



## Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Tian Aditya  
Assignment title: Paper\_1  
Submission title: Paper\_1  
File name: ra\_Hayatin\_Aditya\_-\_analisa\_sentimen\_stemming\_lexicon\_pyth...  
File size: 249.65K  
Page count: 8  
Word count: 3,190  
Character count: 19,836  
Submission date: 24-Jul-2023 06:25PM (UTC+0700)  
Submission ID: 2136044181

REPOSITOR, Vol. 2, No. 7, Juli 2020, Pp. 977-984  
ISSN : 2714-7975  
E-ISSN : 2716-1382

### Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia

Abdul Hady Dyo Fatra<sup>1</sup>, Nur Hayatin<sup>2</sup>, Christian Sri Kusuma Aditya<sup>3</sup>  
Universitas Muhammadiyah Malang  
abdulhady@gmail.com<sup>1</sup>, noorhayatin@gmail.com<sup>2</sup>, christianskaditya@umm.ac.id<sup>3</sup>

#### Abstrak

Penelitian ini mengajukan sebuah klasifikasi terhadap respon masyarakat terhadap keputusan pemerintah untuk memindahkan ibu kota Indonesia menggunakan metode lexicon. Hasil akurasi pengujian diukur dengan menggunakan confusion matrix. Data pada penelitian ini menggunakan data dari twitter yang berupa tweet. Data berisi tweet respon masyarakat terhadap keputusan perpindahan ibu kota Indonesia. Data melewati 5 proses preprocessing yaitu case folding, punctuation removal, stopword removal, stemming, dan tokenizing. Lexicon digunakan karena menghasilkan nilai akurasi yang baik. Pada penelitian ini juga akan mencari kamus yang memiliki hasil klasifikasi paling baik. Hasil penelitian ini menunjukkan hasil klasifikasi yang baik dengan mendekati hasil oleh pakar.

**Kata kunci:** Analisa sentimen, Stemming, Lexicon, Python

#### Abstract

This study proposes a classification of public response to the government's decision to move the Indonesian capital using the lexicon method. The results of testing accuracy are measured using a confusion matrix. The data in this study use data from Twitter in the form of tweets. The data contains tweets of community responses to the decision to move the Indonesian capital. Data passes through 5 preprocessing processes, namely case folding, punctuation removal, stopword removal, stemming, and tokenizing. Lexicon is used because it produces good accuracy values. In this study also will look for a dictionary that has the best classification results. The results of this study show the results of a good classification by approaching the results by experts.

**Keywords:** Sentiment analysis, Stemming, Lexicon, Python

#### 1. Pendahuluan

Respon masyarakat terhadap keputusan pemerintah adalah hal penting dalam kegiatan negara. Seling berjalanya waktu masyarakat memiliki kecenderungan untuk mengutarakan pendapat mereka dalam sosial media dibandingkan dengan mengutarakan aspirasinya secara langsung. Salah satu wadah untuk menampung respon masyarakat adalah sosial media, sehingga pendapat masyarakat di dalam sosial media menjadi aspek yang bisa dijadikan tolak ukur kepuasan masyarakat terhadap keputusan yang di buat oleh pemerintah. Sosial media adalah beberapa perangkat lunak yang membuat individu atau komunitas dapat berbagi, berkomunikasi, bertukar, dan berkolaborasi [1]. Sosial media yang memiliki fokus untuk diskusi topik adalah twitter. Pada kuartal pertama 2018 twitter memiliki 336 juta pengguna aktif [2]. Oleh sebab itu twitter merupakan indikator yang baik dalam penelitian Analisa sentimen [3].

Analisa sentimen merupakan satu contoh dari penelitian teks mining. Analisa sentimen adalah riset komputasional dari opini dan emosi yang terdapat dalam sebuah teks [4]. Tujuan dari analisa sentimen adalah untuk mengetahui opini pengguna terhadap suatu topik. Respon terhadap topik dapat dikelompokkan menjadi dua yaitu positif dan negatif. Sehingga dalam melakukan penelitian analisa sentimen kita dapat mengetahui respon pengguna terhadap tokoh, keputusan, produk, dan lain sebagainya. Metode lexicon adalah salah satu metode klasifikasi teks yang banyak digunakan dalam penelitian. Proses pada metode ini adalah dengan cara memeriksa dokumen yang sudah clean dengan kamus lexicon yang sudah disiapkan sebelumnya. Kamus lexicon terdiri dari banyak kata yang memiliki nilai positif dan negatif. Metode lexicon sudah banyak digunakan dalam banyak penelitian dikarenakan efektif [5].

