Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Text Mining untuk Sentimen Analisis dengan Metode Naïve Bayes, SMOTE, N-Gram dan AdaBoost Pada Twitter CommuterLine

Andreyana Pratama Putra[1], Yuda Pratama[2], Eka Kharisma Krisnadi[3], Indah Purnamasari[4], Dedi Dwi Saputra[5]
[1,2,3,4,5]Information System Study Program, Universitas Nusa Mandiri, Jakarta Timur, Indonesia
e-mail: 11211878@nusamandiri.ac.id[1], 11212257@nusamandiri.ac.id[2], 11211756@nusamandiri.ac.id[3], indah.ihi@nusamandiri.ac.id[4] dedi.eis@nusamandiri.ac.id[5]

Abstract

In the current era, the development of information technology and social media is growing rapidly so that it can provide updated information and various kinds of public opinion. Many internet users in Indonesia use social media for various purposes, such as seeking information and expressing opinions through social media. One of the social media that is widely used by internet users in Indonesia is Twitter. Twitter users can provide information in the form of comments, criticisms, or suggestions for Comutterline services more quickly and easily. Sentiment analysis can help provide an overview of public perception by grouping opinions into positive and negative categories for Commuterline services. Conducting sentiment analysis based on comments or Tweets from the community on Twitter Commuterline to determine the performance of the Naïve Bayes Classifier algorithm, Synthetic Minority Over-sampling Technique (SMOTE), AdaBoost, and N-Gram so that machine learning implementation can help identify public opinion conveyed through Twitter automatically into positive and negative categories. The use of the Naïve Bayes Classifier, Synthetic Minority Over-sampling Technique (SMOTE), AdaBoost, and N-Gram methods which are considered better to generate predictions on tweets sent by CommuterLine users.

Keywords: *CommuterLine, Naive Bayes, SMOTE, AdaBoost, N-Gram*

Abstrak

Di era saat ini perkembangan teknologi informaasi dan media sosial sangat berkembang pesat sehingga dapat menyediakan informasi terupdate serta berbagai macam opini publik. Pengguna internet di Indonesia banyak yang memanfaatkan sosial media untuk berbagai kepentingan seperti mencari informasi dan menyampaikan opini lewat media sosial. Salah satu media sosial yang banyak digunakan oleh pengguna internet di Indonesia adalah Twitter. Pengguna Twitter dapat memberikan informasi berupa komentar, kritik, atau saran terhadap pelayanan Comutterline dengan lebih cepat dan mudah. Analisa sentimen dapat membantu memberikan gambaran bagaimana persepsi masyarakat dengan cara mengelompokan opini menjadi kategori positif dan negatif kepada pelayanan Commuterline. Melakukan sentiment analisis berdasarkan komentar atau Tweet masyarakat pada Twitter Commuterline untuk mengetahui performa algoritma Naïve Bayes Classifier, Synthetic Minority Over-sampling Technique (SMOTE), AdaBoost, dan N-Gram agar implementasi machine learning dapat membantu mengidentifikasi opini masyarakat yang disampaikan melalui Twitter secara otomatis menjadi kategori positif dan negative. Penggunaan metode Naïve Bayes Classifier, Synthetic Minority Over-sampling Technique (SMOTE), AdaBoost, dan N-Gram yang dinilai lebih baik untuk menghasilkan prediksi pada tweet yang dikirimkan oleh pengguna CommuterLine.

Kata kunci: *CommuterLine, Naive Bayes, SMOTE, AdaBosst, N-Gram*

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

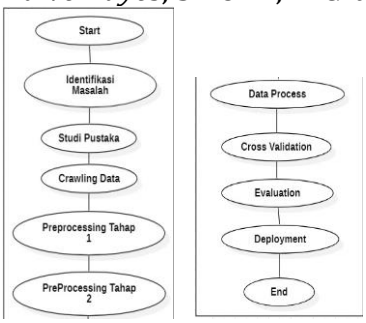
1. PENDAHULUAN

Media sosial Twitter menjadi salah satu media sosial yang di minati masyarakat dunia tidak terkecuali di indonesia. Banyak masyarakat indonesia memanfaatkan media sosial Twitter untuk berbagi informasi dan mencari informasi yang saat ini sedang menjadi topik utama di dunia [1]. Analisa sentimen dapat membantu memberikan gambaran bagaimana persepsi masyarakat dengan cara mengelompokan opini menjadi kategori positif dan negatif kepada pelayanan Commuterline. Analisis sentimen atau bisa disebut opinion mining dalam pengertian secara luas mengacu pada bidang komputasi linguistik, pengolahan bahasa alami dan text mining. Secara umum, tujuannya adalah untuk menentukan attitude dari pembicara ataupun penulis yang berhubungan dengan topik tertentu [2].

