Analisis Tren Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Kota Jakarta Menggunakan Algoritma Support Vector Machine

Novia Rinanti Robynson[1], Yuliant Sibaroni[2]

[1,2,3] Universitas Telkom, Bandung

[1]noviarobynson@students.telkomuniversity.ac.id, [2]yuliant@telkomuniversity.ac.id

Abstrak

Analisis tren sentimen adalah teknik untuk menganalisa suatu pendapat yang diekspresikan dalam bentuk teks dalam satuan waktu tertentu (hari, minggu atau bulan) untuk melihat tren dari sebuah topik apakah positif atau negatif berdasarkan pengelompokan sentimennya. Proses analisis tren sentimen dengan topik mengenai Pembatasan Sosial Berskala Besar kota Jakarta dilakukan dengan menggunakan metode SVM dan TF-IDF untuk pembobotan *term* serta *Grid Search* untuk *hyperparameter optimization* dan curva serta bar diagram untuk melihat tren yang dihasilkan. Setelah dilakukan pengujian dengan mengikuti skenario pengujian menunjukkan bahwa dengan menghilangkan *stopword removal* dan menggunakan semua *preprocessing* lainnya, termasuk *case folding, cleansing data, tokenization* dan *lemmatization,* akan mempengaruhi akurasi terhadap proses klasifikasi yang dilakukan dimana memberikan akurasi yang paling optimal sebesar 87,33% yang menunjukkan akurasi data yang terklasifikasi dengan benar. Setelah mendapatkan model dengan akurasi yang paling optimal, selanjutnya akan digunakan untuk klasifikasi tren pada data prediksi. Dimana diperoleh bahwa tren positif mendominasi secara keseluruhan dibandingkan jumlah negatif dimana hanya terjadi pada 4 hari dari total 19 hari secara keseluruhan. Jadi, sentimen publik terhadap kebijakan pemerintah DKI Jakarta dalam memberlakukan PSBB mendapatkan respon yang positif dari masyarakat. Walaupun pada beberapa titik (perpindahan hari) sedikit mengalami kenaikan, namun hal ini tidak terlalu mencolok atau tidak tajam dan tidak terjadi secara signifikan dibandingkan dengan tren yang mengalami penurunan. Dimana dapat terlihat bahwa tren perlahan-lahan langsung mulai menurun dari hari pertama tanggal 10 September 2020 sampai tanggal 28 September 2020. Tren mengalami peningkatan hanya pada tanggal 13, lalu sedikit naik pada tanggal 21 dan 24 sampai dengan tanggal 25 September 2020. Dimana kenaikan ini tidak terlalu tajam.

Kata kunci : SVM, TF-IDF, Analisis Sentimen, *Grid Search, Cross Validation*

Abstract

Sentiment trend analysis is a technique used to analyze an opinion expressed in the form of text within a certain time unit to see the trend of a topic whether in a positive or negative based on the grouping of sentiments. The sentiment trend analysis process with the topic of Large-Scale Social Restrictions in the city of Jakarta is carried out using the SVM and TF-IDF methods for term weighting and Grid Search for hyperparameter optimization and curve, bar diagrams to see the resulting trends. After testing by following the test scenario, it shows that by eliminating stopword removal and using all other preprocessing, including case folding, data cleansing, tokenization, and lemmatization, will affect the accuracy of the classification process carried out which provides the most optimal accuracy of 87.33% which indicates the accuracy of the data classified correctly. After getting a model with the most optimal accuracy, it will then be used for trend classification on predictive data. Where it is obtained that the positive trend dominates overall compared to the negative number which only occurs on 4 days out of a total of 19 days as a whole. So, public sentiment towards the DKI Jakarta government's policy in enforcing the PSBB received a positive response from the community. Even though at some points (day shifts) there is a slight increase, this is not too conspicuous or not sharp and does not occur significantly compared to the downward trend. Where it can be seen that the trend slowly immediately started to decline from the first day of September 10, 2020, to September 28, 2020. The trend increased only on the 13th, then slightly increased on the 21st and 24th until September 25, 2020. Where this increase is not too sharp.

Keywords: SVM, TF-IDF, *Sentimen Analysis, Grid Search, Cross Validation*

1. Pendahuluan

Perubahan situasi, kondisi serta sosial yang terjadi karena pandemic Covid-19 tentu bukanlah hal yang dapat diprediksi sebelumnya sehingga tentu mengakibatkan *culture shock* yang mengharuskan pemerintah menerapkan Pembatasan Sosial Berskala Besar (PSBB) termasuk kota Jakarta. Berbagai respon masyarakat dapat terlihat salah satunya melalui tweet di media sosial twitter. Banyaknya komentar mengenai PSBB yang ada di twitter dan bahkan sering kali menjadi trending, karena hal inilah penulis merasa perlu untuk menganalisis tren sentimen pada twitter mengenai PSBB. Namun

masalah yang muncul adalah *feature* kata yang terlalu besar membuat proses sentimen akan menjadi sangat lama (tidak efisien) dan sulit serta akurasi yang dihasilkan pun tidak akan maksimal. Diperlukan cara yang dapat mengolah data tersebut untuk mengkategorikan *tweet-tweet* tersebut baik dalam positif maupun negatif agar dapat diproses dengan cepat dan otomatis.

Tren adalah rangkaian rekam jejak terkait objek atau topik yang diteliti misalnya 'harga' yang divisualisasikan melalui grafik dimana saat condong ke atas disebut *uptrend* dan *downtrend* saat condong ke bawah, serta *sideways* yaitu tren tanpa arah yang jelas dan dalam kisaran tertentu perkembangannya sulit diidentifikasi. Analisis tren sentimen merupakan teknik yang digunakan untuk menganalisa suatu pendapat yang diekspresikan dalam bentuk teks. Misalnya review atau tweet mengenai topik tertentu dalam suatu waktu (hari, minggu atau bulan) agar dapat dilihat tren dari sebuah topik yang sedang dianalisis apakah positif atau negatif.

Semakin berkembangnya teknologi dapat membantu manusia dalam mencari informasi apapun yang dibutuhkan ataupun mencurahkan isi hatinya melalui media sosial yang tersedia [1]. Twitter adalah salah satu media sosial dengan pengguna terbanyak di Indonesia dengan lebih dari 29,5 juta dan jumlah tweet per harinya sekitar 383 juta. Melalui twiter setiap *user* dapat membagikan pendapat melalui pesan pendek yaitu tweet. Dari tweet-tweet tersebut apabila dikumpulkan akan menjadi suatu informasi yang dapat dikelola menjadi sumber data untuk sentimen masyarakat untuk membantu dalam pemasaran maupun studi sosial [2]. Proses dalam menemukan suatu pola atau pengetahuan disebut dengan data mining. Pola harus valid, berguna dandapat dipahami dalam sejumlah Langkah: pra-pemrosesan, penambangan data dan pasca-pemrosesan [3]. Pada [4] metode yang digunakan yaitu SVM (*Support Vector Machine*) dimana ektraksi dan pembobotan fiturnya dengan TF-IDF (). Dari pengelitian dapat membuktikan bahwa SVM dapat melakukan klasifikasi dengan baik dan akurasi yang didapatkan sebesar 90,55%.

