

## Implementasi Algoritma *Support Vector Machine* dan *Model Bag-of-Words* dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna *Twitter*

Raja Farhan Ramadha Pohan<sup>1</sup>, Dian Eka Ratnawati<sup>2</sup>, Issa Arwani<sup>3</sup>

Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>rajafarhan1878@gmail.com, <sup>2</sup>dian\_ilkom@ub.ac.id, <sup>3</sup>issa.arwani@ub.ac.id

### Abstrak

Media sosial adalah *platform digital* yang digunakan untuk mengumpulkan informasi. Salah satu contoh media sosial di Indonesia dengan jumlah pengguna paling banyak adalah *Twitter*. *Twitter* merupakan salah satu *platform* media sosial berbentuk *microblog*, yang memiliki fungsi utama untuk bertukar opini dan berbagi ide. Salah satu topik yang sempat ramai diperbincangkan di *Twitter* adalah pelaksanaan Pilkada 2020. Pilkada 2020 ini mengundang perhatian banyak pengguna *Twitter* karena pelaksanaannya dilakukan ditengah pandemi. *Tweets* atau cuitan para pengguna *Twitter* disini dijadikan objek untuk melakukan analisis sentimen terhadap topik terkait. Proses analisis sentimen ini dilakukan dengan menerapkan algoritma *Support Vector Machine* untuk melakukan pengelompokan *tweets* positif dan negative. Penelitian ini menggunakan 400 *tweets* yang terdiri dari 187 *tweets* positif dan 213 *tweets* negatif. Nilai *accuracy*, *precision*, *recall*, dan *f1-score* pada model *machine learning* ini secara berturut-turut adalah 87.5%, 87.4%, 87.4%, dan 87.4%. Nilai ini didapatkan dengan menggunakan parameter C sebesar 5, dan parameter *max\_iter* sebesar 100.

**Kata kunci:** *analisis sentimen, pilkada 2020, pandemi, support vector machine, bag-of-words*

### Abstract

*Social media is a digital platform used to gather information. One of the social media in Indonesia that has the most users is Twitter. Twitter is a social media platform in a form of a microblog, that allows users to exchange opinions and share ideas. One of the topics that was widely discussed on Twitter was the implementation of the Pilkada 2020. Pilkada 2020 attracted the attention of many Twitter users because the implementation was carried out in the midst of a pandemic. Tweets or tweets of Twitter users here are used as objects to conduct sentiment analysis on related topics. This analysis process is carried out by implementing the Support Vector Machine algorithm to classify positive and negative tweets. This study uses 400 tweets, consisting of 187 positive tweets and 213 negative tweets. The scores of accuracy, precision, recall, and f1-score in this machine learning model are 87.5%, 87.4%, 87.4%, and 87.4%, respectively. This value is obtained using the C parameter of 5, and the max\_iter parameter of 100.*

**Keywords:** *sentiment analysis, pilkada 2020, pandemic, support vector machine, bag-of-words,*

## 1. PENDAHULUAN

Media sosial merupakan media interaksi sosial antar individu yang dilakukan secara daring, dan memungkinkan setiap orang untuk saling berkomunikasi tanpa dibatasi ruang dan waktu. Dengan munculnya media sosial, tiap individu dapat saling berkomunikasi kapan saja dan dimana saja, dan media sosial telah memberikan dampak yang sangat besar bagi kehidupan kita di era ini. Melalui media sosial, kita dapat menggunakan akses internet untuk secara bebas mengunggah dan mengunduh semua konten yang ada akun media sosial.

Tetapi harus diingat bahwa semua jenis media sosial mempunyai dampak positif dan negatif dalam penggunaannya. Fungsi media sosial di era ini tidak hanya sekedar memberikan fungsi komunikasi, namun fungsi media sosial telah menjelma menjadi media iklan, media komersial, atau portal berita.

