Jurnal Teknik Komputer AMIK BSI
Volume VI No.1 Januari 2020
P-ISSN 2442-2436, E-ISSN: 2550-0120
Akreditasi Ristekdikti, No: 21/E/KPT/2018
DOI: 10.31294/jtk.v4i2

Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023

Deni Gunawan[1], Dwiza Riana[2], Dian Ardiansyah[3], Fajar Akbar[4], Salman Alfarizi[5]

[1,3]Program Studi Sistem Informasi Universitas Bina Sarana Informatika
[1]e-mail : deni.dee@bsi.ac.id
[3]e-mail : dian.did@bsi.ac.id
[5]e-mail : salman.slz@bsi.ac.id

[2]Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri
[2]dwiza@nusamandiri.ac.id

[4]Program Studi Teknik Informatika STMIK Nusa Mandiri
[4]fajar.fkb@nusamandiri.ac.id

Diterima	Direvisi	Disetujui
12-11-2019	09-01-2020	02-02-2020

Abstrak – Kontestasi politik dalam penentuan menjadi pemimpin tingkat provinsi dalam hal ini gubernur jawa barat 2018-2023. Masyarakat yang memberikan opininya berupa tweet pada media sosial twitter menentukan bentuk dukungan atau tidaknya, sehingga perlu adanya analisis sentimen terhadap calon Gubernur agar mengetahui tingkat kepercayaan masyarakat serta terbentuk citra kepada calon Gubernur Jawa Barat 2018-2023. Akan tetapi membaca keseluruhan tweet yang tersebar dalam twitter yang berkaitan dengan masing-masing calon gubernur akan memakan waktu dan membingungkan dalam pengambilan keputusan. Klasifikasi sentimen akan mengurai masalah mengenai opini, pendapat, emosi dan prilaku dengan studi komputasi. Metode klasifikasi yang akan dibahas dalam penelitian yaitu dengan algoritma Naïve Bayes serta Support Vector Machine. Penentuan fitur menentuka hasil akurasi, dalam penentuan fitur seleksi digunakan Genetic Algorithm agar dapat meningkatan akurasi pengklasifikasian pada Support Vector Machine dan Naive Bayes. Perolehan penelitian ini yaitu klasifikasi teks dalam pola negatif atau positif dari tweet calon gubernur jawa barat 2018-2023. Pada dataset tidak seimbang Support Vector Machine menghasilkan rata-rata akurasi 92.61% dengan AUC 0,950, Naive Bayes menghasilkan rata-rata akurasi 93,29% dengan AUC 0,525, Support Vector Machine berbasis Genetic Algorithm menghasilkan rata-rata akurasi 93,03% dengan AUC 0,869, Naive Bayes berbasis Genetic Algorithm menghasilkan rata-rata akurasi 92,85% dengan AUC 0,543. Hasil ini menunjukan bahwa Support Vector Machine dapat digunakan untuk membangun deteksi tweet klasifikasi positif dan negatif dengan tingkat akurasi yang tinggi. Kebaruan dari penelitian ini adalah bahwa Support Vector Machine dapat digunakan untuk mendeteksi tweet pada dataset twitter berbahasa indonesia penulis.

Kata Kunci : Sentimen Analisis, *Support Vector Machine, Naive Bayes, Genetic Algorithm.*

Abstract *- Political dispute was determines provincial level in West Java Governor 2018- 2023. A society gave their opinions in social media on twitter to support or not, then it needed sentiment analysis against Governor prospective in order to find out trust level of community and formed the image to West Java Governor's prospective e 2018-2023. Reading the whole tweet in the twitter related to each Governor's prospective wasted time and confused on decision making. Sentiment classification decreased the problem using computer literate about opinions, behaviours and emotions of a person against the entity. This research will be discussed about the classification techniques with the Support Vector Machine method and Naive Bayes. The selection features will be affected the accuracy of classification, it was using Genetic Algorithm in order to improve the accuracy of classification on Support Vector Machine and Naive Bayes. This research resulted in texts classification of positive or negative from West Java Governor's prospective tweet s 2018-2023. On the unbalanced datasets Support Vector Machine produces an average of 92.61% accuracy with AUC 0,950, Naive Bayes generates an average of 93.29% accuracy with AUC 0,525, Support Vector Machine-based Genetic Algorithm produces an average accuracy of 93.03% with AUC 0,869, Naive Bayes-based Genetic Algorithm produces an average accuracy of 92.85% with AUC 0,543. These results suggested that Support Vector Machine can be used to build the detection positive and negative classification in tweet with a high accuracy. Support Vector Machine can be used to detect a tweet on twitter dataset speak Indonesian writer as the lastest of this research.*

Keywords: Analysis Sentiment, Support Vector Machine, Naive Bayes, Genetic Algorithm

PENDAHULUAN

Indonesia sebagai pengguna internet yang dengan pesat perkembangannya, menurut Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) survei data statistik pengguna internet indonesia tahun 2016, data yang dilansir kompas 2016 tercatat pengguna internet melalui ponsel pintar berbasis android di Indonesia kini sudah mencapai 132,7 juta pengguna. Mayoritas konsumsi internet di Indonesia adalah untuk menggunakan jejaring media social, dengan pengguna media sosial pada di angka 40%, adapaun pengguna aktifnya sebesar 34%. Dalam mengakses media sosial tersebut berada di angka 39% dengan perangkat mobile.

Kampanye politik akan sering kali terjadi dalam proses perebutan panggung politik sehingga menjadi sesuatu hal yang lumrah dalam suatu negara. Tidak dapat dielakan bahwa kampanye yang dilakaukan para actor politik merupakan pencarian segmentasi pemilih agar mendapat dukungan penuh (Alfiah, Susanti, Kristinna, Ardiansyah, & Pradipta, 2015).