Menurut Jiawei Han, Micheline Kamber, dan Jian Pei Data mining yaitu proses menemukan pola dan pengetahuan yang menarik dari sejumlah besar data. Sumber data bisa meliputi basis data, data warehose, Web, repositori informasi lain, atau data yang dialirkan ke sistem secara dinamis[3]. Saat ini opini yang disampaikan melalui twitter ke CommuterLine masih dianalisis secara manual, karena belum ada alat yang dapat memprediksi kalimat dalam sebuah tweet, sehingga diperlukan alat yang dapat membantu dalam membagi klasifikasi antara kalimat yang mengandung narasi Positif dan Negatif. Hal ini dikarenakan keterbatasan teknologi dan literatur yang membahas tentang text mining dalam mendeteksi pola karakteristik humaniora, khususnya analisis sentimen dalam Bahasa Indonesia. *Naïve Bayes Classifier* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayes classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [4], dengan menambahkan Teknik SMOTE, N-Gram dan *AdaBoost* untuk mengetahui seberapa efektif metode dan teknik tersebut dalam memprediksi tweet dari pengguna, Dalam hal ini kami mengambil studi kasus pada akun Twitter CommuterLine.

2. METODOLOGI PENELITIAN

Berikut tahapan penelitian Text Mining untuk Analisis Sentimen Pelanggan Menggunakan *Naive Bayes*, SMOTE, N-Gram dan *AdaBoost* :



Gambar 1. Tahapan Penelitian

2.1. Identifikasi Masalah

Pada tahapan ini dilakukan identifikasi masalah berdasarkan pengamatan langsung atau observasi pada akun twitter @CommuterLine untuk mengetahui sentimen analisis masyarakat terhadap pelayanan CommuterLine.

2.2. Studi Pustaka

Setelah melakukan identifikasi masalah penulis melakukan Studi Pustaka melalui penelitian yang sudah ada sebelumnya terkait dengan *machine learning*, data mining, serta layanan lainnya terkait dengan penelitian yang akan dilakukan.

2.3. *Crawling* Data

Pada tahapan ini dilakukan pengumpulan atau *colleecting* data melalui aplikasi *RapidMiner* dengan menggunakan koneksi ke akun Twitter dengan mengkoneksikan Twitter *API* melalui akun Twitter *Developer*. Data Twitter dapat diakses melalui *API REST* (*Representational State Transfer*) Twitter yang telah disediakan oleh pihak Twitter dengan terlebih dahulu mengajukan permintaan kepada pihak Twitter untuk memperoleh akses data dari Twitter dengan melakukan pendaftaran sebagai akun *developer*.

2.4. *Preprocessing* Tahap 1

Pada tahapan ini dilakukan pengolahan data melalui aplikasi web Gata Framework terdapat 5 tahapan yaitu:

a) *Annotation Removal*
   Pada *website* Gata *Framework* dilakukan untuk menghilangkan mention atau tanda @ pada kalimat tweet sehingga kalimat yang dihasilkan tidak memiliki tanda @.
b) *Transformation Remove* URL
   Pada proses ini, URL dihilangkan pada kalimat yang terdapat pada *tweet* di *web* Gata *Framework*.
c) *Tokenizasion Regular Expression* (Regexp)
   Pada tahap ini, kalimat-kalimat pada dataset dipecah menjadi kata-kata pada website Gata Framework.
d) Indonesian *Stemming*
   Pada proses ini kalimat yang mengandung imbuhan dihilangkan sehingga kata yang mengandung imbuhan tersebut menjadi kata dasar pada *website* Gata *Framework*.
e) Indonesian *Stop Word Removal*
   Proses terakhir pada preprocessing fase 1 adalah melakukan penghapusan Stopword Bahasa Indonesia pada *website* Gata *Framework*.

2.5. *Preprocessing* Tahap 2

Pada tahap ini akan dilakukan operasi beberapa fungsi, metode dan operator dalam pengolahan data yaitu *read* excel, yaitu membaca data yang telah dilakukan pada *preprocessing* tahap 1, kemudian dilakukan Nominal to *Text* yaitu melakukan konversi atribut nominal menjadi atribut string.

2.6. Evaluation

Pada tahapan ini dilakukan evaluasi pada performance vector dari *machine learning* dengan acuan *Confusion Matrix* berupa tingkat akurasi, presisi, *recall*, *AUC* dan dilakukan stemming pada kata-kata yang tidak bernilai atau mengurangi tingkat performansi pada *performance vector*. Berikut penjelasan mengenai:

a) *Accuracy*
Merupakan rasio prediksi benar ("*positive*" dan "*negative*") dengan keseluruhan data. Akurasi akan menjawab pertanyaan berapa persen positif dan negatif yang disampaikan melalui twitter terhadap akun tweet @CommuterLine.
Accuracy = (TP+TN) / (TP + FP + FN + TN)

b) *Precission*
Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Precission akan menjawab pertanyaan "berapa persen tweet "*positive*" yang disampaikan masyarakat terhadap twitter @CommuterLine.
Precisson = (TP) / (TP + FP)

c) *Recall (Sensitivity)*
Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall akan menjawab pertanyaan " berapa persen tweet "*possitive*" yang disampaikan masyarakat terhadap twitter @CommuterLine.
Recall = (TP) / (TP + FN)

d) *AUC (Area Under The Curve)*
AUC (Area Under The Curve) membuat kemudahan dalam membandingkan model satu dengan model lainnya, *AUC* adalah luas area dibawah *curve ROC (Receiver Operating Characteristics)* atau *integral* dari fungsi *ROC*.

