

SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER

Akhmad Muzaki*¹, Arita Witanti²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Indonesia
Email: 1akhmadmuzaki713@gmail.com, 2arita.mercu-buana@yogya.ac.id

(Naskah masuk: 09 Januari 2021, diterima untuk diterbitkan: 12 Februari 2021)

Abstract

The 2020 regional elections in the midst of the COVID-19 pandemic are starting to get crowded starting from the real world and in cyberspace, especially on Twitter social media. Twitter's existence has been widely used by various communities in recent years. Twitter is one of the media that represents the public response regarding public issue. Ahead of the general election (PEMILU), there are usually some parties who want to know the results of public sentiment or response to the issue, namely academics, intellectuals or even political opponents. Nevertheless, the implementation of local elections is very polemic in the community, therefore this study tries to analyze tweets that talk about issue public, namely the 2020 elections in the wake of the COVID-19 Pandemic. The analysis usually uses the classification of tweets containing public sentiment about the issue. The classification method used in this research is Naive Bayes Classifier (NBC) And Support Vector Machine (SVM). Naive Bayes Classifier is combined with features that can detect weighting using probability. The classification of tweets in this study was obtained based on a combination of two classes namely sentiment class and category class. The classification of sentiment consists of positive and negative. Test results on built-in applications show that accuracy with Naive Bayes delivers better results than Support Vector Machine. However, overall the use of the Naive Bayes method has a good performance to classify tweets with an accuracy rate of 92.2%

Keywords: classification, Naive Bayes Classifier, sentiment analysis.

SENTIMEN ANALISIS MASYARAKAT DI TWITTER TERHADAP PILKADA 2020 DITENGAH PANDEMIC COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER

Abstrak

Pemilihan kepala daerah (Pilkada) serentak 2020 di tengah *pandemic COVID-19* mulai ramai di bicarakan mulai dari dunia nyata maupun di dunia maya, khususnya di media sosial *Twitter*. Keberadaan *Twitter* telah digunakan secara luas oleh berbagai kalangan masyarakat dalam beberapa tahun terakhir. *Twitter* adalah salah satu media yang merepresentasikan tanggapan masyarakat terkait isu publik. Menjelang dilaksanakannya pemilihan umum (PEMILU), biasanya ada beberapa pihak yang ingin mengetahui hasil sentimen atau tanggapan masyarakat terhadap isu tersebut, yaitu akademisi, intelektual atau bahkan lawan politik. Kendati demikian pelaksanaan pilkada sangat menuai polemik di lapisan masyarakat, oleh karena itu penelitian ini mencoba menganalisis *tweet* yang membicarakan tentang isu *public* yaitu pilkada 2020 di tengah *Pandemic COVID-19*. Analisis yang dilakukan biasanya menggunakan klasifikasi *tweet* yang berisi sentimen masyarakat tentang isu tersebut. Metode klasifikasi yang digunakan pada penelitian kali ini adalah *Naive Bayes Classifier (NBC)* Dan *Support Vector Machine (SVM)*. *Naive Bayes Classifier* dikombinasikan dengan fitur yang dapat mendeteksi pembobotan menggunakan probabilitas. Klasifikasi *tweet* dalam penelitian ini diperoleh berdasarkan kombinasi antara dua kelas yaitu kelas sentimen dan kelas kategori. Klasifikasi sentimen terdiri dari positif dan negatif. Hasil pengujian pada aplikasi yang dibangun memperlihatkan bahwa akurasi dengan *Naive Bayes* memberikan hasil yang lebih baik dari pada *Support Vector Machine*. Namun demikian, secara keseluruhan penggunaan metode *Naive Bayes* memiliki performansi yang baik untuk melakukan klasifikasi *tweet* dengan tingkat akurasi 92,2%.

Kata kunci: analisis sentimen, klasifikasi, Naive Bayes Classifier.

