

Sentimen Analisis Masyarakat Terhadap Rakyat Palestina dengan Klasifikasi Naive Bayes

Hafiz Irsyad¹, Akhsani Taqwiym²

¹Teknik Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

²Sistem Informasi, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang

E-mail: *¹hafizirsyad@mdp.ac.id, ²akhsani.taqwiym@mdp.ac.id

Abstrak – Pada zaman teknologi 4.0 media *social* sangat pesat perkembangannya, ada yang menggunakan media *social* untuk berjualan, aksi pengumpulan dana, memberitakan kejadian secara *live*. Beberapa hari yang lalu, palestina mendapatkan agresi dari militer Israel sehingga jagad dunia mengetahui aksi yang sungguh sangat tidak terpuji yang dilakukan oleh militer Israel. Penting untuk menganalisis perasaan agresi militer terhadap Palestina dari banyaknya informasi di media *online*. Data yang digunakan adalah *Twitter*, salah satu *platform* media sosial. Riset ini dilakukan untuk menganalisis reaksi masyarakat terhadap data berupa *tweet* dan mengelompokkannya menurut metode *Naive Bayes*. Berdasarkan *tools* yang digunakan adalah *orange*, maka didapatkan hasil sentiment positif 56%, sentiment *negative* 11% dan sentiment netral 33% dengan tingkatan akurasi 75%. Hasil ini menunjukkan bahwa tingkat emosi positif *tweet* publik lebih tinggi daripada tingkat emosi negatif dan netral.

Kata Kunci — palestina, analisis sentiment, *twitter*, *naive bayes classifier*

Abstract – In the era of technology 4.0, social media is developing very rapidly, some are using social media to sell, fundraising, reporting events live. A few days ago, Palestine received aggression from the Israeli military so that the world knows the very disgraceful action carried out by the Israeli military. It is important to analyze the feelings of military aggression against Palestine from the abundance of information in the online media. The data used is *Twitter*, one of the social media platforms. This research was conducted to analyze the public's reaction to the data in the form of tweets and classify them according to the Nave Bayes method. Based on the tools used are orange, the results obtained are 56% positive sentiment, 11% negative sentiment and 33% neutral sentiment with an accuracy rate of 75%. These results indicate that the level of positive emotion in public tweets is higher than the level of negative and neutral emotion.

Keywords — palestine, sentiment analysis, *twitter*, *naive bayes classifier*

1. PENDAHULUAN

Pada zaman teknologi 4.0 media sosial tidak akan pernah lepas dari kehidupan sehari-hari. Sehingga saat sekarang media sosial sudah setara dengan kebutuhan primer dari kehidupan manusia. Menurut data indeks Tetra pax, sekitar 132 juta orang Indonesia atau lebih dari setengah dari total populasi adalah pengguna internet aktif, 40% di antaranya adalah pengguna jejaring sosial. Jadi masyarakat Indonesia sangat membutuhkan media sosial.

Bagi masyarakat yang mampu mengendalikan media sosial maka akan menuai hasil yang positif, bahkan juga bisa menjadi bumerang bagi mereka yang membuat kegaduhan dan sebagainya. Sehingga pemerintah untuk menertibkan para pengguna media sosial yang “nakal” dengan menggunakan KUHP No. 19 Tahun 2016 dan UU No. 11 Tahun 2008 tentang Perdagangan Informasi Elektronik (ITE) [1].

Palestina adalah memiliki teritorial sekitar 2.400 mil persegi. Palestina mempunyai andil besar terhadap sejarah kuno dan modern dari Timur Tengah. Pada tahun 1948, Palestina memiliki teritorial antara Laut Mediterania dan Sungai Yordan. Secara teoritis, Palestina terdiri dari Tepi Barat (wilayah yang memisahkan Israel modern dari Yordania) dan Jalur Gaza (wilayah yang berbatasan dengan Israel modern dan Mesir). Tetapi beberapa mengklaim bahwa tanah ini adalah apa yang sekarang disebut Israel [2].

