

BAB 2. TINJAUAN PUSTAKA

Penelitian yang dilakukan penulis tidak akan terjadi jika penelitian-penelitian sebelumnya dilakukan, baik dalam segi judul cara dan metode yang dilakukan sebelumnya. Penelitian yang sebelumnya akan digunakan penulis sebagai referensi ataupun acuan dalam penelitian ini. Berikut ini adalah penelitian yang menjadi referensi penulis.

Penelitian pertama dari Pristiyono *et al* [19] yang telah melakukan penelitian dengan topik wabah Covid-19, peneliti melakukan analisis sentimen masyarakat Indonesia terhadap pemberian Vaksin Covid-19. Analisis sentimen tersebut dilakukan dengan menggunakan metode *Naïve Bayes*. Dimana data diambil dari tweet masyarakat di Twitter. Tweet yang dikumpulkan menghasilkan sentimen masyarakat berupa 3.4 ribu tweet bersifat negatif (56%), lebih dari 2,4 ribu tweet bersifat positif (39%), dan 301 tweet sisanya (1%) bersifat netral selama periode minggu ini (15-22 januari 2021).

Penelitian berikutnya dari Mohammed Alhajji *et al* [20] yang telah melakukan penelitian untuk mengetahui pandangan dan nilai dari sikap Negara Saudi Arabia terhadap tindakan pencegahan COVID-19 melalui analisis sentimen konten Twitter. Penelitian ini menggunakan data dari tweet Twitter sebanyak 47000 tweet yang dimana hasil pandangan masyarakat Saudi Arabia terhadap pencegahan Covid-19 adalah 50% positif dan 50% Negatif. Karena data yang ditunjukkan memiliki nilai sebesar 50% dan 50%. Peneliti melakukan penilaian dengan nilai akurasi dari algoritma dengan metode *Naïve Bayes*. Data pelatihan yang digunakan sebesar 11000 data dan mendapatkan hasil akurasi dari *F1-score* sebesar 89%. Karena hasil yang didapatkan belum maksimal peneliti melakukan analisis berikutnya dengan mengubah kata kunci dari tweet untuk mencari data dari Twitter.

Penelitian berikutnya dari R, Muthusami *et al* [21] telah melakukan penelitian analisis sentimen dengan menggunakan metode pembelajaran mesin, SVM, *Max Entropy*, *Decision Tree*, *Boosting*, and *Random Forest*. Peneliti melakukan perbandingan dengan menggunakan data dari Twitter. Peneliti mengambil tweet

dari Twitter. Untuk hasil yang didapatkan *LogitBoost*, pendekatan campuran, berkinerja lebih baik dengan akurasi 74%.

Penelitian berikutnya dari Akib Mohi Ud Din Khanday *et al* [22] telah melakukan analisis sentimen pada tweet di Twitter. Dengan menggunakan perbandingan metode MNB (*Multinomial Naïve Bayes*), SVM (*Support Vector Machine*) dan *Decision Tree*. Dari ketiga metode tersebut metode *Decision Tree* mendapatkan nilai akurasi yang terbaik yaitu akurasi 98,5%, *precision* 99%, *recall* 99% dan Skor *F1-Score* 99%.

Penelitian kali ini dari P. Pulung Hendro, *et al* [23] telah melakukan penelitian dengan sentimen analisis dengan topik wadah Covid-19 dengan menggunakan metode SVM (*Support Vector Machine*), Normalized Poly Kernel dan MNB (*Multinomial Naïve Bayes*). Hasil penelitian menunjukkan bahwa SVM dengan Normalized Poly Kernel mengungguli MNB baik pada data aspek ekonomi maupun aspek umum yang dikategorikan ke dalam dataset dua kelas dan tiga kelas. SVM memperoleh performansi tertinggi pada data aspek umum dengan menggunakan dataset dua kelas dalam hal akurasi rata-rata, presisi, *recall*, dan *f* ukuran masing-masing dengan nilai 82.00%, 82.24%, 82.01%, dan 81.84%.

Penelitian kali ini dari B. Pokharel [24] telah melakukan penelitian dengan topik wadah Covid-19 yang dimana peneliti melakukan analisis sentimen dengan menggunakan Twitter Key Api dan bantuan *Library Python*. Data tweet yang diambil sebanyak 615 di negara lokasi Nepal. Hasil dari tweet tersebut di analisis dengan menggunakan metode *Naïve Bayes* dan mendapatkan hasil yang menunjukkan jumlah korban netral terkait kata kunci virus corona dan covid-19 untuk polaritas cukup tinggi yaitu 58% dan sebagian besar pencatatan bersifat objektif yaitu 62,27%. Dari hasil ini peneliti mengatakan bahwa pandangan masyarakat nepal terhadap Covid-19 dapat berubah-ubah setiap tweet mereka. Peneliti juga memberikan masukan berupa peningkatan dalam *library Textblob* yang menggunakan *Naïve Bayes* perlu ditingkatkan.

