
ANALISIS SENTIMEN *TWITTER* TERHADAP TOKOH PUBLIK DENGAN ALGORITMA *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE*

Tanthy Tawaqalia Widowati

Fakultas Ilmu Komputer, Program Studi Teknik Informatika
Universitas Mercu Buana

Email: 41516010109@student.mercubuana.ac.id

Mujiono Sadikin

Fakultas Ilmu Komputer, Program Studi Teknik Informatika
Universitas Mercu Buana

Email: mujiono@mercubuana.ac.id

ABSTRAK

Salah satu media sosial yang berkembang adalah *Twitter*. Media sosial *Twitter* mempermudah masyarakat untuk bebas berpendapat melalui cuitan atau biasa disebut dengan *tweets*. *Netizen* dengan bebas menyampaikan opini pribadinya untuk topik apapun, termasuk persepsi terhadap tokoh publik. Artikel ini menyajikan hasil penelitian dan analisis sentimen masyarakat (*netizen*) terhadap tokoh publik, Nadiem Makariem sebagai Menteri Kementerian Pendidikan dan Kebudayaan baru. Penelitian ini menggunakan teknik *data mining* yang bertujuan untuk membandingkan hasil klasifikasi dari opini masyarakat yang dituliskan di *Twitter*. *Dataset* yang digunakan berasal dari *tweets* dengan kata kunci “nadiem makariem”, “kemendikbud” dan “pak nadiem”. *Tools RapidMiner* digunakan untuk membantu tahap *pre-processing* dan klasifikasi menggunakan dua metode yaitu, *Naive Bayes* dan *Support Vector Machine* dengan evaluasi *k-fold cross-validation*. Dari hasil uji coba diketahui bahwa untuk kasus yang diteliti, metode *Naive Bayes* menghasilkan kinerja yang lebih baik dengan *accuracy* 91.48%, *precision* 89.28% dan *recall* 91.58%.

Kata kunci: analisis sentimen, *opinion mining*, *naive bayes*, *SVM*, *twitter*, nadiem makariem

ABSTRACT

One of the developing social media is Twitter. Twitter social media makes it easy for people to be free to express their opinions through tweeting or commonly called tweets. Netizens freely express their personal opinions on any topic, including perceptions of public figures. This article presents the results of research and analysis of public sentiments towards public figure, Nadiem Makariem as the new Minister of Ministry of Education and Culture. This study uses data mining techniques that aim to compare the results of the classification of public opinion written on Twitter. The dataset used comes from tweets with the keywords "nadiem makariem", "kemendikbud" and "pak nadiem". RapidMiner tools are used to assist the pre-processing and classification stages using two methods, Naive Bayes and Support Vector Machine with k-fold cross-validation evaluation. From the trial results it is known that for the cases studied, the Naive Bayes method produces better performance with an accuracy of 91.48%, precision 89.28% and recall 91.58%.

Keywords: *sentiment analysis; opinion mining; naive bayes; SVM; twitter, nadiem makariem*

1. PENDAHULUAN

Sebagian besar masyarakat dari berbagai usia dan tempat menghabiskan banyak waktu menggunakan jejaring sosial sebagai media opini, salah satunya adalah *Twitter*. *Twitter* merupakan

salah satu media sosial paling populer di Indonesia dan Indonesia merupakan negara dengan pengguna *Twitter* kelima teraktif di dunia [1]. Pengguna *Twitter* membuat *posting* yang disebut dengan *tweet*, maksimal 140 karakter. Melalui *tweet*, pengguna *Twitter* dapat saling berinteraksi dan berbagi pendapat atau opini.

Sesuai dengan kondisi masyarakat, *tweet* di *Twitter* bisa mewakili berbagai opini, mulai dari opini tentang sebuah produk, layanan, tokoh publik, maupun isu politik. Isu politik yang sedang panas belakangan ini adalah dimulainya periode ke dua pemerintahan yang dipimpin Presiden Joko Widodo. Berawal dari terbentuknya kabinet baru, yaitu Kabinet Indonesia Maju. Kabinet yang baru dibentuk tersebut tidak luput dari komentar publik dalam media sosial *Twitter*. Terpilihnya Nadiem Makariem sebagai Menteri Kementerian Pendidikan dan Kebudayaan (mendikbud) menuai berbagai macam komentar, mulai dari pujian, kritik, saran, sindiran bahkan ujaran ketidaksetujuan karena beliau bukan merupakan tokoh dalam bidang pendidikan.

