



Research Article

Open Access (CC-BY-SA)

Analisis sentimen masyarakat terhadap informasi penerapan PPN atas kegiatan renovasi dan membangun rumah sendiri pada media sosial youtube dengan metode svm dan naive bayes

Muhammad Romadhona Kusuma^{a,1,*}; Windu Gata^{a,2}; Laela Kurniawati^{a,3}; Almay Faiz Rivan^{a,4}; Triadi Kurniawan^{a,5}

^a Universitas Nusa Mandiri, Jl. Kramat Raya No.18, RW.7, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

¹ m.romadhona.kusuma@gmail.com; ² windu@nusamandiri.ac.id; ³ laela@nusamandiri.ac.id; ⁴ almayfaizrivan@gmail.com;

⁵ triadikurniawan560@gmail.com

* Corresponding author

Artikel Histori: Diterima 07/07/2022; Revisi 20/01/2023; Terbit 01/09/2023

Abstrak

Penelitian ini bertujuan untuk menganalisis tanggapan dari masyarakat terhadap informasi penerapan pajak pertambahan nilai atas kegiatan renovasi dan membangun rumah sendiri yang kini telah diumumkan oleh pemerintah melalui kementerian keuangan. Analisis ini dilakukan untuk mengetahui bagaimana persepsi publik terhadap informasi tersebut dengan berfokus pada tata gaya bahasa yang digunakan di dalam komentar serta maksud dari komentar-komentar tersebut. Karena tidak mudah untuk menganalisis sentimen yang ada dari berbagai komentar, hal ini karena komentar mengandung teks yang tidak terstruktur, terutama dalam teks Indonesia. Preprocessing teks adalah bagian penting untuk mendapatkan informasi dasar serta maksud yang terkandung di dalam komentar yang ada. Adapun penelitian yang dilakukan menggunakan pra-pemrosesan teks Indonesia dengan Textmining menggunakan rapid miner. Teknik pengumpulan data dilakukan melalui pengumpulan komentar di media sosial youtube kanal berita viva.co.id dari tanggal 10 april 2022 hingga 30 juni 2022 dengan jumlah 1650 komentar. Kemudian dilanjutkan dengan ekstraksi data menggunakan metode klasifikasi Naive Bayes serta support vector machine yang dioptimasi dengan particle swarm optimization. Kesimpulan dari penelitian ini menunjukkan bahwa respon masyarakat mengenai informasi penerapan pajak pertambahan nilai atas kegiatan renovasi dan membangun rumah sendiri berbeda-beda. Ada yang tidak setuju dan ada juga yang setuju serta ada pula yang tidak peduli. Saran dari penelitian ini, pemerintah melalui kementerian keuangan RI hendaknya memberikan sosialisasi secara luas kepada masyarakat, sehingga masyarakat tidak mengalami kebingungan dengan informasi tersebut sehingga masyarakat tidak akan asal berkomentar dengan pendapatnya masing-masing.

Kata Kunci: Kata Pertama, Kata Kedua, Kata Ketiga, Kata Keempat, Kata Kelima.

Pendahuluan

Perkembangan teknologi yang terjadi saat ini memudahkan akses dan penerimaan informasi oleh masyarakat melalui berbagai jenis media, termasuk media sosial. Dengan adanya media sosial, semua orang bisa dengan mudah berbagi pendapat dan juga informasi serta berinteraksi dengan orang lain secara online tanpa dibatasi oleh waktu maupun tempat. Media sosial adalah salah satu jenis media yang sangat populer di era digital saat ini. Selain mempermudah komunikasi antar individu, media sosial mampu memberikan beberapa kemudahan bagi masyarakat dalam berbagi informasi, menyebarkan pendapat, dan berinteraksi dengan orang lain. Dengan adanya media sosial, masyarakat dapat dengan mudah mengakses dan memperoleh informasi dari berbagai sumber. Namun, di sisi lain, media sosial juga memiliki beberapa risiko yang perlu diwaspadai, seperti penyebaran informasi yang tidak akurat atau hoax, serta potensi masalah privasi. Itulah mengapa penting bagi masyarakat untuk memahami bagaimana menggunakan media sosial dengan bijak dan menyaring informasi yang diperoleh dengan cermat.

Saat ini terdapat banyak jenis media sosial yang tersedia dan memiliki fitur yang beragam. Facebook, Twitter, dan YouTube adalah contoh media sosial yang populer dan memiliki jumlah pengguna yang besar di seluruh dunia. Media sosial YouTube terutama dalam kemudahan penggunaannya dan kesederhanaannya, sebagai kanal media bisa digunakan sebagai penyampaian informasi yang memungkinkan kolaborasi dan berbagi sumber daya berupa gambar suara dan video [1] Sebuah laporan

yang dipublikasikan pada tahun 2022 menyatakan bahwa YouTube merupakan salah satu situs konten video terpopuler di dunia dan merupakan sumber lalu lintas web yang besar saat ini.

YouTube menyediakan fitur berbagi (share) dan berkomentar di dalam kanal video, interaksi sosial seperti komentar ini, memungkinkan pengguna untuk memberikan tanggapan atau pendapat mereka tentang video yang dilihat, Informasi komentar ini berguna dalam mempelajari tanggapan masyarakat terhadap suatu informasi yang ada di kanal video youtube [2].