Konflik Palestina dengan Israel bukan hanya permasalahan agama saja, akan tetapi terkait juga akan tragedi kemanusiaan. Telah banyak korban yang ditunjukkan tentara Israel terhadap rakyat sipil Palestina. Dalam kurun 10 tahun terjadi agresi militer 3 (tiga) kali yaitu pada tahun 2021 berlangsung 2 tahun, 2014 kembali gencatan senjata kurang lebih 60 hari dan terakhir 2021 terjadi gencatan senjata selama 11 hari dan yang berdampak menjadi korban adalah rakyat sipil [3].

Menarik dari sejarah beberapa sejarah dari literatur yang telah beredar dan berhubungan langsung dengan aksi kemanusiaan, maka masyarakat dengan menggunakan media sosial sebagai *people power* untuk berbagai macam yang bisa dilakukan, yakni memberitakan terkait kejadian terkini, informasi korban berjatuhan, bahkan sampai bisa membuat galangan dana bagi rakyat Palestina yang terkena dampak agresi militer dari Israel.

Kehadiran *platform twitter* sehingga dapat kita temui secara mudah berbagai macam opini dan *statement* yang keluar dari masyarakat baik berupa positif, negatif maupun netral. Terkait hal tersebut kita membutuhkan semua opini tersebut agar bisa mengetahui opini mana yang positif, negatif dan netral.

Survei pengklasifikasi opini didasarkan pada total 1.553 data yang digunakan, masing-masing dengan 0,70% dianggap data pelatihan 0,30% selebihnya dianggap data penguji. Metode *naive bayesian* dapat mencapai rata-rata sebesar 0,878% dengan menghasilkan nilai presisi positif sebesar 0,976%, nilai presisi negatif sebesar 0,829%, dan nilai presisi netral sebesar 0,833% [4]. Berdasarkan tragedi kemanusiaan yang telah terjadi di Palestina, sehingga penulis tertarik untuk melakukan penelitian dengan tema “Sentimen Analisis Masyarakat Terhadap Rakyat Palestina di *Twitter* dengan Klasifikasi *Naive Bayes*”.

2. METODE PENELITIAN

2.1. 2.1. Text Mining

Penambangan teks merupakan penggalian informasi yang bermanfaat dari sejumlah sekumpulan dokumen dari suatu masa dengan mengidentifikasi ciri khas yang menarik [5]. Penambangan teks menemukan keputusan yang baru dengan menghasilkan kerangka berpikir secara otomatis dari sumber dokumen yang berbeda [6]. Ada beberapa tingkat keterampilan dalam penambangan teks. Penambangan teks mengikuti pola yang telah tersedia pada aplikasi penambangan data dan dapat dibagi menjadi 4 tingkat *primary function*: pemrosesan awal, manipulasi, pemrosesan pengumpulan dokumen, dan penambangan manipulasi inti [5]. Gambar 1 menunjukkan tingkat arsitektur *text mining*.



Gambar 1. Perancangan metoda dalam *Text Mining* [5]

2.2. Pre-processing

Pre-processing adalah langkah pertama dalam pencarian teks. *word pre-processing* bertujuan untuk menghilangkan data dari yang tidak dibutuhkan sehingga menghasilkan dokumen yang lebih terperinci dan lebih terstruktur. Langkah-langkah dalam *proses pre-processing* teks [5] adalah sebagai berikut:

- a) *Tokenizing* artinya memecah *text* menjadi satu kata dan menghilangkan tanda baca beserta angka sesuai dengan penggunaan kosakata terlibat.
- b) Hapus kata kunci, seperti menghapus kata-kata yang tidak penting dari teks. Kasus ini ditangani untuk meningkatkan ketepatan bobot. Artikel ini menggunakan fitur *stop word removal* untuk menghilangkan dan, atau kata-kata seperti ini.
- c) *Stemming* atau kata dasar digunakan untuk mengubah sufiks lain menjadi kata dasar. Dalam penelitian ini, sufiks dan dihilangkan karena ada bentuk lampau dalam teks bahasa Inggris.