Beberapa tahun terakhir telah melihat gelombang kepentingan dalam metode komputasi yang memiliki pengaruh terhadap opinion mining, untuk deteksi subyektivitas, dan analisis sentiment (Balahur, Mihalcea, & Montoyo, 2014). Twitter adalah layanan microblogging real time populer yang memungkinkan pengguna untuk berbagi informasi singkat dikenal sebagai tweets yang dibatasi 140 karakter. Twitter adalah platform yang ideal untuk ekstraksi pendapat masyarakat umum pada isu-isu spesifik, mengungkapkan sesuatu keluhan dan meluapkan sentimen yang negatif atau positif. (Sarlan, Nadam, & Basri, 2015)(Wahyudi & Putri, 2016). Bahkan terdapat perusahaan mendalami reaksi pengguna dalam sebuah produk manufactur dengan platform twitter (Wahyudi & Putri, 2016). Twitter kini lebih banyak karakter yang disediakan dalam setiap cuitan, yang sebelumnya 140 menjadi 280 karakter (Bohang, 2017).

Jutaan bahkan ratusan juta pengguna dapat mencurahkan pemikiran serta opini mereka mengenai aspek-aspek kehidupan pada platform micro-blogging pribadinya. Oleh karenanya cuitan pada twitter adalah basis informasi yang dapat mempertimbangkan sebuah keputusan serta menganalisis sentimen. Kemunculan media sosial memberikan keleluasaan dalam mengungkapkan pemikiran dan mengekspresikan setiap penguna dalam topik yang bermacam-macam pada wadah yang sama (Balahur et al., 2014). Dengan pengguna telah hampir 600 juta serta pesan per hari mencapai 250 juta, menjadikan twitter sumber informasi serta peluang besar bagi organisasi dalam mengontrol merek mereka dari pesaing dan postingan cuitan di pasar public (Balahur et al., 2014).

Beberapa penelitian sebelumnya mengenai ulasan film menggunakan klasifikasi Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO) atau sentimen opini analisis (Basari, Hussin, Ananta, & Zeniarja, 2013). Machine Learning yang memperkenalkan klasifikasi teks seperti Naive Bayes, K-NN, SVM dan Rocchio Classification (Ramesh & Sathiaseelan, 2015).

Pada penelitian ini akan menjelaskan lebih detail untuk mengoptimasi Algoritma klasifikasi. Menurut Zukhri Optimasi adalah proses menyelesaikan masalah tertentu yang berada pada kondisi yang paling menguntungkan dari sudut pandang. Yang memecahkan masalah terkait erat dengan data yang dapat diekspresikan edinone atau beberapa variabel (Zukhri, 2014). Algoritma Genetika adalah metode heuristik yang dikembangkan berdasarkan prinsip-prinsip genetika dan proses seleksi alam teori evolusi Darwin. Metode optimasi dikembangkan oleh John Holl dan sekitar tahun 1960 dan dipopulerkan oleh salah seorang muridnya, David Gold bergin tahun 1980-an. Penyempurnaan proses pencarian dalam algoritma terjadi seperti pemilihan individu untuk bertahan dalam proses evolusi (Zukhri, 2014).

Dari uraian tersebut diatas, maka dalam penelitian ini akan digunakan metode Support Vector Machine dan Naïve bayes yang dipadu dengan Algoritma Genetika yang akan mengoptimasi parameter Support Vector Machine.

METODOLOGI PENELITIAN

A. Perencanaan Penelitian

Pada dasarnya, penelitian merupakan suatu investigasi yang terorganisasi, yang dilakukan untuk menyajikan suatu informasi dan memecahkan masalah. Metode penelitian yang digunakan penulis menggunakan metode penelitian eksperimen. Adapun metode penelitian yang penulis gunakan melalui beberapa tahapan sebagai berikut:

a. Pengumpulan Data
   Data yang digunakan untuk melakukan eksperimen dikumpulkan melalui *tweet* dari twitter diambil menggunakan aplikasi rapidminer.
b. Pengolahan Data awal
   Menentukan metode yang akan dipakai pada proses pengujian data. Berdasarkan metode terdahulu dengan ketentuan metode yang terbaik dari pengklasifikasian teks. Penelitian ini menggunakan metode Algoritma Support Vector Machine dan Naïve Bayes.
c. Metode yang Diusulkan
   Metode yang diajukan dalam hal ini penulis melakukan komparasi atau perbandingan metode SVM dan NB dengan melakukan peningkatan optimasi yang digunakan, yaitu dengan *Genetic Algorithm* (GA).
d. Eksperimen dan Pengujian Metode

Eksperimen yang dilakukan peneliti, menggunakan framework RapidMiner 8.2. untuk mengolah data *tweet* sebagai alat bantu pengukuran sehingga menghasilkan nilai akurasi yang akurat.

e. Evaluasi Performa dan Validasi Hasil Evaluasi
Evaluasi menggunakan confusion matrix untuk mengetahui akurasi, presisi dan *recall*. Menurut Han confusion matrix adalah alat yang sangat berguna untuk menganalisis seberapa baik classifier dapat mengidentifikasi tupel dari kelas yang berbed (Han, Kamber, & Pei, 2012). Dalam beberapa confusion matrix dikenal istilah True positive yang merujuk pada tuple positif yang diberi label dengan benar oleh classifier, sedangkan True negative adalah tuple negatif yang diberi label dengan benar oleh classifier. Ada juga false positive yang merupakan tuple negatif yang salah diberi label oleh classifier, dan false negative adalah tuple positif yang tidak diberi label dengan benar oleh classifier. Kurva ROC (*Receiver Operating Characteristic*) membagi hasil positif pada sumbu y dan hasil negatif pada sumbu x dalam bentuk AUC (*Area Under the Curve*). Jadi semakin besar area di bawah kurva, semakin baik hasil prediksi. yang bisa mengubah orientasi pendapat juga tidak sebanding dengan yang buruk (Witten, Frank, Hall, & Pal, 2016).