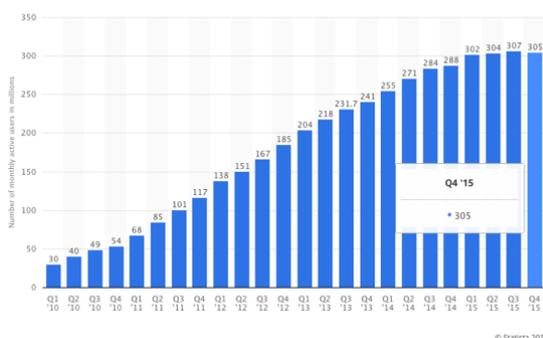
1. PENDAHULUAN

Pemerintah melalui Perpu nomor 2 tahun 2020 mengenai pemilihan kepala daerah diselenggarakan

pada 9 Desember 2020. Pilkada 2020 diselenggarakan pada 23 September untuk pemilihan sembilan gubernur 224 Bupati dan 37 walikota secara serentak. Sebelum Indonesia terparap *COVID-19*

komisi pemilihan umum (KPU) telah melaksanakan beberapa tes tahapan pelaksanaan Pilkada serentak 2020 namun akibat wabah *COVID-19* akhirnya komisi pemilihan umum mengeluarkan surat keputusan KPU Nomor.179/PL.02.-Kpt/01/KPU/III/2020. Yang akhirnya mengundang kontroversial serta polemik dikalangan intelektual, akademisi juga masyarakat pada umumnya sebab bagaimana mungkin tetap dilaksanakan Pilkada sedangkan angka penyebaran *COVID-19* masih sangat tinggi[1].

Media sosial *Twitter* dan *Facebook* bersaing ketat dalam merangkul jumlah pengguna di Indonesia. Kendati demikian, pengguna kedua *platform* tersebut memang merupakan pasar yang penting. Berdasarkan data PT Bakrie Telecom, memiliki angka 19,5 juta pengguna aktif *Twitter* di Indonesia dari total 500 juta pengguna *Twitter* di dunia. *Twitter* adalah salah satu jejaring sosial terbesar di dunia yang mampu meraup keuntungan yang sangat besar yaitu mencapai nilai *USD* 145 juta. Jumlah tersebut tentu saja sudah bertambah hingga September 2020 ini. Selain itu, Indonesia menempati urutan kelima pengguna *Twitter* terbanyak setelah Negara *USA*, *Brasil*, *Jepang*, dan *UK* [2]. Dalam media sosial *Twitter* sekita 1.200 *tweet* pada tanggal 20 Oktober 2020 menyuarakan tentang permasalahan dalam Pilkada serentak yang akan dilaksanakan pada bulan Desember 2020. Dalam media *Twitter* secara tidak langsung kita bisa melihat suara rakyat tentang pilkada 2020. Berikut adalah data pengguna *Twitter* di Indonesia yang ditampilkan pada gambar 1, dimana data pengguna mengalami peningkatan pada tahun ke tahun.



Gambar 1. Data Pengguna *Twitter* [3]

Ketika berkunjung ke Indonesia, bulan Oktober 2015, *CEO Facebook* yaitu *Mark Zuckerberg* mengatakan bahwa Indonesia merupakan pasar yang potensial. *Twitter* sering digunakan untuk menggali informasi yang berharga dan kelebihan penggunaan yang mudah. Salah satu fitur yang paling banyak digunakan adalah *tweet*, akan menjadi sangat sia-sia jika tidak dimanfaatkan dengan benar, sehingga dibutuhkan teknik yang mampu mengolah *tweet-tweet* tersebut sehingga nantinya menghasilkan informasi yang berharga.

Sentimen analisis merupakan suatu bidang studi yang mampu menganalisis opini dari seseorang, sentimen, evaluasi, perilaku dan emosi seseorang melalui entitas seperti layanan umum, organisasi, isu, kejadian, topik dan atribut-atribut lainnya. Sebagian besar sentimen analisis banyak digunakan untuk kepentingan politik, ekonomi, bisnis, pendidikan atau pemerintah dalam menganalisis opini masyarakat atas suatu kejadian sehingga kebijakan yang dibuat mampu menjangkau sampai ke bagian akar dan rumput. Sentimen analisis dalam dunia bisnis kebanyakan digunakan untuk menganalisis tren. atau kebutuhan pasar, atau kebutuhan masyarakat yang diharapkan dapat menyusun strategi pemasaran yang tepat dan dapat meningkatkan pendapatan perusahaan mereka[2].