2.3. 2.3. Naive Bayes

Thomas Bayes merupakan pencetus dalam penggunaan algoritma klasifikasi yang dinamakan dengan *Naive Bayes*. Klasifikasi adalah proses pemeriksaan kumpulan data lama untuk menemukan rumus yang dapat meng-klasifikasikan dokumen terbar. *Naive Bayes* adalah algoritma klasifikasi probabilitas sederhana berdasarkan teorema *Bayes* [5]. Aturan biasanya merupakan bahwa terdapat nilai sebuah atribut independen dan mempengaruhi nilai atribut lainnya. Model *Naive Bayes* memungkinkan setiap atribut untuk berkontribusi sama pada keputusan akhir, membuatnya secara komputasi lebih efisien daripada algoritma klasifikasi *text* lainnya. Dokumen yang diperoleh selama pelatihan berisi dokumen konstan dari setiap data pelatihan. Dokumen digunakan untuk data uji untuk mengetahui nilai keakuratan dokumen yang diperoleh. [6]. *Naive Bayes* menghitung frekuensi dan kombinasi nilai dalam kumpulan data untuk menghitung probabilitas [7]. Probabilitas dihitung menggunakan (1) [6].

$$P(C|X) = \frac{P(X|C)P(c)}{P(x)} \dots\dots\dots (1)$$

dengan

- x* : Data dengan kelas yang tidak ditentukan.
- c* : Data hipotetis adalah kelas khusus *P*
- P(c/x)* : Probabilitas hipotetis (probabilitas posterior) menurut kriteria *P*
- P(c)* : Probabilitas hipotetis (probabilitas sebelumnya)
- P(x/c)* : Probabilitas dalam kondisi hipotetis
- P(x)* : Probabilitas *c*

2.4. Analisa Sentimen

Analisis sentimen, dikenal sebagai pembaharuan terhadap sebuah opini, bidang penelitian yang menganalisis pendapat, perasaan, ulasan, nilai, sikap, serta perasaan tentang seseorang seperti terhadap produk-produk dan layanan-layanan yang ada pada masyarakat, Organisasi, individu, masalah, peristiwa, topik atau diskusi sehingga mencakup ruang masalah yang besar. beberapa tugas sedikit berbeda dari banyak dokumen yang telah tersedia. Tapi sekarang semuanya di bawah kap analisis sentimen dan penambangan opini. Sementara istilah analisis sentimen lebih umum digunakan di industri, analisis sentimen dan penambangan opini biasanya digunakan di dunia akademis [4].

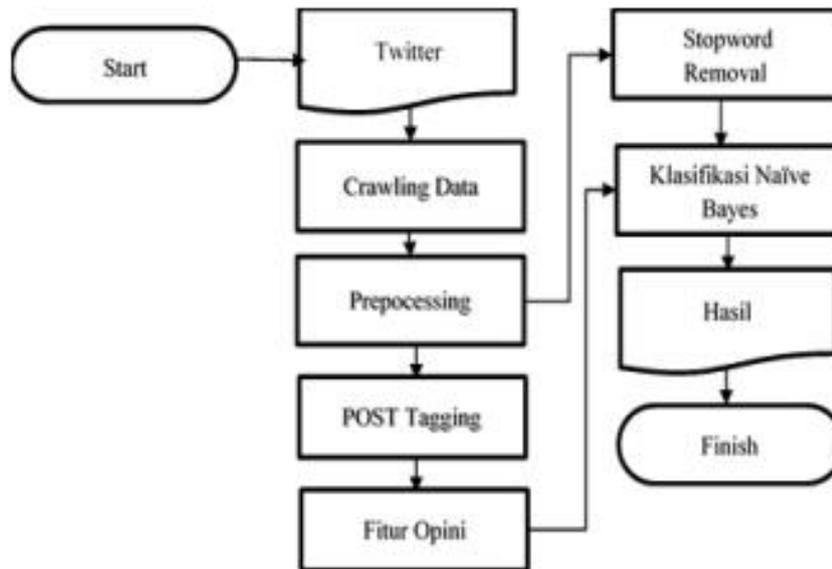
2.5. Twitter Rest API

Twitter adalah interaksi sosial dalam berupa daring sangat populer yang memungkinkan pengguna untuk berbagi pesan singkat, termasuk layanan pesan singkat (SMS). Pengguna berbagi pemikiran, tautan, dan gambar di *Twitter*. Wartawan berkomentar langsung. Bisnis mempromosikan produk mereka dan berinteraksi dengan pelanggan. *Twitter* memiliki banyak metode berbeda, dan karena ada 500 juta *tweet* per hari, itulah jumlah data yang mereka analisis dan gunakan. *Twitter REST API* menyediakan akses terprogram untuk membaca dan mengakses data *Twitter* (menulis

tweet baru, membaca data profil untuk penulis dan pelanggan, dll.). *Twitter* telah menyediakan *REST API* untuk dapat digunakan dalam dunia Pendidikan dan juga tersedia di JSON [8]

3. HASIL DAN PEMBAHASAN

Gambar 2 memberikan gambaran umum tentang proses survei. Skenario penelitian ini adalah sebagai berikut.



