

Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes

Adi Kusuma¹, Agung Nugroho²

^{1,2}Program Studi Teknik Informatika, Universitas Pelita Bangsa
¹adi.kusuma@mhs.pelitabangsa.ac.id, ²agung@pelitabangsa.ac.id

ABSTRAK. Seiring dengan berkembangnya teknologi internet, makin menjamur pula penggunaan sosial media di masyarakat. Sosial media yang masih banyak digemari yaitu twitter. Twitter memberikan layanan kepada penggunanya untuk mengirim dan membaca tweets yang telah dibagikan, sehingga masyarakat lebih memilih menuangkan opininya melalui media sosial daripada menyampaikannya secara langsung. Opini masyarakat yang tertuang dalam media sosial twitter berupa sebuah persepsi, baik itu positif maupun negatif. Melimpahnya opini masyarakat dapat dimanfaatkan sebagai bahan penelitian untuk mencari sebuah informasi. Pemanfaatan informasi tersebut membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu banyak pihak dalam mengambil sebuah keputusan. Untuk mengatasi permasalahan di atas digunakan teknik data mining yang tepat yaitu sentimen analisis. Oleh sebab itu, pada penelitian ini mencoba melakukan analisa sentimen untuk melihat persepsi masyarakat terhadap isu kenaikan tarif dasar listrik pada media sosial twitter menggunakan metode naïve bayes dengan mengklasifikasikan sentimen menjadi positif, negatif dan netral. Dari hasil penelitian yang telah dilakukan dapat diketahui bahwa sentimen negatif paling banyak terbentuk sekitar 60% dalam menanggapi isu kenaikan tarif dasar listrik.

Kata Kunci: Sentimen Analisis, Twitter, Klasifikasi, Naive Bayes

ABSTRACT. Along with the development of internet technology, the use of social media has also mushroomed. The most popular social media is Twitter. Twitter provides services for its users to send and read tweets that have been shared, so people prefer to express their opinions through social media rather than expressing it directly. Public opinion expressed in Twitter social media in the form of perception, both positive and negative. The abundance of public opinion can be used as research material to find information. The utilization of this information requires appropriate analytical techniques so that the information generated can assist many parties in making a decision. To overcome the above problems used the right data mining technique, namely sentiment analysis. Therefore, this research tries to do sentiment analysis to see people's perception of the issue of increasing electricity tariffs on social media Twitter using the Naïve Bayes method by classifying sentiments to be positive, negative, and neutral. From the results of research that has been done, it can be seen that the most negative sentiment is formed in response to the issue of the increase in electricity tariffs.

Keywords: Sentimen Analisis, Twitter, Klasifikasi, Naive Bayes

1. PENDAHULUAN

Dampak perkembangan teknologi informasi yang semakin meningkat membuat pertukaran informasi dan komunikasi semakin mudah dan cepat. Hal tersebut ditandai dengan pertumbuhan media sosial seperti twitter, facebook, instagram, dan lain sebagainya. Pertumbuhan media sosial tersebut mendorong terciptanya informasi tekstual yang besar sehingga diperlukan adanya media untuk menyajikan data yang memudahkan pengguna mendapatkan informasi yang sesuai dan akurat. Media sosial twitter merupakan salah satu media komunikasi yang populer saat ini, hal tersebut dapat dilihat dari peningkatan jumlah penggunaannya yang setiap hari semakin meningkat (Hidayatullah, 2016).

Meningkatnya pengguna twitter dari waktu ke waktu membuat banyak penggunanya menggunakan media tersebut sebagai tempat untuk mencurahkan segala permasalahan yang dialami dalam kehidupan sehari-hari (Setiawan & Setyohadi, 2017). Salah satunya adalah pemanfaat media tersebut untuk menyampaikan kritik, saran ataupun keluhan terhadap layanan dari sebuah produk atau jasa yang mereka gunakan.

