

Sentiment Analysis Pandangan Masyarakat Terhadap Tarif Tol Trans-Jawa Menggunakan *Support Vector Machine* dan *Particle Swarm Optimization*

Afri Yudha^{1*}, Yosep Nuryaman², Iqbalullava Nuddin³, Aulia Andhikawati⁴, Ernawati⁵, Nandang Suwela⁶

^{1,2,3,4,6} Teknik Informatika, Institut Bisnis Muhammadiyah Bekasi

⁵ Ilmu Komputer, STMIK Nusa Mandiri

⁶ Teknik Informatika, Institut Bisnis Muhammadiyah Bekasi

*Email: ibnugazali@gmail.com

Abstrak

Keywords:

sentiment analysis;
Trans-Jawa; svm; pso;
vb.net

Jalan Tol Trans-Jawa adalah jalan tol yang menghubungkan kota-kota di pulau Jawa dimana mempunyai jarak ± 1.000 kilometer, Jalan tol ini menghubungkan dua kota terbesar di Indonesia Jakarta dan Surabaya, banyak opini masyarakat baik dunia nyata maupun di media sosial mengenai tol trans jawa terlebih tarifnya yang hampir mencapai 1 juta rupiah, pada penulisan jurnal ilmiah ini dilakukan sentiment analysis terkait tarif tol trans jawa, data yang digunakan diambil dari jaringan media sosial twiter pada periode bulan juni dan juli, didapatkan data mentah 543 data, setelah dilakukan cleansing dan Preprocessing, data yang dapat diolah 108 data dengan rincian 54 negatif dan 54 positif. Berdasarkan pemodelan menggunakan algoritma support vector machine(SVM) didapat bahwa pandangan masyarakat terhadap jalan tol trans-jawa terlihat lebih banyak yang positif sebesar 52% dengan akurasi 68,36% dengan kurva roc 0,776 dan di optamasi dengan Particle Swarm Optimization (PSO) sehingga mengalami kenaikan akurasi menjadi 75% dan kurva roc 0,802, lalu dari hasil yang didapatkan dibuatlah kamus data dari bobot yang ada menggunakan Visual Basic .NET (VB.Net).

1. PENDAHULUAN

Trans Jawa adalah sebuah proyek yang berencana untuk menyambungkan pulau jawa dengan tol, yang sudah ada sejak 20 tahun yang lalu, yang nanti akan menghubungkan merak di barat Jawa dengan Banyuwangi di timur Jawa [17].

Menurut PP No. 15 Tahun 2005 tentang jalan tol, dijelaskan bahwa definisi jalan tol adalah jalan umum yang merupakan bagian sistem jaringan jalan dan sebagai jalan nasional yang penggunaannya diwajibkan membayar tol.

Tol merupakan sejumlah uang tertentu yang dibayarkan untuk penggunaan jalan tol. Besarnya tarif tol berbeda untuk setiap golongan kendaraan dan ketentuan tersebut telah ditetapkan berdasarkan keputusan presiden.

Sentiment analysis dan opinion mining adalah bidang studi yang menganalisis pendapat seseorang, sentiment seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang ke dalam bahasa tertulis. Sentiment Analysis telah banyak menerima perhatian

semenjak penelitian Pang, Turney, Goldberg dan Zhu [3]. Teknik sentimen dapat mendukung banyak keputusan dalam banyak skenario.

Sentiment analysis dilakukan pada tarif tol trans jawa untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah, apakah cenderung berpandangan negatif atau positif, terlebih banyaknya nada sumbang mengenai tol trans jawa terlebih tarifnya, diantara lain tol yang dibangun dengan hutang, dan tarif yang terlalu tinggi sehingga tidak memperhatikan rakyat.

