

Perbandingan Algoritma Klasifikasi Pada Analisis Sentimen Twitter Menggunakan Framework CRISP-DM

Alfian Ardiansyah ⁽¹⁾

¹Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang

e-mail : 15650063@student.uin-malang.ac.id

* Penulis korespondensi.

Artikel ini diajukan xx xxxx 2021, direvisi xx xxxx 2021, diterima xx xxxx 2021, dan dipublikasikan xx xxxx 2021.

Abstract

Data mining can be used for data analysis of social media users about business affairs to political topics. This study uses data mining techniques to compare classification algorithms to Twitter sentiment analysis. The dataset is derived from tweets of twitter users discussing the results of the 2020 elections. By using the CRISP-DM framework which aims to define workflows for data mining projects such as transform cases, tokenization, filter tokens, filter stopwords, stems and also use a three different classification algorithm approach. Which is Decision Tree, K-NN and Naive Bayes to determine which algorithm that has the highest level of accuracy. The highest result in this study was when using the Naive Bayes classification algorithm with an accuracy of 97.92%, the highest precision is 100% and the highest recall is 100%.

Keywords: *classification, data mining, rapidminer, sentiment analysis, CRISP-DM, pre-processing, twitter*

Abstrak

Data mining dapat digunakan untuk analisis data pengguna media sosial tentang urusan bisnis sampai topik politik. Penelitian ini menggunakan teknik data mining dalam melakukan perbandingan algoritma klasifikasi terhadap analisis sentimen twitter. Kumpulan dataset tersebut berasal dari tweets pengguna twitter yang membahas hasil pilkada 2020. Dengan menggunakan *framework* CRISP-DM yang bertujuan untuk mendefinisikan alur kerja proyek data mining seperti transform cases, tokenisasi, filter tokens, filter stopwords, stem dan juga menggunakan pendekatan tiga algoritma klasifikasi yang berbeda yaitu Decision Tree, K-NN dan Naive Bayes untuk menemukan manakah algoritma yang memiliki tingkat akurasi tertinggi. Hasil tertinggi di dalam penelitian ini adalah ketika menggunakan algoritma klasifikasi Naive Bayes dengan akurasi sebesar 97.92%, presisi tertinggi sebesar 100% dan recall tertinggi juga bernilai 100%.

Kata Kunci: *klasifikasi, data mining, rapidminer, analisis sentimen, CRISP-DM, pre-processing, twitter*

1. PENDAHULUAN

Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) telah melakukan survey dari 2 sampai 25 juni 2020 yang menyatakan bahwa terdapat 196,71 juta pengguna internet di indonesia. Jumlah ini naik secara signifikan dari tahun 2018 yang sebanyak 171,17 juta jiwa. Setidaknya terdapat 10,6 juta pengguna media sosial twitter di Indonesia. (APJII, 2020)

Salah satu topik politik yang trending topik di twitter pada akhir tahun adalah pilkada 2020 secara serentak meskipun pandemi COVID-19 masih berlangsung dan memiliki angka penyebaran yang tinggi namun pilkada serentak tetap harus dilakukan. Dengan terbatasnya kegiatan di luar rumah dan himbuan dari pemerintah menyebabkan masyarakat sangat aktif dalam menyuarakan opini atau pendapat mereka di media sosial tentang pilkada dengan berbagai respon sentimen.

Data mining merupakan proses eksplorasi agar memahami dan memprediksi data menggunakan algoritma *machine learning* agar dapat memprediksi potensi kejadian yang akan terjadi melalui data yang telah diolah. Hal ini membantu dalam pengambilan langkah keputusan untuk memprediksi masa depan. Proses data mining memiliki tiga hal utama, yang pertama adalah statistik yang artinya penelitian numerik terhadap relasi data, kecerdasan buatan yang meniru kecerdasan manusia di dalam sebuah software atau mesin,

dan *machine learning* yang tujuannya memahami guna data dan membuatnya menjadi bermanfaat. Tetapi bukan berarti *data mining* menggunakan algoritma *machine learning* ini dapat digunakan untuk membuat keputusan namun lebih menjadi sebagai data pertimbangan. (Dey et al., 2017)

Analisis sentimen menggunakan twitter merupakan model yang dibuat agar mengetahui sentimen pengguna dari teks opini atau *tweets* yang ada. Didefinisikan sebagai proses teknik *text mining* yang diekstrak lalu selanjutnya dipelajari untuk menentukan polaritas pendapat dari dokumen teks yang diberikan. Dalam prinsip implementasi ini, peneliti berfokus terhadap seberapa akurat pemahaman respon dengan penilaian manusia itu sendiri dan berbagai kesimpulan lain seperti pemahaman kemiripan teks dan fokus topik. Pemahaman dan perbandingan ini terutama dihitung dengan pengukuran yang berbeda tergantung dengan presisi dan keakuratan dengan membagi menjadi tiga variabel dari reaksi negatif, netral, dan positif. (Mundalik, 2018)