2.7. *Deployment*

Pada tahap *deployment* akan dilakukan input *database* dan pengaplikasian *machine learning* ke dalam aplikasi php dari *Gata Framework* sehingga dapat langsung memprediksi tweet yang ditujukan kepada akun Twitter *CommuterLine*.

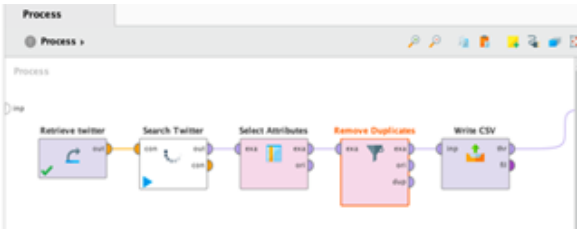
Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

3. HASIL DAN PEMBAHASAN

Berdasarkan tahapan dan metode penelitian yang sudah dilakukan, maka selanjutnya akan menjelaskan hasil dan pembahasan terhadap penelitian yang dikerjakan adalah sebagai berikut:

3.1. Crawling Data Twitter

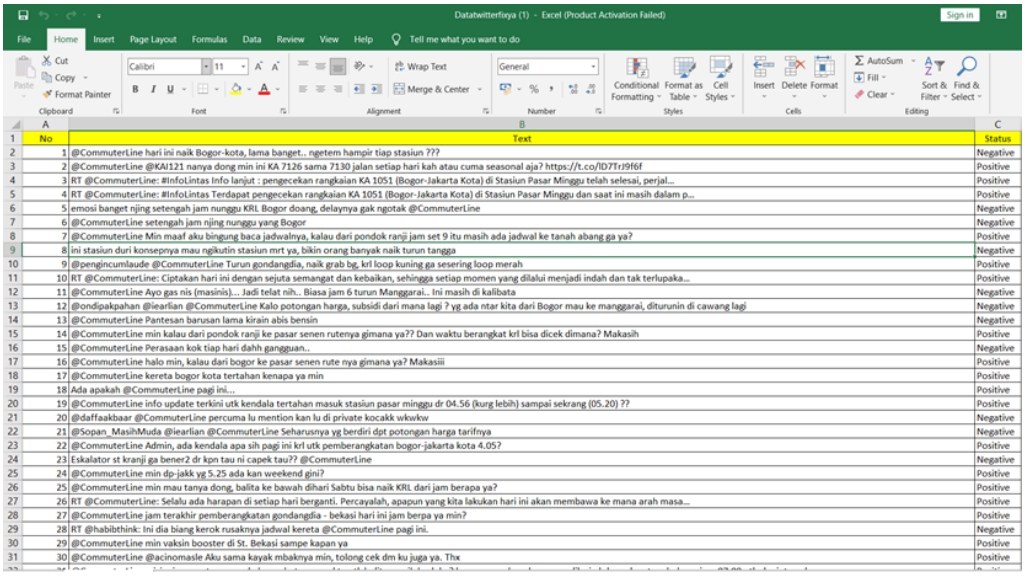
Tahapan ini dilakukan pengumpulan data dengan cara pengambilan data tweet yang tujukan dari akun twitter @CommuterLine dengan aplikasi *RapidMiner* menggunakan operator *Retrieve* Twitter yaitu untuk melakukan koneksi ke Twitter dengan akun Twitter *developer*, kemudian menggunakan operator *search* Twitter dengan *keyword* @CommuterLine, ditambahkan juga operator *remove duplicate* untuk menghapus kalimat atau tweet yang sama, kemudian operator *write to csv* untuk *export* file ke dalam bentuk .csv. Dataset yang di ambil dalam tahapan ini adalah 500 dataset.



Gambar 2. Proses Pengumpulan Data

3.2. Labeling

Pada tahapan ini dilakukan labeling terhadap 500 dataset yang dilakukan oleh publik atau masyarakat dan menghasilkan data dengan klasifikasi "*Positive*" dan "*Negative*".

