

Analisis Sentimen Masyarakat Terhadap Kondisi Perekonomian di Indonesia Pada Masa Pandemi 2020

Krisna Abhi Padhana^{*}, Mujiono Sadikin^{}**

^{*}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Raya Meruya Selatan, Kembangan, Jakarta, 11650
41517010031@student.mercubuana.ac.id

^{**}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Raya Meruya Selatan, Kembangan, Jakarta, 11650
Mujiono@mercubuana.ac.id

ABSTRACT

Masyarakat Indonesia banyak yang memperbincangkan isu ekonomi Indonesia pada masa pandemi Covid-19 2020 di media sosial. Salah satu isu yang diperbincangkan yaitu mengenai resesi Indonesia dimana sudah ada masyarakat yang memperbincangkan hal ini di media sosial sebelum Indonesia mengalami resesi. Masyarakat yang khawatir dan membahas isu ekonomi ini lebih cenderung untuk menuangkan opini negatif di media sosial. Untuk itu dibuat analisis sentimen data opini publik mengenai ekonomi Indonesia di media sosial khususnya twitter. Sehingga dapat diklasifikasikan sentimen setiap opini dan mempermudah mendeteksi isu ekonomi yang ada pada kumpulan opini publik yang bersentimen negatif. Dan dapat dianalisa dan dijadikan acuan oleh pihak yang memerlukan apakah memerlukan penanganan atau tidak. Klasifikasi sentimen pada penelitian ini dibuat menggunakan algoritma Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor dengan menggunakan metode cross validation 10 fold dengan 3 skenario random percentage split. Dari hasil evaluasi dari ketiga model tersebut diketahui bahwa algoritma Support Vector Machine dengan percentage split 90:10 memiliki kinerja paling baik yang menghasilkan akurasi sebesar 95,53%.

PENDAHULUAN

Awal tahun 2020 terdapat virus baru yaitu Covid-19 yang mewabah di berbagai negara sehingga dinyatakan sebagai pandemi. Hingga akhir bulan Februari 2020 Indonesia masih menjadi salah satu negara yang bebas oleh Covid-19, namun kasus pertama Covid-19 akhirnya ditemukan setelah beberapa warga Depok diperiksa dan dinyatakan positif pada tanggal 2 Maret 2020 (Lengkong et al., 2021). Terjadinya pandemi Covid-19 di suatu negara tidak hanya mempengaruhi dari aspek kesehatan masyarakatnya saja. Perekonomian pada suatu negara juga dapat terimbas oleh pandemi ini. Pada awal tahun 2020 terdapat beberapa negara yang sudah terkena dampak ekonomi dan mengalami resesi. Bahkan negara yang mengalami resesi ini tidak hanya dari negara berkembang saja, ada juga negara maju yang mengalaminya seperti Jepang, Korea Selatan, dan

Australia. Resesi sendiri merupakan suatu kondisi di suatu negara dimana ekonominya mengalami pelemahan, yang mengakibatkan pertumbuhan ekonominya berada pada titik minus (Yamali & Putri, 2020).

Dampak ekonomi yang cukup besar pun mulai dirasakan Indonesia setelah beberapa bulan terdampak oleh pandemi Covid-19. Hal ini ditandai bahwa pada triwulan ke-II pertumbuhan ekonomi Indonesia tidak begitu baik dan berada pada titik minus yaitu -5,32% (Lengkong et al., 2021). Dan akhirnya pada kuartal ke-III tepatnya pada hari Kamis 5 November 2020, Badan Pusat Statistik (BPS) mengumumkan bahwa dengan kondisi pertumbuhan ekonomi Indonesia yang masih berada di titik minus maka resmi dapat dikatakan mengalami resesi (Kusuma, 2020).

Tentunya kabar mengenai gejala ekonomi di Indonesia ini banyak diperbincangkan oleh masyarakat. Pada masa pandemi 2020 ini, dimulai dari

awal terkonfirmasi kasus Covid-19 di Indonesia hingga 31 Desember 2020 terdapat 116.018 pengguna Twitter yang memperbincangkan mengenai ekonomi di Indonesia. Dan terdapat 92.334 tweet dibuat sebelum Indonesia resmi dinyatakan mengalami resesi, dimana 5.525 tweet di antaranya memperbincangkan mengenai resesi. Data ini diketahui dari dataset hasil scraping yang digunakan pada penelitian ini. Hal ini menandakan adanya isu ekonomi mengenai resesi dan banyak diperbincangkan oleh masyarakat di media sosial khususnya twitter sebelum Indonesia resmi dinyatakan resesi.