Dalam media sosial, setiap individu dapat menyampaikan opininya terhadap sebuah isu. Bentuk media sosial yang paling banyak digunakan untuk melakukan hal ini disebut *microblog*. *Microblog* adalah bentuk media sosial yang sering digunakan sebagai wadah

bertukar ide dalam bentuk teks, gambar, atau video. *Twitter* merupakan salah satu dari berbagai macam media sosial dengan jumlah pengguna paling banyak. Saat ini, *Twitter* sering digunakan sebagai media untuk mendiskusikan hal yang sedang hangat diberitakan. Salah satunya adalah Pemilihan Umum Kepala Daerah dan Wakil Kepala Daerah atau yang lebih kita kenal dengan Pilkada. Pilkada adalah sebuah proses Pemilihan Umum untuk memilih calon Kepala Daerah dan Wakil Kepala Daerah yang diusulkan oleh sebuah partai politik atau gabungan dari berbagai partai politik. Topik ini sempat menjadi bahan pembicaraan pengguna *Twitter* karena kegiatan ini dilaksanakan ditengah pandemi. Hal ini menarik perhatian warga *Twitter* untuk beropini dikarenakan angkapenderita COVID-19 di Indonesia belum bisa dibilang aman. Untuk menentukan dan mengklasifikasikan opini pengguna *Twitter* terhadap topik ini, dapat menggunakan *analisis sentimen*. *Analisis sentimen* adalah proses untuk mengekstrak dan mengolah data tekstual secara otomatis untuk mengetahui nilai sentimen dari sebuah kalimat. (Dahlan, Pramono, & Rozi, 2012). Pada studi kasus ini, hasil analisis sentimen diharapkan dapat menjadi acuan bagi entitas tertentu dalam pengambilan keputusan yang melibatkan masyarakat luas.

Klasifikasi sentimen yang didapatkan berupasentimen positif dan negatif. Teknik analisis ini menggunakan *Natural Language Processing* (NLP) dan pada masa kini, analisis sentimen banyak membantu entitas tertentu untuk mengumpulkan dan mengungkap opini masyarakat terhadap sebuah isu, kepuasan pelayanan atas sebuah pelayanan, peraturan yang berlaku, dan lain-lain, yang didasari oleh data tekstual.

Analisis sentimen dapat membantu proses perubahan informasi berbentuk data tekstual yang tadinya belum terstruktur menjadi lebih terstruktur. Untuk melakukan pengolahan data tekstual, sebuah model analisis sentimen mengkombinasikan metode *Natural Language Processing* dan *machine learning* untuk menentukan nilai sentimen dari sebuah data tekstual. Algoritma yang akan digunakan pada model *machine learning* ini adalah *Support Vector Machine*, Penggunaan algoritma *Support*

*Vector Machine* dirasa tepat karena algoritma ini masih merupakan salah satu algoritma terbaik untuk melakukan proses klasifikasi, hal ini dibuktikan berdasarkan penelitian yang dilakukan oleh Arifin & Sasongko (2018) yang melakukan komparasi antara *Support Vector Machine* dan *Naïve Bayes*. Hasil penelitian ini menunjukkan algoritma *Support Vector Machine* memiliki nilai akurasi sebesar 96.88%, *precision* 98.62%, dan *recall* 94.95%. *Naïve Bayes* memiliki nilai akurasi sebesar 88.87%, *precision* 89.63%, dan *recall* 87.69%. Selain itu, berdasarkan penelitian Komparasi Algoritma *Naïve Bayes* dan *Support Vector Machine* Untuk Analisa Sentimen Review Film yang dilakukan oleh Indrayuni (2018), Pengujian dilakukan menggunakan *cross validation sebanyak 10 fold*. Nilai akurasi dari penelitian ini adalah 84.50% untuk algoritma *Naïve Bayes*. Dan 90.00% untuk algoritma *Support Vector Machine*.