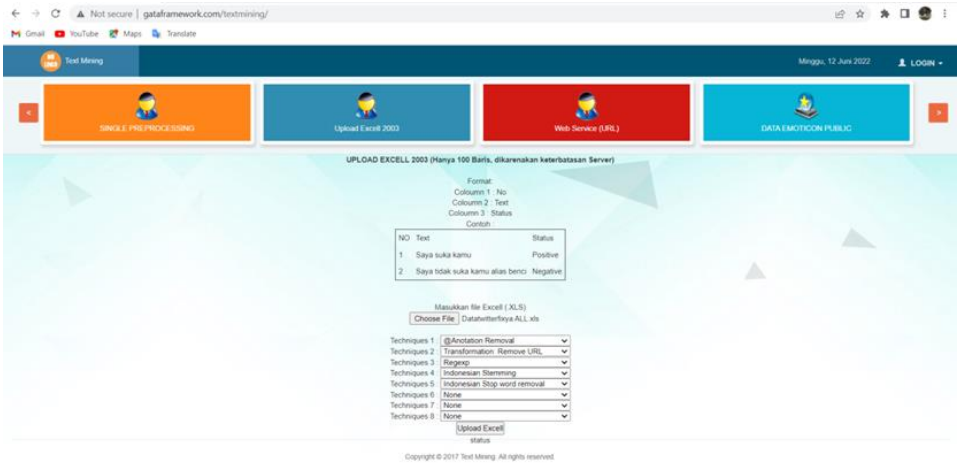


No	Text	Status
1	@CommuterLine hari ini naik Bogor-kota, lama banget.. ngetem hampir tiap stasiun ???	Negative
2	@CommuterLine @KAI121 nanya dong min ini KA 7126 sama 7130 jalan setiap hari kah atau cuma seasonal aja? https://t.co/ID7TrJ9f6f	Positive
3	RT @CommuterLine: #InfoLintas Info lanjut : pengecekan rangkaian KA 1051 (Bogor-Jakarta Kota) di Stasiun Pasar Minggu telah selesai, perjal...	Positive
4	RT @CommuterLine: #InfoLintas Terdapat pengecekan rangkaian KA 1051 (Bogor-Jakarta Kota) di Stasiun Pasar Minggu dan saat ini masih dalam p...	Positive
5	emosi banget njing setengah jam nunggu KRL Bogor doang, delaynya gak ngotak @CommuterLine	Negative
6	@CommuterLine setengah jam njing nunggu yang Bogor	Negative
7	@CommuterLine Min maaf aku bingung baca jadwalnya, kalau dari pondok ranji jam set 9 itu masih ada jadwal ke tanah abang ga ya?	Positive
8	ini stasiun duri konsepnya mau ngikutin stasiun mrt ya, bikin orang banyak naik turun tangga	Negative
9	@pengincumlaude @CommuterLine Turun gondangdia, naik grab bg, krl loop kuning ga sesering loop merah	Positive
10	RT @CommuterLine: Ciptakan hari ini dengan sejuta semangat dan kebaikan, sehingga setiap momen yang dilalui menjadi indah dan tak terlupaka...	Positive
11	@CommuterLine Ayo gas nis (masinis)... Jadi telat nih.. Biasa jam 6 turun Manggarai.. Ini masih di kalibata	Negative
12	@ondipakpahan @iearlian @CommuterLine Kalo potongan harga, subsidi dari mana lagi ? yg ada ntar kita dari Bogor mau ke manggarai, diturunin di cawang lagi	Negative
13	@CommuterLine Pantesan barusan lama kirain abis bensin	Negative
14	@CommuterLine min kalau dari pondok ranji ke pasar senen rutenya gimana ya?? Dan waktu berangkat krl bisa dicek dimana? Makasih	Positive
15	@CommuterLine Perasaan kok tiap hari dahh gangguan..	Negative
16	@CommuterLine halo min, kalau dari bogor ke pasar senen rute nya gimana ya? Makasiii	Positive
17	@CommuterLine kereta bogor kota tertahan kenapa ya min	Positive
18	Ada apakah @CommuterLine pagi ini...	Positive
19	@CommuterLine info update terkini utk kendala tertahan masuk stasiun pasar minggu dr 04.56 (kurg lebih) sampai sekrang (05.20) ??	Positive
20	@daffaakbaar @CommuterLine percuma lu mention kan lu di private kocakk wkwkw	Negative
21	@Sopan_MasihMuda @iearlian @CommuterLine Seharusnya yg berdiri dpt potongan harga tarifnya	Negative
22	@CommuterLine Admin, ada kendala apa sih pagi ini krl utk pemberangkatan bogor-jakarta kota 4.05?	Positive
23	Eskalator st kranji ga bener2 dr kpn tau ni capek tau?? @CommuterLine	Negative
24	@CommuterLine min dp-jakk yg 5.25 ada kan weekend gini?	Positive
25	@CommuterLine min mau tanya dong, balita ke bawah dihari Sabtu bisa naik KRL dari jam berapa ya?	Positive
26	RT @CommuterLine: Selalu ada harapan di setiap hari berganti. Percayalah, apapun yang kita lakukan hari ini akan membawa ke mana arah masa...	Positive
27	@CommuterLine jam terakhir pemberangkatan gondangdia - bekasi hari ini jam berpa ya min?	Positive
28	RT @habibthink: Ini dia biang kerok rusaknya jadwal kereta @CommuterLine pagi ini.	Negative
29	@CommuterLine min vaksin booster di St. Bekasi sampe kapan ya	Positive
30	@CommuterLine @acinomasle Aku sama kayak mbaknya min, tolong cek dm ku juga ya. Thx	Positive