Keluh kesah masyarakat Indonesia pada tweetnya beragam. Ada masyarakat yang khawatir dengan ekonomi di Indonesia dan juga ada yang tetap optimis. Tentunya dari berbagai tweet ini memiliki sentimen masing masing yang dapat menunjukkan apakah yang disampaikan oleh si penulis secara keseluruhan bersifat positif atau negatif. Oleh karena itu dapat digunakan teknik yang sering dilakukan dalam penelitian untuk mengekstrak opini publik yaitu dengan analisis sentimen. Teknik ini termasuk ke dalam natural language processing karena mengolah dan mengekstrak opini publik berbentuk teks untuk menentukan sentimen pada setiap opini (Febrianti et al., 2018).

Pada penelitian ini membahas tentang analisis sentimen mengenai ekonomi di Indonesia pada masa pandemi Covid-19 2020. Dengan dilakukannya analisis sentimen ini dapat diketahui bagaimana tingkat kekhawatiran masyarakat dengan kondisi ekonomi di Indonesia pada masa pandemi 2020. Dan dari yang sudah dijelaskan sebelumnya dimana terdapat 5.525 data tweet yang membahas mengenai resesi yang dibuat sebelum Indonesia resmi dinyatakan mengalami resesi. Hal ini menandakan bahwa terdapat pembahasan isu ekonomi mengenai resesi oleh masyarakat. Pembahasan mengenai isu ekonomi dapat disinyal dengan banyaknya opini publik bersentimen negatif mengenai hal ini. Sehingga selain dapat melihat tingkat kekhawatiran masyarakat, bisa juga untuk mengidentifikasi isu ekonomi yang terjadi. Sehingga dapat dijadikan acuan untuk mengambil tindakan oleh pihak yang memerlukan. Salah satunya yaitu Kementerian Koordinator Bidang Perekonomian yang memiliki tugas menangani isu isu ekonomi yang terjadi di Indonesia (Perekonomian, 2016).

STUDI LITERATUR

Klasifikasi

Terdapat beberapa metode yang masih tergolong ke dalam *machine learning*, salah satunya yaitu klasifikasi yang berdasar pada konsep *supervised learning*. Konsep dari *supervised learning* yaitu memiliki suatu kegunaan dalam melakukan prediksi kedalam kelas manakah data yang ingin di prediksi dari hasil pembentukan model. Sesuai dengan namanya data yang di training diawasi berdasarkan label kelas yang sudah diberikan. Sehingga terbentuk model yang telah di ajari data yang diberikan dan mampu mengklasifikasikan data baru (Ulfah & Anam, 2020). Pada eksperimen ini klasifikasi dilakukan dengan algoritma SVM, NB, dan KNN.

K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma yang memprediksi kelas dari objek baru atau data uji berdasarkan namanya. Algoritma ini memiliki kinerja dengan melakukan pelabelan objek baru kedalam kelas yang memiliki tingkat perbedaan yang paling sedikit dan memiliki banyak kemiripan. Dengan menggunakan euclidean distance maka jarak dari objek yang di prediksi dengan data latih dapat dihitung, sehingga diketahui kelas manakah yang merupakan tetangga terdekat (Wibowo & Fitriah, 2018).

Naïve Bayes

Naïve Bayes adalah algoritma yang sering digunakan dalam kasus prediksi suatu kejadian atau bagaimana tingkat probabilitas dari suatu kejadian. Pada analisis sentimen algoritma ini bekerja dengan asumsi bahwa, pengaruh dari kemunculan kata yang satu dengan yang lainnya tidak ada sama sekali dengan kata lain independen (Andika et al., 2019). Konsep independensi ini berdasarkan dari teorema bayes yang mempelajari prediksi dengan menghitung probabilitas yang sederhana (Sadikin & Alfiandi, 2018).

Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma yang memerlukan label kelas data untuk mengawasi dan melatih data pada proses training data. Pada tahap eksperimen ini terdapat dua label kelas data

yaitu label kelas positif dan negatif. Algoritma SVM bekerja dengan menghasilkan nilai yang berbentuk hyperplane yang berbentuk garis untuk memisahkan antara dua kelas label, yaitu negatif dan positif (Luqyana et al., 2018).