B. Pengumpulan Data Awal

Peneliti menggunakan data *tweet* pemilihan gubernur jawa barat 2018. Data yang dikumpulkan dari www.twitter.com dibantu dengan framework rapid miner, untuk jenis data tweet pasangan calon gubernur jawa barat periode 2018-2023 yang terdapat 4 pasang calon gubernur jawa barat diambil data sebanyak 9637 data, untuk data pasangan calon gubernur Ridwan Kamil-Uu Ruhzanul Ulum sebanyak 3647 data, Tubagus Hasanuddin-Anton Charliyan 2037 data, Sudrajat-Ahmad Syaikhu 2358 data dan Deddy Mizwar-Dedi Mulyadi 1595 data. Kemudian dikelompokkan ke dalam *tweet* positif dan *tweet* negatif. Data *tweet* yang penulis mengunduh merupakan data *tweet* pertanggal 20 Mei, 04 Juni, 22 Juni dan 27 Juni 2018.

C. Pengolahan Data Awal

Untuk mehindari pengolahan data yang tidak diperlukan maka dilakukan remove duplicate, sehingga penulis hanya menggunakan 2010 *tweet* positif dan 633 *tweet* negatif sebagai data training. Kemudian dataset tersebut diolah melalui tahapan preprocessing 4 (empat) proses, diantaranya:

a. Tokenization
Proses memotong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf yang diterima, sedangkan karakter khusus atau tanda baca akan dihilangkan. Jadi hasil dari proses *Tokenization* adalah kata-kata yang merupakan penyusun kalimat atau string yang dimasukan tanpa ada tanda baca (Crc, Hofmann, & Chisholm, 2016).

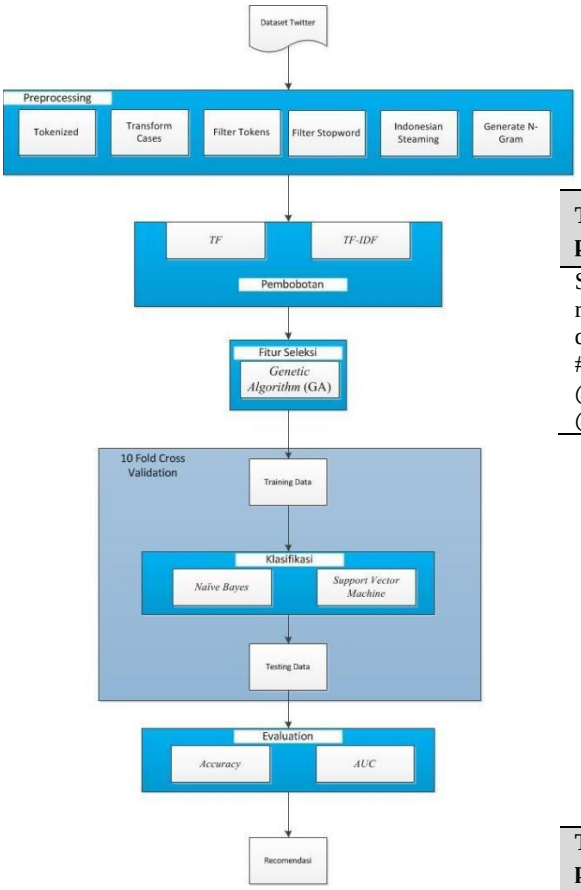
b. Indonesian Stemming
Proses *stemming* adalah variasi dari kata di kelompokkan dengan penghilangan imbuhan sehingga memiliki kata dasar yang sama (Aggarwal, 2015).. Proses pencarian kata dalam sebuah dokumen agar mengetahui seberapa banyak kata yang kemudian dilakukan pembobotan menggunakan TF-IDF.

c. Indonesian Stopword Removal
Proses *Stopword Removal* setiap kata yang selalu muncul namun sedikit memberikan informasi dari isi dokumen, sebagai contoh seperti “apa”,”sebuah”,”untuk”, kata sambung dan sebagainya (Aggarwal, 2015)(Crc et al., 2016).

d. Generate N-Gram
Generate N-Gram merupakan urutan kombinasi kata yang berdekatan dari n item dari ekstrak teks dalam sebuah dokumen dari dua, tiga, empat atau lebih kata. N-gram yang umum digunakan dalam *text mining adalah unigrams, bigrams, dan trigram.* Bigram digunakan untuk mengenerator fitur positif dan negatif (Crc et al., 2016).

D. Metode Yang Usulkan

Dalam penelitian ini metode algoritma yang diusulkan adalah penggunaan 2 (dua) jenis metode algoritma yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) yang kemudian di tambah sebuah seleksi fitur yaitu *Genetic Algorithm* (GA) agar akurasi pengklasifikasi dapat meningkat. Penulis membandingkan ke dua metode algoritma tersebut untuk diketahui metode algoritma terbaik untuk diterapkan bersama dengan seleksi fitur *Genetic Algorithm* (GA). Penulis menggunakan *Support Vector Machine* (SVM) dalam pengklasifikasi karena merupakan teknik *machine learning* yang populer untuk klasifikasi teks serta memiliki performa yang baik pada banyak domain. Kemampuan SVM dalam mengidentifikasi *hyperplane* secara terpisah diantara dua kelas berbeda sehingga termaksimalkan (Indrayuni, 2016). SVM menjamin untuk memaksimalkan jarak antara data yang paling dekat dengan *hyperplane*. Pengkalsifikasi *Naïve Bayes* (NB) merupakan metode klasifikasi teks berdasarkan probilitas kata kunci dalam membandingkan dokumen latih dan dokumen uji. Keudanya dibandingkan melalui beberapa tahanp persamaan, yang akhirnya diperoleh hasil probalitas tertinggi yang di tetapkan sebagai kategori dokumen baru. Lihat gambar 1. untuk model yang diusulkan secara lebih *detail*.