Cuitan “Pak Nadiem” sempat menjadi *rending topic* di *Twitter* pada Sabtu, 23 November 2019. “Nadiem Makariem” juga menjadi *keyword* tokoh yang paling banyak dicari di *google* di akhir tahun 2019 [2]. Dengan banyaknya pengguna *Twitter* yang menyampaikan opini-opini tersebut, maka dapat dimanfaatkan untuk mencari sebuah informasi dari polaritas positif dan negatifnya opini-opini tersebut. Penambahan informasi membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu berbagai pihak [3]. Metode penelitian yang dapat digunakan untuk menganalisis opini – opini *tweets* tersebut adalah analisis sentimen.

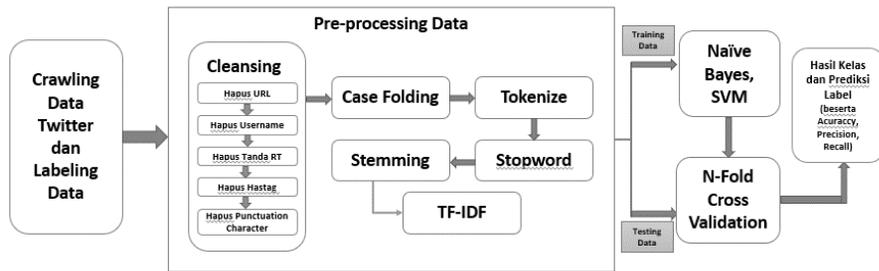
Analisis sentimen atau *opinion mining* adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [4]. Proses analisis data dapat dilakukan melalui pendekatan *machine learning* [5]. Dalam penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap tokoh publik. Analisis sentimen pada penelitian ini menggunakan perbandingan klasifikasi algoritma *Naive Bayes* dan *Support Vector Machine*.

Analisis sentimen ini sudah pernah dilakukan pada beberapa penelitian. Penelitian sebelumnya yang dilakukan oleh Nanang Ruhjana tahun 2019 menganalisis sentimen terhadap penerapan sistem plat nomor ganjil/genap pada twitter dengan metode klasifikasi *Naive Bayes*. Penelitian ini menghasilkan *accuracy* 86,67%, *precision* 71,43% dan *recall* 80,00% [6]. Penelitian selanjutnya juga pernah dilakukan oleh Deni Rusdian dan Didi Rosiyadi tahun 2019. Pada penelitian ini dilakukan analisis sentimen terhadap tokoh publik menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*. Dari hasil penelitian ini diperlihatkan bahwa algoritma *Support Vector Machine* memiliki tingkat keakurasian lebih besar dibanding *Naive Bayes*. *Support Vector Machine* memiliki tingkat keakurasian sebesar 73.96% sedangkan untuk algoritma *Naive Bayes Classifier* memiliki tingkat keakurasian sebesar 71.94% [7].

Berbeda dari penelitian yang telah dilakukan di atas, dalam penelitian ini dilakukan analisis sentimen terhadap tokoh publik Nadiem Makariem, untuk dapat menganalisis opini masyarakat (*netizen*) melalui *tweet* mereka. Lalu mengklasifikasinya ke dalam kategori positif atau negatif menggunakan dua metode, yaitu *Naive Bayes* dan *Support Vector Machine*. Kemudian membandingkan hasilnya untuk mengetahui metode mana yang memberikan hasil terbaik dalam hal *accuracy*, *precision* dan *recall*.

2. METODOLOGI PENELITIAN

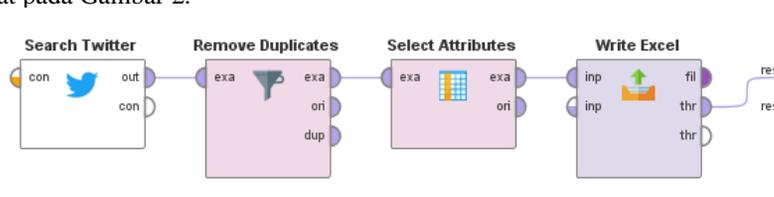
Dalam menganalisa sentimen dan mengetahui akurasinya, ada beberapa tahapan untuk mendapatkan hasil yang terbaik. Tahapan yang dilakukan adalah pengumpulan dan pelabelan data, *pre-processing*, analisis sentimen dan evaluasi untuk mendapatkan *accuracy*, *presision*, dan *recall*. Keseluruhan metodologi analisis sentimen yang dilakukan dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Metodologi Penelitian Analisis Sentimen

2.1 Pengumpulan dan Pelabelan Data

Tahap pertama yang dilakukan pada proses analisis sentimen ini adalah pengumpulan data. Data diambil dari Twitter dengan proses *crawling* yang dilakukan menggunakan *Twitter API* dengan *tools RapidMiner Studio* versi 9.5.0. *Tweets* yang diambil adalah *tweets* dengan kata kunci “nadiem makariem”, “kemendikbud”, dan “pak nadiem”. Kemudian data hasil *crawling* diberi label secara manual, label untuk sentiment *positive* atau *negative*. Model yang digunakan untuk *crawling* data *Twitter* terlihat pada Gambar 2.