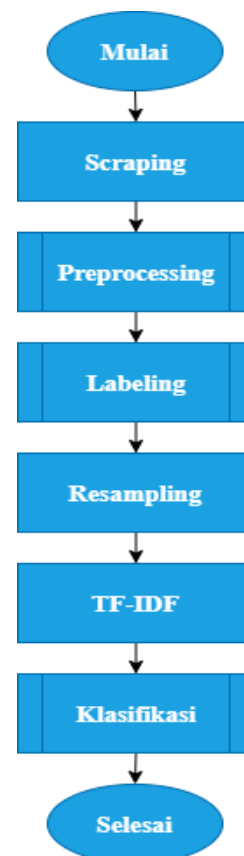
Studi Terkait

Penelitian mengenai analisis sentimen sendiri telah beberapa kali dilakukan. Salah satunya pada penelitian mengenai analisis sentimen ulasan tentang tokopedia yang dipublikasikan pada (Pajri et al., 2020) dapat diketahui bahwa algoritma KNN dapat digunakan untuk mengklasifikasikan sentimen yang menghasilkan akurasi sebesar 88,11%. Kemudian penelitian mengenai analisis sentimen terhadap data tweet saat terjadi peristiwa kritis yang dipublikasikan pada (Ruz et al., 2020) dapat diketahui bahwa dalam mengklasifikasikan sentimen mengenai dua data berbeda dengan beberapa algoritma, didapatkan bahwa algoritma terbaiknya pun bisa berbeda pada masing masing data yang diklasifikasikan. Pada data tweet mengenai gempa bumi Chile algoritma SVM memiliki tingkat akurasi paling tinggi yaitu 81,2% dan pada data tweet mengenai kemerdekaan Catalan algoritma RF memiliki akurasi paling tinggi sebesar 85,8%. Dan pada penelitian analisis sentimen pada data tweet mengenai Go-Pay menggunakan algoritma NB yang dipublikasikan pada (Wisnu et al., 2020) diketahui bahwa algoritma NB dapat mengklasifikasikan analisis sentimen dengan baik dengan akurasi sebesar 70,71%.

METHODOLOGI

Pada penelitian ini dilakukan eksperimen membandingkan kinerja algoritma *Support Vector Machine* (SVM), *Naïve Bayes* (NB), dan *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan atau memprediksi kelas dari data baru berdasarkan model yang dilatih dengan data lama. Tahapan langkah yang dilakukan dalam penelitian meliputi pengumpulan data dengan teknik *scraping*, *pre processing* data, pelabelan kelas data, resampling data, pembobotan TF-IDF, klasifikasi dengan algoritma SVM, NB, KNN, dan evaluasi hasil eksperimen. Pada penelitian ini alat yang digunakan yaitu python dan r dengan google collaboratory.

Rincian dari tahapan dalam penelitian ini yaitu sebagai berikut, pada tahap pengumpulan data dilakukan dengan teknik *scraping* dengan bantuan library *snsrape*. Pada tahap pra pemrosesan dilakukan penyiapan data agar dapat diolah oleh komputer dan dapat menghasilkan akurasi yang lebih tinggi. Setelah itu dilakukan pelabelan kelas data kedalam 2 kelas yaitu negatif dan positif. Pelabelan kelas ini dilakukan karena tahap klasifikasi membutuhkan variabel kelas sebagai prediktor, yang digunakan untuk melatih model dan menguji model. Pada tahap resampling dilakukan untuk menyetarakan jumlah data berdasarkan variabel kelas. Pada tahap pembobotan TF-IDF dilakukan untuk mengubah data yang berbentuk teks menjadi angka agar dapat diolah oleh komputer pada tahap klasifikasi. Untuk tahap klasifikasi dilakukan untuk membentuk model machine learning yang dapat memprediksi kelas dari data baru berdasarkan training model dengan data training dan menguji kinerja dari model dengan data testing. Gambar 1 menyajikan gambar flowchart dari tahapan penelitian.



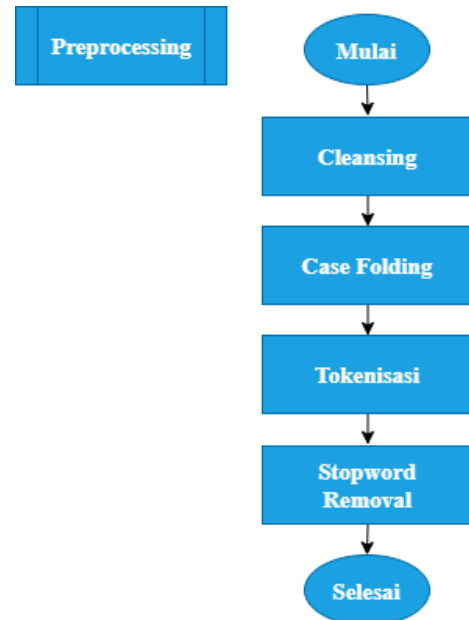
Gambar 1 Flowchart Tahapan Penelitian

Pengumpulan Data

Pada tahap pengumpulan data dilakukan dengan teknik *scraping* menggunakan python dengan bantuan library *snsrape*. Data yang diambil merupakan tweet masyarakat Indonesia dengan kata kunci ekonomi Indonesia, yang dibuat di masa pandemi Covid-19 2020 di Indonesia. Rentang waktu pengambilan data yaitu 1 Maret 2020 hingga 31 Desember 2020, dimana tanggal 2 Maret 2020 merupakan tanggal kasus Covid-19 yang pertama terjadi di Indonesia. Data yang didapatkan dari *scraping* ini sebanyak 116.018. Pengumpulan data ini dilakukan pada 26 Maret 2021. Dari dataset yang didapatkan, diketahui bahwa data ke 23.684 hingga data ke 116.018 sebanyak 92.334 data dibuat sebelum Indonesia resmi dinyatakan resesi. Dan dari 92.334 data tersebut diketahui bahwa terdapat 5.525 data membahas mengenai resesi.