Dari beberapa teknik sentimen anlysis, yang paling sering digunakan untuk klasifikasi data adalah Support Vector Machines (SVM). SVM merupakan metode supervised learning yang menganalisa data dan mengenali pola-pola yang digunakan untuk klasifikasi [2]. Support Vector Machines (SVM) adalah kasus khusus dari keluarga algoritma yang disebut sebagai regularized metode klasifikasi linier dan metode yang kuat untuk meminimalisasi resiko [11]. SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda [4]. Namun Support Vector Machine memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai [2]. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi [12]. Dalam masalah aplikasi tertentu, tidak semua fitur ini sama pentingnya. Kinerja yang lebih baik dapat dicapai dengan membuang beberapa fitur. Dengan demikian, dapat dihilangkannya data yang noise, data yang tidak relevan dan berlebihan [12].

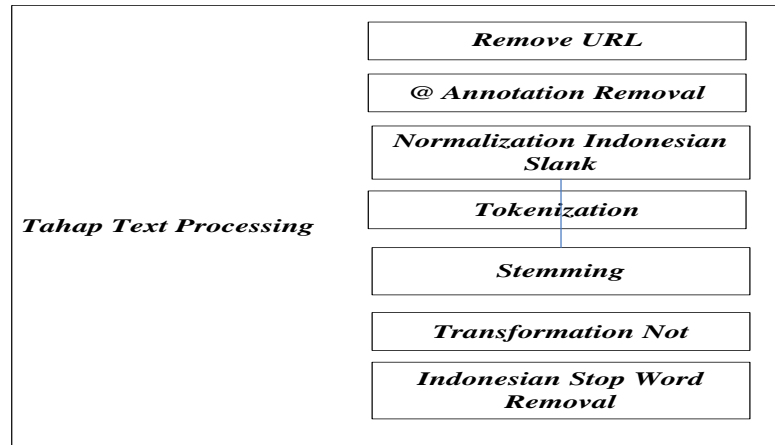
2. METODE

Metode penelitian yang digunakan dalam penelitian ini adalah dengan menggunakan metode eksperimen. Penelitian ini bertujuan mengevaluasi algoritma klasifikasi *data mining* pada sentiment analysis tarif tol trans Jawa.

Menurut Hermawati [15] : “Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya”.

Menurut G. Vinodhini dan RM. Chandrasekaran [10] bahwa “sentiment analysis adalah pengolahan bahasa alami untuk mengetahui pendapat masyarakat tentang produk atau topik tertentu. Analisis sentimen, yang juga disebut opini mining, bagian dalam pembangunan sebuah sistem untuk mengumpulkan dan menganalisa opini publik tentang produk yang dibuat dalam sebuah postingan blog, komentar, ulasan dan tweet”. Ada beberapa tantangan dalam analisis sentimen. yang pertama adalah kata pendapat yang dianggap positif dalam satu situasi dapat dianggap negatif dalam situasi lain. Tantangan kedua adalah karena orang cenderung menggunakan kata positif ketimbang kata negatif untuk sentimen negatif. Tantangan yang lainnya seperti penggunaan bahasa, orang Indonesia cenderung mencampur bahasa yang mereka gunakan [16].

Penelitian tentang sentiment analysis yang menggunakan dataset dari jejaring sosial twitter telah dilakukan sebelumnya oleh [9], dan ini masuk dalam kategori text mining, sedangkan text mining juga dikenal sebagai data mining teks [7] atau penemuan pengetahuan dari database tekstual [13]. Sesuai dengan buku The Text Mining Handbook [5], text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Dari uraian tersebut maka diperlukan adanya penanganan yang ekstra pada saat tahap pre-processing atau tahap persiapan data ditunjukkan gambar 1.



Sumber [16]

Gambar 1 : Tahapan Text Processing

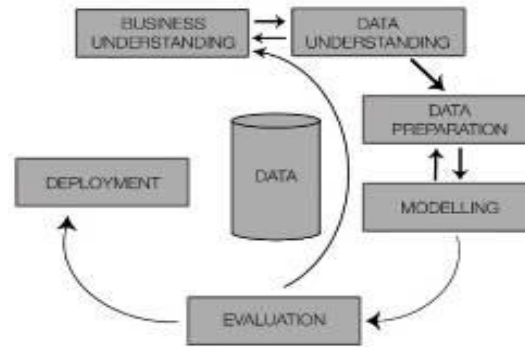
Penjelasan dari gambar diatas adalah :