Gambar 2. Proses Crawling Data

Gambar 2 menunjukkan pengambilan data dari *Twitter* menggunakan operator "*Search Twitter*". Data *tweets* yang didapatkan kemungkinan banyak yang sama, maka hapus semua tweet duplikat dengan operator "*Remove Duplicate*". Peneliti hanya mengambil atribut "*Text*" yang berisi *tweets* dengan menggunakan operator "*Select Attributes*", lalu menyimpan data dalam *file excel* menggunakan operator "*Write Excel*".

Tahap selanjutnya adalah memberi label, pelabelan dilakukan untuk membagi data menjadi dua kelas sentimen yang akan digunakan, *positive* dan *negative*. Kriteria yang digunakan dalam melabeli *tweets* adalah jika *tweet* mengandung kata positif, kalimat yang diungkapkan netral dan bermaksud menyetujui Nadiem Makarim sebagai mendikbud, dikategorikan sebagai *positive*. Sedangkan jika *tweet* mengandung kata negatif dan bermaksud tidak menyetujui Nadiem Makarim sebagai mendikbud, *tweet* akan dikategorikan sebagai *negative*. Terlihat seperti Tabel 1 berikut ini adalah contoh *dataset* yang telah diberi label:

Tabel 1. Contoh Dataset Dengan Label

<i>Text</i>	<i>Sentiment</i>
Gatau kenapa suka aja liat perkembangan bapak nadiem makarim dan kinerja beliau di kemendikbud. Ditunggu ya pak gebrakannya. #nadiemmakarim #kemendikbud https://t.co/Xzf91zHK41	<i>Positive</i>
@detikcom Sangat setuju pak Nadiem Makarim,selama ini terjadi salah kaprah tentang penggunaan IT dlm pembelajaran.Penggunaan komputer,lab bahasa adalah sbg alat bantu bukanlah kunci keberhasilan pembelajaran.Tujuan pembelajaran(kompetensi)sdh dituangkan di RPP,hal itulah yg hrs dicapai.	<i>Positive</i>

RT @blogger_eksis: Kenapa Nadiem Makarim yang jadi Menteri Pendidikan?

Negative

Bukannya @AdamasBelva aja yang udah punya jejak di @ruangguru

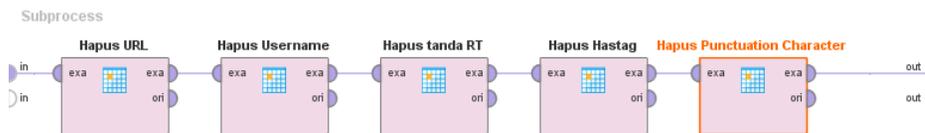
RT @vincenrcrd: Nadiem Makarim Mundur #NadiemMundurAja

Negative

<https://t.co/VyMr5oFvdg>

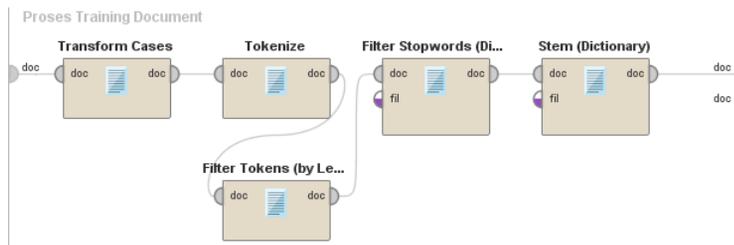
2.2 Pre-processing

Setelah memberi label pada data, tahap selanjutnya adalah *pre-processing*. Kedua *data training* dan *data testing* harus melewati proses *pre-processing* terlebih dahulu. *Pre-processing* diperlukan sebelum proses klasifikasi agar dimensi *vector space model* menjadi lebih kecil. Dengan memperkecil dimensi *vector space model*, proses klasifikasi akan menjadi lebih cepat. Tujuan dilakukannya *pre-processing* ini untuk menyeragamkan dan mengurangi volume kata [8]. Tahap ini dilakukan untuk membuat data siap dan layak untuk dilakukan analisis. Tahapan dalam preprocessing ini dibagi menjadi dua tahap, pada *subprocess* sebelum data dibagi menjadi *data training* dan *data testing*, dan juga *pre-processing* pada *document process*. Tahapan *cleansing* yang dilakukan pada *subprocess* adalah Hapus URL, Hapus Username, Hapus tanda RT, Hapus *Hashtag* dan Hapus *Punctuation Character*. Operator tahap pertama *pre-processing* ditampilkan pada Gambar 3 berikut:



