

Analisis Sentimen Masyarakat Terhadap Kasus Pembobolan Data Nasabah Bank BSI Pada Twitter Menggunakan Metode Random Forest Dan Naïve Bayes

Desti Mualfah¹, Ananda Prihatin², Rahmad Firdaus³, Sunanto⁴

^{1,2,3,4} Teknik Informatika, Ilmu Komputer, Universitas Muhammadiyah Riau

¹destimualfah@umri.ac.id, ²190401261@umri.ac.id, ³rahmadfirdaus@umr.ac.id, ⁴sunanto@umri.ac.id

Abstract

Recently, Indonesia has been shocked by the data breach case that hit Bank Syariah Indonesia (BSI) in May 2023. This case has attracted public reaction and attention, especially on Twitter social media, with various reaction and responses. Most of the public supports the system improvement efforts carried out by BSI, but quite a few also criticize and insult the bank because of the slow repair process carried out after their system was infiltrated by hackers. Therefore, this research aims to conduct sentiment analysis to determine the response of the Indonesian people regarding the case of customer data breach at BSI Bank, so that the community's reactions and responses can be grouped into positive, negative or neutral responses. To achieve this goal, the Naïve Bayes and Random Forest methods were used, because these two methods have been widely used in text data classification and have a high level of accuracy. The dataset used consists of public responses on Twitter, which were taken through a crawling process, with a total of 809 tweets. The results of this research show that the Naive Bayes method has an accuracy rate of 81%, while the Random Forest method has an accuracy rate of 78%.

Keywords: sentiment analysis, random forest, naive bayes, BSI bank, twitter.

Abstrak

Belakangan ini, Indonesia telah dihebohkan dengan adanya kasus pembobolan data yang menimpa Bank Syariah Indonesia (BSI) pada bulan Mei 2023. Kasus ini telah menarik perhatian masyarakat, terutama di media sosial *Twitter*, dengan berbagai macam reaksi dan tanggapan yang diberikan. Sebagian masyarakat mendukung upaya perbaikan sistem yang dilakukan oleh BSI, namun tidak sedikit juga yang mengkritik dan menghujat bank tersebut karena lambatnya proses perbaikan yang dilakukan setelah sistem mereka disusupi oleh *hacker*. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen guna mengetahui reaksi dan tanggapan masyarakat Indonesia terhadap kasus pembobolan data nasabah di Bank BSI, sehingga reaksi dan tanggapan masyarakat tersebut dapat dikelompokkan kedalam tanggap yang bersifat positif, negatif, atau netral. Untuk mencapai tujuan tersebut, metode *Naive Bayes* dan *Random Forest* digunakan, karena kedua metode ini telah banyak digunakan dalam klasifikasi data teks dan memiliki tingkat akurasi yang tinggi. *Dataset* yang digunakan terdiri dari tanggapan masyarakat di *Twitter*, yang diambil melalui proses *crawling*, dengan jumlah data mencapai 809 *tweet*. Hasil penelitian ini menunjukkan bahwa metode *Naive Bayes* memiliki tingkat akurasi 81%, sedangkan metode *Random Forest* memiliki tingkat akurasi 78%.

Kata kunci: analisis sentimen, random forest, naive bayes, bank BSI, twitter

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Perkembangan teknologi informasi memiliki dampak besar terhadap sektor ekonomi, budaya, politik, dan hukum. Seiring dengan kemajuan tersebut, banyak orang dapat merasakan manfaat positifnya. Namun, hal ini juga berpotensi menimbulkan kejahatan baru yang dikenal sebagai kejahatan dunia maya melalui internet. Ada individu yang memanfaatkan kelemahan dalam keamanan pemanfaatan teknologi informasi di internet sebagai sarana untuk melakukan kegiatan kriminal kemudian dikenal dengan istilah kejahatan Siber.[1]

