



## KLASIFIKASI ANALISIS SENTIMEN TERHADAP CALON PRESIDEN 2019 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE ALGORITMA NAÏVE BAYES

Donny Maulana<sup>1</sup>, Rocky Markiano Siagian Siregar<sup>2</sup>

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

<sup>1</sup>donny.maulana@pelitabangsa.ac.id

### Abstraksi

Keberadaan twitter telah digunakan secara luas oleh berbagai lapisan masyarakat dalam beberapa tahun terakhir. Kebiasaan masyarakat mem-posting tweet untuk menilai para calon presiden adalah salah satu media dalam merepresentasikan tanggapan masyarakat terhadap para calon presiden tersebut. Maka dari itu dalam penelitian ini akan dilakukan analisis sentimen masyarakat terhadap calon presiden 2019 yang diungkapkan melalui jejaring sosial twitter. Analisis dilakukan dengan menggunakan klasifikasi tweet yang berisi sentimen masyarakat terhadap pencalonan presiden 2019 yaitu jokowi dan prabowo. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naive Bayes Classification(NBC). NBC digunakan untuk mendapatkan klasifikasi respon positif dan negatif masyarakat di twitter dan mendapatkan nilai preference value dari masyarakat terhadap para calon presiden 2019. Hasil pengujian metode persentase data jokowi 25%,50%,75%,dan 100% jumlah data dari data training menghasilkan akurasi sebesar 64,67%, 70,57%, 87,56%, 97,50% dan untuk hasil pengujian metode persentase data prabowo 25%, 50%, 75%,dan 100% jumlah data dari data training menghasilkan akurasi sebesar 64,57%, 81,67%, 64,22%, 62,67%. Dan untuk hasil pengujian respon positif masyarakat di twitter dengannilai preference value sebesar 53% untuk jokowi dan 48% untuk prabowo. Dengan demikian klasifikasi sentimen dengan menggunakan metode naive bayes classification dapat digunakan untuk mengukur respon masyarakat terhadap kinerja para calon presiden 2019.

**Kata kunci:** Twitter, naive bayes, analisa sentimen

### Abstract

*The existence of Twitter has been widely used by various levels of society in recent years. The public's habit of posting tweets to evaluate the presidential candidates is one of the media in representing the public response to the presidential candidates. Therefore in this study an analysis of public sentiments towards the 2019 presidential candidates will be revealed through the Twitter social network. The analysis was carried out using a tweet classification that contained public sentiment towards the 2019 presidential nomination, namely jokowi and prabowo. The classification method used in this study is Naive Bayes Classification (NBC). NBC is used to get the classification of positive and negative responses to the twitter and get a preference value from the community towards the 2019 presidential candidates. The results of the jokowi data percentage test methods are 25%, 50%, 75%, and 100% of the amount of data from the training data yielding an accuracy of 64.67%, 70.57%, 87.56%,*

*97.50% and for the test results the percentage of Pre -owo data 25%, 50%, 75%, and 100% of the amount of data from the training data resulted in an accuracy of 64.57 %, 81.67%, 64.22%, 62.67%. And for the results of testing the positive response of the people on Twitter with a value of preference value of 53% for Jokowi and 48% for Prabowo. Therefore sentiment classification using the Naive Bayes classification method can be used to measure the public response to the performance of 2019 presidential candidates.*

**Keywords:** Twitter, naive bayes, sentiment analysis

### 1. Pendahuluan

Indonesia merupakan negara demokrasi dengan penduduk terbanyak keempat dan pengguna Twitter terbesar kelima di dunia. Pemilihan presiden 2019 di Indonesia menjadi suatu topik yang menarik bagi para pengguna Twitter. Tweet masyarakat yang berkaitan dengan para calon presiden dapat digunakan untuk melihat gambaran opini

masyarakat terhadap para calon presiden. Banyaknya jumlah *tweet* yang masuk mengenai para calon presiden mendorong perlunya metode yang membantu untuk melihat opini masyarakat secara efektif. Salah satu metode yang dapat digunakan untuk mengklasifikasikan opini masyarakat secara efektif adalah *Support Vector Machine*. Metode ini akan mengklasifikasikan apakah suatu opini masyarakat akan termasuk dalam sentimen positif atau negatif dengan mencari *hyperlane* terbaik dari kedua kelas klasifikasi. Penambahan fungsi *Kernel* pada *Support Vector Machine* berguna untuk mengatasi data yang tidak terpisah secara *linier*. Hasil dari klasifikasi didapatkan akurasi sebesar 86%