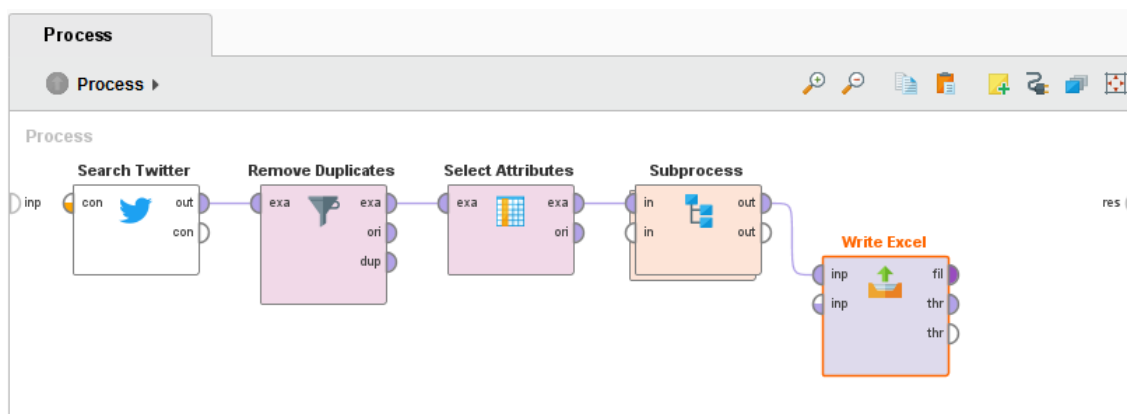
Framework yang digunakan untuk mendefinisikan alur kerja dalam penelitian ini adalah CRISP-DM. Model CRISP-DM (Cross-Industry Standard Process for Data Mining) yang menyediakan alur kerja proyek data mining. Yang memiliki enam fase yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. (Pete et al., 2000)

Penulis pada penelitian ini akan membahas bagaimana melakukan analisis sentimen pada pengguna twitter terhadap hasil pilkada 2020 yang digunakan untuk menentukan opini positif, netral, atau negatif. Dengan menggunakan tiga algoritma berbeda untuk mengekstrak opini atau perasaan pengguna twitter melalui sebuah *tweets* kemudian hasilnya akan dibandingkan untuk mengetahui algoritma klasifikasi mana yang memberikan hasil terbaik dalam hasil *ratio recall*, akurasi dan performa menggunakan data mining dengan program rapidminer.

2. METODE PENELITIAN

2.1. Crawling dan Labeling Data

Dataset didapatkan dari *crawling* dan *filter data* dari twitter menggunakan twitter API. Ada dua macam dataset yaitu data twitter terbaru dan data twitter terbaru serta populer, keduanya dilakukan proses pengolahan lebih lanjut dengan dilabeli polaritas atau sentimen secara manual.



Gambar 1. Crawling dan Filtering di Rapidminer.

Data pertama yaitu data twitter terbaru yang dilabeli secara manual, memiliki 194 tabel dan 3 kolom yang menunjukkan kolom Text dan Polarity Manual. Dengan kolom Polarity Manual sebagai label di dalam program rapidminer.

Tabel 1. Contoh Dataset Crawling dan Filtering Twitter Terbaru

Text	Polarity Manual
RT @Nusantaraku01: Jawa Tengah Kandang Banteng Jawa Tengah Patuh Ibu Mega ???• Banteng membuktikan kegarangannya di Jateng.Dmna dr 21 ge...	Negative
@mochamadariip Bagusnya masyarakat sudah bisa melihat kinerja pimpinan daerah, ke depannya kalau ada pilkada semoga bukan krn partai politik tp karena kompetensi dan kinerjanya yg jadi tolok ukur utk dipilih.	Positive
RT @AriefRohman_838: Bismillah mengikuti Pembekalan Kepemimpinan Pemerintah Dalam Negeri bagi Bupati hasil Pilkada 2020 yang dibuka Bapak M...	Neutral
Cuman pilkada yg ga bawa klaster virus pokoknya. https://t.co/oygU42BJ6V	Negative
RT @RomitsuT: 2022 Pilkada DKI..Anies Baswedan akan menang mudah lawan siapa saja. Sekarang tinggal siapa yg jadi wakil Anies yg pas karena...	Negative
@drpriono1 @ronavioleta @jokowi Heran yah Covid hanya berkluster Mudik! Apa kabar Pilkada? Pesta kawinan Kantoran? Industri? TKA? Yg lebih anehnya, Covid aktif di atas jam 9 malam!Wauuww...luar biasa Tetap pakai Masker yah saudara2	Negative
Siap Maju Pilkada 2024 https://t.co/701cleUVUq	Positive
@Klabangijo @temmymrdla @drpriono1 @jokowi Lu yg ga paham cok.. pilkada kmrn juruwabah pernah ga bahas klaster pilkada. Bego jgn kebangetan cuk.	Negative
RT @triwul82: Sekjend PDIP Hasto Kristiyanto menyampaikan apresiasi kemenangan Pilkada di Jateng dan Yogyakarta. "Prinsip ojo pedhot oyot,...	Positive

Data kedua sama berasal dari data twitter namun dengan kategori terbaru dan terpopuler dan juga telah dilakukan labeling secara manual oleh peneliti. Memiliki 255 tabel dan 3 kolom yang menunjukkan kolom text dan Polarity Manual. Polarity Manual juga sebagai label di dalam rapidminer.

