

ANALISA SENTIMEN PERSEPSI MASYARAKAT TERHADAP PEMINDAHAN IBUKOTA BARU DI KALIMANTAN TIMUR PADA MEDIA SOSIAL TWITTER

Icha Adellia Safra¹, Eri Zuliarso²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: ¹icha.adellia97@gmail.com, ²eri299@edu.unisbank.ac.id

Abstrak

Pada Senin 26 Agustus 2019 atau setelah 74 tahun Indonesia merdeka, presiden terpilih Joko Widodo melalui kanal Youtube resmi Sekretariat Presiden mengumumkan bahwa pemindahan Ibu Kota Indonesia yang baru yaitu di wilayah administratif Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Kalimantan Timur. Penelitian ini bertujuan untuk mengetahui persepsi Masyarakat Indonesia di Twitter mengenai pemindahan Ibukota dengan menggunakan kata kunci #IbuKotaBaru dan #IbuKotaPindah setelah itu data disimpan dalam database MySQL, lalu dilakukan proses text processing. Proses klasifikasi text dibagi menjadi kelas sentimen positif dan negatif, algoritma Naïve Bayes Classifier dibutuhkan untuk itu. Data yang digunakan sebesar 200 data tweets, yang terdiri dari 159 data training dan 41 data testing menghasilkan akurasi sebesar 78%. Visualisasi tab tabel data asli, tabel data test, histogram, wordcloud serta confusion matrix ditampilkan menggunakan R Shiny. Aplikasi web ini dapat diakses pengguna lainnya melalui internet.

Kata Kunci: Pemindahan Ibukota, Naïve Bayes Classifier, Kalimantan Timur

1. PENDAHULUAN

Pada Senin 26 Agustus 2019 atau lebih tepatnya 9 hari setelah kemerdekaan Indonesia atau tepatnya setelah 74 tahun Indonesia merdeka, presiden terpilih Joko Widodo melalui kanal Youtube resmi Sekretariat Presiden mengumumkan bahwa pemindahan Ibu Kota Negara Kesatuan Republik Indonesia yang baru yaitu di Wilayah Administratif Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Kalimantan Timur. Setelah pengumuman itu terdapat banyak pro dan kontra yang terjadi di masyarakat. Dahulu sangat sulit untuk masyarakat mengungkapkan pendapat maupun kritik karena hanya bisa melalui media cetak. Namun sekarang karena perkembangan teknologi yang sangat pesat terutama teknologi komunikasi membuat masyarakat sangat mudah untuk menyampaikan opini, kritik maupun saran dalam media sosial, dalam hal ini masyarakat menyampaikan aspirasinya terhadap pemutusan pemindahan Ibu Kota. Salah satu media untuk menyampaikan aspirasi yang sangat populer di kalangan masyarakat adalah Twitter. Kemunculan twitter menjadikan sarana untuk penggunanya supaya bisa mengirim pesan maupun berita terkini, gagasan, pendapat (opini) maupun curahan hati[1]. Sehingga memunculkan banyak persepsi, tidak hanya persepsi yang positif adapula persepsi negatif. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari *text mining*, dalam hal ini adalah klasifikasi dokumen[2]. *Naïve Bayes Classifier* merupakan salah satu algoritma yang paling banyak digunakan untuk pengklasifikasian suatu dokumen karena algoritma ini mempunyai kecepatan dan akurasi yang tinggi ketika digunakan dalam basis data yang cukup besar dan data yang beragam[3]. Maka dari itu inilah tujuan penulis mengapa menggunakan algoritma tersebut supaya mendapatkan akurasi yang tinggi untuk menentukan suatu opini negatif ataupun positif pada data twitter[4]. Dalam hal ini menggunakan bahasa pemrograman R, software RStudio dan *Shiny*[5]. Sehingga dalam menyelesaikan permasalahan tersebut dilakukan “Analisis Sentiment Pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes Classifier*”, terhadap pemindahan Ibu Kota Indonesia di Kalimantan Timur.

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Pada penelitian ini, metode yang digunakan untuk mendapatkan data informasi dan untuk menyelesaikan masalah adalah sebagai berikut :

a. Studi Kepustakaan

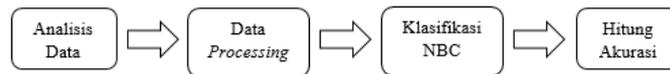
Dalam pengumpulan data menggunakan metode kepastakaan yang dilakukan dengan pengumpulan jurnal, paper, buku, maupun situs internet sebagai acuan yang ada kaitannya dengan materi penulis khususnya klasifikasi dengan algoritma *Naïve Bayes Classifier*.

b. Pengumpulan Data *Tweet* dari Twitter

Untuk mendapatkan data dengan cara mengambil data dari twitter dengan menggunakan kata kunci #IbuKotaBaru dan #IbuKotaPindah. Data yang sudah diperoleh disimpan di database MySQL.