Pada beberapa penelitian sebelumnya terkait dengan klasifikasi tren seperti pada penelitian oleh [11] Semuel L. dkk yang melakukan analisis tren mengenai perubahan curah hujan dan klasifikasi iklim dengan memanfaatkan data curah hujan selama 60 tahun pengamatan. Dan dari penelitian ini berhasil melakukan klasifikasi tren sesuai dengan tujuan dari analisis, dimana hasilnya menunjukkan bahwa dibandingkan dengan periode 30 tahunsebelumnya (1959-1988), dalam 30 tahun terakhir mulai dari tahun 1959 sampai dengan 1988 terjadi peningkatan curah hujan rataan yaitu sebesar 11,2 hingga 15,6%. Pada [12] mengenai dilakukan klasifikasi tren mengenai marketplace yang diklasifikasikan dari ulasan pelanggan dalam klasifikasinya dengan kernel SVM dimana berhasil melakukan klasifikasi tren dimana model klasifikasi terbaik adalah dengan kernel sigmoid dengan akurasi 92%, presisi 92%, recall dan skor F1 92%serta parameter C=100, 0,01 dan R=1. Pada [13] analisis tren penelitian tugas akhir mahasiswa menggunakan data dari 829 karya tugas akhir mahasiswa yang berhasil melakukan klasifikasi tren yang menunjukkan bahwa dalam pengerjaan tugas akhir dengan memanfaatkan buku sebagai sarana referensi mengalami penuruan dan sebalikya sumber yang didapatkan secara olnline melalui internet mengalami peningkatan. Pada [14] melalukan klasifikasi tren dengan topik keilmuan Teknik Industri, data yang digunakan berasal dari suatu media pubilkasi ilmiah Scopus. Dari penelitian ini didapatkan bahwa sebanyak 80 professor telah melakukan tren keilmuan dan juga ada 22 profesor asosiasi, 8 dari asisten professor, 1 berasal dari doctor dan terakhir 1 dari professor terkemuka. Dari beberapa penelitian terkait tren ini menunjukkan bahwa klasifikasi trendapat dilakukan menggunakan data-data yang diperoleh dalam selang waktu tertentu.

Dengan melakukan klasifikasi tren, akan sangat membantu dalam memperluas serta memperdalam informasi untuk keperluan analisis terkait dengan topik yang diteliti. Sehingga penelitian yang dilakukan tidak hanya berpatokan terhadapat akurasi nilai yang didapatkan, namun juga dapat membantu peneliti untuk melihat dari sisi tren sentimen yang dihasilkan. Pada penelitian ini analisis tren akan dilakukan berdasarkan rentang waktu tertentu (harian) yang ditampilkan dalam bentuk curva dan bar diagram untuk melihat apakah positif atau negatif perkembangan tren yang dihasilkan. Oleh karena itu diusulkan analisis tren sentimen terhadap PSBB kota Jakarta dengan menggunakan algoritma SVM dan *Term Frequency-Inverse Document Frequency* (TF- IDF) untuk pembobotan term serta *Grid Search* untuk *hyperparameter optimization* dan curva serta bar diagram untuk melihat tren yang dihasilkan.

Dari hasil analisis tren sentiment yang didapatkan nanti diharapkan dapat membantu pemerintah dalam melihat dan menilai kebijakan yang dilakukan apakah mendapat respon yang positif atau negatif dari masyarakat sehinggadapat menjadi evaluasi kinerja kedepannya atau dimasa mendatang khususnya dalam penanganan situasi pandemiyang sedang dihadapi saat ini.

Pada penelitian [6] menunjukkan bahwa dengan menggunakan metode SVM dalam proses klasifikasi mendapatkan akurasi yang tinggi dan lebih baik dibandingkan dengan Naïve Bayes Classifier (NBC). Dibandingkan dengan metode regresi logistic, *Artificial Neural Network* (ANN), *Naïve Bayes*, dan *Classification and Regression Tree (CART)*, pada penelitian [8-10] dengan menggunakan metode SVM mendapatkan akurasi yang lebih baik. Dari beberapa penelitian diatas menjadi landasan dalam penelitian ini untuk menggunakan metode SVM.

Analisis tren sentimen akan dilakukan mengenai topik PSBB kota Jakarta melalui tweet di media sosial twitter. Namun masalah yang muncul yaitu tweet pengguna yang terlalu banyak yang akan sulit untuk mengkategorikannya dan menganalisisnya. Jika proses klasifikasi atau pengelompokan komentar ataupun informasi dilakukan secara manual akan membutuhkan proses yang lama serta akurasi yang didapatkan tidak akan maksimal [7]. Karena itu dalam mengkategorikan untuk tweet-tweet ke dalam kelas positif atau kelas negative diperlukan suatu cara yang dapat memproses data secara cepat dan otomatis. SVM yaitu algoritma yang digunakan

dan TF-IDF adalah metode yang dipakai untuk pembobotan term serta *Grid Search* untuk *hyperparameter optimization* dan curva serta bar diagram untuk melihat tren yang dihasilkan mengenai topik PSBB Jakarta dengan menggunakan data yaitu *tweet* dari twitter.

Pada penelitian ini terdapat beberapa batasan pekerjaan diantaranya yang pertama analisis dilakukan melalui media sosial twitter. Dibandingkan dengan media sosial yang lain, tulisan-tulisan pada media sosial twitter (*tweet*) memiliki struktur yang sangat cocok untuk digunakan pada analisis. Yang kedua dataset *training* berbahasa Indonesia yang digunakan sebanyak 4340 data terkait dengan PSBB Jakarta. Data akan displit sebesar 90% untuk *train* dan 10% untuk *test* dan untuk dataset yang akan diprediksi atau *data predict* sebanyak 33172 data, yang diperoleh dalam rentang waktu 10 September 2020 sampai dengan 28 September 2020 yaitu pada saat pengumuman dan pelaksaan mengenai PSBB Jakarta yang kedua kalinya (setelah psbb transisi). Dan Batasan yang terakhir yaitu proses pelabelan data tweet ini dilakukan secara manual yang dikelompokkan kedalam du akelas yaitu sentimen positif dan sentimen negatif, karena untuk pelabelan peran semantik pada saat ini saat ini belum ditemukan alat bantu otomatis untuk Bahasa Indonesia yang efektif.

Melaui penelitian yang dilakukan diharapkan dapat untuk mengetahui pengaruh dari metode preprocessing yang digunakan terhadap akurasi untuk menghasilkan SVM paling optimal yang dapat dihasilkan pada penelitian terkait dengan topik PSBB kota Jakarta. Serta untuk mengetahui sentimen dan tren sentimen yang dihasilkan terkait dengan kebijakan PSBB di kota Jakarta dalam rentang waktu 10 September 2020 sampai dengan 28 September 2020 melalui tweet dari media sosial yaitu twitter menggunakan algoritma *Support Vector Machine*.

Organisasi tulisan dalam jurnal ini mengenai pendahuluan yang memuat latar belakang, topik dan batasan, tujuan dari penelitian ini serta organisasi tulisan yang dimuat dalam bab 1. Sedangkan studi terkait dengan penelitian yaitu analisis tren sentimen PSBB di kota Jakarta dengan menggunakan metode SVM ada dalam bab 2. Untuk bab 3 membahas mengenai alur pengerjaan sistem, mulai dari proses awal hingga mencapai tujuan penelitian. Bab 4 berisi tentang pengujian model terbaik dan klasifikasi tren yang dilakukan. Dan terakhir pada bab 5 mengenai kesimpulan dari penelitian yang dilakukan.

2. Studi Terkait

Dalam berbagai publikasi jurnal penelitian banyak membahas mengenai analisis sentimen yang semakin berkembang. Metode yang digunakan pun beragama, seperti metode SVM, ada juga dengan menggunakan algoritma *Decision Tree,* KNN, Naïve Bayes dan metode-metode lainnya.