Tabel 2. Contoh Dataset Crawling dan Filtering Twitter Terbaru Serta Terpopuler

Text	Polarity Manual
Ketum AHY berjanji akan menguatkan koordinasi di akar rumput & kalangan menengah hadapi Pilkada Serentak 2024 mendatang. Terutama menjelang pemilihan presiden. Hal itu disampaikannya saat mengunjungi Kantor Pikiran Rakyat, Jumat (4/6). #AHYNgariungJabar https://t.co/1U2Z7EBNM6	Positive
"Dengan pemilu serentak tahun 2024 otomatis mesti dijauhkan jarak waktunya. Ini kompromi agar KPU dapat bekerja optimal dengan jarak, yang di awal 2024 untuk Pileg dan Pilpres dan 27 Nov 2024 untuk Pilkada," ujar Ketua DPP PKS Mardani Ali Sera. https://t.co/xkFN6HuA8S	Positive
Polisi Usut Dugaan Korupsi Dana Covid di Papua untuk Pilkada https://t.co/VcjPxp4r55	Negative
@RomitsuT: 2022 Pilkada DKI..Anies Baswedan akan menang mudah lawan siapa saja. Sekarang tinggal siapa yg jadi wakil Anies yg pas karena...	Negative
@Humic19: @ManOfMeJOE1 Lha.. trus investasinya kemana ??? Ke Pilkada ??? #AuditDanaHaji Makin seru saja ...	Negative
@korantempo: Pemilihan presiden digelar berbarengan dengan pemilihan legislatif pada 28 Februari 2024, sementara pilkada serentak dilaks...	Negative
Republika. Senin, 07 Juni 2021. "Bagaimana mencari cara supaya proses pemilu kita ini lebih sederhana lebih gampang tapi dengan catatan mereka harus mau mengubah," kata Hadar @HadarNG saat dihubungi, Ahad (6/6). https://t.co/6x9X9Z0PTI	Neutral
@ManOfMeJOE1 Lha.. trus investasinya kemana ??? Ke Pilkada ??? #AuditDanaHaji #AuditDanaHaji Makin seru saja ...	Negative
PPP Setuju Pemilu 2024 Dimajukan: Agar Tak Berhimpitan dengan Pilkada https://t.co/EmJKbORm1U	Positive

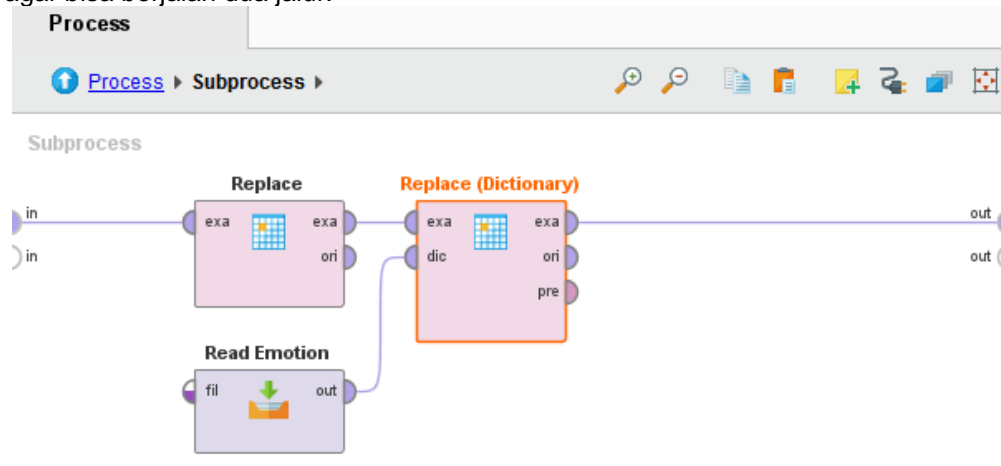
Setelah file data excel dimasukkan ke dalam proses selanjutnya yaitu peneliti melakukan labelling manual untuk mendeteksi emoticon yang gunanya mengenali emosi sedih dan senang. Memiliki 96 tabel dan 2 kolom.

Tabel 3. Contoh Dataset Untuk mendeteksi Emoticon.

sebelum	Sesudah
d:	sedih
o.o	sedih
o_o	sedih
x(sedih
x-(sedih
xp	sedih
%-)	senang
(:	senang
(-:	senang
(^^)	senang
(^.^)	senang
(^^)	senang

2.2. Pre-Processing

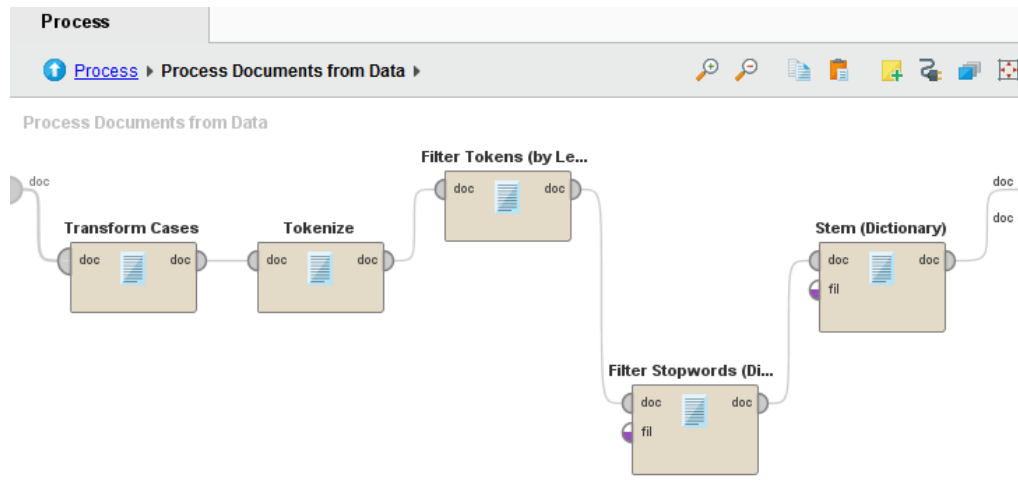
Setelah melabeli data, langkah selanjutnya adalah *pre-processing*. Tahap ini merupakan tahap dimana data disiapkan untuk mejadi data yang siap untuk dianalisa. Ada beberapa tahap dari *pre-processing* dalam penelitian ini termasuk *cleansing*, konversi *emoticons*, *transform cases*, *tokenizing*, *filter token (by length)*, dan *filter stopwords*. Gambar 2 menunjukkan konten dari operator “Subprocess”. Dalam hal ini digunakan untuk *cleansing*, *replace (dictionary)* untuk membaca emoticon. Lalu setelah selesai datanya akan di “Multiply” agar bisa berjalan dua jalur.



