Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019
masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 3 No. 2 (2019) 243 - 250 ISSN Media Elektronik: 2580-0760

Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma Support Vector Machine

Heru Sukma Utama[1], Didi Rosiyadi[2], Bobby Suryo Prakoso[3], Dedi Ariadarma[4]
[1,2,3,4]Magister Ilmu Komputer, Ilmu Komputer, STMIK Nusa Mandiri Jakarta
[5]Pusat Penelitian Informatika, LIPI
hsukmautama@gmail.com

Abstract

Analysis of the odd even-numbered sentiment systems in Bekasi toll using the Support Vector Machine Algorithm, is a process of understanding, extracting, and processing textual data automatically from social media. The purpose of this study was to determine the level of accuracy, recall and precision of opinion mining generated using the Support Vector Machine algorithm to provide information community sentiment towards the effectiveness of the odd system of Bekasi tiolls on social media. The research method used in this study was to do text mining in comments-comments regarding posts regarding even odd oddities on Bekasi toll on Twitter, Instagram, Youtube and Facebook. The steps taken are starting from preprocessing, transformation, datamining and evaluation, followed by information gaon feature selection, select by weight and applying SVM Algorithm model. The results obtained from the study using the SVM model are obtained Confusion Matrix result, namely accuracyof 78.18%, Precision of 74.03%, and Sensitivity or Recall of 86.82%. Thus this study concludes that the use of Support Vector Machine Algorithms can analyze even odd sentiments on the Bekasi toll road.

Keywords : Text Mining, Super Vector Machine

Abstrak

Analisis sentimen sistem ganjil genap di tol Bekasi menggunakan Algoritma *Support Vector Machine*, merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis dari media sosial. Tujuan dari penelitian ini adalah untuk mengetahui tingkat *accuracy, recall* dan *precision* dari *opinion mining yang dihasilkan menggunakan algoritma support vector machine* memberikan informasi sentiment masyarakat terhadap efektifitas sistem ganjil genap tol bekasi di media sosial Metode penelitian yang dilakukan dalam penelitian ini adalah melakukan *text mining* pada komentar-komentar terkait postingan mengenai efektifitas ganjil genap di tol bekasi pada *Twitter, Instagram, Youtube* dan *Facebook*. Tahapan yang dilakukan adalah mulai dari *preprocessing*, *transformation*, *datamining* dan *evaluation*, dilanjutkan dengan seleksi fitur *information gain*, *select by weight* dan penerapan model Algoritma SVM. Hasil yang didapatkan dari penelitian dengan menggunakan model SVM adalah didapatkan hasil *Confusion Matrix*, yaitu *accuracy* sebesar 78,18%, *Precision* sebesar 74,03%, dan *Sensitivity* atau *Recall* sebesar 86,82%. Dengan demikian penelitian ini menyimpulkan bahwa penggunaan Algoritma Support Vector Machine dapat menganlisis sentimen ganjil genap di tol Bekasi.

Kata Kunci :Text Mining, Super Vector Machine

1. Pendahuluan

Penelitian ini membahas tentang analisis sebuah kegiatan atau program. Sebuah kebijakan atau program apalagi program pemerintah tentunya harus dikritisi agar kebijakan tersebut bisa diperbaiki dan hasilnya dapat dinikmati rakyat dengan baik. Analisis sentimen didefinisikan sebagai ilmu untuk melakukan analisis dari pendapat, sikap, emosi seseorang ke dalam bahasa tertulis [1]. Dengan dianalisis, maka kita dapat memahami seperti apa respon masyarakat terhadap program yang sedang berjalan. Analisis sentimen atau

Diterima Redaksi : 22-07-2019 | Selesai Revisi : 02-08-2019 | Diterbitkan Online : 03-08-2019

opinion mining merupakan salah satu solusi mengatasi masalah untuk mengelompokan opini atau review menjadi opini positif atau negatif secara otomotis [2]. Jika opini telah kelompokkan, maka tentunya kita akan mudah untuk menganalisis semua data.. Fokus dari opinion mining adalah melakukan analisis opini dari suatu dokumen teks [4]. Secara umum, *opinion mining* diperlukan untuk mengetahui sikap seorang pembicara atau penulis sehubungan dengan beberapa topik atau polaritas kontekstual keseluruhan dokumen.

