

## **BAB II**

### **TINJAUAN PUSTAKA DAN LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

Dalam eksplorasi literatur, ditemukan penelitian dengan judul "*Sentiment Analysis of Tweets in Saudi Arabia Regarding Governmental Preventive Measures to Contain COVID-19*" [8] memberikan kontribusi signifikan untuk memahami respons masyarakat Saudi Arabia terhadap langkah-langkah pencegahan Covid-19. Penelitian ini dilaksanakan dengan menggunakan metode analisis sentimen, di mana data diperoleh dari media sosial Twitter.

Dalam penggunaan 11.000 data, penelitian ini berhasil mencapai F1-score sebesar 89%, hal ini mencerminkan tingkat akurasi yang memuaskan. Namun, peneliti menunjukkan sikap kritis dengan tidak puas terhadap pencapaian tersebut dan melanjutkan penelitian dengan modifikasi kata kunci. Modifikasi ini diimplementasikan untuk meningkatkan tingkat akurasi dan mendalami pemahaman atas sentimen masyarakat.

Proses modifikasi kata kunci diartikan sebagai upaya peningkatan kualitas analisis sentimen dengan penyesuaian strategis untuk memperoleh wawasan yang lebih mendalam. Dengan penekanan pada modifikasi kata kunci, penelitian ini menonjolkan keuletan peneliti dalam merinci lapisan yang lebih dalam dari data sosial media.

Sebagai pembaca, kita tidak hanya disajikan dengan temuan-temuan signifikan, tetapi juga diundang untuk ikut serta dalam perjalanan eksploratif yang mendalam. Penelitian ini bukan sekadar laporan angka, melainkan narasi yang menciptakan pemahaman yang lebih dalam terhadap kompleksitas pandangan masyarakat. Penelitian ini memberikan kontribusi penting dalam memperkaya literatur mengenai analisis sentimen.

Kemudian dalam merinci perkembangan analisis sentimen selama pandemi COVID-19, penelitian yang menonjol adalah karya Akib Mohi Ud Din Khanday dan rekannya yang berjudul “ *Identifying Propaganda from Online Social Networks during COVID-19 using Machine Learning Techniques*” [9]. Penelitian ini memusatkan perhatian pada analisis sentimen dalam cuitan Twitter dengan menerapkan tiga metode analitik, yaitu MNB (*Multinomial Naïve Bayes*), SVM (*Support Vector Machine*), dan *Decision Tree*.

Dalam kajian ini, *Decision Tree* menunjukkan hasil luar biasa dengan mencapai akurasi sebesar 98,5%, presisi dan *recall* mencapai 99%, dan *F1-Score* mencapai 99%. Keberhasilan metode *Decision Tree* ini memunculkan sorotan atas potensinya dalam mengidentifikasi propaganda dalam cuitan di platform Twitter selama periode pandemi.

Penggunaan *Decision Tree* bukan hanya mencerminkan keunggulan teknisnya, tetapi juga sebagai alat analisis yang memahami kompleksitas sentimen manusia dalam konteks informasi yang melimpah dan dinamis. *Decision Tree* bukan hanya sebuah algoritma, tetapi senantiasa menjadi pionir dalam meretas kompleksitas analisis sentimen dengan memberikan pengetahuan yang lebih mendalam.

Penelitian ini bukan sekadar penyajian data dan fakta, tetapi merupakan narasi yang mengundang pembaca untuk meresapi keunggulan metode analisis sentimen yang tidak hanya berbasis angka, melainkan juga terwujud dalam kemampuan mendalam dalam memahami dan merespons sentimen masyarakat di ranah maya.

Penelitian lainnya datang dari Nepal oleh B. Pokharel yang memandu kita melalui perjalanan analitik dengan tergambar didalam penelitiannya yang berjudul “*Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal*”[10]. Seakan seorang arkeolog digital, Pokharel memanfaatkan metode *Naïve Bayes* sebagai instrumen utama untuk menggali informasi dari serangkaian tweet selama periode wabah Covid-19.