Tahun 2020 merupakan periode akhir masa jabatan kepada daerah, sehingga perlu diselenggarakan pemilihan kepala daerah (Pilkada) periode 2020-2025. Pada Pilkada ini terdapat pasangan cagub (calon gubernur) beserta cawagub (calon wakil gubernur). Dalam pra-pelaksanaan maupun pelaksanaan Pilkada terdapat berbagai opini dan tanggapan dengan sentimen positif dan negatif pada media sosial maupun di dunia nyata terkait diselenggarakannya Pilkada serentak 2020, Pemilihan akan sangat beresiko tertular *COVID-19* jika pemerintah tetap melaksanakannya. Permasalahan yang muncul adalah ketika pemerintah menghimbau agar masyarakat tetap dirumah namun disisi lain pilkada harus tetap dilaksanakan. Tujuan penelitian ini dimaksudkan untuk mengetahui *sentiment* masyarakat terhadap pelaksanaan pilkada 2020 yang akan dilaksanakan pada tanggal 9 Desember dan ditengah pandemic *COVID-19* dengan menggunakan metode *Naïve Bayes Classifier*.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Naive bayes classifier* karena dalam proses klasifikasi dan perhitungan probabilistik, *naive bayes* memiliki lebih banyak keunggulan. Salah satu keunggulannya adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas. *Naive Bayes* berdasarkan pada *teorema Bayes* yang memiliki kemampuan untuk mengklasifikasi yang sama dengan *decision tree* dan *neural network*. *Naive Bayes* terbukti memiliki tingkat akurasi dan kecepatan yang tinggi ketika diaplikasikan ke dalam *database* dengan data yang sangat besar[4].

2.1 Text Mining

Text mining adalah bagian dari *data mining* dimana proses yang dilakukan adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen *text* menggunakan alat analisis tertentu. *Text mining* dapat juga dimanfaatkan menjadi beberapa macam kebutuhan diantaranya yaitu untuk *summarization*,

pencairan dokumen teks dan sentiment analisis[2]. *Text mining* berfungsi untuk mencari dan mematerialkan kata-kata yang dapat mewakili apa yang ada didalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. *Text mining* mempunyai 4 tahapan yaitu *Tokenizing*, *Filtering*, *Stemming*, dan *Analyzing*.

Tahapan *tokenizing* merupakan proses pemotongan *string* masukan berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses ini ialah memisahkan setiap kata yang menyusun suatu dokumen. Tahapan *Filtering* merupakan suatu proses dimana data diambil sebagian dari data tertentu, dan membuangnya pada *frekuensi* yang lain. Tahapan *Stemming* adalah proses dimana pemetaan atau penguraian berbagai bentuk teks dari sebuah kata menjadi bentuk kata dasarnya. Tahapan *Tagging* adalah metode yang belum lama dilahirkan. Dahulu sebelum ditemukanya metode *tagging*, data informasi yang beredar di internet berantakan dan tidak tersusun dengan rapih berdasarkan kategorinya. Tahapan *Analyzing* adalah tahapan untuk mencari seberapa jauh keterhubungan data antar kata-kata disetiap dokumen.

2.2 Sentimen Analisis

Sentimen analisis merupakan salah satu bidang studia tau metode yang fungsinya menganalisis opini masyarakat, sentimen, evaluasi, penilaian, perilaku dan emosi melalui entitas seperti produk, layanan umum, organisasi, individual, issu, kejadian dan topik. Sentimen analisis berfokus kepada opini yang mengandung pesan negatif atau positif [2]

Proses *data mining* terbentuk dari beberapa tahapan, salah satunya adalah pengumpulan data *data collection*, *data transformation*, dan *data analysis*[4]. Proses tersebut dimulai melalui *pre-processing* yang terbentuk dari pengumpulan data untuk menghasilkan data mentah (*raw data*) yang akan dibutuhkan oleh *data mining*, yang kemudian dilanjutkan dengan proses transformasi data untuk mengubah data mentah berubah menjadi format yang lebih mudah diproses oleh *data mining*, misalnya melalui *filtrasi* atau *agregasi*. Selanjutnya hasil *transformasi* data akan di manfaatkan oleh analisis data untuk membangkitkan pengetahuan dengan menggunakan teknik seperti *analisis statistik*, *machine learning*[7], dan *visualisasi* informasi.

2.3 Klasifikasi

Klasifikasi tergolong dalam area pembelajaran *machine learning* dimana mesin mempelajari data yang sudah diberi label atau data yang sudah terklasifikasi. Terdapat beberapa algoritma yang potensial digunakan dalam klasifikasi sentimen misalnya adalah *Naive Bayes*, *K-NN*, *Logistic Regression*, *Support Vector Machine*, *Artificial Neural Network* dan lain-lain[8].