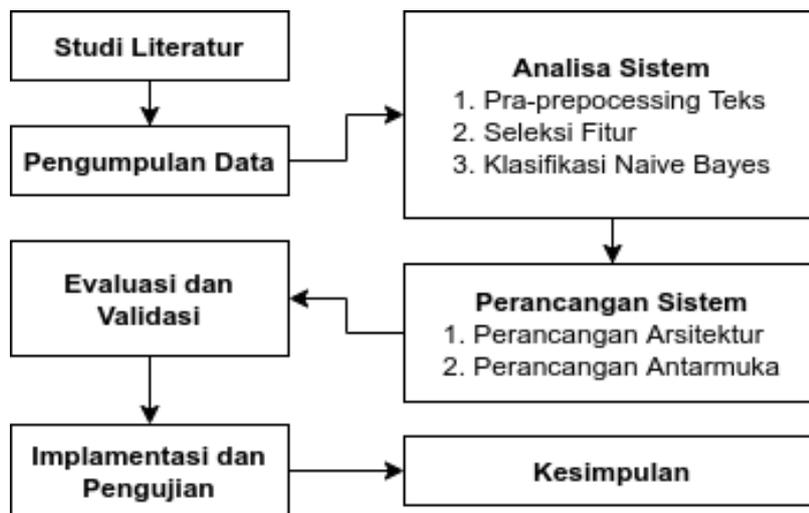
Diantara banyak tweet yang menjadi trending topik adalah masalah isu kenaikan tarif dasar listrik untuk golongan 900VA. Sehingga bermunculan pendapat masyarakat terhadap isu tersebut. Beragam pendapat yang muncul ada yang bernilai positif ada pula yang bernilai negatif. Analisis sentimen dapat membantu untuk mendapatkan gambaran secara umum tentang persepsi masyarakat dengan mengelompokkan jenis opini yang muncul menjadi kategori positif, negatif dan netral (Ernawati, 2016). Analisis sentimen adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik (Medhat & Korashy 2014).

Sentimen analisis melakukan proses pengelompokan polaritas teks yang ada dalam sebuah dokumen, kalimat, atau tingkat aspek dan menentukan apakah sebuah pendapat yang diungkapkan dalam sebuah dokumen atau

kalimat tersebut bersifat positif, negatif atau netral. Sentimen analisis yang lebih kompleks dapat menyatakan emosional sedih, gembira ataupun marah berdasarkan teks yang ada pada dokumen (Liu, Bing, 2012). Pada penelitian ini digunakan pendekatan teorema bayes yang menggunakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip tersebut kemudian digunakan untuk memecahkan masalah klasifikasi. salah satu metode klasifikasi yaitu naive bayes classifier. naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman, R and Sanger, J. 2007).

2. METODE

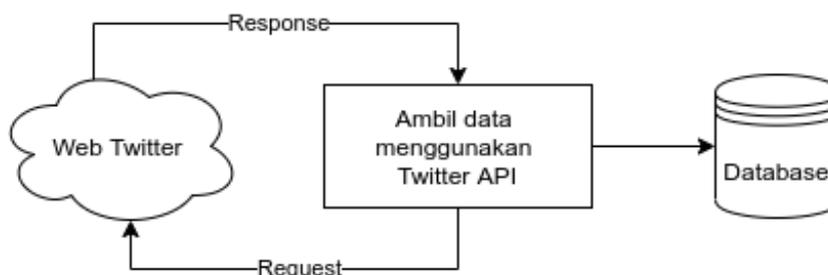
Jenis penelitian yang dipakai dalam penelitian ini merupakan kualitatif, yaitu penelitian yang dilakukan dengan mengumpulkan data dalam bentuk kata-kata atau kalimat dari sumber informasi yang penulis jadikan objek, dalam hal ini twitter. Tahapan yang dilakukan pada penelitian ini sebagai berikut:



Gambar 1. Tahapan Penelitian

Pengumpulan data

Pada penelitian ini data dikumpulkan dari website twitter dengan mengambil berdasarkan kata kunci yang berhubungan dengan tarif dasar listrik. Data kemudian disimpan kedalam database. Penarikan data dari twitter dilakukan menggunakan *Twitter API Search*.



Gambar 2. Proses Pengambilan Data Twitter

Data tweet yang diambil dengan memasukkan kata kunci tentang tarif dasar listrik rentang waktu mulai bulan mei sampai dengan juli 2017. Setelah proses pengambilan data selesai dilakukan, data tweet akan dibagi menjadi dua bagian, yaitu data latih dan data uji.

Text Mining

Menurut Younis, Text Mining adalah proses otomatis untuk mendeteksi dan mengungkapkan pengetahuan baru dan tidak terungkap serta hubungan dan pola dalam sumber data teks yang tidak terstruktur (Younis & Eman, 2015)

Menurut Feldman dan Sanger, Text mining adalah area penelitian baru dan menarik yang mencoba memecahkan masalah informasi dalam jumlah besar (overload information) dengan menggunakan teknik data mining, machine learning, natural language processing (NLP), information retrieval (IR), dan knowledge management. Text mining melibatkan preprocessing dokumen (text categorization, information extraction, term extraction)” (Liu, 2012).