- a. *Remove URL* : Seringnya muncul sebuah url dari data twitter membuat data tidak efektif dan tidak memiliki arti. Untuk itu perlu adanya penghapusan url tersebut.
- b. *@Annotation Removal* : Bertujuan menghapus nama user biasanya diawali dengan simbol "@" karena dalam suatu kasus dapat dianggap tidak penting maka perlu dihilangkan.
- c. *Normalization Indonesian Slank* : Pada proses ini untuk mengkonversi kata-kata yang tidak baku (bahasa gaul) menjadi kata baku, contoh: ngeyel artinya keras kepala.
- d. *Tokenization* : Dalam tahap tokenize ini dilakukan pemotongan string input berdasarkan tiap kata yang menyusunnya atau dengan kata lain tokenize memecah sekumpulan karakter dalam suatu teks kedalam satuan kata dengan menghilangkan simbol dan karakter yang bukan merupakan huruf.
- e. *Stemming* : Pada tahapan ini dicari kata dasar (root) dari tiap kata hasil filtering yang terdapat dalam suatu

dokumen dengan menggunakan aturan-aturan tertentu.

- f. *Transformation Not Negative* : sebenarnya prosesnya tidaklah menghapus kata melainkan diambil untuk menilai bahwa kalimat yang diproses mengandung kalimat negative contoh : tidak, jangan , bukan.
- g. *Indonesian Stop Word Removal* : diproses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan di anggap tidak penting, seperti kata penghubung. contoh : dengan, dan, juga.

Dalam mendesain metode penelitian eksperimen ini peneliti menggunakan metode penelitian standar yang digunakan pada *data mining* yaitu *Cross Industry Standart Process for Data Mining* (CRISP-DM) ditujukan pada gambar 2, yaitu terdiri dari 6 fase dengan langkah-langkahnya adalah *Business Understanding, Data Uderstanding, Data Preparation, Modelling, Evaluation* dan *Deployment* [7].



Sumber [16].

Gambar 2 : CRISP-DM

2.1 Pemahaman Bisnis (*Business Understanding*)

Tahap pertama dari CRISP-DM adalah *Business Understanding*. Pada tahap ini peneliti membuat satu pertanyaan. Pertanyaannya adalah apa tujuan dari penelitian ini? Jawabannya adalah untuk Mengetahui pandangan masyarakat terhadap tarif tol trans-jawa yang sudah ditetapkan pemerintah.

2.2. Pemahaman Data (*Data Understanding Phase*)

Tahap kedua dari CRISP-DM adalah *Data Understanding*. Pada tahap ini peneliti bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki. Data tersebut juga harus dapat diverifikasi kebenarannya. Data yang digunakan dalam penelitian ini adalah data komentar yang terdapat di jejaring sosial media twitter mengenai tarif tol trans Jawa.

2.3 Persiapan Data (*Data Preparation*)

Jumlah data awal diperoleh dari pengumpulan data yaitu sebanyak 543 record. Persiapan juga dapat berupa konversi nilai-nilai redundan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Dari semua data yang sudah terkumpul, tidak semua data dapat digunakan. Oleh karena itu data tersebut akan melewati beberapa tahap persiapan awal data (*preparation data*) antara lain menggunakan transformasi data dengan cara melakukan pre-processing, lalu melihat hasilnya.

2.4 Pemodelan (*Modelling*)

Dalam tahap Modelling ini akan dilakukan teknik pengklasifikasian data yang paling akurat. Pada penelitian ini akan digunakan algoritma *Support Vector Machine* (SVM) yang ada pada framework Rapid Miner 7.3 dan di optimasi dengan algoritma *Particle Swarm Optimization*.

Support vector machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, SVM pertama kali dikenalkan oleh Vapnik, Noser dan Guyon pada tahun 1992 dan sejak itu SVM mulai tumbuh berkembang pesat. SVM merupakan salah satu teknik yang baru bila dibandingkan dengan teknik yang lain, namun memiliki performa yang lebih baik di berbagai aplikasi seperti klasifikasi teks, dan pengenalan tulisan tangan.