2.4 Naive Bayes Classifier

Naive Bayes adalah klasifikasi linear yang sangat efisien. Model probabilitas dan klasifikasi *Naive Bayes* didasarkan atas *teorema bayes*, dan *adjective naive* berasal dari asumsi yaitu fitur yang ada pada *dataset* saling independen. Model probabilitas yang telah diformulasikan oleh *Thomas Bayes* (1701-1761) [2] sangat sederhana berikut adalah formulasinya:

$$\frac{\text{posterior probability} = \text{conditional probability.prior probability}}{\text{evidence}} \quad (1)$$

Notasi umum dari posterior probability juga bisa dituliskan seperti berikut

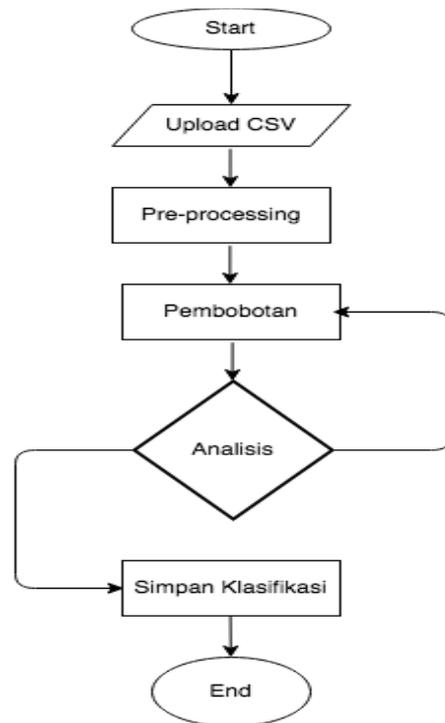
$$P(Y_j|X_i) = \frac{P(Y_j|X_i).P(Y_j)}{P(X_i)} \quad (2)$$

Di mana :

X_i = fitur sampel *Vector i*, $i \in \{1, 2, \dots, n\}$

Y_j = notasi kelas j , $j \in \{1, 2, \dots, m\}$

$P(Y_j|X_i)$ = probabilitas dari sampel X_i milik kelas W_j .



Gambar 2. Tahapan Analisis

Pada proses ini data yang sudah melewati proses *pre-processing* dengan baik akan dilanjutkan dengan pembobotan [9] dan akan diberikan algoritma *Naive Bayes*. Proses pembelajaran akan dilakukan dengan data yaitu data latih. Data latih merupakan data yang sudah mempunyai label positif dan negatif, sebagai pembelajaran untuk mengetahui label dari suatu dokumen. Jumlah keseluruhan data adalah 1200 data dengan pembagian data latih dan data uji. Perbandingan antara data latih dan data uji adalah sebanyak 70% untuk data latih dan 30% untuk data

uji. Data latih dan data uji akan sama-sama diimplementasikan menggunakan algoritma *Naive Bayes*. Adapun tahapan analisis diunjukkan pada gambar 2.

2.5 Pre-processing

Pre-processing adalah tahapan yang paling penting dalam proses *text mining*, tahapan ini adalah tahap pertama dalam pemrosesan *text*. Pada tahapan ini *text* akan dinormalisasikan sehingga nantinya dapat memudahkan pada tahap pembobotan. *Pre-processing* dilakukan untuk menghindari data yang kurang sempurna atau cacat, atau data-data yang tidak konsisten [2]. Tahapan *text pre-processing* Pada penelitian ini diantaranya ditunjukkan pada tabel 1.

Table 1. *Pre-processing*

Sebelum	Sesudah
Tetap gunakan hak pilih andam, tapi jangan lupa disiplin mematuhi protokol kesehatan. Pilkada te... https://t.co/RWRvnlTCY	Tetap gunakan hak pilih andam, tapi jangan lupa disiplin mematuhi protokol kesehatan. Pilkada

1. Menghapus *URL*

Url (<http://www.situsname.com>) dan *email* nama (nama@situs.com)

2. Mengganti *Emoticon*

Pada proses *convert emoticon* adalah mengganti *emoticon* yang ada pada *tweet* berupa kata yang mencerminkan *emoticon*. Daftar konversi *emoticon* dapat dilihat pada tabel 2:

Table 2. Konversi *Emoticon*

Emoticon	Konversi
:) :-) :) :-) =))	Senyum
:D :-D=D	Tawa
:-(:(Sedih
:-) :)	Kedip
:-P :P	Ejek
:-/ :/	Ragu
! :-	Haru

3. *Case Folding*

Tahapan ini adalah tahap dimana *case folding* akan mengubah semua huruf yang ada pada dokumen menjadi bentuk yang sama seperti mengubah semua dokumen ke bentuk *lower case*[10].

4. *Remove number or Punctuation*

Tahapan ini adalah tahapan dimana dokumen dihilangkan nomor dan tanda baca yang di dalam teks.

