ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN PENERAPAN PPKM DI MEDIA SOSIAL TWITTER DENGAN MENGGUNAKAN METODE XGBOOST

(*Analysis Of Community Sentiment On The Policy Of Implementation Of PPKM On Twitter Social Media Using Xgboost Method*)

I Putu Angga Purnama Widiarta, Ramaditia Dwiyansaputra, Arik Aranta

[1]Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

*Email:* anggawidiarta55@gmail.com, [rama, arikaranta]@unram.ac.id

***Abstract***

*Abstrak dan kata kunci ditulis dalam bahasa inggris dengan* ***format Italic*** dan ***font Calibri 10*** *yang memuat ringkasan singkat dari makalah sehingga membantu pembaca dalam mengetahui tujuan, metode, hasil dan kesimpulan dari artikel. Abstrak harus singkat, padat dan jelas serta informatif. Panjang abstrak tidak lebih dari 200 kata. Abstrak tidak boleh memuat singkatan dan simbol yang belum didefinisikan. Menggunakan kata kunci (keywords) yang terkait dengan topik penelitian agar artikel mudah untuk di indeks. Kata kunci minimum 5 kata yang berkaitan dengan topik artikel dan dipisahkan dengan tanda koma (,).*

***Keywords:*** *Sistem Cerdas, Jaringan Syaraf Tiruan, Citra Sidik Jari*

# Pendahuluan

*Corona Virus Disease* (Covid-19) merupakan virus yang dapat menyebabkan infeksi saluran pernafasan pada manusia, mulai dari flu biasa hingga penyakit yang lebih serius seperti *Severe Acute Respiratory Syndrome* (SARS) dan *Middle East Respiratory Syndrome* (MERS). Wabah ini bermula di Wuhan, Provinsi Hubei, China.

Indonesia merupakan salah satu negara yang terjangkit oleh virus *covid-19*, oleh karena itu pemerintah Indonesia menerapkan kebijakan pemberlakuan pembatasan kegiatan masyarakat (PPKM) per tanggal 11 Januari 2021 demi menekan angka persebaran dari virus ini. Kebijakan tersebut berdampak signifikan pada berbagai sektor kehidupan masyarakat, salah satunya adalah dari sektor ekonomi. Pembatasan kemampuan untuk melakukan kegiatan skala besar akan mengakibatkan perekonomian menjadi semakin sulit.

Pro dan kontra bermunculan di kalangan masyarakat, khususnya pada media sosial. *Twitter* merupakan salah satu *platform* yang sering digunakan selama periode PPKM. Dilansir dari *Global Digital Statistic* ”Digital, Social & Mobile in 2019”, pada tahun 2019 pengguna media sosial di Indonesia berjumlah 150 juta pengguna, twitter menyumbang angka terbanyak yang mencakup lebih dari 52% [1]. Opini publik bersifat tidak dibatasi dan bebas di *twitter* [2]. Dalam PPKM, opini publik dinyatakan dalam bentuk reaksi positif, negatif, atau netral terhadap pemerintah. Analisis sentimen dilakukan untuk mengolah opini – opini tersebut dan memperoleh inferensi tekstual dari isi benak masyrakat Indonesia.

Analisis sentimen adalah metode untuk memahami, menganalisis, dan memproses *input* tekstual secara otomatis untuk memperoleh informasi sentimen dari suatu opini [3]. *Machine learning* merupakan metode pendekatan yang menonjol untuk memproses sentimen. Penelitian mengenai sentimen analisis dengan menggunakan pendekatan *machine learning* terkait dengan kebijakan PPKM sebelumnya dengan metode *Support Vector Machine* (SVM) sudah dilakukan dengan nilai akurasi yang diperoleh sebesar 64% [4].

Berdasarkan pada hal - hal yang telah dipaparkan pada paragraf sebelumnya. Pada penelitian ini, akan dilakukan implementasi penggunaan dari metode XGBoost sebagai algoritma klasifikasi, dan mengimplementasikan metode *TF-RF* (*Term Frequency – Relevance Frequency*) sebagai metode untuk menentukan bobot dari suatu term pada teks. XGBoost menghemat waktu, mengoptimalkan sumber daya memori, dan dapat diterapkan secara paralel selama proses implementasi untuk mengelola sentimen.