Sosial media adalah sebuah media untuk bersosialisasi satu sama lain dan dilakukan secara online yang memungkinkan manusia untuk saling berinteraksi tanpa dibatasi ruang dan waktu [5]. Kalau dulu media itu sifatnya terbatas hanya beberapa saja, tetapi sekarang seiring dengan perkembangan jaman media khususnya media sosial menjadi sarana favorit yang ingin dimikiki warga masyarakat. Banyak manfaat yang bisa diperoleh dari media sosial, akan tetapi banyak juga mudaratnya. Salah satu manfaat yang paling jelas dengan adanya media sosial adalah semua masyarakat dengan cepat dapat menyampaikan pendapat, kritik, dan saran dan bisa juga menyampaikan unek-uneknya secara bebas. Dan salah satu dampak buruknya adalah banyaknya *hate speech*.

Media sosial terdiri atas *Twitter, instagram, youtube, facebook.* Salah satu media sosial yang sangat banyak menampilkan opini masyarakat adalah twitter. Pada umumnya kebanyakan kebijakan pemerintah akan lebih disorot dengan media twitter. Berbeda dengan media sosial yang lain. Setiap hari server Twitter menerima data tweet dengan jumlah yang sangat besar, dengan demikian, kita dapat melakukan data mining yang digunakan untuk tujuan tertentu [6].

Ada banyak macam jalan, yang kita jumpai di negeri kita. Salah satu jalan yang sangat diminati masyarakat adalah jalan tol. Banyak jalan tol yang sudah dibuat pemerintah. Termasuk di Bekasi juga sudah banyak jalan tol yang dibangun pemerintah. Akan tetapi kalo berbicara jalan tol di Bekasi, tentu tidak lepas juga bicara dengan kemacetan. Karena biarpun banyak jalan tol, jalanan di Bekasi tetaplah macet. Jalan Tol diselenggarakan untuk mendukung pergerakan lalu lintas secara optimal serta meningkatkan efisiensi pelayanan jasa distribusi guna menunjang peningkatan pertumbuhan ekonomi terutama di wilayah yang tingkat perkembangan ekonominya tinggi. [7] Untuk menggunakan fasilitas ini, para pengguna jalan tol harus membayar sesuai tarif yang berlaku. Penetapan tarif didasarkan pada golongan kendaraan. Sistem ganjil genap adalah satu konsep pembatasan kendaraan yang mengacu pada dua nomor terakhir pelat nomor kendaraan. Dengan begitu, nantinya setiap kendaraan yang melintas akan bergantian sesuai hari pemberlakuan dua digit angka terakhir pelat nomornya. Seperti kita ketahui bersama sistem genap ganjil adalah satu konsep pembatasan kendaraan yang mengacu pada dua nomor terakhir pelat nomor kendaraan. Dengan begitu, nantinya setiap kendaraan yang melintas akan bergantian sesuai hari pemberlakuan dua digit angka terakhir pelat nomornya.

Dilihat dari permasalahan yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran maupun keluhan yang diterima oleh Jasa Marga selaku operator jalan tol Bekasi sehingga dapat diketahui informasi sentimen mengenai efektifitas ganjil genap di tol Bekasi. Adapun permasalahan pengklasifikasian sebuah kalimat sentimen ke dalam kelas–kelas tertentu dapat diselesaikan dengan *Support Vector Machine*. Salah satu sumber informasi yang dibutuhkan oleh Pemerintah untuk dapat meningkatkan kinerjanya adalah umpan balik dari masyarakat [8]. Oleh karena itu penelitian ini sangat diperlukan untuk membantu pemerintah dalam mengevaluasi program yang dijalankan. Tanpa adanya evaluasi dari masyarakat, maka pemerintah tidak akan mengetahui sejauh mana keberhasilan program yang dijalankan. Harapannya penelitian ini bisa menjadi masukan utamanya buat pemerintah dalam mencari solusi mengatasi kemacetan terutama di Bekasi.

Text mining adalah salah satu teknik penambangan data yang berupa teks [1]. Teks yang diambil dari media sosial yang tentu tidak menggunakan bahasa baku perlu sekali diolah menggunakan *text mining*. Proses *text mining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep/entitas, produksi taksonomi *granular, sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas [9]. Tidak semua orang bisa dengan mudah memahami ungkapan atau kalimat yang disampaikan orang lain. Untuk itulah perlu adanya teknik yang bisa menafsirkan makna dari sebuah kalimat. Hal ini sesuai dengan pendapat yang menyebutkan text mining merupakan salah satu teknik yang digunakan untuk menggali kumpulan dokumen text sehingga dapat diambil intisarinya [11]. Dalam melakukan *text mining* akan dilakukan tahap-tahap sebagai berikut : *selection, preprocessing, transformation, datamining* dan *evaluation.* Gartner group menyebutkan bahwa data mining adalah proses menelusuri pengetahuan yang baru, pola, dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan tehnik pengenalan pola serta statistik dan teknik matematika[12].