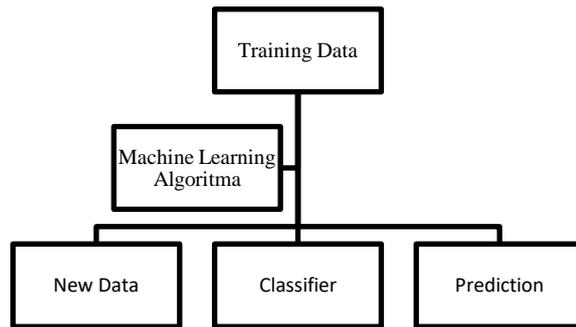
5. *Stopwords*

Stopwords atau kata sambung yaitu kata-kata yang sangat sering dipakai dalam Bahasa Indonesia. Contoh *stopwords* dalam bahasa Indonesia antara lain “saya”, “anda”, “dia”, dan “kita”. *Stopwords* bisa

dihilangkan sebab terkadang kata-kata tersebut terdapat hampir di setiap *tweet* yang ada dengan demikian, menghilangkan *stopwords* tidak akan berpengaruh apapun pada tahap pelatihan model[11].

6. *Tokenizing*

Tahapan ini merupakan tahapan membagi kalimat ke dalam beberapa token[12] contoh kalimat “Tolak Pilkada 2020” setelah dilakukan tokenisasi akan menjadi “Tolak” “Pilkada” “2020”.

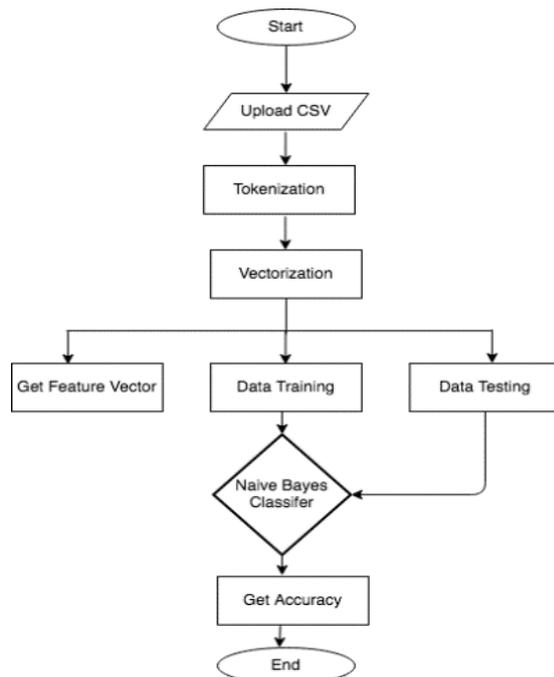


Gambar 3. Sistem *Flow Diagram*

Sistem *Flow Diagram* merupakan bagan yang menunjukkan aliran atau arus dokumen dari satu tahap ke tahap selanjutnya di dalam sistem secara logika, dan dapat menggambarkan tiap-tiap bagian[13] yang ditunjukkan pada gambar 3.

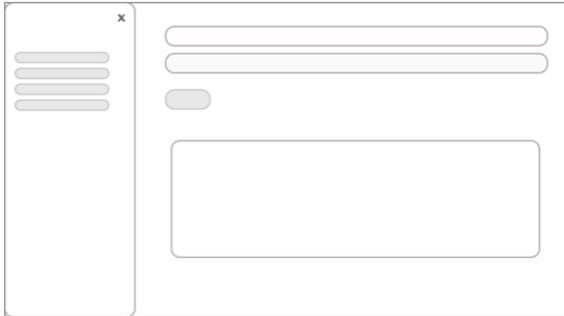
2.6 Desain

Tahap desain yang dilakukan adalah, diagram *activity*, dan perancangan antar muka.



Gambar 4. Diagram *Activity*

Activity diagram adalah tipe diagram khusus dari *statechart* yang berfokus pada sebuah aktivitas atau rangkaian proses yang sesuai dengan metode atau fungsi aktivitas yang telah terjadi. Adapun *activity diagram* ditunjukkan pada Gambar 4[14].