Selain algoritma *Support Vector Machine*, pada penelitian diterapkan model *bag-of-words* yang digunakan untuk menemukan dan menghitung kata yang paling sering di *tweet* oleh pengguna *Twitter* terhadap topik terkait. Model *bag-of-words* merupakan representasi sederhana dari sebuah teks yang digunakan pada *natural language processing* untuk menghitung jumlah kemunculan sebuah kata pada sebuah dokumen, yang juga dikenal dengan model *vector space* (McTear, Callejas, & Griol, 2016).

Berdasarkan latar belakang yang sudah dipaparkan, rumusan masalah yang dapat ditarik dari penelitian ini adalah untuk merancang model machine learning yang menerapkan algoritma *Support Vector Machine* dan *Model Bag-of-Words* untuk menentukan sentimen pengguna *Twitter* terhadap topik terkait, dari sinilah dapat ditentukan tujuan dari penelitian ini, yaitu Merancang model sebuah model machine learning yang menerapkan algoritma *Support Vector Machine* dan *Model Bag-of- Words* untuk menentukan sentimen pengguna *Twitter* terhadap topik terkait Menentukan tingkat akurasi algoritma *Support Vector Machine* untuk melakukan proses klasifikasi *tweets*, dan diharapkan penelitian bisa sebagai sebuah gambaran atas dampak dari sebuah kebijakan yang dibuat oleh entitas tertentu, serta juga apat

dijadikan dasar evaluasi oleh entitas terkait dalam pembuatan kebijakan yang melibatkan masyarakat luas.

## 2. LANDASAN KEPUSTAKAAN

### 2.1. Twitter

Twitter merupakan sebuah platform media sosial berbentuk *microblog* yang memiliki fitur untuk bertukar pesan yang berbentuk teks, gambar, dan video. Karena fiturnya ini, Twitter dirasa tepat sebagai tempat untuk menyuarakan gagasan, opini, dan mendapatkan informasi terbaru, untuk mencari ide, atau untuk melihat apa yang sekarang sedang menjadi tren. (Brogan, 2010). Karena popularitasnya sebagai media sosial berbentuk *microblog*, Twitter saat ini banyak dimanfaatkan untuk berbagai macam kebutuhan, contohnya adalah sarana berbagi opini, penyampaian protes terhadap sebuah

2020). Pilkada digelar oleh Komisi Pemilihan Umum (KPU) tingkat provinsi dan kabupaten/kota dan diawasi langsung oleh Badan Pengawas Pemilu (Bawaslu).

### 2.3. Analisis Sentimen

Analisis sentimen merupakan sebuah metode untuk menentukan nilai sentimen atas sebuah objek. Analisis sentimen dapat digunakan untuk membedakan teks menjadi dua kategori, yakni fakta dan opini. Fakta merupakan ungkapan objektif terhadap suatu hal, yang mana opini adalah ungkapan subjektif yang mendeskripsikan sentimen dan perasaan terhadap suatu hal. Opini yang dikumpulkan menggunakan teknik ini kemudian dapat di klasifikasikan menjadi tiga golongan, yaitu positif, negatif, dan netral.

### 2.4. Terms Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency merupakan sebuah metode yang digunakan untuk bobot masing-masing dari setiap kata yang ada pada sebuah dokumen, dan merupakan metode yang paling sering digunakan dalam proses *information retrieval*. Metode ini juga dikenal mudah untuk diimplementasikan dan memiliki hasil akhir yang baik. Metode ini bekerja dengan

cara menghitung nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* pada setiap kata yang ada pada dokumen. (Hidayat, 2016).