Pre Processing Data

Pre processing data merupakan proses penyiapan dan membersihkan data mentah menjadi data siap pakai. Pada tahap eksperimen ini dilakukan penyiapan data dimana data yang digunakan berupa teks dan memiliki beberapa permasalahan seperti duplikat data, noise, dan data tidak lengkap. Oleh karena itu data harus diolah kembali agar dapat digunakan untuk diproses pada tahap klasifikasi dan mendapatkan hasil akurasi model yang lebih baik (Pravina et al., 2019). Gambar 2 menyajikan *flowchart* tahap eksperimen *pre processing* data.



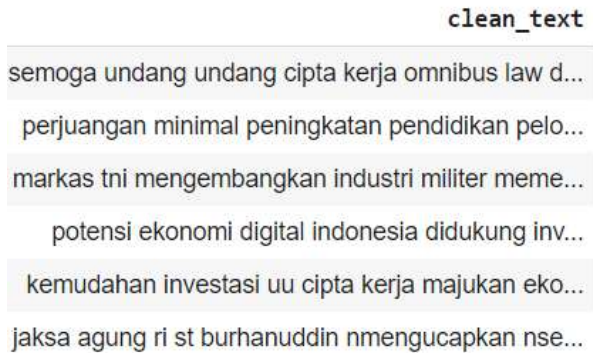
Gambar 2 Flowchart Preprocessing

Berikut rincian dari tahap eksperimen ini dan langkah yang dilakukan yaitu :

1. *Cleansing*, pada data yang baru saja diperoleh dari *scraping* twitter ada yang terdapat unsur yang tidak terpakai untuk digunakan pada tahap penelitian selanjutnya seperti hashtag, url, dan tanda baca. Sebagai contoh digunakan teks awal yang belum melalui tahap *pre processing* yaitu “Banyak cara untuk membangkitkan kembali ekonomi Indonesia. #KembangkanEkonomiDigital”. Dan setelah melalui tahap *cleansing* teks menjadi “Banyak cara untuk membangkitkan kembali ekonomi Indonesia KembangkanEkonomiDigital”.
2. *Case folding*, pada tahap ini dilakukan penyetaraan dimana data masih ada yang pada kalimatnya terdapat huruf besar dan kecil. Sehingga harus disamakan standar dari huruf dengan mengubah menjadi huruf kecil. Dan setelah contoh kalimat pada tahap *cleansing* melalui tahap ini maka kalimatnya menjadi “banyak cara untuk membangkitkan kembali ekonomi indonesia kembangkanekonomidigital”.
3. *Tokenizing*, pada tahap ini dilakukan memotong kata yang ada pada kalimat menjadi kata yang terpisah antara satu dengan yang lainnya. Sehingga setiap kata pada kalimat

dapat di cek satu persatu pada langkah berikutnya. Berikut contoh kalimat pada tahap *case folding* yang sudah melalui tahap ini “ ‘banyak’ ‘cara’ ‘untuk’ ‘membangkitkan’ ‘kembali’ ‘ekonomi’ ‘Indonesia’ ‘kembangkanekonomidigital’ ”.

4. *Stopword removal*, pada tahap ini telah disediakan *stopword* yang berisi kata tidak bermakna. Dan apabila kata dalam data cocok dengan yang ada di *stopword* maka di hapus pada tahap ini. Dan hasil akhir dari contoh kalimat yang di *pre processing* menjadi seperti ini “bangkit ekonomi indonesia kembangkan ekonomi”. Gambar 3 menyajikan hasil dari tahap eksperimen *pre processing*.