Data *text mining* diolah menggunakan *Knowledge Discovery Database* (KDD**).** KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [13]. Data yang diambil dari penelitian biasanya tidak sedikit. Oleh karena itu bila data-data tersebut dianalisis secara manual tentunya akan memakan waktu yang cukup lama**.** Jumlah yang banyak dan waktu yang cepat akan menyulitkan editor mengkategorikan secara manual

[14]. Sehingga diperlukanlah sebuah tool untuk memudahkan peneliti dalam mengolah data.

Pengurutan data atau *sorting* merupakan suatu proses dimana suatu susunan data yang semula dalam kondisi acak dapat menjadi urut, baik dari data terkecil sampai dengan data yang terbesar, atau sebaliknya dari data terbesar sampai dengan data terkecil [13]. Dengan diurutkan maka peneliti akan mudah mengolah data. Selain itu orang lain juga akan membacanya. Pada tahap *selection* dicari data yang berhubungan dengan tema penelitian kali ini dengan seleksi fitur dan mengujinya dengan algoritma *support vector machine*. Kemudian pemahaman tersebut diubah menjadi sebuah rencana awal data mining yang dirancang untuk mencapai tujuan. Sedangkan tahap Preprocessing terdiri dari beberapa tahap yaitu *cleansing, tokenizing, stopword removal*, dan *stemming* [15]. Tahapan-tahapan ini dilalui untuk mendapatkan intisari dari data yang diambil.

Cleansing data yaitu mengurangi *noise* pada data tweet, *transform case* tahapan ini merupakan proses merubah bentuk huruf menjadi huruf kecil (*lower case*) atau dapat disebut juga penyeragaman bentuk huruf. *Stopword Removal*, merupakan proses menghilangkan daftar kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan seperti "yang", "di", "ke", "itu" dan lain sebagainya , *tokenizing* atau *parsing* adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses *tokenizing* adalah pemenggalan kalimat menjadi kata, *filter tokens* adalah tahap dimana menghilangkan kata yang dikonfigurasikan untuk dihilangkan berdasarkan jumlah hurufnya. Sebagai contoh *filter tokens* dengan minimal 3 huruf maka kata "ya", "yg", dan "kk" akan hilang pada kalimat tersebut. Tahap ini juga sangat penting karena akan mempengaruhi akurasi penelitian

Pembobotan dilakukan untuk mendapatkan nilai dari kata *term* yang telah diekstrak. *Term* dapat berupa kata atau frasa dalam kalimat tentunya untuk mengetahui maksud, tujuan dan konteks kalimat tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata tersebut diberikan sebuah indikator, yaitu *term weight. Term weighting* atau pembobotan kata sangat dipengaruhi oleh hal-hal berikut ini : 1) *Document Frequency (df),* metode *document frequency (df)* merupakan salah satu metode pembobotan dalam bentuk sebuah metode yang merupakan perhitungan jumlah dokumen yang mengandung suatu term tertentu. Tiap term akan dihitung nilai *document frequency-nya (df), 2) Term Frequency (tf), term frequency (tf)* yaitu faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah memunculkan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* (tf tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar, 3) *Inverse Document Frequency (idf), Inverse Document Frequency (idf)* yaitu pengurangan dominasi term yang sering muncul di berbagai dokumen. Hal ini diperlukan karena *term* yang banyak muncul di berbagai dokumen, dapat dianggap sebagai *term* umum sehingga tidak penting nilainya. Sebaliknya, faktor jarangnya munculnya kata dalam kumpulan dokumen harus diperhatikan dalam pemberian bobot. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*Inverse Document Frequency*).

Setelah dilakukannya proses pembobotan, data tersebut akan melalui tahap pengklasifikasian dengan metode *Support Vector Machine* agar dapat melakukan analisis dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasikan sebelumnya. Algoritma klasifikasi yang dapat melakukan teks mining diantaranya *Support Vector Machine* (SVM), *Naïve Bayessian classification* (NBC) dan *K-Nearest Neighbor* (K-NN) [3]. *Support Vector Machine* adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi [15]. Berdasarkan data yang ada algoritma ini banyak digunakan untuk penelitian. *Support Vector Machine* (SVM) dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learn-ing Theory* [16].