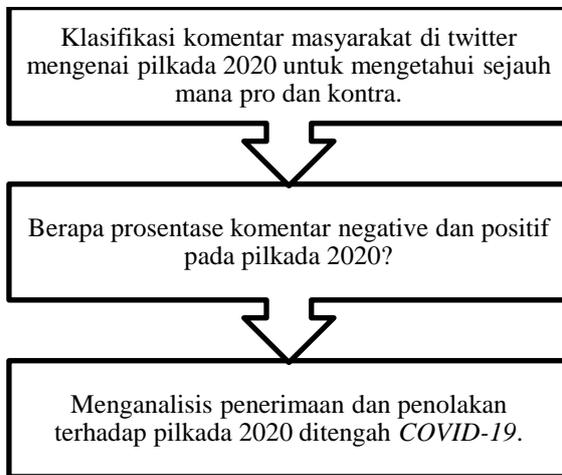


Gambar 5. Perancangan Antar muka

Pada perancangan antar muka atau *user interface* dimaksudkan agar supaya *user* dapat mengerti dengan mudah isi konten pada aplikasi *web* tersebut yang ditunjukkan pada gambar 5.

2.6 Kerangka Berfikir

Berikut adalah gambar dari proses kerangka berfikir dari aplikasi yang di bangun. Mulai dari klasifikasi hingga hasil akhir apakah masyarakat menyetujui pelaksanaan pilkada atau malah sebaliknya yang ditunjukkan pada gambar 6.



Gambar 6. Kerangka Berfikir

2.7 Implementasi

Sentimen analisis berbasis web yang bertujuan untuk mengetahui persentasi positif dan negatif terkait pelaksanaan Pilkada 2020 akan dibangun dengan menggunakan bahasa pemrograman (PYTHON) [14] adalah bahasa pemrograman umum yang ditafsirkan tingkat tinggi. Dibuat oleh Guido van Rossum, dengan *framework Streamlit*. *Streamlit* merupakan pustaka dari *python* sumber terbuka (*open source*) yang mampu mengubah skrip menjadi aplikasi *web* yang dapat ditampilkan dalam beberapa menit. Ini membuatnya sangat mudah untuk

membangun sebuah aplikasi *web* khususnya untuk mereka yang memiliki sedikit pengetahuan tentang kerangka kerja desain *front-end*.

2.8 Pengujian

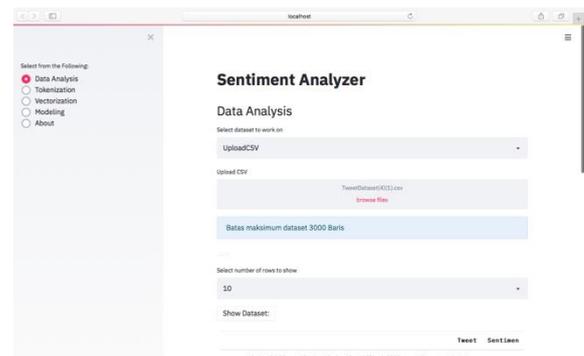
Dalam penelitian ini akan dilakukan pengujian dengan menggunakan metode *Naive Bayes Classifier*. Pengujian dengan metode *Naive Bayes* digunakan untuk mendapatkan sentiment *probability* dalam suatu perangkat lunak, sedangkan mengukur akurasi menggunakan *NMF Vector (Non-negatif Matriks Faktorisasi)*[15]. adalah sekelompok algoritma di analisis multivariat dan aljabar linear di mana matriks V yang di faktorkan ke dalam (biasanya) dua matriks W dan H, dengan properti bahwa ketiga matriks tidak memiliki elemen negatif. *Non-negativitas* ini membuat matriks yang dihasilkan lebih mudah diperiksa dioperasikan dengan benar[16].

3. Hasil Dan Pembahasan

Dalam penelitian ini akan dilakukan implementasi dari hasil perancangan sistem serta perangkat lunak yang dibuat. Implementasi ini menjelaskan mengenai bentuk *interface* dari Aplikasi Sentimen Analisis yang telah di buat dengan *framework streamlit*.

3.1. Halaman Homepage

User pada halaman ini dapat memasukkan data dan mengelola *dataset*, melakukan data analisis, *Tokenization*, *Vectorization* dan *Modeling* yang ditunjukkan pada gambar 7.



Gambar 7. Halaman Depan

Pada halaman gambar 7, menunjukkan *dataset* yang diambil dari *Twitter* menggunakan *Twitter API*. Data yang di dapat sebanyak 1200 *tweet* yang ditunjukkan pada gambar 8.

Pada halaman dataset dibersihkan dari *lowercase* dan huruf *alpha* tanda kutip dan lain-lain contoh “Tolak” “Pilkada” “2020” menjadi Tolak Pilkada 2020 dan ditunjukkan pada gambar 9.

Show Dataset:

	id_str	Text
0	1335541882802249700	jadi duga langgar tempat residen one tangsel bagi c bagi bingkis isi gelas
1	1334429821833601000	bareskrim jaga netralitas tugas polri tenang kampanye h pilkada
2	1335124245605703700	gelar pilkada tinggal hitung salur apd petug
3	1335412397708734500	adama dannypomanto fatmawatirusdi pilwalkotmakassar pilkada
4	1333801656618680300	happy holiday abis uas teenserang worldaidsday wweraw imaceleb ikawangliwanagatligaya oneus snow davido
5	1336125026391584800	kait polemik beri gelar adat tanggap ikeedwin bandarlampung pilkada pilkadasehatnogolput
6	1335150062196813800	cek rekomendasi calon kepala daerah pilkada thread
7	133360335228184000	ikut politik uang ya pilkada cc
8	1335430749227348000	ayo tps rabu desember selamat pekan ya sahabatpenyiaran cc pilkada milu
9	1335865338164998100	gak takut tps terap protokol sehat kota tangan

Gambar 8. Dataset

Select 'TEXT/Reviews' Column:

Text

```

0 bawaslu kota bandar lampung rencana patroli anti...
1 instal sirekap pilkada sirekap kpumelayani pilka...
2 halo sobat pilih yuk simak video ajq perhati sob...
3 bobby-aulia wujud medan yg adab santun harmonis ...
4 besok lupa tps ya pilkada wajib terap pilkada
5 ketua bidang hubung sayap badan dpp partai nasde...
6 bobby-aulia bantu kembang kelompok usaha medan u...
7 desember lupa pilih pimpin yg baik lupa protokol...
8 jadi duga langgar tempat residen one tangsel bag...
9 depok depokjam rumahsakitmahal corona coronaviru...
10 debat publik putar pilih walikota wakil walikota...
    
```

Gambar 9. Halaman Tokenization

Show Top Words:

Select no of Top words

10

	Word
0	pilkada
1	pilih
2	desember
3	pilkadaserentak
4	medan
5	coblos
6	bawaslu
7	covid
8	tps
9	serentak

Gambar 10. Vectorsization Top Words

Pada halaman gambar 10, dataset di cari kata-kata teratas atau terbanyak yang di tweet oleh masyarakat. Untuk setiap kata atau fitur yang dimiliki tweet-tweet tersebut pada selang waktu tertentu akan

dihitung nilai $Tf.Idf$ -nya. Setelah itu diambil beberapa kata atau fitur yang memiliki nilai tertinggi. Hasil akhir dari sistem ini adalah kata atau term yang memiliki nilai paling tinggi yang nantinya akan merepresentasikan topik pemicu nilai sentimen pada tweet tersebut yang ditunjukkan pada gambar 10.

$$W_{x,n} = TfIdf_{x,n} + \alpha W_{x,n-1} \quad (3)$$

Dimana

$W_{x,n}$ = Bobot kata x pada interval ke-n

$TfIdf_{x,n}$ = Nilai Tf-Idf kata x pada interval ke-n

α = Discounted cumulative factor

Select training and testing split: 0.30

Select training and validation split: 0.70

Select one of the pipeline for Model Training: NMF

Load Results

	Type	NMF
0	Train	92.23300970873787
1	Validation	50.79365079365079
2	Test	54.10798122065727

Gambar 11. Halaman Modeling

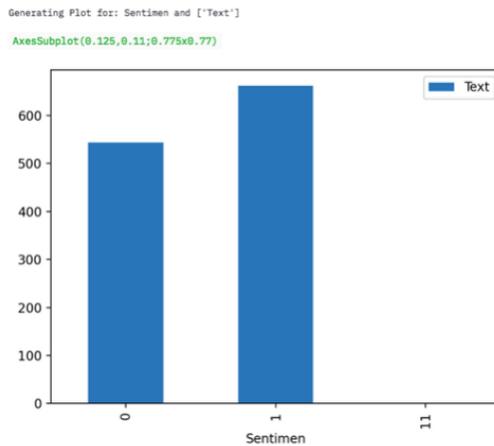
Pada halaman ini dataset dibagi menjadi dua yaitu data *training* dan data uji dengan masing-masing perbandingan 70% dan 30%. Yang menghasilkan nilai akurasi 92,2%. maka dengan penggunaan algoritma *Naive Bayes Classifier* untuk melakukan proses klasifikasi sangat tepat karna menghasilkan jumlah akurasi data yang tinggi yang ditunjukkan pada gambar 11 dan 12.

$$Akurasi = \frac{TP+TN}{TP+FP+TF+FN} \quad (4)$$

	Type	NMF
0	Train	92.23300970873787
1	Validation	50.79365079365079
2	Test	54.10798122065727

Gambar 12. Nilai Accuracy Data

Pada halaman ini menunjukkan hasil *sentiment* masyarakat dimana 0 adalah positif dan 1 adalah negatif, maka dari hasil *sentiment* berikut ini menunjukkan bahwa lebih dari 50% data yang saya ambil menunjukkan hasil *sentiment* negatif yang ditunjukkan pada gambar 13.