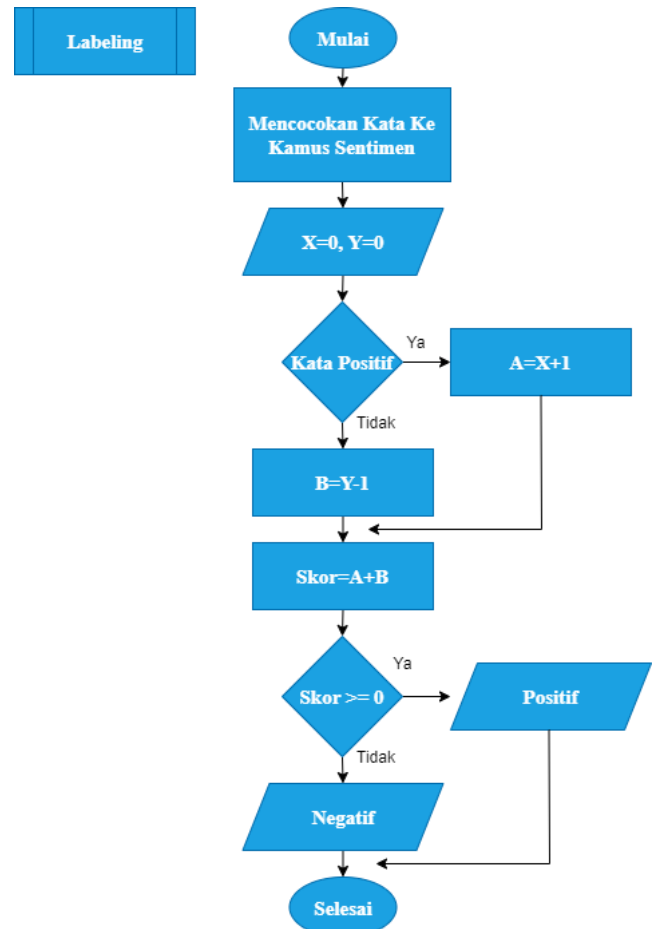


Gambar 3 Screenshot Hasil Tahap *Pre Processing*.

Pelabelan Kelas Data

Pada tahap ini data yang sudah bersih dan hanya memiliki kata yang bermakna diolah dengan bahasa r di *google collaboratory*. Data kemudian dicocokkan ke dalam kamus korpus bahasa Indonesia sentimen negatif dan positif yang disiapkan sebelumnya dalam bentuk csv. Setiap kata yang cocok pada tweet dengan corpus sentimen positif maka diberi nilai 1, jika ada kata pada tweet yang cocok dengan corpus sentimen negatif maka diberi nilai negatif 1, dan jika tidak ditemukan kata tersebut di korpus negatif dan positif maka diberi nilai 0. Kemudian dilakukan penghitungan skor dengan menjumlahkan semua nilai dari kata. Jika skor bernilai <0 maka tweet dilabeli negatif dan jika jumlah skor ≥ 0 maka tweet dilabeli positif (Ulfah & Anam, 2020). Proses pelabelan otomatis ini memiliki kelebihan dan kekurangan sendiri. Dengan pelabelan otomatis ini

tingkat keakuratan pemberian label pada setiap teks lebih kecil. Dan pelabelan otomatis ini memiliki kelebihan untuk mempermudah melabeli data dengan jumlah yang banyak. Gambar 4 flowchart alur pelabelan data dan gambar 5 screenshot hasil tahap pelabelan data.



Gambar 4 Flowchart Tahap Pelabelan Data

	score	clean_text	label_data
0	6	semoga undang undang cipta kerja omnibus law d...	Positif
1	4	perjuangan minimal peningkatan pendidikan pelu...	Positif
2	6	markas tni mengembangkan industri militer meme...	Positif
3	3	potensi ekonomi digital indonesia didukung inv...	Positif
4	4	kemudahan investasi uu cipta kerja majukan eko...	Positif
5	4	jaksa agung ri st burhanuddin nmengucapkan nse...	Positif
6	1	ekonomi indonesia diyakini bangkit 2021	Positif
7	7	iklim ramah investasi kemudahan berusaha uu ci...	Positif
8	2	berkat uu cipta kerja investor global tertarik...	Positif
9	2	selamat nsmg ekonomi bangsa indonesia lg nsmg ...	Positif
10	-1	pemerintah menargetkan indonesia xc2 xa0masuk ...	Negatif
11	-1	mi instan indonesia alat transaksi seks ghana ...	Negatif

Gambar 5 Screenshot Hasil Pelabelan Kelas Data

Resampling

Pada tahap ini dilakukan menyeimbangkan jumlah data antar kelas label jika tidak seimbang (Wijanarto & Brilianti, 2020) dan memang ketidak seimbangannya terlalu jauh. *Resampling* sendiri memiliki tujuan agar data yang di training tidak memiliki kecenderungan untuk memprediksi kepada salah satu kelas yang jumlah datanya lebih banyak. Pada tahap ini terdapat dua cara untuk menyelesaikan masalahnya. Pertama data dari kelas yang jumlahnya sedikit ditambah sampai setara dengan kelas yang datanya lebih banyak yaitu *up sampling*. Dan pemangkasan data yang berlebih dari suatu kelas yaitu *under sampling* (Pujianto, 2016). *Up sampling* dilakukan jika jumlah data yang diolah sedikit sehingga mempengaruhi pada model yang dibuat dan begitu pula sebaliknya pada *under sampling*. Dilakukan atau tidak dilakukannya *resampling* berdasarkan hasil perbandingan jumlah data berdasarkan label kelas data. Setelah melakukan tahap pelabelan data diketahui jumlah data kelas positif sebanyak 65.450 dan data kelas negatif berjumlah 33.466. Sehingga pada penelitian ini dilakukan tahap eksperimen *resampling* dengan cara *under sampling* karena jumlah data yang banyak, dan perbedaan jumlah data pada kelas negatif dan positif yang jauh yaitu di atas 30%.