2.2 Metode Analisis Data

Metode yang digunakan untuk menganalisis data *tweet* untuk mengetahui kelas sentimennya adalah sebagai berikut :



Gambar 1. Alur Sistem Analisis Sentimen

Berikut adalah penjelasan mengenai alur sentimen pada Gambar 1:

- a. Analisis Sumber Data
 Dalam penelitian ini menggunakan API Twitter untuk mengambil data dari twitter dan disimpan dalam bentuk file Ms.Excel dalam format CSV lalu diimport ke database, langkah selanjutnya yaitu menentukan klasifikasi secara manual terhadap data yang sudah ada. Data berjumlah 200 data tweet yang terdiri dari 159 data *training* dan 41 data *testing*.
- b. Data *Preprocessing*
 Dalam proses ini, klasifikasi tweet di twitter yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri *case folding*, *normalisasi*, *tokenizing*, dan *stopword removal*. Tujuannya adalah agar data yang telah didapatkan lebih tersusun rapi dan akan lebih memudahkan dalam melakukan pengolahan data.
- c. Klasifikasi *Naïve Bayes Classifier*
 Dalam proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier* dengan cara menghitung probabilitas pada setiap tweet yang akan menentukan opini tersebut masuk dalam opini yang negatif atau opini yang positif.
- d. Hasil Akurasi
 Tahap terakhir adalah menghitung tingkat akurasi, data yang akan dilakukan pengujian merupakan data yang sudah diketahui, sehingga setelah melakukan proses pengklasifikasian maka langkah selanjutnya menghitung tingkat akurasi atau ketepatan sistem dalam proses pengklasifikasian. Menggunakan metode pengukuran tingkat akurasi *Confusion Matrix*.

2.3 Metode Klasifikasi *Naive Bayes Classifier*

Secara umum proses ini dibagi menjadi beberapa tahap. Antara lain sebagai berikut :

- a. Data uji yang belum diklasifikasi akan melalui tahap *preprocessing text*. Tahapan *preprocessing text* terdiri dari *case folding*, *normalisasi*, *tokenizing*, dan *stopword removal*.
- b. Setelah dilakukan *preprocessing* maka data *tweet* selanjutnya akan dilakukan penghitungan frekuensi *term*.
- c. Kemudian menghitung nilai V_{map} untuk tiap kelas dengan persamaan rumus :

$$V_{map} = \underset{V_j \in V}{argmax} \prod_{i=1}^n P(x_i|V_j)P(V_j) \tag{1}$$

Keterangan :

- V_j Kategori tweet $j=1, 2, \dots, n$. Dimana dalam penelitian ini
 $j1$ = kategori tweet sentimen positif,
 $j2$ = kategori tweet sentiment negatif
- $P(x_i|V_j)$ Probabilitas x_i pada kategori V_j
- $P(V_j)$ Probabilitas dari V_j

Kelas suatu *tweet* ditentukan berdasarkan nilai V_{map} terbesar.

- d. Hitung akurasi
 Untuk menghitung tingkat akurasi dari klasifikasi sistem menggunakan *confusion matrix*. Data hasil klasifikasi akan dihitung berapa banyak data yang benar dan berapa banyak data yang salah. Untuk menghitung tingkat akurasi maka menggunakan persamaan rumus :

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

Keterangan :

- TP : True Positif
- TN : True Negatif
- FP : False Positif
- FN : False Negatif

3. HASIL DAN PEMBAHASAN

Data *tweet* diperoleh dari media sosial Twitter dengan kata kunci pencarian #IbuKotaBaru dan #IbuKotapindah. Data yang diunduh dari Twitter berupa *tweet* bahasa Indonesia dengan jumlah data 200 *tweet*. kemudian data akan disimpan ke dalam database MySQL. Data yang dibutuhkan dalam penelitian ini terdiri dari dua jenis, yaitu data latih dan data uji. Pembagian data latih dan data uji dapat dilihat pada Tabel 1.

Tabel 1. Pembagian Data

Jenis Sentimen Tweet	Positif	Negatif
Latih	103	56
Uji	29	12

3.1 Pembacaan Data

Setelah menentukan pembagian data latih dan data uji, selanjutnya pembacaan data ke dalam RStudio. Tabel 2 adalah partisi data latih, Tabel 3 adalah partisi dari data uji.