Analisis sentimen adalah proses yang digunakan untuk mengidentifikasi perasaan atau opini dalam teks serta mengelompokkan dan mengklasifikasikannya menjadi tiga kategori: positif, negatif, atau netral [3][4]. Tujuannya adalah untuk mengetahui bagaimana orang memandang suatu topik tertentu. Analisis sentimen juga dikenal sebagai pertambangan sentimen, yaitu sebuah bidang studi untuk mempelajari pendapat, penilaian, sentimen serta emosi yang terkandung dalam teks terkait dengan entitas atau atribut tertentu [5]. Analisis sentimen dapat digunakan untuk mengevaluasi respon masyarakat terhadap suatu produk atau layanan, mengukur popularitas sebuah brand, atau untuk menilai sentimen masyarakat terhadap suatu peristiwa atau isu.

Naive Bayes merupakan teknik untuk memprediksi kemungkinan menggunakan data yang telah tersedia [6]. Ini merupakan metode yang cukup sederhana untuk menangani sejumlah besar data dengan akurasi yang tinggi [7]. Dalam mengklasifikasikan dokumen menggunakan metode Naive Bayes, beberapa tahap yang dilakukan antara lain seperti menentukan kemungkinan pada setiap kata yang muncul dalam proses klasifikasi dokumen, menghitung probabilitas pada setiap kata yang muncul pada kategori dokumen, dan mengidentifikasi kelas-kelas dokumen berdasarkan proses hitungan yang dilakukan pada tahap sebelumnya [8]. SVM adalah algoritma yang digunakan untuk mengklasifikasikan dan melakukan regresi pada masalah yang tidak linear dan linear. Keuntungan dari menggunakan SVM adalah dengan menggunakan kernel yang tepat, dapat menangani masalah pada dimensi yang tinggi [9], [10].

Penelitian yang sejenis tentang analisis sentimen yang menggunakan algoritma klasifikasi sebelumnya banyak juga peneliti yang telah melakukannya. seperti yang telah dilakukan oleh Sapulrohman et al dengan topik analisis sentimen kepuasan pengguna whatsapp, penelitian tersebut menggunakan beberapa algoritma seperti algoritma klasifikasi Naive Bayes dan Support Vector Machine. Langkah-langkah preprocessing teks seperti tokenization, cleaning, dan filtering digunakan. Hasil penelitian tersebut memberikan metode klasifikasi support vector machine dengan skor 77 n naive bayes 70,40%, namun penelitian ini kekurangan Stemmer yang berbahasa Indonesia serta data stop word yang ternyata masih terbatas sehingga perlu pengembangan lebih lanjut [11].

Algoritma Naive Bayes serta Support Vector Machine juga digunakan oleh Kurniawan et al dengan topik penelitian menganalisis sentimen terhadap tokoh politik dalam komentar-komentar yang ada pada media berita online, Penelitian-penelitian tersebut, tidak berbeda dengan penelitian lain yang telah dilakukan yaitu dilakukan pemrosesan seperti pre-processing teks dan klasifikasi teks menerapkan metode algoritma machine learning. Dari hasil penelitian sebelumnya yang telah dilakukan tersebut, kesimpulannya yaitu algoritma Naive Bayes yang dioptimalisasi dengan pso maupun svm yang dioptimalisasi menggunakan pso menghasilkan akurasi yang lebih tinggi dibandingkan naive bayes dan svm yg tanpa pso sehingga optimalisasi pso bisa menjadi solusi untuk klasifikasi analisis sentimen berbasis text [12].

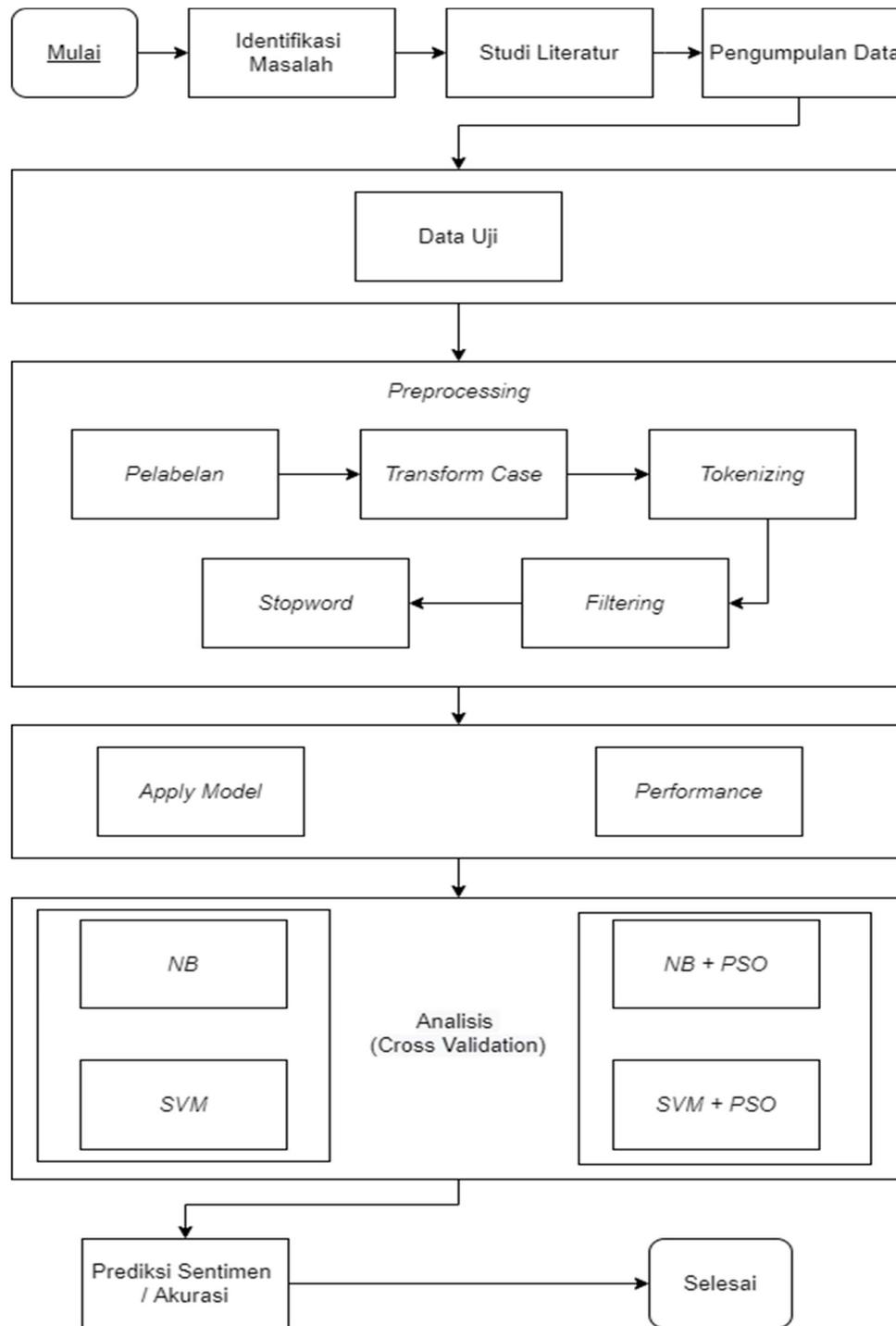
Berdasarkan penjelasan dari beberapa hasil penelitian sebelumnya telah dilakukan, maka penulis bermaksud untuk melakukan penelitian tentang analisis sentimen masyarakat terhadap informasi PPN atas kegiatan renovasi dan membangun rumah sendiri belakangan ini, penelitian dilakukan dengan menggunakan sumber data komentar yang ada di media sosial YouTube.