Gambar 2. Gambaran umum penelitian [4]

Pada Gambar 2 dapat dijabarkan:

- Tools Orange* mendapatkan INPUT berupa *dataset Twitter*. Data tersebut dapat diperoleh dengan menggunakan kata kunci *hashtag #INASTandsWithPalestine*.
- Melakukan *pre-processing* terhadap *dataset* agar dapat memenuhi klasifikasi dan mempermudah dalam memproses data baik berupa dokumen.
- POST Tagging* berfungsi sebagai pencarian opini. Selain berfungsi untuk opini, *Post Tagging* jg dapat mengekstrasi fitur opini.
- Stopword Removal* Penghapusan kata derau digunakan untuk menghilangkan kata-kata yang umum tetapi tidak relevan dalam penelitian ini.
- Langkah kemudian lakukan klasifikasi *sentiment* data dengan menggunakan *Naive Bayes*.

3.1. Crawling Data Twitter

Dataset yang digunakan dalam penelitian ini adalah *tweet* dan *retweet* dari *hashtag* yang digunakan adalah *#INASTandsWithPalestine*. *Dataset* diperoleh dengan memanfaatkan API dari *Twitter* yang terhubung dengan aplikasi *Orange*.

Crawling data dimulai pada tanggal 20 Mei 2021 sampai tanggal 22 Mei 2021, sehingga memperoleh data *tweet* maupun *retweet* adalah sebanyak 697. Adapun hasil dari perolehan *crawling* data dapat dilihat pada Tabel 1.

Tabel 1. Hasil Crawling Data Twitter

No	Tweet
1	GEROMBOLAN BIPANG REPUBLIK INDONESIA, ADALAH GEROMBOLAN PENCURI UANG RAKYAT/NKRI.
2	Umat Islam,Mari Kita Bersatu,Berdonasi Kpd Rakyat Negara Republik Palestina.
3	#INASTANDSWITHPALESTINE #INASTANDSWITHPALESTINE https://t.co/DMc7I2ywWp
4	https://t.co/uZU4RDMZBq ,2021-05-27 04:56:37,in,,0.0,0.0,,frogho Axson,"Hidup,adalah Perjuangan.",2363.0,6170.0,119.0,56.0,0.0,False,,,@FroghoA
5	Apapun Alasan dr Binatang Bani Bipang,EGP. Yg Jelas,Tetap Dukung: #INASTANDSWITHPALESTINE

Dari hasil *crawling* data pada *twitter*, terdapat kata-kata yang bakal mempengaruhi hasil dari klasifikasi, kata-kata tersebut dapat kita temui berupa link, *emoticon* dan lain-lain. Untuk membersihkan kata-kata tersebut kita membutuhkan *pre-processing* sehingga setelah melakukan *pre-processing* dapat menghasilkan kata-kata yang terstruktur [9].

3.2. Pre-processing

Pada tahapan ini kita akan melakukan *case folding*. *Case folding* proses mengubah semua karakter dalam dokumen menjadi satu subjek. Tidak semua dokumen teks (yang di-*crawling*) menggunakan semua huruf besar secara konsisten. Salah satu hasil dari proses dari *case folding* adalah mengubah huruf menjadi kecil.

3.3. Tokenizing

Tokenisasi adalah proses memecah teks menjadi kata-kata tunggal dan menghilangkan tanda baca dan angka, menurut definisi kamus enkripsi. Fokusnya adalah pada pemilihan data yang dihilangkan karena pengguna melewati lebih banyak kata dan menggunakan kata-kata yang lebih standar karena mereka hanya dapat membuat 280 karakter di *Twitter* sesuai dengan *dataset* yang dijelaskan. Kata-kata yang Anda dapatkan dari *dataset* memiliki bentuk 'sekarang' menjadi 'sekarang', 'Mdh2an' menjadi "Harapan", "konteks" menjadi 'Kontak' dan seterusnya. [4]. Hasil dari *tokenizing* dapat dilihat pada tabel 2.