Di Indonesia, terdapat beberapa insiden kejahatan di dunia maya seperti penyebaran informasi yang merusak reputasi atau perilaku tidak pantas melalui media sosial, penyalahgunaan kartu kredit, pencurian data, manipulasi data dan pengiriman email yang tidak

diizinkan yang dilakukan oleh pelaku dengan mengambil alih komputer pengembang.[2]

Baru-baru ini pada tanggal 08 Mei 2023, BSI menjadi salah satu target utama. Saat ini, bank BSI sedang menghadapi masalah pencurian data yang dilaporkan oleh JawaPos.com. Pencurian data ini dilakukan oleh geng hacker bernama *LockBit*, kabarnya data yang dicuri mencapai 1,5 TB. Setelah terjadi kasus pencurian data nasabah bank BSI, masyarakat ramai memberikan tanggapan di platform komunikasi *Twitter*. Tanggapan tersebut bisa berupa dukungan, kritik, atau protes terhadap masalah ini. *Twitter* merupakan media komunikasi yang populer dan banyak digunakan oleh masyarakat untuk berekspresi. Unggahan-unggahan di *Twitter* disebut dengan istilah "*tweet*" dan digunakan oleh pengguna lain untuk menginformasikan bahwa mereka memiliki

pembaruan terbaru, mengungkapkan pendapat, atau mengikuti topik yang sedang populer.[3][4][5]

Untuk menemukan berbagai pendapat publik tentang kasus pembobolan data nasabah Bank BSI, *Twitter* dapat menjadi sumber yang baik. Meskipun pendapat yang ditemukan dapat bervariasi antara positif dan negatif, hal ini dapat menimbulkan kekhawatiran. Oleh karena itu, digunakan solusi berupa metode *Random Forest* dan *Naïve Bayes* untuk mengklasifikasikan komentar menjadi positif, netral, atau negatif dengan tujuan menganalisis sentimen publik terkait pembobolan data nasabah Bank BSI. Algoritma *Random Forest* merupakan perkembangan dari metode *CART* yang menerapkan teknik penggunaan *bootstrap aggregating (bagging)* dan pemilihan fitur acak.[6] Klasifikasi *Naïve Bayes* merupakan suatu teknik dalam pembelajaran algoritma data mining yang didasarkan pada metode probabilitas dan statistika..[7][8]

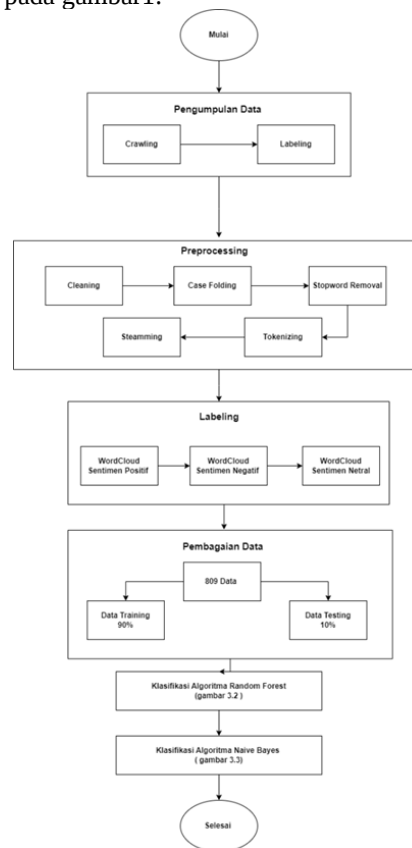
Menurut penelitian yang dilakukan [9] dengan judul “perbandingan metode *Random Forest* dan *Support Vector Machine* pada analisis sentimen *Twitter* (studi kasus : kaburnya selebgram Rachel Vennya dari karantina)”, dari penelitian tersebut didapatkan hasil Algoritma *Random Forest* mendapatkan hasil yang lebih optimal dalam melakukan analisis sentimen. *Random Forest* mencapai nilai *precision*, *recall*, *f1-score*, dan akurasi sebesar 94%, sedangkan *SVM* mendapatkan akurasi sebesar 93%, dengan nilai rata-rata presisi 93%, *recall* 94%, dan *f1-score* 93%. Dengan demikian, dapat disimpulkan bahwa *Random Forest* menghasilkan kinerja yang lebih baik dibandingkan dengan *SVM*.