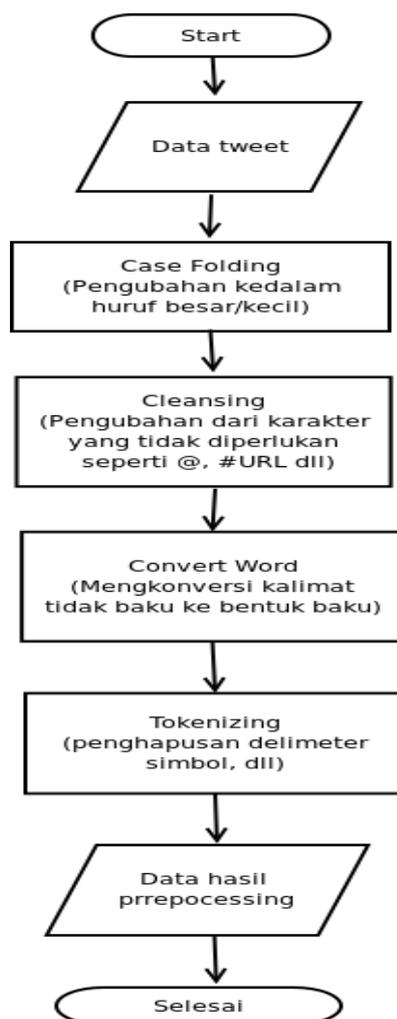
Sedangkan Menurut Kumar dan Bhatia (2013) “Text mining adalah bidang baru yang mencoba untuk menggali Informasi yang berguna dari teks bahasa alami (natural language). Hal ini dapat diartikan sebagai proses menganalisa teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu” (Kumar & Bhatia, 2015).

Pengolahan Data

a. Text Preprocessing

Text preprocessing dilakukan untuk menghindari data yang kurang sempurna. Tahapan yang dilakukan pada text preprocessing yaitu antara lain sebagai berikut:

1. Langkah pertama, semua huruf akan diubah kedalam huruf kecil atau besar semua atau disebut case folding.
2. Kemudian dilakukan cleansing yaitu untuk membersihkan dokumen dari kata yang tidak diperlukan untuk mengurangi noise (Hadna et al., 2016). Kata yang dihilangkan adalah karakter HTML, kata kunci, emoticon, hashtag (#), username (@username), url, RT.
3. Kemudian dilakukan convert word adalah untuk mengkonversi kalimat yang tidak baku kedalam bentuk baku, saat ini penggunaan kalimat alay atau bahasa gaul mengakibatkan penggunaan Bahasa Indonesia tidak baku.
4. Selanjutnya diteruskan dengan proses tokenizing yaitu proses dimana dokumen yang masih berupa kalimat dipecah perkata dan secara bersamaan dihilangkan semua karakter maupun tanda baca yang ada pada kalimat tersebut, hasil dari proses inilah yang disebut token. pada proses tokenize ini, semua tandabaca, simbol, atau apapun yang bukan huruf dihilangkan sehingga menjadi sekumpulan kata secara utuh (Indrayuni. 2016).
5. Hasilnya adalah kata-kata penyusun dari dokumen teks.



Gambar 3. Proses Text Preprocessing

b. Seleksi Fitur

Setelah melalui tahapan text preprocessing, selanjutnya dilakukan proses seleksi fitur pada dokumen teks. Berikut tahapan yang dilakukan pada proses seleksi fitur:

Tahap awal adalah Filtering data, Filtering data dilakukan untuk menghilangkan kata-kata yang tidak berpengaruh pada proses klasifikasi. Proses ini menggunakan stopwords removal. Stopwords dapat diartikan sebagai menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan (Ling et al.2014). Stopword removal juga dapat menghilangkan kata-kata yang ambigu (Widodo at al., 2016).

Tahap berikutnya adalah Stemming. Proses stemming dilakukan untuk mengambil kata dasari dari setiap kata yang ada (Novitasari, 2017). Pada proses ini dilakukan stemming menggunakan library sastrawi stemmer yang dibangun berdasarkan algoritma Nazief & Andriani (Mardiana et al, 2016).

Tahap selanjutnya adalah dilakukan proses ekstraksi fitur menggunakan n-gram pada dokumen teks. Proses ini dilakukan untuk menguraikan kata menjadi sejumlah n karakter sebagai satuan term dan menghitung probabilitas kemunculan tiap-tiap karakter. N-gram adalah potongan n karakter dalam suatu string tertentu atau potongan n kata dalam suatu kalimat tertentu (Sadida at al. 2017).

c. Model Klasifikasi Naive Bayes

Proses selanjutnya adalah dilakukan klasifikasi berdasarkan data hasil ekstraksi fitur menggunakan algoritma Naive Bayes. Naive bayes merupakan metode klasifikasi dengan konsep probabilitas bayesian.