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Hal ini ditandai dengan diadakannya suatu pemilihan umum terhadap presiden dan wakil presiden. Pemilihan umum pada suatu negara yang menganut demokrasi biasanya diselenggarakan secara periodik. Pada tahun 2019 akan dilaksanakan pemilihan umum presiden. Seorang tokoh politik yang ingin maju sebagai calon presiden tentu akan melihat atau mempertimbangkan popularitas mereka berdasarkan opini dari masyarakat.

Seperti akhir – akhir ini yang menjadi topic berita yaitu isu mengenai calon presiden untuk memilih mendatang yaitu Jokowi dan Prabowo. Memang semua yang berhubungan dengan pemilihan presiden menjadi banyak di perbincangkan di media sosial. Sehingga bermunculan pendapat positif maupun negatif dari masyarakat terhadap isu tersebut. Pendapat-pendapat tersebut bisa diolah dan dianalisa agar menghasilkan data atau informasi yang sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan.

Media sosial yang dimaksud pada peneliti ini adalah twitter,” Twitter adalah layanan jaringan sosial dan mikroblog daring yang memungkinkan penggunaannya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan *tweet* (wikipedia)”. *Tweet* dapat dilihat secara publik, namun, penggunaannya dapat membatasi pengiriman status ke daftar teman yang diinginkan. Pengguna juga dapat melihat *kicauan* penggunaan lain, dan “#” bisa disebut *hashtag* berfungsi sebagai topik pembahasan yang akan dibahas (Haewoon Kwak, et.al.2010).

Sedangkan untuk mengklasifikasikan kumpulan data *tweet* tersebut bisa menggunakan salah satu metode klasifikasi yaitu *naive bayes classifier*. *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada katekatori yang paling tepat (Feldman, R and Sanger, J. 2007).

## 2. Landasan Pemikiran

Twitter adalah layanan jejaring sosial dan

mikroblog daring yang memungkinkan penggunaannya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan *kicauan* (*tweet*). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan “pesan singkat dari Internet.” Di Twitter, pengguna tak terdaftar hanya bisa membaca *kicauan*, sedangkan pengguna terdaftar bisa memosting *kicauan* melalui antarmuka situs web, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler.

Twitter mengalami pertumbuhan yang pesat dan dengan cepat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif. Lonjakan penggunaan Twitter umumnya berlangsung saat terjadinya peristiwa-peristiwa populer. Pada awal 2013, pengguna Twitter mengirimkan lebih dari 340 juta *kicauan* per hari, dan Twitter menangani lebih dari 1,6 miliar permintaan pencarian per hari. Hal ini menyebabkan posisi Twitter naik ke peringkat kedua sebagai situs jejaring sosial yang paling sering dikunjungi di dunia, dari yang sebelumnya menempati peringkat dua puluh dua.

Tingginya popularitas Twitter menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai aspek, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat. Twitter juga dihadapkan pada berbagai masalah dan kontroversi seperti masalah keamanan dan privasi pengguna, gugatan hukum, dan penyensoran.

### 2.1. Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan proses *crawling* data twitter menggunakan software *rapidminer*, data yang dikumpulkan merupakan *tweet* yang berbahasa Indonesia yang diambil dari akun twitter @Jokowi dan @Prabowo yang berisi komentar para pengguna twitter.



Gambar 1. Proses Pengumpulan Data

### 2.2 Seleksi Data

Seleksi Data bertujuan untuk menyeleksi data *tweet* yang akan digunakan nantinya,

dikarenakan data tweet yang didapatkan dari proses crawling menggunakan rapidminer belum sepenuhnya bersih dan memiliki data yang tidak diperlukan [2].