Pada penelitian yang dilakukan oleh [10] dengan judul “Sentimen Analisis Pengguna *Twitter* pada *Event Flash Sale* Menggunakan Algoritma *K-NN*, *Random Forest*, dan *Naïve Bayes*”. Diperoleh hasil bahwa metode *Naïve Bayes* menunjukkan akurasi yang lebih tinggi pada kedua sampel data, yaitu 83.53% untuk kata kunci “*flash sale*” dan 81.48% untuk kata kunci “*flash sale shopee*”. Dengan demikian, dapat disimpulkan bahwa metode *Naïve Bayes* mampu memberikan akurasi yang lebih baik dalam konteks klasifikasi Analisis Sentimen ini.

Konsep dasar dari penelitian sentimen adalah untuk mengidentifikasi polaritas teks pada dokumen atau kalimat pendek dan mengklasifikasikannya berdasarkan premis ini. Polaritas sentimen kemudian dikategorikan sebagai “positif”, “Negatif” atau “tidak memihak” (netral). Berdasarkan adanya masalah-masalah seperti uraian diatas. Hasil klasifikasi tersebut kemudian dikategorikan sebagai positif, negatif, atau netral, dan selanjutnya akan dijadikan sebagai parameter untuk menilai seberapa baik atau buruk pendapat masyarakat Indonesia terkait dengan pembobolan data nasabah Bank BSI tersebut.

2. Metode Penelitian

Berdasarkan paparan dalam pendahuluan, berikut merupakan metode dan tahapan pada penelitian ini seperti pada gambar1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Langkah awal dari metode penelitian ini adalah pengumpulan data yang diperoleh melalui proses *crawling* pada *platform* media sosial *Twitter*. Pencarian data dilakukan dengan kata kunci “BSI” dalam rentang waktu antara tanggal 13 Mei 2023 hingga 20 Juni 2023. Proses *crawling* data dilakukan dengan menggunakan Bahasa pemrograman Python melalui emulator Google Colab.

2.2 Pengolahan Data

Pengolahan data atau *Pre-processing* adalah langkah dalam proses klasifikasi dengan tujuan untuk membersihkan serta mempersiapkan data sehingga dapat diolah dengan lebih baik pada tahap klasifikasi selanjutnya..[11][12]

Dalam penelitian ini, pengolahan data dapat dilaksanakan dengan empat tahapan, yaitu:

1. Data Cleansing

Pada langkah ini data dibersihkan dari *noise* seperti tanda baca dan karakter yang tidak berhubungan seperti *hashtag*, *RT*, dan karakter yang tidak penting lainnya. Proses pembersihan data melibatkan penghapusan teks berisi awalan (Prefix), tag HTML, tanda baca (*Punctuation*), dan angka menggunakan

ekspresi reguler (regex)[13][14]. mengubah *tweet* dapat juga meningkatkan konsistensi data.

Contoh : "Tingkatkan lagi" menjadi "tingkatkan lagi".

2. Tokenizing

Tokenisasi adalah proses yang dilakukan untuk memisahkan serangkaian kata dalam kalimat, paragraf, atau halaman menjadi token atau potongan kata tunggal, atau kata dasar (*temmed word*). [15][16]

Pada langkah ini kalimat atau dokumen di potong menjadi bagian-bagian, dan memisahkan frase, kata, simbol dan entitas penting lainnya

Contoh: "Tingkatkan lagi" menjadi "[tingkatkan, lagi]".