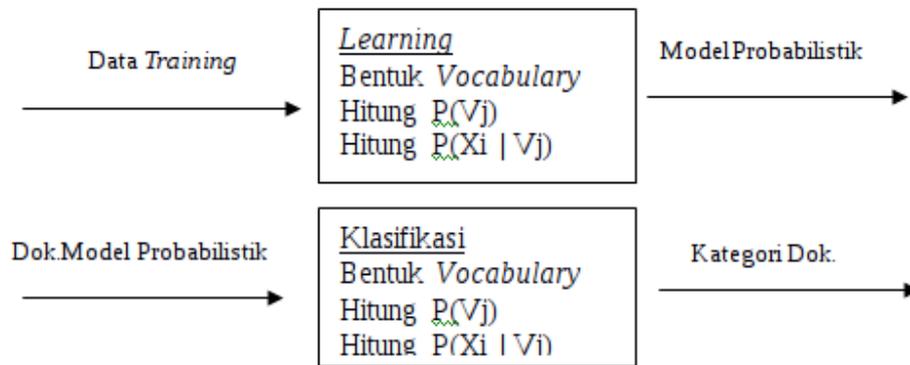
Naive Bayes merupakan teknik klasifikasi berbasis probalistik sederhana yang berdasarkan teorema bayes dengan asumsi independensi yang kuat. Dengan kata lain pada naive bayes, model yang digunakan adalah model fitur independent (Kusrini & Luthfi, 2010).

Naive bayes adalah salah satu model algoritma klasifikasi yang merupakan metode machine learning yang menggunakan konsep perhubungan probabilitas dan statistika yang dikemukakan oleh Thomas bayes. Algoritma tersebut digunakan untuk memprediksi probabilitas dimasa depan berdasarkan pengalaman masa lalu. Pendekatan naive bayes membuat asumsi sederhana bahwa semua atribut bersifat independen. Hal ini menyebabkan penggolongan yang jauh lebih sederhana, ini membuat efektif dalam praktiknya. Berikut persamaan dari teorema bayes:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{1}$$

Keterangan:

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan class spesifik
- P(H|X) = Probabilitas Hipotesis H berdasarkan kondisi X (posteriori probability)
- P(H) = Probabilitas Hipotesis (prior probability)
- P(X|H) = Probabilitas X berdasarkan kondisi Hipotesis H
- P(X) = Probabilitas X



Gambar 4. Proses Klasifikasi

Pada tahap ini data hasil ekstrasi fitur n-gram kemudian diklasifikasikan dengan menggunakan metode naive bayes classifier untuk memperoleh hasil sentimen.

d. Pengujian akurasi klasifikasi

Pengujian akurasi klasifikasi menggunakan *confusion matrix* untuk mengetahui nilai akurasi, *precision* dan *recall*. *Confusion matrix* merupakan metode yang digunakan untuk menghitung nilai akurasi pada data mining (Gunawan at al. 2018). *Confusion matrix* digambarkan dalam bentuk tabel yang menyatakan jumlah data uji yang benar dan jumlah data uji yang salah pada proses klasifikasi (Indrayuni, 2016).

Tabel 1. confusion matrix

	Classification	
	Positive	Negative
Prediction Positive	True Positive	False Positive
Prediction Negative	False Negative	True Negative

True Positives (TP) adalah jumlah record data positif yang dikalsifikasikan dengan benar. **False Positives (FP)** adalah jumlah record data negatif dikalsifikasikan sebagai nilai negatif. **False Negatives (FN)** adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif. Dan **True Negatives (TN)** adalah jumlah record data negatif yang diklasifikasikan dengan benar.

Sehingga nilai akurasi didapatkan dengan menghitung nilai yang benar dengan rumus:

$$Accuracy = \frac{TN+TP}{Total\ Data} \tag{1}$$

Nilai precision didapatkan dengan rumus:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