**2.3 Remove Duplicates**

Pada proses ini bertujuan untuk memfilter data tweet yang sama, hal ini disebabkan karena twitter memiliki fitur retweet yang menyebabkan banyaknya tweet yang sama di retweet oleh pengguna lain dan menimbulkan teks berulang dengan topik yang sama.

**2.4 Cleansing**

Cleansing bertujuan untuk membersihkan data dari noise. Seperti karakter HTML, ikon emosi, hashtag, username, url dan email [3]

**2.5 Normalisasi**

Mengganti kata yang tidak baku Kata tidak baku merupakan kata yang tidak terdapat pada Kamus Besar Bahasa Indonesia (KBBI)

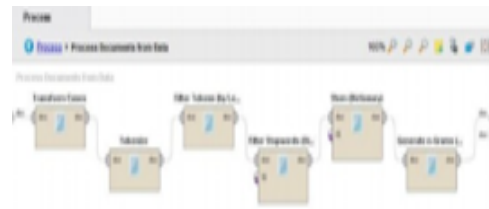
**2.6 Penentuan Kelas Atribut**

Data yang sudah melalui proses pada tahapan preprocessing kemudian akan dilakukan penentuan kelas padaatributnya, penentuan ini diberikan sesua subjektifitas peneliti, pembagian atribut pada penelitian ini sendiri hanya akan dibagi menjadi positif dan negatif,

**2.7 Document Process**

- 1) Transform Case Tahapan transform cases sendiri dalam penelitian ini bertujuan untuk menyeragamkan bentuk huruf menjadi lower case karna pada komentar Twitter sendiri memiliki bentuk huruf yang beragam
- 2) Tokenize Tokenize merupakan sebuah proses untuk memisahkan atau pemisahan dokumen menjadi per kata atau beberapa kata dalam suatu token/unit kecil.
- 3) Filter Token by Length Filter Tokens by Length merupakan proses penghapusan kata atau token berdasarkan panjang karakter yang ditentukan dan pada penelitian ini minimum panjang karakter yang akan digunakan adalah empat karakter pada setiap kata.
- 4) Filter Stopwords Tahapan filter stopword (dictionary) merupakan tahapan proses yang bertujuan untuk menghapus atau memfilter kata yang tidak memiliki arti dan tidak berhubungan dengan ucapan atau kata sentiment [4]
- 5) Stemming Tahap Stemming adalah tahap menghilangkan imbuhan yang masih melekat sehingga dapat diperoleh kata dasar, dalam penelitian ini proses stemming akan dilakukan dengan

- 6) Generate N-Grams Proses ini merupakan proses pemotongan string atau dokumen (n-gram) berdasarkan limit angka tertentu[5]. Dan dalam penelitian ini sendiri n-gram yang akan digunakan adalah bigram. Tujuan penggunaan N-gram sendiri karena dalam Bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata.



**Gambar 2.** Document Process

- 7) Klasifikasi Dalam penelitian ini algoritma yang digunakan adalah Naïve Bayes Classifier dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator cross validation dengan menambahkan Naïve Bayes pada tab training dan pada tab testing akan ditambahkan apply model dan performance yang nantinya akan digunakan sebagai penghitung performa klasifikasi.



**Gambar 3.** Klasifikasi NBC

- 8) Evaluasi Dalam penelitian ini proses evaluasi akan menggunakan Confusion Matrix, Evaluasi bertujuan untuk mengetahui kinerja dari algoritma Naïve Bayes yang nantinya akan dijadikan sebagai acuan dari perbandingan akurasi pengujian [6].

**Tabel 1.** Perbandingan akurasi pengujian

		Actual Class	
		Class 1	Class 2
Predicted	Class 1	True Positive	False Negative
	Class 2	False Positive	True Negative

**3. Metode Penelitian**

**3.1. Objek Penelitian**

Dalam penyusunan penelitian ini peneliti melakukan penelitian tentang Klasifikasi Analisis Sentimen Terhadap calon presiden 2019

Menggunakan Naive Bayes Classifier, yang dimana pada penelitian ini dilakukan untuk mengetahui kinerja dari optimasi metode dan seleksi fitur yang digunakan serta mengetahui sentimen dari para calon presiden pada media sosial twitter 2019.