Gambar 3. Proses *Cleansing Data*

Sedangkan tahapan kedua, preprocessing dalam "*Document Process*" menggunakan operator "*Transform Case*", "*Tokenize*", "*Filter Stopword (Dictionary)*" dan "*Stem (Dictionary)*" yang menggunakan *dictionary* dalam bahasa Indonesia. Seperti yang terlihat pada Gambar 4 berikut:



Gambar 4. Proses *Pre-processing dalam Document Process*

Berikut ini adalah penjelasan dari tahapan-tahapan *preprocessing* di atas:

1) *Cleansing*

Cleansing adalah tahap di mana karakter dan tanda baca yang tidak diperlukan dihapus dari teks [9]. *Cleansing* bertujuan untuk mengurangi gangguan/noise pada *dataset*. Tahap-tahap pada *cleansing* ini adalah Hapus URL, Hapus *Username*, Hapus tanda RT dan Hapus *Hashtag*. Proses tersebut menggunakan *regular expression* (regex) dalam parameter operator RapidMiner. Regrex yang digunakan adalah “`http\S+|\S-co\S+`”, “`@\S+`”, “`RT`”, “`#\S+`”, dan “`#\S+[-!\"#$%&'()*+.,/:;<=>?@\[\]_`{}~]`”. Contoh karakter yang dihilangkan seperti URL, tag (#), *username* (@), RT, juga tanda baca seperti titik (.), koma (,) dan tanda baca lainnya.

2) *Case Folding*

Transform case merupakan operator untuk *case folding*. Proses ini akan merubah seluruh karakter dalam dokumen menjadi *upper case* atau *lower case*, untuk penelitian ini menggunakan *lower case*. *Case folding* dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil) [8]. Contohnya “Nadiem Makariem berpotensi” menjadi “nadiem makariem berpotensi”

3) *Tokenization*

Tokenisasi merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna [8]. Tokenisasi memisah setiap kata yang menyusun kalimat. Misalkan “kementerian pendidikan dan kebudayaan”, setelah di tokenisasi menjadi “kementerian”, “pendidikan”, “dan”, “kebudayaan”.

4) *Filter Token*

Filter Token setelah tahap tokenisasi adalah menyeleksi hasil token-token dengan syarat minimal dan maksimal karakter. Untuk penelitian ini, kata hasil tokenisasi yang diambil hanya kata yang memiliki jumlah 2-25 karakter. Jadi, hanya kata-kata yang memiliki 2 sampai 25 karakter yang akan diambil dari proses ini.

5) *Stopword Removal*

Stopword Removal merupakan proses menghilangkan kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan. Kata yang dihilangkan adalah kata yang ada dalam *dictionary* yang berisi daftar kata (*stopword list*). Contoh *stopword* yang digunakan adalah: atau; ada; adalah; ke; kembali; kemudian; kepada; oleh; yang; dll.

6) *Stemming*

Stemming merupakan tahapan untuk melakukan proses mengubah kata yang berimbuhan infix maupun suffix menjadi sebuah kata dasar yang akan lebih mengandung sebuah makna untuk peroleh suatu informasi sehingga komentar akan menjadi lebih spesifik dalam pengkategorian [10]. Contoh *stem list* yang digunakan adalah: menepatkan | tepat; penyogokan | sogok; ketakutan | takut; menyepakati | sepakat; kesukarelaan | sukarela; strategisnya | strategis; pembedulan | betu; menjengkelkan | jengkel; dll.

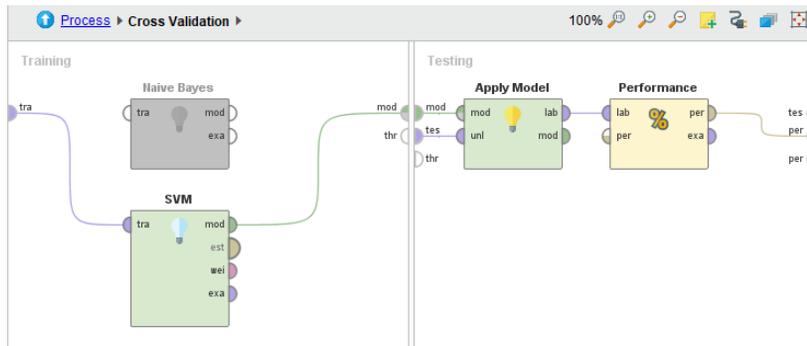
7) *Pembobotan Kata*

Pembobotan kata adalah mekanisme untuk memberi skor pada frekuensi kemunculan kata dalam dokumen teks [9].