Tabel 2. Hasil *Tokenizing*

Weight	Word
266	Palestina
172	Inastanswithpalestine
93	Dukung
91	Negara
77	Ayo

3.4. Stopword Removing

Stopword removing suatu proses untuk membuang kata-kata yang tidak penting dalam *text*. Kenapa ini dilakukan, agar pada akhirnya dapat mencegah terjadinya akurasi yang berlebihan dari pembobotan. Dalam penelitian ini, *stopword removing* digunakan untuk menghilangkan kata-kata seperti dan, atau, mungkin, ini, dan itu [4]. Pada penelitian ini menggunakan *library Python* Sastrawi dalam proses *steeming* Bahasa Indonesia.

3.5. Post Tagging

Tahapan semua sebagaimana yang telah dijabarkan pada poin-poin diatas, maka selanjutnya melakukan pelabelan data, pelabelan data ini yang diambil hanya atribut teks saja. Pelabelan ini dilakukan secara manual, sehingga dapat memasukan pada kategori positif, *negative* maupun netral. Kategori positif berupa kata-kata yang memberikan penyemangatan terhadap warga palestina, sementara untuk kata-kata negatif berupa unsur SARA dan yang mengakibatkan kearah adu domba. Untuk pelabelan sendiri dapat dilaksanakan dengan pemberian tanda label -1 yang artinya adalah *negative*, label 1 adalah kategori bernilai positif dan sementara untuk kategori bernilai netral diberikan label 0.

Dalam tahapan *pre-processing* banyak yang dibuang kata-kata didalam *tweet* dari masyarakat itu sendiri. Kata-kata yang dibuang berupa yang tidak mempunyai arti, *emoticon*, *retweet*, *link*, *no handphone* serta *tweet* yang berisikan video maupun gambar sehingga saat melakukan *crawling* terdapat *tweet* menjadi kosong. Untuk pelabelan secara manual bisa ditampilkan pada contoh di tabel 3.

Tabel 3. Contoh pelabelan data berdasarkan fitur

Dukung	Negara	Palestina	Donasi	Dukung
1	1	0	1	1
-1	0	1	0	-1
0	-1	-1	-1	0

Evaluasi ini diberlakukan untuk memvalidasi hasil klasifikasi serta mengukur atribut kinerja. Parameter uji yang digunakan adalah derajat. Perhitungan parameter ini diambil dari tabel *Confusion Matrix* [7]. Hitung tingkat kecocokan presisi, tingkat pemulihan dari matriks konfusi [8]. *Akurasi*, *fit rate*, *recovery rate* dihitung menggunakan (2) sampai (4).

Nilai akurasi didapatkan dari jumlah data bernilai positif yang diprediksi positif dan data bernilai negatif yang diprediksi negatif dibagi dengan jumlah seluruh data di dalam dataset [10].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots (2)$$

Precision sebenarnya adalah probabilitas bahwa kasus yang diprediksi positif termasuk dalam kasus tipe positif.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3)$$

Recall adalah probabilitas bahwa instance kategori positif diprediksi dengan benar menjadi positif.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (4)$$

F1 Score juga dikenal sebagai nilai skor kunci F1 atau F, berasal dari hasil *fit* and *recall* antara kategori yang diprediksi dan kategori yang sebenarnya.

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall+Precision} \dots\dots\dots (4)$$

dengan

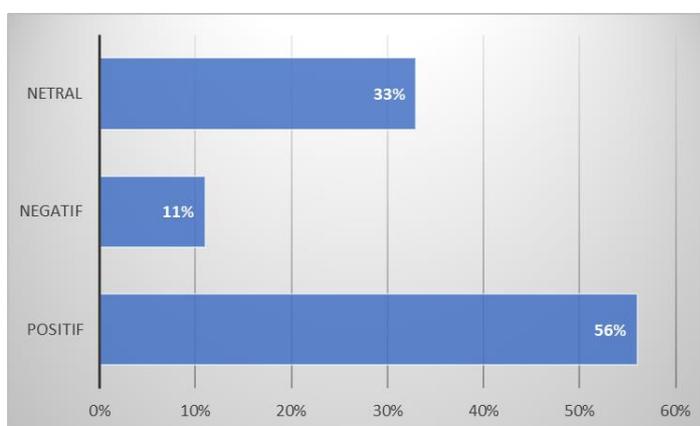
- TP = *True Positif Count*.
- TN = *True Negative Count*.
- FP = *False Positive Count*.

