

Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi *Naïve Bayes* dan *Support Vector Machine*

Enos Dwiarto¹, Mujiono Sadikin*²

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana^{1,2}

Jl. Raya Meruya Selatan, Kembangan, Jakarta, 11650

E-mail : 141516010105@student.mercubuana.ac.id¹, mujiono@mercubuana.ac.id²

Abstrak -- Transportasi online merupakan salah satu pilihan bagi masyarakat untuk berkegiatan sehari-hari baik saat bekerja, bepergian dan melakukan aktivitas lain. Salah satu cara untuk mengetahui persepsi masyarakat terhadap layanan transportasi online adalah dengan analisis sentimen seperti yang dilakukan pada penelitian ini. Data yang digunakan merupakan data valid dari sosial media Twitter untuk Transportasi online GrabId dan GojekIndonesia. Teknik analisis sentimen yang digunakan adalah *Naïve Bayes Classifier* dan metode *Support Vector Machine (SVM)*. Keduanya digunakan untuk membandingkan tanggapan masyarakat dari analisis sentimen data tweet yang telah diklasifikasikan menjadi positif dan negatif. Berdasarkan penelitian ini didapatkan bahwa GrabId menggunakan metode SVM memberikan hasil class precision positif dan negatif yaitu 86.47% dan 46.67%, class recall positif dan negatif yaitu 96.21% dan 18.06%, accuracy 84.08%. Sedangkan untuk GojekIndonesia, metode SVM memberikan hasil yaitu class precision positif dan negatif yaitu 73.90% dan 35.65%, class recall positif dan negatif yaitu 89.84% dan 15.07%, accuracy 69.50%. Dari akurasi yang dihasilkan, metode SVM menghasilkan kinerja terbaik.

Kata Kunci: Transportasi online, analisis sentimen, Twitter, *Naïve Bayes Classifier*, *Support Vector Machine*

Abstract - Online transportation is an option for people to carry out their daily activities, both while working, traveling and doing other activities. One way to find out people's perceptions of online transportation services is by analyzing sentiment as was done in this study. The data used is valid data from social media Twitter for online transportation, GrabId and GojekIndonesia. The sentiment analysis technique used is the *Naïve Bayes Classifier* and the *Support Vector Machine (SVM)* method. Both are used to compare public responses from the sentiment analysis of tweet data that have been classified into positive and negative. Based on this study, it was found that GrabId using the SVM method gave positive and negative precision class results, namely 86.47% and 46.67%, positive and negative class recall namely 96.21% and 18.06%, accuracy 84.08%. Whereas for GojekIndonesia, the SVM method gives results, namely positive and negative precision class, namely 73.90% and 35.65%, positive and negative class recall, namely 89.84% and 15.07%, accuracy 69.50%. From the resulting accuracy, the SVM method produces the best performance.

Keywords: Online transportation, sentiment analysis, Twitter, *Naïve Bayes Classifier*, *Support Vector Machine*

I. PENDAHULUAN

Transportasi umum merupakan fasilitas yang digunakan masyarakat dalam aktivitas sehari-hari. Transportasi umum terbagi menjadi dua yaitu transportasi konvensional (*offline*) dan transportasi *online*[1]. Transportasi konvensional adalah transportasi umum yang digunakan masyarakat yang tanpa perlu melakukan pemesanan menggunakan aplikasi. Sedangkan transportasi *online* adalah transportasi umum berbasis aplikasi sebagai media untuk melakukan pemesanan sebelum menggunakannya. Saat ini masyarakat mulai terbiasa menggunakan jasa transportasi online karena kemudahan dan efisiensinya yang membantu dalam beraktivitas