Objek penelitian yang akan diteliti oleh peneliti ialah tentang sentimen dengan pengambilan datanya berdasarkan query (@jokowi dan @prabowo) yang berasal dari media sosial twitter. Pemilihan objek didasari karena ingin mengetahui bagaimana pendapat masyarakat terhadap calon presiden, yang mana banyak muncul berbagai macam opini soal calon presiden ini pada media sosial twitter baik itu opini kepuasan ataupun ketidakpuasan.

**3.2. Jenis Data**

Dalam penelitian ini untuk jenis datanya ini sendiri peneliti mengambil data yang diperoleh dari proses crawling data pada API Search Twitter yang bersifat kualitatif, data itu berupa kata-kata. Kemudian akan diproses nantinya untuk mengetahui sentimen yang terdapat didalamnya.

**3.3. Data yang digunakan**

**A. Data twitter/tweet**

Dataset berupa teks berbahasa Indonesia yang diambil dari website <http://www.twtiiter.com>. Data yang diambil untuk penelitian ini data yang diambil menggunakan query „@jokowi dan @prabowo“, dengan memanfaatkan API Search Twitter. Tweet yang diambil merupakan postingan dari pelanggan akun dari calonpresiden tersebut.

**B. Data Kata**

Dasar Data kata dasar diperoleh dari <https://github.com/nolimitid/nolimit-kamus> yang mana data dari kata dasar ini berjumlah 28526 kata dasar, yang mana nantinya data kata dasar ini akan disimpan dan digunakan untuk proses stemming. Berikut contoh kata dasar:

**Tabel 2. Data Kata Dasar**

No	Kata Dasar
1	Temu
2	Kasih
3	Terima
4	Dating

**C. Data Stopword**

Data stopwords juga diperoleh dari <https://github.com/nolimitid/nolimitkamus> yang mana data stopwords ini berjumlah 758 data stopwords yang mana nantinya data ini akan disimpan dan digunakan pada proses stopwords nantinya. Berikut contoh kata

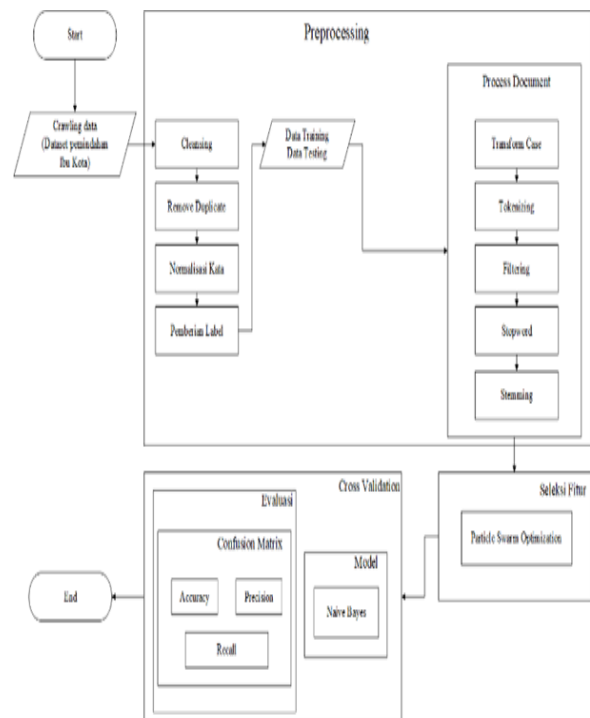
stopword:

**Tabel 3. Data Stopword**

No	Stopword
1	Adapun
2	Agar
3	Akan
4	Akhir
5	Adalah

**3.4. Metode yang digunakan**

Penelitian ini akan melalui beberapa proses, alur urutan proses pada penelitian ini akan diuraikan pada metode penelitian ini, metode penelitian ini dapat dilihat sebagai mana pada gambar di bawah ini:

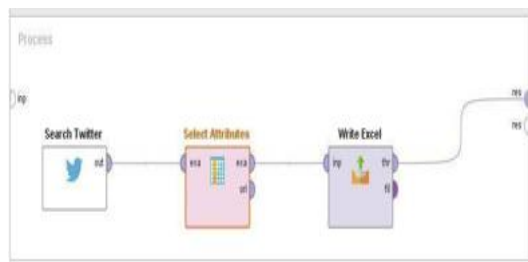


**Gambar 4. Alur Metode Penelitian**

**3.5. Pengumpulan Data**

Data yang digunakan dalam penelitian ini diambil dari jejaring sosial twitter. Pengambilan data dengan memanfaatkan Tools Rapid Miner menggunakan operator Search twitter yang dihubungkan dengan API pencarian twitter dengan mendapatkan kode API twitter yang berhubungan dengan topik terkait calon presiden 2019 dengan menggunakan kata kunci akun twitter “@jokowi” dan “@prabowo”. Kita bisa menentukan batas maksimal data yang kita inginkan. Namun data yang diberikan hanya sebatas berapa banyak tweet dengan kata kunci terkait dalam jangka waktu seminggu sebelum tanggal pencarian. Sedangkan untuk seleksi bahasa digunakan library bawaan twitter (lang = ‘in’) yang merupakan code untuk teks bahasa indonesia. Berikut data sampel yang peneliti sebanyak 1200 sampel data:





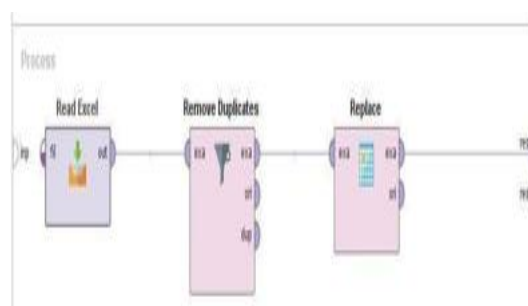
**Gambar 5.** Alur Pengambilan Data Menggunakan RapidMiner

Keterangan:

1. Search Twitter  
Operator yang berfungsi untuk mengambil data dari sosial media Twitter. Operator search twitter terdiri dari beberapa Parameter yaitu: connection, query, result type, limit, since id, max id, language, dan local.
2. Select Attributes  
Operator yang berfungsi untuk memfilter atribut yang ingin digunakan. Operator select attributes terdiri dari beberapa parameter yaitu: attribute filter type, attribute, invert selection, dan include special attributes.
3. Write Excel  
Operator yang berfungsi untuk meng export data ke file excel. Operator write excel terdiri dari beberapa parameter yaitu: excel file, file format, encoding, sheet name, date format, dan number format.

### 3.5.1 Preprocessing

Preprocessing merupakan tahap awal untuk mendapatkan data yang bersih supaya dapat lanjut ke tahap berikutnya, tahap preprocessing pada penelitian ini terdiri dari seleksi data, remove duplicate, cleansing, normalisasi kata, pemberial label, pembagian data, dan proses dokumen. Dibawah ini merupakan gambar alur tahap preprocessing.



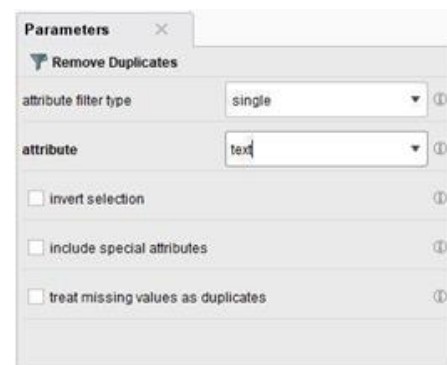
**Gambar 6.** Tahap *Preprocessing* menggunakan Rapidminer

Keterangan:

- a) Read excel Read excel berfungsi untuk membaca file dataset saat proses pengambilan data. Operator read excel memiliki beberapa parameter yakni: import configuration wizard, excel file, sheet selection, sheet number, sheet name, imported cell range, first row as names, annotations, date format, time zone, locale,

read all values as polynominal, data set meta data information,read not matching values as missings, dan data management