*Term Frequency* merupakan jumlah kemunculan sebuah kata *i* pada sebuah dokumen *j*, dibagi dengan semua *term* pada dokumen *j*. (Yunus, 2020). Ditulis dalam persamaan(1)

$$Term\ Frequency = \frac{Term\ i\ pada\ dokumen\ j}{Total\ term\ pada\ dokumen\ j} \quad (1)$$

### 2.2. Pemilihan Kepala Daerah

Pemilihan Umum Kepala Daerah dan Wakil Kepala Daerah atau biasa disebut dengan Pilkada adalah tahapan Pemilihan Umum untuk memilih bakal calon Kepala Daerah yang diusulkan oleh sebuah partai politik atau gabungan dari beberapa partai politik. Pilkada merupakan sebuah pemilihan yang dilakukan secara langsung oleh para penduduk yang telah memenuhi syarat dan ketentuan. (Simkada,

*Inverse Document Frequency* dipergunakan untuk mengurangi bobot sebuah kata, jika kata tersebut muncul terlalu sering pada sebuah dokumen. (Yunus, 2020). Ditulis dalam persamaan (2)

$$Inverse\ Document\ Frequency = \frac{Total\ dokumen}{log(2)} \quad Total\ dokumen\ yang\ memiliki\ term\ i$$

Nilai akhir dari *TF-IDF* diperoleh dari hasil perkalian antara nilai *Term Frequency (TF)* dan

$x^T$  = Transpose dari nilai  $x$

(5) *Inverse Document Frequency (IDF)* yang ditulis  $j^i$  pada persamaan (3)

$$TF-IDF = \frac{Term\ i\ pada\ dokumen\ j}{log\ Total\ dokumen}$$

Total dokumen yang memiliki term  $i$

2.5. Support Vector Machine

$x$

(3)

Nilai pada  $y$  didapatkan dari Persamaan 6

$$\sum_{i=1, j=1}^n y_i y_j^T, (i, j = 1, \dots, n) \quad (6)$$

Memberi nilai *bias* sebesar 1 pada hasil perhitungan dari nilai  $x$  dan  $y$ . Hal ini dilakukan agar bisa mendapatkan jarak tegak lurus yang optimal.

Support Vector Machine merupakan salah satu algoritma *machine learning* yang tergolong kedalam *supervised learning* dan pertama kali dipopulerkan oleh Boser, Guyon, & Vapnik di tahun 1992. Pada umumnya, Support Vector Machine digunakan untuk melakukan klasifikasi dan regresi. Support Vector Machine adalah algoritma *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM), yang bertujuan untuk mencari *hyperplane* terbaik yang menjadi pemisah kedua buah *class* pada sebuah *input space*. (Nugroho, Witarto, & Handoko, 2003). *Hyperplane* pada algoritma

5. Mencari nilai dari  $a_i$  dengan menggunakan Persamaan 7, yang dilanjutkan dengan substitusi antara nilai hasil perhitungan dengan menggunakan Persamaan 8.

$$\sum_{i=1, j=1}^n a_i S_i^T S_j \quad (7)$$

$$\sum_{i=1, j=1}^n a_i S_i^T S_j = y_i \quad (8)$$

$a_i$  = lagrange multiplier

$S_i$  = nilai dari support vector bias

Gambar 1 Diagram Alir Proses Penelitian

$S^T$  = Transpose dari  $S$

SVM terletak ditengah-tengah objek dan berbeda  $i$

dengan kelas yang berdekatan. Dalam algoritma SVM, objek terluar yang paling mendekati *hyperplane*, disebut dengan *support vector*.

6. Mencari nilai dari  $\tilde{W}$  dengan menggunakan Persamaan 9, yaitu

(9)

Berikut adalah tahapan penerapan algoritma Support Vector Machine dalam analisis sentimen (Syukuran, 2016):

1. Setelah semua terms berhasil dinilai bobotnya menggunakan TF-IDF, tahap selanjutnya adalah melakukan perhitungan  $x_i x^T$  dimana  $i, j = 1, \dots, n$ .
2. Hasil perhitungan  $x_i x^T$  akan diubah menjadisebuah *matriks*.
3. Menghitung nilai  $y_i y^T$  dimana  $i, j = 1, \dots, n$ . Hasil perhitungan dari  $y_i y^T$  juga akan diubah menjadi sebuah *matriks*.
4. Mencari nilai dari  $a_i$  atau *lagrange multiplier* dengan cara mengubah setiap kalimat pada dokumen menjadi *support vector* = ( $x$ ). Yang selanjutnya bentuk *support vector* dari setiap kalimat dimasukkan ke dalam Persamaan 4:

Keterangan :

$\tilde{W}$  = *hyperlane*

$a_i$  = *lagrange multiplier*

$S_i$  = *support vector bias*

Mencari nilai  $b$  atau *hyperplane* dengan menggunakan Persamaan 10:

$$y = wx + b \quad (10)$$

Keterangan :

$w$  = *vector* bobot

$b$  = nilai *hyperplane*

$y$  = sentimen *tweets*

Sehingga pada akhirnya akan didapatkan nilai dari vector bobot dan *hyperplane* yang digunakan untuk melakukan klasifikasi.

$$S_i = \phi \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{x_n^2 + y_n^2} - x + (x - y)^2 \\ \sqrt{x_n^2 + y_n^2} - y + (x - y)^2 \end{bmatrix}$$

Nilai  $x$  didapat dari Persamaan 5, yaitu:

$$\tilde{W} = \sum_{i=1}^n a_i S_i$$

(4)

$$\sum_{i=1, j=1}^1 x_i x_j^T, (i, j = 1, \dots, n)$$

### 2.6. Bag-of-Words

Bag-of-Words merupakan sebuah model darisebuah proses yang ada didalam *Natural Language Processing*, dan banyak digunakan untuk mengambil nilai dari sebuah kata yang sebelumnya diolah pada sebuah model *machine learning*. Model *Bag-of-Words* bekerja dengan cara mempelajari sebuah kata dari pada sebuah dokumen, kemudian menginterpretasikan setiap dokumen dengan menghitung jumlah kemunculan tiap kata dari dokumen tersebut (Deepu, Pethuru, & Rajaraajeswari, 2016).

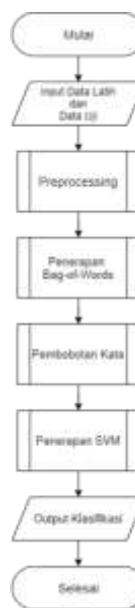
“saya”, “suka”, “film”, “horor”
“saya”, “suka”, “film”, “komedi”
“saya”, “suka”, “menonton”, “film”

Tabel 1 Sebelum penerapan model *Bag-of-Words*

saya	suka	film	horor	komedi	menonton
1	1	1	1	0	0
1	1	1	0	1	0
1	1	1	0	0	1

Tabel 2 Sesudah penerapan model *Bag-of-Words*

### 3. P E R A N C A N G A N



Gambar 1. Diagram alir proses analisis sentimen

Langkah pertama yang harus dilakukan adalah melakukan scraping *tweets*, yaitu pengambilan data yang berbentuk *tweets* menggunakan *library python* dan mengintegrasikannya dengan *API Twitter*. Parameter yang digunakan dalam pengambilan *tweets* adalah *keyword* terhadap topik terkait dan jangka waktu *tweets* yang ditentukan berdasarkan tanggal. Selanjutnya adalah proses *data labeling*, yaitu membuat kolom pada data yang sudah berhasil diambil dengan tujuan untuk menentukan polaritas dari sebuah tweet. Nilai polaritas disini digolongkan menjadi positif dan negatif.