Gambar 2. Isi Konten dari Operator Subprocess.

Data yang akan peneliti gunakan masih butuh untuk disaring dan disiapkan untuk proses selanjutnya. Tentunya *data crawling* dan *filtering* dari twitter sendiri masih belum rapi dan bersih sehingga penyaringan ulang masih sangat dibutuhkan. Namun sebelum itu masih dibutuhkan mengubah data nominal ke bentuk teks, lalu peneliti akan menggunakan operator yang bernama “Process Document From Data”. Di dalam operator tersebut akan digunakan berbagai macam operasi untuk menyaring data. Proses ini dilakukan

dua kali di aliran yang berbeda karena operator *multiply* sebelumnya. Gambar 3 merupakan operator-operator yang ada di dalam proses dokumen dari sebuah data.



Gambar 3. Isi Operator Proses Dokumen.

1. Transform cases

Operator ini digunakan untuk mengubah semua karakter menjadi huruf kecil.

2. Tokenize

Operator ini akan memisahkan kalimat menjadi tiap kata atau yang disebut tokenisasi.

3. Filter Tokens (by length)

Operator ini memfilter token tergantung dengan panjangnya. Penulis akan memberikan range atau jarak dari 4 sampai 25. Jadi, operator akan filter setiap kata kecuali kata yang ada di range tersebut.

4. Filter Stopwords (Dictionary)

Operator ini akan menghilangkan *stopwords*. Namun untuk mengenali dan bekerja pada bahasa indonesia butuh pengenalan secara manual. Oleh karena itu peneliti memakai 'Filter Stopwords (Dictionary)' dengan menggunakan file "stopword_indo.txt" yang katanya sudah ditentukan sebelumnya.

5. Stem (Dictionary)

Guna operator adalah mengubah semua kata mejadi bentuk dasarnya. Dengan menggunakan file "stem.txt". Contohnya seperti kata lah,kah,ku,mu,isme,di, dan lain-lain.

6. Pembobotan Kata

Pembobotan kata atau *Weighting Word* merupakan mekanisme untuk memberikan skor pada frekuensi kemunculan data dalam dokumen. Salah satu pembobotan yang populer adalah TF-IDF (Term Frequency-Inverse Document Frequency) yang merupakan kombinasi dari dua konsep. *Term frequency* merupakan konsep pembobotan dengan mencari seberapa sering istilah dalam suatu dokumen. Sedangkan *Document Frequency* adalah jumlah dokumen di mana istilah tersebut muncul. Semakin kecil frekuensi kejadiannya, semakin kecil bobot nilai. Dengan rumus pembobotan ini adalah :

$$TF-IDF = TF(t,d) \times IDF(t)$$

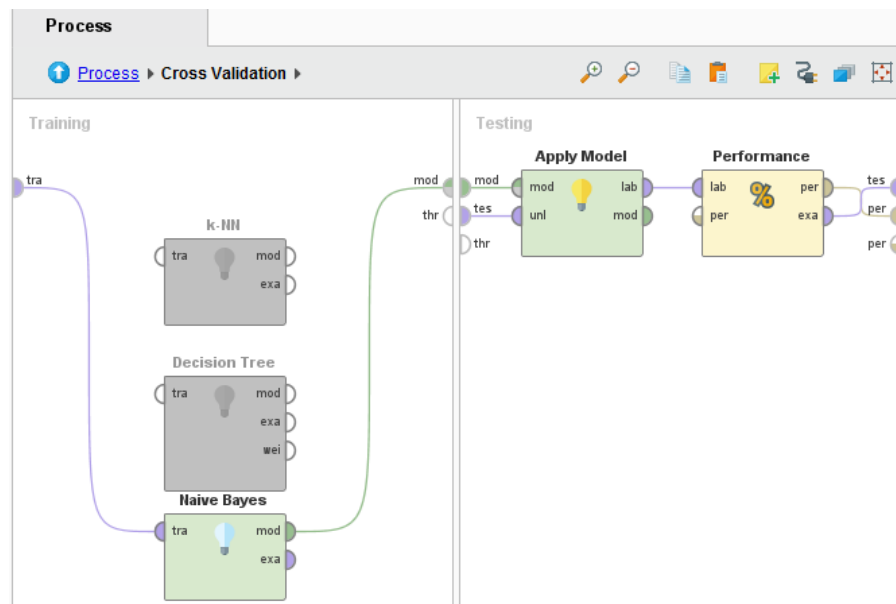
Dengan term frequeny (tf) berapa banyak frekuensi terjadinya istilah (t) dalam sebuah dokumen (d). (Bayhaqy et al., 2018)

