**SEBARAN LOKASI PENGGUNA TWITTER BERDASARKAN SENTIMEN TERHADAP OPINI PEMINDAHAN IBUKOTA REPUBLIK INDONESIA**

Fathiyarizq Mahendra Putra1**,** I Wayan Santiyasa2**,** Ida Bagus Gede Dwidasmara 3

*Program Studi Teknik Infomatika*, Universitas Udayana

Jl. Kampus Bukit Jimbaran, Indonesia

1fathiyarizq.mahendra@gmail.com

2santiyasa67@gmail.com (Corresponding author)

3dwidasmara@cs.unud.ac.id

*Program Studi Teknik Infomatika*, Universitas Udayana

Jl. Kampus Bukit Jimbaran, Indonesia

***Abstrak***

*Pemindahan lokasi ibukota republik indonesia menjadi isu terhangat saat ini, Upaya pemindahan ibu kota Indonesia dimulai pada tahun 2019 pada masa kepresidenan Joko Widodo. Selain alasan umum yakni pertimbangan sosial ekonomi, pertimbagan ekonomi, serta pertimbangan politik. pemilihan lokasi baru ibukota yakni wilayah di Penajam Paser Utara. Namun dari pernyataan dan keputusan pemerintah mengenai pemindahan lokasi ini terdapat pro dan kontra dikalangan masyarakat indonesia, terutama pengguna media sosial twitter. Dalam penelitian ini, mencoba mencari sebaran lokasi pengguna twitter terhadap pemindahan Ibukota Republik Indonesia dari pengguna twitter yang memberikan opini, serta melakukan klasifikasi sentimen menggunakan metode SVM (Support Vector Machine), hal yang dilakukan pada penelitian ini adalah pengumpulan data, preprocessing, klasifikasi sentimen, pengujian dan evaluasi, pencarian lokasi, dan visualisasi kedalam bentuk peta. Dari hasil pengujian dengan kernel linear didapatkan hasil terhadap data uji dan data latih dapat mempengaruhi kinerja dari sistem yang dibangun. Pada hal ini didapatkan hasil dengan menggunakan rasio data uji dan data latih sebesar 67:34 dengan akurasi sebesar 73.2%, dan rata rata akurasi yang didapatkan dari keseluruhan pengujian yakni sebesar 68.01%. Pengujian juga dilakukan dengan mengganti kernel linear dengan kernel Polinomial dumana hasil dari pergantian kernel tersebut dapat mempengaruhi kinerja dari sistem, pada hal ini didapatkan akurasi tertinggi sebesar 62.5% dengan rata rata dari keseluruhan pengujian sebesar 46.4%. dengan penggunaan kernel linear memiliki peran penting dalam klasifikasi sentiment ini, pada pengujian terhadap kernel polinomial. kernel linear mendapatkan hasil akurasi lebih tinggi daripada kernel polinomial yang berpengaruh terhadap label kelas yang digunakan. penelitian ini juga mampu menvisualisasikan data lokasi ke dalam 3 bentuk yakni peta marker, peta cluster, dan HeatMap.*

***Kata kunci:*** *Klasifikasi, Sentimen, SVM, Sebaran lokasi, Peta Lokasi*

1. **Pendahuluan**

Pemindahan lokasi ibukota Republik Indonesia menjadi isu terhangat saat ini, Upaya pemindahan ibu kota Indonesia dimulai pada tahun 2019 pada masa kepresidenan Joko Widodo, dalam keputusannya untuk memindahkan ibu kota negara ke luar Pulau Jawa , dimana sebelumnya lokasi ibukota semula berada Provinsi DKI Jakarta ke Penajam Pasar Utara, Kutai Kartanegara Kalimantan Timur. Pemindahan ibu kota ini tertuang dalam Rencana Pembangunan Jangka Menengah Nasional 2020-2024 (Media, 2019).

Selain alasan umum yakni pertimbangan sosial ekonomi, pertimbagan ekonomi, serta pertimbangan politik. pemilihan lokasi baru ibukota yakni wilayah di Penajam Paser Utara dijadikan lokasi ibu kota baru adalah kecilnya risiko bencana alam di wilayah itu, dan juga melihat pengalaman beberapa negara di dunia yang sudah memindahkan ibukotanya (Hutasoit, 2018). Dengan adanya pemindahan lokasi ibukota ini berdampak dari segi perekonomian seperti pertumbuhan Produk Domestik Regional Bruto (PDRB) baik di daerah provinsi yang ditunjuk sebagai daerah ibukota baru dan juga nasional. (Bappenas, 2019)

Namun dari pernyataan dan keputusan pemerintah mengenai pemindahan lokasi ini terdapat pro dan kontra dikalangan masyarakat indonesia, terutama pengguna media sosial twitter. Twitter sendiri adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan kicauan (tweet) (Twitter, 2019). Pengguna twitter di Indonesia saat ini menempati peringkat 5 pengguna Twitter terbesar di dunia, dimana berdasarkan data PT Bakrie Telecom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global (Kominfo, 2019).