Penelitian Pokharel mengungkapkan bahwa sebagian besar tweet yang terkait dengan kata kunci virus corona dan Covid-19 menunjukkan polaritas netral, mencapai puncak pada angka 58%. Layaknya sebuah kanvas seni abstrak, catatan dalam tweet juga bersifat objektif mencapai proporsi signifikan sebesar 62,27%.

Penelitian ini bukan sekadar kumpulan data, melainkan sebuah karya seni analitis yang membedah dinamika sentimen masyarakat Nepal selama pandemi, tercermin melalui corak tweet di dunia maya. Penggunaan metode *Naïve Bayes* seperti sapuan kuas seorang seniman, menggambarkan nuansa sentimen dengan ketajaman dan presisi.

Sebagai seorang peneliti, B. Pokharel tidak sekadar menyajikan data, tetapi menciptakan sebuah karya orisinal yang memberikan pencerahan dalam memahami kompleksitas respons masyarakat terhadap keadaan darurat. Dengan menempatkan fokus pada media sosial Twitter, penelitian ini tidak hanya menganalisis kata-kata, melainkan mencitrakan sebuah lukisan kolektif dari pemikiran dan perasaan masyarakat Nepal selama periode sulit ini.

Lebih lanjut Pristiyono dan rekan-rekannya menyajikan penelitian berjudul "*Sentiment Analysis of COVID-19 Vaccine in Indonesia using Naïve Bayes Algorithm*" [11] sebagai kontribusi signifikan dalam memahami pandangan masyarakat Indonesia terhadap vaksin Covid-19. Penelitian ini memfokuskan pada analisis sentimen sebagai bentuk eksplorasi mendalam terhadap respons kolektif terhadap program vaksinasi.

Penelitian ini memilih metode *Naïve Bayes* sebagai instrumen utama. Penggunaan metode ini tidak hanya sebagai alat analisis, tetapi juga sebagai pensil yang menggoreskan peta orisinalitas penelitian. Data tweet yang dihimpun selama periode 15-22 Januari 2021 diuraikan menjadi kisah kompleks sentimen masyarakat. Sebanyak 3,4 ribu tweet menyuarakan sentimen negatif sebesar 56%, lebih dari 2,4 ribu tweet menghadirkan nuansa positif sebesar 39%, sementara 301 tweet (1%) merangkum opini netral.

Penelitian ini menjadi tidak sekadar telaah data, melainkan peta eksplorasi emosional yang menelusuri dimensi perasaan dan persepsi masyarakat. Pilihan metode *Naïve Bayes* memberikan kedalaman analisis yang memperkaya pemahaman terhadap kompleksitas pandangan publik terhadap vaksinasi Covid-19 di Indonesia.

Selanjutnya dalam penelitian yang berjudul "Analisis Sentimen dengan SVM, *Naïve Bayes*, dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter" [12], berfokus pada pemahaman mendalam terhadap respons dan perasaan masyarakat Indonesia melalui platform media sosial Twitter. Penelitian ini menggambarkan suatu penjelajahan yang cermat terhadap algoritma klasifikasi, yaitu *Support Vector Machine (SVM)*, *Naive Bayes*, dan *K-Nearest Neighbor (KNN)* dalam merinci kompleksitas tanggapan masyarakat.

*Support Vector Machine (SVM)* muncul sebagai bintang utama, menghasilkan tingkat akurasi rata-rata sebesar 90,01% dengan penggunaan *kernel linear*. *Naive Bayes* menunjukkan kekuatannya dengan akurasi rata-rata 79,20%, sementara *K-Nearest Neighbor (KNN)*, dengan nilai K optimal menyajikan pemahaman unik dengan akurasi 62,10%.