3. Remove Stopward

Penghapusan *stopword* adalah tahap dalam pengolahan teks yang bertujuan untuk menghilangkan kata-kata yang dianggap tidak relevan atau umumnya tidak memberikan kontribusi signifikan pada hasil parsing. Langkah ini membantu meningkatkan akurasi dan relevansi analisis dengan membuang kata-kata yang umumnya tidak membawa makna khusus atau informasi penting. [17]. Pada langkah ini dilakukan penghapusan stopward atau kata-kata yang tidak unik didalam dokumen. contohnya seperti lata "oleh", "pada", "karena", "di", dll. penghapusan ini bertujuan untuk meningkatkan kinerja analisis sentimen.

Contoh : "Tingkatkan lagi" menjadi "[tingkatkan]".

4. Stemming

Pada tahap ini, dilakukan proses penguraian suatu kata menjadi bentuk kata dasarnya. Proses ini melibatkan penghilangan awalan, akhiran, sisipan, dan konfiks (kombinasi awalan dan akhiran). Metode stemming yang diterapkan adalah stemming Sastrawi, sebuah stemmer yang merupakan pengembangan dari Algoritma Nazief dan Adriani. Awalnya, stemmer ini dibangun menggunakan bahasa pemrograman PHP, namun kemudian berkembang juga dalam bahasa pemrograman Python, Java, C, Go, Ruby.[18][19]

2.3 Labeling

Pada tahap ini data yang sudah diolah kita visualisasi dengan *wordCloud*, Wordcloud adalah suatu sistem yang dapat memvisualisasikan kata-kata dengan menyoroti frekuensi penggunaan kata dalam suatu teks tertulis. [20][21]. *Wordcloud* memberikan gambaran visual yang ringkas dan menarik tentang kata-kata yang paling umum dalam teks. Ini membantu dalam memahami tema, fokus, atau konten utama dalam teks tersebut. *Wordcloud* juga dapat berfungsi sebagai alat komunikasi yang efektif untuk menyampaikan informasi kepada audiens secara visual.

2.4 Implementasi Algoritma Random Forest

Tahap Setelah melakukan analisis sentimen, yang melibatkan pengolahan data yang sudah dilabel, langkah selanjutnya adalah melakukan pengklasifikasian dengan menggunakan metode *Random Forest*. Algoritma *Random forest* adalah sebuah algoritma gabungan antara *decision tree* dilakukan dengan cara mengelompokkan “*weak learners*” bersama-sama untuk membentuk “*strong learners*” melalui teknik gabungan (*ensemble techniques*) [22]. Klasifikasi sentimen dengan menggunakan metode *Random Forest* melibatkan pembangunan model *ensemble tree*. Setiap *decision tree* yang dibangun dalam *Random Forest* dikonstruksi menggunakan subset data latih yang berbeda. Pada setiap pemisahan (split) dari simpul (*node*), digunakan perhitungan Gini Index dengan nilai terkecil dari setiap kelas untuk mendapatkan nilai gain terbesar.[23][24]

2.5 Implementasi Algoritma Naïve Bayes

Dalam tahapan pengklasifikasian algoritma Naive Bayes, jumlah kata yang muncul pada dataset pada setiap kelas menjadi faktor penentu. Algoritma Naive Bayes adalah salah satu algoritma klasifikasi yang sering digunakan dalam bidang *Data Mining* dan *Text Mining*. [25][26]. Pada waktu sebelumnya, teorema Naive Bayes digunakan untuk melakukan estimasi peluang dengan memanfaatkan probabilitas dan statistik. Algoritma Naive Bayes memanfaatkan asumsi yang sangat sederhana (naif) dan bergantung pada setiap kondisi atau kejadian secara terpisah. Ini merupakan ciri utama dari algoritma Naive Bayes. [27]. Data yang akan diklasifikasikan dengan algoritma Naive Bayes adalah data hasil preprocessing yang sudah diberi label sesuai dengan sentimen dari opini. Proses pelabelan dilakukan secara manual untuk meningkatkan keakuratan analisis sentimen. Data ini berupa teks yang akan dihasilkan dalam dua kategori, yaitu teks yang mengandung sentimen positif dan teks yang mengandung sentimen negatif.