Dalam padanannya pengguna twitter di Indonesia selain menggunakannya sebagai sarana komunikasi dan berinteraksi antar pengguna juga sebagai sarana mengutarakan pendapatnya dan opini terhadap pemerintah. Opini yang diutarakan diperlukan Klasifikasi Sentimen, Klasifikasi Sentimen merupakan salah satu ini beragam mulai dari opini negatif dan opini positif, dimana opini tersebut tersebar di berbagai daerah. Untuk dapat mengetahui opini yang saat ini berkembang peercabangan dari Analisis Sentimen, dimana digunakan untuk mengolah berbagai macam opini yang telah diberikan oleh masyarakat atau para pakar melalui berbagai media yang ada dan membaginya ke sejumlah kelas, opini tersebut diberikan untuk sebuah produk, jasa maupun sebuah instansi. Pada Klasifikasi Sentimen terdapat 3 jenis opini, yaitu opini negatif, opini positif dan opini netral (Kontopoulos, et al., 2013). Terdapat beberapa algoritma yang dapat digunakan untuk metode sentimen ini, Salah satunya Support Vector Machine dimana metode ini dapat melakukan menganalisis dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya.

Dalam beberapa penelitian Mengenai klasifikasi sentiment sebelumnya (Go, et al., 2009). penelitian yang berjudul “Twitter Sentiment Classification using Distant Supervision” dalam penelitiannya menjelaskan mengenai analisis sentimen terhadap Twitter dengan berbagai metode seperti Naïve Bayes Classification, Maximum Entropy, ataupun Support Vector Machine. Didapatkan kesimpulan bahwa penggunaan metode SVM (Support Vector Machine) memberikan hasil dengan tingkat keakuratan hingga 82,2%.

Pada penelitian lainnya (Rofiqoh, et al., 2017) yang berjudul “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features”. Dalam penelitiannya, tentang opini masyarakat mengenai penyedia layanan telekomunikasi seluler. Hasil akurasi sistem yang diperoleh dari analisis sentimen dengan metode Support Vector Machine tanpa menggunakan Lexicon Based Features menghasilkan tingkat akurasi sebesar 84%.

Pada penelitian mengenai Sentimen Analisis yang berjudul “Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm” dalam penelitiannya, tentang opini yang ada di masyarakat maroko serta melakukan visualisasi data area asal tweet tersebut kedalam peta interaktif yang disebut “Folium” yang menghasilkan data secara grafik yang diklasifikasikan di peta Maroko dengan menggunakan koordinat yang diekstrak dari dari data tweet mereka.

Dalam penelitian ini, mencoba mencari sebaran lokasi pengguna twitter terhadap pemindahan Ibukota Republik Indonesia dari pengguna twitter yang memberikan opini, serta melakukan klasifikasi sentimen menggunakan metode SVM (Support Vector Machine) untuk mengklasifikasikan opini yang bersifat positif dan negatif, Sehingga menjadi informasi yang dapat membantu pemerintah untuk mengetahui opini masyarakat pengguna twitter serta persebaran lokasi pengguna tersebut. Selain itu penelitan ini dilakukan untuk mengetahui akurasi yang didapatkan nantinya efektif dengan menggunakan metode SVM (Support Vector Machine).

1. **Metodologi Penelitian**

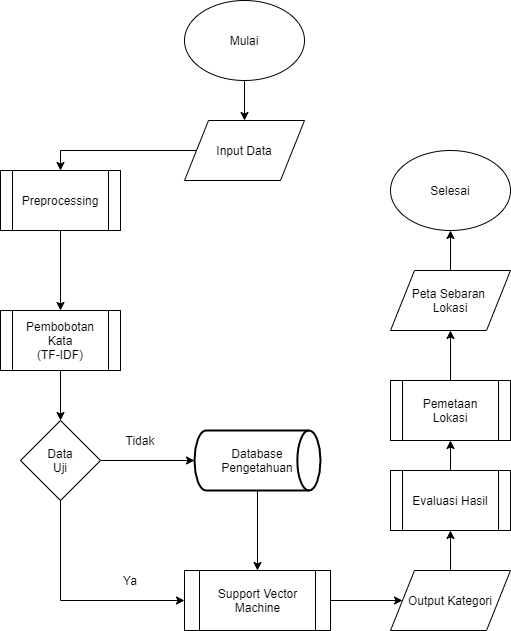
## 2.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan pengumpulan data sekunder. Data diperoleh berdasarkan pencarian atas term objek pada tweet berbahasa Indonesia. Pengambilan data dilakukan berdasarkan pencarian tweet dengan kata kunci “pemindahan ibukota”, “ibu kota pindah” atau “ibukota baru” dari tanggal 1 Januari 2019 sampai tanggal 31 Desember 2019 Hasil query berupa tweet mentah baik untuk data training maupun data testing akan diolah dengan metode preprocessing yang sama. Tweet yang didapatkan dikategorikan menjadi dua jenis yaitu positif untuk tweet yang memberikan respon setuju, masukan, dan dukungan dan negative untuk tweet yang memberikan respon tidak setuju dari isu tersebut. Data tweet yang akan dikumpulkan sebanyak 338 data tweet dengan proporsi sebanyak 43% dari data sebesar 143 tweet sebagai data bersentimen positif dan 57% sisanya, sebanyak 195 tweet sebagai data bersentimen negatif. Dari data tersebut akan dibagi menjadi data latih (data training) dan data uji (data testing) menggunakan metode K-Fold Cross Validation. Data yang digunakan hanyalah data tweet berbahasa Indonesia, dan juga data tersebut termuat nama pengguna, id pengguna, isi tweet, link url, jumlah like dan retweet