Penelitian ini bukan sekadar sekumpulan data dan hasil analisis, tetapi sebuah narasi yang membuka jendela ke kompleksitas perasaan masyarakat. SVM, *Naive Bayes*, dan KNN seolah menjadi pencerita dalam melukis peta sentimen yang memandang dalam dari respons masyarakat terhadap pandemi. Evaluasi melalui *10-Fold Cross Validation* melengkapi dan memberikan fondasi yang kuat untuk mendukung kesimpulan penelitian ini.

Ditemukan juga penelitian yang berjudul "Analisis Sentimen dengan SVM, *NAIVE BAYES* dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter" [13], penelitian ini memperlihatkan esensi lebih dari sekadar angka. Analisis awal menunjukkan bahwa *Naive Bayes* berhasil mencapai tingkat akurasi sebesar 78,02% dengan *Area Under the Curve* (AUC) 0,714. Sementara itu, SVM menonjol dengan tingkat akurasi yang lebih tinggi, mencapai 80,23%, dan AUC 0,904, menunjukkan perbedaan signifikan sebesar 2,21%.

Optimasi menggunakan *Particle Swarm Optimization* (PSO) memberikan perubahan yang signifikan pada hasil. *Naive Bayes* (PSO) mengalami peningkatan akurasi menjadi 79,07%, dengan AUC mencapai 0,729. Di sisi lain, SVM (PSO) memuncaki hasil dengan akurasi mencengangkan sebesar 81,16%, dan AUC 0,903, mengungkap selisih akurasi yang tetap mengagumkan sebesar 2,09%.

Lebih dari sekadar pemrosesan angka, penelitian ini mengeksplorasi sentimen yang tertanam dalam setiap komentar. Ini bukan hanya perbandingan metode klasifikasi, melainkan pemahaman mendalam tentang kerumitan pandangan masyarakat. SVM, sebagai pilihan utama tidak hanya menawarkan keunggulan angka tetapi juga kepekaan uniknya dalam merespons dan menggambarkan nuansa sentimen. Dengan pendekatan ini, penelitian ini membawa angka ke dalam wacana yang hidup, memberikan pandangan orisinal dan mendalam pada respons masyarakat Indonesia terhadap pandemi Covid-19.

Kemudian dalam rangkaian penelitian mengenai "*Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel*" [14], fokus utama penelitian ini adalah merinci tanggapan masyarakat terhadap upaya penanganan pandemi Covid-19 oleh pemerintah Indonesia. Melalui penerapan metode inovatif seperti SVM (*Support Vector Machine*) dengan *Normalized Poly Kernel*, penelitian ini telah berhasil merancang pendekatan analisis sentimen yang tidak hanya canggih tetapi juga orisinal.

Penelitian ini bukanlah sekadar kumpulan data dan algoritma, sebaliknya ini merupakan sebuah narasi baru dalam pemahaman kompleksitas respons masyarakat terhadap kebijakan pemerintah. SVM dengan *Normalized Poly Kernel* menjadi elemen kunci dalam menganalisis nuansa yang rumit dari dua kelas data aspek umum, memberikan wawasan mendalam tentang pemikiran dan perasaan masyarakat.

Signifikansi penelitian ini tidak hanya terletak pada keunggulan metodenya, tetapi juga pada kemampuannya menghadirkan gambaran yang jelas dan terperinci tentang sentimen masyarakat. Dengan akurasi rata-rata mencapai 82.00%, presisi 82.24%, *recall* 82.01%, dan nilai F1 sebesar 81.84%, penelitian ini tidak hanya menyajikan data, tetapi juga menggambarkan sebuah kisah. Kisah tentang bagaimana pandemi dipandang, dirasakan, dan direspons oleh masyarakat Indonesia.