Terdapat penelitian sebelumnya yang dilakukan menggunakan metode lexicon. Penelitian yang dilakukan oleh Ai Reza Alasi, dkk [6] menyatakan bahwa perbandingan banyak metode

# Paper\_1

*by* Tian Aditya

---

**Submission date:** 24-Jul-2023 06:25PM (UTC+0700)

**Submission ID:** 2136044181

**File name:** ra\_Hayatin\_Aditya\_-\_analisa\_sentimen\_stemming\_lexicon\_python.pdf (249.65K)

**Word count:** 3190

**Character count:** 19836

## Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia

Abdul Hadiy Dyo Fatra<sup>1</sup>, Nur Hayatin<sup>2</sup>, Christian Sri Kusuma Aditya<sup>3</sup>

Universitas Muhammadiyah Malang

abdulhadiy8@gmail.com<sup>1</sup>, noorhayatin@gmail.com<sup>2</sup>, christianskaditya@umm.ac.id<sup>3</sup>

### Abstrak

Penelitian ini mengajukan sebuah klasifikasi terhadap respon masyarakat terhadap keputusan pemerintah untuk memindahkan Ibu kota Indonesia menggunakan metode lexicon. Hasil akurasi pengujian diukur dengan menggunakan confusion matrix. Data pada penelitian ini menggunakan data dari twitter yang berupa tweet. Data berisi tweet respon masyarakat terhadap keputusan perpindahan Ibu kota Indonesia. Data melewati 5 proses preprocessing yaitu case folding, punctuation removal, stopword removal, stemming, dan tokenizing. Lexicon digunakan karena menghasilkan nilai akurasi yang baik. Pada penelitian ini juga akan mencari kamus yang memiliki hasil klasifikasi paling baik. Hasil penelitian ini menunjukkan hasil klasifikasi yang baik dengan mendekati hasil oleh pakar.

**Kata kunci:** Analisa sentimen, Stemming, Lexicon, Python

### Abstract

This study proposes a classification of public response to the government's decision to move the Indonesian capital using the lexicon method. The results of testing accuracy are measured using a confusion matrix. The data in this study use data from Twitter in the form of tweets. The data contains tweets of community responses to the decision to move the Indonesian capital. Data passes through 5 preprocessing processes, namely case folding, punctuation removal, stopword removal, stemming, and tokenizing. Lexicon is used because it produces good accuracy values. In this study also will look for a dictionary that has the best classification results. The results of this study show the results of a good classification by approaching the results by experts.

**Keywords:** Sentiment analysis, Stemming, Lexicon, Python

### 1. Pendahuluan

Respon masyarakat terhadap keputusan pemerintah adalah hal penting dalam kegiatan negara. Seiring berjalannya waktu masyarakat memiliki kecenderungan untuk mengutarakan pendapat mereka dalam sosial media dibandingkan dengan mengutarakan aspirasinya secara langsung. Salah satu wadah untuk menampung respons masyarakat adalah sosial media, sehingga pendapat masyarakat di dalam sosial media menjadi aspek yang bisa dijadikan tolak ukur kepuasan masyarakat terhadap keputusan yang di buat oleh pemerintah. Sosial media adalah beberapa perangkat lunak yang membuat individu atau komunitas dapat berbagi, berkomunikasi, berkumpul, dan berkolaborasi [1]. Sosial media yang memiliki fokus untuk diskusi topik adalah twitter. Pada kuartal pertama 2018 twitter memiliki 336 juta pengguna aktif [2]. Oleh sebab itu twitter merupakan indikator yang baik dalam penelitian Analisa sentimen [3].