Riset tentang sentimen berbasis twitter telah dilakukan dalam beberapa penelitian [1-6]. [1] juga memanfaatkan data dari media sosial Facebook, dimana proses klasifikasi sentimen menggunakan algoritma SVM dengan TF-IDF untuk pembobotan *term.* Dari penelitian ini berhasil mengklasifikasikan sentimen masyarakat dan diperoleh sebesar 45% bersentimen netral, 27% bersentimen negative dan 28% bersentimen positif. Penelitian [2] dengan tujuan untuk mendapatkan performansi atau kinerja terbaik dari setiap kernel yang digunakan menggunakan algoritma SVM dimana empat parameter kernel yang akan dibandingkan yaitu: gaussian, polynomial, sigmoid dan kernel linear. Melalui penelitian didapatkan hasil terbesar 80% untuk presisi menggunakan kernel linear, 85% untuk *recall* menggunakan kernel *Sigmoid* dan 81% akurasi yang didapatkan dengan menggunakan kernel *Sigmoid*. Pada penelitian [3] menggunakan tiga algoritma KNN, Naïve Bayes dan *Decision Tree* yang akan dibandingkan dam didapatkan hasil bahwa algoritma dengan menggunakan *Decision Tree* memiliki nilai tertinggi dari nilai *accuracy* dan *recall* yaitu sebesar 83,3%. [4] mendapatkan akurasi tertinggi sebesar 83,2% dengan menggunakan metode SVM yaitu sebesar 83,2%. [5] menggunakan metode SVM dan didapatkan hasil pengujian yang dilakukan sebesar 86%, 67,44% untuk prediksi yang benar pada sentiment negative dan 100% utnuk sentiment positif. [6] meotde SVM dan Nauve Bayes Claasifier dibandingkan dengan menggunakan data yang akan diambil dengan batas waktu tertentu. Dan hasilnya menunjukkan bahwa tidak lebih dari 50% bersentimen negatif dan berhasil membuktikan bahwa dengan menggunakan SVM mendapatkan akurasi lebih baik baik dari NBC dalam klasifikasi yang dilakukan. [7] menggunakan metode SVM dengan daya yang dipakai yaitu review dari pengguna operasi *windows phone* dengan tipe kernel *polynomial* untuk mengklasifikasikan *data review*. Selain menggunakan SVM, penelitian ini juga menggunakan empat metode padatahap tokenisasi yaitu *unigram, bigram, trigram* dan *n-gram*. Pada data *preprocessing*, dilakukan delapan eksperimen dimana masing-masing dari keempat metode tersebut akan menggunakan dua algoritma *stemming, yaitu algoritma Snowball Stmmer* dan *Iterated-Lovin Stemmer*. Hasil dari *preprocessing* digunakan sebagai input dalam proses klasifikasi. Proses klasifikasi dari tiap-tiap input dilakukan sebanyak 3 kali dengan menggunakan nilai C yang berbeda-beda, yaitu C= 0,25, C=0,5 dan C=1,0. Tingkat akurasi tertinggi diperoleh dengan menggunakan *n-gram, algoritme Snowball Stemmer* dan nilai parameter C=1,0.

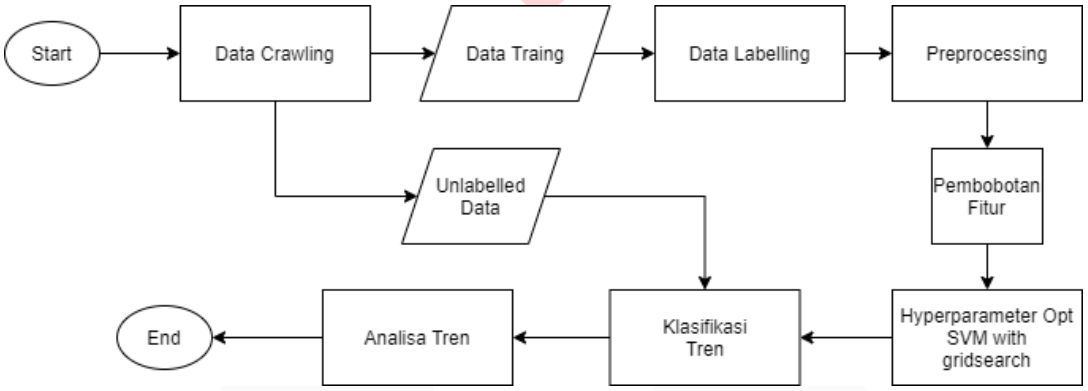
Pada penelitian ini akan melakukan klasifikasi tren, klasifikasi tren sangat membantu dalam memperluas serta memperdalam informasi untuk keperluan analisis terkait dengan topik yang diteliti. Sehingga penelitian yang dilakukan tidak hanya berpatokan terhadapat akurasi nilai yang didapatkan, namun juga dapat membantu peneliti untuk melihat dari sisi tren sentimen yang dihasilkan. Perkembangan tren sangat bermanfaat dalam menilai suatu topik atau objek yang sedang diteliti, yang membantu dalam memperbaiki ataupun meningkatkan kinerjanya. Misalnya dalam tren penjualan, dengan melihat tren yang dihasilkan contohnya tren penjualannya yaitu awalnya besar lalu menjadi kecil, dari tren ini kitab isa menilai bahwa ada hal yang perlu untuk diperbaiki untuk meningkatkan penjualan agar meningkat. Sama halnya dalam penelitian ini, dari tren yang dihasilkan diharapkan dapat membantu pemerintah dalam melihat dan menilai kebijakan yang dilakukan apakah mendapat respon yang positif atau negatif dari masyarakat agar dapat menjadi perbaikan dan evaluasi kinerja kedepannya atau dimasa mendatang khususnya dalam penanganan situasi pandemi yang sedang dihadapi saat ini.

Pada penelitian ini analisis tren akan dilakukan berdasarkan rentang waktu tertentu (harian) yang ditampilkan dalam bentuk curva dan bar diagram untuk melihat perkembangan tren yang dihasilkan apakah positif atau negatif.

3. Sistem yang Dibangun

Metode yang digunakan dalam penelitian ini yaitu *Support Vector Machine* dan untuk ekstraksi dan pembobotan fiturnya digunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) serta *Grid Search* untuk *hyperparameter optimization* dan bar diagram untuk melihat tren yang dihasilkan mengenai topik Pembatasan Sosial Berskala Besar (PSBB) kota Jakarta.

Berikut *flow chart* dalam penelitian ini untuk melihat tahapan analisis tren sentimen yang dilakukan mengenai topik PSBB kota Jakarta.



Gambar 1. *Flow-chart* Analisis Tren Sentimen PSBB Jakarta

3.1 Data Crawling

Pengumpulan data berupa *tweet* dari twitter diperoleh dalam kurun waktu 10 September sampai dengan 28 September 2020 menggunakan bahasa pemrograman yaitu python melalui API yang telah disediakan oleh Twitter. *Tweet* yang akan dipakai berkaitan dengan topik Pembatasan Sosial Berskala Besar di Kota Jakarta dengan menggunakan beberapa *keyword* yang dapat dillihat dalam tabel dibawah. Pada rentang waktu ini tren *tweet* mengenai PSBB Jakarta sedang ramai dibicarakan karena kebijakan pemerintah DKI Jakarta yang Kembali memberlakukan PSBB Jakarta gelombang kedua setelah sebelumnya diberhentikan (diberikan kelonggaran atau PSBB transisi).