Pembobotan TF-IDF

Agar data yang digunakan dapat diolah oleh mesin dan di klasifikasi dengan algoritma maka dilakukan pembobotan dengan TF-IDF (*Term Frequency-Inverse Document Frequency*). Pembobotan TF-IDF merupakan ekstraksi fitur dimana dataset tweet yang

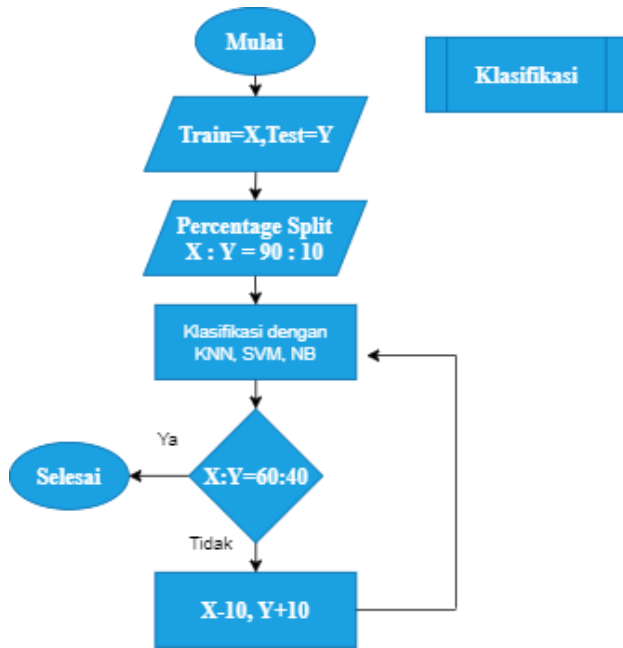
digunakan berupa kalimat. Oleh karena itu harus ada pengubahan dari setiap kata dalam bentuk angka, sehingga mesin dapat mengolah data yang digunakan (Oyebode & Orji, 2019). Pemberian nilai pada setiap kata di tahap ini dilakukan dengan berdasar pada seberapa sering kata tersebut muncul pada dataset (Pravina et al., 2019).

Klasifikasi

Pada eksperimen ini klasifikasi dilakukan dengan algoritma SVM, NB, dan KNN. Pada tahap ini juga dilakukan menggunakan tiga skenario pemisahan data antara data train dan data uji dan juga digunakan metode *cross validation*. Hal ini dilakukan supaya dapat menemukan model dan skenario terbaik untuk mengklasifikasikan dataset.

HASIL DAN DISKUSI

Pada penelitian ini dilakukan 3 skenario percentage split pada tahap klasifikasi. Percentage split merupakan pemisahan data train dan tes berdasarkan perbandingannya, contohnya 80:20 yang berarti data train yang digunakan sebanyak 80% dan 20% nya sebagai data test. Pada eksperimen tahap klasifikasi ini dilakukan percentage split dengan 3 skenario terbaik dan paling banyak digunakan berdasarkan studi pustaka yang telah dilakukan. Percentage split 90:10, 80:20, dan 70:30 merupakan tiga skenario percentage split yang memiliki hasil kinerja paling baik dalam pembentukan model. Gambar 6 menunjukkan flowchart tahap klasifikasi dengan 3 skenario percentage split.



Gambar 6 Flowchart Klasifikasi Dengan 3 Skenario Percentage Split.

Selain percentage split pada tahap klasifikasi juga digunakan metode cross validation 10 fold. Cross validation merupakan metode untuk merotasi urutan langkah train dan test data. Contohnya pada cross validation 10 fold maka dilakukan proses sebanyak k^2 yaitu 100 kali proses yang terdiri dari proses test data sebanyak k fold yaitu 10 kali dan sisanya untuk train data. Pada cross validation 10 fold dilakukan k proses pada setiap iterasi hingga iterasi ke k, dan test data dilakukan di proses pertama pada iterasi pertama kemudian dilakukan di proses ke dua pada iterasi ke dua hingga iterasi ke k. Pada eksperimen ini digunakan cross validation k=10 atau 10 fold karena diketahui dari hasil studi pustaka bahwa menghasilkan akurasi yang paling besar. Tabel 1 menunjukkan cara kerja cross validation 10 fold.