Beberapa Tujuan penelitian ini adalah:

- 1) Melihat Respon masyarakat pada media sosial youtube terkait munculnya informasi penerapan pajak pertambahan nilai (PPN) atas kegiatan renovasi dan membangun rumah sendiri.
- 2) Mendapatkan hasil dari pengujian data analisis sentimen dengan algoritma Support Vector Machine (SVM) dan Algoritma Naive Bayes tanpa Particle Swarm Optimization (PSO) maupun dengan PSO.
- 3) Mendapatkan hasil pengujian yang membandingkan performa algoritma SVM dengan algoritma Naive Bayes tanpa PSO atau dengan particle swarm optimization.

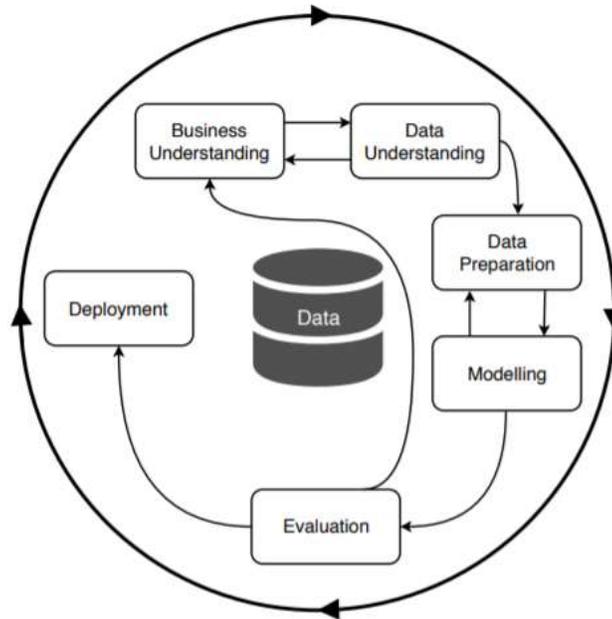
Metode Penelitian

Untuk menganalisa akurasi terbaik pada komentar penerapan PPN kegiatan membangun rumah sendiri, metode algoritma Naive Bayes dan Support Vector Machine digunakan. Langkah-langkah untuk mencapai akurasi yang terbaik ditunjukkan pada Gambar 1 di bawah ini:



Gambar 1. Diagram Alir Penelitian

Pada Penelitian berikut ini termasuk dalam Machine Learning, khususnya di bagian penambahan kata atau disebut Text Mining. Kemudian Sumber informasi yang menjadi dataset penelitian ini berasal dari komentar masyarakat terhadap informasi penerapan pajak pertambahan nilai (ppn) atas kegiatan renovasi dan membangun rumah sendiri pada media sosial youtube. Cross-Industry Standard Process for Data Mining (CRISP-DM) digunakan sebagai metode penelitian. Metode ini terdiri dari beberapa langkah seperti yang ditunjukkan pada Gambar 2 [13].