Particle Swarm Optimization (PSO) Menurut Basari [2] adalah merupakan teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasikan beberapa parameter, diperkenalkan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995, merupakan algoritma optimasi yang meniru proses yang terjadi dalam kehidupan populasi burung (flock of bird) dan ikan (school of fish) dalam bertahan hidup.

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan

menggunakan bahasa java sehingga dapat bekerja di semua sistim operasi [1].

2.5 Evaluasi (*Evaluation*)

Model yang terbentuk akan diuji menggunakan *confusion matrix* untuk mengetahui tingkat akurasi. Akurasi akan dihitung dari seluruh hasil prediksi yang benar (baik prediksi positif dan negatif) dibandingkan dengan seluruh data testing. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Pengujian juga diukur dengan menggunakan ROC Curve. ROC Curve akan menggambarkan kelas positif dalam bentuk kurva. Pengujian dilakukan dengan menghitung nilai AUC (*Area Under Curve*), semakin tinggi nilai AUC dan ROC Curve, maka semakin baik pula model klasifikasi yang terbentuk.

Confusion Matrix [6] adalah alat (tool) visualisasi yang biasa digunakan supervised learning. Tiap kolom pada matriks adalah contoh dalam kelas prediksi, sedangkan setiap baris mewakili kejadian di kelas yang sebenarnya. Satu keuntungan dari *Confusion Matrix* adalah mudah untuk mengetahui jika data ada diantara dua kelas (mislabeling). *Confusion Matrix* berisi informasi tentang aktual (actual) dan prediksi (predicted) pada sistem klasifikasi

2.6 Penerapan (*Deployment*)

Tahapan ini adalah tahap terakhir dari CRISP-DM, yaitu hasil dari seluruh tahapan yang sebelumnya digunakan secara nyata. Maknanya adalah melakukan sesuatu berdasarkan pengetahuan yang di dapatkan dari kegiatan mining terhadap data, dan mengaplikasikannya kedalam suatu bentuk aplikasi yang peneliti akan gunakan adalah pemrograman Visual Basic.Net (.NET).

Microsoft Visual Studio adalah sebuah Integrated Development Environment buatan Microsoft Corporation. Microsoft Visual Studio dapat digunakan untuk mengembangkan aplikasi dalam native code (dalam bentuk bahasa mesin yang berjalan diatas Windows) ataupun managed code (dalam bentuk Microsoft Intermediate Language di atas .NET Framework), Microsoft Visual Studio didukung bahasa pemrograman yang berbeda. Adapun bahasa pemrograman yang didukung oleh Visual

Basic Studio adalah bahasa pemrograman C++, Visual Basic, Visual C#. Visual Studio juga dapat mendukung bahasa pemrograman lain seperti M, python dan ruby yang semuanya itu terdapat pada pack extra yang terpisah dari visual studio [14].

3. HASIL DAN PEMBAHASAN

3.1 Fase Pemahaman Bisnis (*Business Understanding Phase*)

Pada tahapan business understanding, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan menggali informasi mengenai tol trans-Jawa dan tarifnya. Motivasi pada fase ini yaitu isi komentar warganet dari berita yang disajikan, biasanya dalam bentuk teks pada media digital yang dikelompokkan berdasarkan isi pembahasan dari masing-masing kategori berita.

Pada tahap ini juga dilakukan pemahaman untuk mencari metode klasifikasi yang terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma yang digunakan dan untuk meningkatkan performa dari metode klasifikasi.

3.2 Fase Pemahaman Data (*Data Understanding Phase*)

Pada tahap *data understanding*, dilakukan proses pengambilan dataset (contoh gambar 3) dari jejaring media sosial twitter dari periode bulan Juli sampai dengan Juni awal 2019, didapatkan lah total data primer 543 record, lalu setelah dilakukan *pre-processing*, maka data yang dapat diolah menjadi 117 record dengan 63 record data negatif dan 54 data positif, dan diolah dengan data seimbang 54 negatif dan 54 positif.

3.3 Fase Pengolahan Data (*Data Preparation Phase*)

Tahap data preparation merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Dalam text mining tahapan awal yang akan dilakukan adalah tahap preprocessing. Berikut merupakan tahapan yang dilakukan dalam preprocessing. Contoh Tabel 1-7.