Gambar 3. Hasil Labeling

3.3. *Preprocessing* Tahap 1

Dari hasil labeling terhadap dataset yang telah dikumpulkan tersebut kemudian dilakukan proses *preprocessing* tahap I dengan membagi file menjadi sebanyak 10 file, masing-masing file berisi 50 dataset dikarenakan keterbatasan server pada aplikasi *gata framework*. Pada tahapan ini *preprocessing* dan *cleansing* dengan menggunakan beberapa teknik pada *website gataframework* untuk membuat *design preprocessing* dan *cleansing* pada dataset *local*. *@Annotation Removal, Transformation Remove URL, Regexp, Indonesian Stemming, Indonesian Stop Word Removal*.



Gambar 4. *Preprocessing* Tahap 1

Hasil *preprocessing* pada aplikasi *Gata Framewok* dimana terdapat penghilangan tanda @, penghapusan URL, pencacahan kalimat, penghapusan imbuhan dan penghapusan kata yang tidak bermakna.

Tabel 1. Hasil *Preprocessing* menggunakan *Website gataframework*

No	Text	Status	@Anotation Removal	Transformation: Remove URL	Regexp	Indonesian Stemming	Indonesian Stop Word Removal
1.	@CommuterLine hari ini naik Bogor-kota, lama banget.. ngetem hampir tiap stasiun ???	Negative	hari ini naik bogor-kota, lama banget.. ngetem hampir tiap stasiun ???	hari ini naik bogor-kota, lama banget.. ngetem hampir tiap stasiun ???	hari ini naik bogor kota lama banget ngetem hampir tiap stasiun	hari ini naik bogor kota lama banget ngetem hampir tiap stasiun	bogor kota banget ngetem stasiun
2.	@CommuterLine @KAI121 nanya dong min ini KA	Positive	nanya dong min ini ka 7126 sama 7130 jalan setiap	nanya dong min ini ka sama jalan setiap hari	nanya dong min ini ka sama	nanya dong min ini ka sama	nanya min ka jalan kah seasonal

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

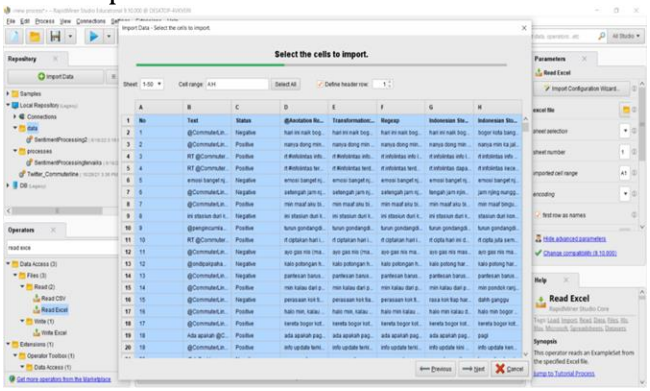
No	Text	Status	@Anotation Removal	Transformation: Remove URL	Regexp	Indonesian Stemming	Indonesian Stop Word Removal
	7126 sama 7130 jalan setiap hari kah atau cuma seasonal aja? https://t.co/lD7TrJ9f6f		hari kah atau cuma seasonal aja?	kah atau cuma seasonal aja	jalan tiap hari kah atau cuma seasonal aja	jalan tiap hari kah atau cuma seasonal aja	
....	....	....	.....	.....	.....		.....
500	@CommuterLine Kalo terakhir ke Cikarang dari manggarai terakhir jam berapa ya min?	Positive	kalo terakhir ke cikarang dari manggarai terakhir jam berapa ya min?	kalo terakhir ke cikarang dari manggarai terakhir jam berapa ya min?	kalo terakhir ke cikarang dari manggarai terakhir jam berapa ya min	kalo akhir ke cikarang dari manggarai akhir jam berapa ya min	kalo cikarang manggarai jam min

3.4. *Preprocessing* Tahap 2

Pada tahapan *preprocessing* tahap 2 dilakukan beberapa tahapan untuk pengolahan dataset yang telah di *preprocessing* pada tahap 1. Tahapan *preprocessing* tahap 2 dilakukan beberapa proses sebagai uji coba pada dataset sehingga menghasilkan data yang lebih akurat untuk pengimplementasian *machine learning*.

a) *Import* Data

Pada tahapan ini dilakukan dengan cara melakukan *import* data yang sudah di *preprocessing* pada tahap 1 ke aplikasi *Rapid Miner* dengan menggunakan operator *read* excel.