Tabel 4. Hasil pengujian model

Kelas	Prediksi Positif	Prediksi Negatif	Prediksi Netral	Total	Recall	Precision	F-Measure
Positif	100	9	10	119			
Negatif	1	22	1	24	0,6445	0,8889	0,7350
Netral	2	4	65	71			
Akurasi				214	0,7561		

Pada tabel 4 diatas dapat dijelaskan hasil pengolahan dokumen dari perbandingan kelas positif, *negative* dan netral. Pada perhitungan *recall* dengan orange menghasilkan 0,6445. Kemudian pada perhitungan *precision* dengan orange menghasilkan 0,8889. Dan perhitungan terakhir perhitungan *F-Measure* menghasilkan 0,7350. Sehingga untuk akurasi mendapatkan 0,7561.

Perolehan data *sentiment* dari kelas positif adalah 0,556. Pada perolehan data untuk *sentiment* dari kelas *negative* adalah 0,112. Untuk perolehan *sentiment* dari kelas netral adalah 0,331. Hasil dari sentimen dapat kita lihat pada gambar 4.



Gambar 4. Perbandingan hasil sentimen

4. KESIMPULAN

Berdasarkan hasil dari penelitian ini yang terkait dengan sentiment masyarakat terhadap rakyat palestina dengan menggunakan *navie bayes* untuk menentukan kelas, maka didapatkan hasil *sentiment* positif adalah 56%, *sentiment* negatif mendapatkan hasil 11% dan untuk *sentiment* yang netral menghasilkan 33% dan dengan akurasi 75%. Dengan penelitian ini banyak masyarakat Indonesia berpartisipasi untuk mendukung warga palestina berupa doa, donasi serta aksi damai menyuarakan kemerdekaan terhadap Palestina yang sedang mengalami agresi militer Israel berdasarkan data *tweet* maupun *retweet* dari *twitter*.

Dalam penenilitan ini masih banyak terdapat kekuarangan berupa pemberian label, penentuan data latih dan data uji, pengujian model dan lain-lain, yang mana dikemudian hari bisa diperbaiki atau diperbarui oleh para peneliti selanjutnya

DAFTAR PUSTAKA

- [1] Kontan.co.id, "Pidana di UU ITE efektif menjerat pengguna Medsos, hingga Oktober ada 324 kasus," Kontan.co.id, -, 2021.
- [2] Kompas.com, "Mengapa Palestina Tidak Diakui Sebagai Negara?," Jakarta, 2021.
- [3] C. D. tvOne, "PANAS! Adu Argumen Pro Israel vs Pro Palestina Bongkar-bongkar Sejarah," tvOne, Jakarta, 2021.

- [4] A. F. M. R. P. Hafiz Irsyad, "Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naïve Bayes," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 1, pp. 30-34, 2019.
- [5] J. S. R. Feldman, *The Text Mining Handbook*, New York, USA: Cambridge University Press, 2007.
- [6] S. D. H. P. D. N. C. M. Kini M., "Text Mining Approach to Classify Technical Research Document Using Naive Bayes," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 7, pp. 286-391, 2015.
- [7] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creative Information Technology Journal*, vol. 2, no. 3, pp. 207-217, 2015.
- [8] J. A. M. Rani, "Twitter Data Predicting Geolocation Using Data Mining Techniques," *International Journal of Innovative Research in Computer*, vol. 4, no. 6, pp. 10446-10453, 2016.
- [9] V. & A. R. Srividhya, "Evaluating preprocessing techniques in text categorization," *International journal of computer science and application*, vol. 47, no. 11, pp. 49-51, 2010.
- [10] L. Afifah, "ilmudatapy," *ilmudatapy*, [Online]. Available: <https://ilmudatapy.com/apa-itu-confusion-matrix/>. [Accessed 22 May 2021].