# Tinjauan Pustaka

Bryan Pratama, dkk. pada tahun 2019 melakukan studi analisis dengan judul “Sentiment Analysis Of The Indonesian Police Mobile Brigade Corps Based On Twitter Posts Using The SVM And NB Methods” pada studi tersebut dilakukan analisa pada *tweet* dengan kata kunci “Brimob” dimana total *tweet* yang digunakan sebanyak 1000 *tweets*. Studi ini menggunakan *text mining* dengan didukung oleh *support vector machine* (SVM) untuk mengklasifikasikan sentimen publik terhadap brimob di *twitter*. Akurasi yang diperoleh dengan SVM mencapai 86,96% sedangkan dengan *Naive Bayes* diperoleh akurasi sebesar 86,48% [5].

Tahun 2019, Eka dkk. melakukan studi analisis sentimen pada contoh Gojek dan Grab, menggunakan algoritma *Naive Bayes Classifier*, dan menemukan bahwa akurasi, *recall*, dan presisi metode *Naive Bayes Classifier* masing-masing adalah 72,33%, 73,95%, dan 73,24% [6]. Penelitian tersebut kemudian dilanjutkan oleh D. A. Al-Qudah et al. melakukan penelitian analitik sentimen terhadap penyedia layanan *e-payment* menggunakan algoritma yang disebut *XGBoost* dan membandingkan hasilnya dengan J84, *Naive Bayes*, dan KNN. Akurasi maksimum didapatkan oleh KNN dan *XGBoost* yang masing-masing memiliki nilai *recall* 85,2 persen dan 82,8 persen. Sedangkan dengan menggunakan nilai presisi *Naive Bayes* didapatkan akurasi tertinggi sebesar 72% [7].

ajar Fathur Rachman pada tahun 2020, dalam penelitiannya yang berjudul “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada media sosial *Twitter*” melakukan penelitian sentimen analisis dengan menggunakan algoritma *Latent Dirichlet Allocation* (LDA) untuk mengelompokkan opini masyarakat dengan tujuan mengetahui topik pembicaraan yang sering dibahas masyarakat terkait dengan wacana vaksinasi, hasil analisis menunjukkan bahwa masyarakat lebih banyak memberikan respon positif terhadap wacana tersebut (30%) dibandingkan dengan respon negatifnya (26%) [8].

Angelina Puput Giovani, dkk. Pada tahun 2020, dalam penelitian dengan judul “Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi” melakukan komparasi beberapa algoritma yaitu *Naive Bayes*, *Support Vector Machine*, dan K-*Nearest Neighbour* yang menggunakan *feature selection* dengan yang tidak menggunakan *feature selection*, serta juga membandingkan nilai Area *Under Curve* dari metode – metode tersebut untuk mengetahui algoritma mana yang paling optimal, hasil pengujian menunjukkan bahwa algoritma SVM dengan *feature selection* menjadi algoritma terbaik dengan nilai akurasi 78,55% dan AUC 0,853 [9].

Sulaiman Ainin, dkk. Pada tahun 2020, dengan penelitian berjudul “Sentiment Analyses Of Multilingual Tweets On Halal Tourism” menuliskan tentang penelitian yang mereka lakukan pada *tweet – tweet* dari rentang waktu 2008 - 2018 yang berkaitan dengan multilingual halal *tourism* dimana konten dan sentimen dari *tweet – tweet* tersebut dianalisa, mereka menggunakan 19 kata kunci untuk mengesktrak data dari *tweet* dimana 5 kata kunci tersebut adalah bahasa Malaysia, dan sisanya bahasa Inggris. Setelah dilakukan analisa diperoleh kesimpulan bahwa *tweet* terkait pariwisata halal pada negara non muslim melebihi jumlah *tweet* pada negara muslim, penelitian ini menunjukkan bahwa pariwisata halal mulai populer di negara seperti Inggris, Kanada, dan Spanyol [10].

Elena, Podasca pada tahun 2021 dengan penelitian berjudul “Predicting The Movement Direction Of OMXS30 Stock Index Using *XGBoost* and Sentiment Analysis” melakukan prediksi pada indeks pasar saham Swedia menggunakan metode *XGBoost* yang disertakan dengan sentimen analisis dari berita keuangan guna membantu meningkatkan kinerja klasifikasi ketika memprediksi tren harga harian dari indeks pasar saham Swedia yaitu OMXS30. Hasil pada penelitian ini menunjukkan bahwa *XGBoost* memiliki kinerja yang baik dalam mengklasifikasikan tren hari OMXS30 dimana akurasi yang diperoleh mencapai 73% [11].