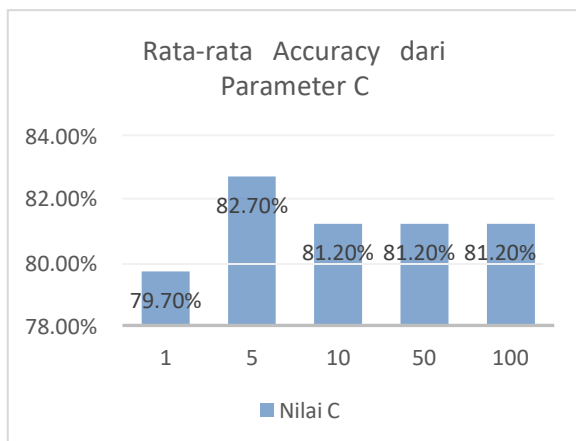
Selanjutnya adalah tahap *data preprocessing* yang terbagi menjadi lima proses, yaitu *case folding*, untuk merubah huruf menjadi huruf kecil, *data cleaning*, untuk membersihkan data dari karakter yang tidak memiliki nilai signifikan, *stopwords removal*, untuk menghapus kata yang tidak memiliki makna, *stemming*, untuk merubah kata menjadi bentuk dasar, dan *tokenization*, untuk memecah kalimat menjadi sebuah kata. Tujuan dari proses ini adalah untuk mendapatkan bentuk data yang awalnya tidak terstruktur, menjadi bentuk yang lebih terstruktur, dan dapat disesuaikan dengan kebutuhan proses penelitian selanjutnya

#### 4. PENGUJIAN DAN ANALISIS

Pada bagian ini akan dipaparkan hasil pengujian menggunakan algoritma *Support Vector Machine* dengan menggunakan parameter C dan parameter max\_iter, yang kemudian dilanjutkan dengan analisa model *Bag-of-Words*. Jumlah data yang digunakan pada pengujian terdiri dari 400 tweets, yang terbagi menjadi 187 tweets positif dan 213 tweets negatif.

##### 4.1. Pengujian Menggunakan Parameter C

Pada pengujian Parameter C, digunakan nilai C = 1, 5, 10, 50 dan 100. Hasil rata-rata Accuracy dari 5 kali percobaan untuk setiap nilai dari Parameter C adalah sebagai berikut:

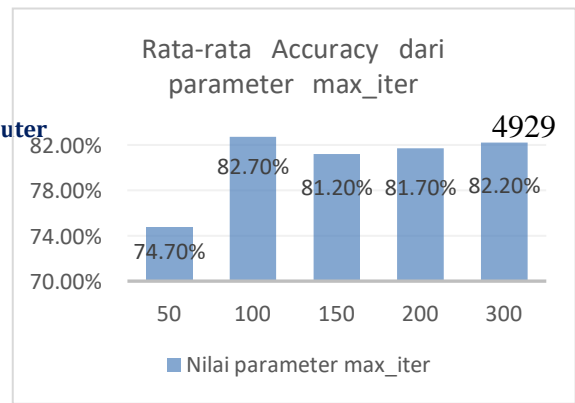


Gambar 2. Rata-rata accuracy dari parameter C

Berdasarkan Gambar 2, dapat dilihat bahwa nilai rata-rata akurasi tertinggi adalah 82.7%, dan nilai ini diperoleh saat nilai parameter C = 5, dan semakin tinggi nilai C, maka *hyperplane* akan semakin lebar dan nilai akurasi akan semakin turun dan stabil di angka 81.2%.

##### 4.2. Pengujian Menggunakan Parameter max\_iter

Pada pengujian parameter max\_iter, digunakan nilai max\_iter = 50, 100, 150, 200 dan 300, dan nilai C yang akan digunakan adalah 5.



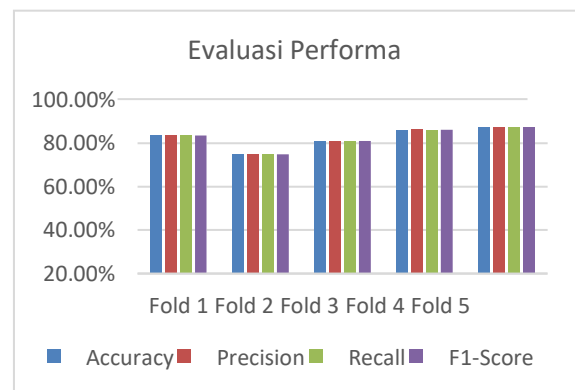
Berdasarkan Gambar 3, dapat dilihat bahwa rata-rata nilai akurasi tertinggi didapatkan saat