Gambar 13. Sentimen Masyarakat

4. KESIMPULAN

Dari penelitian yang telah dilakukan, maka kesimpulan bahwa analisis sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya *Netizen Twitter* terhadap pelaksanaan pilkada 2020 ditengah pandemi *COVID-19*. Setelah dilakukan analisis sentimen, terlihat berapa banyak sentimen yang ditujukan kepada *hashtag* *pilkada2020*. Jika di lihat dari hasil data yang ditunjukkan di atas maka ada baiknya pemerintah menunda pelaksanaan Pilkada dikarenakan adanya pandemi sehingga masyarakat dapat memilih calon yang bermutu dengan tenang serta gembira tanpa adanya ketakutan akan tertular *COVID-19*.

DAFTAR PUSTAKA

- [1] F. D. Anisa, "Enam Bulan Pandemi, Angka Penularan Covid-19 Semakin Tinggi," 2020. <https://www.beritasatu.com/kesehatan/672143/enam-bulan-pandemi-angka-penularan-covid19-semakin-tinggi> (accessed Jan. 14, 2021).
- [2] P. Rizqiah, "Klasifikasi Komentar Twitter Tentang Pengesahan UUMD3 Menggunakan Metode Naïve Bayes," 2018.
- [3] Yudianzer, "Menurun, Segini Jumlah Pengguna Twitter Sekarang - Pos Harian." <https://yudianzer.blogspot.com/2016/12/menurun-segini-jumlah-pengguna-twitter.html> (accessed Jan. 14, 2021).
- [4] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [5] L. Y. T. Suarez, "Machine Learning Prediksi Karakter Pengguna Hastag (#)," pp. 1–27, 2015.
- [6] A. P. Wijaya and H. A. Santoso, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," *J. Appl. Intell. Syst.*, vol. 1, no. 1, pp. 48–55, 2016, [Online]. Available: <https://publikasi.dinus.ac.id/index.php/jais/issue/view/76>.
- [7] A. T. J. Harjanta, A. Syukur, and C. Supriyanto, "Penerapan Pembobotan Atribut Pada Algoritma Naive Bayes Untuk Analisis Sentimen Review Aplikasi Android Dari Google Play," *Jurnal Cyberku*, vol. 11, no. April, pp. 78–89, 2015.
- [8] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.
- [9] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. Sekar, T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 4, pp. 334–339, 2019.
- [10] H. Najjichah, A. Syukur, and H. Subagyo, "Pengaruh Text Preprocessing Dan Kombinasinya," *J. Teknol. Inf.*, vol. 15, no. 1, pp. 1–11, 2019.
- [11] A. R. Nugraha and G. Pramukasari, "Sistem Informasi Akademik Sekolah Berbasis Web Di Sekolah Menengah Pertama Negeri 11 Tasikmalaya," *J. Manaj. Inform.*, vol. 4, no. 2, pp. 1–10, 2017.
- [12] R. P. Mahardikawati and Nurgiyatna, "Sistem Informasi Industri Kecil Menengah Pemerintahan Kabupaten Boyolali Berbasis Website," *J. Tek. Inform (JUTIF)*, vol. 1, no. 2, pp. 53–60, 2020.
- [13] S. Suryono, E. Utami, and E. T. Luthfi, "Klasifikasi Sentimen Pada Twitter Dengan Naive Bayes Classifier," *Angkasa J. Ilm. Bid. Teknol.*, vol. 10, no. 1, pp. 89–96, 2018, doi: 10.28989/angkasa.v10i1.218.
- [14] ICHI.PRO, "Streamlit - Merevolusi Pembuatan Aplikasi Data." <https://ichi.pro/id/streamlit-merevolusi-pembuatan-aplikasi-data-248940993548562> (accessed Jan. 09, 2021).
- [15] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naive Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, pp. 145–148, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [16] I. Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik," *J. Sarj. Inst. Teknol. Bandung Bid. Tek. Elektro dan Inform.*, vol. 1, no. 2, pp. 200–206, 2012.