Pada tahun 2021. Aldiansyah Putra, dkk. dalam penelitiannya berjudul “Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM” melakukan penelitian terhadap respons masyarakat di *Twitter* berupa pro dan kontra mereka kepada kebijakan pemerintah dalam pemberlakuan pembatasan kegiatan masyarakat (PPKM), metode yang digunakan pada penelitian tersebut adalah *Support Vector Machine*, dengan memanfaatkan 3000 dataset yang kemudian diperoleh hasil akurasi sebesar 64%. Dari penelitian tersebut, algoritma SVM dapat mengenali *tweet* yang berisikan penolakan PPKM sebagai *tweet* bertendensi negatif dan juga kata – kata yang memiliki hubungan terhadap tendensi negatif tersebut [4].

# Metode Penelitian

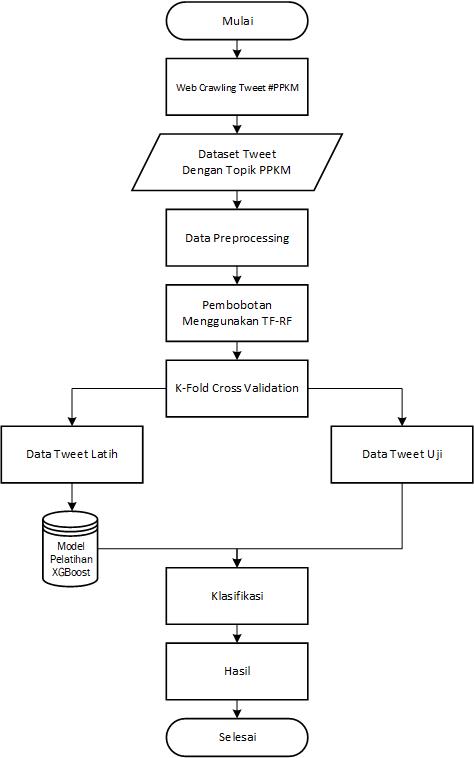
## Dataset

*Dataset* yang digunakan pada penelitian ini memiliki jumlah total sebanyak 5.938 *tweet* dengan jumlah masing-masing yaitu 1958 *tweet* positifdan 3980 *tweet* negatif.



1. Jumlah tweet pada setiap kategori

## Alur Klasifikasi Sentimen Tweet.

Analisis sentimen masyarakat Indonesia pada media sosial *twitter* terhadap kebijakan pemerintah dalam melaksanakan penerapan PPKM di Indonesia menggunakan *XGBoost* dilakukan dalam beberapa tahapan. 

1. Diagram proses klasifikasi

### Web crawling

*Tweet* dikumpulkan melalui jejaring media sosial *Twitter*. *Tweet* yang dikumpulkan adalah *tweet* yang menggunakan *hashtag* “#ppkm”. “#ppkm” kemudian dimasukkan pada *query* pencarian *tweets* yang digunakan oleh *library snscrape* untuk melakukan *crawling* data pada *tweet – tweet* berbahasa Indonesia yang memuat tagar “ppkm” di dalamnya.

### Input dataset tweet training dan testing

*Tweet* yang digolongkan sebagai data *training* digunakan untuk membuat model klasifikasi sedangkan *tweet* yang digunakan sebagai data *testing* digunakan untuk menguji model yang telah dibuat.

1. *Tweet testing* sama seperti *tweets training*,yang membedakan adalah *tweet testing* tidak diketahui jenis kategori sentimennya yang kemudian dimasukkan ke dalam sistem untuk diproses pada *training tweets*.
2. *Tweet training* adalah *tweets* yang telah dilakukan *preprocessing* pada seluruh *tweets*-nya, lalu dilanjutkan dengan melakukan *term weighting* dengan TF-RF, yang kemudian di-*training* menggunakan algoritma *XGBoost*.

### Text preprocessing tweets dataset

*Text preprocessing* yang dilakukan pada penelitian ini dibagi menjadi 4 tahap yaitu *casefolding, tokenization, stopword removal,* dan *stemming*.