Tabel 1 Cara Kerja Cross Validation 10 Fold (TS=Tes dan TR=Train)

TS	TR	TR	TR	TR	TR	TR	TR	TR	TR
TR	TS	TR	TR	TR	TR	TR	TR	TR	TR
TR	TR	TS	TR	TR	TR	TR	TR	TR	TR
TR	TR	TR	TS	TR	TR	TR	TR	TR	TR
TR	TR	TR	TR	TS	TR	TR	TR	TR	TR
TR	TR	TR	TR	TR	TS	TR	TR	TR	TR

TR	TR	TR	TR	TR	TR	TS	TR	TR	TR
TR	TR	TR	TR	TR	TR	TR	TS	TR	TR
TR	TR	TR	TR	TR	TR	TR	TR	TS	TR
TR	TR	TR	TR	TR	TR	TR	TR	TR	TS

Hasil evaluasi dari model dengan setiap skenario yang dilakukan tentu saja berbeda. Kinerja masing masing model yang terbentuk dari hasil training didapatkan dari hasil uji dengan data test. Pengujian terhadap setiap model dilakukan menggunakan konsep confusion matrix. Pada confusion matrix terdapat istilah TP (*True Positif*), TN (*True Negatif*), FP (*False Positif*), dan FN (*False Negatif*). TP berarti hasil prediksi kelas positif yang dinyatakan prediksinya benar dengan kelas aslinya, dan begitu juga pada TN untuk prediksi kelas negatif yang dinyatakan prediksinya benar sesuai kelas aslinya. Kemudian FP adalah hasil prediksi kelas positif yang dinyatakan salah prediksi atau tidak sesuai dengan kelas aslinya, dan begitu pula untuk FN pada kelas negatif. Setelah didapatkan jumlah dari TP, TN, FP, dan FN maka barulah model dapat diketahui tingkat akurasi. Karena akurasi dari model adalah bagaimana model memprediksi keseluruhan dokumen secara benar.

Dari eksperimen pada tahap klasifikasi yang dilakukan dengan algoritma SVM, NB, dan KNN. Diketahui beberapa rincian kinerja dari model dengan percentage split 90:10 dan cross validation 10 fold. Bahwa untuk algoritma NB memiliki nilai akurasi 81,55%, untuk algoritma SVM memiliki nilai akurasi 95,53%, dan untuk algoritma KNN memiliki nilai akurasi 74,30%. Gambar 7 menunjukkan kinerja model dengan percentage split 90:10.

	precision	recall	f1-score	support
0	0.7847	0.8670	0.8238	3330
1	0.8531	0.7646	0.8064	3364
accuracy			0.8155	6694
macro avg	0.8189	0.8158	0.8151	6694
weighted avg	0.8191	0.8155	0.8151	6694
0.8155064236629818				
[[2887 443]				
[792 2572]]				
	precision	recall	f1-score	support
0	0.9592	0.9527	0.9559	3404
1	0.9514	0.9581	0.9547	3290
accuracy			0.9553	6694
macro avg	0.9553	0.9554	0.9553	6694
weighted avg	0.9554	0.9553	0.9553	6694
0.9553331341499851				
[[3243 161]				
[138 3152]]				
	precision	recall	f1-score	support
0	0.7599	0.7079	0.7330	3335
1	0.7285	0.7779	0.7524	3359
accuracy			0.7431	6694
macro avg	0.7442	0.7429	0.7427	6694
weighted avg	0.7441	0.7431	0.7427	6694
0.743053480729011				
[[2361 974]				
[746 2613]]				

Gambar 7 Kinerja Model Percentage Split 90:10 dan CV 10 Fold

Dari eksperimen pada tahap klasifikasi, diketahui beberapa rincian kinerja dari model dengan percentage split 80:20 dan cross validation 10 fold. Bahwa untuk algoritma NB memiliki nilai akurasi 80,76%, untuk algoritma SVM memiliki nilai akurasi 80,74%, untuk algoritma SVM memiliki nilai akurasi 95,47%, dan untuk algoritma KNN memiliki nilai akurasi 73,71%. Gambar 8 kinerja model dengan percentage split 80:20.

	precision	recall	f1-score	support
0	0.7800	0.8581	0.8172	6715
1	0.8412	0.7564	0.7966	6672
accuracy			0.8074	13387
macro avg	0.8106	0.8073	0.8069	13387
weighted avg	0.8105	0.8074	0.8069	13387
0.8074251139164862				
[[5762 953]				
[1625 5047]]				
	precision	recall	f1-score	support
0	0.9541	0.9547	0.9544	6641
1	0.9554	0.9548	0.9551	6746
accuracy			0.9547	13387
macro avg	0.9547	0.9547	0.9547	13387
weighted avg	0.9547	0.9547	0.9547	13387
0.9547322028833943				
[[6340 301]				
[305 6441]]				
	precision	recall	f1-score	support
0	0.7451	0.7130	0.7287	6628
1	0.7300	0.7608	0.7451	6759
accuracy			0.7371	13387
macro avg	0.7375	0.7369	0.7369	13387
weighted avg	0.7375	0.7371	0.7370	13387
0.7371330395159483				
[[4726 1902]				
[1617 5142]]				

Gambar 8 Kinerja Model Percentage Split 80:20 dan CV 10 Fold

Dari eksperimen pada tahap klasifikasi, diketahui beberapa rincian kinerja dari model dengan percentage split 70:30 dan cross validation 10 fold. Bahwa untuk algoritma NB memiliki nilai akurasi 80,76%, untuk algoritma SVM memiliki nilai akurasi 94,90%, dan untuk algoritma KNN memiliki nilai akurasi 73,92%. Gambar 9 kinerja model dengan percentage split 70:30.