Penelitian dari Bryan Pratama *et al* [25] telah melakukan analisis sentimen publik terhadap korps Brimob Polri. Penelitian ini menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) dalam melakukan klasifikasi berdasarkan 1000 data di twitter. *RapidMiner* dan *Gataframework* digunakan untuk mengumpulkan data twitter, dengan acuan kata yaitu “Brimob” dan “Mako Brimob”. Setelah dilakukan pengumpulan data lalu dilakukan pembersihan terhadap data yang tidak dibutuhkan serta melakukan pelabelan, apakah tweet tersebut positif atau tweet tersebut negatif. Selain pembersihan dan pelabelan, data yang diambil juga di *preprocessing* kembali dengan menghilangkan *noise* pada data, serta melakukan tokenisasi dan *word* normalisasi subproses. Hasil penelitian ini menyatakan bahwa model SVM mengungguli model NB, dengan akurasi 86.96%, *precision* 91.06%, dan 83.1% *recall*, sedangkan untuk NB akurasi yang didapat adalah 86.48%, dengan *precision* 88.86%, dan 84.68% *recall*. Penelitian ini menyimpulkan bahwa analisis sentiment terhadap korps Brimob Polri dilakukan dengan baik menggunakan SVM model, kesuksesan dari model ini dipengaruhi oleh beberapa hal seperti komposisi antara data *training* dan data *testing*, jumlah data yang digunakan, serta komposisi jumlah data positif dan negatif.

Penelitian dari Raisa Sianipar *et al* [16] telah melakukan analisis sentimen terhadap pendekatan kekuatan suatu teks tweet yang berbahasa Indonesia dengan menggunakan SentiStrength. Tweet atau data yang dikumpulkan berupa tweet yang berisi dari enam akun *unofficial fanbase* yang membahas kehidupan mahasiswa, yaitu @yeahmahasiswa @gagalmahasiswa, @ngampusaja, @skrip_shit, @deritamahasiswa, dan @maulasam. Masing-masing dari akun diambil sebanyak 100 tweets secara acak, sehingga keseluruhan tweet berjumlah 600 tweets. Tweet data tersebut akan dilakukan pelabelan terhadap data tweet dengan memberdayakan sejumlah *human coder* untuk memberikan nilai *positive* dan *negative* emotion rating pada teks. Hasil yang didapatkan dari 600 teks ada sebanyak 344 teks yang berhasil sama yang bisa dikeluarkan sistem terhadap hasil pelabelan dan 256 teks yang tidak berhasil sama yang bisa dikeluarkan sistem terhadap hasil pelabelan. Dimana hasil akurasi yang didapatkan oleh SentiStrength sebesar 57,33%.

Penelitian dari Roque López *et al* [26] melakukan perbandingan pengecekan klasifikasi kata bahasan Spayol dengan bantuan *SentiStrength* dan *Classifier*. Dimana data yang diambil merupakan komentar berjumlah 282 kata baik dari platform Twitter dan Facebook. Hasil yang didapatkan dimana hasil dengan *Classifier* dan *SentiStrength* mendapatkan akurasi sebesar 72% dimana data dari 282 menunjukkan bahwa komentar dari *Facebook* mengguli dari pada data komentar dari Twitter. Prediksi dari Facebook mendapatkan nilai positif sebesar 85,37%, negatif sebesar 62,34% dan netral sebesar 26,83%.

Dari penelitian di atas terdapat kekurangan dan kelebihan. Kekurangan dan kelebihan tersebut dapat dirangkum dan ditambahkan pada penelitian ini. Penulis akan melakukan analisis sentimen terhadap penyebaran Covid-19 yang terus meningkat setelah vaksin disebar, penyelenggaraan vaksin Covid-19 dan vaksin Covid-19 yang diinginkan dan tidak diinginkan oleh masyarakat. Data tweet akan diambil dengan menggunakan RapidMiner dengan kata kunci Covid-19, vaksin Covid-19, vaksin *Sinovac* dan vaksin *AstraZeneca*. Data ini menghasilkan opini masyarakat berupa positif, negatif dan Netral. Lalu data ini akan di *training* dengan bantuan *GoogleColab* dengan menggunakan metode *Multinomial Naïve Bayes* (MNB), *Support Vector Machine* (SVM), *Multinomial Naïve Bayes* (MNB) *feature extraction SentiStrength*, dan *Support Vector Machine* (SVM) *feature extraction SentiStrength*. Hasil akhir yang didapatkan penulis adalah keempat metode tersebut apakah metode tersebut dapat digunakan dalam melakukan analisis tersebut. Dan peneliti mendapatkan metode mana yang paling cocok untuk analisis tersebut. Hasil dari metode ini akan dibandingkan dengan bantuan *Accuracy* dari *Confusion Matrix* yang menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar dari penghitungan *Confusion Matrix*.