2.6 Hasil

Tahap yang terakhir yaitu hasil, dilangkah ini didapatkan hasil berdasarkan pengujian yang sudah dilakukan dengan metode *Random Forest* dan *Naïve Bayes* dan selanjutnya setelah itu dapat dilakukan evaluasi klasifikasi *Random Forest* dan *Naïve Bayes* dengan melakukan pengujian nilai akurasi.

3. Hasil dan Pembahasan

Bagian ini menjelaskan alur penelitian mulai dari pengumpulan dataset hingga evaluasi hasil performa

3.1 Pengumpulan Data

Proses mendapatkan data opini masyarakat *Twitter* melalui *Twitter* dengan bantuan kata kunci berupa hastag yaitu #BSI, "Bank Syariah Indonesia" dan mendapatkan data sebanyak 809 *tweet* dari tanggal 08 Mei 2023 hingga 1 bulan kebelakang yang kemudian data tersebut di simpan kedalam format .csv,

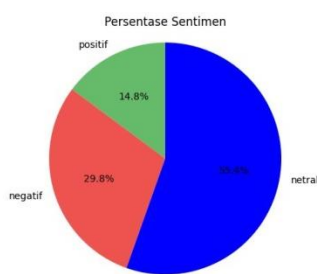
ditampilkan pada gambar hasil scraping data dibawah ini :

URL	Date	Twitter	Retweet	Like	Reply	Quote	Bookmark	Text
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada yang sama kayak aku gak bis...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...
https://twitter.com/infoda/with_replies/1671880484	2023-08-22 12:00:00	Infoda (@infoda) Retweeted @infoda	0	2	7	0	0	guys butuh infoada kayak gak narik tunai bsi e...

Gambar 1. Hasil Crawling Data

3.2 Preprocessing Data

Pada Gambar 3 dapat dilihat hasil yang didapatkan dari data tweet kasus pembobolan data nasabah bank BSI sebanyak 809 tweet , yaitu sentimen netral sebesar 55,4 % , sentimen positif sebesar 14,8% dan sentimen negatif sebesar 29,8%.



Gambar 2. Diagram Perbandingan Sentimen

1) Case Folding

Case folding adalah tahapan dalam pemrosesan teks di mana semua huruf yang berada dalam teks akan diubah menjadi huruf kecil (*lowercase*) atau huruf besar (*uppercase*) untuk mencapai konsistensi dalam teks.

Links	Media	Retweeted	Bookmarks	Username	Kategori	Column1	Column2	Column3	Column4	Column5	text_clean
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada yang sama kayak aku gak bis...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...
https://pbs.twimg.com/media/yg3YAUAAAt7?%3D	None	0.0	0.0	infoda_11	netral	None	None	None	None	None	guys butuh infoada kayak gak narik tunai bsi e...

Gambar 3. Hasil Case Folding

2) Stopword Removal

Stopword removal (penghapusan *stopword*) adalah proses dalam pemrosesan teks di mana kata-kata umum yang memiliki sedikit atau tanpa nilai informasi atau kontribusi dalam analisis teks dihilangkan dari teks yang sedang diproses. *Stopwords* merupakan kata yang sangat sering ditemukan dalam teks.

text_clean	text_StopWord
guys butuh infoada yang sama kayak aku gak bis...	guys butuh infoada kayak gak narik tunai bsi e...
mau tarik tunai tapi bsi eror lagi tidak bisa ...	tarik tunai bsi eror tarik tunaibsi bsieror
id min saya tr uang tadi dari bsi ke dana kok...	id min tr uang bsi dana gak masuk ya slip udah...
ini yang nyambar malah akun nggak jelas begin...	nyambar akun nggak tahukah bsi reply akun meny...
menyorot security system bsi yang letoy reqne...	menyorot security system bsi letoy reqnews bsi...