Tabel 1. Kata kunci yang digunakan untuk *Crawling*

No.	Kata Kunci
1.	psbb jakarta
2.	psbb jkt
3.	psbb dki
4.	psbb ibukota
5.	#psbbjakarta
6.	#psbbjkt
7.	#psbbdki
8.	#psbbibukota

Hasil dari proses *crawling* ini yaitu dataset yang akan dibagi menjadi dua. Dataset pertama akan digunakan dalam proses *learning* yaitu *data training* untuk menghasilkan model SVM dengan akurasi terbaik yaitu sebanyak 4340 data *tweet*. Sedangkan dataset kedua sebanyak 33172 dataset *unlabelled*, sebagai data yang akan diprediksi untuk digunakan dalam melihat dan menentukan pola tren sentimen.

3.2 Data Labelling

Proses pelabelan dilakukan secara manual pada dataset pertama sebanyak 4340 data *tweet* yang dikelompokkan apakah kelas positif atau negative dengan memperhatikan kriteria pelabelan datanya sebagai berikut.

Tabel 2. Kriteria pelabelan data

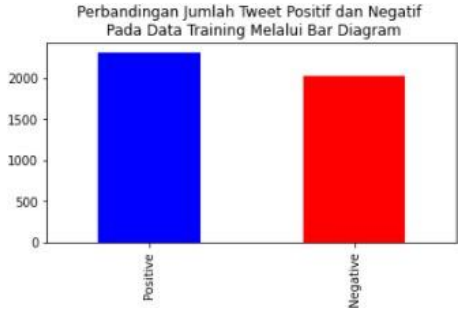
Label Positif	Label Negatif
Kalimat yang menunjukkan dukungan terhadap PSBB Jakarta	Kalimat yang menunjukkan penolakan terhadap PSBB Jakarta
Kalimat tweet yang bersentimen positif jumlah katanya lebih banyak dibandingkan dengan negatif dan memiliki makna emosi yang positif	Kalimat tweet yang bersentimen negatif jumlah katanya lebih banyak dibandingkan dengan positif dan memiliki makna emosi yang negatif

Berdasarkan kriteria pelabelan yang telah ditentukan diatas, maka dilakukan proses pelabelan untuk *data training* yang dapat dilihat dalam tabel berikut beberapa contoh hasil pelabelan data *tweet.*

Tabel 3. Beberapa contoh data tweet berlabel *Positif* dan *Negatif*

No.	Tweets	Sentimen
1.	Mendukung penuh PSBB Jakarta @aniesbaswedan. Salam dari wisma atlet. No-sleep hazmat mode https://t.co/Cu7LUo2elm	Positif
2.	Meskipun dalam waktu dan suasana yang serba terbatas, kangen-kangenan sama sahabat via video call seengaknya bisa membuat mood kita jadi lebih baik loh. Nah kalau mood baik, imunitas juga baik kan? :) #CrystallineWater #PSBB #PSBBJakarta #Pandemi #Covid19 https://t.co/XFXNRXnvN8	Positif
3.	Daerah jakarta mau psbb ketat lagi ðŸ¥º cuman mau bilang please stay safe ya! Please jangan nongkrong2, sering pake masker kalo keluar rumah, sering cuci tangan & jangan sembarangan kucek mata & ambil makanan pake tangan, jaga jarak	Positif
4.	PSBB diperketat atau apapun namanya itu di Jakarta, tidak berpengaruh. https://t.co/BrmAdcn50S	Negatif
5.	PSBB Jakarta diperpanjang berulangkali tapi tdk dikerjakan dg nyata dilapangan, banyak kerumunan dibiarkan bahkan diijinkan, akibatnya penderita covid meningkat, rakyat jakarta yg sengsara #BerantasMafiaDitengahPandemi ðŸ˜· https://t.co/E1udEBKryV	Negatif

Dari hasil pelabelan diperoleh data *tweet* positif sebanyak 2312 dan data tweet negatif sebanyak 2028. Berikut visualisasi perbandingan jumlah data *tweet* berlabel *positif* dan data *tweet* berlabel *negatif* pada *data training* melalui *Bar Chart* dibawah ini.



Gambar 2. Perbandingan jumlah data berlabel positif dan negatif melalui *bar chart*

3.3 Preprocessing Data

Melalui tahap *preprocessing* ini diharapkan dapat mengurangi atribut yang tidak relevan dalam proses klasifikasi dengan mengurangi data-data yang tidak dibutuhkan menjadi bentuk yang dapat dipahami sistem untuk memudahkan proses analisa data.

Berikut ini *preprocessing data* yang dilakukan pada penelitian ini untuk memudahkan proses analisa sesuai dengan kebutuhan data yang dipakai.

1. **Case Folding**
   Untuk mengubah setiap huruf menjadi *lowercase* atau huruf kecil dalam dokumen yang dimiliki digunakan metode *Case Folding*. Walaupun sederhana, namun metode ini tidak boleh untuk diabaikan karena tidak memerlukan *library external* apapun, cukup dengan menggunakan python dengan modul yang sudah disediakan.

2. ***Tokenization* dan *Cleansing Data***
   Untuk memisahkan suatu kalimat menjadi bagian atau yang biasa disebut dengan token dapat menggunakan Teknik yaitu *tokenization* agar kalimat dapat terpisah dan dapat dibedakan. *Cleansing data* atau pembersihan data dilakukan untuk menghilangkan *number and punctuation, url, username/mention, hashtag, replace* karakter berulang, menghilangkan *whitespace, nonASCII chars,* menghilangkan simbol dan karakter aneh serta menghapus kata yang mengandung dua huruf.

3. **Stopword Removal**
   Untuk menghilangkan kata yang kurang memiliki makna dalam suatu kalimat ataupun informasi yang diberikan dalam kalimat tersebut rendah dapat dihilangkan dengan menggunakan metode yaitu *stopword removal,* sehingga kata-kata penting yang lainnya menjadi lebih terfokuskan. Misalnya "yang", "dan", "di", "dari", dan lainnya termasuk ke dalam *stopword* yang akan dihilangkan. Proses ini dilakukan dengan menambahkan *stopword removal* dengan menggunakan *library* sastrawi.

4. **Lemmatization**
   Untuk mengembalikan sebuah katan menjadi bentuk dasarnya digunakan metode yaitu *lemmatization. Lemmatization* berkaitan erat dengan *stemming*, namun perbedaanya adalah *stemmer* beroperasi pada satu kata tanpa pengetahuan tentang konteksnya bahkan walaupun tidak menjadi *root* yang *valid*. Sedangkan pada *lemmatization*, akan mengembalikan bentuk kamus dari kata yang harus menjadi kata yang *valid*. Dibandingkan dengan *stemming, lemmatization* dipilih karena lebih cocok dengan dataset yang ada. Dimana apabila dilakukan *stemming* malah akan menghilangkan bagian dari kata sehingga berbeda arti, misalnya elektabilitas menjadi elektabilita.

3.4 Ekstraksi dan Pembobotan Fitur

Agar dapat diproses setelah tahapan *preprocessing* maka harus dilakukan pengambilan ciri atau *feature* dalam proses yang disebut ekstraksi fitur. Setelah melewati tahapan *preprocessing,* data *tweet* akan masuk tahap ekstraksi fitur. Dalam pembobotan fiturnya digunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengukur seberapa berpengaruh kata dalam dokumen. Dimana setiap *term* diasumsikan memiliki proporsi kepentingan sesuai dengan seberapa seringnya *term* muncul dalam dokumen. Apabila kemunculannya pada himpunan dokumen jarang, maka akan dimasukkan ke dalam nilai *Inverse Document Frequency* (IDF).