Sumber : media sosial Twiter (2019)

Gambar 3 : Contoh Tweet Negatif dan Tweet Positif

Tabel 1: *Remove URL*

Text Sebelum pre-processing	Text Sesudah Pre-processing
@irvaneenk Teman gua ceceb juga, ngeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia ogah. Katanya tarif tol trans jawa mahal, dompet ga cukup https://t.co/nPAuSJwHrl	@irvaneenk teman gua ceceb juga, ngeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia ogah. katanya tarif tol trans jawa mahal, dompet ga cukup

Tabel 2: *@Annotation Removal*

Text Sebelum pre-processing	Text Sesudah Pre-processing
@irvaneenk teman gua ceceb juga, ngeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia ogah. katanya tarif tol trans jawa mahal, dompet ga cukup	teman gua ceceb juga, ngeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia ogah. katanya tarif tol trans jawa mahal, dompet ga cukup

Tabel 3 : *Normalization Indonesian Slank*

Text Sebelum pre-processing	Text Sesudah Pre-processing
teman gua ceceb juga, ngeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia ogah. katanya tarif tol trans jawa mahal, dompet ga cukup	teman saya ceceb juga, mengeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia tidak bersedia. katanya tarif tol trans jawa mahal, dompet tidak cukup

Tabel 4 : *Tokenization*

Text Sebelum pre-processing	Text Sesudah Pre-processing
teman saya ceceb juga, mengeluh ada kena macet di jalan pantura, suruh lewat jalan tol, dia tidak bersedia. katanya tarif tol trans jawa mahal, dompet tidak cukup	teman saya ceceb juga mengeluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak bersedia katanya tarif tol trans jawa mahal dompet tidak cukup

Tabel 5 : *Stemming*

Text Sebelum pre-processing	Text Sesudah Pre-processing
teman saya ceceb juga mengeluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak bersedia katanya tarif tol trans jawa mahal dompet tidak cukup	teman saya ceceb juga keluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak sedia kata tarif tol trans jawa mahal dompet tidak cukup

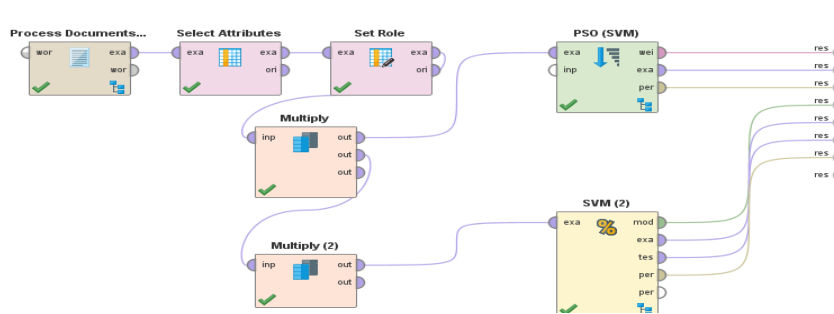
Tabel 6 : *Transformation Not Negative*

Text Sebelum pre-processing	Text Sesudah Pre-processing
teman saya ceceb juga keluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak sedia kata tarif tol trans jawa mahal dompet tidak cukup	teman saya ceceb juga keluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak_sedia kata tarif tol trans jawa mahal dompet tidak_cukup

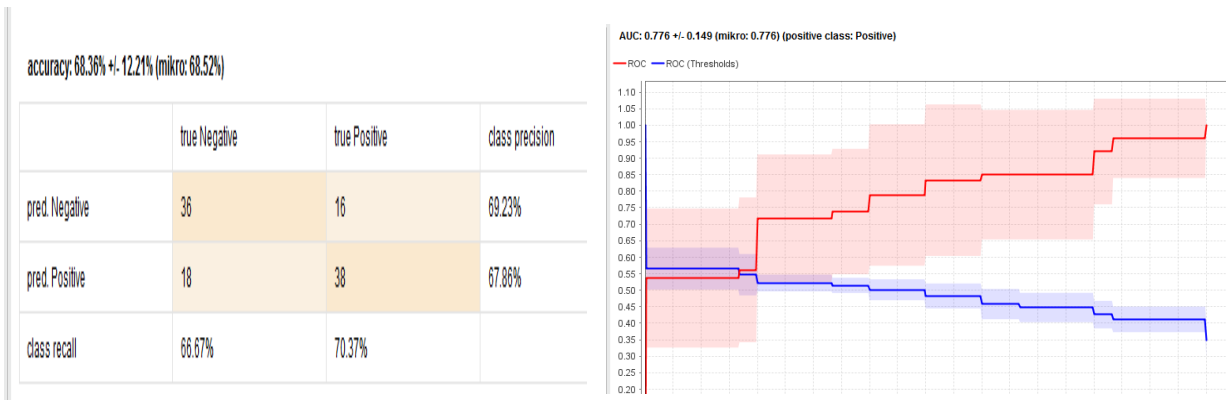
Tabel 7 : Indonesian Stop Word Removal

Text Sebelum pre-processing	Text Sesudah Pre-processing
teman saya ceceb juga keluh ada kena macet di jalan pantura suruh lewat jalan tol dia tidak_sedia kata tarif tol trans_jawa mahal dompet tidak_cukup	teman ceceb keluh kena macet jalan pantura suruh jalan tol tidak_sedia tarif tol trans jawa mahal dompet tidak_cukup

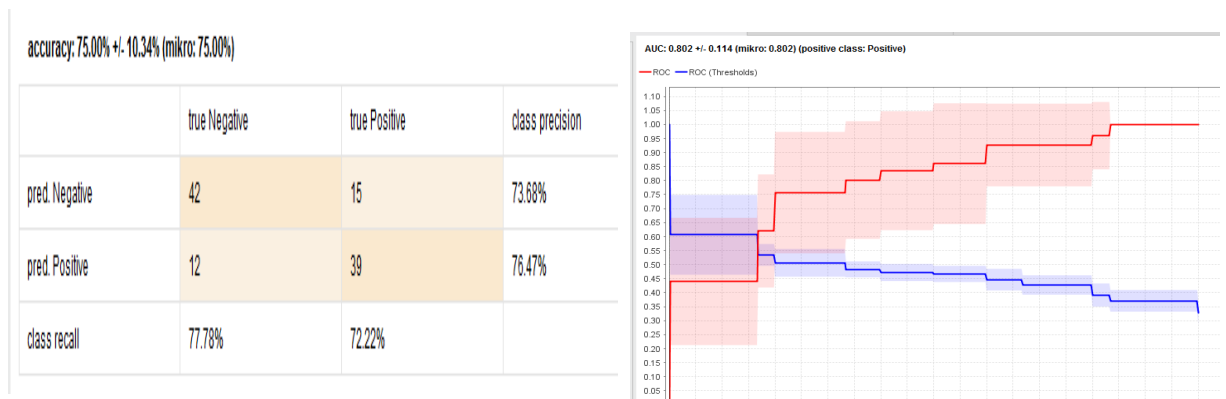
Sumber : Hasil Penelitian (2019)



Gambar 4. Pemodelan pada Rapid Miner



Gambar 5. Akurasi & Kurva ROC SVM



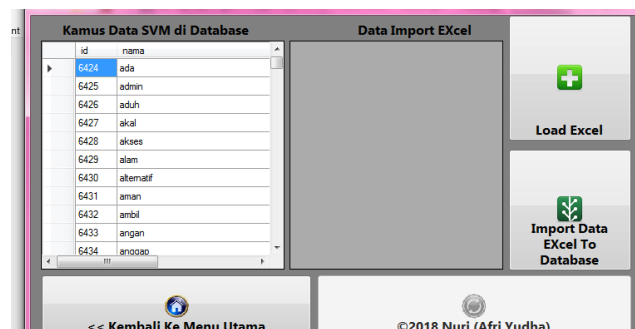
Gambar 6. Akurasi & Kurva ROC PSO-SVM

Attribute	Weight
ada	-0.011
admin	-0.024
aduh	0.013
akal	-0.010
akses	0.024
alam	-0.024
alternatif	0.016
aman	0.029
ambil	0.009
angan	-0.024
anggap	-0.017
anggaran	-0.017
anggur	-0.018
annin	-0.024