Gambar 2. Metode CRISP-DM

a. Business Understanding

Pada tahap business understanding dilakukan sebagai pemahaman terhadap objek penelitian yang diwujudkan dengan cara mencoba menggali informasi dari media sosial YouTube dengan objeknya yaitu komentar-komentar tanggapan mengenai informasi penerapan pajak pertambahan nilai (ppn) atas kegiatan renovasi dan membangun rumah sendiri. Motivasi dalam fase ini adalah komentar di kanal youtube yang telah disediakan berupa teks komentar yang kemudian dikelompokkan, berdasarkan detail yang ada di dalam komentar tersebut. Media online youtube lebih dikenal sebagai media untuk menonton video, tetapi juga bisa dimanfaatkan untuk mengetahui masalah-masalah yang terjadi serta untuk melihat sentimen masyarakat terhadap suatu informasi. Analisis sentimen (sentiment analysis) adalah teknik yang digunakan untuk mengekstrak opini atau perasaan dari teks. Dalam analisis sentimen, teknik klasifikasi digunakan untuk mengklasifikasikan teks sebagai positif, negatif, atau netral. Tujuan dari Analisis sentimen ini melibatkan penemuan metode klasifikasi yang memungkinkan untuk membedakan antara komentar positif dan negatif dengan tingkat akurasi yang tinggi. Proses ini juga dapat membantu dalam pengolahan data dengan menggunakan algoritma yang telah ada seperti Naive Bayes dan SVM. Pemilihan fitur yang sesuai dapat juga digunakan untuk meningkatkan kinerja metode klasifikasi.

b. Data Understanding

Pada tahap Data Understanding akan dilakukan pengumpulan data-data berdasarkan dengan atribut yang dibutuhkan. Data yang diperoleh ini berasal dari media online youtube kanal viva news judul berita video segini tarif pajak bangun & renovasi rumah sendiri. Data dikumpulkan dari 10 April 2022, hingga 10 juni 2022. Data primer yang diperoleh adalah 1650 data komentar. Kemudian setelah data telah terkumpul akan dilakukan pembersihan dan hasilnya diperoleh jumlah data sebanyak 1494 data komentar yang hanya berbahasa Indonesia.

c. Data Preparation

Tahap Data Preparation merupakan langkah dalam proses persiapan data yang mempunyai maksud supaya data menjadi bersih dan siap pakai, Pada tahap pertama text mining akan dilakukan langkah text preprocessing,

Pada tahap ini akan dilakukan proses pre-processing teks pada data komentar, seperti : Labeling, Transform Case dan Tokenizing, Stopword serta Filtering, kemudian pada pembahasan langkah-langkah akan dijelaskan secara lebih lanjut pada tahapan berikut.

d. *Modeling*

Modeling adalah tahap untuk menerapkan teknik pemodelan penambangan data yang telah dipilih dan diimplementasikan untuk digunakan dalam analisis sentimen. Dalam penelitian ini, alat RapidMiner versi 9.4 dipergunakan sebagai alat bantu untuk melakukan proses pemodelan sesuai dengan teknik yang telah di tentukan. Dua jenis algoritma klasifikasi seperti Naive Bayes (NB) serta Support Vector Machine (SVM) digunakan sebagai model dalam penelitian ini. Algoritma SVM dioptimalkan dengan menggunakan Particle Swarm Optimization (PSO) akan digunakan untuk memperbaiki kinerja algoritma. Setiap model digunakan untuk mengklasifikasikan komentar sebagai positif atau negatif, dan diuji untuk menentukan nilai akurasi terbaik dari masing-masing algoritma.

Tabel 1. Confusion Matrix

Fact	Prediction	
	Positif (P)	Negatif (N)
True (T)	TP	TN
False (F)	FP	FN

e. *Evaluation*

Tahap evaluation bermaksud untuk menguji hasil manfaat dari model yang telah dibuat pada tahap pemodelan sebelumnya. Dalam penelitian ini, 10-fold cross validation digunakan sebagai metode evaluasi. Proses validasi meliputi pengujian data dengan menggunakan model algoritma yang sudah ditentukan. Kemudian dilakukan proses silang dengan membagi data menjadi 10 bagian dan setiap bagian dari 10 bagian yang diuji menggunakan model algoritma. Ini dilakukan sebanyak 10 kali untuk memastikan hasil yang valid dan dapat diandalkan dari model yang dibuat.

Hasil dari pengujian adalah kelas prediksi yang dihasilkan dari model yang dibuat, kemudian dibandingkan berdasarkan kelas aktual yang ada pada data uji. Confusion matrix digunakan untuk mengevaluasi hasil ini. Confusion matrix menunjukkan hasil dari pengujian dengan menggambarkan hubungan antara kelas prediksi dan kelas aktual. Ada beberapa indikator yang digunakan dalam confusion matrix, seperti true positive rate (TP rate), true negative rate (TN rate) dan false positive rate (FP rate), false negative rate (FN rate), untuk mengevaluasi performa dari model.

f. *Deployment*

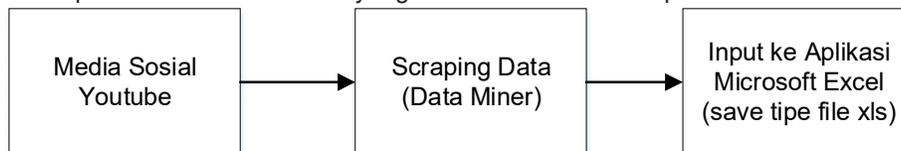
Tahap deployment adalah tahap pembuatan model implementasi yang dapat dibangun dengan tools pemrograman yang berbeda. Saat membangun model implementasi, kemudian hasil proses pengujian dan evaluasi digunakan sebagai bagian sumber data referensi.