Untuk mendapatkan nilai TF-IDF setidaknya diperlukan tiga nilai yaitu *Term Frequency* (TF) yang menunjukkan jumlah kemunculan kata dalam suatu dokumen, *Document Frequency* (DF) menunjukkan jumlah dokumen dimana suatu kata muncul, yang akan menunjukkan seberapa penting dan umumnya kata -kata tersebuut.

Perhitungan TF-IDF dapat diilustrasikan sebagai berikut. Misalnya terdapat 3 kalimat dimana setiap kalimat akan dianggap sebagai dokumen. Proses pertama yaitu TF, setiap kata diberi nilai berdasarkan posisi atau dokumen asal kata tersebut dan jumlah kata tersebut dalam dokumen. Misalnya, terdapat 2 kata 'penuh' dalam dokumen 1, maka nilai TF kata 'penuh' pada dokumen 1 yaitu 2. Kemudian proses yang selanjutnya dokumen dimana suatu *term* (t) muncul akan dihitung sebagai nilai dari DF. Misalnya, dari tiga kalimat tadi terdapat kata 'masyarakat' yang muncul sebanyak 1 kali pada setiap dokumen, maka nilai DF dari kata 'masyarakat' yaitu 3. Selanjutnya menghitung IDF = 1/df atau dengan persamaan idf=log(N/df), dimana nilai df telah didapatkan dari proses sebelumnya. Dan terakhir menghitung TF-IDF dengan mengalikan antara nilai TF dan IDF (tf x idf).

Contoh perhitungan TF-IDF akan diilustrasikan dengan menggunakan tiga kalimat dari data tweet yang digunakan dalam penelitian ini. Dimana setiap kalimat akan dianggap sebagai sebuah dokumen. Dokumen ini diantaranya "mendukung penuh psbb salam wisma atlet", "pemberlakuan psbb Jakarta selama hari menampakkan hasil", "psbb diperketat apapun Namanya Jakarta tidak berpengaruh".

Berikut ini tabel perhitungan TF-IDF untuk tiga kalimat diatas.

Tabel 4. Contoh Perhitungan TF-IDF

Term (t)	Nilai Term Frekuensi (tf)			df	D/df	idf	TF.IDF (tf x idf)		
	D1	D2	D3			log (N/df)	D1	D2	D3
mendukung	1	0	0	1	3	0,477121	0,477121	0	0
penuh	1	0	0	1	3	0,477121	0,477121	0	0
psbb	1	1	1	3	1	0	0	0	0
salam	1	0	0	1	3	0,477121	0,477121	0	0
wisma	1	0	0	1	3	0,477121	0,477121	0	0
atlet	1	0	0	1	3	0,477121	0,477121	0	0
pemberlakuan	0	1	0	1	3	0,477121	0	0,477121	0
jakarta	0	1	1	2	1,5	0,176091	0	0,17609	0,17609
selama	0	1	0	1	3	0,477121	0	0,477121	0
hari	0	1	0	1	3	0,477121	0	0,477121	0
menampakkan	0	1	0	1	3	0,477121	0	0,477121	0
hasil	0	1	0	1	3	0,477121	0	0,477121	0
diperketat	0	0	1	1	3	0,477121	0	0	0,477121
apapun	0	0	1	1	3	0,477121	0	0	0,477121
namanya	0	0	1	1	3	0,477121	0	0	0,477121
berpengaruh	0	0	1	1	3	0,477121	0	0	0,477121
banyak	0	0	1	1	3	0,477121	0	0	0,477121

3.5 Hyperparameter Optimization

Proses klasifikasi dalam penelitian ini menggunakan algoritma *Support Vector Machine* yang adalah salah satu algoritma *machine learning* yang bertujuan untuk untuk memisahkan dua kelas yang berbeda atau biasa disebut dengan *hyperplane* sebagai metode pemisah terbaik yang dapat mengelompokkan kelas positif (1) dan kelas negatif (-1). Untuk mendapatkan *hyperplane* yaitu dengan memaksimalkan jarak antar kelas atau pemisah antar kelas atau *margin* untuk menemukan titik optimum *hyperplane* tersebut. Inti dari proses pembelajaran SVM nya yaitu usaha untuk mencari lokasi *hyperplane* ini. Untuk menentukan nilai *hyperplane,* terlebih dahulu memaksimalkan nilai margin dengan persamaan sebagai berikut.

$$\frac{1}{2}||w||^2 \quad (1)$$

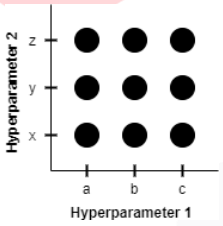
Dengan syarat:

$$w.x_i + b = 0 \quad (2)$$

Untuk mendapatkan garis yang tgak lurus antara garis *hyperplane* dan titik *support vetor* dinyatakan dengan w sebagai nilai dari parameter *hyperplane* yang dicari, untuk atribut ke I pada data dinyatakan dengan x_i, dan untuk bias dinyatakan sebagai b. Dari persamaan 2, kelas *hyperplane* dibedakan ke dalam dua kelas yaitu kelas positif (1) dan negatif (-1). Penelitian ini akan menggunakan model LinearSVC dengan optimasi *hyperparameter* menggunakan *Grid Search Cross Validation* yang akan diuji untuk mendapatkan model dengan akurasi terbaik dengan parameter yang diberikan akan disesuaikan dengan kebutuhan data yang dipakai dan menerapkan *hyperparameter turning* akan menggunakan *pipeline. Pipeline* merupakan fungsi yang mentransformasikan *final estimator.* Tujuan *pipeline* adalah mengumpulkan beberapa langka yang dapat divalidasi silang Bersama-sama sambal mengatur parameter berbeda. *Pipeline* akan berisi pengklasifikasi yang terdiri dari *CountVectorizer, TfidfTransformer,* dan *LinearSVC* sebagai model yang digunakan. Kemudian menentukan kombinasi parameter yang inigin diuji pada *pipeline* untuk mendapatkan hasil terbaik. Dalam melakukan analisis dengan model LinearSVC diperlukan optimasi *hyperparameter* dimana akan digunakan parameter n_gram, tfidf, normalisasi dan nilai tol dengan nama variabel
masing-masing dalam code yaitu, clf__tol, tfidf__norm, kemudian tfidf__use_idf dan vect__ngram_range.

Salah satu permasalahan yang ditemukan dalam proses pembangunan suatu model yaitu pencarian *hyeperparameter* yang optimal. Misalnya berapa nilai gamma dan C untuk mendapatkan model dengan parameter terbaik, berapa nilai yang harus diberikan pada parameter yang digunakan agar dapat memberikan hasil yang optimal dan *hyperparameter* lainnya yang harus dioptimasi. Jika menggunakan *trial-error* dengan mencoba mengganti satu persatu parameter tersebut untuk mencari model yang terbaik maka tentu akan sangat lama dan tidak akan efisien, diperlukan suatu solusi dan dalama penelitian ini digunakan *Grid Search* untuk menentukan kombinasi *hyperparameter*. Sesuai dengan namanya, *Grid Search* akan mencari parameter *"grid"* yang diberikan dan *cross validation* (CV) untuk menguji performanya. *Grid Search Cross Validation* adalah metode yang akan mengkombinasikan parameter yang dimasukkan kemudian pemilihan kombinasi model dengan *hyperparameter* masing-masing dilakukan secara otomatis dan sistematis dengan cara menguji coba satu persatu kombinasi dan melakukan validasi untuk setiap kombinasi agar dapat menentukan kombinasi yang menghasilkan performa model terbaik yang dapat dipilih untuk dijadikan model untuk prediksi. *Grid Search* akan memakan waktu yang lama jika menggunakan banyak parameter. Namun, meskipun *Grid Search* bisa sangat mahal secara komputasi, sebagai pencarian lengkap, ini sangat berguna untuk melihat semua kombinasi *hyperparameter* yang ditentukan.