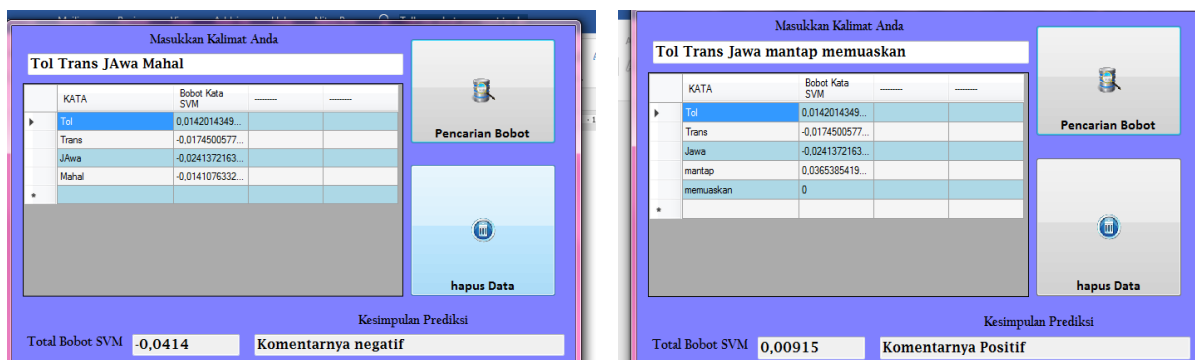
Gambar 7. Hasil Bobot Data



Gambar 8. Tampilan Aplikasi



Gambar 9. Halaman Upload Bobot Data



Gambar 10. Hasil Pengujian Teks Bebas

3.4 Fase Pemodelan (*Modelling*)

Pada fase ini secara langsung melibatkan teknik data mining. Tool yang digunakan pada fase pemodelan ini adalah Rapidminer versi 7.3.0. Adapun hasil dalam pengujian model yang dilakukan adalah mengklasifikasi komentar positif dan komentar negative menggunakan algoritma *Support Vector Machine* (SVM) dan dioptimasi dengan *Particle Swarm Optimization* (PSO) Adapun desain Rapidminer yang digunakan seperti ditunjukkan gambar 4 diatas.

3.5 Fase Evaluasi (*Evaluation Phase*)

Tahapan evaluasi bertujuan untuk mengetahui nilai akurasi dan confusion matrix dari model yang telah berhasil dibuat pada langkah sebelumnya, seperti gambar 5 diatas yang menunjukkan akurasi dan kurva ROC dari algoritma SVM setelah dilakukan pemodelan.

Mendapatkan akurasi 68,36%, Data komentar negatif yang sesuai prediksi yaitu 36 data. Data komentar negatif dalam prediksi positif yaitu 16 data. Data komentar positif yang termasuk kedalam prediksi negatif yaitu 18 data, data positive yang sesuai prediksi yaitu 38 data.

Kurva ROC SVM dengan nilai AUC (Area Under Curve) yang dihasilkan sebesar 0.776 dimana diagnosa hasilnya adalah *fair classification*.

Dioptimasi dengan algoritma PSO menghasilkan akurasi dan kurva ROC yang ditunjukkan seperti gambar 6 diatas.

Mendapatkan akurasi 75.00%, Data komentar negatif yang sesuai prediksi yaitu 42 data. Data komentar negatif dalam prediksi positif yaitu 15 data. Data komentar positif yang termasuk kedalam prediksi negatif yaitu 12 data, data positive yang sesuai prediksi yaitu 39 data.

Kurva ROC SVM dengan nilai AUC (Area Under Curve) yang dihasilkan dari sebesar 0.802 dimana diagnosa hasilnya adalah *good classification*.

Setelah dilakukan perhitungan akhir dengan algoritma SVM didapatkan 56 data positif dan 52 data negatif, dengan prosentase positif 52% dan prosentase positif 48%

dengan akurasi 68,36% dan mencapai 75% setelah dioptimasi dengan PSO.