Hasil dan Pembahasan

Dalam penelitian ini, sumber data yang digunakan berasal dari data komentar di media sosial YouTube yang bersumber dari kanal berita di youtube. Secara keseluruhan, data tersebut terdiri dari 1650 data komentar. Kemudian dilakukan pembersihan data dan setiap komentar diberi label positif atau negatif. Setelah proses pembersihan data selesai, 1450 komentar diterima dan ditandai sehingga data ini akan menjadi kumpulan data untuk penelitian ini.

a. *Pengumpulan data*

Pengumpulan data menggunakan ekstensi browser chrome yang disebut dengan Dataminer, Ekstensi ini digunakan untuk mengambil komentar di YouTube yang sesuai dengan kriteria yang ditentukan dan menyimpan data komentar tersebut dalam format yang sesuai. data yang diperoleh kemudian digunakan dalam proses analisis sentimen yang akan dilakukan dalam penelitian ini.



Gambar 3. Proses Pengumpulan Komentar

b. *Pemrosesan Data*

Pada saat melakukan pengolahan data untuk kemudian mendapatkan pemodelan yang sesuai pada kasus penelitian tentang sentimen analisis komentar informasi penerapan pajak pertambahan nilai (PPN)

atas kegiatan renovasi dan membangun sendiri, algoritma klasifikasi Naive Bayes serta Support Vector Machine digunakan. Tools RapidMiner versi 9.4 digunakan untuk implementasi algoritma tersebut, karena memiliki fitur-fitur yang akan dibutuhkan pada proses text-mining. Sebelum mencapai sebuah model yang baik, beberapa tahap harus dilakukan terlebih dahulu, seperti data understanding, preprocessing, modeling dan evaluation untuk dapat menentukan kegunaan dari model yang telah ditentukan dan dibuat.

Pada tahap ini pembahasan yang akan dilakukan yaitu melakukan proses pengolahan dataset terlebih dahulu sebelum dilakukan proses untuk klasifikasi menggunakan algoritma Naive Bayes (NB) serta Support Vector Machine (SVM) yang dioptimasi dengan PSO (Particle Swarm Optimization). Pada penelitian ini akan menerapkan beberapa langkah preprocessing untuk data teks komentar, dan berikut langkah-langkahnya:

a) Pelabelan

Pada tahap pelabelan, penulis akan melakukan proses pelabelan yang akan dilakukan secara manual yaitu menggunakan Aplikasi pengolah Excel pada ulasan komentar masyarakat di YouTube. Pelabelan dilakukan dengan memberikan label pada data pendapat komentar masyarakat di aplikasi Microsoft Excel. Kemudian Data tersebut akan dikategorikan ke dalam dua jenis label sentimen seperti positif dan negatif. Label positif diberikan pada ulasan yang memiliki sentimen positif terhadap aspek yang diteliti, label negatif akan diberikan pada ulasan yang memiliki sentimen negatif terhadap aspek yang diteliti. Ini dilakukan untuk membantu proses pemodelan dan evaluasi data yang akan digunakan dalam penelitian.

b) Transform Case

Tahap Transform Case merupakan proses melakukan konversi teks yang isinya tidak beraturan seperti pada pemilihan huruf dalam penulisan komentar, dengan proses ini maka teks komentar yang akan ditulis menjadi konsisten. Kemudian Proses case folding digunakan untuk mengubah huruf di dalam teks komentar ke dalam bentuk yang standar, yaitu mengubah huruf besar menjadi huruf kecil. Hal ini dilakukan untuk menghindari kesalahan dalam proses tokenisasi dan untuk menjamin keseragaman data yang digunakan dalam proses pemodelan klasifikasi. Case folding membuat proses analisis lebih mudah dan hasilnya lebih akurat karena data yang digunakan dalam analisis sesuai dengan standar yang ditentukan.

c) Tokenizing

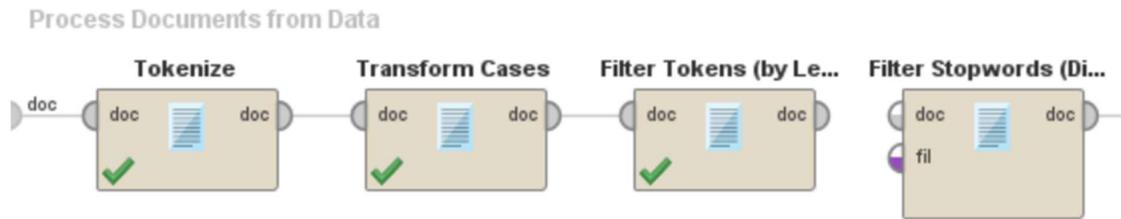
Tahap Tokenizing adalah tahap pemotongan string pada hasil inputan berdasarkan tiap kata yang membentuknya. Proses tokenisasi pada data teks komentar adalah pemecahan kata-kata di dalam susunan kalimat komentar menjadi potongan-potongan karakter maupun satuan kata dan sesuai kebutuhan dan sering disebut token. Pada proses tokenizing, Kemudian dilakukan proses filtering data untuk kebutuhan mengambil kata-kata yang memiliki maksimal 25 karakter dan minimal 3 karakter. Kata yang lebih dari 25 karakter atau terdiri dari kurang dari 3 karakter akan dihapus. Proses ini dilakukan agar kata yang digunakan dalam proses analisis sesuai dengan kriteria yang ditentukan dan hanya kata yang relevan yang digunakan dalam analisis

d) Filtering

Tahap Filtering adalah proses yang digunakan untuk mengambil isi kata yang dianggap penting dari hasil tokenisasi menerapkan algoritma stoplist atau wordlist. Algoritma stopword digunakan untuk menghilangkan kata yang tidak penting atau kurang relevan dalam analisis, sedangkan wordlist digunakan untuk menyimpan kata yang penting. Pada tahap ini, slang atau kosakata yang digunakan yaitu dalam Bahasa Indonesia, kemudian harus dilakukan proses normalisasi menjadi bentuk yang standar seperti "tdk" menjadi "tidak". Hal ini dilakukan agar tidak menghambat proses analisis dan hasil analisis lebih akurat.