- *Term Frequency – Inverse Document Frequency (TF-IDF)*

TF-IDF merupakan algoritma pembobotan tersusun dari dua nilai yang berasal dari dua algoritma dengan pembobotan yang berbeda, yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* [11]. Pada dasarnya TF-IDF bekerja dengan menghitung frekuensi relatif dari suatu kata yang muncul pada sebuah dokumen dibandingkan dengan proporsi *inverse* dari kata tersebut yang muncul pada seluruh kumpulan dokumen [12]. Secara intuitif, perhitungan ini dapat digunakan untuk mengetahui seberapa relevan kata tersebut pada sebuah dokumen tertentu.

2.3 Klasifikasi Analisis Sentimen



Gambar 5. Klasifikasi Analisis Sentimen

Setelah *pre-processing data*, tahap selanjutnya adalah klasifikasi analisis sentimen. Tahap ini adalah tahap untuk memberikan pelatihan dan mengimplementasikan berbagai algoritma *machine learning*. Gambar 5 menunjukkan isi dari operator "Cross Validation" dalam aplikasi RapidMiner. Dalam proses ini, digunakan dua operator klasifikasi berbeda untuk perbandingan, operator klasifikasi "Naive Bayes" dan "SVM". Setelah eksperimen dan model yang diusulkan dibuat, selanjutnya dilakukan eksperimen dengan menguji model yang ada dengan dataset yang sudah dikelompokkan yang menjadi *data training* dan *data testing*. Operator "Performance" adalah operator untuk menampilkan hasil *accuracy*, *precision* dan *recall*.

2.3.1 Naive Bayes

Naive Bayes adalah salah satu algoritma *machine learning* yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayesian. Penggunaan teorema Bayes dalam algoritma *Naive Bayes* adalah dengan menggabungkan probabilitas sebelumnya dan probabilitas kondisional dalam formula yang dapat digunakan untuk menghitung probabilitas setiap klasifikasi yang mungkin [14].

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

- X = data dengan kelas tidak dikenal
- H = hipotesis data X adalah kelas khusus
- P(H|X) = probabilitas hipotesis H didasarkan pada kondisi X
- P(H) = probabilitas hipotesis H
- P(X|H) = probabilitas hipotesis X didasarkan pada kondisi H
- P(X) = probabilitas X

1.1.7 Support Vector Machine

Support Vector Machine (SVM) adalah metode untuk menganalisa data dan mengenali pola yang bisa digunakan untuk pengklasifikasian. Penelitian ini berpusat pada klasifikasi biner yang terdiri dua kelas. Kelas tersebut adalah *positive* atau *negative* [15]. SVM adalah metode klasifikasi yang memisahkan data secara optimal dengan membangun *hyper-planes* dalam ruang input multidimensi. *Hyper-planes* terbaik dapat didefinisikan dengan mengukur batas-batas di antara keanggotaan kelas [16]. SVM pada awalnya digunakan untuk klasifikasi data numerik, tetapi ternyata SVM juga sangat efektif dan cepat untuk menyelesaikan masalah-masalah data teks. Data teks cocok untuk dilakukan klasifikasi dengan algoritma SVM karena sifat dasar teks yang cenderung mempunyai dimensi yang tinggi, dimana terdapat beberapa fitur yang tidak relevan, tetapi

akan cenderung berkolerasi satu sama lain dan umumnya akan disusun dalam kategori yang terpisah secara linear [17].

2.4 Evaluasi Analisis Sentimen

Setelah proses klasifikasi, diperlukan adanya evaluasi untuk menentukan kualitas proses yang dilakukan. Pada tahap evaluasi ini, kinerja proses klasifikasi yang telah dilakukan akan diuji dengan parameter *accuracy*, *precision*, dan *recall*. Evaluasi pada analisis sistem ini dilakukan dengan menghitung tingkat keakuratan suatu metode dalam menganalisis opini, *k-fold cross validation* dipilih untuk menguji keakuratan metode Naive Bayes dan *Support Vector Machine*. Membagi data menjadi dua bagian yaitu *data training* dan *data testing* adalah prinsip dari *cross-validation*. Dalam penelitian ini menggunakan *10-fold cross validation*. Data akan dibagi menjadi 10 bagian, bagian 1 menjadi *data training* dan bagian 2-10 menjadi *data testing*. Selanjutnya dilakukan proses silang, dimana *data training* dijadikan *data testing* dan sebaliknya, proses ini diulang sebanyak 10 kali.