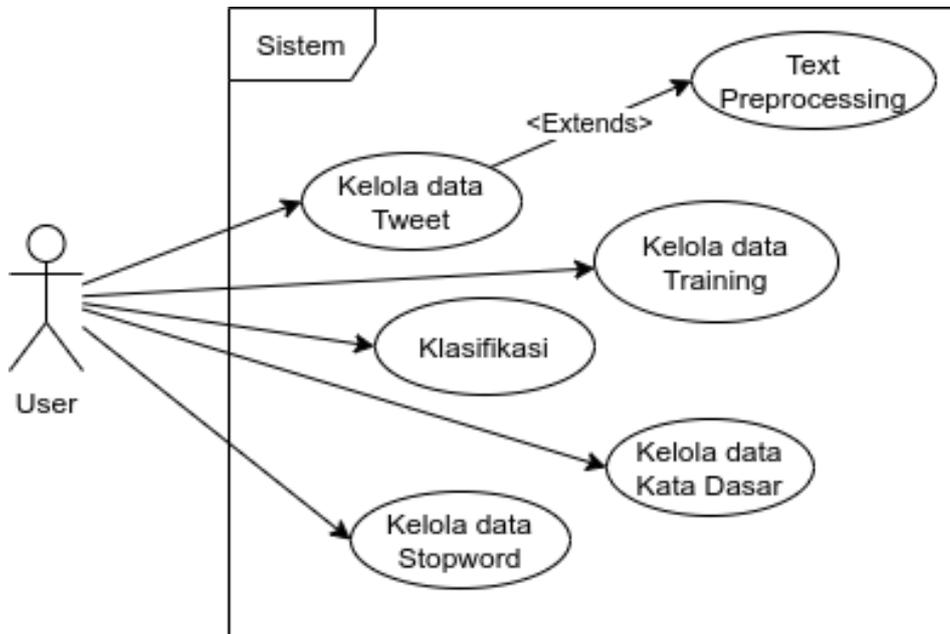
Dan nilai recall didapatkan dengan rumus:

$$Recall = \frac{TP}{TP+FN}$$

Persamaan (3)

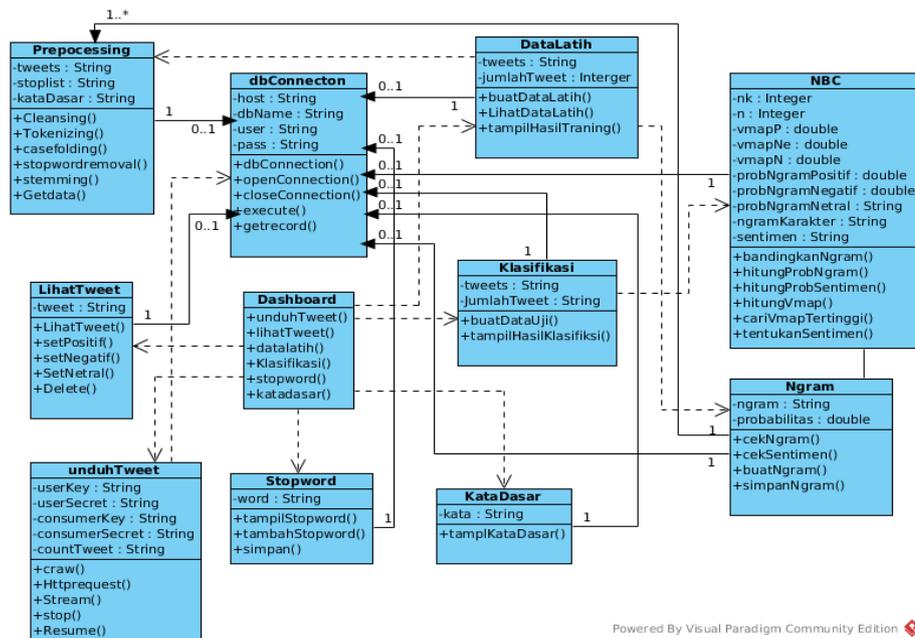
Perancangan Sistem

Pada tahap selanjutnya adalah dilakukan perancangan sistem untuk membangun sistem klasifikasi sentimen analisis. Berikut gambaran kebutuhan sistem yang digambarkan dengan diagram Use Case.