Analisa sentimen merupakan satu contoh dari penelitian teks mining. Analisa sentimen adalah riset komputasional dari opini dan emosi yang terdapat dalam sebuah teks [4]. Tujuan dari analisa sentimen adalah untuk mengetahui opini pengguna terhadap suatu topik. Respon terhadap topik dapat dikelompokkan menjadi dua yaitu positif dan negatif. Sehingga dalam melakukan penelitian analisa sentimen kita dapat mengetahui respon pengguna terhadap tokoh, keputusan, produk, dan lain sebagainya. Metode lexicon adalah salah satu metode klasifikasi teks yang banyak digunakan dalam penelitian. Proses pada metode ini adalah dengan cara memeriksa dokumen yang sudah *clean* dengan kamus lexicon yang sudah disiapkan sebelumnya. Kamus lexicon terdiri dari banyak kata yang memiliki nilai positif dan negatif. Metode lexicon sudah banyak digunakan dalam banyak penelitian dikarenakan efektif [5].

Terdapat penelitian sebelumnya yang dilakukan menggunakan metode lexicon. Penelitian yang dilakukan oleh Ali Reza Alaei, dkk [6] menyatakan bahwa perbandingan banyak metode

sentimen analisis dalam topik pariwisata memiliki kesimpulan yaitu penelitian yang menggunakan metode lexicon memiliki peningkatan hasil sentimen paling tinggi dibandingkan dengan metode lain. Penelitian yang dilakukan oleh Hidayatullah Himawan, dkk [7] menunjukkan bahwa pengujian pada sistem memperlihatkan metode lexicon memberikan akurasi dan presisi yang lebih baik daripada metode support vector machine dengan nilai akurasi 87.78% dan presisi sebesar 94.23%. Penelitian yang dilakukan oleh Veronikha Effendy [8] menunjukkan bahwa secara umum, nilai akurasi diatas 70% bisa dikatakan bahwa analisa sentimen dapat diterapkan dalam kasus pengelolaan sampah. Namun, nilai akurasi tergantung pada jumlah kata pada kamus lexicon yang digunakan. Penelitian yang dilakukan oleh Yufis Azhar [9] menunjukkan metode lexicon dapat meningkatkan nilai recall dan mempertahankan nilai recall. Metode lexicon merupakan metode yang tepat dalam sentimen analisis teks.

Pada penelitian ini, diajukan perbandingan hasil klasifikasi dengan menggunakan beberapa kamus. Kamus yang digunakan yaitu kamus dengan topik politik, kamus gabungan, dan kamus yang memiliki bobot diatas rata-rata. Hal tersebut bertujuan untuk mengetahui kamus mana yang memiliki nilai akurasi paling baik. Lexicon digunakan karena merupakan metode analisa sentimen yang baik. Dengan menggunakan lexicon diharapkan akan mendapatkan hasil yang akurat dan dapat mendapatkan fakta tentang respon masyarakat Indonesia terhadap keputusan perpindahan ibu kota.

## 2. Metode Penelitian

### 2.1. Pengumpulan Data

Data yang akan digunakan dalam penelitian ini adalah data *tweet* masyarakat Indonesia pada topik perpindahan Ibu kota Indonesia 2019. Data *tweet* yang diambil adalah yang menggunakan *hashtag* "#IbuKotaBaru". *Hashtag* tersebut digunakan karena merupakan *hashtag* resmi dari akun twitter Kementerian Perencanaan Pembangunan Nasional dengan *user name* BappenasRI

Data diambil menggunakan proses *crawling* melalui API *twitter*. API *twitter* adalah fasilitas dimana *twitter* mengizinkan pihak *developer* lain untuk mengakses informasi dari *twitter*. Untuk dapat menggunakan API *twitter* terlebih dahulu harus mendaftar menjadi *developer* *twitter* di situs <http://dev.twitter.com> dengan menggunakan akun *twitter* yang telah kita daftarkan. Setelah login, kita dapat membuat App baru untuk mendapatkan *consumer key*, *consumer secret*, *access token*, dan *access token secret* yang digunakan untuk dapat mengakses informasi dari *twitter*. Pada proses *crawling* ini akan menggunakan library *tweepy*.

Tabel 1. Contoh Data Hasil Crawling

No	Tweet
1.	pendekatan perencanaan #ibukotabaru yang top down, tidak partisipatif cerminan dari bakal ibukota elitis
2.	jangan sampai #ibukotabaru justru jadi pulau sendiri, yg terisolasi. jauh dari warganya, dari harapan dan aspirasi
3.	PKS Kritik Keputusan Presiden Jokowi Tentang Investasi Ibu Kota Baru. PKS lebih baik kebanjiran daripada pindah ibu
4.	Permasalahkan Investor Jepang Di Ibukota Baru, PKS Tutupi Keberhasilan Jokowi Gandeng Pangeran Arab.
5.	dukung pembangunan ibu kota baru

Pada Tabel 1 merupakan contoh data yang sudah di *crawling* menggunakan API *twitter*. Data tersebut belum siap untuk di klasifikasikan karena masih memiliki banyak noise. Maka selanjutnya akan dilakukan *preprocessing* untuk menghasilkan data yang clean.

### 2.2. Preprocessing

Tahapan pada *preprocessing* yang dilakukan pada penelitian ini disesuaikan pada kebutuhan penelitian yang dilakukan. Proses *preprocessing* yang sesuai dengan data set berupa teks. Tahapan tersebut terdiri dari case folding, punctuation removal, stopword removal, stemming, dan tokenizing.

### 2.2.1. Case Folding

Pada tahap ini akan dilakukan untuk menyamakan kecil besarnya huruf pada teks. Proses dilakukan untuk mempermudah dalam pencarian data. Hasil dari proses ini adalah semua huruf menjadi huruf kecil.

### 2.2.2. Punctuation Removal

Tahap ini akan menghapus tanda baca dan angka dalam proses analisa sentimen, seperti koma (,), titik (.), 1, 2, dan lainnya. Hasil yang di dapatkan pada proses ini adalah kata-kata yang bersih dari tanda baca dan angka.

### 2.2.3. Stopword Removal

Pada tahap ini akan menghapus kata yang tidak digunakan pada proses analisa sentimen Seperti "dan", "atau", dan lainnya. Hasil yang di dapatkan dalam proses ini adalah kata kata yang tidak termasuk dalam *stopword*.

### 2.2.5. Stemming

Proses ini akan mengubah kata berimbuhan menjadi kata dasar. Imbuhan yang dimaksud terdiri dari imbuhan pada awal kata (*prefiks*), imbuhan pada akhir kata (*suffiks*), imbuhan pada depan dan akhir kata (*konfiks*), imbuhan pada tengah kata (*kemilau*). Proses ini bertujuan agar dapat memudahkan proses klasifikasi. Contoh dari proses ini seperti "mengetik" menjadi "ketik". Pada penelitian ini digunakan *library* sastrawi stemmer dalam proses *stemming*.

### 2.2.6. Tokenizing

Proses ini akan dilakukan pemisahan kata dalam kalimat sehingga dapat dilakukan pembobotan pada setiap kata. Hasil pada proses ini adalah kata kata yang sudah terpisah dari kalimat.

## 2.3. Klasifikasi Dengan Menggunakan Metode Lexicon

Metode lexicon adalah salah satu metode klasifikasi teks yang banyak digunakan dalam penelitian. Proses pada metode ini adalah dengan cara memeriksa dokumen yang sudah clean dengan kamus lexicon yang sudah disiapkan sebelumnya. Kamus lexicon terdiri dari banyak kata yang memiliki nilai positif dan negatif. Metode lexicon sudah banyak digunakan dalam banyak penelitian dikarenakan efektif [5].

Kamus Lexicon sendiri terdiri dari kumpulan kata beserta nilai sentimennya yang digunakan untuk mengetahui emosi di dalam suatu kalimat. Nilai sentimen pada kamus lexicon terdiri dari nilai positif dan nilai negatif. Klasifikasi menggunakan metode Lexicon digunakan dengan cara pengecekan kata yang ada pada kalimat dengan kamus lexicon yang disiapkan. Setelah itu kata yang cocok dengan kamus lexicon lalu dicatat skornya.

$$Sentence_{sentiment} \begin{cases} positive & \text{if } S_{positive} > S_{negative} \\ neutral & \text{if } S_{positive} = S_{negative} \\ negative & \text{if } S_{positive} < S_{negative} \end{cases}$$