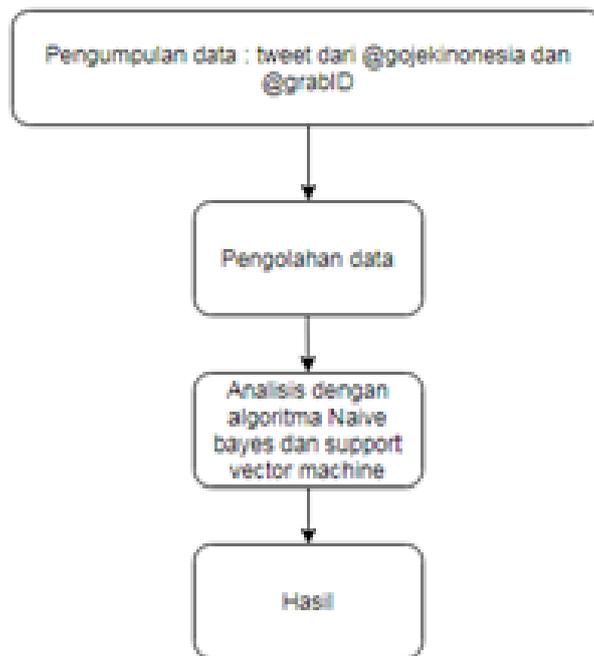
Banyak perusahaan transportasi online yang menawarkan berbagai layanan sehingga menciptakan persaingan yang ketat. Perusahaan transportasi online yang saat ini sering digunakan oleh masyarakat antara lain Gojek dan Grab[2]. Gojek adalah sebuah perusahaan yang berasal dari Indonesia yang mengandung konsep teknologi yang menyediakan layanan jasa transportasi berbasis aplikasi Online. Gojek adalah sebagai pelopor transportasi online yang berdiri sejak tahun 2010[3]. Grab didirikan pada tahun 2012 di Singapura sebagai kantor pusatnya dan beroperasi di Indonesia sejak tahun 2014[4]. Terjadinya persaingan pada kedua perusahaan transportasi online tentu perlu saran dan kritik dari pengguna untuk meninjau kekurangan dari setiap layanan kedua transportasi online ini, sehingga dilakukan sentimen analisis untuk mengetahui tingkat kepuasan pengguna transportasi online antara 2(dua) aplikasi tersebut.

Adanya persaingan antara kedua aplikasi transportasi online mengakibatkan masyarakat banyak menuangkan cerita dan protes mengenai hal tersebut yang ada di Indonesia menggunakan media sosial seperti twitter. Oleh karena itu, sentimen analisis yang digunakan dalam penelitian ini mengambil data tweet masyarakat yang berkomentar tentang transportasi online seperti di twitter[5]. Penelitian ini ditujukan untuk mengetahui tingkat kepuasan dari pengguna yang salah satu manfaatnya adalah bagi pihak transportasi online dapat digunakan sebagai bahan evaluasi peningkatan pelayanan. Dalam penelitian ini digunakan metode *Naïve Bayes Classifier* dan *SVM* untuk klasifikasi data.

Menurut penelitian Fauziah Afshoh yang berjudul Analisa Sentimen Menggunakan Naïve Bayes Classifier untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial twitter diketahui apabila metode Naïve Bayes Classifier sangat berperan baik dalam menghasilkan nilai akurasi klasifikasi yang lebih baik dibandingkan dengan metode Lexicon Based. Dimana penelitian tersebut bertujuan untuk melihat hasil sentimen dari persepsi masyarakat terhadap kenaikan harga rokok di Indonesia[6].

II. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis data komentar customer pada GrabId dan GojekIndonesia di twitter dengan membandingkan metode *Naïve Bayes Classifier* dan algoritma *Support Vector Machine*. Hasil penelitian ini untuk tahap edukasi pada kedua perusahaan ojek online terhadap sistem perusahaan tersebut dan sebagai perbandingan kedua metode yang lebih baik pada penelitian ini. Untuk menjalankan eksperimen kinerja algoritma, digunakan *tool* Rapid Miner. Tahapan dari proses penelitian ini dituangkan pada gambar 1.