Hyperparameters dan parameters adalah hal yang berbeda, yang membedakannya yaitu pada nilai yang didapatkan dimana nilainya bisa untuk didapatkan secara langsung dari proses *training* jika parameters sedangkan hyperparameters tidak bisa karena nilainya sebelum proses *training* nilainya sudah ditentukan. Optimasi *hyperparameter* diantaranya vect_ngram_range dengan mencari kombinasi dari (1,2), (1,3), (1,4) lalu tfidf_use_idf dengan parameter True, False kemudian tfidf__norm dengan parameter (l1,l2) dan terakhir ada clf_tol dengan nilai 1, 0.1, 0.01, 0.001. *Grid Search* akan memilih *hyperparameter* mana yang akan memberikan model performa yang terbaik. cara kerja dari *grid search* dapat diilustrasikan sebagai berikut. Misalnya terdapat, Hyperparameter_satu = [a, b, c] dan Hyperparameter_dua = [x, y, z], yang dengan gambar berikut.



Gambar 3. Ilustrasi dengan metode *grid search*

Dengan metode *grid search*, maka pencarian akan dilakukan dengan membandingkan semua kemungkinan kombinasi parameter yang terjadi (dengan membandingkan satu persatu kombinasi yang ada dalam parameter), yaitu antara satu parameter dengan parameter yang lainnya. Misalnya dari contoh diatas, maka akan dilakukan pencarian dengan kombinasi yaitu (a,x) , (a,y) , (a,z) , (b,x) , (b,y) , (b,z) , (c,x) , (c,y) , (c,z). Dari pencarian dengan kombinasi ini maka akan diambil kombinasi dengan nilai tertinggi.

Metode evaluasi dilakukan dengan *cross validation* (CV) yang jenisnya sendiri menyesuaikan dengan nilai yang diberikan. Pada penelitian ini akan diset 10, artinya setiap kombinasi model dan parameter divalidasi sebanyak 10 kali dengan membagi data sebanyak 10 bagian sama besar secara acak. Metode ini akan melakuakan evaluasi kinerja dari model dengan ukuran yang sama, cara kerjanya yaitu dengan memisahkan data ke dalam dua subset yaitu untuk proses *training* yang akan digunakan untuk proses *learning* sebagai k-1 *subset* dan yang kedua yaitu 1 *subset* yang tersisa sebagai data *testing* yang digunakan untuk evaluasi atau validasi. Dengan menggunakan CV *k-fold* akan membantu dalam proses komputasinya agar dapat lebih berkurang keakuratan estimasinya pun akan tetap terjaga. Setiap *fold* dapat dipakai untuk menghitung nilai dari kinerja klasifikasi dari rata-rata yang diperoleh. *K-fold* yang umum digunakan yaitu 10[15]. Hal yang penting dan perlu diperhatikan dalam penentuan nilai *k-fold* yaitu *k-fold* yang dapat memberikan estimasi yang kurang bias apabila dibandingkan dengan jenis atau nilai yang lain jika digunakan dan nilai 10 ini sangat direkomendasikan untuk itu. Apabila nilai yang diberikan yaitu 10 fold CV, hal ini berarti data displit menjadi 10 bagian yang berukuran sama kemudian dari 9 bagian ini digunakan untuk pelatihan atau *training* dan sisanya yaitu 1 fold digunakan untuk pengujian atau *testing*.

3.6 Klasifikasi Tren

Setelah proses *hyperparameter optimazation* dengan *Grid Search* dilakukan, didapatkan model dengan akurasi tertinggi dengan parameter terbaik yang selanjutnya akan digunakan dalam proses klasifikasi tren. Klasifikasi tren dapat dilakukan dalam rentang waktu tertentu, baik secara harian, mingguan, bulanan, tahunan atau bahkan dalam satuan jam. Setiap waktu memiliki pembahasan yang berbeda-beda sehingga tren yang dihasilkan pun dapat berbeda mengikuti topik yang sedang dibahas saat itu. Pada beberapa kasus, klasifikasi tren dapat membantu dalam memprediksi tren dimasa mendatang atau pada saat tertentu, misalnya pada kasus prediksi polusi udara yang dapat diprediksi karena memiliki siklus. Berbeda dengan kasus yang dapat diprediksi seperti polusi udara itu, pada penelitian ini hal itu tidak dapat diterapkan karena topik yang dibahas tidak

berkelanjutan dimana tidak terjadi setiap tahun (tidak memiliki siklus) dan hanya terjadi pada saat itu saja atau karena suatu pemicu. Dalam hal ini yaitu sebagai tanggapan masyarakat terhadap kebijakanpemerintah terhadap PSBB Jakarta.

Pada penelitian ini analisis tren dilakukan berdasarkan rentang waktu tertentu (yaitu secara harian) yang ditampilkan dalam bentuk curva dan bar diagram untuk melihat perkembangan tren yang dihasilkan apakah positif atau negatif. Dimana sumbu X sebagai satuan waktu yang telah dikelompokkan secara harian (per-hari) dan sumbu Y yaitu jumlah data tweet positif serta jumlah data tweet negatif yang dihasilkan.

Dari hasil klasifikasi tren yang dilakukan dapat dilihat bagaimana perkembangan tren untuk mengetahui tanggapan masyarakat terhadap topik PSBB Jakarta apakah positif atau negatif.

4. Evaluasi

Proses klasifikasi yang dilakukan menggunakan algoritma SVM dengan optimasi hiperparamteter *Grid Search* dilakukan untuk melihat dan membandingkan akurasi dan parameter terbaik dari model yang digunakan berdasarkan kombinasi parameter yang paling optimal.

4.1 Pengujian Model Terbaik

Dari klasifikasi yang dihasilkan akan dilakukan beberapa skenario pengujian terhadap metode *preprocessing* dengan membandingkan akurasi yang didapatkan. Pengujian ini dilakukan untuk melihat pengaruh *preprocessing* terhadap hasil akurasi dari proses klasifikasi. Skenario pengujian tersebut diantaranya yang pertama yaitu dengan menghilangkan proses *case folding* dan *cleansing* kemudian yang kedua yaitu tanpa *tokenization* dan *lemmatization* lalu yang ketiga yaitu tanpa *stopword removal* dan yang terakhir dengan menggunakan semua *preprocessing.*

Berikut ini tabel yang menunjukkan hasil pengujian berdasarkan masing-masing skenario yang telah dijalankan yang menunjukkan akurasi serta nilai parameter optimal yang dihasilkan.