Gambar 3. Rata-rata accuracy dari parameter max\_iter

menggunakan parameter max\_iter sebesar 100, dan semakin tinggi nilai iterasi maksimum, maka nilai rata-rata akurasi akan semakin fluktuatif

##### 4.3. Evaluasi Performa Menggunakan Cross Validation

Pada sub-bab ini akan dilakukan pengujian menggunakan Cross Validation dengan *fold* = 5, yang artinya data akan diuji sebanyak 5 kali yang sudah dibagi menjadi 5 kelompok, dan Parameter yang digunakan adalah parameter C sebesar 5, dan parameter max\_iter sebesar 100, seperti yang sudah dijelaskan pada sub bab sebelumnya

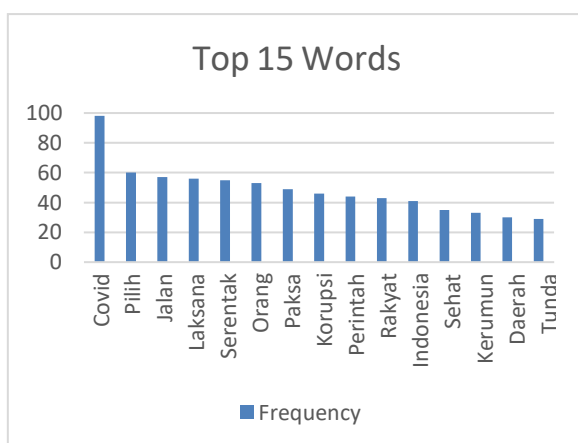


Gambar 4 Nilai rata-rata Evaluasi Performa dengan *fold* = 5

Setelah dilakukan evaluasi performa menggunakan cross validation yang ditampilkan pada Gambar, dapat dilihat bahwa nilai rata-rata *accuracy* yang didapatkan adalah sebesar 82.7% yang artinya model machine learning ini berhasil melakukan prediksi klasifikasi terhadap 331 tweets dari total 400 tweets. Nilai rata-rata *precision* yang didapatkan adalah 82.7%, yang artinya model machine learning ini berhasil memprediksi 155 tweets positif dari

total 187 tweets yang diprediksi sebagai kelas positif. Nilai rata-rata *recall* yang didapatkan adalah 82.7%, yang artinya model *machine learning* ini berhasil memprediksi 155 tweets dari total 187 tweets yang memiliki kelas asli positif. Kemudian nilai rata-rata *F1-Score* dari model *machine learning* ini adalah sebesar 82.6%, nilai ini merupakan nilai kombinasi dari *precision* dan *recall*.

### 4.3. Analisa Model Bag-of-Words



Gambar 5 Top 15 Words

Model *Bag-of-Words* pada penelitian ini digunakan untuk menghitung 15 kata yang paling sering dicuitkan oleh pengguna *Twitter* terhadap topik terkait. Dapat dilihat bahwa mayoritas pengguna *Twitter* merasa kurang setuju dengan kebijakan pemerintah yang memilih untuk tidak menunda Pilkada 2020 untuk dilakukan pada tahun berikutnya, dan tetap melanjutkan Pilkada 2020 di tengah pandemi, yang mana kemungkinan besar kasus pandemi akan terus naik karena kemunculan *cluster* baru akibat banyaknya kerumunan. Hal ini dapat dibuktikan dengan seberapa seringnya kata ‘laksana’, ‘paksa’, ‘kerumun’, dan ‘tunda’ muncul.

### 5. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat ditarik beberapa kesimpulan seperti:

1. Dalam pembuatan model *machine learning* untuk melakukan analisis sentimen menggunakan algoritma *Support Vector Machine*, dilakukan beberapa tahapan mulai dari *tweets scraping*, *labeling*, *preprocessing*, pembobotan kata

menggunakan *TF-IDF*, dan evaluasi performa.

