

Sentiment Analysis On The Presidential Threshold Policy In Elections As A Principle Of Democracy

Jovianus Abel Andreas^{a*)}, Erick Dazki^{b)}

^{a)} Universitas Pradita, Tangerang, Indonesia

^{b)} Universitas Pradita, Tangerang, Indonesia

^{*)}jovianus.abel@student.pradita.ac.id

Article history: received 00 0000000 0000; revised 00 0000000 0000; accepted 00 0000000 0000

Abstract. *Indonesia is a country that adheres to a presidential system of government. Elections are the basis of a democratic country in its implementation. Democracy is the government of reason by the people, for the people, and of the people, hence the people have the highest position in a democracy. The existence of a threshold is controversial in political dynamics because it is considered as a suppression of democratic values. Sentiment analysis is used to evaluate public assumptions about the application of the threshold by filling out a questionnaire that will be filled by subjects who have attitudes towards politics. Naïve Bayes and SVM are the methods used in solving sentiment analysis classification. Data collected through Twitter crawling is integrated with 1500 data from public assumptions about the presidential threshold. Naïve Bayes and SVM methods will be used to classify comment data. Through testing and classification, 784 comments were obtained which will be used as training data. The accuracy obtained from processing is 75.13% for Naïve Bayes and 83.29% for SVM.*

Keywords: *Presidential Threshold;Naïve Bayes; Laplacian;Sentiment Analysis;SVM*

I. INTRODUCTION

Pemilihan umum merupakan faktor utama yang menjadi perwujudan dari kehidupan berdemokrasi dalam sistem pemerintahan di suatu negara[1]. Kedaulatan demokrasi tertinggi berada di tangan rakyat sebagai konsituen. Pasal 1 ayat (2) UUD NRI Tahun 1945 menegaskan bahwa kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar[2]. Peralihan sistem ke era yang lebih demokratis terjadi setelah jatuhnya rezim Orde baru di era Soeharto pada tahun 1998. Perubahan pada aspek amandemen konstitusi 1945 menghasilkan sebuah sistem presidensial. Presiden dan wakil presiden yang dipilih langsung oleh rakyat, peran MPR sebagai Lembaga negara tertinggi tersubordinasi, presiden dan wakil presiden memiliki batas maksimum dua kali masa jabatan selama lima tahun.

Demokrasi sendiri itu adalah sebuah *value* dari hasil perjuangan yang mengedepankan aspek kebebasan, kesetaraan dan persaudaraan, dan sebagai hasil dari consensus antar negara dan rakyat yang merupakan bagian dari konflik atau interaksi kepentingan. Sebuah demokrasi bukan hanya tentang good governance, tetapi bagaimana proses praktis yang paling mungkin untuk dicapai. Proses demokrasi menjadi tantangan tersendiri bagi Indonesia sendiri, karena adanya disparitas sosial yang begitu tinggi dari berbagai wilayah di Indonesia. Pada prakteknya dilapangan terjadi *money politics* di kalangan bawah, dengan adanya presidential threshold, menjadi

semakin meluas praktek politik praktis ini. Penetapan ambang batas mengakibatkan pro-kontra yang memicu diskursus diantara para *elite* politik dan para ahli terkait efektifitas pada perkuatan pada sistem presidensial di Indonesia[3]. Penerapan Presidential Threshold tercantum dalam aturan Bab VI Pasal 222 Undang-Undang Nomor 7 tahun 2017 mengenai prosedural ambang batas calon Presiden yang artinya: “Pasangan Calon Presiden dan Wakil Presiden yang diusung oleh partai politik atau koalisi partai politik peserta pemilu yang memenuhi perolehan kursi paling sedikit 20% dari jumlah kursi DPR atau memperoleh 25% suara sah nasional pada pemilu DPR sebelumnya”[4].

Presidensial Threshold menjadi hal yang diperdebatkan tiap tahun politik sebelum pemilu. Ambang batas pemilu di Indonesia tergolong unik karena pada dasarnya, diterapkan sebagai batas pencalonan presiden, bukan sebagai batas keterpilihan presiden, bahkan pada UUD tidak ada dirumuskan mengenai ambang batas. Presiden terpilih jika mendapat suara (50%+1%), tersebut lebih dari dan sekurang-kurangnya (20% dari 50%+1%), tidak ada batas dengan syarat sebagai perserta pemilu dan partai/gabungan partai. Secara substansial presidensial threshold membatasi jumlah calon presiden itu sendiri menjadi sangat terbatas, seperti pada pemilu 2014 dan 2019, hal ini menyebabkan para pemilih harus memilih diantara calon yang sudah disiapkan oleh para oknum politisi dengan diskusi tukar tambah diruang tertutup yang menyebabkan hegemoni pada kekuasaan.