1. *Casefolding* merupakan tahap dimana *dataset* yang ada akan disamaratakan penggunaan huruf kapitalnya. Seluruh *tweet* akan diubah hurufnya menjadi huruf kecil, ini bertujuan agar *tweets* menjadi konsisten pada penggunaan hurufnya, dan mencegah sistem mengalami kebingungan dikarenakan kata yang sama apabila penulisan hurufnya berbeda.
2. *Tokenization* merupakan tahap untuk mentransformasikan *tweets* menjadi kumpulan kata yang disebut *terms*. Pada *tokenization* juga dilakukan *cleaning* berupa penghilangan tanda baca. Hal ini dilakukan karena tanda baca tidak dapat digunakan sebagai *terms* karena terdapat pada hampir seluruh dokumen. Sebelum proses *tokenization*, terlebih dahulu dilakukan proses *casefolding*. Karakter selain huruf dihilangkan dan dianggap *delimiter.*
3. *Stopword removal* yaitu tahapan untuk menghapus kata – kata yang diyakini memiliki sedikit makna, bahkan hampir tidak bermakna, kata – kata tersebut seperti ‘yang’, ‘yaitu’, ‘di’, ‘tempat’, ‘terus’, ‘walau’, dan masih banyak lainnya.
4. *Stemming* merupakan tahapan untuk memperoleh kata dasar pada suatu kalimat dengan cara melakukan pemotongan pada imbuhan (*affix*) kata, baik itu *prefix* maupun *suffix* [12]. Proses *stemming* dilakukan dengan menggunakan algoritma Nazief dan Adriani.

### Feature weighting

Penelitian ini menggunakan metode *Term Frequency - Relevance Frequency* (TF-RF)dalam melakukan pembobotan *term*, dimana prosesnya akan dibagi menjadi tahapan -tahapan berikut ini :

1. *Term Frequency* (TF),yaitu tahapan untuk menghapus kata – kata yang diyakini memiliki sedikit makna, bahkan hampir tidak bermakna, kata – kata tersebut seperti ‘yang’, ‘yaitu’, ‘di’, ‘tempat’, ‘terus’, ‘walau’, dan masih banyak lainnya.

(1)

1. *Relevance Frequency* (RF),yaitu tahapan untuk menghapus kata – kata yang diyakini memiliki sedikit makna, bahkan hampir tidak bermakna, kata – kata tersebut seperti ‘yang’, ‘yaitu’, ‘di’, ‘tempat’, ‘terus’, ‘walau’, dan masih banyak lainnya.

(2)

1. *Relevance Frequency* (RF),yaitu tahapan untuk menghapus kata – kata yang diyakini memiliki sedikit makna, bahkan hampir tidak bermakna, kata – kata tersebut seperti ‘yang’, ‘yaitu’, ‘di’, ‘tempat’, ‘terus’, ‘walau’, dan masih banyak lainnya.

### Klasifikasi dengan XGBoost

Penelitian ini menggunakan metode *Term Frequency - Relevance Frequency* (TF-RF)dalam melakukan pembobotan *term*, dimana prosesnya akan dibagi menjadi tahapan -tahapan berikut ini :

# Hasil dan Pembahasan

Bagian ini berisi hasil dan pembahasan penelitian. Hasil penelitian disajikan bentuk tabel atau grafik yang selanjutnya diberikan deskripsi dan pembahasan atas fakta yang diperoleh dikaitkan teori pendukung penelitian dan atau dibandingkan dengan hasil penelitian yang sangat terkait lainnya.

# Kesimpulan dan Saran

## Kesimpulan

Pada penelitian yang telah dilakukan, diperoleh hasil penelitian yang dapat ditarik kesimpulannya bahwa implementasi metode *XGBoost* dalam melakukan klasifikasi sentimen pada *tweets* dengan menggunakan *dataset* yang telah diaugmentasi dan telah melewati proses *stemming*, menghasilkan nilai akurasi sebesar 85.27%, presisi sebesar 86.07%, dan *recall* sebesar 85.23% setelah diuji dengan menggunakan teknik *cross validation* dengan nilai *k-fold* yaitu sama dengan 10.

Kemudian pada *dataset* yang mengalami augmentasicenderung menghasilkan nilai akurasi, presisi, dan *recall* yang lebih tinggi dibandingkan dengan *dataset* yang tidak mengalami augmentasi, hal ini dikarenakan augmentasi *dataset* merujuk pada teknik penambahan variasi atau modifikasi pada *dataset* sehingga meningkatkan jumlah sampel yang tersedia yang menyebabkan model memiliki lebih banyak variasi dan informasi yang dapat dipelajari.

Sementara pada skenario pengujian yang menggunakan tahapan *stemming* cenderung menghasilkan nilai *accuracy, precision,* dan *recall*  yang lebih tinggi dibandingkan dengan pengujian tanpa tahapan *stemming*. Hal ini terjadi karena imbuhan berupa *prefiks* maupun *sufiks* pada kata yang memiliki kata dasar yang sama malah menambah variasi fitur saat melakukan pembobotan, sehingga memberikan nilai bobot yang berbeda – beda.

## Saran

Untuk penelitian selanjutnya, perlu digunakan lebih banyak *dataset* berupa *tweets* sehingga dapat memperoleh performa model yang lebih baik lagi dikarenakan banyaknya *dataset* maka dapat meningkatkan representasi data yang digunakan pada pelatihan model. Selanjutnya, pada metode pembobotan fitur yang digunakan, dapat dicoba mengimplementasikan metode pembobotan selain TF-RF. Kemudian untuk metode augmentasi yang diterapkan juga dapat dicoba dengan menggunakan metode selain dari melakukan substitusi pada kata untuk melihat bagaimana perbandingan performa pada masing – masing metode yang ada.

# Ucapan Terima Kasih

Ucapan terima kasih diberikan kepada *annotator* yang telah membantu dalam melakukan pelabelan pada *dataset*.

# Daftar Pustaka

[1] R. Yanuarti, “Jurnal Sistem dan Teknologi Informasi Analisis Media Sosial Twitter Terhadap Topik Vaksinasi Covid-19,” vol. 6, no. 2, 2021, [Daring]. Tersedia pada: http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO

[2] N. D. Asih dan M. Rosit, “Opini Publik di Media Sosial: Analisis Isi Opini Kandidat Ahok-Djarot dan Anies-Sandi di Twitter,” *CoverAge: Journal of Strategic Communication*, vol. 8, no. 2, hlm. 45–56, Mar 2018.

[3] L. Ardiani, H. Sujaini, dan T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 2, hlm. 183, Apr 2020, doi: 10.26418/justin.v8i2.36776.

[4] A. Putra, D. Haeirudin, H. Khairunnisa, dan R. Latifah, “Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma Svm,” *Prosiding Semnastek*, 2021.

[5] B. Pratama *dkk.*, “Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods,” dalam *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mei 2019. doi: 10.1088/1742-6596/1201/1/012038.

[6] E. Y. and W. A. D. and S. A. Sari, “Sentiment analysis of customer satisfaction on transportation network company using naive Bayes classifier,” *019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, hlm. 1–6, 2019.

[7] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, dan H. Faris, “Sentiment analysis for e-payment service providers using evolutionary extreme gradient boosting,” *IEEE Access*, vol. 8, hlm. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.

[8] F. F. Rachman dan S. Pramana, “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter,” *Health Information Management Journal ISSN*, vol. 8, no. 2, hlm. 2655–9129, 2020.

[9] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, vol. 14, no. 2, hlm. 115–123, Jul 2020, doi: 10.33365/JTI.V14I2.679.

[10] S. Ainin, A. Feizollah, N. B. Anuar, dan N. A. Abdullah, “Sentiment analyses of multilingual tweets on halal tourism,” *Tour Manag Perspect*, vol. 34, hlm. 100658, Apr 2020, doi: 10.1016/J.TMP.2020.100658.

[11] E. Podasca, “Predicting the Movement Direction of OMXS30 Stock Index Using XGBoost and Sentiment Analysis,” 2021, [Daring]. Tersedia pada: www.bth.se

[12] D. Wahyudi, T. Susyanto, dan D. Nugroho, “Implementasi dan analisis algoritma stemming nazief & adriani dan porter pada dokumen berbahasa indonesia,” *Jurnal Ilmiah SINUS*, vol. 15, no. 2, 2017.