e) Stopword

Tahap Stopword yaitu menghilangkan kata-kata yang tidak relevan atau tidak penting dalam analisis sentimen. Hal ini dilakukan dengan menghapus kata-kata seperti "tetapi", "untuk", "dengan" yang tidak memiliki makna tersendiri ketika dipisahkan dari kata lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen. Pada tahap ini, contoh teks akan berbeda dari teks sebelumnya karena kata-kata yang tidak penting sudah dihapus. Ini dilakukan untuk memastikan hanya kata-kata yang relevan yang digunakan dalam analisis dan meningkatkan akurasi dari hasil analisis.

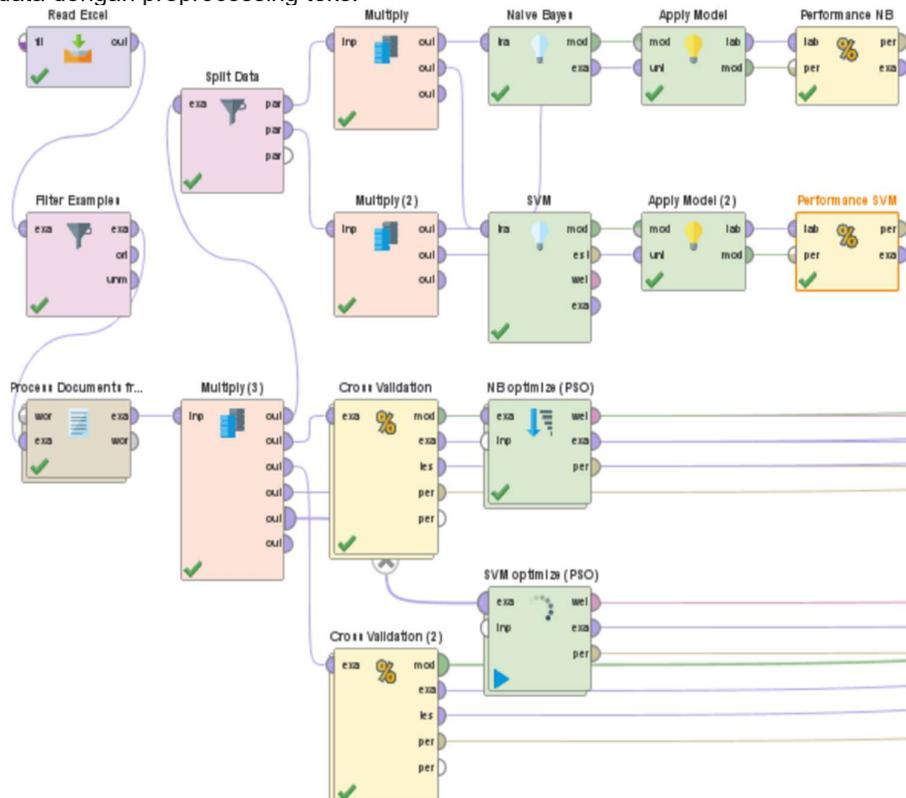


Gambar 4. Model Proses Data di Rapid Minner

c. Pembuatan Model Klasifikasi

Pada tahap penelitian ini, algoritma klasifikasi digunakan untuk membuat model pada materi teks komentar yang sudah melewati tahap pre-processing. Pada langkah ini digunakan dua algoritma klasifikasi yaitu Naive Bayes (NB) dan Support Vector Machine (SVM) yang dioptimalkan dengan Particle Swarm Optimization (PSO). Pada penelitian ini tools Rapidminer versi 9.40 digunakan untuk mengolah materi teks anotasi yang telah melewati tahap persiapan data dengan preprocessing teks.

Pada tahap penelitian ini, algoritma klasifikasi digunakan untuk membuat model pada materi data teks komentar yang telah melewati tahap pre-processing. Pada langkah ini digunakan dua jenis algoritma klasifikasi yaitu Naive Bayes (NB) serta Support Vector Machine (SVM) yang dioptimalkan menggunakan Particle Swarm Optimization (PSO). Pada penelitian ini tools Rapidminer versi 9.40 digunakan untuk mengolah materi teks anotasi, Langkah pertama yang dilakukan yaitu mengunggah data teks komentar menggunakan file Excel, yang kemudian akan diproses menggunakan algoritma Naive Bayes (NB) dan Support Vector Machine (SVM) untuk mendapatkan hasil awal setiap dari algoritma ini. Setelah menyelesaikan langkah pertama, penelitian ini dilanjutkan dengan membandingkan kedua algoritma tersebut dengan menambahkan algoritma optimasi menggunakan Particle Swarm Optimization (PSO). Langkah penggunaan PSO dalam proses pemodelan adalah meningkatkan hasil nilai akurasi klasifikasi algoritma Naive Bayes dan SVM, prosesnya dapat dilihat pada Gambar 4 hasil telah melewati tahap persiapan data dengan preprocessing teks.



Gambar 5. Evaluasi Desain Cross-Validation sebanyak 10 kali untuk Algoritma NB dan SVM tanpa PSO dan dengan PSO

Untuk hasil dari perbandingan akurasi serta Area Under Curve (AUC) dari hasil pada kedua jenis algoritma tersebut ditampilkan pada tabel 1.