Penelitian ini memberikan nilai tambah dengan menyuguhkan pandangan manusiawi terhadap data. Dengan menggabungkan kecanggihan metode analisis sentimen dengan kepekaan terhadap realitas sosial, penelitian ini mewakili langkah maju dalam memahami interaksi yang kompleks antara masyarakat dan kebijakan pemerintah saat menghadapi krisis global.

Selanjutnya dalam penelitian yang berjudul "*COVID-19 Outbreak: Tweet-based Analysis and Visualization towards the Influence of Coronavirus in the World*" [15], peneliti secara teliti dan cermat menjalankan analisis sentimen melalui berbagai metode pembelajaran mesin seperti SVM, *Max Entropy*, *Decision Tree*, *Boosting*, dan *Random Forest*. Metode ini dipilih dengan seksama untuk merespons tantangan unik yang ditawarkan oleh data tweet dari platform Twitter.

Penting untuk mencatat bahwa penelitian ini bukan sekadar pengumpulan data dan analisis angka, melainkan suatu persembahan intelektual yang memadukan kecerdasan buatan dengan pendekatan analitis yang mendalam. Melalui pemilihan metode *LogitBoost* sebagai pendekatan campuran, peneliti berhasil mencapai tingkat akurasi sebesar 74%, menciptakan suatu karya penelitian yang tidak hanya informatif namun juga memiliki nilai seni intelektual.

Pendekatan unik dalam penelitian ini tidak hanya memperkaya pemahaman kita tentang dampak global pandemi, tetapi juga menghadirkan dimensi artistik dalam memahami cara masyarakat merespons peristiwa luar biasa ini melalui media sosial. Data dari Twitter diolah sepenuhnya tidak hanya sebagai entitas angka, melainkan sebagai palet yang menampilkan ragam warna respons dan opini global.

Dengan merangkul kecerdasan buatan dan menunjukkan kreativitas analitis, penelitian ini memainkan peran krusial dalam membongkar lapisan-lapisan kompleksitas pandemi di era digital. Sebagai sebuah karya seni intelektual, penelitian ini bukan hanya sekadar memberikan wawasan, melainkan membawa kita pada pemahaman mendalam tentang bagaimana masyarakat global secara kolektif merespons dan menginterpretasikan peristiwa luar biasa ini.

Penelitian ini bersifat lebih dari sekadar analisis data dari tweet berbahasa Indonesia. Dengan merujuk pada kekuatan analisis *SentiStrength* [16], penelitian ini merintis pendekatan yang lebih mendalam dalam memahami kompleksitas dan variasi emosi yang terkandung dalam setiap kalimat. Dengan memfokuskan perhatian pada keenam akun fanbase yang membahas kehidupan mahasiswa secara acak, penelitian ini menghadirkan suatu kerangka kerja yang mencerminkan beragam pengalaman dan perspektif.

Tak hanya sebatas analisis sentimen, penelitian ini menciptakan suatu perjalanan eksploratif melalui keragaman pemikiran dan perasaan, mengajak pembaca untuk terlibat dalam realitas kehidupan sehari-hari yang kaya dan beraneka ragam.

Dengan memanfaatkan kebijaksanaan manusia, penelitian ini memperoleh dimensi tambahan yang tak ternilai dengan melibatkan human coder dalam memberikan penilaian terhadap aspek emosi positif dan negatif. Pendekatan ini bukan hanya merespon keputusan algoritma, melainkan juga menyajikan wawasan yang lebih luas dan holistik.

Dalam konteks pengambilan keputusan yang bijaksana, penelitian ini tidak sepenuhnya bergantung pada otomasi. Keterlibatan manusia dalam memberikan penilaian emosi menambah nilai signifikan pada hasil analisis, membawa perspektif manusia ke dalam dunia digital.

Dengan keberhasilan sistem dalam mencocokkan penilaian emosi pada 344 dari 600 teks, penelitian ini membawa pemahaman baru tentang bagaimana teknologi dan manusia dapat berkolaborasi. Ini bukan sekadar analisis sentimen, melainkan eksplorasi mendalam tentang peran manusia dalam mengartikan dan memahami emosi di era media sosial berbahasa Indonesia.

Dengan demikian, penelitian ini melewati batas-batas analisis sentimen konvensional, menjadi pelopor dalam menggali kompleksitas bahasa dan ekspresi emosional di ranah media sosial berbahasa Indonesia.

Penelitian ini menjelajahi kerumitan analisis sentimen dalam konteks bahasa Spanyol dengan pendekatan holistik dan inovatif. Pemahaman mendalam terhadap respons pengguna diperoleh melalui perbandingan kinerja dua alat analisis sentimen, yaitu *SentiStrength* dan *Classifier* [17], yang digunakan untuk menganalisis 282 komentar dalam bahasa Spanyol dari platform Twitter dan Facebook.

Dengan mencapai tingkat akurasi sebesar 72%, penelitian ini menunjukkan bahwa keduanya, *SentiStrength* dan *Classifier* secara konsisten mampu mengklasifikasikan ekspresi emosional dengan efektif. Penelitian ini lebih jauh lagi dengan mengungkap perbedaan signifikan dalam hasil analisis sentimen antara kedua platform media sosial tersebut. Komentar dari Facebook, dengan prediksi sentimen positif sebesar 85,37%, sentimen negatif sebesar 62,34%, dan sentimen netral sebesar 26,83%, menunjukkan tingkat keberhasilan yang lebih tinggi dibandingkan dengan komentar dari Twitter.

Pendekatan ini tidak hanya memberikan temuan konvensional tentang analisis sentimen, tetapi juga membuka wawasan baru. Kesadaran bahwa respon terhadap konten berbahasa Spanyol dapat bervariasi secara signifikan tergantung pada platformnya memberikan kontribusi penting terhadap pemahaman kontekstual dalam analisis sentimen. Temuan ini memberikan dasar yang kuat untuk penelitian lanjutan dalam bidang analisis sentimen, menunjukkan arah yang menarik untuk eksplorasi lebih lanjut.

Dari penelitian-penelitian sebelumnya, dapat ditarik suatu benang merah bahwa metode SVM mendapatkan akurasi lebih tinggi dari metode lainnya. Penelitian yang sebelumnya merupakan sentiment analisis yang dilakukan saat dilaksanakannya pemberian vaksin dosis kedua dengan mengambil data berkaitan akan hal tersebut. Hal ini yang membuat penulis ingin melakukan pengembangan mengenai analisis sentimen mengenai vaksin booster di Indonesia menggunakan metode SVM dengan di kombinasikan dengan *Sentistrength*. Hasil yang akan didapatkan berupa sentimen masyarakat Indonesia perihal pemberlakuan vaksinasi booster kepada masyarakat secara masif dan inklusif.

**Tabel 2. 1. Studi Terdahulu**

<b>No.</b>	<b>Pembanding</b>	<b>Algoritma</b>	<b>Pokok Penelitian</b>	<b>Klasifikasi</b>	<b>Hasil</b>
1.	[8]	Naïve Bayes	Sentimen Positif, Negatif dan Netral Serta Nilai Akurasi dari algoritma	Data Twitter	Menggunakan pendekatan Naïve Bayes, hasil pengukuran akurasi mencapai 89% dalam F1-score, dengan rincian akurasi 89%, presisi 92%, dan recall 86%
2.	[9]	Multinomial Naïve Bayes, Support Vector Machine dan Decision tree	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Setelah membandingkan Multinomial Naïve Bayes, Support Vector Machine, dan Decision Tree, ditemukan bahwa Decision Tree memberikan hasil terbaik dengan tingkat akurasi sebesar 98,5%. Selain itu, Decision Tree juga menunjukkan presisi sebesar 99%, recall 99%, dan mencapai nilai F1-Score sebesar 99%.
3.	[10]	Naïve Bayes	Nilai Akurasi dari algoritma	Data Twitter	Penggunaan metode Naïve Bayes dalam analisis menunjukkan bahwa sekitar 58% dari catatan terkait kata kunci covid-19 bersifat netral, dan sebagian besar dari catatan tersebut bersifat objektif, sekitar 62,27%.
4.	[11]	Naïve Bayes	Sentimen Positif, Negatif dan Netral	Data Twitter	Pemanfaatan library Naïve Bayes dalam Rapid Miner menghasilkan kesimpulan bahwa sentimen masyarakat terhadap vaksin Covid-19 adalah negatif..