2. Jumlah data yang di uji adalah 20% data (80 tweets) dari total data (400 tweets). Hasil pengujian terbaik didapatkan saat menggunakan parameter C dengan nilai 5, dan parameter *max\_iter* dengan nilai 100, dengan melakukan *cross validation* sebanyak 5 *fold*. Nilai rata-rata yang didapatkan saat menggunakan parameter dan jumlah *fold* ini adalah *accuracy* sebesar 82.7%, *recall* sebesar 82.7%, *precision* sebesar 82.7%, dan *f1-score* sebesar 82.6%.
3. Berdasarkan analisa model *Bag-of-Words*, kata yang sering muncul rata-rata memiliki sentimen negatif, yang artinya pengguna *Twitter* kurang setuju dengan kebijakan pemerintah yang tetap melaksanakan Pilkada 2020 di tengah pandemi.

### 6. DAFTAR PUSTAKA

- Annisa, T., 2020. *Mengenal peran sentiment analysis beserta cara kerjanya*. [online] Available at: <<https://www.ekrut.com/media/sentiment-analysis-adalah>>.
- Arifin, O., Sasongko, T.P., 2018, *Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA*. Universitas Amikom Yogyakarta.
- Appkey, 2020. *Apa itu Natural Language Processing (NLP) Pemrosesan Bahasa Alami*. [online] Available at: [https://appkey.id/pembuatan-aplikasi/mobile-programming/natural-language-processing-adalah/#Kegunaan\\_serta\\_Use\\_Case\\_dalam\\_Lingkup\\_yang\\_Lebih\\_Luas](https://appkey.id/pembuatan-aplikasi/mobile-programming/natural-language-processing-adalah/#Kegunaan_serta_Use_Case_dalam_Lingkup_yang_Lebih_Luas)
- Geitgey, A., 2018. *Natural Language Processing is Fun!* [online] Available at: <https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e>.
- Kasim, A.A., Sudarsono, M., 2019. *Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah*. Fakultas

Teknik Universitas Tadulako.

Lestari, L., 2018. *Mengenal Teknologi Natural Language Processing*. [online] Available at:

<https://blog.kata.ai/id/post/mengenal-teknologi-nlp/>

Liani, A., Enri, U., Umaidah, Y., 2020. *Analisis Perbandingan Kernel Algoritma Support Vector Machine dalam Mengklasifikasikan Skripsi Teknik Informatika Berdasarkan Abstrak*. Universitas Singaperbangsa Karawang.

Munir, M.M., 2018. *Implementasi Metode Backpropagation Neural Network Berbasis Lexicon Based Features Dan Bag Of Words Untuk Identifikasi Ujaran Kebencian Pada Twitter*. Sarjana thesis, Universitas Brawijaya Springer, London.

Nugroho, A.S., Witarto, A.B., Handoko, D., 2003. *Support Vector Machine -Teori dan Aplikasinya dalam Bioinformatika-*. IlmuKomputer.com

Olla, K., 2017. *Cari Tahu Apa Bedanya Supervised vs Unsupervised Learning*. [online] Available at: <https://www.jagoanhosting.com/blog/cari-tahu-apa-bedanya-supervised-vs-unsupervised-learning/>

Pakpahan, S.R.M., 2019. *Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku*. Sarjana thesis, Universitas Brawijaya.

Putri, W.T.H., Hendrowati, R., 2018. *Penggalian Teks Dengan Model Bag Of Words Terhadap Data Twitter*. Universitas Paramadina Jakarta

Syukuran, G., 2016. *Implementasi Metode Support Vector Machine Dan Metode Maximum Marginal Relevance Untuk Menghasilkan Rangkuman Dari Kumpulan Dokumen Berita Dengan Topik Sejenis*. Universitas Komputer Indonesia.

Zuhriel, H.P.PR., Fahrurrozi, A., 2021. *Implementasi Algoritma Klasifikasi Support Vector Machine Untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan PSBB*. Universitas Gunadarma.