2.3. Klasifikasi Sentimen Analisis

Setelah *pre-processing* data langkah selanjutnya adalah klasifikasi dari sentimen analisis yang akan menggunakan 3 algoritma *data mining* yang berbeda dalam proses untuk menentukan model mana yang

memiliki tingkat keakuratan paling tinggi dengan dataset yang telah ditentukan polaritasnya secara manual. Setelah mencoba berbagai macam algoritma, model algoritma yang paling akurat akan digunakan untuk evaluasi dan penerapan.

Disini peneliti menggunakan *cross validation* yang cocok untuk data yang sudah diolah dan tidak terlalu besar. Juga dikarenakan sebelumnya sudah banyak dilakukan *pre-processing* seperti labeling manual, *replace* menggunakan emosi, proses dokumen dari data, dan *dictionary* bahasa indonesia yang ditentukan secara manual. Digunakan k-fold sebanyak 10 kali, jumlah ini direkomendasikan sebagai jumlah standar dalam menentukan model mana yang paling baik karena kecenderungan memberikan estimasi akurasi yang kurang bias akhirnya tidak berat sebelah terhadap salah satu model tertentu. (Mierswa, 2017)



Gambar 4. Isi Operator Cross Validation.

2.4. Evaluasi Sentimen Analisis

Sebelum ke proses selanjutnya perlunya melihat ulang semua proses yang telah dilakukan. Ada beberapa langkah untuk memeriksa apakah telah mencapai tujuan yang diinginkan dengan meninjau model. Juga, akan memeriksa apakah ada masalah potensi kesalahan dan apakah hal tersebut dipertimbangkan diperbolehkan atau tidak. Langkah-langkah evaluasi tersebut sebagai berikut :

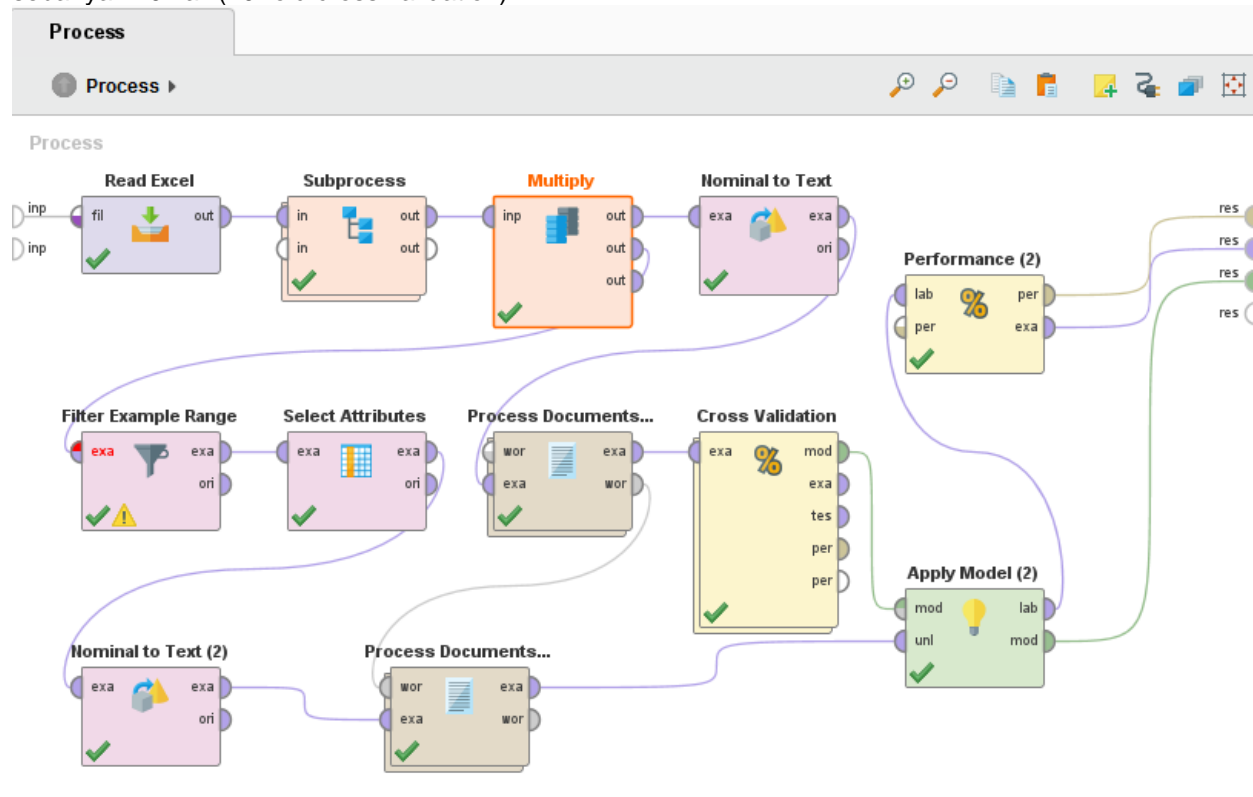
- Evaluasi Operator
- Review proses
- Menentukan langkah selanjutnya

Perlunya perbandingan dengan dataset lainnya juga membantu meningkatkan performa dan kebenaran data yang akan dilakukan, sebelumnya telah dinyatakan bahwa ada dataset pilkada lain yaitu dataset pilkada yang berasal dari tipe hasil *tweets* terbaru melainkan juga terpopuler yang telah dilabeli secara manual polaritasnya.