Gambar 4. Hasil Stopword Removal

3) Tokenizing

Tokenizing adalah proses dalam pemrosesan teks di mana teks atau kalimat dipecah menjadi unit-unit yang lebih kecil yang disebut "*token*". *Token* bisa berupa kata, frasa, simbol, atau entitas lainnya yang memiliki makna dalam konteks tertentu.

text_clean	text_StopWord	text_tokens
guys butuh infoada yang sama kayak aku gak bis...	guys butuh infoada kayak gak narik tunai bsi e...	[guys, butuh, infoada, kayak, gak, narik, tuna, bsi, e...]
mau tarik tunai tapi bsi eror lagi tidak bisa ...	tarik tunai bsi eror tarik tunaibsi bsieror	[tarik, tunai, bsi, eror, tarik, tunaibsi, bsieror]
id min saya tr uang tadi dari bsi ke dana kok...	id min tr uang bsi dana gak masuk ya slip udah...	[id, min, tr, uang, bsi, dana, gak, masuk, ya, slip, udah...]
ini yang nyambar malah akun nggak jelas begin...	nyambar akun nggak tahukah bsi reply akun meny...	[nyambar, akun, nggak, tahukah, bsi, reply, akun, meny...]
menyorot security system bsi yang letoy reqne...	menyorot security system bsi letoy reqnews bsi...	[menyorot, security, system, bsi, letoy, reqnews, bsi...]

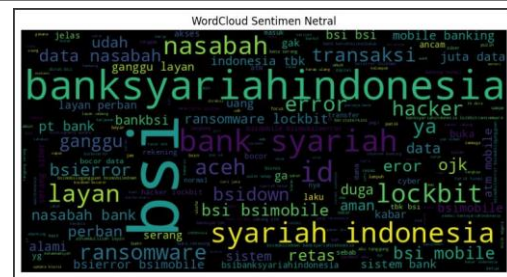
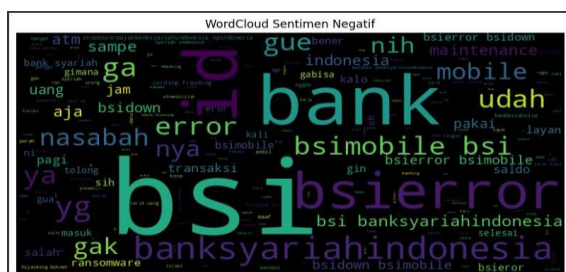
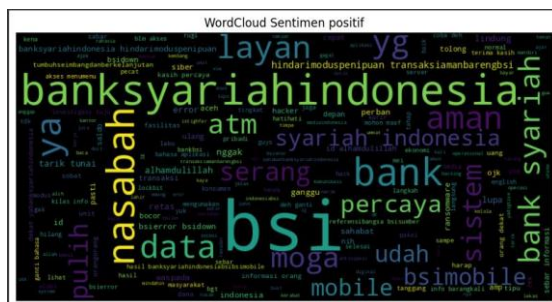
Gambar 5. Hasil Tokenizing

4) Stemming

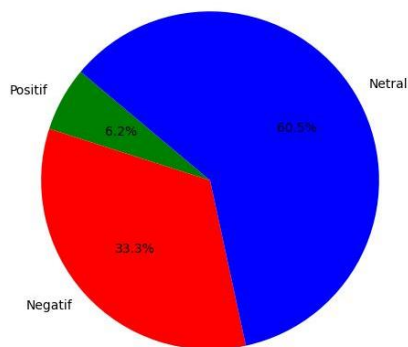
Stemming adalah proses dalam pemrosesan teks di mana bentuk dasar (*root*) suatu kata diambil dengan menghapus afiks atau akhiran kata. Tujuan dari tahapan *stemming* adalah mengurangi kata yang berasal dari bentuk yang sama sehingga mereka dapat dianggap sebagai satu kesatuan atau representasi yang sama.