Terdapat beberapa penelitian terdahulu yang telah dilakukan sebelumnya oleh para peneliti berkaitan dengan sentimen analisis, diantaranya:

(Try Iryanto Saputra,2019) memberikan penelitian analisis sentimen keluhan pengguna kartu sim Indosat menggunakan metode *K-Means* dimana mengklasterisasi kemiripan data yang diperoleh dari hasil data *crawling*. Penelitian dilakukan dengan menampilkan kelompok dari cluster yang didalamnya ada anggota masing-masing cluster yang berbentuk *wordcloud* ke 3 buah yang berbeda. Perbedaan dari penelitian ini terdapat di analisis sentimennya hanya mengklusterisasi kata dari hasil metode yang diterapkan dimana tidak menampilkan hasil keseluruhan dari statement yang ada. Pada penilitian ini analisis sentimen dilakukan untuk mendapat hasil secara keseluruhan dari data yang diolah dengan bantuan klasifikasi *Naïve Bayes* yang memberi hasil lebih luas[5].

(Erwin Yudi Hidayat,2021) menggunakan metodologi DNN (*Deep Neural Network*) untuk analisis sentimen opini terhadap perusahaan publik, dikatakan bahwa pengujian analisis sentimen menggunakan metodologi DNN memberikan proposisi data training dan testing sebesar 90:10 merupakan hasil terbaik. Dengan model 3 hidden layer dengan susunan node tiap *layer* pada model yaitu 128,256,128 node dan menggunakan *learning rate* sebesar 0,005, model mampu menghasilkan nilai akurasi 88.72%. Penggunaan metode DNN untuk analisis sentimen memiliki kompleksitas tersendiri dan proses yang signifikan. Pada penelitian ini dengan kasus yang sama yaitu analisis sentimen, penggunaan metode *Naïve Bayes* untuk penganalisaannya terdapat perbedaan yang membuat hasil dan efektivitas didapatkan sedikit lebih mudah dalam implikasinya[6].

(M.Khairul Anam,2022) memaparkan penelitian mengenai analisis sentimen menggunakan metodologi *Support Vector Machine* (SVM) dan *Social Network Analysis* (SNA) terhadap opini masyarakat mengenai kebijakan publik BPJS. Pengujian dilakukan menggunakan SVM tetapi hasil yang didapatkan masih kurang signifikan, maka digabungkan dengan metode SNA untuk menambah tingkat akurasi yang lebih signifikan.. Hasil dari berbagai analisis dengan penggabungan berbagai metode melalui proses yang cukup memakan waktu, dapat dilihat bawa ke efektifan dalam penggunaan metode SVM ini mempunyai tingkat yang cukup rendah dibanding metode lainnya dalam efisiensi waktu[7].

(Erwin Yudi Hidayat,2022) penggunaan metodologi 1D-Convolutional Neural Network(1D-CNN) dalam proses analisis sentimen terhadap website *female daily review*, dimana platform tersebut memberi paparan dari berbagai kalangan yang menggunakan skincare untuk diberi ulasan

pada website tersebut. Penggunaan CNN digunakan untuk mengolah NLP, dimana menghasilkan sebuah ekspresi dan di identifikasi. Layer pada CNN menggunakan 3 jenis, *convolutional, pooling* dan *fully connected layer*. Penggunaan metode CNN dibanding dengan Naïve Bayes memberi sebuah kompleksitas tersendiri, dimana membutuhkan akurasi yang tepat untuk algoritma dalam mengklasifikasi sentimennya[8].

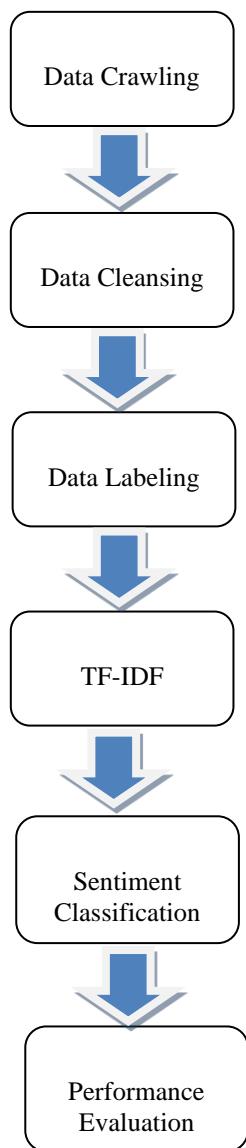
(Habib Hakim Sinaga,2022) mendefinisikan hasil analisis sentimen dengan membandingkan metodologi *Decision tree* dan XGBoost. Penelitian dengan perolahan dataset menggunakan *Random Forest* yang merupakan pengembangan algoritma *Decision Tree* yang menghasilkan akurasi sebesar 75,96%. Masalah pada analisis sentimen dari penilitian ini mengacu pada Covid-19 dimana diperoleh data melalui twitter. Klasifikasi dari data yang diperoleh menjadi acuan utama dalam memproduksi respon masyarakat terhadap vaksin dalam kasus Covid-19[9].