Gambar 1. Model yang diusulkan

Tahapan *preprocessing* merupakan awalan dalam proses pengolahan data agar didapatkan kata-kata yang relevan untuk diklasifikasikan. Validasi dilakukan dengan pengujian *10 Fold Cross Validation.* Pengukuran akurasi, presisi dan *recall* diukur dengan *Confusion matrix* serta Kurva ROC untuk mengukur nilai AUC.

E. Evaluasi dan Validasi Hasil

Model yang diusulkan pada penelitian tentang *tweet* analisis sentimen calon gubernur jawa barat 2018-2023 ini adalah dengan menerapkan *Support Vector Machine* (SVM), *Naive Bayes* (NB), *Support Vector Machine* (SVM) berbasis *Genetic Algorithm* (GA) dan *Naive Bayes* berbasis *Genetic Algorithm* (GA). Penerapan evaluasi algoritma tersebut menggunakan *confusion matrix* untuk menghitung Akurasi, Presisi dan *Recall*. Dengan semakin besar area dibawah kurva (AUC), semakin baik hasil prediksi.

HASIL DAN PEMBAHASAN

A. *Tokenization*

Proses *Tokenization* berfungsi untuk menghilangkan tanda baca, symbol dan katakter yang bukan berupa huruf pada setiap *tweet* . Salah satu hasil dari proses *Tokenization* pada *GataFramework* adalah sebagai berikut.

Tabel 1. Perbandingan teks sebelum dan sesudah dilakukan proses Tokenization dengan data pasangan Ridwan Kamil – Uu Ruhzanul Ulum

Teks sebelum dilakukan proses tokenization	**Teks setelah dilakukan proses tokenization**
Semoga tanggal 27 pasangan rindu menang pilgub jabar 2018 dan tetap juara #RINDUJabarjuara1 @ridwankamil @uuruzhan @infojabar	Semoga tanggal pasangan rindu menang pilgub jabar dan tetap juara rindujabarjuara ridwankamil uuruzhan infojabar

B. *Indonesian Stopword Removal*

Proses *Stopword Removal* setiap kata yang selalu muncul namun sedikit memberikan informasi dari isi dokumen, sebagai contoh seperti “apa”,”sebuah”,”untuk”, kata sambung dan sebagainya.

Tabel 2. Perbandingan teks sebelum dan sesudah dilakukan proses Indonesian Stopword Removal dengan data pasangan Ridwan Kamil – Uu Ruhzanul Ulum

Teks sebelum dilakukan proses Indonesian Stopword Removal	**Teks setelah dilakukan proses Indonesian Stopword Removal**
Semoga tanggal pasangan rindu menang pilgub jabar dan tetap juara rindujabarjuara ridwankamil uuruzhan infojabar	Semoga tanggal pasangan rindu menang pilgub jabar juara rindujabarjuara ridwankamil uuruzhan infojabar

C. *Indonesian Stemming*

Proses *stemming* adalah variasi dari kata di kelompokkan dengan penghilangan imbuhan sehingga memiliki kata dasar yang sama. Proses pencarian kata dalam sebuah dokumen agar mengetahui seberapa banyak kata yang kemudian dilakukan pembobotan menggunakan TF-IDF.

Tabel 3. Perbandingan teks sebelum dan sesudah dilakukan proses Indonesian Stemming dengan data pasangan Ridwan Kamil – Uu Ruhzanul Ulum

Teks sebelum dilakukan proses Indonesian Stemming	**Teks setelah dilakukan proses Indonesian Stemming**
Semoga tanggal pasangan rindu menang pilgub jabar juara rindujabarjuara ridwankamil uuruzhan infojabar	Semoga tanggal pasang rindu menang pilgub jabar juara rindujabarjuara ridwankamil uuruzhan infojabar

D. *Generate N-Gram*

Generate N-Gram merupakan urutan kombinasi kata yang berdekatan dari n item dari ekstrak teks

dalam sebuah dokumen dari dua, tiga, empat atau lebih kata. N-gram yang umum digunakan dalam text mining adalah unigrams, bigrams, dan trigram. Bigram digunakan untuk mengenerator fitur positif dan negatif.

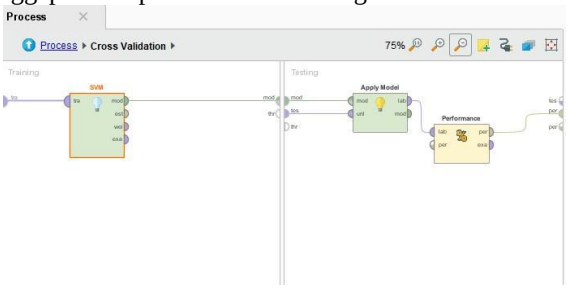
Tabel 4. Perbandingan teks sebelum dan sesudah dilakukan proses *N-Gram* dengan data pasangan Ridwan Kamil – Uu Ruhzanul Ulum

Teks sebelum dilakukan proses N-Gram	**Teks setelah dilakukan proses N-Gram**
Semoga tanggal pasang rindu menang pilgub jabar juara rindujabarjuara ridwankamil uuruzhan infojabar	semoga semoga_tanggal tanggal tanggal_pasang pasang pasang_rindu rindu rindu_menang menang menang_pilgub pilgub pilgub_jabar jabar jabar_juara juara juara_rindujabarjuara rindujabarjuara rindujabarjuara_ridwankamil ridwankamil ridwankamil_uuruzhan uuruzhan uuruzhan_infojabar infojabar

E. Proses Klasifikasi

Klasifikasi didefinisikan sebagai proses menetapkan kategori atau beberapa kategori di antara yang sudah ditentukan sebelumnya untuk setiap item data (Jo, 2018). Dengan menentukan dari sebuah kalimat untuk menjadi sebuah kelas positif dan kelas negatif berdasarkan nilai probabilitas dari yang lebih besar. Termasuk kedalam *class* positif ketika nilai probabilitas kalimat lebih besar pada *class* positif. Jika dalam kategori *class* negatif ketika nilai probabilitas kalimat lebih besar pada *class* negatif. Model pengklasifikasian yang di implementasikan pada Rapidminer 8.2 bisa di lihat pada gambar 2 dan 3.