Tabel 2. Perbandingan Akurasi dan AUC

Algoritma	Akurasi	AUC
Naïve Bayes	76.41%	0.500
Naiva Bayes + PSO	75.07%	0.567
SVM	89.75%	0.843
SVM + PSO	73.26%	0.656

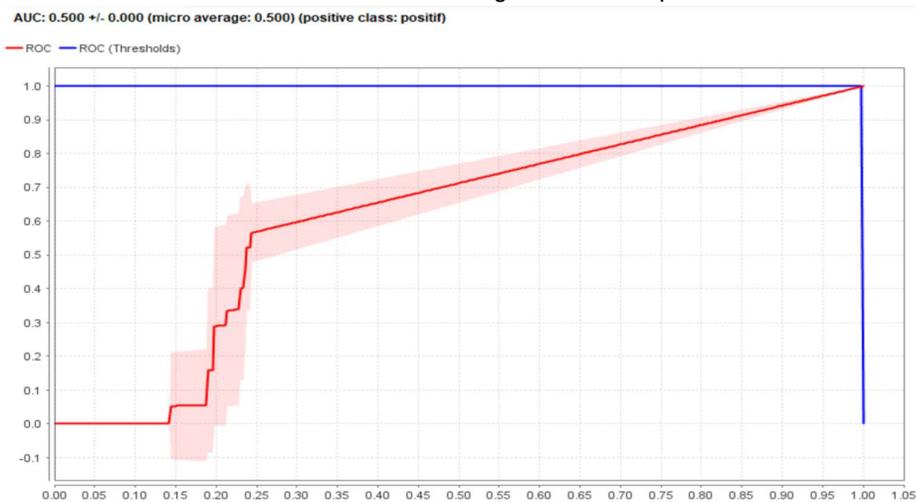
d. Hasil Klasifikasi Sentimen Masyarakat

Berikut ini adalah hasil klasifikasi sentiment masyarakat terkait topik Penerapan Pajak Bangun Rumah Sendiri:

- 1) Hasil akurasi menggunakan algoritma Naïve Bayes telah menghasilkan presentase sebesar 76.41% serta AUC 0.500 dapat dilihat pada gambar berikut ini:
accuracy: 76.41% +/- 2.96% (micro average: 76.41%)

	true negatif	true positif	class precision
pred. negatif	1045	79	92.97%
pred. positif	273	95	25.82%
class recall	79.29%	54.60%	

Gambar 6. Hasil akurasi Algoritma NB tanpa PSO

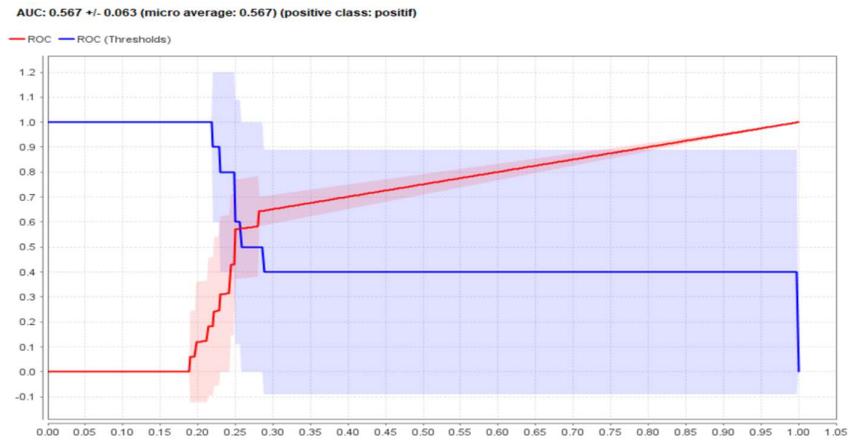


Gambar 7. Hasil AUC Algoritma Naïve Bayes Tanpa PSO

- 2) Hasil akurasi dengan menggunakan algoritma NB yang telah dioptimalisasikan menggunakan PSO menghasilkan presentasi sebesar 75.07% dan AUC 0.567 pada gambar berikut ini:
accuracy: 75.07% +/- 2.40% (micro average: 75.07%)

	true negatif	true positif	class precision
pred. negatif	1013	67	93.80%
pred. positif	305	107	25.97%
class recall	76.86%	61.49%	

Gambar 8. Hasil akurasi Naïve Bayes dengan PSO

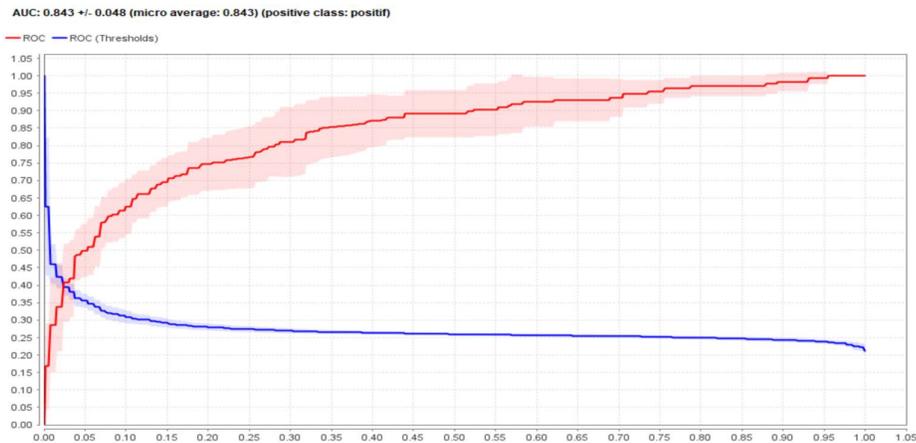


Gambar 9. Hasil AUC Algoritma Naive Bayes dengan PSO

- 3) Hasil akurasi dengan menggunakan algoritma SVM menghasilkan presentase sebesar 89.75% dan AUC 0.843 dapat dilihat pada gambar berikut:
accuracy: 89.75% +/- 1.40% (micro average: 89.75%)

	true negatif	true positif	class precision
pred. negatif	1310	145	90.03%
pred. positif	8	29	78.38%
class recall	99.39%	16.67%	