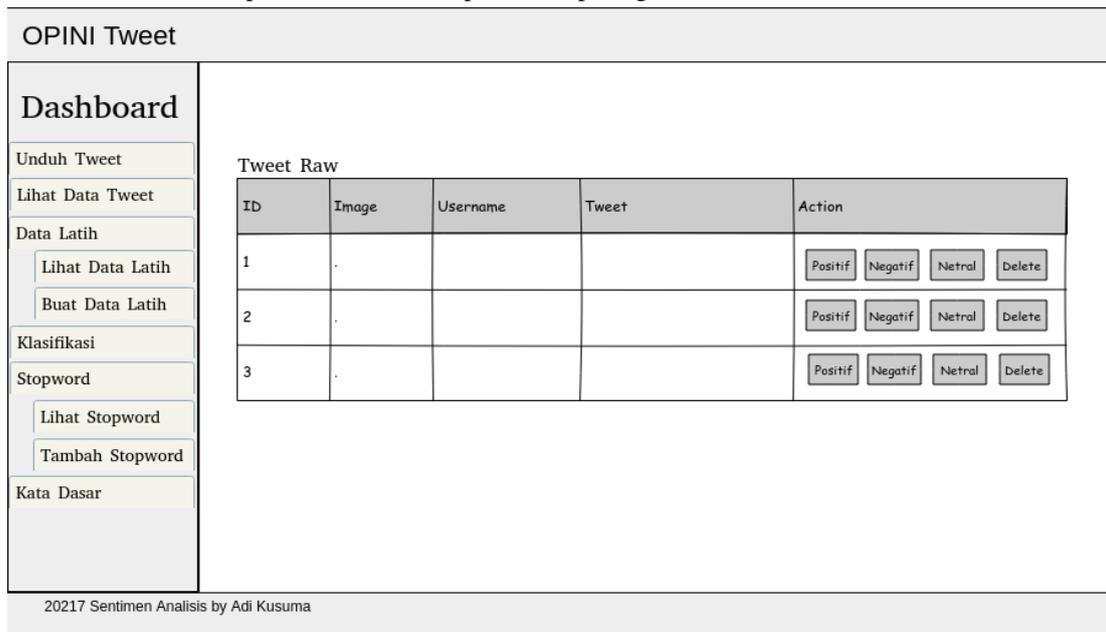
Gambar 5. Use Case Diagram

Dari use case diagram tersebut dihasilkan rancangan arsitektur sistem berupa class diagram yang digunakan untuk tahap pengembangan database dan antarmuka sistem. Berikut gambaran diagram class terlihat pada gambar 6.



Gambar 6. Class Diagram

Antarmuka sistem untuk proses klasifikasi dapat dilihat pada gambar 7 berikut ini.



Gambar 7. Antarmuka Proses Klasifikasi

3. HASIL DAN PEMBAHASAN

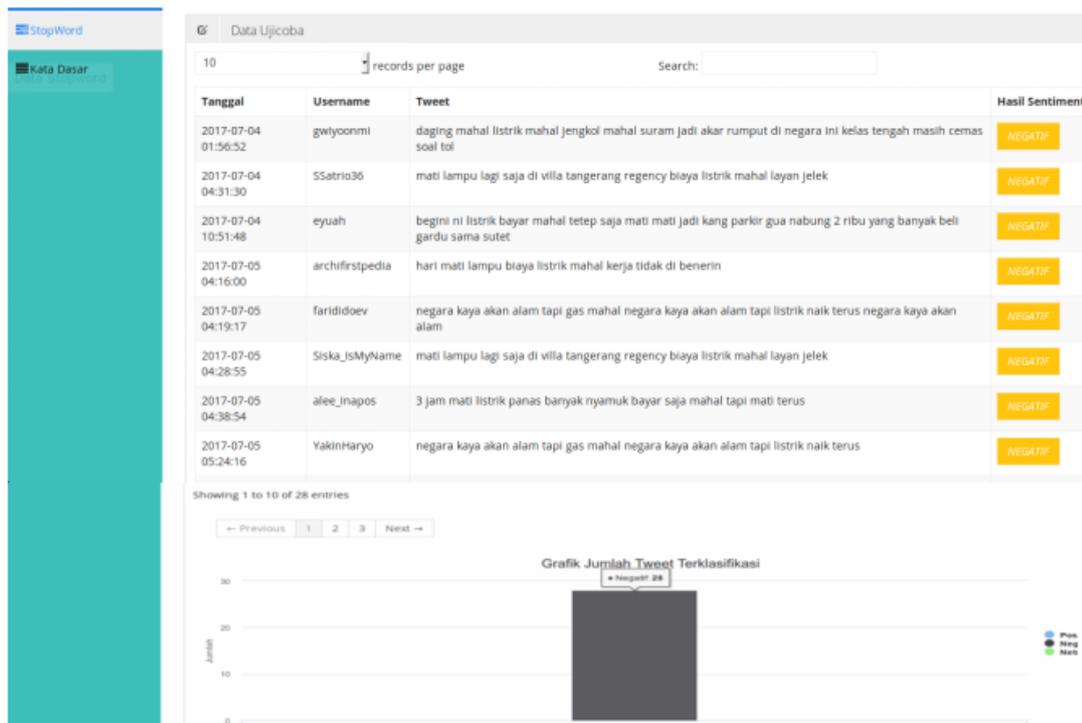
Hasil penelitian ini dilakukan proses pengujian terhadap sistem yang telah dibangun, diantaranya adalah pengujian fungsional sistem, pengujian akurasi klasifikasi, dan hasil analisis sentimen terhadap isu kenaikan tarif dasar listrik.

Pengujian Fungsional

Pengujian ini dilakukan dengan menjalankan aplikasi yang dibangun berbasis web. Pengujian dengan menjalankan secara detil pada setiap menu yang ada dengan tujuan untuk mengetahui fungsi mana yang sudah berfungsi dengan baik sesuai dengan sebagaimana mestinya.

Tabel 2. Pengujian Fungsional Sistem

No	Item Pengujian (Input/Output)	Hasil
1.	Login sistem	Sukses
2.	Unduh data Tweet	Sukses
3.	Kelola data latih	Sukses
4.	Kelola data kata dasar	Sukses
5.	Kelola data stopword	Sukses
6.	Proses Klasifikasi data	Sukses



Gambar 8. Hasil kalsifikasi

Pengujian Akurasi Klasifikasi

Pengujian yang telah dilakukan terhadap model klasifikasi naive bayes dengan ekstrasi fitur n-gram dengan data sebanyak 300 tweet didapatkan hasil klasifikasi menjadi tiga kelas yaitu positif, negatif dan netral dengan hasil klasifikasi positif sebanyak 100 tweet, negatif sebanyak 87 tweet dan netral sebanyak 89 tweet. Perhitungan akurasi dengan menghitung nilai precision dan recall ditampilkan dalam tabel berikut:

Tabel 3. Hasil Pengujian Akurasi

	True netral	True positif	True negatif	Class precision
Pred. Netral	89	0	13	87.25%
Pred. Positif	2	100	0	98.04%
Pred. Negatif	9	0	87	90.62%
Class recall	89.00%	100.00%	87.00%	

Pengujian akurasi klasifikasi naïve bayes menggunakan *confusion matrix* untuk mencari nilai akurasi, presisi dan *recall*. Dari tabel 2 tersebut diketahui nilai hasil klasifikasi terbagi menjadi tiga kelas yaitu positif, negatif dan netral. Pada proses pengujian klasifikasi terdapat nilai klasifikasi yang tidak sesuai. Nilai akurasi dari proses kalsifikasi didapatkan 92.00% berdasarkan perhitungan

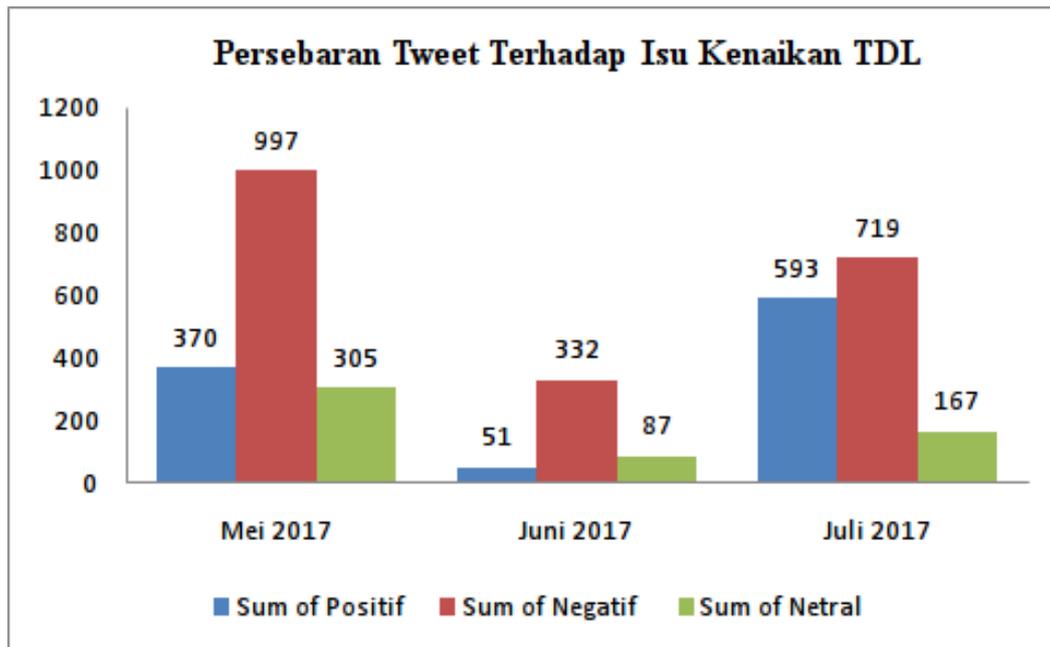
$$Accuracy = \frac{89 + 100 + 87}{300}$$

$$Accuracy = 0,92 * 100\%$$

$$Accuracy = 92.00\%$$

Hasil Analisa Sentimen

Data yang telah dikumpulkan dari periode mei sampai dengan juli 2017 didapatkan data tweet sebagai 4000 tweet yang kemudian dilakukan proses klasifikasi. Dari hasil klasifikasi tersebut dapat dilihat pada tampilan grafik pesebaran tweet terhadap isu kenaikan tarif dasar listrik sebagai berikut:



Gambar 9. Grafik Pesebaran Twitter

Dari grafik diatas bisa kita lihat pada bulan Mei, tweet negatif terhadap isu kenaikan tarif dasar listrik sangat tinggi. Seperti yang kita ketahui di awal bulan Mei tersebut pemerintah melakukan pencabutan subsidi listrik golongan 900 VA yang masuk dalam kategori mampu untuk tahap ketiga Mei dan Juni 2017, hal tersebutlah yang menyebabkan banyaknya tweet negatif di twitter terhadap kebijakan pemerintah tersebut.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa Penerapan proses text mining kedalam sistem yang dibangun cukup berhasil. Terbukti dengan sistem yang dibangun bisa melakukan preprocessing text dan melakukan klasifikasi data tweet secara otomatis dalam jumlah besar.

Metode *naive bayes classifier* dapat diterapkan sebagai proses klasifikasi untuk membangun sentimen analisis dengan tingkat akurasi 92.00% dengan menerapkan ekstrasi fitur n-gram karakter.

DAFTAR PUSTAKA

- Ernawati, S. (2016). Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naïve Bayes. *J. Evolusi*, 4(1), 45-54.
- Feldman, R and Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press:NewYork.
- Gunawan, B., Sastypratiwi, H., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naïve Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 4(2), 113-118.
- Hadna, Nurrun Muchammad Shiddieqy, Santosa, Paulus Insap dan Winarno, Wing Wahyu.2016. Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter.Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016) Yogyakarta, 18-19 Maret 2016.
- Hidayatullah, A. F. (2016). Twitter sebagai media dakwah. *Teknoin*, 22(1).
- Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *EVOLUSI: Jurnal Sains dan Manajemen*, 4(2).
- Kumar, Lokesh and Bhatia, Parul Kalra. 2015. Text Minig: Concepts, Process And Application. *Journal of Global Research in Computer Science*. Volume 4, No. 3, March 2013
- Kusrini dan Luthfi, Emha Taufiq. 2010. *Algoritma Data Mining*.Penerbit Andi:Yogyakarta
- Ling, Juen, Kencana, I Putu Eka N dan Oka, Tjokorda Bagus.2014. Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square.E-Jurnal Matematika Vol. 3 (3), Agustus 2014.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher:San Rafael, California
- Mardiana, T., Adji, T. B., & Hidayah, I. (2016). Stemming influence on similarity detection of abstract written in Indonesia. *Telkomnika*, 14(1), 219.

- Medhat, Walaa, Hassan, Ahmed, & Korashy, Hoda.2014. Sentiment Analysis Algorithms And Applications: A Survey. *Ain Shams Engineering Journal* (2014) 5, 1093-1113
- Novitasari, D. (2017). Perbandingan Algoritma Stemming Porter dengan Arifin Setiono untuk Menentukan Tingkat Ketepatan Kata Dasar. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 1(2), 120-129.
- Sadida, Rizqon dkk.2017. Perancangan Sistem Analisis Sentimen Masyarakat Pada Sosial Media Dan Portal Berita.Seminar Nasional Teknologi Informasi dan Multimedia 2017 STMIK AMIKOM Yogyakarta, 4 Februari 2017.
- Setiawan, R. A., & Setyohadi, D. B. (2017). Analisis Komunikasi sosial media twitter sebagai saluran layanan pelanggan provider internet dan Seluler di Indonesia. *Journal of Information Systems Engineering and Business Intelligence*, 3(1), 16-25.
- Widodo, P., Putra, J. A., Afiadi, S., Arifin, A. Z., & Herumurti, D. (2016). Klasifikasi Kategori Dokumen Berita Berbahasa Indonesia dengan Metode Kategorisasi Multi-Label Berbasis Domain Specific Ontology. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 2(2).
- Younis, Eman M.G .2015. SentimentAnalysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study.*International Journal of Computer Applications* (0975 - 8887). Volume 112 - No. 5, February 2015