Table 2. 1. Perbandingan dengan penelitian yang telah dilakukan

No.	Pembandingan	Algoritma	Pokok Penelitian	Klasifikasi Pada	Hasil
1.	[19]	Naïve Bayes	Sentimen Positif, Negatif dan Netral	Data Twitter	Dengan bantuan Rapid Miner yang memiliki Library Naïve Bayes menghasilkan sentimen masyarakat terhadap vaksin Covid-19 adalah Negatif.
2.	[20]	Naïve Bayes	Sentimen Positif, Negatif dan Netral Serta Nilai Akurasi dari algoritma	Data Twitter	Dengan menggunakan metode Naïve Bayes memperoleh akurasi F1-score sebesar 89%. Untuk lebih lengkapnya akurasi sebesar 89%, presisi sebesar 92%, dan <i>recall</i> sebesar 86%.
3.	[21]	Naïve Bayes, SVM, Max Entropy, Decision Tree, Boosting, dan Random Forest	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Dari beberapa metode yang digunakan LogiBoots berkinerja lebih baik dengan akurasi 74% dari pada metode lainnya.

No.	Pembandingan	Algoritma	Pokok Penelitian	Klasifikasi Pada	Hasil
4.	[22]	Multinomial Naïve Bayes, Support Vector Machine dan Decision tree	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Dengan melakukan perbandingan dengan metode Multinomial Naïve Bayes, Support Vector Machine dan Decision tree menghasilkan bahwa Decision Tree menunjukkan nilai terbaik dengan akurasi 98,5%, presisi 99%, <i>recall</i> 99% dan Skor <i>F1-Score</i> 99%.
5.	[23]	Support Vector Machine, Multinomial Naïve Bayes dan Normalized Poly Kernel	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Hasil penelitian menunjukkan SVM dengan Normalized Poly Kernel mengungguli MNB baik pada data aspek. SVM memperoleh hasil akurasi rata-rata, presisi, <i>recall</i> , dan <i>f</i> ukuran masing-masing dengan nilai 82.00%, 82.24%, 82.01%, dan 81.84%.

No.	Pembandingan	Algoritma	Pokok Penelitian	Klasifikasi Pada	Hasil
6.	[24]	Naïve Bayes	Nilai Akurasi dari algoritma	Data Twitter	Dengan metode Naïve Bayes menunjukkan jumlah korban netral dengan polaritas nilai 58% dan sebagian besar pencatatan bersifat objektif yaitu 62,27%.
7.	[25]	Naïve Bayes Dan Support Vector Machine	Nilai Akurasi dari algoritma	Data Twitter	Hasil penelitian menunjukkan SVM mengungguli model NB, dengan akurasi 86.96%, <i>precision</i> 91.06%, dan 83.1% <i>recall</i> .
8.	[16]	Classifier SentiStrength	Nilai Akurasi dari algoritma	Data Twitter	Hasil penelitian menunjukkan dengan data sebanyak 600 jumlah pendekatan teks dengan bantuan <i>SentiStrength</i> mendapatkan akurasi sebesar 57,33%

No.	Pembandingan	Algoritma	Pokok Penelitian	Klasifikasi Pada	Hasil
9.	[26]	Classifier SentiStrength	Nilai Akurasi dari algoritma	Data Komentar Twitter dan Komentar Facebook	Hasil penelitian menunjukkan bahwa komentar Facebook lebih unggul dari Twitter dengan nilai positif sebesar 85,37 % dan mendapatkan nilai akurasi Classifier SentiStrength sebesar 72%
10.	Damanik (2022) *	Multinomial Naïve Bayes, Support Vector Machine, Multinomial Naïve Bayes SentiStrength, dan Support Vector Machine SentiStrength	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Dengan menggunakan keempat diharapkan dapat mengetahui metode mana yang menghasilkan nilai tertinggi dari metode tersebut.

* Penelitian Yang Dilakukan