Sosial media merupakan platform interaksi dan komunikasi antar manusia tanpa Batasan jarak dan waktu, perkembangan yang sangat masif ini menimbulkan konflik baru dan permasalahan baru dalam berkomunikasi dan bercengkrama dengan orang lain,mengingat setiap persoalan dan konflik dari manapun dapat dilihat dalam sekejap melalui sosial media[10].Dalam penelitian ini digunakan sebuah sosial media yaitu twitter yang dimana salah satu platform inklusif yang terdapat membahas mengenai isu kebijakan ini secara masif, dibandingkan dengan wawancara secara langsung, melalui platform ini sangat efisien dan efektif[11].

Parameter dalam menilai hal yang berbasis pada penilaian subjektif atau opini terkait permasalahan ini, diperlukan alat ukur terhadap asumsi penilaian sentimen masyarakat terkait kebijakan presidential threshold dalam pemilu 2024. Dalam penelitian ini digunakan 2 metode untuk melihat perbandingan yang memiliki persentase dan akurasi lebih tinggi terkait analisis sentimen. Mekanisme yang diterapkan secara prosedural dengan mengukur dengan sentimen analisis menggunakan metodologi *Naïve Bayes* dan *Support Vector Machine*(SVM). *Naïve bayes* membantu dalam mengelompokan data berdasarkan kesamaan label,serta *Laplacian* digunakan untuk menghindari hasil 0 dalam penggunaan klasifikasi *Naïve Bayes*. *Support Vector Machine* yang merupakan machine learning untuk mengenali pola dalam data dan memprediksi kategori yang tepat dalam teks sebagai yang akan di analisis.

II. METODE PENELITIAN

A. Tahapan Penelitian



Alur 1 Tahapan analisis sentimen

Tahapan yang terjadi pada alur diagram **Alur 1** akan menjadi acuan dari proses klasifikasi data sentimen menggunakan metodologi Naïve Bayes. Dalam tiap proses memiliki sub proses dan penjelasan yang terdifinisikan pada **Table 1**. Pelaksanaan penelitian meliputi 6 tahap. Pada tahap pertama data akan dikumpulkan dari penggunaan *google collab* yang menggunakan Bahasa python, tahap kedua tiap data yang sudah di crawl akan dimodifikasi untuk membersihkan text dari berbagai jenis selain teks sentimen, seperti *link* atau *symbol* yang tak dibutuhkan,pada tahap ketiga, dimana data akan diberi label sentimen secara

manual sebagai data latih dan diberi label positive dan negative pada penerapannya yang akan digunakan nantinya untuk data training, tahap keempat dokumen akan dibersihkan dengan beberapa sub-proses didalamnya berupa, *tokenize, transform cases, filter stopwords dan filter tokens*, tahap kelima dimana data akan diaplikasikan dengan metode sentimen, yaitu *naïve bayes* dan *SVM*, setelah data diaplikasikan metode,pada tahap ke-enam hasil sentimen akan dianalisis menggunakan *K-Fold Cross Validation* dimana data latih akan dievaluasi menjadi beberapa bagian (fold)[12].

Tahapan Penelitian	Detail
Data Crawling	<i>Crawling</i> menggunakan bahasa python menggunakan software <i>Google Collab</i> .
Data Cleansing	Data yang diperoleh setelah crawling, data akan dibersihkan dari berbagai instrument yang tidak diperlukan dalam data sentimen dan melakukan penghapusan data yang duplikat.
Data Labeling	Setelah pemrosesan dan tahapan diatas, text sentimen akan diberi label berupa positif dan negative
Procces Document (TF-IDF)	Terdiri dari sub process yang memiliki 4 tahapan, <i>tokenize</i> , memecah teks menjadi fragmen yang lebih subtil; <i>transform cases</i> yang berperan sebagai proses mengubah teks yang kapital menjadi kecil; <i>filter stopwords</i> yang berfungsi menghapus kata yang bersifat umum dan tak memiliki makna, seperti “dan”, “atau”, dan “yang” atau teks lain yang irrelevant; <i>filter tokens</i> yang berfungsi menghapus teks yang terlalu Panjang atau pendek.
Sentiment Analysis	Bagian proses dari sentiment analisis mengadopsi 2 metode utama sebagai fondasi dasar dan parameter utama yaitu <i>Naïve Bayes</i> dan <i>Support Vector Machine (SVM)</i> . Text sentimen yang terpolarisasi

	menjadi <i>positive</i> dan <i>negative</i> .
Performance Evaluation	Penggunaan metode <i>K-Fold Cross Validation</i> merupakan model yang dilatih menggunakan fold akan diuji dengan fold lainnya, dan akan terus diulang beberapa kali dengan memvariasikan fold yang diugunakan untuk pengujian dan pelatihan. Nantinya hasil pengujian akan digabungkan untuk menghasilkan estimasi kinerja model yang lebih akurat.