Gambar 1. Pemberian Label Terhadap Data yang Sudah Memiliki Bobot

Berdasarkan Gambar 1 menunjukkan pelabelan terhadap data yang telah diberi bobot dengan metode lexicon. Jika data memiliki nilai sentimen positif lebih besar dari sentimen negatif maka akan dilabeli dengan data dengan sentimen positif. Jika data memiliki nilai sentimen positif lebih kecil dari sentimen negatif maka akan dilabeli dengan data dengan sentimen negatif. Jika data memiliki nilai sentimen positif sama dengan sentimen negatif maka akan dilabeli dengan data dengan sentimen netral.

## 2.4. Skenario Pengujian

Pada penelitian ini akan dilakukan pengujian secara intrinstik dengan menggunakan confusion matrix. Hasil klasifikasi yang dihasilkan dengan metode lexicon dibandingkan dengan hasil pakar yang dilakukan oleh Dewi Larasetiani S.pd dengan NIM 201510080311009. Beliau dipilih sebagai pakar dikarenakan merupakan lulusan strata 1 jurusan Pendidikan Bahasa

Indonesia yang memiliki kompetensi pada bidang ini. Lalu akan dilakukan penghitungan nilai akurasi, presisi, recall, dan f-measure [10]. Confusion matrix pada Tabel 2 merupakan salah satu contoh parameter yang biasa digunakan sebagai indikator untuk mengukur dan membandingkan kemiripan hasil prediksi dan data asli.

Tabel 2. Confusion Matrix

		Lexicon	
		Positif	Negatif
Pakar	Positif	BP	SP
	Negatif	SN	BN

$$Akurasi = \frac{BP + BN}{total\ tweet} \quad (1)$$

$$Presisi = \frac{BP\ atau\ BN}{BP + SN\ atau\ SP + BN} \quad (2)$$

$$Recall = \frac{BP\ atau\ BN}{BP + SP\ atau\ SN + SN} \quad (3)$$

$$F - Measure = 2 \times \frac{Presisi \times Recall}{(Presisi + Recall)} \quad (4)$$

Untuk menghitung nilai akurasi, presisi, recall, dan f-measure digunakan persamaan seperti diatas. Akurasi merupakan nilai keberhasilan pengukuran dengan nilai sebenarnya (1). Presisi adalah nilai ketepatan antar informasi yang dicari dengan hasil pengukuran (2). Recall adalah tingkat keberhasilan pengukuran dalam menemukan kembali sebuah informasi (3). F-measure adalah harmonic mean dari presisi dan recall (4)

### 3. Hasil Penelitian dan Pembahasan

Pembahasan terdiri dari 5 skenario. Skenario pertama adalah klasifikasi menggunakan kamus dengan topik politik. Skenario kedua adalah klasifikasi menggunakan kamus gabungan. Skenario ketiga adalah klasifikasi menggunakan kamus dengan yang menyisakan kata yang memiliki bobot diatas rata-rata. Skenario keempat adalah memvisualisasikan perbandingan hasil klasifikasi dengan menggunakan banyak kamus. Skenario kelima adalah membandingkan hasil klasifikasi menggunakan stemming dan tanpa menggunakan stemming. Berikut adalah penjabaran masing-masing skenario:

#### 3.1. Klasifikasi dengan kamus lexicon dengan topik politik

Pada pengujian Tabel 3 ini akan dilakukan pembuatan kamus lexicon dengan topik politik. Pertama, Data yang sudah *clean* di pisahkan menjadi kata yang terpisah satu sama lain. Lalu kata yang sama akan dihapus menjadi satu kata saja. Lalu kata yang termasuk pada *stopword removal* juga dihapus karena tidak memiliki makna. Lalu kata yang dihapus selanjutnya adalah nama seseorang, instansi, perusahaan, dll. Setelah mendapatkan kata lalu akan dilakukan pemberian bobot secara manual seperti yang dilakukan pada penelitian yang dilakukan oleh Fajri Koto [16]. Dari proses pembuatan kamus ini dihasilkan total 537 kata dengan topik politik yang sudah diberikan bobot dengan nilai antara -5 sampai 5. Bobot negatif adalah nilai yang diberikan pada kata yang memiliki makna negatif. Bobot positif adalah nilai yang diberikan pada kata yang memiliki makna positif.

Lalu akan dilakukan proses klasifikasi menggunakan kamus dengan topik politik. Hasil klasifikasi lalu akan dihitung keberhasilan analisa sentimennya dengan confusion matrix. Pada proses klasifikasi ini menghasilkan 872 tweet positif, 127 kata negatif, dan 1 kata netral. Hasil klasifikasi menggunakan kamus dengan topik politik memiliki akurasi sebesar 70%.

Tabel 3. Contoh Kamus Politik

Kata	Bobot
Dungu	-4
Ahli	5
Calon	3
Biaya	-4
Absurd	-5

### 3.2. Klasifikasi dengan menggunakan kamus gabungan

Selanjutnya akan dilakukan proses penggabungan kamus dengan topik politik dengan kamus yang dibuat oleh Fajri Koto, seperti pada Tabel 4. Kamus tersebut memiliki jumlah kata sebanyak 10218 kata. Penggabungan kamus ini dilakukan dengan cara memindahkan kata dan bobot pada kamus politik ke kamus yang dibuat oleh Fajri Koto. Setelah digabungkan dengan kamus dengan topik politik jumlah kata yang dihasilkan berjumlah 10.755 kata.

Tabel 4. Contoh Perbandingan Kamus

	Kamus Topik Politik	Kamus Gabungan
Jumlah Kata	537	10755
Jumlah Kata Positif	366	3976
Jumlah Kata Negatif	171	6780

Lalu akan dilakukan proses klasifikasi menggunakan kamus gabungan. Hasil klasifikasi lalu akan dihitung keberhasilan analisa sentimenya dengan confusion matrix. Pada proses klasifikasi ini menghasilkan 729 tweet positif, 222 kata negatif, dan 49 kata netral. Hasil klasifikasi menggunakan kamus yang saya gabungkan memiliki akurasi sebesar 77%.

### 3.3. Klasifikasi dengan menggunakan kamus yang memiliki kata dengan bobot diatas rata-rata

Selanjutnya akan dilakukan pemotongan kata pada kamus. Kata yang akan dihapuskan adalah kata yang memiliki bobot dibawah rata rata bobot dari kamus tersebut. Hal ini bertujuan untuk menguji kamus seperti apa yang lebih baik digunakan pada proses klasifikasi yang akan dilakukan.

Tabel 5. Contoh Perbandingan Kamus yang Dipotong

	Kamus Topik Politik	Kamus Gabungan
Jumlah Kata	285	5961
Jumlah Kata Positif	177	2663
Jumlah Kata Negatif	108	3298

Setelah mendapatkan kamus dengan kata yang memiliki bobot diatas rata-rata. Lalu akan dilakukan proses klasifikasi menggunakan kamus campuran yang menyisakan kata yang memiliki bobot diatas rata-rata. Hasil klasifikasi lalu akan dihitung keberhasilan analisa sentimenya dengan confusion matrix. Pada proses klasifikasi ini menghasilkan 866 tweet positif, 122 kata negatif, dan 12 kata netral. Hasil klasifikasi menggunakan kamus gabungan yang dipotong menyisakan kata yang memiliki bobot diatas rata-rata memiliki akurasi sebesar 70%.

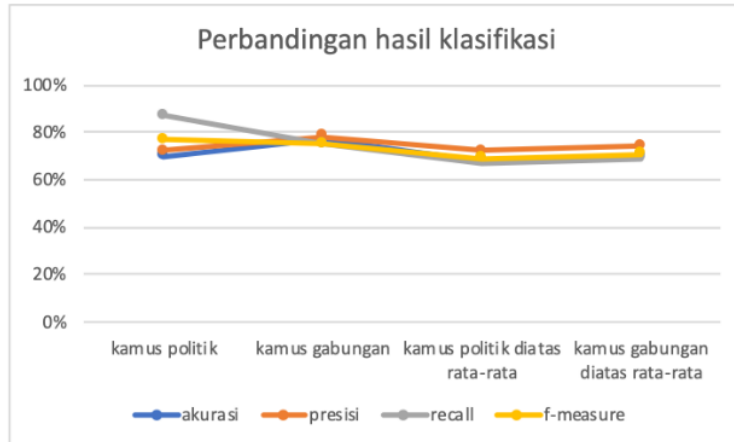
Lalu akan dilakukan proses klasifikasi menggunakan kamus dengan topik politik yang menyisakan kata yang memiliki bobot diatas rata-rata. Hasil klasifikasi lalu akan dihitung keberhasilan analisa sentimenya dengan confusion matrix. Pada proses klasifikasi ini menghasilkan 766 tweet positif, 164 kata negatif, dan 70 kata netral. Sedangkan Hasil klasifikasi menggunakan kamus yang dipotong menyisakan kata yang memiliki bobot diatas rata-rata memiliki nilai akurasi sebesar 68%.

Tabel 6. Contoh Perbandingan Nilai Akurasi Setelah Dilakukan Pemotongan Kamus

Jenis Kamus	Nilai akurasi
Kamus topik gabungan diatas rata rata	70%
Kamus topik politik diatas rata-rata	68%

### 3.4. Menvisualisasikan hasil klasifikasi dengan menggunakan banyak kamus

Untuk melihat perbandingan hasil klasifikasi dari banyak kamus yang digunakan. Maka akan dilakukan proses visualisasi hasil klasifikasi. Proses ini dilakukan dengan menggunakan Microsoft excel. Dengan cara mencatat hasil dari banyak klasifikasi menjadi satu dan mengubahnya menjadi bentuk grafik. Berikut grafik hasil perbandingan hasil klasifikasi.



Gambar 2. Perbandingan Hasil Klasifikasi

Pada Gambar 2 dapat diketahui nilai akurasi menunjukkan bahwa proses klasifikasi menggunakan kamus gabungan tanpa pemotongan kata yang berbobot diatas rata-rata memiliki nilai akurasi yang paling tinggi. Maka kamus yang akan digunakan dalam proses klasifikasi adalah kamus gabungan tanpa scenario pemotongan kata pada kamus.

### 3.5. Perbandingan klasifikasi menggunakan stemming dan tanpa stemming

Lalu akan dilakukan pengaruh stemming dalam proses klasifikasi menggunakan metode lexicon. Untuk mengetahui pengaruh stemming terhadap proses klasifikasi dengan menggunakan metode lexicon.

**Positive: 840 (84.1%)**  
**Negative: 132 (13.2%)**  
**Neutral: 27 (2.7%)**

Gambar 3. Hasil Klasifikasi Tweet dengan Stemming menggunakan Metode Lexicon

Pada Gambar 3 menunjukkan hasil klasifikasi yang dilakukan oleh pakar mengklasifikasikan jumlah *tweet* bernilai positif berjumlah 602, *tweet* bernilai negatif berjumlah 370, dan *tweet* bernilai netral berjumlah 28. Proses klasifikasi menggunakan kamus gabungan dan menggunakan stemming mengklasifikasikan jumlah *tweet* bernilai positif berjumlah 841, *tweet* bernilai negatif berjumlah 132, dan *tweet* bernilai netral berjumlah 27. Dari hasil tersebut di dapatkan nilai sentimen negatif paling tinggi bernilai -17. Untuk nilai sentimen positif paling tinggi bernilai 41. Dari hasil pengujian analisa sentimen menggunakan *stemming* yang menggunakan rumus yang terdapat pada bab 2 dapat diketahui bahwa nilai akurasi adalah 58%. Nilai rata-rata presisi adalah 52%. Nilai rata rata recall adalah 56%. Nilai f-measure adalah 53%.

**Positive: 728 (72.9%)**  
**Negative: 222 (22.2%)**  
**Neutral: 49 (4.9%)**

Gambar 4. Hasil Klasifikasi Tweet dengan Stemming menggunakan Metode Lexicon



Pada Gambar 4 menunjukkan hasil klasifikasi menggunakan kamus gabungan tanpa menggunakan *stemming* dalam proses *preprocessing* mengklasifikasikan jumlah *tweet* bernilai positif berjumlah 729, *tweet* bernilai negatif berjumlah 222, dan *tweet* bernilai netral berjumlah 49. Dari hasil tersebut di dapatkan nilai sentimen negatif paling tinggi bernilai -24. Untuk nilai sentimen positif paling tinggi bernilai 47. Dari hasil pengujian analisa sentimen tanpa menggunakan *stemming* yang menggunakan rumus yang terdapat pada bab 2 dapat diketahui bahwa nilai akurasi adalah 77%. Nilai rata-rata presisi adalah 78%. Nilai rata-rata recall adalah 75%. Nilai f-measure adalah 75%.

Dari hasil perbandingan perhitungan diatas diketahui bahwa nilai rata-rata akurasi, presisi, recall, dan f-measure proses analisa sentimen menggunakan kamus lexicon gabungan menunjukkan hasil yang paling baik dibandingkan menggunakan kamus hanya dengan topik politik. Melihat hasil pengujian diatas, dapat diketahui bahwa nilai akurasi pada proses klasifikasi menggunakan lexicon tergantung pada jumlah kata pada kamus yang digunakan. Semakin kata pada kamus banyak dan relevan dengan data set yang di klasifikasikan akan meningkatkan nilai akurasi pada proses klasifikasi yang dilakukan.

#### 4. Kesimpulan

Hasil dari penelitian analisa sentimen pada topik perpindahan ibu kota Indonesia dengan menggunakan kamus gabungan memiliki nilai akurasi paling baik dibandingkan menggunakan kamus lainnya. Pada proses klasifikasi menggunakan kamus gabungan menghasilkan 729 tweet positif, 222 kata negatif, dan 49 kata netral. Hasil klasifikasi menggunakan kamus digabungkan memiliki akurasi sebesar 77%. Nilai rata-rata presisi adalah 78%. Nilai rata-rata recall adalah 75%. Nilai rata-rata f-measure adalah 75%.

Dari proses klasifikasi dapat disimpulkan bahwa kamus yang digunakan sangat menentukan nilai akurasi yang dihasilkan. Proses *stemming* juga memiliki dampak yang besar dalam proses klasifikasi.

#### Daftar Notasi

Keterangan Confusion Matrix:

BP	: Benar Positif
SP	: Salah Positif
SN	: Salah Negatif
BN	: Benar Negatif

#### Referensi

- [1] A. C. Sari *et al.*, "Komunikasi dan Media Sosial," no. December, 2018.
- [2] A. Majumdar and I. Bose, "Do tweets create value? A multi-period analysis of Twitter use and content of tweets for manufacturing firms," *Int. J. Prod. Econ.*, vol. 216, no. August 2018, pp. 1–11, 2019.
- [3] J. Weng, E. P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topic-sensitive influential twitterers," *WSDM 2010 - Proc. 3rd ACM Int. Conf. Web Search Data Min.*, pp. 261–270, 2010.
- [4] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017.
- [5] S. Deng, A. P. Sinha, and H. Zhao, "Adapting sentiment lexicons to domain-specific social media texts," *Decis. Support Syst.*, vol. 94, pp. 65–76, 2017.
- [6] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2019.
- [7] H. Himawan, W. Kaswidjanti, A. Sentimen, M. Sosial, and L. Based, "Metode Lexicon Based Dan Support Vector Machine Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit," vol. 2018, no. November, pp. 235–244, 2018.
- [8] V. Effendy, "Analisis Sentimen Berbahasa Indonesia Dengan Pendekatan Lexicon Based ( Studi Kasus : Solusi Pengelolaan Sampah )," *J. Ilm. Komput. dan Inform. ( KOMPUTA )*, vol. 4, no. 1, pp. 55–60, 2015.
- [9] Y. Azhar, "Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 3, p. 237, 2018.
- [10] F. Krüger, "Activity, Context, and Plan Recognition with Computational Causal Behaviour Models," *ResearchGate*, no. August, 2018.



# Paper\_1

---

## ORIGINALITY REPORT

---

17%

SIMILARITY INDEX

15%

INTERNET SOURCES

5%

PUBLICATIONS

6%

STUDENT PAPERS

---

## MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

2%

★ Submitted to Universitas Brawijaya

Student Paper

---

Exclude quotes  On

Exclude matches  Off

Exclude bibliography  On