Dalam proses *10-fold cross validation* terciptalah *confusion matrix* yang terdiri dari *True Positives* (TP) yaitu jumlah data kelas positif yang diklasifikasikan sebagai kelas positif dan *True Negatives* (TN) yaitu jumlah data kelas negatif yang diklasifikasikan sebagai kelas negatif. Sedangkan *False Positives* (FP) merupakan jumlah data kelas negatif yang diklasifikasikan sebagai kelas positif, dan *False Negatives* (FN) merupakan jumlah data kelas positif yang diklasifikasikan sebagai kelas negative [9]. Nilai rata-rata *accuracy*, *precision*, *recall* dapat didapatkan dari tabel *confusion matrix* seperti pada Tabel 2 berikut.

Tabel 2. Confussion Matrix

	<i>True Yes</i>	<i>True No</i>
<i>Prediction Yes</i>	<i>TP</i>	<i>FP</i>
<i>Prediction No</i>	<i>FN</i>	<i>TN</i>

Accuracy adalah jumlah dokumen yang diklasifikasikan dengan benar, ketika hasil prediksi dari model dengan sentimen yang telah ditentukan sebelumnya. Baik *True Positive* maupun *True Negative* [9]. Nilai *accuracy* dapat dihitung dengan rumus:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

Precision adalah klasifikasi *True Positive* dari semua data yang diprediksi sebagai kelas *positive*. Nilai *precision* dapat dihitung dengan rumus:

$$precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

Sedangkan *Recall* adalah berapa banyak dokumen yang relevan dalam koleksi yang dihasilkan oleh sistem. Dengan kata lain, *recall* adalah jumlah dokumen yang memiliki klasifikasi *True Positive* dari semua dokumen yang benar-benar positif (termasuk *False Negative*) [9]. Menghitung nilai *recall* dapat menggunakan rumus:

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

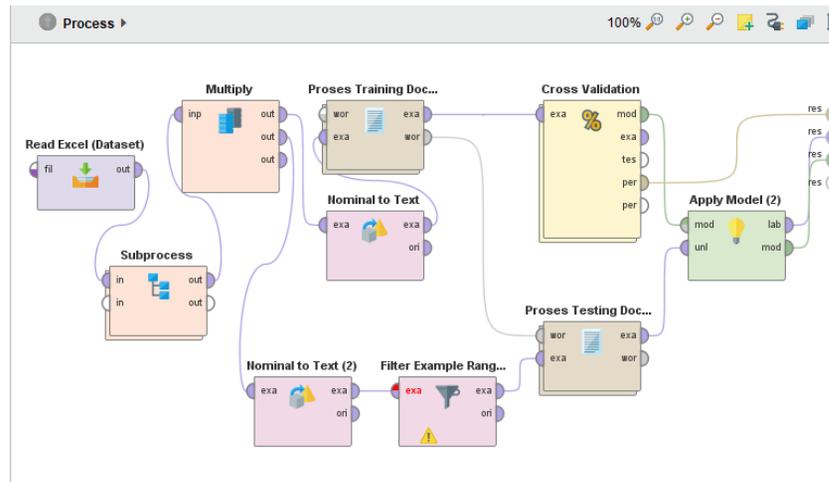
3. HASIL PENELITIAN DAN PEMBAHASAN

Bagian ini menjelaskan hasil dari penelitian yang dilakukan. Penelitian ini bertujuan untuk mengetahui sentimen analisis dari *tweets* opini terpilihnya Nadiem Makariem sebagai menteri Kementerian dan Kebudayaan baru. Kemudian membandingkan hasil klasifikasi antara metode *Naive Bayes* dan *Support Vector Machine*. Penelitian ini menggunakan *dataset tweets* yang berjumlah 2500 *tweets*. 2200 untuk *data training* dan 300 untuk *data testing*.