	precision	recall	f1-score	support
negatif	0.72	0.78	0.75	27
netral	0.86	0.88	0.87	49
positif	1.00	0.40	0.57	5
accuracy			0.81	81
macro avg	0.86	0.69	0.73	81
weighted avg	0.82	0.81	0.81	81



	precision	recall	f1-score	support
negatif	0.81	0.63	0.71	27
netral	0.77	0.90	0.83	49
positif	0.67	0.40	0.50	5
accuracy			0.78	81
macro avg	0.75	0.64	0.68	81
weighted avg	0.78	0.78	0.77	81



Gambar 14. Diagram Sentimen

Pada gambar diatas merupakan gambar diagram dari data yang telah diklasifikasi menggunakan kedua metode tersebut, diagram sentimen presentase dari hasil klasifikasi metode *Naïve bayes* dan *random forest*. dapat dilihat pada gambar 14 hasil dari klasifikasi menggunakan metode *naïve bayes* dan *random forset* yaitu sentimen positif sebesar 6,2%, sentimen negatif sebesar 33,33% dan sentimen netral sebesar 60,5%.

4. Kesimpulan

Berdasarkan data komentar yang diperoleh dari *Twitter* tentang kasus pembobolan data nasabah bank BSI kecenderungan komentar berisi komentar netral. Dari 809 data pengujian dalam penelitian ini, didapat 60,5% komentar dengan sentimen netral, 6,2% komentar dengan sentimen positif dan 33,33% komentar dengan sentimen negatif. Disini dapat disimpulkan bahwa pengguna media sosial *Twitter* di Indonesia bersikap netral terhadap kasus pembobolan data nasabah bank BSI, tetapi lebih banyak yang mengkritik bank BSI mengenai kasus ini dibandingkan dengan yang membela bank BSI.

Hasil pengujian akurasi menggunakan Algoritma Naive Bayes menghasilkan sebesar 81% dan hasil Algoritma Random Forest sebesar 78%, berdasarkan hasil tersebut Algoritma *naïve bayes* lebih optimal digunakan untuk kasus pembobolan data nasabah bank BSI di media sosial *twitter* dibandingkan dengan algoritma *random forest*.

Daftar Rujukan

- [1] I. kadek odie kharisma Putra, I. made adi Darmawan, I. putu gede Juliana, and Indriyani, "Tindakan Kejahatan Pada Dunia Digital Dalam Bentuk Phising," vol. 5, no. 2, pp. 77–83, 2022.
- [2] T. Sutabri, "Analisis Kejahatan Carding Pada BNI 46," vol. 2, no. 1, pp. 33–43, 2023.
- [3] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine Steam sentiment analysis using logistic regression algorithm and support vector machine," vol. 10, pp. 154–164, 2023, doi: 10.37373/tekno.v10i2.440.
- [4] D. Winarso, Yanda Noor Yudha, and Syahril, "Analisis Sentimen Masyarakat Pada Twiter Terhadap Isu Covid-19 Menggunakan Metode Lexicon Based," *J. Fasilkom*, vol. 11,