	precision	recall	f1-score	support
0	0.7812	0.8525	0.8153	9999
1	0.8391	0.7632	0.7994	10081
accuracy			0.8077	20080
macro avg	0.8102	0.8079	0.8073	20080
weighted avg	0.8103	0.8077	0.8073	20080
0.8076693227091634 [[8524 1475] [2387 7694]]				
	precision	recall	f1-score	support
0	0.9504	0.9471	0.9488	9998
1	0.9477	0.9510	0.9494	10082
accuracy			0.9491	20080
macro avg	0.9491	0.9490	0.9491	20080
weighted avg	0.9491	0.9491	0.9491	20080
0.9490537848605578 [[9469 529] [494 9588]]				
	precision	recall	f1-score	support
0	0.7504	0.7130	0.7312	9987
1	0.7294	0.7653	0.7469	10093
accuracy			0.7393	20080
macro avg	0.7399	0.7392	0.7391	20080
weighted avg	0.7398	0.7393	0.7391	20080
0.7392928286852589 [[7121 2866] [2369 7724]]				

Gambar 9 Kinerja Model Percentage Split 70:30 dan CV 10 Fold.

KESIMPULAN

Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa algoritma yang memiliki kinerja paling baik dari hasil cross validation 10 fold adalah Support Vector Machine (SVM) dengan percentage split 90:10. Hal ini ditunjukkan bahwa algoritma SVM dengan percentage split 90:10 memiliki tingkat akurasi model yang paling tinggi yaitu 95,53%. Dan secara keseluruhan dari ketiga skenario, percentage split 80:20 memiliki nilai akurasi model yang paling tinggi.

DAFTAR PUSTAKA

Andika, L. A., Azizah, P. A. N., & Respatiwan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan

- Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1), 34. <https://doi.org/10.13057/ijas.v2i1.29998>
- Febrianti, Y. M., Indriati, I., & Widodo, A. W. (2018). Analisis Sentimen Pada Ulasan “Lazada ” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3689–3698. <http://j-ptiik.ub.ac.id>
- Kusuma, H. (2020). *Indonesia Resmi Resesi! Ekonomi Kuartal III-2020 Minus 3,49%*. <https://finance.detik.com>
- Lengkok, N. C., Safitri, O., Machsus, S., Putra, Y. R., Syahadati, A., & Nooraeni, R. (2021). Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg. *Jurnal Teknoinfo*, 15(1), 20. <https://doi.org/10.33365/jti.v15i1.866>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(11), 4704–4713.
- Oyebode, O., & Orji, R. (2019). Social Media and Sentiment Analysis: The Nigeria Presidential Election 2019. *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019, October*, 140–146. <https://doi.org/10.1109/IEMCON.2019.8936139>
- Pajri, D., Umaidah, Y., & Padilah, T. N. (2020). K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(2), 242–253. <https://doi.org/10.28932/jutisi.v6i2.2658>
- Perekonomian, K. K. B. (2016). *Kementerian Koordinator Bidang Perekonomian*. <https://www.ekon.go.id/>
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan*

- Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 3(3), 2789–2797.
- Pujianto, U. (2016). Strategi Resampling Berbasis Centroid Untuk Menangani Lunak. *Tekno*, 25(Maret), 1–6.
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106(January), 92–104. <https://doi.org/10.1016/j.future.2020.01.005>
- Sadikin, M., & Alfiandi, F. (2018). Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(6), 4763. <https://doi.org/10.11591/ijece.v8i6.pp4763-4771>
- Ulfah, A. N., & Anam, M. K. (2020). Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM). *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(1), 1–10. <https://doi.org/10.35957/jatisi.v7i1.196>
- Wibowo, A. T., & Fitriah, D. (2018). a K-Nearest Algorithm Based Application To Predict Snmptn Acceptance for High School. *International Research Journal of Computer Science (IRJCS)*, 5(01), 9–20. <https://doi.org/10.26562/IRJCS.2018.JACS10083>
- Wijanarto, W., & Brilianti, S. P. (2020). Peningkatan Performa Analisis Sentimen Dengan Resampling dan Hyperparameter pada Ulasan Aplikasi BNI Mobile. *Jurnal Eksplora Informatika*, 9(2), 140–153. <https://doi.org/10.30864/eksplora.v9i2.333>
- Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes. *Journal of Physics: Conference Series*, 1444(1), 0–10. <https://doi.org/10.1088/1742-6596/1444/1/012034>
- Yamali, F. R., & Putri, R. N. (2020). Dampak Covid-19 Terhadap Ekonomi Indonesia. *Ekonomis: Journal of Economics and Business*, 4(2), 384. <https://doi.org/10.33087/ekonomis.v4i2.179>