Tabel 5. Hasil pengujian skenario

No	Metode Preprocessing	Parameters				Rata-rata Akurasi
		clf__tol	tfidf__norm	tfidf__use_idf	vect__ngram_range	
1.	Tanpa *Case Folding* dan *Cleansing*	0,1	l2	True	(1,3)	85,94%
2.	Tanpa *Tokenization and Lemmatization*	1	l2	True	(1,2)	85,71%
3.	Tanpa *Stopword Removal*	0,1	l2	True	(1,2)	87,33%
4.	Dengan *preprocessing* lengkap	0,01	l2	True	(1,3)	86,41%

Dari skenario pengujian yang telah dijalankan diatas, diketahui bahwa akurasi tertinggi didapatkan dengan klasifikasi tanpa *stopword removal* yang akan dipakai atau diimplementasikan untuk klasifikasi tren pada data *predict* yaitu dengan akurasi sebesar 87,33% dan dengan optimasi parameter svm yang diperoleh dimana parameter clf__tol yaitu 0,1, tfidf__norm yaitu l2, kemudian tfidf__use_idf yaitu bernilai True dan parameter vect_ngram_range yaitu (1,2). Dapat dilihat pada tabel 5, dua pengujian yang mendapatkan akurasi tertinggi yaitu pengujian tanpa menggunakan *stopword removal* dan pengujian dengan menggunakan semua preprocessing lengkap yaitu dengan akurasinya 86,41%. Pengujian dengan *stopword removal, case folding* dan *cleansing* saja mendapatkan akurasi yang lebih kecil yaitu 85,71%. Dan ketika menggunakan *stopword removal, tokenization* dan *lemmatization* saja mendapatkan akurasi sebesar 85,94%. Semua pengujian scenario yang dilakuka yang melibatkan metode *stopword removal* didalamnya mendapatkan akurasi yang rendah. Hal ini karena kata-kata yang terdapat dalam *library* sastrawi yang digunakan misalnya kata kurang, jangan, bukan dan lainnya malah mengurangi suatu informasi dari kalimat yang ada sebagai suatu kesatuan yang utuh sehingga kurang efektif apabila diterapkan karena dapat menghilangkan makna yang sebenarnya atau kehilangan sentimennya yang menjadi salah satu alas an mengapa nilai sentiment akhirnya salah, contohnya pada table 6.

Kata-kata tersebut awalnya dianggap sebagai *stopword*, tapi nyatanya kata tersebut malah membantu proses pelabelan untuk menentukan sentimen. Dari kasus ini juga menunjukkan bahwa untuk memilih dan membuat kamus yang berisi kata-kata *stopword* yang tepat sulit untuk dilakukan.

Tabel 6. Contoh Tweet yang Salah diklasifikasikan

Tweet Sebenarnya	Tweet Setelah Preprocessing	Label Sebenarnya	Prediksi Label	Keterangan
PSBB akan sangat tidak ada hasilnya jika di tengah karantina masih biarkan ttp adakan orasi.	psbb sangat hasilnya tengah karantina biarkan ttp adakan orasi	Negative	Positive	Sistem gagal untuk memahami dan mengelompokkan kata ke dalam kelas yang seharusnya karena dalam kamus sastrawi yang digunakan terdapat kata 'tidak' yang berarti akan dihilangkan.
Semoga PSBB kali ini tidak terlalu mengguncang sektor ekonomi.. masih bisa mencari nafkah.. aamiin..	semoga psbb kali terlalu mengguncang sektor ekonomi mencari nafkah aamiin	Positive	Negative	Sama dalam kasus pada *tweet* diatas pengelompokkan katanya menjadi salah sehingga membuat *tweet* masuk ke dalam kelas negatif.

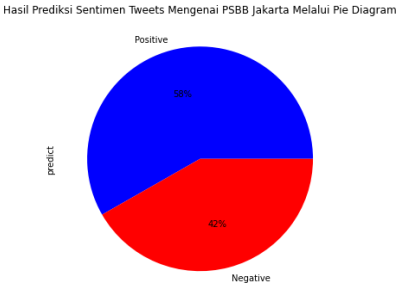
Pengujian yang terdapat teknik *preprocessing*, *case folding*, *cleansing*, *tokenization* dan *lemmatization* menunjukkan bahwa *preprocessing* tersebut memberikan pengaruh yang cukup baik dalam meningkatkan kinerja sistem klasifikasi. Walaupun pengujian dengan menggunakan semua *preprocessing* lengkap (*case folding, cleansing, tokenization,lemmatization* dan *stopword removal*) yang menerapkan *stopword removal* mendapatkan akurasi tertinggi kedua, hal ini karena dilakukan pengkombinasian keempat teknik *preprocessing* yaitu *case folding*, *cleansing*, *tokenization* dan *lemmatization* yang membuat akurasi yang dihasilkan menjadi optimal. *Case folding* yang mampu mengubah huruf menjadi *lowercase* dan *cleansing* mampu yang mengurangi fitur-fitur yang kurang informatif dalam kalimat *tweet* seperti simbol, angka, serta link url yang akan digunakan dalam klasifikasi, sehingga dihasilkan fitur yang memiliki relevansi dengan kelasnya. *Tokenization* dan *lemmatization* akan mengembalikan kata dalam *tweet* menjadi bentuk dasarnya sehingga variansi fitur yang memiliki makna yang sama dapat dikurangi. Meskipun hasil akurasipada pengujian yang melibatkan *stopword removal* yaitu pada pengujian dengan menggunakan *preprocessing* lengkap menghasilkan akurasi yang cukup tinggi namun penggunaan *stopword removal* di dalamnya memberikan pengaruh yang kurang baik yang ditunjukkan dari akurasi rendah yang didapatkan pada pengujian melalui tabel 5. Pengaruh terhadap akurasi ini karena berkurangnya informasi dari keutuhan suatu kalimat dalam *tweet* yang menyebabkan fitur-fitur hasil dari *stopword removal* belum mampu untuk memberikan gambaran dalam penentuan kelasnya.

Meskipun perbedaan akurasinya tidak begitu besar, namun tetap saja dari proses pengujian ini juga membuktikan bahwa penambahan *stopword removal* pada penelitian ini berpengaruh terhadap akurasi dari proses klasifikasi yang dilakukan dan dari pengujian yang dihasilkan dapat menunnjukkan bahwa walaupun menggunakan *preprocessing* yang lebih banyak, namun hal tersebut tidak menjadi jaminan bahwa akurasi dan kinerja yang dihasilkan akan menjadi lebih baik.

4.2 Klasifikasi Tren

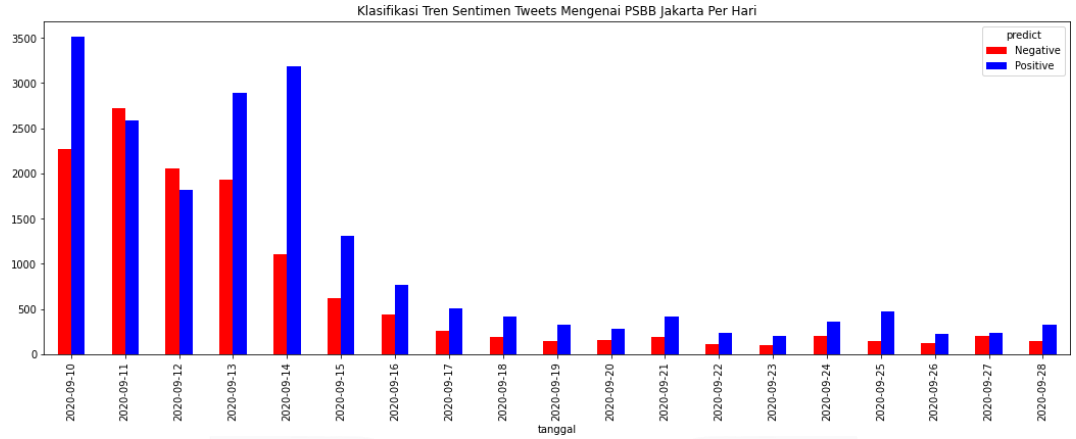
Telah dilakukan pengujian sesuai berdasarkan skenario pengujian dan didapatkan hasil bahwa metode paling optimal yang akandigunakan dalam proses data prediksi yaitu dengan menghilangkan proses *stopword removal* dan akan menggunakan semua *preprocessing* lainnya, termasuk *case folding, cleansing data, tokenization* dan *lemmatization.* Dan didapatkan hasil bahwa secara keseluruhan yang dimulai dari tanggal 10 sampai 28 September 2020 mendapatkan tren yang positif dari masyarakat yang dapat dilihat melalui tren dari bar diagram dimana tren positif mendominasi secara keseluruhan. Hal ini berarti sebagian besar masyarakat memberikan respon yang positif terkait dengan kebijakan pemberlakuan PSBB kota Jakarta terbukti dari jumlah tweet positif yang lebih banyak.

Berikut ini dapat dilihat visualisasi perbandingan akurasinya melalui pie diagram dibawah ini.



Gambar 4. Perbandingan hasil prediksi sentimen positif dan negatif melalui Pie Diagram

Untuk melihat tren yang dihasilkan dalam waktu tertentuberikut ini dapat dilihat hasil dari klasifikasi tren secara harian yang dilakukan pada data prediksi melalui multi histogram berikut ini.



Gambar 5. Visualisasi klasifikasi tren melaui bar diagram

Dari visualisasi diatas dapat terlihat bahwa mulai pada hari pertama tanggal 10 September 2020 sampai dengan tanggal 12 September 2020 tren langsung mengalami penurunan yang cukup tajam dan kemudian naik pada tanggal 13 namun pada tanggal 15 September mengalami penurunan yang sangat tajam. Bahkan dapat dilihat dalam kurun waktu 14 sammpai 28 September 2020 perlahan-lahan tren terus menurun. Pada beberapa titik (hari) mengalami sedikit peningkatan, namun hal ini tidak terlalu signifikan jika dibandingan dengan tren yang mengalami penurunan.

Dari visualisasi tren menunjukkan bahwa respon masyarakat terhadap kebijakan pemerintah DKI Jakarta untuk memberlakukan kembali PSBB mendapatkan respon yang tinggi hanya pada minggu pertama setelah kebijakan ini diumumkan yaitu mulai tanggal 10 September 2020 sampai dengan 15 September 2020 dan tren yang dihasilkan pada minggu pertama pun adalah tren yang positif yang dapat dilihat dari visualisasi bar diagram. Pada minggu kedua terlihat bahwa perkembangan tren tweet langsung mengalami penurunan yang sangat tajam mulai tanggal 15 sampai dengan tanggal 28 September 2020 tren terus menurun secara perlahan, namun dapat dilihat bahwa tren yang dihasilkan pada minggu kedua masih mendapatkan tren yang positif yang dapat dilihaat dari visualisasi diatas.

5. Kesimpulan

Berdasarkan pengujian dan analisis yang telah dilakukan dalam penelitian ini, menunjukkan hasil bahwa dengan menghilangkan *stopword removal* dan menggunakan semua *preprocessing* lainnya, termasuk *case folding, cleansing data, tokenization* dan *lemmatization*, dapat meningkatkan hasil akurasi terhadap proses klasifikasi yang dilakukan dimana memberikan akurasi yang paling optimal sebesar 87,33% yang menunjukkan akurasi data yang terklasifikasi dengan benar. Pengaruh akurasi yang dihasilkan terkait dengan penggunaan *stopword removal* ini karena kalimat *tweet* kehilangan sentimen dan maknanya pun menjadi berbeda karena kata-kata yang dianggap *stopword* ternyata membantu dalam proses pengelompokan ke dalam sentimennya. Karena itu dalam penentuan daftar kata yang tepat yang akan dimasukkan ke dalam kamus *stopword* akan menjadi sulit untuk dilakukan.

Setelah mendapatkan model dengan akurasi yang paling optimal, selanjutnya akan digunakan untuk klasifikasi tren pada data prediksi. Dimana diperoleh bahwa tren positif mendominasi secara keseluruhan dibandingkan angka sentimen negatif yang hanya terjadi pada dua hari dari total 19 hari secara keseluruhan. Jadi, sentimen publik terhadap kebijakan pemerintah DKI Jakarta dalam memberlakukan PSBB mendapatkan respon yang positif dari masyarakat.

Tren yang dihasilkan dari penelitian ini yaitu tren turun yang dapat dilihat dari hasil visualisasi data melalui curva dan bar diagram. Walaupun pada beberapa titik (perpindahan hari) sedikit mengalami kenaikan, namun hal ini tidak terlalu mencolok atau tidak tajam dan tidak terjadi secara signifikan dibandingkan dengan tren yang mengalami penurunan. Dimana dapat terlihat bahwa tren perlahan-lahan langsung mulai menurun dari hari pertama tanggal 10 September 2020 sampai tanggal 28 September 2020. Tren mengalami peningkatan hanya pada tanggal 13, lalu sedikit naik pada tanggal 21 dan 24 sampai dengan tanggal 25 September 2020. Dimana kenaikan ini tidak terlalu tajam.

Referensi

[1] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," *J. Inf. Syst. Dev.*, vol. 4, no. 1, 2020, [Online]. Available: https://122.200.2.179/index.php/isd/article/view/381.

[2] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentimen Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *J. Inform. Polinema*, vol. 4, no. 2, p. 93, 2018, doi: 10.33795/jip.v4i2.152.

[3] I. T. S. A. Pamungkas, "Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (Svm)," *Log!K@*, vol. 8, no. 1, pp. 69–79, 2018.

[4] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.

[5] nusadaily.com, "Indef Ungkap PSBB Picu Banyak Sentimen Negatif ke Pemerintah," *nusadaily.com*, 2020. https://nusadaily.com/news/indef-ungkap-psbb-picu-banyak-sentimen-negatif-ke-pemerintah.html (accessed Oct. 22, 2020).

[6] M. Syarifuddinn, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.

[7] S. H. Thorik, "Efektivitas Pembatasan Sosial Berskala Besar Di Indonesia Dalam Penanggulangan Pandemi Covid-19," *J. Adalah Bul. Huk. dan Keadilan*, vol. 4 No. 1, pp. 115–120, 2020.

[8] Rachman, F., & Purnami, S. W. (2012). Klasifikasi Tingkat keganasan Breast Cancer dengan menggunakan Regresi Logistik Ordinal dan Support Vector Machine (SVM). Jurnal Sains dan Seni ITS Vol.1 NO.1, D130.

[9] Huang, J., Jingjing, L., & Ling, C. (2003). Comparing Naive Bayes, Decision Trees and SVM with AUC and accuracy. International Conference on Data Mining.

[10]Byvatov, E. e. (2003). Comparison of Support Vector Machine and Artificial Neural Network System for Drug/Nondrug classification. Chem Inf Compu Sci, 1882-1889.

[11] S. Laimeheriwa, E. L. Madubun, and E. D. Rarsina, "Analisis Tren Perubahan Curah Hujan dan Pemetaan Klasifikasi Iklim Schmidt - Ferguson untuk Penentuan Kesesuaian Iklim Tanaman Pala (Myristica fragrans) di Pulau Seram," *Agrologia*, vol. 8, no. 2, 2020, doi: 10.30598/a.v8i2.1012.

[12] D. L. Rianti, Y. Umaidah, and A. Voutama, "Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 98, 2021, doi: 10.30998/string.v6i1.9993.

[13] "Analisis Tren Penelitian Tugas Akhir Mahasiswa Jenjang S1 Stmik Akakom.pdf." .

[14] M. Syarifuddinn, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.

[15] Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques 3rd Edition. USA: Morgan Kaufmann