Gambar.1 Tahapan penelitian

A. Pengumpulan Data

Pengumpulan data penelitian ini hasil dari penarikan dataset melalui media sosial twitter. Proses pengambilan dengan hasil dari Tag dan Retweet dari tanggal 2018-01-01 kedua perusahaan tersebut. Data yang diperoleh sejumlah 1000 GojekIndonesia Dan 1000 GrabID dalam format csv. Contoh *script* untuk crawling data pada gambar 2.

```

import tweepy
import csv

consumer_key = 'CCHGoCYD5xveYLwfmHTEUC8sP'
consumer_secret = '8ku54fWXP58digJ06ufUPNgBEDc7KvMDdG5zbpn0uso11CrAM0'
access_token = '983243081452437504-ksA8131in7eB1Mz9R37a8XZ1bZwaQ7B'
access_token_secret = 'JNkyYIrfR5coBeUuqpvYEnKX02xmBcxpyf6gKelq57enn'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

csvFile = open("gojeksekarang.csv", "a")
csvWriter = csv.writer(csvFile)

# COLLECTING AND FILTERING RETWEETS
for tweet in tweepy.Cursor(api.search, q="gojekindonesia -filter:retweets", count=50,
                           lang="id",
                           since="2018-01-01", tweet_mode="extended").items(1000):
    if (not tweet.retweeted) and ('RT @' not in tweet.full_text):
        print(tweet.full_text)
        csvWriter.writerow([tweet.full_text.encode('utf-8')])
    
```

Gambar 2. Crawling Data

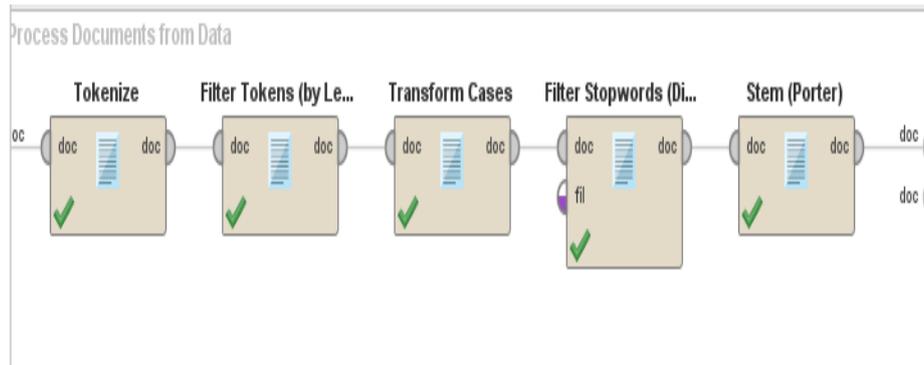
Setelah data sudah diambil dengan script diatas lalu selanjutnya diberi label sentimen yang berfungsi membedakan kelas, dengan dua kelas sentimen yaitu sentimen positif dan negatif. Berikut pada tabel 1 berupa contoh dataset yang sudah dicrawling dan diberi label.

Tabel 1. Dataset

Tweets	sentimen
@GrabID Aku pernah order grabcar padahal aku order yg biasa gak pake grabprotect ehh malah dapet yg grabprotect tadinya mau aku cancel kasian sm driver krn aku dpt fasilitas yg aman tapi bayar dgn harga yg lebih murah #CeritaNgeGrab #SiapJalanLagibbb	positive
@GrabID Keren deh buat @GrabID untuk inovasi keamanan pelanggannya sekarang aku jadi gak takut dan #SiapJalanLagi #CeritaNgeGra	positive
@GrabID Paman gimana ini dah dua hari keliling kemana mana tidak dapat orderan sama sekali. Mohon pencerahanyabbb	negative
@GrabID Pengen cerita, tapi selama pandemi ga pernah pake ojol lagi kecuali beli makanbbb	positive
@GrabID Aku mau ceritakk!jadi selama new normal, tiap weekend aku tuh suka pergi buat olahraga. Berhubung aku anak rantau yang gak ada kendaraan jadi yaaa mau gamau pilihannya ya naik angkutan umum. Tapi, kalo semisal naik busway/KRL agak worry gt kan #CeritaNgeGrab #SiapJalanLagibbb	positive

B. Data Preprocessing

Tahap ini data tweet yang hasil di Crawling data akan dilakukan tahap Preprocessing untuk pembersihan data yang tidak terpakai yang dilakukan Folding case, Tokenize, Filtering, Stemming[1]. Dapat dilihat pada gambar 3. Tahap Preprocessing .



Gambar 3. Tahap Preprocessing

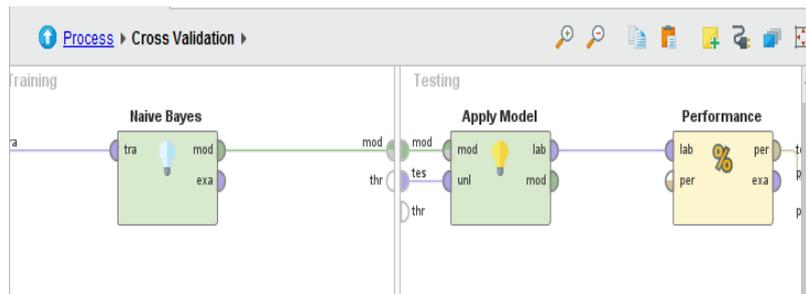
Berikut merupakan penjelasan dari tahap Preprocessing :

1. Tahap Transform case / folding case adalah tahap mengubah semua huruf pada dataset menjadi huruf kecil dan menghilangkan seluruh tanda baca pada dataset.
2. Tahap Tokenize / filter tokens adalah tahap memisahkan kata pada dataset berdasarkan spasi yang terbaca.
3. Tahap Filtering adalah tahap membuang kata-kata yang tidak diperlukan pada dataset.
4. Tahap Stemming adalah tahap pengubahan kata pada dataset menjadi kata dasar[8].

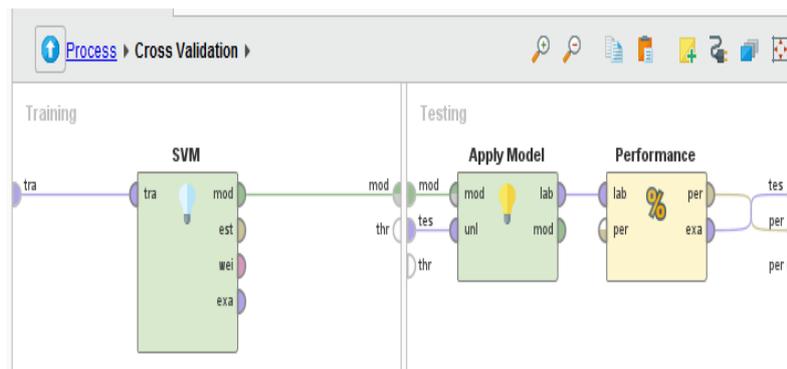
C. Klasifikasi Analisis Sentimen.

Tahap ini pengolahan bahasa pada dataset yang sudah di Preprocessing untuk mendapatkan akurasi pada penelitian ini dengan menggunakan 2 metode *Naïve Bayes Classifier* dan *SVM* untuk perbandingan persentase

akurasi yang terbaik. Pada gambar 4. Proses Tahap metode *Naïve Bayes Classifier* dan Pada gambar 5. Proses Tahap metode SVM :



Gambar 4. Tahap proses Metode Naïve Bayes Classifier.



Gambar 5. Proses tahap metode Support Vector Machine

D. Evaluasi Klasifikasi analisis sentimen

Setelah selesai klasifikasi pada kedua metode ini, selanjutnya melakukan evaluasi perhitungan kinerja klasifikasi telah diuji dengan parameter *Precesion*, *Recall*, dan *Accuracy*. Pada keakuratan telah dilakukan pada klasifikasi dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine*, dikelompokkan pada kelas positif terdapat *True Positif(TP)*, *False Positif(FP)* dan bagian kelas negative terdapat *True Negatif(TN)*, *False Negatif(FN)*. Bentuk *confusion matrix* seperti berikut.:

Table 2. *Confusion Matrix*

	True Positive (1)	True Negative (0)
Prediction Positive (1)	TP	FP
Prediction Negative (0)	FN	TN

Keterangan :

TP: prediksi *True Positive* dengan data sentimen positif yang sudah diklasifikasikan dengan hasil positif.

TN: prediksi *True Negative* dengan data sentimen negatif yang sudah diklasifikasikan dengan hasil negatif.