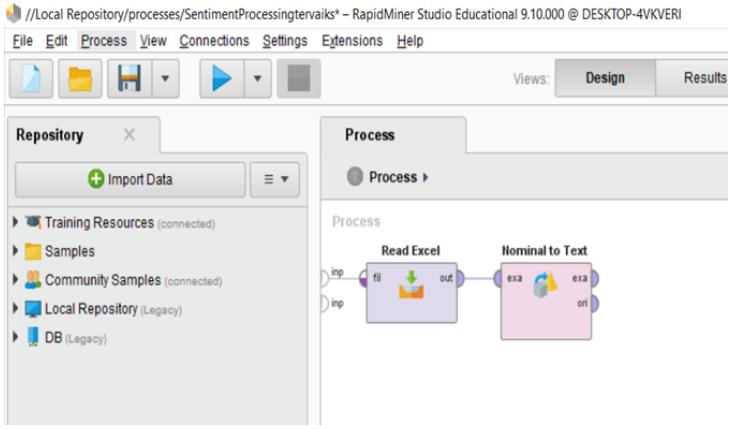


Gambar 5. Proses *Import* Data Excel

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

b) *Nominal To Text*

Pada tahapan ini berfungsi untuk mengubah jenis atribut nominal yang dipilih menjadi *text* dan memetakan semua nilai atribut ke string yang sesuai.



Gambar 6. Operator *Nominal to Text*

c) *Process Document From Data*

Penggunaan operator *Process Document from Data*, dilakukan beberapa proses yang digunakan untuk membersihkan data agar menjadi vektor yang dapat digunakan sebagai perhitungan algoritma diantaranya yaitu:

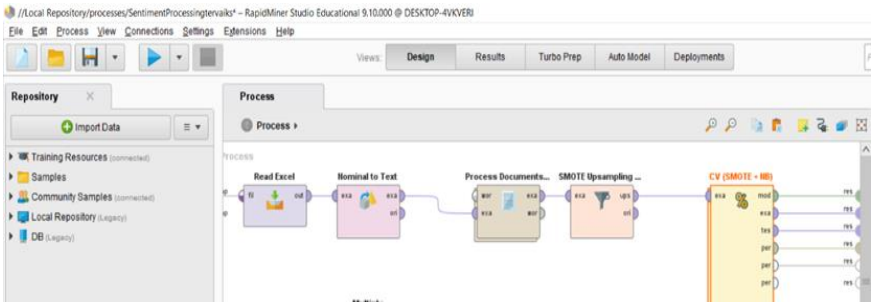
1. *Transform case* pada tab parameter kita ubah menjadi *transform* to *lower case* yang buerfungsi untuk mengubah karakter menjadi huruf kecil.
2. *Tokenize* berfungsi untuk menghilangkan tanda baca, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf dan memecah kalimat menjadi perkata.
3. *Filter Tokens (by Length)* untuk memilih panjang karakter yang 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.
4. *Filter Stopwords (Dictionary)* berfungsi untuk membuang kata yang memiliki informasi yang rendah dalam sebuah teks.
5. *Steming (dictionary)* berfungsi untuk mereduksi kata berimbuhan menjadi kata dasar. Dengan proses *stemming*, kata yang dimasukkan ke dalam *index* adalah dalam bentuk umum, sehingga dapat menghasilkan dokumen yang lebih relevan.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

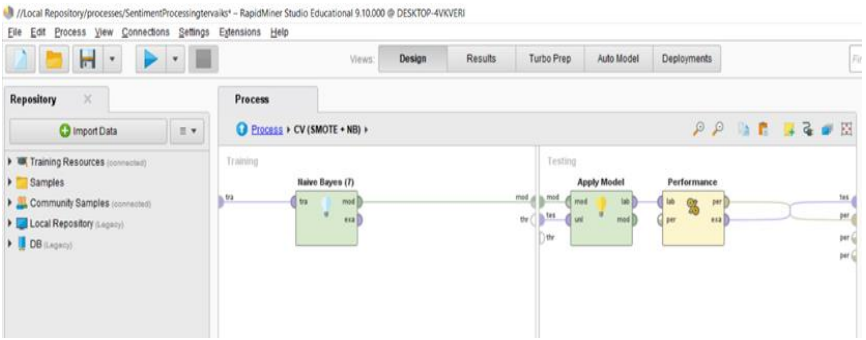
Gambar 7. Operator *Process Document*

d) *Cross Validation*

Fungsi operator ini adalah untuk melakukan *training* dan juga *testing* pada *dataset* yang di dalamnya terdapat operator *algoritma Naive Bayes* untuk *training* data, sedangkan *apply model*, dan *performance*.

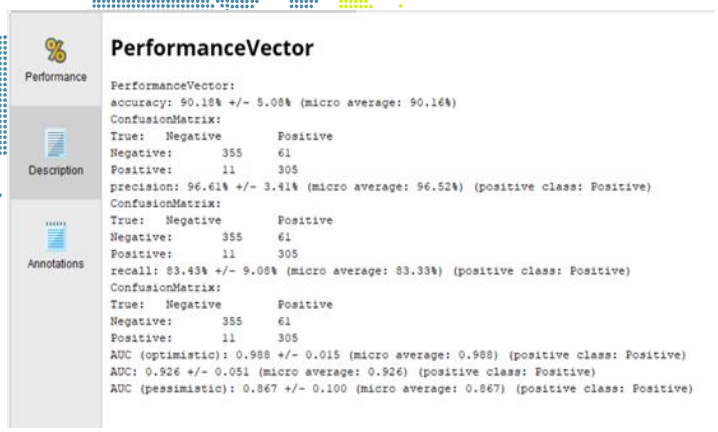


Gambar 8. Operator *Cross Validation*



Gambar 9. *Training* dan *Testing Operatator Cross Validation*

Merupakan *testing* terhadap data untuk mengetahui tingkat *accuracy*, *performance*, *recall* dan *AUC (Area Under Curve)* pada *dataset*.



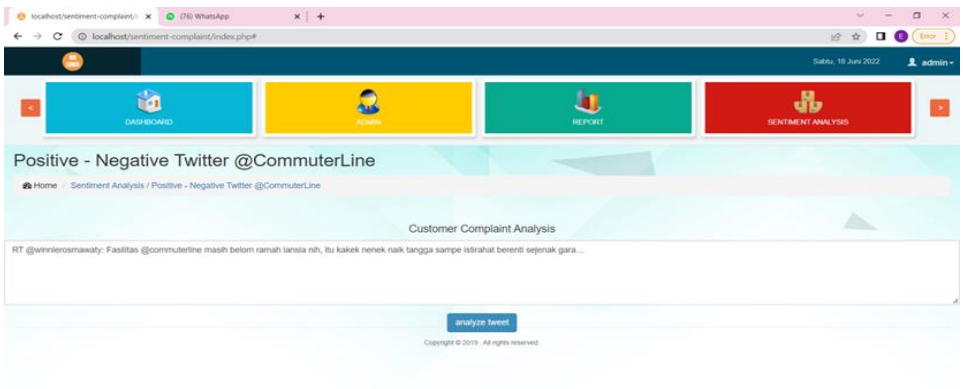
Gambar 10. Hasil Penambahan Teknik *Naïve Bayes*, SMOTE, N-Gram dan *AdaBoost*

Dan didapatkan hasil perbandingan menggunakan Teknik *Naïve Bayes*, SMOTE, N-Gram dan *AdaBoost* adalah sebagai berikut:

1. *Accuracy* : 90.18 %
2. *Precission* : 96.61 %
3. *Recall* : 83.43%
4. *AUC* : 0.988 %

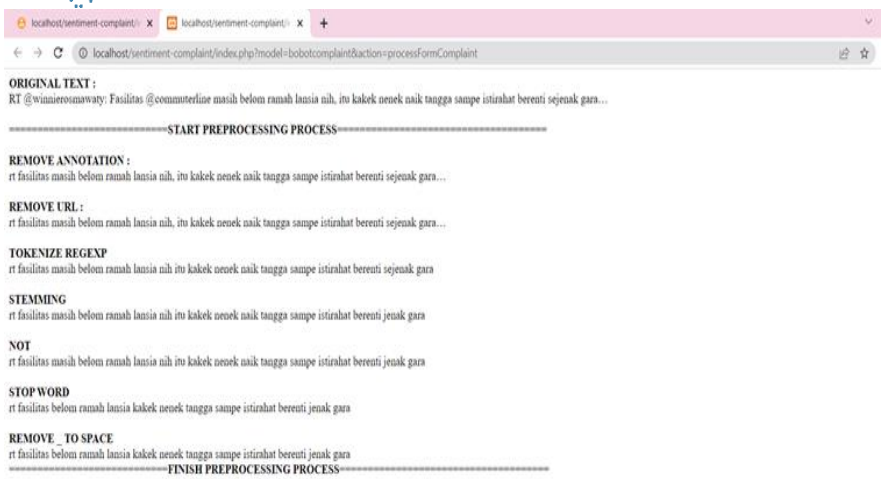
3.5. *Deployment*

Pada tahap deployment dilakukan input database dari dataset yang telah di *preprocess* pada tahap 1 dan tahap 2, *deployment* juga merupakan penerapan *machine learning* ke dalam pemograman php dari *Gata Framework* sehingga dapat langsung memprediksi tweet yang ditujukan ke CommuterLine Akun Twitter. *Depolyment* berdasarkan hasil evaluasi proses pengujian model antara algoritma *Naive Bayes* dengan model algoritma *Naive Bayes* ditambah dengan fitur *Syntethic Minority Over Sampling Technique Method* (SMOTE), N-Gram dan *AdaBoost*. Karena bobot ini akan digunakan dalam memprediksi kalimat atau tweet di aplikasi.



Gambar 11. *Get API* Twitter

Hasil *deployment* dalam mengambil data dari Twitter dari akun @CommuterLine dengan menggunakan *API* (*Application Programming Interface*). Kemudian pada tahapan selanjutkan ketika dilakukan tekan pada tombol *Analyze* Tweet akan menuju ke tahap *text processing*.



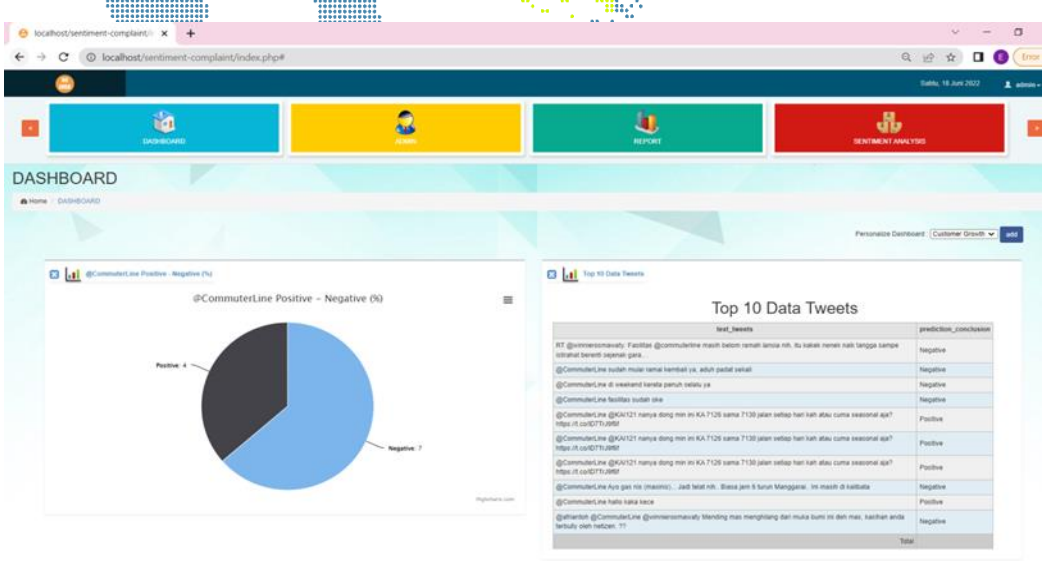
Gambar 12. *Preprocessing*

Merupakan tahapan *Preprocessing* dan *cleansing* pada tweet tersebut, dengan menggunakan teknik *Remove @annotation, Remove URL, Tokenize Regexp, Stemming, Not Transformation Negatif, Stopword, Remove _ to Spac.=*

Gambar 13. Perhitungan Bobot Kata

Merupakan tahapan penilaian bobot kata pada tweet yang ditujukan kepada akun @CommuterLine, hasil penghitungan tersebut menghasilkan kesimpulan *Negative* dikarenakan bobot *negative* dari kata pada tweet tersebut lebih besar daripada *Positive*.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 6 Nomor 2, September 2022, pp. 961-973
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti



Gambar 14. *Result Summary Positive* dan *Negative*

Hasil keseluruhan dari kalimat tweet *Positive* dan *Negative* yang ditujukan kepada akun @CommuterLine dan top 10 data *Tweets* yang ditujukan kepada akun @CommuterLine berdasarkan bobot kata yang telah dilakukan pada *preprocessing* menampilkan *graphic* untuk mengetahui perbandingan antara kelas *Positive* dan *Negative*.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai permasalahan pengkategorian tweet dengan sentiment "*Positive*" dan "*Negative*" terhadap pelayanan PT KAI Commuter, di sosial media Twitter dengan menggunakan *mention* @CommuterLine dapat ditarik kesimpulan yaitu, pendekatan dengan menggunakan metode *text mining* dan pemodelan *algoritma Naïve Bayes* yang ditambahkan dengan *feature selection SMOTE*, *N-Gam*, dan *Adaboost* terbukti efektif dalam hal klasifikasi pengkategorian narasi tweet "*Positive*" dan "*Negative*", hal ini didukung dengan dihasilkannya kategori tertinggi pada hasil penelitian *algoritma Naïve Bayes* yang ditambahkan dengan *feature selection SMOTE*, *N-Gram*, dan *Adaboost* didapati tingkat *Acuraccy* 90.18%, *Precission* 96.61%, *Recall* 83.43%, *AUC* 0.988% .

DAFTAR PUSTAKA

[1] N. Hannani, "Pengertian Twitter," *2019*, 2019. https://www.nesabamedia.com/ (accessed Apr. 25, 2022).

[2] Y. Cahyono and S. Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic," *J. Inform. Univ. Pamulang*, vol. 4, no. 3, p. 87, 2019,

doi: 10.32493/informatika.v4i3.3186.

[3] E. Miranda and Julisar, "Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Mengklasifikasikan Pelanggan," *Infotech*, vol. 4, no. 9, pp. 6–12, 2018.

[4] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.