Tabel 3. Polaritas Sentimen Tweets

	<i>Jumlah Tweet</i>	<i>Persentase</i>
<i>Tweets Positive</i>	1743	69,72%
<i>Tweets Negative</i>	757	30,28%

Tabel 3 merupakan tabel hasil presentase polaritas dari tweets yang didapat. Analisis sentiment terhadap polaritas *tweets* yang didapatkan adalah jumlah *tweets positive* 1743 dan jumlah *tweets negative* 757. Persentase opini *positive* sebesar 69,72% dan *negative* sebesar 30,28%. Dari hasil tersebut, dapat dikatakan bahwa opini masyarakat (*netizen*) terhadap Nadiem Makariem didominasi oleh opini positif. Meskipun beberapa *netizen* kurang setuju dengan terpilihnya Nadiem Makariem sebagai Mendikbud, karena pendidikan bukanlah bidang yang digelutinya. Keseluruhan hasil penelitian dijabarkan seperti berikut:



Gambar 6 Proses Utama dalam RapidMiner

Gambar 6 menunjukkan proses utama dalam aplikasi RapidMiner. Operator "Dataset" adalah operator "Read Excel" yang digunakan untuk membaca dataset dalam file excel. Operator "Subproses" adalah operator untuk tahap pertama *pre-processing* (*cleansing data*). Operator "Proses Training" dan operator "Proses Testing" adalah operator "Document Process" yang digunakan untuk *pre-processing* dan pembobotan data. Operator "Cross Validation" digunakan untuk klasifikasi dan evaluasi analisis sentimen dengan eksperimen yang dilakukan sepuluh kali (*10-fold cross validation*).

Row No.	Sentiment	prediction(S...	confidence(...	confidence(...	text
1	Positive	Positive	0.749	0.251	kurikulum indonesia sex edu pak urgent pak
2	Positive	Positive	0.752	0.248	pak nadiem makarim
3	Negative	Positive	0.507	0.493	edisi kesel pidatonya pak nadiem dipake alas...
4	Positive	Positive	0.746	0.254	setuju pak
5	Positive	Positive	0.749	0.251	ki hajar dewantara menggambarkan belajar ta...

Gambar 7 Hasil Prediksi dari RapidMiner

Gambar 7 menunjukkan hasil prediksi sentimen dalam aplikasi RapidMiner. Membandingkan label sentimen yang sudah ditentukan dengan label hasil prediksi dari proses. Tabel 4 berikut merupakan hasil *confusion matrix* masing-masing algoritma:

Tabel 4. Confusion Matrix untuk Setiap Metode

Metode	TP	FP	TN	FN
Naive Bayes	1591	62	659	151
SVM	621	7	256	15

Berdasarkan tabel *confusion matrix* dalam Tabel 3, nilai rata-rata *accuracy*, *precision* dan *recall* dalam RapidMiner terlihat pada Tabel 4 berikut:

Tabel 5. Tabel Hasil Accuracy, Precision, Recall

Metode	Accuracy	Precisi on	Recall
Naive Bayes	91.48%	89.28 %	91.58%

		90.95	
SVM	85.47%	%	76.18%

Tabel 4 menunjukkan hasil *accuracy* dari *Naive Bayes* adalah 91.48%, sedangkan SVM 85.47%. Hasil *precision* dari *Naive Bayes* adalah 89.28% dan hasil dari SVM adalah 90.95%. Sementara hasil *recall* dari *Naive Bayes* adalah 91.58% sedangkan SVM 76.18%. Jadi dapat diperhatikan bahwa klasifikasi menggunakan *Naive Bayes* adalah *classifier* yang lebih baik dalam analisis sentimen dengan *dataset tweets* dari media sosial *Twitter* karena memberikan prediksi yang lebih akurat dan tepat.

Perbedaan hasil tersebut dapat terjadi karena kesesuaian karakteristik dari *dataset* dan parameter-parameter yang berbeda dari masing-masing algoritma. Dalam hal ini, *Naive Bayes* bekerja pada variabel independen, yaitu data teks *tweet*. Dari kesesuaian jenis data dan variabel ini, maka *Naive Bayes* dapat bekerja lebih baik dibanding *Support Vector Machine*. Hal tersebut juga dapat terjadi karena dalam tahapan klasifikasi *Naive Bayes* memproses satu persatu data atribut, sedangkan SVM melakukan klasifikasi secara general sehingga cakupan SVM lebih luas. Perbedaan hasil polaritas yang didapatkan juga dapat dikarenakan SVM memisahkan data menurut pola linear yang ditentukan berdasarkan pembobotan kata yang dilakukan, bukan berdasarkan dari probabilitas kata yang sering muncul.