Table 1 Detail Proses Tahapan Penelitian

2.2 Naive Bayes Classification

Metodologi utama yang digunakan sebagai pengklasifikasian sentiment pada penelitian mengenai penerapan ambang batas pencalonan presiden. Klasifikasi *Naïve bayes* dicetuskan oleh Thomas Bayes lahir pada tahun (1701-1706) dan ditemukan sendiri oleh Pierre-Simon Laplace (1749-1827). Metode naïve bayes merupakan sebuah metode yang teliti untuk menafsirkan bukti dalam konteks pengalaman atau pengetahuan yang terjadi di masa lampau[13]. Metode klasifikasi yang berbasis pada teorema Bayes. Metode ini digunakan untuk memprediksi kelas dari suatu data berdasarkan fitur-fitur yang dimilikinya[14][15]. Naive Bayes bekerja dengan menghitung probabilitas setiap fitur dan kemudian mengalikan probabilitas tersebut untuk mendapatkan probabilitas kelas yang diinginkan[16], [17]. Bentuk dari rumus teorema naïve bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Pada rumus diatas dapat didefinisikan bahwa, $P(H|X)$ adalah probabilitas posteriori dari hipotesis H berdasarkan data X , $P(X|H)$ adalah probabilitas likelihood dari data X berdasarkan hipotesis H , $P(H)$ adalah probabilitas priori dari hipotesis H . dan $P(X)$ adalah probabilitas evidence dari data X [17][18][19].

2.3 Laplacian Smoothing

Penggunaan metode Laplacian mengehendaki penambahan nilai 1 pada tiap data set yang terdapat pada data training. Hal ini dilakukan dalam rangka menghindari probabilitas bernilai nol pada data training. Variable v adalah jumlah pseudo-counts yang ditambahkan ke setiap

kemungkinan nilai variabel acak untuk menghindari nilai probabilitas nol.

$$P(H|X) = \frac{P(X|H).P(H)+v}{P(X)+v} \quad (2)$$

2.4 Support Vector Machine (SVM)

Merupakan salah satu metode yang merupakan Teknik pembelajaran mesin untuk mengenali pila dalam data dan memprediksi kategori yang presisi pada tiap teks. Secara konsepsional, SVM membagi data kedalam ruang dimensi yang lebih tinggi dan mencari garis pemisah antar kategori berbeda yang terpolarisasi, SVM menganalisa *hyperlane* paling akurat untuk memisahkan data kedalam tiap kategori[20]. *Hyperlane* merupakan sebuah garis yang membelah data dengan jarak terbesar antara dua kategori. Lalu, memprediksi kategori yang presisi untuk tiap teks berdasar posisi terhadap *hyperlane*.

$$f(x) = \text{sign}(\sum_{i=1}^n y_i \alpha_i K(x_i, x) b) \quad (3)$$

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Bagian krusial yang difungsikan sebagai Teknik untuk mengevaluasi pentingnya kata dalam struktur suatu dokumen. TF-IDF melakukan kalkulasi poin untuk tiap kata dalam dokumen yang menunjukkan seberapa penting secara substansial kata tersebut atau secara keseluruhan, dengan perannya poin yang dihitung dengan mengalikan frekuensi tiap kemunculan data dalam dokumen (TF) dilain sisi merupakan frekuensi kemunculan kata tersebut dalam kumpulan dokumen. Semakin tinggi poin TF-IDF suatu kata dapat diartikan bahwa semakin penting peran kata tersebut dalam dokumen secara keseluruhan. Rumus TF-IDF adalah sebagai berikut.

$$tf - idf(t, d) = tf(t, d) * idf(t) \quad (4)$$

Di mana $tf(t, d)$ adalah frekuensi kemunculan kata t dalam dokumen d ; $idf(t, d)$ adalah kebalikan dari frekuensi kemunculan kata t dalam seluruh dokumen dalam kumpulan dokumen. Rumusnya sebagai berikut.

$$idf(t) = \log \left(\frac{N}{df(t)} \right) \quad (5)$$

Dimana N adalah jumlah seluruh dokumen dalam Kumpulan dokumen; $df(t)$ adalah jumlah dokumen dalam Kumpulan dokumen yang mengandung kata t .

III. HASIL DAN PEMBAHASAN

3.1 Data Crawling

Perolehan data melalui tools google collab yang menggunakan Bahasa python melalui sosial media twitter. Pencarian dengan keywords “presidential threshold lang:id”, lang:id yang membantu untuk filter Bahasa Indonesia, selain Bahasa Indonesia tidak akan tercrawling dalam output yang akan dihasilkan. Data yang didapatkan sejumlah 1500 data. Data duplikat dan yang tak berhubungan akan difilterasi menggunakan operator replace dan remove duplicates. Berikut,dalam **Table 2** dibawah telah dipaparkan contoh data yang telah di crawling.

Tanggal Pembuatan	Text
Wed Nov 22 11:55:23	Saya sebenarnya setuju perubahan syarat usia capres dan cawapres selama akar masalahnya diselesaikan dulu yaitu PT 20% (Presidential Threshold) yang membuat pilihan semakin terbatas dan membuat rakyat tidak bisa menggunakan hak politik sepenuhnya itu dihapuskan.
Thu Oct 26 09:08:41	RT@Fahrihamzah Cuma mau bilang #tebalmuka ngomongin presidential threshold kan ente yang ada di senayan dan ikut menyetujui. Rakyat di papua kelaparan dan meninggal dunia puluhan orang tuh . #tolakdungu
Tue Oct 17 04:38:48	Masyarakat Mendesak presidential threshold Nol (0) persen agar Demokrasi Bermanfaat bagi Rakyat dan Oligarki Tak Bisa Kuasai Pemerintah @RamliRizal
Mon Nov 13 13:14:48	@primawansatrio @ainunnajib Kita harus melihat mekanisme konstitusional yang ada bahwa ketentuan presidential threshold itu merupakan ketentuan yang sah secara kontitusi tidak boleh diganggu gugat karena sudah berulang kali dilakukan judicial

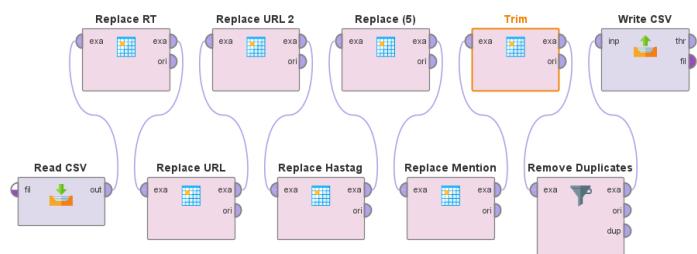
review

Sun Nov 19 17:45:55	@odsandy @mediocrickey kalo presidential threshold digugat kek gitu berarti bakal banyak dong yang daftar jadi bacapres wkwk berpotensi senjata makan tuan. modal dengkul & popularitas non-politic related bakal dengan mudah kepilih kalo gitu wkwk tau sendiri kualitas warga indo demen yang viral-viral.
---------------------	---

Table 2 Contoh dari data crawling

3.2 Data Cleansing

Hasil dari data crawling akan diolah dan dimodifikasi sebagaimana prosesnya untuk menganalisa sentiment. Pada tahap ini data yang tak signifikan dan tak memiliki kaitannya dengan proses analisis sentiment akan di anulir. Data yang didapatkan setelah proses cleansing adalah sebesar 784 total data dari 1500 data crawling yang diambil melalui X (twitter) .Berikut rangkaian operator yang digunakan untuk memproses data adalah sebagai berikut.



Gambar 1 Proses membersihkan data

Proses read CSV yang mana penginputan file yang telah di crawling, operator replace yang melakukan substitusi seperti yang sudah di instruksikan, operator trim yang menghapus spasi kosong di awal dan akhir nominal dan operator remove duplicate agar tidak ada data yang serupa. Lalu, hasil akan di tulis Kembali ke file csv melalui operator Write CSV. Berikut adalah hasil setelah data Cleansing.

Text
Saya sebenarnya setuju perubahan syarat usia capres dan cawapres selama akar masalahnya diselesaikan dulu yaitu PT Presidential Threshold yang membuat pilihan semakin terbatas dan membuat rakyat tidak bisa menggunakan hak politik sepenuhnya itu dihapuskan
Cuma mau bilang ngomongin presidential threshold kan ente yang ada di senayan dan ikut menyetujui Rakyat di papua kelaparan dan meninggal dunia puluhan orang tuh

Masyarakat Mendesak presidential threshold Nol 0 persen agar Demokrasi Bermanfaat bagi Rakyat dan Oligarki Tak Bisa Kuasai Pemerintah

Kita harus melihat mekanisme konstitusional yang ada bahwa ketentuan presidential threshold itu merupakan ketentuan yang sah secara kontitusi tidak boleh diganggu gugat karena sudah berulang kali dilakukan judicial review

kalo presidential threshold digugat kek gitu berarti bakal banyak dong yang daftar jadi bacapres wkwk berpotensi senjata makan tuan. modal dengkul & opularitas non-politic related bakal dengan mudah kepilih kalo gitu wkwk tau sendiri kualitas warga indo demen yang viral-viral.

merupakan ketentuan yang sah secara kontitusi tidak boleh diganggu gugat karena sudah berulang kali dilakukan judicial review

kalo presidential threshold digugat kek gitu berarti bakal banyak dong yang daftar jadi bacapres wkwk berpotensi senjata makan tuan. modal dengkul popularitas non politic related bakal dengan mudah kepilih kalo gitu wkwk tau sendiri kualitas warga indo demen yang viral viral.

Positive

Table 3 Hasil dari pemrosesan data cleansing

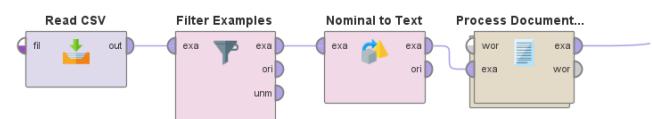
3.3 Data Labeling

Hasil dari data cleansing selanjutnya akan di labeli sentiment yang terpolarisasi menjadi positive dan negative. Tolak ukur positivenya adalah jika text sentiment setuju dengan kebijakan presidential threshold dan sebaliknya negative jika text sentiment tidak setuju. Pelabelan dilakukan secara manual,dengan pembagian dari total 784 data yang diperoleh, 500 data akan diambil sebagai data latih dan 284 data akan diambil sebagai data uji. Berikut contoh text yang sudah dilabeli sentiment.

Table 4 Text yang sudah dilabeli oleh sentimen negative dan positive

3.4 Process Document (TF-IDF)

Tahapan selanjutnya pada proses ini adalah untuk mengolah data yang padanya akan dilakukan tahap akhir sebelum pengklasifikasian sentimen. Proses ini difungsikan agar dapat membantu klasifikasi lebih mudah dibaca secara teknis, yaitu dengan menghilangkan berbagai aspek dan faktor yang irrelevant pada prosesnya atau tidak memiliki nilai yang dibutuhkan dalam pengklasifikasian sentimen. Sebelum itu ada beberapa proses yang perlu dilakukan sebelumnya, dapat dilihat pada proses dibawah ini.



Gambar 2 Tahapan pra-proses sebelum TF-IDF

Pada **Gambar 2** dapat dilihat terdapat beberapa tahapan, pertama adalah operator read CSV yang didapatkan dari proses sebelumnya pada **Gambar 1** yaitu write CSV; kedua adalah filter examples yang merupakan pengambilan data yang telah diberi label sentimen; ketiga adalah nominal to text untuk mengubah parameter yang masih berbentuk polynominal menjadi text agar bisa diproses dalam tahapan TF-IDF, lalu keempat yaitu pemrosesan dokumen yang terdapat 4 sub-proses didalamnya. Pada **Gambar 3** bisa dilihat bahwa ada 4 operator tahapan sub-proses pada process document yang terjadi yaitu :

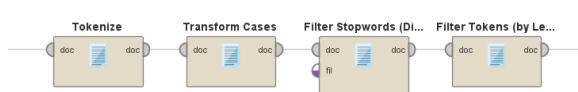
1. Tokenize : Melakukan fragmentasi teks, khalayak kata atau frasa. Tokenize dilakukan dengan menggunakan spasi, tanda baca, atau karakter khusus lainnya sebagai pemisah. Contoh: “Cuma”, “mau”, “bilang”, “ngomongin”, “presidential”, “threshold”, “kan”, “ente”, “yang”, “ada”, “di”, “senayan”, “dan”, “ikut”, “menyetujui”, “Rakyat”, “di”,

“papua”, “kelaparan”, “dan”, “meninggal”, “dunia”, “puluhan”, “orang”, dan “tuh”.

2. Transform Cases : Melakukan transformasi semua karakter menjadi huruf kecil atau besar, yang dalam kontekstualisasi prosesnya membantu analisis teks. Contoh :“cuma mau bilang ngomongin presidential threshold kan ente yang ada di senayan dan ikut menyetujui rakyat di papua kelaparan dan meninggal dunia puluhan orang tuh”.

3. Filter Stopwords : Sebuah proses menghapus kata yang umum dan tak memiliki makna yang terkandung dalam teks analisis. Contoh :“Cuma mau bilang ngomongin presidential threshold kan ente yang ada di senayan dan ikut menyetujui Rakyat di papua kelaparan dan meninggal dunia puluhan orang tuh”.

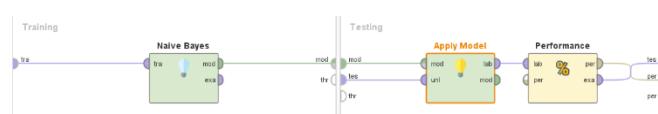
4. Filter Tokens : Proses menghapus token yang tidak relevan atau tidak diinginkan dari teks. Berupa angka,karakter khusus, atau kata-kata yang tidak relevan dalam teks yang di analisa.



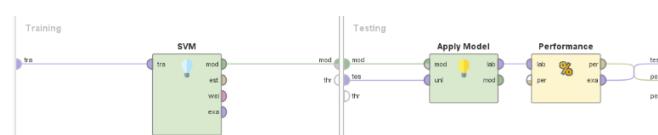
Gambar 3 Tahapan Sub-proses dalam process document

3.5 Klasifikasi Sentiment

Proses yang dilakukan pada tahap ini adalah mengolah data sentiment yang berjumlah 500 data latih dan 284 data uji. Beberapa tahap yang telah dilakukan sebelumnya adalah agar pada proses klasifikasi dapat memberi akurasi dan perhitungan yang tepat untuk model ini dengan metode naïve bayes dan SVM. Berikut operator yang digunakan dalam proses klasifikasi sentiment.



Gambar 4 Operator untuk klasifikasi menggunakan metode Naïve Bayes



Gambar 5 Operator untuk klasifikasi menggunakan metode SVM

3.6 Performance Evaluation

Evaluasi yang dilakukan untuk mengetahui performa statistik yang dihasilkan dari model machine learning pada **Gambar 4** dan **Gambar 5** diperlukan menggunakan metode K-Fold Cross Validation. Parameter yang digunakan dimana nilai K merupakan jumlah besaran data yang terpisah menjadi latih dan uji. Pada hal ini digunakan 10-fold cross validation. Sub proses yang dapat dilihat pada **Gambar 4** dan **Gambar 5**. Disebelah kiri yang merupakan latihan dapat dilihat operator metode yang difungsikan untuk membuat algoritma untuk klasifikasi. Disebelah kanan yang merupakan bagian dari pengujian dari data yang tidak berlabel dari bagian latihan untuk menguji performa model. Apply model yang difungsikan untuk menguji data yang belum dilabeli oleh sentiment. Setelahnya data dari pemrosesan akan masuk ke tahap pengevaluasian performa untuk memberi hasil dari data yang diuji. Berikut hasil dari tabel confusion matrix dari pemrosesan menggunakan metode K-Fold Cross Validation.

	True Positive	True Negative	Class Precision
Pred. Positive	262 (TP)	65 (FP)	80.12%
Pred. Negative	130 (FN)	327 (TN)	71.55%
Class Recall	66.84%	83.42%	

Table 5 Hasil dari perhitungan menggunakan operator Cross Validation berupa Conffusion Matrix menggunakan perhitungan Naïve Bayes

	True Positive	True Negative	Class Precision
Pred. Positive	318 (TP)	60 (FP)	83.72%
Pred. Negative	74 (FN)	332 (TN)	82.87%
Class Recall	82.65%	83.93%	

Table 6 Hasil dari perhitungan menggunakan operator Cross Validation berupa Conffusion Matrix menggunakan perhitungan SVM

Pada **Table 5** dan **Table 6**. True Positive (TP) adalah jumlah data positif yang diprediksi benar oleh model. False Positive (FP) adalah jumlah data negatif yang diprediksi salah oleh model. False Negative (FN) adalah jumlah data positif yang diprediksi salah oleh model. True Negative (TN) adalah jumlah data negatif yang diprediksi benar oleh model. Hasil dari confusion matrix **Table 5** ditunjukkan bahwa jumlah True Positive adalah 262, False Positive adalah 65, True Negative adalah 327, dan False Negative 130. Pada **Table 6** ditunjukkan bahwa jumlah True Positive adalah 318, False Positive adalah 60, True Negative adalah 332, dan False Negative 74. Class Precision mengukur seberapa akurat model dalam mengkalkulasikan kelas positive,sedangkan Class Recall mengidentifikasi

semua kelas positive dalam data set. Berikut perhitungan rumus dari Class precision dan Class Recall.

$$\text{Class Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Class Recall} = \frac{TP}{TP+FN} \quad (7)$$

Perhitungan yang dilakukan oleh operator menggunakan K-Fold Cross Validation, dengan pelatihan dan pengujian pada model sebanyak k kali. Dalam rangka menemukan akurasi terbaik yang dihasilkan dalam metode ini, data akan di uji menggunakan K-2 sampai K-10, data akan dipaparkan pada **Table 7**.

K-Fold	SVM	Naïve Bayes
K-2	79.46%	75.13%
K-3	79.08%	74.36%
K-4	81.63%	74.62%
K-5	81.38%	73.85%
K-6	83.29%	72.96%
K-7	82.53%	74.23%
K-8	82.14%	72.96%
K-9	82.91%	73.85%
K-10	82.28%	72.72%

Table 7 Perbandingan hasil dari perhitungan K-Fold dengan dua metode yang berbeda

Pada **Table 7** dapat dilihat hasil dari K-Fold dari metode Naïve Bayes dan SVM, hasil akurasi terbesar dari 2 metode tersebut didapatkan pada K-6 adalah 83.29% menggunakan metode SVM dan K-2 adalah 75.13% menggunakan metode Naïve Bayes. Pada **Table 5** dan **Table 6** yang mana perwakilan dari tiap metode yang memiliki akurasi paling tinggi. Dapat disimpulkan metode yang secara total memiliki akurasi terbaik dalam pengklasifikasian sentiment adalah metode Support Vector Machine (SVM).

IV. CONCLUSIONS

The paper Penelitian ini mengadaptasi untuk Analisa sentiment mengenai isu kebijakan presidential threshold dalam pemilu sebagai asas demokrasi yang terpolarisasi menjadi sentiment positive dan negative. Dari hasil yang didapatkan dari pengujian menggunakan metode Naïve Bayes dan SVM, didapatkan akurasi sebesar 83.29% menggunakan metode SVM dan 75.13% menggunakan

metode Naïve Bayes. Pada hasilnya SVM mendapat akurasi yang paling tinggi, bahkan dengan keterbatasan data yang dimiliki.