Tabel 2 Data Latih

class	text	username	argumentasi
28 neg	Pemindahan ibukota Indonesia gak perlu jauh2 ke Kalimantan...	inihansedbert	Pemindahan ibukota tak perlu jauh2
80 pos	Setuju bgt aku dengan ibu baru dengan alasan dan sebab.k...	cntyafm	Setuju bgt aku
150 neg	Mohon maaf sy menolak #bukotabaru Indonesia pindah ke...	hansedbert18	Menolak
101 pos	Moeldoko Sebut PLTA Sungai Kayan Dapat Mendukung Pas...	TribunNew5	Mendukung
111 neg	Dari Januari hingga Agustus 2019 belanja modal di Kement...	detikcom	Masih sangat rendah
137 pos	Ridwan Kamil jadi juri sayembara desain ibukota di Kaltim #I...	KontanNews	Ridwan Kamil jadi juri sayembara desain
133 pos	Wakil Bupati Penajam Hamdam Sebut ASN di Calon Ibu Kot...	tribunkaltim	Calon ibukota baru sudah mumpuni
166 pos	Demi bangsa Indonesia yang semakin maju #IbuKotaBaru	BappenasRI	Maju
144 pos	Ibu Kota Baru Bisa Tampung 3 Juta Jiwa pada tahap awal na...	beritaKBR	Ibukota baru bisa tampung 3 juta jiwa pad

Tabel 3 Data Uji

class	text	username	argumentasi
68 neg	Warga Penajam Paser Utara Tolak Jargas, Pemkab Edukasi P...	tribunkaltim	Tolak
91 pos	Panitia Khusus (Pansus) pemindahan Ibu Kota Negara (IKN) ...	detikfinance	Menggelar rapat kajian
55 pos	Mengikuti jejak Indonesia, kini, Thailand pun ingin pindah Ib...	RiefqiMuna	Thailand pun ingin pindah
12 pos	Menteri Perhubungan Budi Karya Sumadi hari ini melakukan...	detikfinance	Meninjau
47 pos	Saya senang ibukota pindah #IbuKotaBaru	RedaksiMerdeka	Saya senang
156 pos	Ibu Kota Pindah ke Kaltim, Balikpapan Tak Mau Bernasib sep...	beritaKBR	Tak mau bernasib seperti Jakarta
125 pos	Pemerintah dapat mengurangi impor LPG dan pengurangan...	minhyunsc	Mengurangi impor
44 neg	4. Pemerintah melalui Kementerian Lingkungan Hidup dan K...	notbuzzer	Mempersiapkan Pelepasan Status
18 pos	Malah bikin lancar jaya. Gak akan ada konflik dgn masyarak...	bumiboy	Gak akan ada konflik

3.2 Text Processing

Langkah selanjutnya yaitu *text processing*, tindakan yang dilakukan pada tahap ini adalah tokenisasi, mengubah semua huruf menjadi huruf kecil, menghapus tanda baca, menghapus angka dan menghapus kata – kata (*stopword removal*) yang tidak memiliki makna sehingga tidak berpengaruh terhadap proses analisis sentimen. *Script* untuk pemrosesan teks dapat dilihat pada Gambar 2.

```
corpus.clean <- corpus %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removewords, stopwords(kind="en")) %>%
  tm_map(stripwhitespace)
```

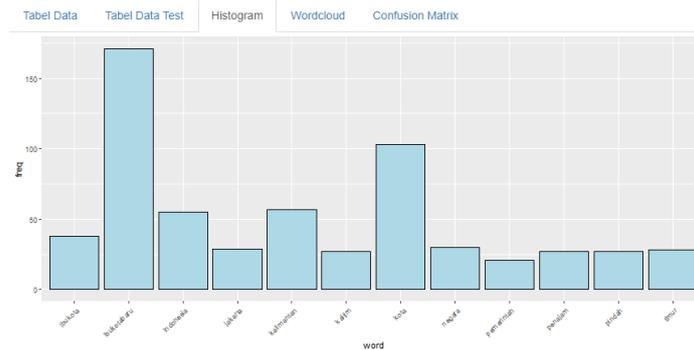
Gambar 2. Script Pemrosesan Teks

3.3 Pelatihan Model Naive Bayes

Untuk melatih model, penelitian ini menggunakan fungsi *naive bayes* dari paket ‘e1071’. Karena *Naive Bayes* mengevaluasi probabilitas, maka memerlukan beberapa cara untuk menetapkan probabilitas tidak nol pada kata-kata yang tidak muncul dalam sampel. Dengan menggunakan *Laplace 1 smoothing* untuk tujuan ini. Gambar 3 merupakan *script* model *naive bayes classifier* sedangkan hasil klasifikasi data uji dapat dilihat pada Gambar 4.

```
# Apply the convert_count function to get final training and testing DTMs
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)
# Train the classifier
system.time( classifier <- naiveBayes(trainNB, df.train$label, laplace = 1) )
# Use the NB classifier we built to make predictions on the test set.
system.time( pred <- predict(classifier, newdata=testNB) )
#Create a truth table by tabulating the predicted class labels with the actual class labels
table("Predictions"= pred, "Actual" = df.test$label)
print((pred))
```