Peneliti sangat tertarik menggunakan algoritma ini, karena sudah lama ditemukan. Konsep dasar SVM sebenarnya merupakan kombinasi harmonis dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperplane* [16]. *Support Vector Machine* berada dalam satu kelas dengan *Artificial Neural Network (ANN)* dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan. Selain itu algoritma ini juga merupakan algoritma terbaik. Algoritma *Support Vector Machine termasuk kedalam* kelas *supervised learning.* Hal ini tercermin dari hasil studi yang dilakukan oleh ICDM pada tahun 2006, mengenai top 10 algoritma dalam Data Mining. Studi yang dilakukan oleh Xindong Wu dan Vipin Kumar ini mengidentifikasikan SVM pada peringkat ke-3, setelah C4.5 dan k-Nearest Neighbor Classifier [16]. Algoritma *Support Vector Machine* juga memiliki beberapa karakteristik. Karakteristik dari *Support Vector Machine* ada 2 yaitu : *Pattern recognition,* dilakukan dengan mentransformasikan data pada ruang input (*input space*) ke ruang yang berdimensi lebih tinggi (*feature space*), dan optimisasi dilakukan pada ruang vector yang baru tersebut [15].

Prinsip kerja *SVM* pada dasarnya hanya mampu menangani klasifikasi dua kelas, namun telah dikembangkan untuk klasifikasi lebih dari dua kelas dengan adanya *pattern recognition. Pattern*

Recognition (PR) didefinisikan sebagai proses pemetaan suatu data ke dalam konsep tertentu yang telah didefinisikan sebelumnya [16]. Prinsip dasar SVM adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada problem *non-linear*. [16] dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi Dalam penelitian ini, teknik. *Support Vector Machine* juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network (NN)* [18].

Tujuan Penelitian ini adalah mengetahui tingkat *accuracy, recall* dan *precision* dari *opinion mining yang dihasilkan menggunakan algoritma SVM,* memberikan informasi sentimen masyarakat terhadap efektifitas sistem ganjil genap tol bekasi di media sosial. Pengklasifikasi *Support Vector Machine* adalah teknik *machine learning* yang populer untuk klasifikasi teks karena dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya dan juga mampu memberikan hasil yang baik [19].

Penelitian ini pernah dilakukan sebelumnya oleh Nanang Ruhyana, 2019 menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap, dalam penelitian ini menghasilkan *accuracy* 86,67%, *precision* 71,43% dan *recall* 80,00% dan Dwi Suci Ariska Yanti, Indriati, Putra Pandu Adikara, 2019 berdasarkan hasil pengujian, sistem ini memiliki nilai *F-Measure* tertinggi sebesar 66,1% dan nilai akurasi sebesar 66,5%. Untuk mendapatkan solusi dari permasalahan di atas, penelitian difokuskan untuk menjawab pertanyaan riset sebagai berikut : Bagaimana cara memanfaatkan *Twitter, instagram, youtube, facebook* untuk menganalisis sentiment efektifitas sistem ganjil genap tol bekasi menggunakan seleksi fitur *Information Gain* dengan *Support Vector Machine* ? dan Bagaimana mengukur tingkat *accuracy, recall* dan *precision* dari *opinion mining* yang dihasilkan menggunakan seleksi fitur *Support Vector Machine* Dan melihat sentiment masyarakat terhadap efektifitas sistem ganjil genap tol bekasi ?

2. Metode Penelitian

Metode penelitian yang dilakukan dalam penelitian ini terbagi menjadi beberapa tahapan, berikut ini adalah penjelasannya.

2.1 Pengambilan Dataset

Pada kegiatan ini, peneliti mengambil data mengenai efektifitas ganjil genap di tol Bekasi dari media sosial, baik dari Twitter, Instagram, Youtube dan Facebook.

2.2 *Teks Pre-Processing*

Pada tahapan ini, peneliti mengolah data yang telah dikumpulkan dari awal sampai mendapatkan data yang akan dirubah menjadi bentuk nominal.

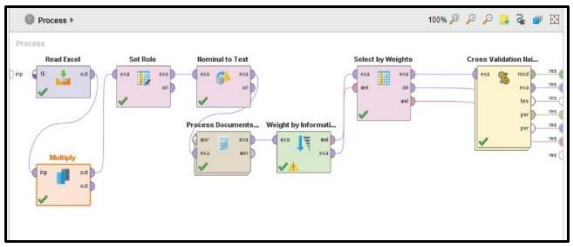
2.3 Seleksi Fitur *Information Gain*

Dalam penggunaan seleksi fitur penulis menggunakan Algoritma *Information Gain* dan *Select by weight. Information Gain* merupakan metode seleksi fitur paling sederhana dengan melakukan perangkingan atribut dan banyak digunakan dalam aplikasi kategorisasi teks, analisis data *microarray* dan analisis data citra [19]. Metode ini sangat diperlukan untuk mengurangi noise karena fitur tidak relevan dan mengetahui fitur yang tidak sesuai. Penentuan atribut dilakukan dengan menghitung nilai *entropy* terlebih dahulu.

Lalu dilakukan sentimen analisis dari tweet, komentar dan postingan yang disampaikan oleh masyarakat luas. Penelitian ini memiliki hipotesis bahwa penggunaan algoritma *Support Vector Machine* dapat menentukan pola terbaik dengan mempelajari sentiment analisis berupa *text* yang ada saat ini dan mempercepat dalam proses perhitungannya. Penelitian ini juga diharapkan dapat memberikan kontribusi dalam bidang *text mining* berbahasa Indonesia. Objek penelitian ini berkaitan dengan data mining dan *text mining*, yang bersumber datanya berupa data teks yang diambil dari media *social*, yang bertemakan ganjil genap Bekasi. Media *social* tersebut terdiri dari *facebook, twitter, instagram* dan *youtube*. Keempat media *social* tersebut termasuk dalam paling banyak terakses oleh para pengguna internet yang bisa memberikan gambaran penggunaan dalam dataset yang digunakan.

3. Hasil dan Pembahasan

Tujuan dari penelitian ini adalah untuk mengetahui tingkat *accuracy, recall* dan *precision* dari *opinion mining yang dihasilkan menggunakan algoritma support vector machine* memberikan informasi sentiment masyarakat terhadap efektifitas sistem ganjil genap tol bekasi di media sosial : *Twitter, instagram, youtube, facebook.* Proses pemodelan menggunakan metode *Suport Vector Machine (SVM)* seperti tampak pada Gambar 1.



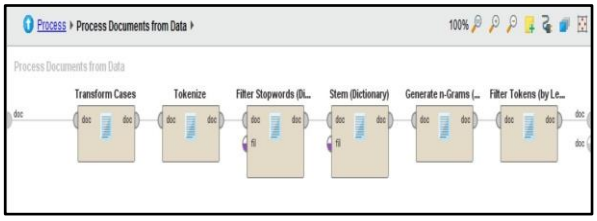
Gambar 1. Proses Pemodelan Menggunakan SVM

Data set yang digunakan pada penelitian ini berjumlah 440 dataset, terbagi menjadi dua yang terdiri dari 220 set bernilai positif dan 220 set bernilai negatif seperti pada Tabel 1.

Tabel 1. Dataset Penelitian

No	Nilai	Ulasan
1	positif	Pak, kl mobil dari Bandung menuju Jakarta bagaimana pengenaan ganjil genapny
2	positif	Semoga solusi pemerintah bisa berjalan seperti apa yg diharapkan.. terlalu banyak pengguna roda 4 mugkin ya di ibu kota
3	positif	Nah begini negara bisa maju kayanya eheheh
4	positif	Bagus pak polisi lanjutkan
5	positif	ayo galakkan naik krl dan busway walau penuh tp itulah yg namanya nyari duit di ibukota
6	negatif	Ganjil genap kayak gimana???
7	negatif	Puyeng puyang dah yang kaya gini . Kasian yang gak paham dan tbatba di tilang ... ahhhsudahlah!
8	negatif	Demen bgt ribettttttttt
9	negatif	Polisi nya jd maen gala asin gtu…
10	negatif	Tambah ruwet ju
Dst.		

Tabel 1. menunjukkan betapa banyaknya kata-kata yang tidak baku di medsos, karena dimedsos biasanya orang menuliskan kalimat dan kata dengan bahasa tidak resmi. Demikian juga dalam penggunaan tanda bacanya. Dari 440 data set, peneliti hanya menyampaikan sampel sebanyak 10 data. Selanjutnya dataset tersebut diolah dengan menggunakan tool sebagaimana terlihat pada Gambar 2.



Gambar 2. Tahap Preprocesing

Gambar 2 menunjukkan tahapan-tahapan dalam mengolah data sampai akhir.

3.1 Transformation

Pada tahap *Transformation* dilakukan seleksi fitur atas data yang telah melalui tahap *pre-processing* dengan menggunakan algoritma *Support Vector Machine* (SVM). Tahapan yang dilakukan adalah data yang diolah diberikan bobot dari informasi yang telah tersedia dari *dataset*. Seleksi Fitur *Information Gain* merupakan suatu teknik dalam mengurangi jumlah fitur yang sesuai atau relevan, lalu mengurangi dimensi fitur pada data yang akan digunakan. Untuk menghitung *infomation gain* menggunakan hitungan sebagai berikut [20] :

$$info(D) = -\sum_{i=1}^{c} p_i log_2(p_i) \qquad (1)$$

Keterangan dari rumus tersebut adalah:

c : jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi)

p_i : jumlah sampe untuk kelas i

$$info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} x\ info(D_j) \qquad (2)$$

Keterangan dari rumus tersebut adalah:

A : atribut

$|D|$: jumlah seluruh sampel data

$|D_j|$: jumlah sampel untuk nilai j

v : suatu nilai yang mungkin untuk atribut A

Selanjutnya nilai *information gain* yang akan dipakai dengan dihitung menggunakan rumus dibawah ini[21]

$$Gain\ (A) = |info(D) - info_A(D)| \qquad (3)$$

3.3 *Data Mining*

Pada tahapan ini dilakukan proses mengelolah data yang telah siap dikelola, dengan pengujian *10-fold cross validation*. Data *training* dibagi secara acak ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh error rate untuk mendapatkan error rate secara keseluruhan. Pada penelitian ini penulis menggunakan metode pengujian 10 *fold cross validation* yang akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian, karena hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa 10 *fold cross validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat.

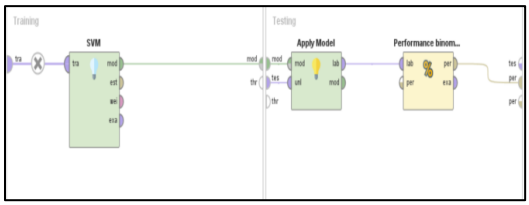
Metode tf-idf merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembanding terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot term t dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*. Metode tf-idf dapat dilihat pada Rumus 1 berikut :

weig *t (t,d)* = *TermFreq(t,d)* x id*f* ……..(1)

Idf = log (*N/df(t)*)

Sumber : Feldman dan Sanger (2007)

Notasi Term Freq (t,d) adalah jumlah kemunculan kata t dalam dokumen d, N adalah jumlah seluruh dokumen dan *DocFreq(t)* adalah jumlah dokumen yang mengandung term t**.** Sebuah dokumenteks dapat dilihat sebagai kumpulan pernyataan subjektif dan objektif. Pernyataan objektif tersebut berkenaan dengan informasi faktual yang ada dalam teks dan subjektivitas berkaitan dengan ekspresi dari opini dan spekulasi. Berikut adalah hasil pengolahan data yang dilakukan dengan menggunakan algoritma *Support Vector Machine.*



Gambar 3. *Cross Validation SVM*

3.4 Evaluation

Tahap evaluasi merupakan tahapan dimana dilakukan interpretasi terhadap hasil *text mining* yang telah dihasilkan pada tahapan sebelumnya. Evaluasi yang dilakukan, akan dilakukan secara mendalam dengan tujuan agar hasil pada tahapan sebelumnya sudah sesuai dengan tujuan yang text data mining yang dilakukan, telah sesuai dengan keinginan perusahaan. Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC* serta akan ditampilkan dalam bentuk kurva *ROC.* Berikut ini adalah alur penelitian yang akan dilakukan. Dalam penelitian ini *performance* diukur. Berikut ini adalah alur penelitian yang akan dilakukan. Pada tahapan evaluasi dilakukan berdasarkan dua langkah, dimana langkah tersebut adalah:

3.4.1 Evaluate Results

Pada tahapan ini dilakukan evaluasi terhadap keluaran yang dihasilkan oleh proses *text mining*. Hasil tersebut dibandingkan dengan tujuan yang terdapat pada tahap *business understanding*, kemudian diketahui sejauh mana hasil dari proses *text mining* ini memenuhi tujuan yang ingin dicapai. Hasil pengujian yang dilakukan melalui model SVM menghasilkan *Confusion Matrix*, yaitu *accuracy* sebesar 78,18%, *Precision* sebesar 74,03%, dan *Sensitivity* atau *Recall* sebesar 86,82% seperti terlihat pada gambar 4.

Table View Plot View

accuracy: 78.18% +/- 5.05% (micro average: 78.18%)

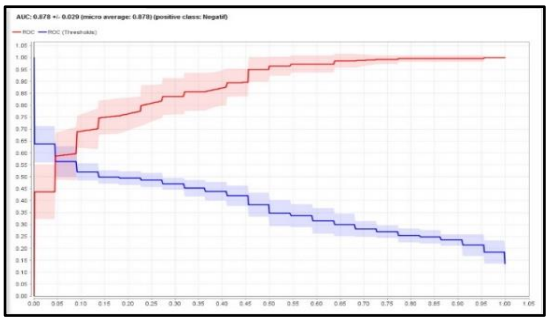
	true Positif	true Negatif	class precision
pred. Positif	191	67	74.03%
pred. Negatif	29	153	84.07%
class recall	86.82%	69.55%	

Gambar 4. *Accuracy* dari Model SVM

Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC* serta akan ditampilkan dalam bentuk kurva *ROC.* Berikut ini adalah alur penelitian yang akan dilakukan *Area Under Curve* dari model Algoritma SVM dapat dilihat pada Gambar 5.

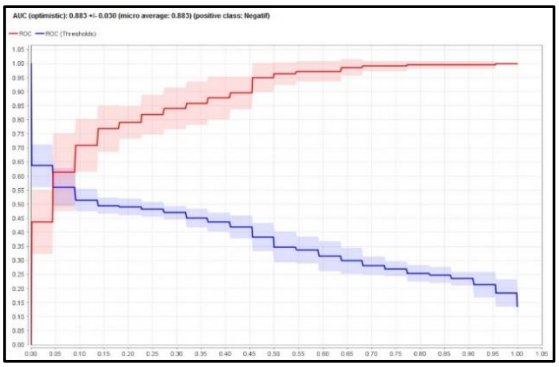
3.4.2 Review Process

Setelah dilakukan pemeriksaan terhadap keseluruhan proses, dimana pemeriksaan dilakukan dengan kembali ke tahapan awal. Dipastikan bahwa tidak ada faktor atau parameter penting yang terlewatkan dari proses *text mining*.



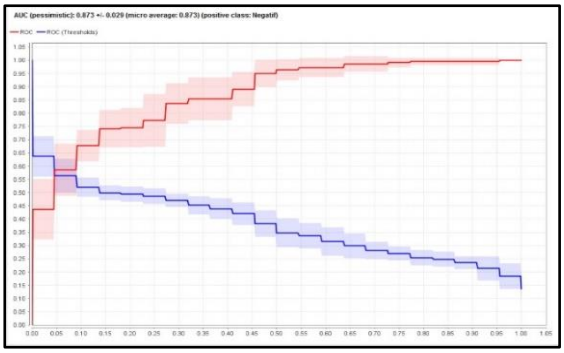
Gambar 5. AUC dari model SVM

curve optimis juga terdapat SVM yang hasilnya :



Gambar 6. AUC Optimis dari model SVM

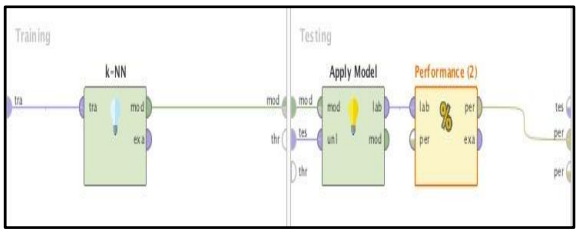
Sedangkan Area Under Curve pesimistis dari algoritma SVM dapat dilihat pada Gambar 7.



Gambar 7. AUC Pesimistis dari model SVM

Selain itu peneliti juga membandingkan hasil penelitian SVM ini dengan metode *K-Nearest Neighbor* (K-NN). Hasilnya ternyata SVM lebih baik dibandingkan dengan K-NN. Tingkat akurasi SVM lebih tinggi dibandingkan dengan K-NN. Algoritma *k-Nearest Neighbor* (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN [22]. Sedangkan pendapat yang lain menyebutkan bahwa K-NN merupakan metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan

[23], dan ada juga pendapat yang menyebutkan bahwa metode K-Nearest Neighbor (K-NN), yaitu metode yang memperhitungkan kemiripan jumlah kemunculan kata antara satu dokumen dengan dokumen lain [24]. Pada penelitian ini tingkat Accuracy SVM adalah sebesar 78.18% +/- 4.79% (micro average: 78.18%) sedangkan akurasi K-NN 57.05% +/- 6.54% (micro average: 57.05%). Berikut adalah gambar praprocessing dengan menggunakan K-NN.



Gambar 8. Tahap praprocesing dengan menggunakan metode K-NN

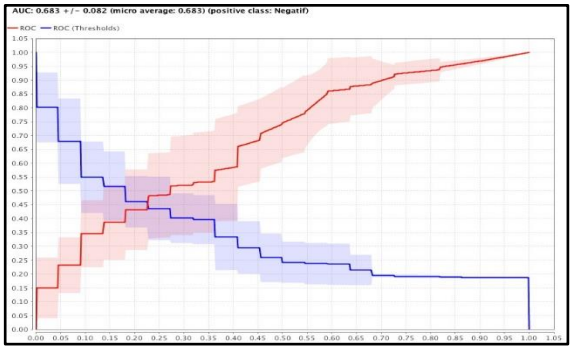
Tingkat akurasi terlihat hanya sekitar 57,05% seperti pada Gambar 9.

accuracy: 57.05% +/- 6.54% (micro average: 57.05%)

	true Positif	true Negatif	class precision
pred. Positif	183	152	54.63%
pred. Negatif	37	68	64.76%
class recall	83.18%	30.91%	

Gambar 9. Tingkat akurasi menggunakan K-NN

Berikut ini adalah alur penelitian yang akan dilakukan *Area Under Curve* dari model Algoritma K-NN :



Gambar 10. AUC dari model K-NN

4. Kesimpulan

Setelah dilakukan penelitian maka dapat disimpulkan bahwa tingkat Hasil pengujian yang dilakukan melalui model SVM menghasilkan *Confusion Matrix*, yaitu *accuracy* sebesar **78,18%**, *Precision* sebesar **74,03%**, dan *Sensitivity* atau *Recall* sebesar **86,82%.** Penelitian selanjutnya yang akan dilakukan adalah membandingkan metode algoritma antara SVM dan NB dengan menambahkan seleksi fitur *Information Gain* dan *Select By Weight* Untuk mencari hasil akurasi mana algoritma terbaik diantara keduanya.

Daftar Referensi

[1] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Niave Bayes Classifier," *Pros. SNST*, 2016.

[2] E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," *J. Evolusi Vol. 4 Nomor 2 - 2016*, 2016.

[3] J. Ipmawati, Kusrini, and E. Taufiq Luthfi, *Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen*, vol. 6, no. 1. 2017.

[4] I. F. Rozi, S. Hadi, and E. Achmad, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," vol. 6, no. 1, pp. 37–43, 2012.

[5] Rafi Saumi Rustian, "Apa itu Sosial Media," *01 maret 2012*, 2012.

[6] "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, 2013.

[7] N. Winarsih and J. Kusumaningrum, "Analisis Kapasitas Gerbang Tol Karawang Barat," *Psikologi, Ekon. Sastra, Arsit. Tek. Sipil*, 2013.

[8] N. Y. A. Faradhillah, R. P. Kusumawardani, and I. Hafidz, "Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin," *Pros. Semin. Nas. Sist. Inf. Indones. 2016*, 2016.

[9] P. Pascasarjana and U. Udayana, "Text mining dengan metode naïve bayes classifier dan support vector machines untuk sentiment analysis," *Univ. Stuttgart*, 2011.

[10] A. Fathan Hidayatullah, M. Rifqi Ma, and arif Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta Jl Ringroad Barat, "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," *Semin. Nas. Apl. Teknol. Inf. Agustus*, 2016.

[11] S. Budi, "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means," *Techno.Com*, 2017.

[12] U. Pauziah and D. Michael Sonny, "Kajian Algoritma Naive Bayes Dalam pemilihan Penerimaan Beasiswa Tingkat SMA," *Semin. Nas. Teknol. Inf. dan Multimed.*, 2014.

[13] D. Dwinavinta, C. Nugraha, M. Fahmi, Z. Naimah, and N. Setiani, "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means," *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, 2014.

[14] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. Sains Dan Seni Its*, 2015.

[15] andi nurul Hidayat, "Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes," *J. Elektron. Sistim Inf. Dan Komput.*, 2015.

[16] A. S. Nugroho, "Pengantar Support Vector Machine ∗," *J. Data Mining, Jakarta*, 2007.

[17] M. Abdi, D. Herumurti, and I. Kuswardayan, "Analisis Perbandingan Kecerdasan Buatan pada Computer Player dalam Mengambil Keputusan pada Game Battle RPG," *JUTI J. Ilm. Teknol. Inf.*, 2017.

[18] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine – Teori dan Aplikasinya dalam Bioinformatika," *Kuliah Umum IlmuKomputer.Com*, 2003.

[19] I. Mathilda Yulietha and S. Al Faraby, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4740–4750, 2017.

[20] S. Chormunge and S. Jena, "Efficient feature subset selection algorithm for high dimensional data," *Int. J. Electr. Comput. Eng.*, vol. 6, no. 4, pp. 1880–1888, 2016.

[21] L. Dini Utami and R. S. Wahono, "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk

Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes," *J. Intell. Syst.*, vol. 1, no. 2, pp. 120–126, 2015.

[22] N. Krisandi, B. Prihandono, and Helmi, "Algoritma K - Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu," *Bul. Ilm. Math.Stat.dan Ter.*, 2013.

[23] N. Hermaduanti and S. Kusumadewi, "Sistem Pendukung Keputusan Berbasis Sms Untuk Menentukan Status Gizi Dengan Metode K- Nearest Neighbor," *Semin. Nas. Apl. Teknol. Inf. ISSN 1907-5022*, 2008.

[24] C. Darujati, "Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naïve Bayes Dengan K-Nearest Neighbor Abstrak," *Univ. Stuttgart*, 2010.