4. KESIMPULAN

Dalam penelitian ini, dilakukan proses klasifikasi sentimen analisis *tweets* pada berbagai pendapat tentang menteri kemendikbud baru, Nadiem Makariem. Kesimpulan polaritas sentimen opini *netizen* terhadap Nadiem Makariem adalah dominan positif. Dengan persentase opini *positive* sebesar 69,72% dan *negative* sebesar 30,28%. Dari hasil tersebut, sentimen ini bisa digunakan untuk menilai kinerja Nadiem Makariem sejak terpilihnya hingga penelitian ini rampung. Meskipun untuk menilai dan mempertimbangkan kinerja selanjutnya.

Penelitian ini menggunakan data dengan dua pengklasifikasi yang berbeda yaitu *Naive Bayes* dan SVM. Dua pengklasifikasi memprediksi *label* dalam *dataset* berdasar pada pola data dalam *dataset*. Hasilnya menunjukkan *accuracy* dari *Naive Bayes* adalah 91.48%, sedangkan SVM 85.47%. Hasil *precision* dari *Naive Bayes* adalah 89.28% dan hasil dari SVM adalah 90.95%. Sementara hasil *recall* dari *Naive Bayes* adalah 91.58% sedangkan SVM 76.18%. Jadi dapat disimpulkan bahwa metode *Naive Bayes* lebih baik dari SVM karena memberikan prediksi yang lebih akurat dan tepat. Perbedaan tersebut terjadi dikarenakan oleh kesesuaian karakteristik dari dataset dan parameter-parameter yang berbeda dari masing-masing algoritma. Dalam penelitian ini, *Naive Bayes* bekerja pada variabel independen, yaitu data teks *tweet*. Dari kesesuaian jenis data dan variabel ini, maka *Naive Bayes* dapat bekerja lebih baik dibanding *Support Vector Machine*.

DAFTAR PUSTAKA

- [1] Indonesia Peringkat Lima Pengguna Twitter [Internet]. Available from: https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media
- [2] Daftar Pencarian Populer di Google Sepanjang Tahun 2019 [Internet]. Available from: <https://teknologi.id/teknodaftar-pencarian-populer-di-google-sepanjang-tahun-2019>
- [3] Afshoh, F. 2017. "Analisa Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter".
- [4] Rozi, I., Pramono, S., Dahlan, E. 2012. "Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi". *Jurnal EECCIS*. 6. 1, 37-43.
- [5] Firdaus, D. 2017. "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran

- Berbantuan Komputer". *Jurnal Format*. 6. 2, 91–7.
- [6] Ruhyana, N. 2019. "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes". *Jurnal IKRA-ITH Informatika*. 3. 1, 94–9.
- [7] Rusdian, D., Rosiyadi, D. 2019. "Analisa Sentimen terhadap Tokoh Publik Menggunakan Metode Naive Bayes Classifier Dan Support Vector Machine". *Journal of Computer Engineering System and Science*. 4. 2, 230–235.
- [8] Vidya, NA. 2015. "Twitter Sentiment Analysis terhadap Brand Reputation: Studi Kasus PT XL Axiata Tbk. Universitas Indonesia".
- [9] Bayhaqy, A., et al. 2018. "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naive Bayes". *International Conference on Orange Technologies, ICOT 1–6*.
- [10] Salam, A., Zeniarja, J., Khasanah, RSU. 2018. "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)". *Prosiding SINTAK*. 480–6.
- [11] Gumilang, ZAN. 2018. "Implementasi Naive Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee pada Situs Google Play".
- [12] Manning, CD., Raghavan, P., Schütze, H. (2009). *An Introduction to Information Retrieval*. Vol. 38, Cambridge University Press. Cambridge University Press: 156–164 .
- [13] Adi, S. 2018. "Perancangan Klasifikasi Tweet Berdasarkan Sentimen Dan Fitur Calon Gubernur DKI Jakarta 2017". *Journal Of Informatic Pelita Nusantara*. 3. 1, 10–16.
- [14] Tripathi, P., Vishwakarma, SK., Lala, A. 2016. "Sentiment Analysis of English Tweets Using Rapid Miner". *Proceedings - International Conference on Computational Intelligence and Communication Networks, CICN*. 668–672.
- [15] Basari, ASH., et al. 2013. "Opinion Mining of Movie Review Using Hybrid Method of Support Vector Machine and Particle Swarm Optimization". *Procedia Engineering*. 53, 453–462.
- [16] Ramayanti, D., Salamah, U. 2018. "Complaint Classification Using Support Vector Machine for Indonesian Text Dataset". *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 3. 7, 179–184.
- [17] Aggarwal, CC., Xhai, C. (2012). *Mining Text Data*. Kluwer Academic Publishers. Vol 8. 77–128.