



Komparasi Metode Text Mining Terhadap Masalah Pengklasifikasian Narasi Informative & Non Informative Pada twitter @PLN_123

Lina Hermawati¹, Vincentius Berland², Anissa Rahmadiyah³, Exaudi Hutabarat^{4✉}, Dedi Dwi Saputra⁵

^{1,2,3,4}Program Studi Sistem Informasi, Universitas Nusa Mandiri

⁵Program Studi Sains Data, Universitas Nusa Mandiri

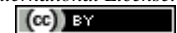
exaudybarth@gmail.com

Abstrak

PT. PLN sebagai Badan Usaha Milik Negara yang bergerak dibidang penyedia tenaga listrik memiliki tujuan untuk melayani kepentingan umum dan menjaga mutu serta kualitas dari daya dan energi itu sendiri. Melalui akun resmi twitternya, PT.PLN menggunakan informasi yang ada untuk bisa mengumpulkan segala keluhan,kritikan maupun dukungan guna memberikan pelayanan terbaik kepada masyarakat. Berdasarkan penelitian yang telah dilakukan, data hasil hasil *crawling* dari media sosial twitter dengan query @PLN_123 dengan algoritma *Naïve Bayes* telah berhasil dilakukan. Algoritma *Naïve Bayes* & Adaboost yang dioptimasi dengan fitur *Synthetic Minority Over-Sampling Technique* (SMOTE) adalah komparasi yang terbaik dengan nilai accuracy: 90,68%, precision: 84,85%, recall: 99,30% dan AUC: 0,910. Hasilnya kemudian diimplementasikan dalam aplikasi berbasis web yang dibangun menggunakan bahasa pemrograman HeidiSQL.

Kata kunci: Sentimen Analisis, PLN, Twitter, *Naïve Bayes*, Adaboost

JSISFOTEK is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

PT. Perusahaan Listrik Negara (PLN) adalah perusahaan milik pemerintah dengan misi untuk menyediakan tenaga listrik bagi kepentingan umum dalam jumlah dan mutu yang memadai serta memupuk keuntungan [1]. Dikarenakan cakupan wilayah PT PLN tersebar diseluruh Indonesia maka analisis opini publik penting dilakukan untuk dapat menganalisis opini masyarakat tentang kondisi pelayanan ketersediaan listrik ataupun permasalahan yang terjadi dapat diminimalisir kesalahan dan kekurangan. Di era globalisasi saat ini, media sosial sangat umum digunakan dan sudah menjadi kebutuhan sehari-hari. Media sosial menjadi layanan internet yang paling sering diakses oleh masyarakat Indonesia, salah satunya adalah Twitter [2]. Pengguna twitter di Indonesia mencapai 18,45 juta pada tahun 2022, sehingga menempatkan Indonesia di peringkat kelima negara pengguna twitter terbesar di dunia. Melalui akun twitter @pln 123 segala jenis opini publik yang ingin diungkapkan oleh masyarakat, seperti mengenai pelayanan, kualitas, mutu, serta hal yang ingin dilaporkan lainnya dapat diungkapkan melalui akun tersebut [3]. Pendapat yang berupa data tekstual tersebut memiliki pengetahuan tersembunyi yang dapat diekstraksi dengan *text mining* menggunakan seperangkat *tools* analisis [4]. Guna melakukan analisis layanan diperlukan suatu metode yang mampu mengklasifikasikan opini atau keluhan masyarakat yaitu dengan menggunakan algoritma *Naïve Bayes*. Sehingga opini yang berisi narasi negative maupun narasi positive dapat diklasifikasikan [5]. Pada penelitian ini penulis ingin melakukan analisis sentiment *tweet* warganet di twitter @PLN_123, dan melakukan komparasi metode text mining terhadap masalah pengklasifikasian narasi informative & non informative. Melalui komparasi algoritma *Naïve Bayes*, Smoote dan Adaboost tersebut diharapkan dapat diketahui metode mana yang lebih optimal.

Penelitian tentang analisis sentimen sudah dilakukan sebelumnya, seperti yang dilakukan oleh (Gunawan et al., 2018). Dengan judul “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes”, menjelaskan sistem analisis sentimen dibagi menjadi 5 tahap, yaitu crawling, pre-processing, pembobotan kata, pembentukan model dan klasifikasi sentiment. Data yang sudah diklasifikasikan akan dievaluasi menggunakan pengujian confusion matrix dengan parameter akurasi, recall dan precision. Penelitian yang dilakukan tentang analisis sentiment juga dilakukan oleh (Ratnawati, 2018). dengan judul “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter” Dimana dalam penelitian ini algoritma *Naïve Bayes* dapat digunakan untuk mengklasifikasikan tweet positif atau negative tentang opini film pada twitter. Penerapan metode *naive Bayes* terdapat juga pada judul “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan” (Putro et al., 2020). Berdasarkan master pelanggan yang dijadikan data latih, telah berhasil mengklasifikasikan 23 dari 25 data yang diuji dengan nilai *precision* mencapai 100%, nilai *recall* 91%, dan nilai *accuracy* mencapai 92%. Sementara itu pada judul “ Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter

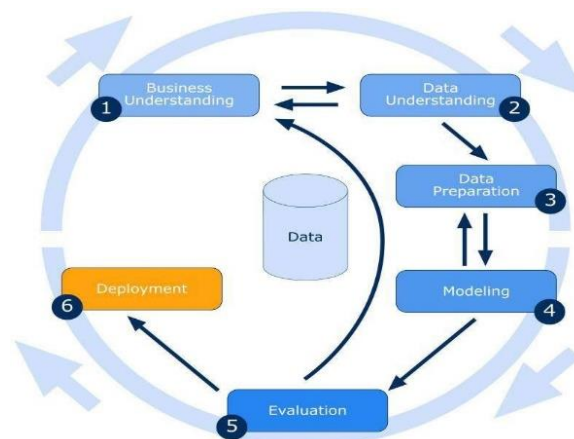
(Studi Kasus Ovo dan Dana)”(Ogi et al., 2021). Model klasifikasi *Support Vector Machine*, merupakan model machine learning yang mendapatkan hasil akurasi paling tinggi diantara *Naïve Bayes* dan *K-Nearest Neighbor* berdasarkan nilai metrics, akurasi presisi. Penelitian sebelumnya yang telah dilakukan membuktikan bahwa analisis sentimen berhasil diterapkan untuk mengetahui sentimen masyarakat terhadap suatu masalah yang menjadi topik utama pembicaraan. Berdasarkan hal tersebut, maka dilakukan penelitian mengenai analisis sentimen narasi informatif dan non informatif pada twitter @PLN_123 dan komparasi metode text mining.

2. Metodologi Penelitian

Metode yang digunakan dalam penelitian ini adalah metode *Naïve Bayes*. Algoritma *Naive Bayes Classifier* merupakan pengklasifikasian dengan metode probabilitas, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya dikenal juga dengan istilah teorema Bayes [10]. Salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam Text Mining adalah metode *Naïve Bayes* yang sering disebut dengan *Naive Bayes Classifier*. Selain itu Metode penelitian yang diusulkan dalam penelitian ini menggunakan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang telah disesuaikan dengan kebutuhan penelitian, metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM), metode *Industry Standard Process for Data Mining* (CRISP-DM) terdiri dari 6 tahapan yaitu: *Business Understanding*, *Data Understanding*, *Preprocessing*, *Modelling*, *Evaluation*, *Deployment*.

Tabel 1. Tabel Software dan Hardware Pendukung

Informative	Non Informative
599	401



Gambar 1. Metodologi CRISP-DM

2.1. Business Understanding (Pemahaman Bisnis)

Pemahaman bisnis adalah studi tentang subjek penelitian yang sedang dilakukan. Pada saat ini, pemahaman terhadap objek penelitian dilakukan dengan menggunakan informasi dari twitter [11]. Penelitian yang diusulkan mengambil data dari sosial media twitter dengan menggunakan *keyword* @pln_123. Pada tahapan ini dijelaskan terlebih dahulu mengenai informasi dari sosial media twitter yang digunakan oleh PLN Persero. Pengumpulan data dilakukan dengan menggunakan data yang bersumber dari komentar berbahasa Indonesia pada akun Twitter @pln_123 menggunakan Rapid Miner Studio 9.10 untuk membuat data set berisi 1.000 ulasan kemudian melakukan labeling. dengan menghasilkan data 599 status Informative dan 401 Status Non Informatif.

2.2. Data Understanding (Pemahaman Data)

Tahap *data understanding* adalah bagian yang dibutuhkan untuk pemahaman mengenai data yang akan digunakan dalam proses penelitian [11]. Data yang akan digunakan untuk eksperimen penelitian ini dikumpulkan melalui data tweet yang berasal dari twitter, yang dimana pada data tweet tersebut mengandung narasi Informative & Non Informative data pada twitter tersebut akan diambil dengan menggunakan *tools RapidMiner* dan disimpan agar data tersebut dapat digunakan sebagai *dataset local*.

2.3. Data Preparation (Persiapan Data)

Data Preparation adalah proses menyiapkan data untuk mendapatkan informasi yang akurat, bersih dan siap untuk digunakan dalam penelitian [12]. Setelah mempunyai *dataset local* yang dikumpulkan dari sosial media twitter dengan menggunakan keyword @pln_123, pada data tersebut dikelompokkan mana yang mengandung narasi Informative & Non Informative data pada twitter yang simpan pada *dataset local* adalah bentuk data yang tidak terstruktur sehingga pada data tersebut diperlukan *preprocessing* dahulu.

2.4. Modelling (Pemodelan)

Pemilihan teknik pemodelan adalah langkah pertama yang akan diambil dan diikuti untuk pembuatan skenario pengujian dan memvalidasi kualitas sebuah model. Berdasarkan hal tersebut, metode yang diusulkan pada penelitian ini akan menggunakan pendekatan metode klasifikasi dengan algoritma *Naïve Bayes* yang ditambahkan *feature Synthetic Minority Oversampling Technique Method* (SMOTE) untuk menangani ketidak seimbangan kelas [1] dan adaboost dikarenakan algoritma tersebut merupakan teknik *machine learning* yang populer untuk permasalahan klasifikasi teks serta memiliki performa yang baik pada banyak domain.

2.5. Evaluation (Pengujian)

Penelitian ini akan menggunakan teknik *Cross Validation* dengan menggunakan *tools* Rapidminer untuk melakukan pengujian model dengan *Naïve bayes*, *Synthetic Minority Oversampling Technique Method* (SMOTE) dan adaboost berdasarkan data Pengklasifikasian tweet yang sudah dilakukan.

2.6. Deployment (Penyebaran)

Hasil pengujian melalui tahapan eksperimen dan evaluasi dari model yang akan digunakan, akan menghasilkan nilai-nilai yang dapat digunakan sebagai data acuan dalam pembuatan model implementasi. Pada tahap ini akan membuat kesimpulan dari penelitian setelah mendapatkan hasil nilai tingkat akurasi Pengklasifikasian Informatif & Non Informatif. Dengan pendekatan metode *Naïve bayes* ditambahkan dengan *feature Synthetic Minority Oversampling Technique Methode* (SMOTE) dan AdaBoost pada *dataset local*. Tahap Ini adalah langkah terakhir dalam metode CRISP-DM, yang melibatkan pembuatan atau membangun aplikasi berdasarkan model dan data yang dihasilkan pada langkah sebelumnya [13]. Model implementasi usulan ini rencananya akan dibuatkan dalam sebuah *tools* aplikasi yang berbasis pemrograman PHP dan MySQL.

3. Hasil dan Pembahasan

Sesuai dengan metodologi yang sudah disampaikan akan dijelaskan tahapan implementasi metodologi penelitian yang dikerjakan, adalah sebagai berikut:

3.1. Business Understanding

Pada tahapan *business understanding*, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan mengambil data dari sosial media twitter dengan menggunakan *keyword* @pln_123 yang mengandung narasi atau kalimat Informatif & Non Informatif. Pada fase ini data tweet dari para pelanggan diklasifikasi berdasarkan tweet “Informatif & Non Informatif” yang dialami oleh pelanggan PLN. Pada tahap ini pun dilakukan pemahaman untuk mencari metode dengan pendekatan model Pengklasifikasian terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma. Adapun algoritma yang digunakan adalah algoritma *Naïve bayes* ditambahkan fitur *Synthetic Minority Oversampling Technique Methode* (SMOTE) dan adaboost, Untuk meningkatkan akurasi dari metode Pengklasifikasian yang digunakan dalam *dataset local*.

3.2. Data Understanding

Tahap ini adalah proses memahami data yang akan digunakan sebagai bahan yang akan diteliti untuk bisa dilakukan ke tahap setelahnya yaitu *Preprocessing*. Dibawah ini adalah langkah-langkah yang akan dikerjakan ; menyiapkan data tweet yang diambil dari sosial media twitter, dimana data tersebut sebanyak 1.000 data. Proses selanjutnya adalah *cleaning* data, dengan cara melakukan *remove duplicated* terhadap data-data tersebut, agar menjadi data yang unik. Jumlah data yang awalnya 1.000 setelah di *cleaning* dengan melakukan *remove duplicated* data tersebut menjadi 999 data. Proses Pengklasifikasian dan *labelling* dilakukan oleh kuesioner. Dengan adanya bantuan pada proses pelabelan dan pengklasifikasian yang dilakukan oleh kuesioner, hal ini sangat efisien karena dapat menekan biaya jika dibandingkan dengan menggunakan jasa para ahli/ pakar bahasa Indonesia untuk melakukan hal tersebut. Hasil dari *labelling* dan pengklasifikasian yang dilakukan oleh kuesioner bertujuan agar status data menjadi normal dalam penelitian *text mining*. Adapun hasil dari tahap ini adalah sebagai berikut:

Tabel 2. Hasil proses pelabelan

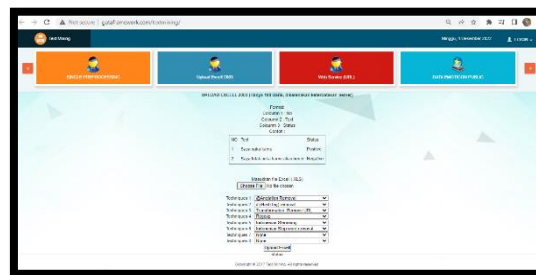
Informative	Non Informative
599	401

Tabel 3. Hasil proses pelabelan

No	Text	Status
1	@tantompas aribu Hi Kak Tantom, Mhn maaf terkait PJU dpt hub Pemda setempat dikarenakan dikelola oleh Dinas Pertamanan Kota. Agar semua makin mudah, ayo download aplikasi New PLN Mobile di Play Store atau Apps Store.	Informatif
2	@LadikhaPuteri Hi Kak Ladikha, mohon maaf atas ketidaknyamanannya. Padam diwilayah tsb karena gangguan menengah dan saat ini sudah normal kembali.	Informatif
3	pagi2 udah mati aja listriknya oom @pln_123	Non informatif
4	@pln_123 hi min tolong cek dm yah????	Non informatif

3.3. Data Preparation

Dari hasil proses pelabelan terhadap dataset yang telah dikumpulkan tersebut kemudian langkah-langkah selanjutnya adalah melakukan *text-preprocessing*, dengan menggunakan dua aplikasi *preprocessing*. Pertama menggunakan Gataframework yang dapat diakses melalui link <http://www.gataframework.com/>. Langkah kedua menggunakan *preprocessing* dari Rapidminer. tahapan *preprocessing* mencakup prosedur memperbaiki data serta menghapus data yang tidak diperlukan supaya dapat diolah ke tahapan berikutnya [14]. Dan yang ketiga dilakukan pembobotan nilai pada dataset distribusi yang dihasilkan, berikut adalah tahapannya:

**Gambar 2.** Desain Model *Preprocessing* tahap 1 *Data Local* menggunakan *Gataframework*

Pada gambar 2, menunjukan tahap *preprocessing* dan *cleansing* dengan menggunakan beberapa teknik pada website gataframework untuk membuat *design preprocessing* dan *cleansing* pada *dataset local*. Dengan urutan pertama *@anotation removal*, *# (HashTag) removal*, *transform url*, *regex*, *Indonesian stemming*, dan *Indonesian stopword removal*

3.3.1. Annotation Removal

Teks diurai berdasarkan *white space*. Dalam proses ini, semua anotasi yang terkandung dalam *tweet* akan dihilangkan dan mengubah seluruh huruf kapital menjadi huruf kecil.

Tabel 4. Hasil proses Annotation Removal

Teks sebelum proses <i>Removal</i>	Teks setelah proses <i>Anotation Removal</i>
@jochomint ♦ Halo Kak, maaf ya atas kendalanya dan telah admin respon via DM. Thanks - D	♦ halo kak,maaf ya atas kendalanya dan telah admin respon via dm thanks -dion

3.3.2. Remove Hastag

Hasil dari *@anotation removal* akan di lanjutkan ke proses fungsi Remove hastag dalam proses ini akan menghapus tanda hastag (#) pada twitter.

Tabel 5. Hasil proses Remove Hastag

Teks sebelum proses <i>Remove Hastag</i>	Teks setelah proses <i>Remove Hastag</i>
@pln_123 diwilayah kota banda aceh dan sekitarnya. Gak ada angin, gak ada hujan Listrik kok padam..?Kenapa..? #PLNUntukIndonesia	diwilayah kota banda aceh dan sekitarnya. gak ada angin, gak ada hujan listrik kok padam..? kenapa..?

3.3.3. Transformation Remove URL

Hasil dari @ Remove hastag akan di lanjutkan ke proses *Transformation Remove URL*, dalam proses ini link atau URL yang terkandung pada tweet akan dihilangkan.

Tabel 6. Hasil proses Transformation Remove URL

Teks sebelum proses <i>Transformation Remove URL</i>	Teks setelah proses <i>Transformation Remove URL</i>
@pln_123 tropodo, waru sidoarjo mati lagi setelah tadi hampir 5 jam lampu padam	tropodo, waru sidoarjo mati lagi setelah tadi hampir 5 jam lampu

3.3.4. Tokenize Regular Expression (Regex)

Hasil dari *Transformation Remove URL* dilanjutkan oleh proses *Tokenization (Regex)* yaitu semua kata yang ada didalam tiap dokumen dikumpulkan dan dihilangkan tanda baca, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf

Tabel 7. Hasil proses *Tokenization (Regex)*

Teks sebelum proses <i>Tokenization (Regex)</i>	Teks setelah proses <i>Tokenization (Regex)</i>
@erikmoer Halo Kak Erik, sebelumnya Admin mohon maaf ya. Untuk pembelian token silakan bisa dicoba kembali. Jika masih terkendala Kakak dapat hubungi Admin via DM atau Kakak juga bisa melakukan pelaporan di apk PLN mobile dengan memilih menu "pengaduan". Thanks -Reka	halo kak erik sebelumnya admin mohon maaf ya untuk pembelian token silakan bisa dicoba kembali jika masih terkendala kakak dapat hubungi admin via dm atau kakak juga bisa melakukan pelaporan di apk pln mobile dengan memilih menu pengaduan thanks reka

3.3.5. Indonesian Stemming

Hasil dari *Tokenization (Regex)* akan dilakukan proses *stemming* yaitu mengubah kata berimbuhan menjadi kata dasar menggunakan *indonesian stemming* untuk tweet berbahasa Indonesia.

Tabel 8. Hasil proses *Indonesian stemming*

Teks sebelum proses <i>Stemming</i>	Teks setelah proses <i>Stemming</i>
@tuaice Hai Kak, Mohon maaf atas gangguan yang terjadi, Kakak bisa info detailnya melalui DM ke Mimin atau Lapor langsung melalui PLN Mobile di menu Pengaduan. Mimin tunggu ya konfirmasinya lebih lanjut. Tks- Ichsan	hai kak mohon maaf atas ganggu yang jadi kakak bisa info detail lalu dm ke mimin atau lapor langsung lalu pln mobile di menu adu mimin tunggu ya konfirmasi lebih lanjut tks ichsan

3.3.6. Indonesian Stopword Removal

Hasil dari *indonesian stemming* akan di lanjutkan ke proses *Indonesian stopwords removal*, dalam proses ini kata-kata yang tidak relevan akan dihapus, seperti kata tetapi, untuk, dengan, yang dimana kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan *sentiment* [15].

Tabel 9. Hasil proses *Indonesian stopwords removal*

Teks sebelum proses <i>Stopword removal</i>	Teks setelah proses <i>Stopword removal</i>
@pln_123 Min, tolong cek DM. Padam listrik lagi nih Min. Info sampai kapan ya ini?	min tolong cek dm padam listrik min info

Setelah selesai dilakukan *preprocessing* tahap 1, kemudian dilakukan pengujian dilakukan menggunakan *tool Rapidminer*. Tahapan *pre-processing* tahap 2 dilakukan beberapa proses uji coba pada dataset sehingga menghasilkan data yang lebih akurat untuk pengimplementasian *machine learning*.



Gambar 3. Desain Model *Preprocessing* tahap 2 Data Local Menggunakan Rapid Miner

Pada gambar 3, menunjukan *process modelling* pada aplikasi rapidminner dalam *process document*, dengan urutan *transform-cases*, *tokenize*, *filter token by length* dan *list stopwords tambahan*.

3.3.7. Tokenization

Hasil dari *indonesian stopwords removal* dilanjutkan oleh proses *Tokenization* dari RapidMiner. Tujuan dari proses tokenisasi adalah untuk menghapus spasi kosong, tanda baca serta simbol non huruf, dan karakter dari setiap dokumen ulasan. Setiap karakter yang tidak diperlukan akan dihapus serta memecah kalimat menjadi perkata [15].

Tabel 10. Hasil proses *Tokenization*

Teks sebelum proses <i>Tokenization</i>	Teks setelah proses <i>Tokenization</i>
mati lampu di metro serpong desa cibogo kecamatan cisauk kabupaten tangerang banten sudah jam lebih hujan dan petir sudah berlalu	Mati Lampu Dimetro Serpong Desa Cibogo Kecamatan Cisauk Kabupaten Tangerang Banten Sudah Jam Lebih Hujan Dan Petir Sudah berlalu

3.3.8. Filter Token (by length)

Hasil dari proses Tokenization pada Rapidminer dilanjutkan oleh proses Filter Tokens (by Length). *Filter tokens by length* digunakan untuk menghapus kata dengan jumlah karakter yang kurang dari nilai yang ditentukan [16]. Pada proses ini kata-kata yang memiliki panjang karakter kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata di, ada, oleh, yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentiment

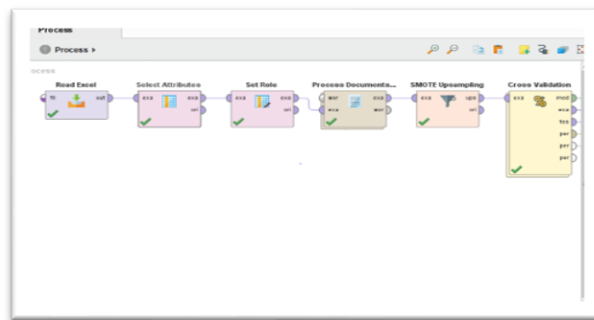
Tabel 11. Hasil proses Token (by length)

Teks sebelum proses Tokens (by length)	Teks setelah proses Stemming
Mati	Mati
Lampu	Lampu
Dimetro	Serpong
Serpong	Desa
Desa	Cibogo
Cibogo	Kecamatan
Kecamatan	Cisauk
Cisauk	Kebupaten
Kebupaten	Tangerang
Tangerang	Banten
Banten	Sudah
Sudah	Lebih
Jam	Hujan
Lebih	Petir
Hujan	Sudah
Dan	Berlalu
Petir	
Sudah	
berlalu	

3.4. Modelling Metode

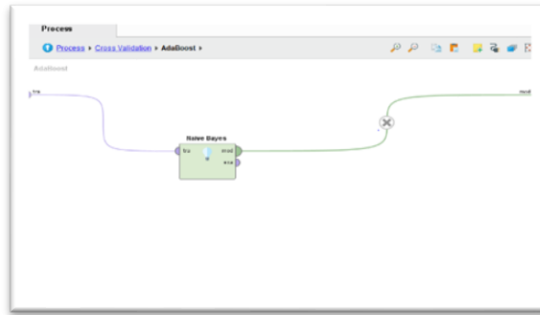
Pada tahap yang ke-3 ini dataset pada tahap sebelumnya digunakan sebagai masukan untuk algoritma klasifikasi, yakni dijadikan sebagai *dataset training* dan *testing*. Pada penelitian ini akan digunakan dua jenis algoritma sekaligus untuk di komparasi, yaitu *Naïve bayes* dengan ditambahkan dengan *feature Synthetic Minority Over Sampling Technique Methode (SMOTE)* dan *Adaboost*.

3.4.1. Modelling naïve bayes, Smote dan Adaboost

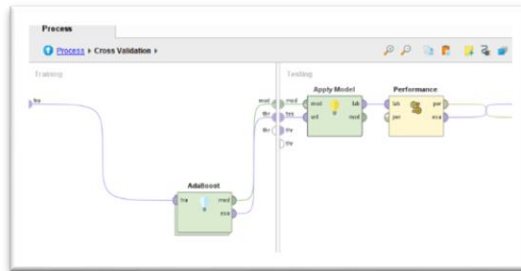


Gambar 4. Desain Model Proses Menggunakan Rapid Miner

Pada desain model tersebut diawali dengan operator *Read Excel*, *Select Attribute*, *Set Role*, *Process Document*, *Smote Upsampling*, dan *cross Validation*



Gambar 5. Desain Model Naïve bayes Menggunakan Rapid Miner



Gambar 6. Desain Model Adaboost Menggunakan Rapid Miner

3.4.2. Confusion Matrix Algoritma Naïve Bayes, Smoote dan Adaboost

	true non_informatif	true informatif	class precision
pred. non_informatif	700	6	99.15%
pred. informatif	153	847	84.70%
class recall	82.06%	99.30%	

precision: 84.80% +/- 3.10% (micro average: 84.70%) (positive class: informatif)

	true non_informatif	true informatif	class precision
pred. non_informatif	700	6	99.15%
pred. informatif	153	847	84.70%
class recall	82.06%	99.30%	

recall: 99.30% +/- 1.25% (micro average: 99.30%) (positive class: informatif)

	true non_informatif	true informatif	class precision
pred. non_informatif	700	6	99.15%
pred. informatif	153	847	84.70%
class recall	82.06%	99.30%	

Gambar 7. Hasil Proses Confusion Matrix Algoritma Naïve Bayes, Smoote dan Adaboost



Gambar 8. Grafik ROC algoritma Naïve Bayes, Smote dan Adaboost

3.5. Evaluation

Tahap evaluasi bertujuan untuk menentukan seberapa baik model yang telah dibuat pada awal proses langkah sebelumnya [16]. Dari hasil pengujian yang telah dilakukan dalam beberapa komparasi algoritma, dalam penelitian ini menghasilkan hasil evaluasi yang akan dilihat perbandingannya agar mendapatkan model klasifikasi yang terbaik pada tabel berikut :

- [1] Hasil dengan Algoritma Naïve Bayes saja

Tabel 12. Hasil dengan algoritma *Naïve Bayes*

Ket.	Accuracy	Precision	Recal	AUC
Persentase	83,89%	47,74%	80,19%	0,500

- [2] Hasil dengan Algoritma *Naïve Bayes* ditambah fitur SMOTE

Tabel 33. Hasil dengan algoritma *Naïve Bayes* ditambah fitur *SMOTE*

Ket.	Accuracy	Precision	Recal	AUC
Persentase	91,73%	85,96%	99,88%	0,534

- [3] Hasil dengan Algoritma *Naïve Bayes & Adaboost* ditambah fitur SMOTE

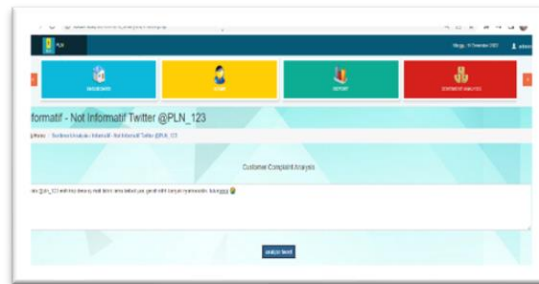
Tabel 44. Hasil dengan algoritma *Naïve Bayes & Adaboost* ditambah fitur *SMOTE*

Ket.	Accuracy	Precision	Recal	AUC
Persentase	90,68%	84,85%	99,30%	0,910

Berdasarkan dari data hasil uji tersebut, data yang menggunakan algoritma *Naïve Bayes & Adaboost* ditambah fitur **SMOTE** memiliki kinerja sangat baik dengan *Accuracy* yang diperoleh yaitu 90.68%, *precision* 84.85%, *Recall* 99.30% dan AUC 0.910% (*good classification*)

3.6. Deployment

Berdasarkan hasil evaluasi dari proses pengujian model antara algoritma Naïve Bayes, *feature Synthetic Minority Over-sampling Technique Methode* (SMOTE) dan Adaboost didapati hasil pengujian model tertinggi dari seluruh hasil pengujian algoritma adalah Naïve Bayes, *feature Synthetic Minority Over-sampling Technique Methode* (SMOTE) dan Adaboost. Maka dari itu, bobot yang akan digunakan dalam tahap deployment pemodelan aplikasi berdasarkan dari hasil pengujian algoritma yang ditambahkan *feature Synthetic Minority Over-sampling Technique Methode* (SMOTE) dan Adaboost



Gambar 9. Get Tweet dengan API Twitter

Pada gambar 9, menunjukan hasil *deployment* untuk *get* tweet ke twitter menggunakan *API (Application Programming Interface)* twitter dengan mention @pln_123. Pada tahap ini data tweet diambil agar dapat dilakukan pada step berikutnya yaitu *text preprocessing*



Gambar 10. Text Preprocessing

Pada gambar 10, setelah data tweet diambil step berikutnya adalah melakukan *preprocessing* dan *cleansing* pada *text* tersebut, dengan menggunakan teknik *Remove @annotation*, *Remove Hastag*, *Remove URL*, *Tokenize Regexp*, *Indonesian Stemming*, *Indonesian Stopword*



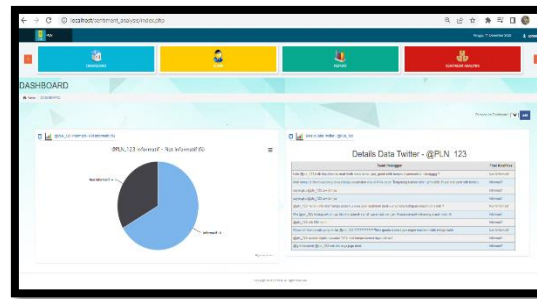
Gambar 11. Proses Tokenization

Pada gambar 11, setelah *preprocessing* dan *cleansing* pada *text* tersebut, berikutnya Proses Tokenization adalah semua kata yang ada didalam tiap dokumen dikumpulkan dan dihilangkan tanda baca, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf dan memecah kalimat menjadi perkata



Gambar 12. Hasil Proses Tokenization

Pada gambar 12 Hasil Proses Tokenization tahap selanjutnya ialah menghitung bobot kata untuk Klasifikasi Informatif dan Non Informatif. Pada gambar 12, Hasil proses dari perhitungan untuk Klasifikasi tersebut menghasilkan kesimpulan Non Informatif



Gambar 12. Data Grafik dan hasil summary

Pada gambar 12 menunjukkan data grafik dan hasil summary prediksi kategori yang sudah dilakukan pengklasifikasian berdasarkan data tweet dan sudah dihitung bobot tiap kata dari masing-masing pengklasifikasian. Disamping itu, fungsi grafik adalah untuk memonitor jumlah data tweet yang sudah diklasifikasikan agar dapat dipantau perkembangan datanya.

4. Kesimpulan

Pengujian terhadap data hasil hasil *crawling* dari media sosial twitter dengan query @PLN_123 dengan algoritma *Naïve Bayes* telah berhasil dilakukan. Berdasarkan dari penelitian yang telah dilakukan, menunjukkan bahwa Algoritma *Naïve Bayes* & *Adaboost* jika dioptimasi dengan menggunakan Synthetic Minority Over-sampling Technique (SMOTE) menghasilkan nilai accuracy: 90,68%, precision: 84,85%, recall: 99,30% dan AUC: 0,910. Pada penelitian ini juga menggunakan algoritma *Naïve Bayes* dengan optimasi SMOTE menghasilkan accuracy: 91,73%, precision: 85,96%, recall: 99,88% dan AUC: 0,534. Sedangkan dalam penggunaan algoritma *Naïve Bayes* tanpa optimasi SMOTE menghasilkan accuracy 83,89%, precision: 47,74%, recall: 80,19% dan AUC: 0,500. Berdasarkan hasil penelitian ini dapat disimpulkan bahwa Algoritma *Naïve Bayes* & *Adaboost* yang dioptimasi dengan fitur *Synthetic Minority Over-Sampling Technique* (SMOTE) adalah komparasi yang lebih baik digunakan dibandingkan dengan Algoritma *Naïve Bayes* dengan optimasi SMOTE maupun Algoritma *Naïve Bayes* tanpa optimasi SMOTE.

Daftar Rujukan

- [1] M. S. Alrajak, I. Ernawati, and I. Nurlaili, "Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K- Nearest Neighbor (K-NN)," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 1, no. 2, pp. 110–122, 2020.
- [2] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [3] T. Prasetyo, H. Zakaria, and P. Wiliantoro, "Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier," *OKTAL : Jurnal Ilmu Komputer dan Sains*, vol. 1, no. 6, pp. 573–582, 2022, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [4] H. B. Tambunan and T. W. D. Hapsari, "Analisis Opini Pengguna Aplikasi New PLN Mobile Menggunakan Text Mining," *Petir*, vol. 15, no. 1, pp. 121–134, 2021, doi: 10.33322/petir.v15i1.1352.
- [5] T. Prasetyo, H. Zakaria, and P. Wiliantoro, "Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier," *OKTAL : Jurnal Ilmu Komputer dan Sains*, vol. 1, no. 6, pp. 573–582, 2022, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [6] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, 2018, doi: 10.26418/jp.v4i2.27526.
- [7] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Informatika*, vol. 3, no. 1, 2018, doi: 10.35314/isi.v3i1.335.

- [8] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal Teknologi Informasi dan Komunikasi (TIKomsin)*, vol. 8, no. 2, Oct. 2020, doi: 10.30646/tikomsin.v8i2.500.
- [9] E. Ogi, I. Pratiwi1, and W. Yustanti2, "Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi Kasus Ovo dan Dana)," *Jeisbi*, vol. 02, no. 03, 2021.
- [10] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram," 2018. [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [11] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis Smote," *Jurnal Teknik Komputer AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [12] S. Kurniawan, W. Gata, D. Puspitawati, Nurmalasari, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," *Jurnal Resti*, vol. 3, no. 2, pp. 176–183, 2019, doi: <https://doi.org/10.29207/resti.v3i2.935>.
- [13] I. Santoso, W. Gata, and A. Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *Jurnal Resti*, vol. 3, no. 3, pp. 364–370, 2019, doi: <https://doi.org/10.29207/resti.v3i3.1084>.
- [14] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021, doi: <https://doi.org/10.24246/aiti.v18i2.173-184>.
- [15] R. Ariyanti, A. Saepudin, E. Fitriani, R. Permana, and D. Saefudin, "Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway," *JURNAL TEKNIK KOMPUTER AMI BSI*, vol. 5, no. 2, pp. 227–234, 2019, doi: <https://doi.org/10.31294/jtk.v5i2.5406>.
- [16] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.