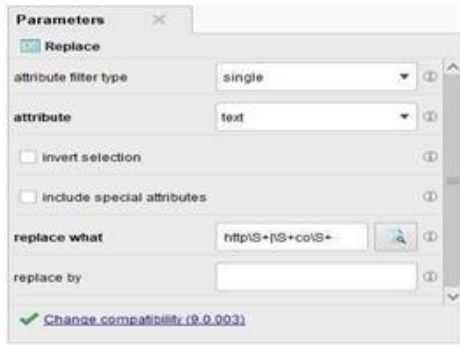
- b) Remove duplicate Remove duplicate berfungsi untuk menghapus data yang sama. Operator remove duplicate memiliki beberapa parameter yakni: attribute filter type, include special attributes, invert selection, dan treat missing values as duplicates.
- c) Replcae Replace berfungsi untuk menghapus kata, simbol, karakter unik pada dataset. Operator replace memiliki beberapa parameter yakni: attribute filter type, invert selection, include special attributes, replace what, dan replace by.
  1. Seleksi data Seleksi data bertujuan untuk menyeleksi data tweet yang akan digunakan, data hasil crawling masih memiliki data yang tidak diperlukan. Dalam penelitian ini hanya memerlukan komentar yang berhubungan dengan pemindahan Ibu Kota. Gambar berikut ini merupakan contoh proses seleksi data.
  2. Remove duplicate Remove duplicate bertugas untuk menghilangkan data (komentar tweet) yang sama (duplicate). Data yang didapat dari hasil crawling masih terdapat data berulang, hal itu akan berdampak pada lamanya waktu pemrosesan data. Gambar dibawah ini merupakan contoh duplikat data pada data dan proses menghapus data yang duplikat.



**Gambar 7.** Proses Remove Duplicate Data

### 3.5.2 Cleansing

Tahapan cleansing pada penelitian ini untuk menghapus kata, karakter dan simbol yang tidak diperlukandalam penelitian, yaitu antara lain sebagai berikut:



Gambar 8. Proses Replace Data

Keterangan:

- a. *Attribute filter type*  
*Attribute filter type* berfungsi untuk memilih jenis atribut yang akan di filter. Attribute filter type memiliki beberapa pilihan diantaranya yaitu: beberapa pilihannya yaitu: *all*, *single*, *subset*, *regular expression*, *value type*, *block type*, *no missing value*, dan *numeric value filter*.
- b. *Attribute*  
*Attribute* berfungsi memilih atribut kelas yang ada pada dataset. Dalam penelitian ini atribut yang digunakan adalah atribut kelas *text*.
- c. *Replace what*  
*Replace what* berfungsi untuk menghapus simbol, karakter spesial, dan URL. Dalam penelitian ini simbol dan karakter yang akan dihapus adalah *https*, *@*, *#*, dan *RT* (*retweet*).

### 3.5.3 Pembagian Data

Pembagian data bertugas untuk membagi data keseluruhan menjadi dua bagian, data training dan data testing. Membagi data menjadi Data training dan testing diperlukan untuk proses klasifikasi, jumlah data keseluruhan sebesar 1000 tweet yang mana 800 tweet sebagai data training dan 200 tweet sebagai data testing.

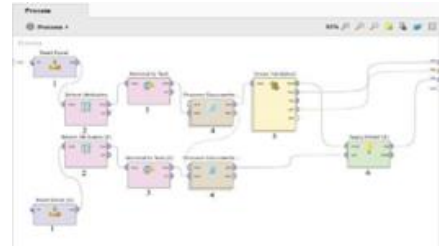
## 4. Hasil dan Pembahasan

### A. Dataset

Dataset yang digunakan dalam penelitian ini sebanyak ada 1200 tweet yang kemudian dibagi 600 tweet yang masing-masing diambil dari pengguna yang mention ke akun @jokowi dan @prabowo data diambil dari tanggal yang berbeda-beda dikarenakan twitter sendiri membatasi pengambilan data atau crawling selama 7 hari dari waktu pengambilan data. Untuk lebih jelasnya dapat dilihat pada tabel berikut:

- B. Proses Pengujian  
 Proses Pengujian dilakukan dengan mengikuti alur yang telah ditentukan yaitu dengan metode klasifikasi Naïve Bayes dengan Cross Validation, proses ini terdiri dari 5 kali proses pengujian dimasing-masing akun yang mana nantinya akan dijadikan tolak ukur dari penelitian ini. Dalam proses ini peneliti menggunakan RapidMiner untuk melakukan pengujian dengan merangkai beberapa proses yaitu

read excel, Select Attribute, Nominal to text, Process Document, Cross Validation dan Apply Model. Proses tersebut dapat dilihat pada gambar dibawah ini:



Gambar 9. Proses Pengujian

### C. Analisa Akurasi

Dari hasil pengujian yang telah dilakukan menggunakan algoritma Naïve Bayes dan dilakukannya pengujian menggunakan teknik *cross validation* peneliti mendapatkan hasil pengujian seperti pada tabel berikut:

Tabel 4. Analisa Akurasi Jokowi

Nilai	PERSENTASE DATA			
	25%	50%	75%	100%
Akurasi	64,67%	70,57%	87,56 %	97,50%
Recall	34,57%	65,76%	74,77%	94,83%
Precision	56,56%	73,17%	80,28%	95,89%

Dari tabel diatas maka dapat disimpulkan bahwa nilai Akurasi , *Precision* dan *recall* tertinggi terdapat pada pengujian 100% yang mana mendapatkan hasil paling tinggi dengan nilai *Accuracy* 97.50% *Precision* 94.83% dan *Recall* sebanyak 95,89%. Hal ini menunjukkan bahwa penggunaan teknik *cross validation* cukup berpengaruh pada hasil pengujian yang didapatkan.

Tabel 5. Analisa Akurasi Prabowo

Nilai	PERSENTASE DATA			
	25%	50%	75%	100%
Akurasi	64,57%	81,67%	64,22%	62,67%
Recall	56,56%	63,33%	58,61%	59,83%
Precision	34,57%	32,31%	27,48%	99,70%

Dari tabel diatas maka dapat disimpulkan bahwa nilai Akurasi , *Precision* dan *recall* tertinggi terdapat pada pengujian 50% yang mana mendapatkan hasil paling tinggi dengan nilai *Accuracy* 81.67% *Precision* 32.31% dan *Recall* sebanyak 63,33%. Hal ini menunjukkan bahwa penggunaan teknik *cross validation* cukup berpengaruh pada hasil pengujian yang didapatkan.

### D. Preference Value jokowi.

Preference Value bertujuan untuk mengetahui respon positif masyarakat dari data yang telah diambil dan di proses sebelumnya, jumlah data positif pada data training dan testing berjumlah 328 dan jumlah data negatif pada data training dan testing berjumlah 272, total dari data tersebut berjumlah 600 data, dan berikut adalah perhitungan preference value dari

data tersebut dikalikan dengan nilai akurasi terbaik:

$$\text{preference value} = \frac{328}{328 + 272} * 97,50\% = 0,533$$

Semakin tinggi nilai preference value yang dihasilkan maka dapat diartikan semakin tinggi pula respon positif yang didapatkan[7]. Dan dari perhitungan diatas maka dapat disimpulkan bahwa respon positif masyarakat terhadap kepemimpinan ridwan kamil berdasarkan data yang diambil dari Twitter adalah 53%.

#### E. Preference Value Prabowo

Preference Value bertujuan untuk mengetahui respon positif masyarakat dari data yang telah diambil dan di proses sebelumnya, jumlah data positif pada data training dan testing berjumlah 330 dan jumlah data negatif pada data training dan testing berjumlah 270, total dari data tersebut berjumlah 600 data, dan berikut adalah perhitungan preference value dari data tersebut dikalikan dengan nilai akurasi terbaik:

$$\text{preference value} = \frac{330}{330 + 270} * 87,67\% = 0,482185$$

## 5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa penggunaan algoritma Naïve Bayes disertai dengan penggunaan teknik Cross Validation menghasilkan hasil yang cukup baik. Hasil klasifikasi sentimen data tweet calon presiden 2019 yaitu dari akun jokowi dan prabowo dengan data testing masing-masing 600 data dengan komposisi 328 data tweet positive dan 272 data tweet negative yang ditentukan secara manual, diperoleh jumlah data tweet dengan popularitas sentimen positive lebih banyak dibandingkan dengan data tweet dengan popularitas sentimen negative. dan untuk menghitung nilai preference value dengan akurasi jokowi sebesar 97,50% dan tingkat respon positif masyarakat sebesar 53% dan akurasi prabowo sebesar 87,67% dan tingkat respon positif masyarakat sebesar 48%.

## Daftar Pustaka

- [1] A. Maulana, "Twitter Rahasiakan Jumlah Pengguna di Indonesia," 2016. [Online]. Available: <https://www.cnnindonesia.com/teknologi/20160322085045-185118939/twitter-rahasiakan-jumlahpengguna-di-indonesia>.
- [2] Hasanudin, "Analisa sentimen terhadap tokoh publik pada twitter menggunakan metode klasifikasi naive bayes classifier, term frequency, tf-idf dan n-gram," p. 49, 2018.
- [3] G. A. Buntoro, "Analisis Sentimen Calon Gubernur Jawa Timur 2018 di Twitter," *Sci. Comput. Sci.Informatics J.*, vol. 1, no. 2, p. 37, 2019.
- [4] R. S. Putra, "Analisis sentimen twitter dengan klasifikasi naïve bayes menggunakan seleksi fitur mutual information dan inverse document frequency riky sutriadi putra," 2017.
- [5] W. C. Indhiarta, "Penggunaan N-Gram Pada Analisa Sentimen," 2017.
- [6] Ahmad Turmudi Zy, "COMPARISON ALGORITHM CLASSIFICATION NAIVE BAYES, DECISION TREE, AND NEURAL NETWORK FOR ANALYSIS SENTIMENT" *Progr. Stud. Tek. Inform. Sekolah Tinggi Teknol. Pelita Bangsa*, vol. 12, no. April, p. 200, 2017.
- [7] H. S. Ginting, K. M. Lhaksana, and D. T. Murdiansyah, "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," *ISSN 2355-9365 e-Proceeding Eng. Vol.5, No.1 Maret 2018, Univ. Telkom*, vol. 5, no. 1, pp. 1793–1802, 2018.
- [8] A. Novantirani1, M.K. Sabariah, and V. Effendy, 2015. "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine".
- [9] Abdul, A.M. 2015. Penerapan Algoritma TF-IDF Untuk Pencarian Karya Ilmiah. Universitas Dian Nuswantoro. Semarang.
- [10] Buntoro, A. Ghulam, 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. Universitas Muhammadiyah Ponorogo.
- [11] Charu, C. Aggarwal. 2015. Data Mining The Text Book. IBM T.J. Watson Research Center YorktownHeights. New York USA.
- [12] Feldman, R and Sanger, J. 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press:NewYork.
- [13] Ginting, H. S., Lhaksana, K. M., & Murdiansyah, D. T. (2018). Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes, 5(1), 1793–1802.
- [14] Hidayatullah, A. F., & SN, A. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. Seminar Nasional Informatika 2014, (August 2013), 0–8.
- [15] Kent, M. L. 2013. Public Relations Review Using social media dialogically :Kumar, Lokesh and Bhatia, Parul Kalra. 2015. Text Mining: Concepts, Process And Application. Journal of Global Research in Computer Science. Volume 4, No. 3, March 2013.

- [16] Lesmana, P. I. (2013). Analisis Sentimen Pengguna Layanan Media Sosial Twitter Di Indonesia, Karya Akhir Ilmu Komputer. Jakarta: Universitas Indonesia.
- [17] Mujilawati, S. 2016. PRE-PROCESSING TEXT MINING PADA DATA TWITTER. Universitas Islam Lamongan. Lamongan. 18–19.
- [18] M.Y. Nur and D.D. Santika, “Analisis Sentimen pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine”. Konferensi Nasional Sistem dan Informatika, 2014.
- [19] Patil, T. R., & Sherekar, S. S. (2013). Performance Analysis of Naive Bayes and J48 ClassificationAlgorithm for Data Classification, 6(2).