FP: Prediksi *False Positive* dengan data sentimen positif yang sudah diklasifikasikan namun dengan hasil negatif.

FN: : Prediksi *False Negative* dengan data sentimen negatif yang sudah diklasifikasikan namun dengan hasil positif.

Pada hasil yang ditunjukkan dari dua metode sentimen dengan klasifikasi data yang merupakan sentimen dapat dihitung dengan rumus *accuracy*, *precession*, *recall* :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

accuracy merupakan rumus rasio prediksi *True (Positive dan Negative)* dengan keseluruhan data.

$$precession = \frac{TP}{TP + FP} \times 100\%$$

Precession merupakan rumus rasio prediksi *True Positive* dengan keseluruhan data yang diprediksi positif.

$$recall = \frac{TP}{TP + FN} \times 100\%$$

Recall merupakan rumus rasio prediksi *True Positive* dibandingkan dengan keseluruhan data keseluruhan *True Positive*.

III. HASIL DAN PEMBAHASAN

A. Naïve Bayes Classifier

Pada *performance* vektor dari hasil akurasi klasifikasi GrabId yang menggunakan metode *Naïve Bayes Classifier* dengan hasil dari nilai TP=749, FP=97, TN=95, FN=58 dan menunjukkan nilai akurasi. *Class precision* positif dan negatif yaitu 88.53% dan 37.91%, *Class recall* positif dan negatif yaitu 88.74% dan 37.42%,. Maka hasil yang didapatkan Rata-rata *class precision* 63.22%, Rata-rata *class recall* 63.32%, *accuracy* 80.78%, pada gambar 6. Hasil akurasi GrabId metode *Naïve Bayes Classifier*.

accuracy: 80.78% +/- 2.79% (micro average: 80.78%)

	true positive	true negative	class precision
pred. positive	749	97	88.53%
pred. negative	95	58	37.91%
class recall	88.74%	37.42%	

Gambar 6. Hasil akurasi GrabId metode *Naïve Bayes Classifier*.

Pada *performance* vektor dari hasil akurasi klasifikasi GojekIndonesia yang menggunakan metode *Naïve Bayes Classifier* dengan hasil dari nilai TP=604, FP=198, TN=124, FN=74 dan menunjukkan nilai akurasi. *Class precision* positif dan negatif yaitu 75.31% dan 37.37%, *Class recall* positif dan negatif yaitu 82.97% dan 27.21%. Maka hasil yang didapatkan rata-rata *class precision* 56.34%, rata-rata *class recall* 55.09%, *accuracy* 67.80%, pada gambar 7. Hasil akurasi *GojekIndonesia* metode *Naïve Bayes Classifier*.

accuracy: 67.80% +/- 4.52% (micro average: 67.80%)

	true positive	true negative	class precision
pred. positive	604	198	75.31%
pred. negative	124	74	37.37%
class recall	82.97%	27.21%	

Gambar 7. Hasil akurasi *GojekIndonesia* metode *Naïve Bayes Classifier*.

B. SVM Classifier

Pada *performance* vektor dari hasil akurasi klasifikasi GrabId yang menggunakan metode SVM dengan hasil dari nilai TP=812, FP=127, TN=32, FN=28 dan menunjukkan nilai akurasi. *Class precision* positif dan negatif yaitu 86.47% dan 46.67%, *Class recall* positif dan negatif yaitu 96.21% dan 18.06%. Maka hasil yang didapatkan rata-rata *class precision* 66.57%, rata-rata *class recall* 57.14%, *accuracy* 84.08%, pada gambar 8. Hasil akurasi GrabId metode *Support Vector Machine*.

accuracy: 84.08% +/- 2.13% (micro average: 84.08%)

	true positive	true negative	class precision
pred. positive	812	127	86.47%
pred. negative	32	28	46.67%
class recall	96.21%	18.06%	

Gambar 8. Hasil akurasi GrabId metode Support Vector Machine.