Tahap pengklasifikasian teks menggunakan data training dari masing- masing data pasangan calon Gubernur Jawa Barat yang terdiri dari Ridwan Kamil- Uu Ruhzanul Ulum 987 data *tweet* , Tubagus Hasanuddin-Anton Charliyan 643 data *tweet* , Sudrajat-Ahmad Syaikhu 541 data *tweet* dan Deddy Mizwar-Dedi Mulyadi 470 data *tweet* . Masing-masing data pasangan calon terdapat *tweet* yang dianggap *tweet* positif dan *tweet* negatif.

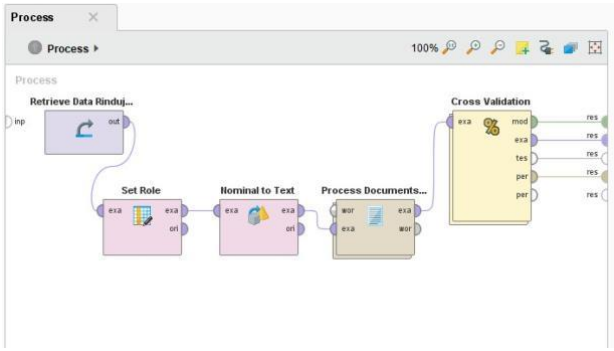


Gambar 2. Model Klasifikasi SVM

Gambar 3. Model Klasifikasi NB

F. Pembobotan dan pemilihan fitur seleksi

Metode pembobotan Fitur yang akan digunakan adalah *Term Frequency Invers Document Frequency* (TF-IDF) dan pemilihan seleksi fitur yang akan diujicoba pertama kali yaitu model algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dengan pengujian 10 *Fold Cross Validation*. Selanjutnya pemilihan seleksi fitur yang akan di ujicoba berikutnya yaitu *Genetic Algorithm* (GA) dengan model algoritma *Support Vector Machine* (SVM) dan *Genetic Algorithm* (GA) dengan model *Naïve Bayes* (NB) pengujian 10 Fold Cross Validation. Model pembobotan yang di implementasikan pada Rapidminer 8.2 bisa di lihat pada gambar 3.

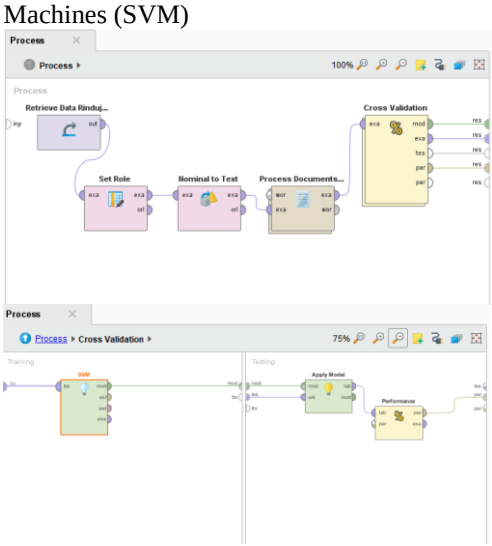


Gambar 4. Model Pembobotan TF-IDF dan K-*Fold Cross Validation*

G. Hasil Eksperimen Pengujian Metode.

a. Metode Support Vector Machine (SVM).

1) Model Klasifikasi Support Vector Machines (SVM)



Gambar 5. Model Klasifikasi SVM

2) Hasil Eksperimen

Nilai *accuracy*, *precision* dan *recall* dari data training pasangan calon Ridwan Kamil-Uu Ruhzanul Ulum dapat dihitung dengan menggunakan RapidMiner. Hasil terbaik pada eksperimen SVM di atas adalah dengan C = 0.0 dan Epsilon = 0.9 serta population size=5 dihasilkan *Accuracy* 89.08% dan AUC = 0.947.

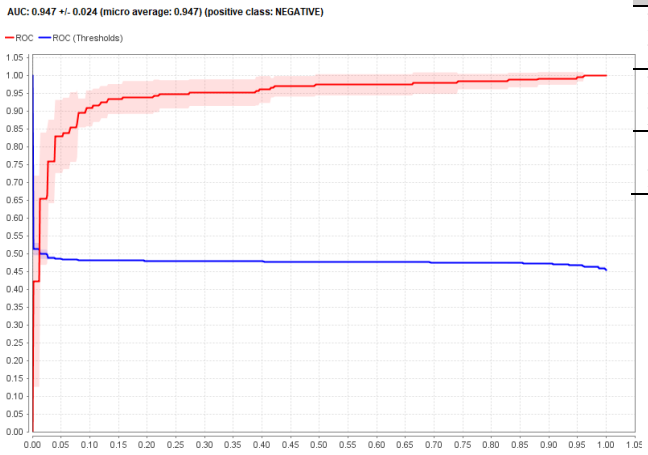
3) Confusion Matrix

Hasil uji terbaik pada pengklasifikasian data *tweet* Ridwan Kamil-Uu Ruhzanul Ulum menggunakan Algoritma *Support Vector Machine* (SVM) dapat di lihat pada gambar berikut:

Tabel 5. Model Confusion Matrix Untuk Metode SVM Rindu

***Accuracy* : 89.08 %, +/- 2.95 % (Mikro avarage: 89.08 %)**			
	True Positif	**True Negatif**	**Class Precission**
Prediksi Positif	754	100	88,29 %
Prediksi Negatif	8	127	94,07 %
Class Recall	98.95%	55.95 %	

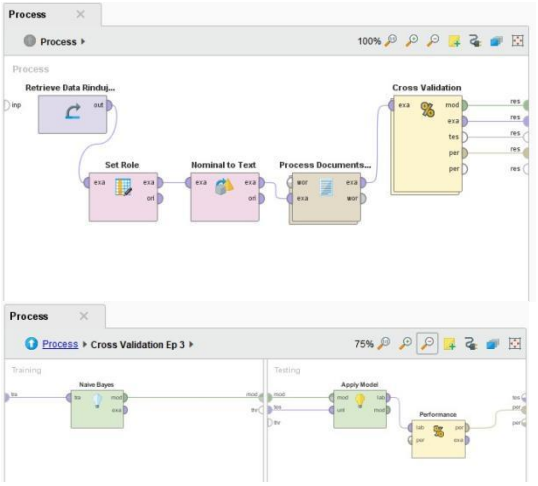
Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat di lihat pada gambar 6 yang mengekspresikan *confusion matrix* dari tabel 5. Garis horizontal adalah *false positive* dan garis vertikal *true positive.*



Gambar 6. Kurva ROC SVM

b. Metode Naïve Bayes (NB)

1) Model Klasifikasi NB



Gambar 7. Model Klasifikasi NB

2) Hasil Eksperimen

Nilai *accuracy*, *precision* dan *recall* dari data *training* pasangan calon Ridwan Kamil-Uu Ruhzanul Ulum dapat dihitung dengan menggunakan RapidMiner. Hasil terbaik pada eksperimen NB di atas dihasilkan *Accuracy* 90.19% dan AUC = 0.531.

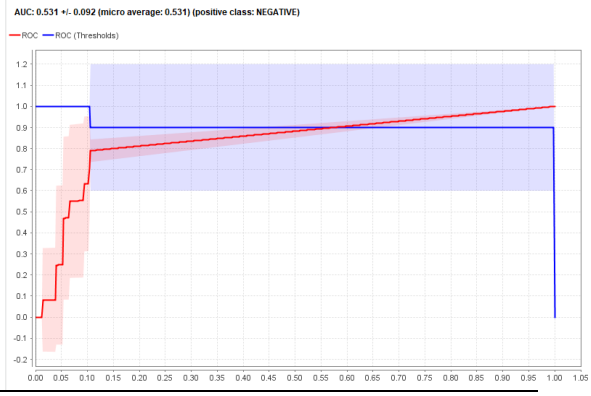
3) *Confusion Matrix*

Hasil uji terbaik pada pengklasifikasian data *tweet* Ridwan Kamil-Uu Ruhzanul Ulum menggunakan Algoritma *Naïve Bayes* (NB) dapat di lihat pada gambar berikut :

Tabel 6. Model Confusion Matrix Untuk Metode NB Rindu

***Accuracy* : 90.19 %, +/- 2.74 % (Mikro avarage: 90.19 %)**			
	True Positif	**True Negatif**	**Class Precission**
Prediksi Positif	715	50	93,46 %
Prediksi Negatif	47	177	79,02 %
Class Recall	93.83%	77.97 %	

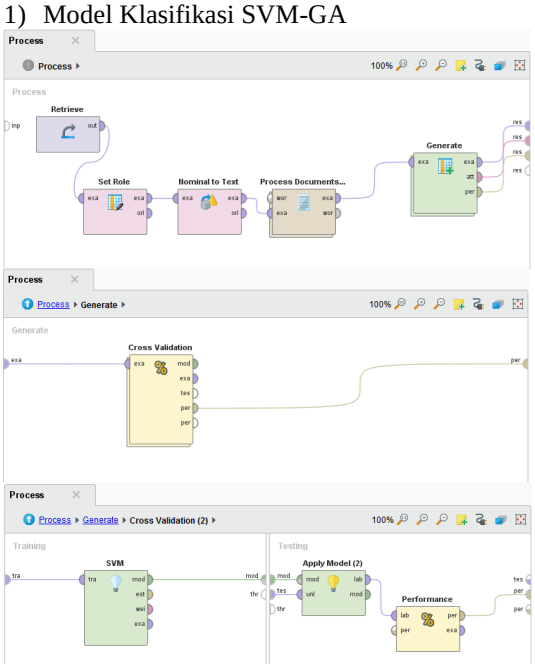
Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat di lihat pada gambar 8 yang mengekspresikan *confusion matrix* dari tabel 6. Garis horizontal adalah *false positive* dan garis vertikal *true positive.*



Gambar 8. Kurva ROC NB

c. Model SVM-GA

1) Model Klasifikasi SVM-GA



Gambar 3. Model Klasifikasi SVM-GA

2) Hasil Eksperimen

Nilai *accuracy*, *precision* dan *recall* dari data training pasangan calon Ridwan Kamil-Uu Ruhzanul Ulum dapat dihitung dengan menggunakan RapidMiner. Hasil terbaik pada eksperimen SVM di atas adalah dengan C = 0.0 dan Epsilon = 0.0 serta population size=5 dihasilkan *Accuracy* 88.98% dan AUC = 0.955.

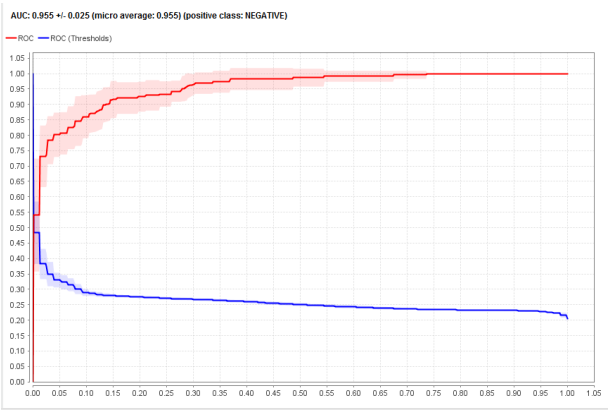
3) *Confusion Matrix*

Hasil uji terbaik pada pengklasifikasian data *tweet* Ridwan Kamil-Uu Ruhzanul Ulum menggunakan Algoritma *Support Vector Machine* berbasis *Genetic Algorithm* (SVM-GA) dapat di lihat pada gambar berikut:

Tabel 7. Model *Confusion Matrix* Untuk Metode SVM – GA Rindu

***Accuracy* : 88.98 %, +/- 2.13 % (Mikro avarage: 88.98 %)**			
	True Positif	**True Negatif**	**Class Precission**
Prediksi Positif	759	106	87,75 %
Prediksi Negatif	3	121	97,57 %
Class Recall	99.61%	53.30 %	

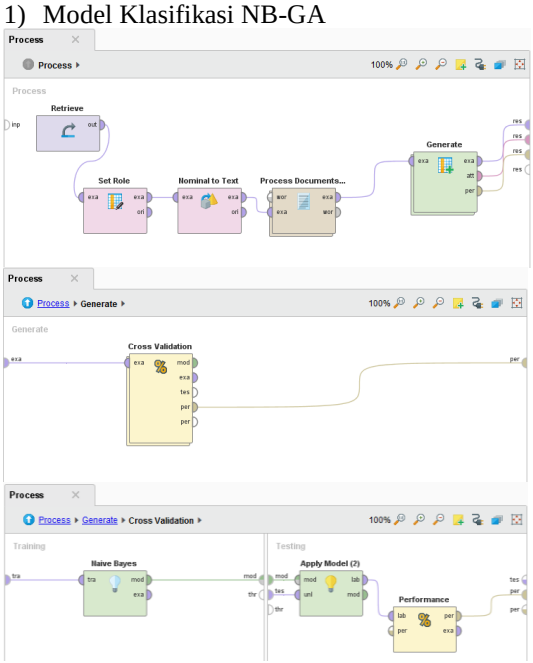
Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat di lihat pada gambar 10 yang mengekspresikan *confusion matrix* dari tabel 7. Garis horizontal adalah *false positive* dan garis vertikal *true positive.*



Gambar 10. Kurva ROC SVM-GA

d. Model NB-GA

1) Model Klasifikasi NB-GA



Gambar 11. Model Klasifikasi NB-GA

2) Hasil Eksperimen

Nilai *accuracy*, *precision* dan *recall* dari data training pasangan calon Ridwan Kamil-Uu Ruhzanul Ulum dapat dihitung dengan menggunakan RapidMiner. Hasil terbaik pada eksperimen NB-GA di atas adalah dengan C = 0.0 dan Epsilon = 0.0 serta population size=5 dihasilkan *Accuracy* 89.99% dan AUC = 0.528.

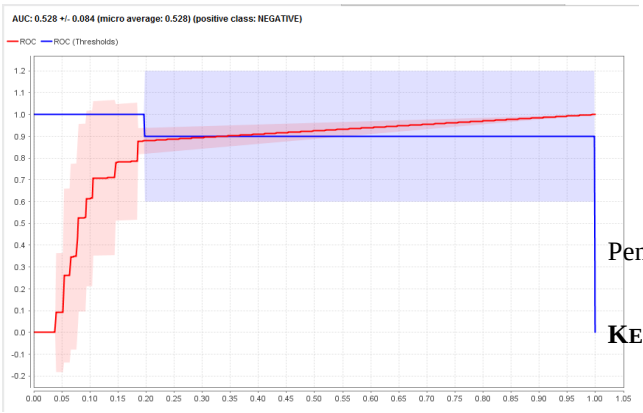
3) Confusion Matrix

Hasil uji terbaik pada pengklasifikasian data *tweet* Ridwan Kamil-Uu Ruhzanul Ulum menggunakan Algoritma *Naive Bayes* berbasis *Genetic Algorithm* (NB-GA) dapat di lihat pada gambar berikut :

Tabel 8. Model *Confusion Matrix* Untuk Metode
NB – GA Rindu

Accuracy : 89.99 %, +/- 3.87 % (Mikro avarage: 89.99 %)			
	True Positif	**True Negatif**	**Class Precission**
Prediksi Positif	694	31	95,72 %
Prediksi Negatif	68	196	74,24 %
Class Recall	91.08%	86.34 %	

Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat di lihat pada gambar 12 yang mengekspresikan *confusion matrix* dari tabel 8. Garis horizontal adalah *false positive* dan garis vertikal *true positive*.



Gambar 12. Kurva ROC NB-GA

H. Analisis Evaluasi Hasil dan Validasi Model

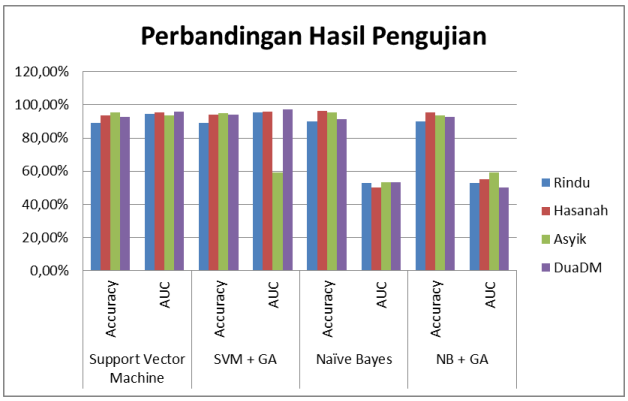
Dari hasil pengujian di atas, pengukuran akurasi menggunakan *confusion matrix* dan kurva ROC terbukti bahwa hasil pengujian algoritma NB memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma SVM, SVM berbasis GA dan NB berbasis GA. Setelah dilakukan pengujian data calon Gubernur periode 2018-2023 dengan calon Gubernur Rindu pada hasil pengujian tertinggi yaitu dengan model algoritma NB sebesar 90.19 %, untuk calon Gubernur Hasanah hasil pengujian tertinggi terdapat pada model algoritma NB sebesar 96.10 %, pada calon Gubernur Asyik sebesar 95.57 % pada model algoritma NB sedangkan calon Gubernur DuaDM akurasi tertinggi pada model algoritma SVM berbasis GA.

Perbandingan hasil pengujian terbaik dapat di lihat pada tabel 9 dan gambar 13 di bawah ini

Tabel 9. Perbandingan hasil pengujian terbaik

Calon Gubernur	**Support Vector Machine**		**SVM + GA**	
	Accuracy	**AUC**	**Accuracy**	**AUC**
Rindu	89.08%	0.947	88.98%	0.955
Hasanah	93.62%	0.953	94.25%	0.958
Asyik	95.19%	0.938	94.83%	0.594
DuaDM	92.55%	0.960	94.04%	0.970

Calon Gubernur	**Naïve Bayes**		**NB + GA**	
	Accuracy	**AUC**	**Accuracy**	**AUC**
Rindu	90.19%	0.531	89.99%	0.528
Hasanah	96.10%	0.500	95.33%	0.550
Asyik	95.57%	0.532	93.53%	0.594
DuaDM	91.28%	0.535	92.55%	0.500



Gambar 13. Diagram Perbandingan Hasil Pengujian Terbaik

KESIMPULAN

Dalam penelitian ini penulis melakukan pengujian model dengan algoritma Naïve Bayes, *Support Vector Machine*, Naïve Bayes berbasis *Genetic Algorithm*, dan *Support Vector Machine* berbasis *Genetic Algorithm* dengan menggunakan 4 (empat) jenis data calon gubernur jawa barat periode 2018-2023 (Pasangan calon gubernur pertama Ridwan Kamil-Uu Ruhzanul Ulum, kedua Tubagus Hasanuddin-Anton Charliyan, ketiga Sudrajat-Ahmad Syaikhu dan keempat Deddy Mizwar-Dedi Mulyadi) dan masing-masing data dikelompokkan menjadi positif dan negatif dengan total data sebanyak 2643 data *tweet* . Model yang diuji akan menghasilkan nilai *accuracy*, *precision*, *recall* dan AUC dari setiap algoritma.

Hasil pengujian data *tweet* mengenai calon gubernur jawa barat periode 2018-2023 dengan Algoritma *Support Vector Machine* menghasilkan rata-rata akurasi 92,61% dengan AUC 0,950, Algoritma Naïve Bayes menghasilkan rata-rata akurasi 93,29% dengan AUC 0,525, Algoritma *Support Vector Machine* berbasis *Genetic Algorithm* menghasilkan rata-rata akurasi 93,03% dengan AUC 0,869 dan Algoritma *Naive Bayes* berbasis *Genetic Algorithm* menghasilkan rata-rata akurasi 92,85% dengan AUC 0,543.

Dengan demikian model algoritma *Support Vector Machine* berbasis *Genetic Algorithm* adalah model algoritma terbaik dalam penelitian ini dan dapat memberikan hasil terbaik dalam pengujian dan pengklasifikasian analisis sentiment *tweet* calon

gubernur jawa barat periode 2018-2023 dibandingkan dengan model algoritma Naïve Bayes berbasis *Genetic Algorithm* (NB-GA).

REFERENSI

Aggarwal, C. C. (2015). *Data Mining*. https://doi.org/10.1007/978-3-319-14142-8

Alfiah, F., Susanti, E., Kristinna, J., Ardiansyah, O. R., & Pradipta, D. (2015). *Manfaat Menganalisis Pengaruh Sosial Media*. 6–8.

Balahur, A., Mihalcea, R., & Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech and Language*, *28*(1), 1–6. https://doi.org/10.1016/j.csl.2013.09.003

Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, *53*, 453–462. https://doi.org/10.1016/j.proeng.2013.02.059

Bohang, F. K. (2017). Twitter 280 Karakter Resmi di Seluruh Dunia. Retrieved from tekno.kompas.com website: https://tekno.kompas.com/komentar/2017/11/08/08340057/twitter-280-karakter-resmi-di-seluruh-dunia

Crc, H., Hofmann, M., & Chisholm, A. (2016). *Text Mining and Visualization Case Studies Using Open Source Tools*.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *San Francisco, CA, itd: Morgan Kaufmann*. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi Volume 4 Nomor 2 - 2016*, *4*(2), 20–27.

Jo, T. (2018). Text Mining. In *Springer, Cham*. https://doi.org/10.1016/B978-0-12-396963-7.00010-6

Ramesh, B., & Sathiaseelan, J. G. R. (2015). An Advanced Multi Class Instance Selection based Support Vector Machine for Text Classification. *Procedia Computer Science*, *57*, 1124–1130. https://doi.org/10.1016/j.procs.2015.07.400

Sarlan, A., Nadam, C., & Basri, S. (2015). Twitter Sentiment Analysis. *ArXiv:1507.00955 [Cs, Stat]*, 212–216. https://doi.org/10.1109/ICIMU.2014.7066632

Wahyudi, M., & Putri, D. W. I. A. (2016). *ALGORITHM APPLICATION SUPPORT VECTOR MACHINE WITH GENETIC ALGORITHM OPTIMIZATION TECHNIQUE FOR SELECTION FEATURES FOR THE ANALYSIS OF*. *84*(3).

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. *Data Mining: Practical Machine Learning Tools and Techniques*, 1–621.

Zukhri, Z. (2014). *Algoritma Genetika : Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*. Yogyakarta: Andi.