3. HASIL DAN PEMBAHASAN

Peneliti akan menjelaskan hasil dari eksperimen dan analisa performa terhadap proses yang telah dilakukan. Dimulai dengan operator "Read Excel" yang digunakan untuk membaca berkas excel. Operator "Subprocess" dan operator "Process Documents" digunakan untuk *pre-processing*. Operator "Cross

Validation” digunakan untuk klasifikasi dan evaluasi analisis sentimen dengan eksperimen yang dilakukan sebanyak 10 kali (10-fold cross validation).



Gambar 5. Proses Utama di Rapidminer.

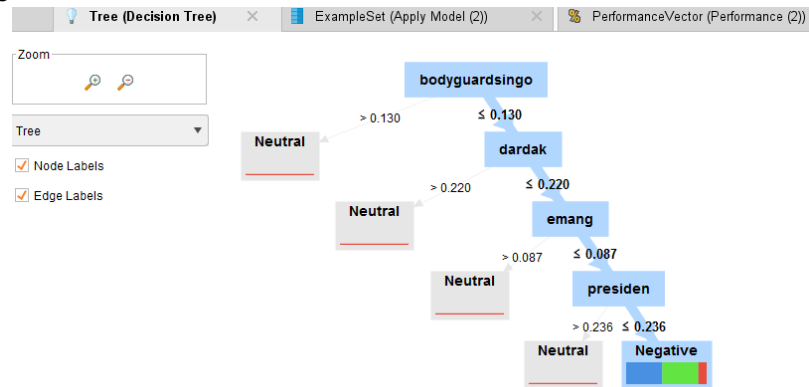
Bagian ini adalah proses akhir dalam *framework* CRISP-DM. Peneliti tidak hanya melakukan berbagai macam model algoritma terhadap satu dataset. Namun di bagian ini akan dilakukan perbandingan dengan melakukan percobaan dengan dataset pilkada kedua yang tipe hasil *tweets* terbaru dan juga terpopuler. Jadi peneliti juga akan melakukan proses model algoritma klasifikasi terhadap *dataset* tersebut. Dengan percobaan pertama menggunakan dataset pilkada yang hanya berasal terhadap data *tweets* terbaru.

1. k-NN

KNNClassification (k-NN)				
ExampleSet (Apply Model (2))				
PerformanceVector (Performance (2))				
Criterion	Table View Plot View			
accuracy	accuracy: 80.21%			
	true Positive	true Neutral	true Negative	class precision
pred. Positive	75	8	19	73.53%
pred. Neutral	0	17	1	94.44%
pred. Negative	8	2	62	86.11%
class recall	90.36%	62.96%	75.61%	

Gambar 6. Nilai Akurasi, Presisi, Recall Average Pada Algoritma k-NN Dataset Pertama.

2. Decision Tree



Gambar 7. Gambaran Decision Tree Pada Dataset Pertama.

Tree (Decision Tree) ExampleSet (Apply Model (2)) PerformanceVector (Performance (2))

Criterion: accuracy

Table View Plot View

accuracy: 46.88%

	true Positive	true Neutral	true Negative	class precision
pred. Positive	0	0	0	0.00%
pred. Neutral	0	8	0	100.00%
pred. Negative	83	19	82	44.57%
class recall	0.00%	29.63%	100.00%	

Gambar 8. Nilai Akurasi, Presisi, Recall Average Pada Decision Tree Dataset Pertama.

3. Naive Bayes

SimpleDistribution (Naive Bayes) ExampleSet (Apply Model (2)) PerformanceVector (Performance (2))

Criterion: accuracy

Table View Plot View

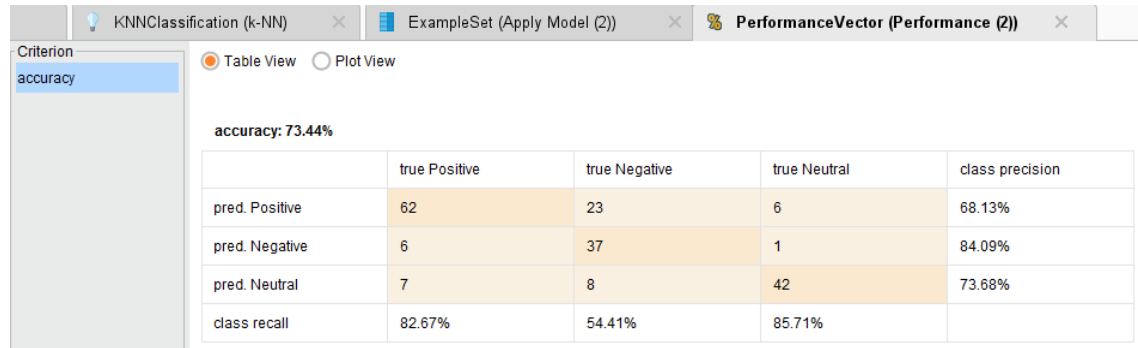
accuracy: 97.92%

	true Positive	true Neutral	true Negative	class precision
pred. Positive	82	0	1	98.80%
pred. Neutral	1	27	2	90.00%
pred. Negative	0	0	79	100.00%
class recall	98.80%	100.00%	96.34%	

Gambar 9. Nilai Akurasi, Presisi, Recall Average Pada Naive Bayes Dataset Pertama

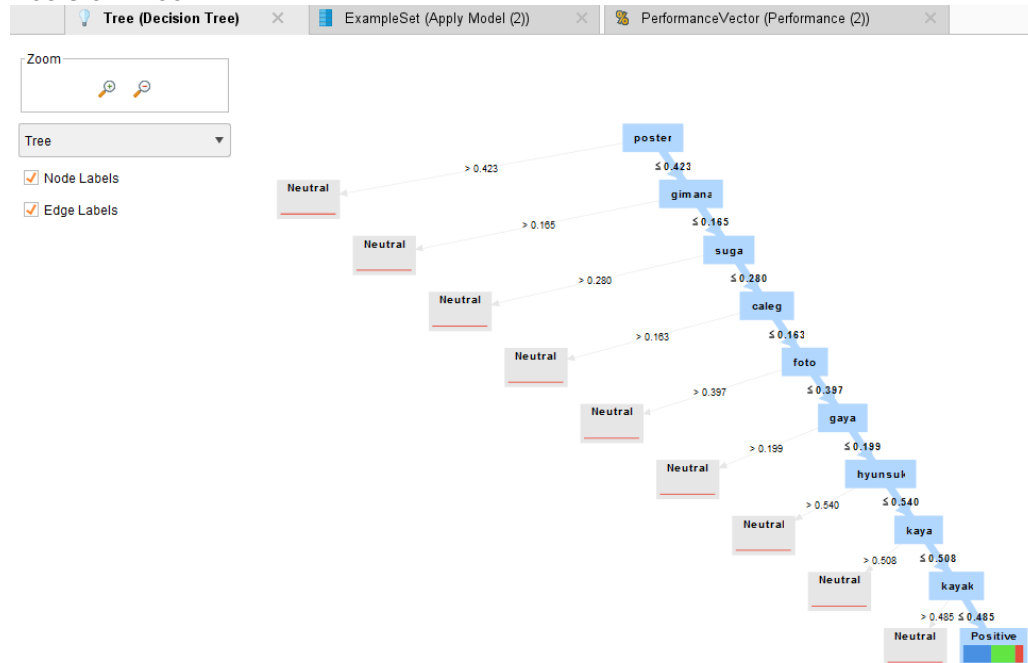
Setelah melakukan percobaan dengan dataset pilkada pertama yang hanya berdasarkan *tweets* terbaru, selanjutnya akan dilakukan percobaan terhadap dataset pilkada kedua yang berdasarkan tipe hasil *tweets* terbaru dan terpopuler. Disini akan juga digunakan *range* pengambilan data yang sama dengan dataset pertama yaitu 193 tabel meskipun isi dataset kedua sebenarnya lebih banyak daripada dataset pertama dengan jumlah 255 tabel. Tujuannya agar performa yang dihasilkan akan lebih akurat.

1. k-NN

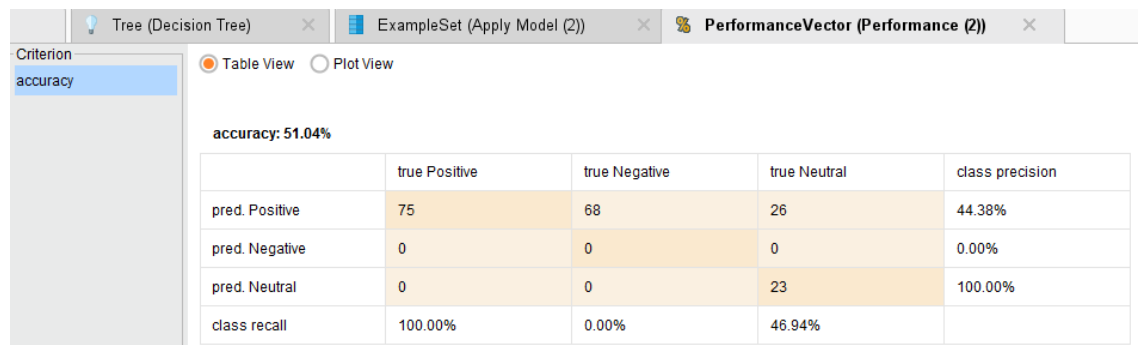


Gambar 10. Nilai Akurasi, Presisi, Recall Average Pada Algoritma k-NN Dataset Kedua

2. Decision Tree

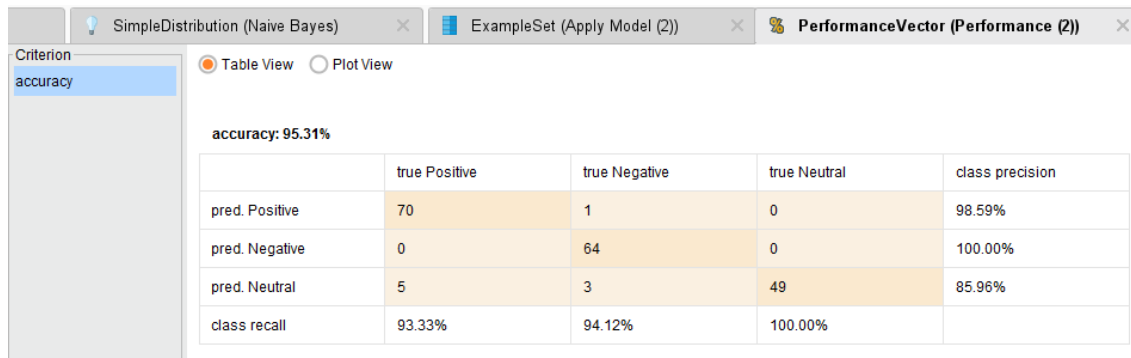


Gambar 11. Gambaran Decision Tree Pada Dataset Kedua



Gambar 12. Nilai Akurasi, Presisi, Recall Average Pada Decision Tree Dataset Kedua.

3. Naive Bayes



SimpleDistribution (Naive Bayes) ExampleSet (Apply Model (2)) PerformanceVector (Performance (2))

Criterion: accuracy

Table View Plot View

accuracy: 95.31%

	true Positive	true Negative	true Neutral	class precision
pred. Positive	70	1	0	98.59%
pred. Negative	0	64	0	100.00%
pred. Neutral	5	3	49	85.96%
class recall	93.33%	94.12%	100.00%	

Gambar 13. Nilai Akurasi, Presisi, Recall Average Pada Naive Bayes Dataset Kedua

4. KESIMPULAN

Dapat disimpulkan bahwa sentimen analisis pada twitter ini dapat diselesaikan dengan *framework* CRISP-DM dalam melakukan penelitian yang memiliki topik *data mining*. Walaupun kedua dataset memiliki tipe hasil yang berbeda, yang pertama hanya berdasarkan dataset pilkada dengan *tweets* terbaru dan dataset pilkada berdasarkan dengan *tweets* terbaru dan terpopuler. Keduanya juga telah dilabeli secara manual.

Pendefinisian *dataset* kedua terlalu banyak memiliki nilai polaritas netral, dikarenakan proses *crawling* dan *filtering* dari twitter tersebut juga mengenali hal-hal yang populer selama itu berkaitan dengan pilkada , maka dari itu dataset tersebut terlalu banyak mengandung data *tweets* yang menyebutkan pilkada namun tidak memiliki kaitan dengan isu politik.

Penggunaan algoritma klasifikasi k-NN, decision tree dan naive bayes telah dilakukan terhadap tiap dataset dengan hasil akurasi keseluruhan dataset pertama secara berurutan yaitu 80.21%, 46.88% dan 97.92%. Sedangkan akurasi keseluruhan dataset kedua secara berurutan yaitu 73.44%, 51.04% dan 95.31%.

Naive bayes merupakan keluarga *probabilistic classifier* yang didasarkan penerapan teorema bayes dengan asumsi independensi yang kuat (naive) diantara fitur-fitur model lainnya. Hal ini menjadikan naive bayes cocok sebagai algoritma pendeteksi spam, sentimen analisis dan sistem rekomendasi. Namun pernyataan ini masih perlu di buktikan apakah naive bayes masih cocok dengan dataset lainnya. Maka dari itu dibutuhkannya perbandingan dengan menggunakan dataset telah disediakan sebelumnya yaitu dengan menggunakan dataset kedua dengan polaritas atau sentimen telah dilabeli. Telah dibuktikan di penelitian ini bahwa akurasi naive bayes ini memiliki nilai tertinggi dari model algoritma klasifikasi lainnya di kedua dataset yang disediakan.

Secara keseluruhan, prediksi akurat sesuai dengan dataset yang dipakai dengan naive bayes memiliki tingkat keakuratan yang paling tinggi. Namun bukan berarti keakuratan tinggi sesuai dengan tujuan pengolahan data seperti kasus dalam penelitian ini. Tujuan *data mining* adalah mencari nilai data yang paling benar bukan keakuratannya saja. Hal ini dapat dibuktikan dengan nilai *class precision* yang tinggi algoritma naive bayes pada dataset pertama yang memiliki nilai prediksi positif 98.80%, 90.00% dan 100%. Semakin tinggi nilainya maka kebenaran data semakin dapat terbukti. Penelitian ini juga membuktikan bahwa dataset kedua yang lebih spesifik tidak selalu memiliki tingkat akurasi, performa, dan presisi yang lebih tinggi.

DAFTAR PUSTAKA

- APJII. (2020). Laporan Survei Internet APJII 2019 – 2020. *Asosiasi Penyelenggara Jasa Internet Indonesia*, 2020, 1–146. <https://apjii.or.id/survei>
- Dey, N., Ashour, A. S., & Nguyen, G. N. (2017). Deep learning for multimedia content analysis. *Mining Multimedia Documents*, 1(4), 193–203. <https://doi.org/10.1201/b21638>
- Mundalik, A. (2018). *Aspect Based Sentiment Analysis Using Data Mining Techniques Within Irish Airline Industry* Aishwarya Mundalik Supervisor :
- Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.
- Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies, ICOT 2018*, 1–6. <https://doi.org/10.1109/ICOT.2018.8705796>
- Mierswa, I. (2017). *How to Correctly Validate Machine Learning Models*. 26. <https://rapidminer.com/resource/correct-model-validation/>