Sentiment analysis terhadap presidential threshold dapat membantu dalam melihat polarisasi mengenai dinamika politik khususnya pada kebijakan ini. Penggunaan metode ini juga merupakan salah satu algoritma yang dibutuhkan dalam menganalisa sentiment. Mengenai keterbatasan mengenai data yang di crawling menggunakan Bahasa python melalui media sosial twitter, yang mana dari kebijakan twitter membatasi jumlah maksimum yang dapat di ambil datanya, untuk mengambil data lebih banyak diperlukan subscription lebih yang memiliki harga cukup tinggi.

Peneliti berharap analisis mengenai kebijakan presidential threshold ini dapat dikembangkan kedepannya, terutama dalam menganalisis agar memiliki akurasi yang lebih tinggi, khususnya dalam mencari data yang lebih berkualitas, agar dapat terpampang dengan jelas polarisasi yang lebih asli mengenai kebijakan ini. Seiring dengan perkembangan kebijakan dan terlebih lagi memasuki tahun politik, kebijakan ini menjadi sangat substansial dalam dinamika politik di Indonesia dan akan terus berlanjut sampai hasil yang diinginkan oleh mayoritas adalah presidential threshold 0% tanpa adanya ambang batas apapun.

REFERENCES

- [1] A. M. Al Mas'udah, "The Presidential Threshold As An Open Legal Policy In General Elections In Indonesia," *Prophetic Law Review*, vol. 2, no. 1, Jun. 2020, doi: 10.20885/plr.vol2.iss1.art3.
- [2] A. Fitri and W. Setiadi, "Presidential Threshold dalam Pemilihan Umum Serentak: Kemunduran Demokrasi Konstitusional?," 2022.
- [3] S. D. Ambarwati, M. R. Saifulloh, and S. M. S. Artonang, "REKONSTRUKSI SISTEM PRESIDENTIAL THRESHOLD DALAM SISTEM PEMILU DI INDONESIA (STUDI PERBANDINGAN SISTEM PRESIDENTIAL THRESHOLD INDONESIA DAN BRAZIL)," Bulan Kedelapan, 2020. [Online]. Available: <https://jhlg.rewangrencang.com/>
- [4] V. Anggara, "Dinamika Presidential Threshold dalam Sistem Presidensial di Indonesia," 2019.
- [5] T. I. Saputra and R. Arianty, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS SENTIMEN KELUHAN PENGGUNA INDOSAT," *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 191–198, 2019, doi: 10.35760/ik.2019.v24i3.2361.
- [6] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp.

- 108–118, Sep. 2021, doi:
10.25077/teknosi.v7i2.2021.108-118.
- [7] M. K. Anam, M. I. Mahendra, W. Agustin, R. Rahmaddeni, and N. Nurjayadi, “Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA),” *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 11–28, Feb. 2022, doi:
10.29407/intensif.v6i1.15870.
- [8] E. Y. Hidayat and D. Handayani, “Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 153–163, Jan. 2023, doi:
10.25077/teknosi.v8i3.2022.153-163.
- [9] H. H. Sinaga and S. Agustian, “Pebandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 107–114, Dec. 2022, doi:
10.25077/teknosi.v8i3.2022.107-114.
- [10] M. O. Odim, A. O. Ogunde, B. O. Oguntunde, and S. A. Phillips, “Exploring the performance characteristics of the naïve bayes classifier in the sentiment analysis of an Airline’s social media data,” *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 4, pp. 266–272, Jul. 2020, doi:
10.25046/aj050433.
- [11] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis,” *J Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-018-0164-1.
- [12] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes,” *Information (Switzerland)*, vol. 12, no. 5, May 2021, doi: 10.3390/info12050204.
- [13] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, “Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2019. doi: 10.1088/1742-6596/1192/1/012024.
- [14] M. Z. Nafan and A. E. Amalia, “Kecenderungan Tantangan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 3, no. 4, p. 268, Oct. 2019, doi:
10.30865/mib.v3i4.1283.
- [15] N. Ramadhan and N. Fajarianto, “Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naïve Bayes dan Smoothing Laplace,” *JURNAL SISTEM INFORMASI BISNIS*, vol. 10, no. 2, pp. 228–234, Dec. 2020, doi:
10.21456/vol10iss2pp228-234.
- [16] D. Berrar, “Bayes’ theorem and naive bayes classifier,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 403–412. doi:
10.1016/B978-0-12-809633-8.20473-1.
- [17] B. Liu, “Sentiment Analysis and Opinion Mining,” Morgan & Claypool Publishers, 2012.
- [18] M. Nababan *et al.*, “The diagnose of oil palm disease using Naive Bayes Method based on Expert System Technology,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1742-6596/1007/1/012015.
- [19] S. Samsir, K. Kusmanto, A. H. Dalimunthe, R. Aditiya, and R. Watrifiantos, “Implementation Naïve Bayes Classification for Sentiment Analysis on Internet Movie Database,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 1–6, Jun. 2022, doi: 10.47065/bits.v4i1.1468.
- [20] Y. Surya, S. Al Faraby, and M. Dwifebri, “Analisis Sentimen Terhadap Ulasan Film Menggunakan Word2Vec dan SVM.”