Gambar 10. Hasil Akurasi SVM Tanpa PSO



Gambar 11. Hasil AUC Algoritma SVM tanpa PSO

- 4) Hasil akurasi dengan menggunakan algoritma SVM dengan PSO menghasilkan presentase sebesar 89.75% dan AUC 0.843 pada gambar berikut:
accuracy: 73.26% +/- 3.44% (micro average: 73.26%)

	true negatif	true positif	class precision
pred. negatif	995	76	92.90%
pred. positif	323	98	23.28%
class recall	75.49%	56.32%	

Gambar 12. Hasil Akurasi SVM dengan PSO



Gambar 13. Hasil AUC Algoritma SVM dengan PSO

Simpulan

Secara keseluruhan, kesimpulan dari penelitian ini menjelaskan bahwa pengolahan data dilakukan melalui beberapa tahap yang meliputi pelabelan, transform case, tokenizing, filtering dan stopwords removal, untuk menghasilkan data yang berkualitas untuk digunakan dalam proses analisis sentimen. Hasil dari penelitian ini menunjukkan bahwa algoritma Support Vector Machine yang dioptimalkan dengan Particle Swarm Optimization (PSO) memberikan hasil yang lebih baik daripada Naive Bayes dalam klasifikasi sentimen analisis informasi penerapan pajak pertambahan nilai pada kegiatan renovasi dan membangun rumah sendiri.

Selain itu, penelitian ini juga memberikan saran untuk penelitian selanjutnya, seperti melakukan pengujian dengan menggunakan algoritma yang berbeda serta metode lainnya, dan tidak hanya meneliti topik yang relevan dengan penerapan pajak pertambahan nilai pada kegiatan renovasi dan membangun rumah sendiri, namun juga bisa melakukan penelitian dengan topik-topik lain yang berbeda.

Daftar Pustaka

- [1] X. Cheng, C. Dale, and J. Liu. Understanding the characteristics of internet short video sharing: YouTube as a case study. In Technical Report arXiv:0707.3670v1 cs.NI, New York, NY, USA, 2007. Cornell University, arXiv e-prints
- [2] Bridge, C. 2011. Unstructured Data and the 80 Percent Rule. [Online]. Tersedia di: <http://www.clarabridge.com/default.aspx?tabid=137&ModuleID=635 &ArticleID=551>
- [3] Jannti, R., Mahendra, R., Wardhana., W, C. Adriani, M, 2018. Stance Classification Towards Political Figures on Blog Writing. In: IEEE, 2018 International Conference on Asian Language Processing (IALP). Bandung, Indonesia 15-17 November 2018. Indonesia.
- [4] Oxford Dictionaries, 2018. Sentiment Analysis. [Online] Tersedia di: https://en.oxforddictionaries.com/definition/sentiment_analysis/
- [5] Liu, B., 2015. Sentiment Analysis Opinions, Sentiment, and Emotion in Text. Chicago, USA: Cambridge University Press.
- [6] D. C. Rini, Y. Farida, and D. Puspitasari, "Klasifikasi Menggunakan Metode Hybrid Bayesian-Neural Network (Studi Kasus: Identifikasi Virus Komputer)," J. Mat. "MANTIK," vol. 1, no. 2, p. 38, 2016, doi: 10.15642/mantik.2016.1.2.38-43.
- [7] A. P. Fadillah and B. Hardiyana, "Penerapan Naïve Bayes Classifier Untuk Pemilihan Konsentrasi Mata Kuliah," J. Teknol. dan Inf. UNIKOM, vol. 8, no. 2, 2018, doi: 10.34010/jati.v8i2.1039.
- [8] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 12, pp. 1718–1724, 2017.
- [9] D. J. Haryanto, L. Muflikhah, and M. A. Fauzi, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," J. Pengembangan. Teknologi. Inf. dan Ilmu Komput. Univ. Brawijaya, vol. 2, no. 9, pp. 2909–2916, 2018.
- [10] I. Santoso, W. Gata, and A. B. Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 1, no. 10, pp. 5–11, 2019
- [11] Saepulrohman, A. Saepudin, S, Gustian, D., 2019. Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp menggunakan Algoritma Naïve Bayes dan Support Vector Machine, Volume 6, Nomor 2

-
- (2021) Hal. 91-105 ISSN: 2252-9853 (Print) | ISSN: 2656-808X (Online)
<https://ojs.unikom.ac.id/index.php/aisthebest/index>
- [12] Kurniawan, S. Gata, W, Puspitawati, A D. , Nurmalasari., , Tabrani, M., , Novel K., ,
2019.Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Berita Online Pada Komentar
Berita Online, Vol. 3 No. 2 (2019) 176 – 183 ISSN Media Elektronik: 2580-0760 <http://jurnal.iaii.or.id>
- [13] Kenneth Jensen, 2016. IBM SPSS Modeler CRISP-DM Guide. [Online] Tersedia di:
[ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.
pdf](ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf)