasi Gojek sama Grab namanya Remy wkwkwk
Positif	Apa sepi sehingga gak bisa buat kontrak lagi Kalau sepi gak usah disesali mas mungkin Tuhan punya cara membantu Mas jgn menyerah suruh istri tetap jualan dari rumah dan bisa lewat aplikasi grab atau gojek yg mas jalani skrg tempat tdk selamanya menjamin Jadi ttp jualan
Positif	Bahkan ketika covid melanda gojek tetap memberikan penyuluhan ke driver dlm bentuk kopdar online melalui aplikasi zoom Gw...

Sentimen	Text
Positif	Makanannya banyak jenis dan enak Aplikasi kayak gojek dan sejenisnya memudahkan banget Senjata api ga dibebasin penggunaannya ATM dimanamana Tingkat kepeduliannya masih tinggi dll banyak sih sebenarnya
Negative	ya Allah meresahkan nian bapak22 gojek yg sudah off tapi dk matike aplikasi nak dicancel dk biso
Positif	PROMO GOFOOD Mulai sekarang kamu udah bisa order Sejuk dengan lebih hemat dan cepat lewat aplikasi Gojek loh Dengan order Sejuk lewat Gofood kamu bakalan nemuin 9 promo menarik mulai dari diskon presentase potongan harga gratis ongkir dan masih banyak lainnya
Negative	Hai Kak terkait laporan kamu mengenai kendala Akun Gojek kami sarankan kamu untuk membuat laporan melalui halaman Bantuan di aplikasi Gojek dengan cara ketik "Akun saya dibajak orang lain" di kolom pencarian dan isi formulir yang ada di aikel tersebut Terima kasih
Positif	Definisi kenyang bego Garagara hujan kenyang habis makan buka aplikasi Gojek buat pesan gocar Untung sadar kan bawa motor Nyaris
Positif	nder ada yg namanya apotik k24 bisa beli lewat aplikasi/website dianter ke rumah manja bgt atau bisa lewat goma di gojek apalagi gojek suka ngasih voucher potongan goma
Positif	Ini nih yang bikin kamu serba bisa! Bisa bayar apa aja dan lebih hemat tentunya Beli paket voucher GoPay Serba Bisa mulai Rp9900 bisa dapet CASHBACK hingga Rp930000 lho Yuk beli paket vouchernya di aplikasi Gojek sekarang SK berlaku cek di
Negative	Simple nya kalau orang yg km mintain tolong gakbisa atau ga mau Coba buka hp mu Aplikasi gojek punya Ada goma disitu Ga ada gojek Grab punya dong Ada grabma atau grab jastip disitu :Hal kekini ga perlu didebatin Tiap orang punya hak sendiri buat menilai sesuatu
Positif	Tinggal buka aplikasi grab atau gojek
Negative	Gatau mau cerita kesiapa tapi emng wajar cewe minta jemput ke cowo yg udah punya pacar Tau banget loh klo si cowonya udh punya cewe ga ngehargain perasaan pacarnya mbbaaaaa!!! download aplikasi goje mbaaa



Gambar 3. Labeling positif negatif

4.2. Preprocessing



Gambar 4. Tahap Preprocessing

Pada tahapan ini diterapkan pola preprocessing yaitu pembersihan data guna meminimalisir data yang tidak lengkap. Berdasarkan Gambar 4, Tahap *preprocessing data* mempunyai 4 tahapan.

4.2.1. Tokenize

Tahap ini berfungsi untuk menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks.

4.2.2. Transform Cases

transform cases dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil

(*lowercase*) atau menjadi huruf kapital (*Uppercase*), pada penelitian ini semua huruf dirubah kedalam huruf kecil, karena mayoritas teks berupa tulisan opini.

4.2.3. Filter Tokens

Operator ini memungkinkan kita untuk memangkas kata-kata yang terlalu panjang

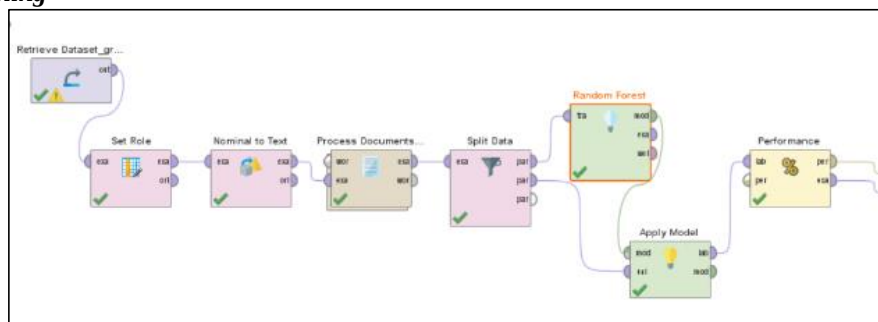
4.2.4. Filter Stop Word

Pada tahap ini, dilakukan eliminasi kata-kata yang termasuk dalam kategori stopwords. Stopwords adalah kata-kata yang sering muncul dalam teks tetapi dianggap tidak memiliki makna signifikan.[8]

4.3. Transformasi Data

Tahap transformasi, melibatkan proses perubahan atau transformasi data yang memiliki entitas yang ambigu menjadi bentuk data yang sah dan dapat diproses ke tahap berikutnya.

4.4. Data Mining



Gambar 5. Penerapan Algoritma Random Forest

Pada langkah ini, dilakukan proses klasifikasi menggunakan algoritma Random Forest. Pada tahap ini dilakukan *Split Data*, sehingga data akan dibagi menjadi dua bagian, yakni data *training* dan data *testing*, dengan presentase data *training* sebesar 60% dan data *testing* sebesar 40%

Setelah membagi data menjadi dua bagian, yaitu data *training* dan data *testing*, langkah selanjutnya adalah melakukan proses klasifikasi menggunakan RapidMiner. Untuk mengevaluasi performa klasifikasi, kita akan melakukan perhitungan

menggunakan *confusion matrix*. Hasil dari matriks konfusi yang dihasilkan menggunakan algoritma Random Forest dapat dilihat dalam Gambar 6.

accuracy: 76.25%

	true positif	true negatif	class precision
pred. positif	61	19	76.25%
pred. negatif	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 6. Confusion matrix

4.5. Evaluasi Model

Umumnya confusion matrix mempunyai 3 metode untuk pengujian, sebagai Berikut:

1) Accuracy

Merupakan metode untuk menguji suatu algoritma menurut tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Rumus untuk menghitung nilai akurasi adalah :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$= \frac{61+0}{61+0+19+0} = \frac{61}{80} = 0.7625 = 76.25\%$$

2) Precision

Metode tes algoritma yang membandingkan data benar yang diperoleh sistem dengan seluruh jumlah data yang diperoleh sistem yang benar atau salah. Rumus untuk menghitung nilai precision adalah:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$= \frac{61}{61+19} = \frac{61}{80} = 0.7625 = 76.25\%$$

3) Recall

Merupakan salah satu proses pengujian dari suatu Algoritma yang bisa digunakan untuk membandingkan data yang tepat yang didapat sistem dengan seluruh data yang tepat diambil atau tidak diambil oleh sistem. Rumus untuk menghitung nilai recall adalah :

$$\text{Recall} = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$= \frac{61}{61+0} = \frac{61}{61} = 1 = 100\%$$

Pada Gambar 2 diketahui bahwa nilai class precision sebesar 76.25% dan nilai class recall sebesar 100%. Maka dari evaluasi klasifikasi menggunakan algoritma Random Forest menghasilkan nilai akurasi sebesar 76.25%

5. KESIMPULAN DAN SARAN

Setelah melakukan implementasi dan pengujian, beberapa temuan penting yang dapat disimpulkan termasuk fakta bahwa model yang diterapkan untuk mengklasifikasikan sentimen di Twitter terkait Transportasi online yaitu Grab dan Gojek cenderung menunjukkan lebih banyak komentar yang bersifat positif dibandingkan dengan komentar yang bersifat negatif. Akurasi yang dihasilkan mencapai 76.25% yang menandakan bahwa algoritma random forest cukup baik dalam melakukan klasifikasi sentimen terkait Grab dan Gojek pada Twitter.

Dalam studi berikutnya, disarankan untuk mempertimbangkan penggunaan algoritma alternatif sebagai metode untuk mengevaluasi sejauh mana

performa algoritma tersebut dibandingkan dengan algoritma random forest dalam hal klasifikasi.

DAFTAR PUSTAKA

- [1] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta Pada 2022," *DataIndonesia.Id*, 2022.
- [2] K. A. Rokhman, B. Berlilana, and P. Arsi, "PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN DECISION TREE UNTUK ANALISIS SENTIMEN REVIEW KOMENTAR PADA APLIKASI TRANSPORTASI ONLINE," *Journal of Information System Management (JOISM)*, vol. 3, no. 1, 2021, doi: 10.24076/joism.2021v3i1.341.
- [3] S. Mandasari, B. H. Hayadi, and R. Gunawan, "Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier," *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, vol. 5, no. 2, 2022, doi: 10.53513/jsk.v5i2.5635.
- [4] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *Jurnal Informatika Upgris*, vol. 7, no. 1, 2021, doi: 10.26877/jiu.v7i1.7099.
- [5] M. M. Mala Olhang, S. Achmadi, and F. X. A. Wibisono, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP COVID-19 DI INDONESIA MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER (NBC)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 2, 2020, doi: 10.36040/jati.v4i2.2695.
- [6] H. Sujadi, "ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE," *INFOTECH journal*, vol. 8, no. 1, 2022, doi: 10.31949/infotech.v8i1.1883.
- [7] A. Bastian, Ardi Mardiana, and Dinda Sri Wulansari, "Sentiment Analysis Against Political Figure's Billboard During Pandemic Using Naïve Bayes Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, 2023, doi: 10.29207/resti.v7i1.4643.
- [8] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.