Gambar 3. Model Naive Bayes Classifier



Gambar 9. Visualisasi Wordcloud

Dalam Gambar 7 menampilkan daftar kata yang paling sering muncul di dalam data *tweet*. Gambar 7 menampilkan daftar kata yang paling sering muncul dalam bentuk wordcloud. Gambar 9 Gambar 7 menampilkan daftar kata yang paling sering muncul dalam bentuk histogram. Daftar 10 kata dengan kemunculan yang paling sering yaitu ibukotabaru, kota, indonesia, ibukota, negara, jakarta, timur, pindah, kaltim.

3.6 Hasil Penelitian

Dari hasil *confusion matrix* tersebut, maka didapatkan akurasi sebesar :

$$\frac{TP+TN}{TP+FP+TN+FN} = \frac{29+3}{29+9+3+0} = \frac{32}{41} = 0,78$$

Hasil akurasi algoritma *naive bayes classifier* sebesar 0,76 atau 76% dalam mengklasifikasikan *tweet* ke dalam sentimen kelas positif dan negatif. Dengan hasil *confusion matrix* tersebut, maka didapat hasil *precision* untuk masing-masing kelas dan *rate error* sistem sebagai berikut :

a. Precision Positif

$$\frac{TP}{TP+FP} = \frac{29}{29+9} = \frac{29}{38} = 0,763$$

Hasil akurasi sentimen positif sebesar 0,763 atau 76% sehingga tingkat kesalahan klasifikasi kelas yaitu sebesar 24%.

b. Precision Kelas Negatif

$$\frac{TN}{TN+FN} = \frac{3}{3+0} = \frac{3}{3} = 1$$

Tingkat ketepatan kelas sentimen positif sebesar 1%

c. Error Rate Sistem

$$\frac{FP+FN}{TP+TN+FN+FP} = \frac{9+0}{29+3+0+9} = \frac{9}{41} = 0.219$$

Berdasarkan hasil dari *presicion* kelas positif dan *precision* kelas negatif, tingkat kesalahan mengklasifikasikan kelas sentimen dipengaruhi oleh tingkat akurasi dari jumlah data positif yang benar diklasifikasi sebagai data positif dan jumlah data negatif yang benar diklasifikasi sebagai data negatif. Sedangkan hasil dari *error rate* sistem, kesalahan hasil prediksi klasifikasi yang terdapat dalam tabel *confusion matrix* dipengaruhi oleh tidak seimbangnya jumlah data antara kelas sentimen negatif dengan kelas sentimen positif yang ada dalam data latih mengakibatkan sedikitnya referensi untuk kata-kata positif yang akan berpengaruh pada tahap klasifikasi untuk data uji mengalami beberapa kesalahan prediksi.

4. KESIMPULAN

Penelitian ini melakukan klasifikasi tweet sentimen negatif dan sentimen positif dengan Metode *Naive Bayes Classifier* dengan data latih 159 dan data uji 41 dan mendapatkan akurasi sebesar 78%. Analisis Sentimen terbukti dapat digunakan untuk mengetahui sentimen masyarakat khususnya kepada pengguna twitter terhadap Pemandahan Ibu Kota Indonesia ke Kalimantan Timur, sehingga membantu masyarakat awam untuk mengetahui sentimen masyarakat lainnya terhadap Pemandahan Ibu Kota Indonesia. Penelitian ini menggunakan web aplikasi yang interaktif R Shiny yang bisa diakses seluruh pengguna dengan alamat https://ichaadellia.shinyapps.io/PemandahanIbukota_NBC/

DAFTAR PUSTAKA

- [1] Castillo ,C. , Mendoza, M., Poblete , B. (2011). Information credibility on twitter. Proceedings of the 20th international conference on World wide web, March 2011 Pages 675–684<https://doi.org/10.1145/1963405.1963500>.
- [2] Bolen, Johan., Mao, Huina., Zeng, Xiaojung. (2011) Twitter mood predicts the stock market. *Journal of Computational Science*, 2, pp.1–8.
- [3] Natalius, Samuel. (2010) Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen, : *Makalah II2092 Probabilitas dan Statistik*.
- [4] Liu, B. (2012) Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, Morgan & Claypool Publisher, 5(1), pp.1-167.
- [5] Muthia, D. A. (2017) Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes, *Jurnalilmu Pengetahuan Dan Teknologi Komputer*, 2, (2), pp. 39-45.