5	[12]	SVM, NAIVE BAYES dan KNN	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Jika melihat hasil evaluasi model dengan 10-Fold Cross Validation, dapat disimpulkan bahwa algoritma SVM memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan Naive Bayes dan KNN. Akurasi rata-rata yang dicapai adalah sekitar 90,01% untuk SVM dengan kernel linear, 79,20% untuk Naive Bayes dengan laplace 1, dan 62,10% untuk KNN dengan K 20 dan menggunakan kernel optimal.
6	[13]	Support Vector Machine dan Naïve Bayes	Sentimen Positif, Negatif dan Netral	Instagram	Dalam proses klasifikasi, metode Naïve Bayes berhasil mencapai akurasi sekitar 78,02% dengan AUC 0,714. Sementara itu, Support Vector Machine (SVM) mencapai akurasi lebih tinggi, yaitu sekitar 80,23% dengan AUC 0,904. Perbedaan akurasi antara keduanya adalah sekitar 2,21%. Namun, setelah melalui proses optimasi menggunakan operator Particle Swarm Optimization (PSO), akurasi Naïve Bayes (PSO) meningkat menjadi sekitar 79,07% dengan AUC 0,729, sedangkan SVM (PSO) mencapai akurasi sekitar 81,16% dengan AUC 0,903. Dengan optimasi, perbedaan akurasi antara keduanya menjadi sekitar 2,09%. Hasil pengujian menunjukkan bahwa SVM,

					baik dengan atau tanpa optimasi PSO, selalu memiliki akurasi yang lebih tinggi.
7	[14]	Support Vector Machine, Multinomial Naïve Bayes dan Normalized Poly Kernel	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	Hasil penelitian mengindikasikan bahwa SVM dengan Normalized Poly Kernel berhasil mengungguli Multinomial Naïve Bayes (MNB), terutama dalam konteks data yang berfokus pada berbagai aspek. SVM menunjukkan performa terbaik pada data yang berhubungan dengan aspek umum, terutama ketika menggunakan dataset dengan dua kelas, dengan mencapai akurasi rata-rata sekitar 82.00%. Selain itu, SVM juga memiliki presisi, recall, dan nilai f-measure yang tinggi, yaitu sekitar 82.24%, 82.01%, dan 81.84% masing-masingnya.
8	[15]	Naïve Bayes, SVM, Max Entropy, Decision Tree, Boosting, dan Random Forest	Nilai Akurasi dari algoritma untuk membandingkan metode mana yang terbaik	Data Twitter	LogiBoots menunjukkan kinerja yang lebih unggul dibandingkan dengan berbagai metode lainnya, dengan tingkat akurasi sebesar 74%.

9	[16]	Classifier SentiStrength	Nilai Akurasi dari algoritma	Data Twitter	Hasil penelitian menunjukkan bahwa dari 600 pendekatan teks, penggunaan SentiStrength menghasilkan tingkat akurasi sebesar 57,33%.
10	[17]	Classifier SentiStrength	Nilai Akurasi dari algoritma	Data Komentar Twitter dan Komentar Facebook	Hasil penelitian mengindikasikan bahwa komentar dari Facebook memiliki keunggulan, mencapai tingkat positif sebesar 85,37%, dengan akurasi klasifikasi SentiStrength sebesar 72%.