Pada *performance* vektor dari hasil akurasi klasifikasi GojekIndonesia yang menggunakan metode SVM dengan hasil dari nilai TP=654, FP=231, TN=74, FN=41 dan menunjukkan nilai akurasi. *Class precision* positif dan negatif yaitu 73.90% dan 35.65%, *Class recall* positif dan negatif yaitu 89.84% dan 15.07%. Maka hasil yang didapatkan rata-rata *class precision* 54.78%, rata-rata *class recall* 52.46%, *accuracy* 69.50%, pada gambar 9. Hasil akurasi GojekIndonesia metode *Support Vector Machine*.

accuracy: 69.50% +/- 3.21% (micro average: 69.50%)

	true positive	true negative	class precision
pred. positive	654	231	73.90%
pred. negative	74	41	35.65%
class recall	89.84%	15.07%	

Gambar 9. Hasil akurasi GojekIndonesia metode Support Vector Machine.

Table 3. Kinerja Akurasi Model

Studi kasus/Metode Klasifikasi	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
GrabId, <i>Naïve Bayes Classifier</i>	63.22%	63.32%	64.36%
GojekIndonesia, <i>Naïve Bayes Classifier</i>	56.34%	55.09%	50.90%
GrabId, <i>SVM</i>	66.57%	57.14%	84.08%
GojekIndonesia, <i>SVM</i>	54.78%	52.46%	69.50%

Pada Tabel 3 sebuah hasil perbandingan metode *Naïve bayes classifier* dan *SVM* dari dua studi kasus yaitu GrabId dan GojekIndonesia. Berdasarkan penelitian ini didapatkan bahwa GrabId menggunakan metode *SVM* memberikan hasil yaitu *class precision* 66.57%, *class recall* 57.14%, *Accuracy* 84.08%. Sedangkan GojekIndonesia menggunakan metode *SVM* memberikan hasil yaitu *class precision* 54.78%, *class recall* 52.46%, *Accuracy* 69.50%.

IV. KESIMPULAN

Berdasarkan hasil dari analisis sentimen pada penelitian ini, terlihat jasa transportasi *online* yang telah diprediksi dengan menggunakan metode *Naïve Bayes Classifier* dan *SVM* menunjukkan hasil yang signifikan. GrabId menggunakan metode *SVM* memberikan hasil yaitu *class precision* 66.57%, *class recall* 57.14%, *Accuracy* 84.08%. sedangkan GojekIndonesia menggunakan metode *SVM* memberikan hasil yaitu *class precision* 54.78%, *class recall* 52.46%, *Accuracy* 69.50%. Disimpulkan metode *SVM* memiliki *performance* yang baik dibandingkan metode *Naïve Bayes Classifier*.

DAFTAR PUSTAKA

- [1] D. G. Nugroho, Y. H. Chrisnanto, and A. Wahana, "Analisis Sentimen Pada Jasa Ojek Online ... (Nugroho dkk.)," pp. 156–161, 2015.
- [2] A. Chan, M. Maharani, and P. W. Tresna, "Perbandingan Pengalaman Pengguna Pada Aplikasi Mobile Go-Jek Dan Grab (Studi Pada Konsumen Pt Go-Jek Dan Pt Grab Indonesia Di Dki Jakarta)," *AdBispreneur*, vol. 2, no. 2, 2017, doi: 10.24198/adbispreneur.v2i2.13183.
- [3] "Gojek - Wikipedia bahasa Indonesia, ensiklopedia bebas." <https://id.wikipedia.org/wiki/Gojek>.
- [4] "Grab (perusahaan) - Wikipedia bahasa Indonesia, ensiklopedia bebas." [https://id.wikipedia.org/wiki/Grab_\(perusahaan\)](https://id.wikipedia.org/wiki/Grab_(perusahaan)).
- [5] B. M. Pintoko and K. M. L, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2017.
- [6] F. AFSHOH, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter," pp. 1–17, 2017.
- [7] M. . Imelda A.Muis & Muhammad Affandes, "Penerapan Metode SVM (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet," *Sains, Teknol. dan Ind. Sultan Syarif Kasim Riau*, vol. 12, no. 2, pp. 189–197, 2015.
- [8] A. T. J. H, "Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining," *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.