- no. 2, pp. 97–103, 2021, doi: 10.37859/jf.v11i2.2772.
- [5] S. Jurnal Pipin and H. Kurniawan, "Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM," *J. SIFO Mikroskil*, vol. 23, no. 2, pp. 197–208, 2022, doi: 10.55601/jsm.v23i2.900.
- [6] A. Miftahusalam, adinda febby Nuraini, A. A. Khoirunisa, and P. Hasih, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," *Semin. Nas. Off. Stat.* 2022, pp. 563–572, 2022.
- [7] P. A. Permatasari and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," vol. 20, no. 2, 2021.
- [8] R. A. Wildan, R. A. Rajagede, and R. Rahmadi, "Analisis Sentimen Politik Berdasarkan Big Data dari Media Sosial Youtube: Sebuah Tinjauan Literatur," *Automata*, vol. 2, 2021.
- [9] P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Method On Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Vennya Escape From Quarantine)," vol. 3, no. 1, pp. 141–145, 2022.
- [10] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes," vol. 5, no. September, pp. 651–665, 2021.
- [11] D. Muallfah, Ramadhoni, R. Gunawan, and D. Mulyadipa Suratno, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM," *J. Fasilkom*, vol. 13, no. 01, pp. 72–80, 2023, doi: 10.37859/jf.v13i01.4920.
- [12] A. Arham, E. R. Swedia, M. Cahyanti, and M. R. D. Septian, "Implementasi Sentiment Analysis Pada Opini Masyarakat Indonesia Di Twitter Terhadap Virus Covid-19 Varian Omicron Dengan Algoritma Naïve Bayes, Decision Tree, dan Support Vector Machine," *Sebatik*, vol. 26, no. 2, pp. 565–572, 2022, doi: 10.46984/sebatik.v26i2.1961.
- [13] Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, F. E. N. Saputro, and Nisrina Akbar Rizky P, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *J. Fasilkom*, vol. 13, no. 02, pp. 326–333, 2023, doi: 10.37859/jf.v13i02.5170.
- [14] F. Rizqi Irawan, "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.v5i1.4267.
- [15] Anief Fauzan Rozi and Agus Sidiq Purnomo, "Analisis Sentimen Untuk Respon Masyarakat Terhadap Universitas (Studi Kasus : Universitas Mercu Buana Yogyakarta)," *J. Inf. Syst. Artif. Intell.*, vol. X, no. X, pp. 53–60, 2020.
- [16] G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *Citee*, no. June, pp. 39–43, 2014.
- [17] A. Rozaq, Y. Yunitasari, K. Sussolaikah, E. R. N. Sari, and R. I. Syahputra, "Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 746, 2022, doi: 10.30865/mib.v6i2.3554.
- [18] S. Nanda, D. Muallfah, and D. A. Fitri, "Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 3, no. 2, pp. 210–219, 2022, doi: 10.31102/jatim.v3i2.1592.
- [19] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [20] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [21] M. I. Putri and I. Kharisudin, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace

-
- Tokopedia,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 759–766, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [22] J. A. Nursiyono and Q. Huda, “Analisis Sentimen Twitter Terhadap Perlindungan Data Pribadi Dengan Pendekatan Machine Learning,” *J. Pertahanan Bela Negara*, vol. 13, no. 1, p. 1, 2023, doi: 10.33172/jpbh.v13i1.1877.
- [23] I. N. Arlan, K. Ardika, and I. G. Arta, “Analisis Sentimen Ulasan Pengguna Aplikasi Pelayanan Masyarakat Dengan Menggunakan Algoritma Random Forest,” vol. 1, no. November, pp. 361–372, 2022.
- [24] R. D. Kartiko and K. Umam, “Implementasi Machine Learning pada Alat Pendeteksi Benda untuk Penderita Tunanetra dengan Algoritma Random Forest,” vol. 2, pp. 40–48, 2018.
- [25] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [26] T. Ernayanti, M. Mustafid, A. Rusgiyono, and A. R. Hakim, “Penggunaan Seleksi Fitur Chi-Square Dan Algoritma Multinomial Naive Bayes Untuk Analisis Sentimen Pelanggan Tokopedia,” *J. Gaussian*, vol. 11, no. 4, pp. 562–571, 2023, doi: 10.14710/j.gauss.11.4.562-571.
- [27] D. Era, S. Andryana, and A. Rubhasy, “Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 263–272, 2023.
-