3.6 Fase Deployment (*Deployment Phase*)

Pada fase deployment akan dilakukan perancangan aplikasi dengan menggunakan bobot kata komentar mengenai tarif tol trans jawa yang telah dilakukan olah data di aplikasi rapid miner, dimana bobot kata ini akan di jadikan kamus data, yang nanti akan digunakan untuk menentukan komentar negatif dan positif dari kalimat yang diketikkan dalam deployment.

Bobot data ditunjukkan diatas pada gambar 7 yang didapatkan setelah dilakukan nya pemodelan algoritma.

Tampilan aplikasi kamus data ditunjukkan diatas pada gambar 8.

Tampilan upload bobot data ditunjukkan diatas pada gambar 9.

Tampilan percobaan aplikasi dari bobot data yang telah diolah dengan pemodelan algoritma, ditunjukkan pada diatas pada gambar 10.

4. KESIMPULAN

Dari hasil pembahasan yang penulis lakukan, maka didapatkan hasil positif 52% dan pandangan negatif 48% terhadap tarif tol yang berlaku di trans Jawa, dengan akurasi 68,36% menggunakan algoritma SVM dan mencapai 75% setelah dioptimasi dengan PSO. Dari hasil yang didapatkan dibuat kamus data dan dapat dilakukan pengujian terhadap data yang lainnya.

UCAPAN TERIMA KASIH

Alhamdulillah kami bersyukur kepada rektor Institut Bisnis Muhamaddiyah Bekasi (IBMB) bapak H.Jaenudin S.Ag.,M.PD yang telah memberikan support atas penelitian yang telah kami lakukan.

REFERENSI

Jurnal, Bulletin, dan Majalah Ilmiah

- [1]. Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner. Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1

- and 2. <https://doi.org/10.1007/s13398-014-0173-7.2>.
- [2]. 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.
- [3]. B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found Trends Inf Retr*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.
- [4]. Chou, j.-s., cheng, m.-y., wu, y.-w., & pham, a.-d. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert systems with applications*, 41(8), 3955–3964.
- [5]. Feldman, R & Sanger, J. (2007) *The Text Mining Handbook-Advanced Approaches in Analyzing Unstructured Data*, USA: New York.
- [6]. Gerunescu, florin (2011). *Data mining: concept, model, and techniques*. Verlag scoring berlin heidelberg: springer.
- [7]. Hearst, M. A. (1997) *Text data mining: Issues, techniques, and the relationship to information access*. Presentation notes for UW/MS workshop on data mining, July 1997.
- [8]. Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data*.
- [9]. Parikh, R., & Movassate, M. (2009). *Sentimen Analysis of User Generated Twitter Updates using Various Classification*.
- [10]. Vinodhini.G, Chandrasekaran.RM.(2012) *Sentiment Analysis and Opinion Mining: A Survey*. *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Vol 2.
- [11]. Weiss, S. M., Indurkha , Nitin dan Zhang, Tong (2010). *Fundamentals of predictive text mining*. London: Springer-Verlag, 2010.
- [12]. Zhao, M., Fu, C., Ji, L., Tang, K., dan Zhou, M. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, 38(5), 5197–5204 (2011).
- [13]. Feldman, R & Dagan, I. (1995) *Knowledge discovery in textual databases (KDT)*. Dalam *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, Montreal, Canada, August 20-21, AAAI Press, 112-117.
- [14]. Edy Winarno ST, M.Eng, Ali Zaki, SmitDev Community (2010:8), dalam buku “*Web Programming dengan Visual Basic 2010*”.
- [15]. Hermawati, F. A., 2013. *Data Mining*. Yogyakarta: Andi.
- [16]. Yudha.Afri (2018). *Analisis Sentiment Untuk Menilai Kepuasan Masyarakat Terhadap Kinerja Puan Maharani Di Kabinet Kerja Menggunakan Support Vector Machine Dan Naive Bayes*, STMIK Nusa Mandiri.
- [17]. <https://tirto.id/ambisi-menyambungkan-jawa-dengan-tol-bUXX>