## 2.2 Alur Penelitian

Pada bagian ini akan digambarkan alur secara umum dari penelitian yang akan dilakukan penulis, yaitu dimulai dari pengumpulan data kotor dari media sosial, yang setelah itu akan dilakukan tahap preprocessing,

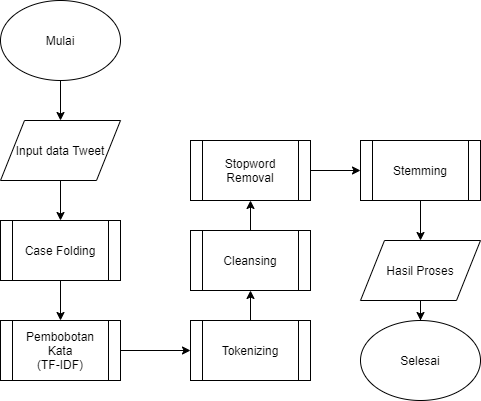
Selanjutnya akan dilakukan proses ekstraksi fitur dan lalu masuk ke tahap klasifikasi sentimen positif atau negatif, dan pada tahap akhir nantinya akan dilakukan evaluasi terhadap performa sistem dan metode yang digunakan serta visualisasi data kedalam bentuk peta



Gambar 2.1 Alur Umum Sistem

## 2.3 Pre Processing

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, Varma, & Govardhan, 2012). Data yang telah dikumpulkan ini akan masuk ke proses preprocessing yang terdiri dari Case Folding, Tokenization, Cleansing, Stopword Removal, dan Stemming yang dapat dilihat pada gambar 2.2.



Gambar 2. 2 Alir Proses Pre-Processing

Berikut akan dijelaskan masing-masing dari tahapan text preprocessing yang akan dilakukan

### 2.3.1 Case Folding

Case Folding merupakan proses mengkonversi keseluruhan teks menjadi bentuk yang seragam, dalam penelitian ini akan di konversi menjadi huruf kecil (*lower case*). Hasil dari *Case Folding* sebagai berikut :

Tabel 2.1 Case Folding

|  |  |
| --- | --- |
| Contoh Tweet | Hasil Case Folding |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah | visi yg disampaikan @tsamaradki cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @sherlyannavita . kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ilcperlukahibukotapindah |

### 2.3.2 Tokenization

Tahap ini adalah pemotongan string masukan berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen Hasil dari *Tokenization* adalah sebagai berikut :

Tabel 2.2 Tokenization

|  |  |  |  |
| --- | --- | --- | --- |
| Contoh Tweet | Hasil | | |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah | Visi  Yg  Disampaikan  @TsamaraDKI Cukup  Bisa  Dipahami  Tapi  Lebih  Setuju  Sama  Yg | Disampaikan  @SherlyAnnavita Kita  Ini  Bukan  Negara  Kaya  Banyak  Hutang  Buat  Apa  Pindah | Ibu  Kota  Hasil  Jual  Aset  &  Ngutang  Lagi  #ILCPerlukahIbuKotaPindah |

### 2.3.3 Cleansing

Tahapan Cleansing yakni proses yang dilakukan untuk membersihkan data teks berupa komponen khas di twitter seperti URL (Uniform Resource Locator), dan username. Hasil dari *Cleansing* adalah sebagai berikut :

Tabel 2.3 Cleansing

|  |  |  |  |
| --- | --- | --- | --- |
| Contoh Tweet | Hasil | | |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah | Visi  Yg  Disampaikan  Cukup  Bisa  Dipahami  Tapi  Lebih  Setuju  Sama  Yg | Disampaikan  Kita  Ini  Bukan  Negara  Kaya  Banyak  Hutang  Buat  Apa  Pindah | Ibu  Kota  Hasil  Jual  Aset  &  Ngutang  lagi |

### 2.3.4 Stopword Removal

Setelah melakukan *Cleansing* , masuk ke tahapan *Stopword Removal* untuk menghilangkan semua kata-kata yang merupakan *Stopword* pada Bahasa Indonesia , didapatkan hasil sebagai berikut :

Tabel 2.4 Stopword Removal

|  |  |  |  |
| --- | --- | --- | --- |
| Contoh Tweet | Hasil | | |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah | Visi  Disampaikan  Cukup  Bisa  Dipahami  Tapi  Lebih  Setuju  Sama | Disampaikan  Kita  Bukan  Negara  Kaya  Banyak  Hutang  Buat  Pindah | Ibu  Kota  Hasil  Jual  Aset  Ngutang |

### 2.3.5 Stemming

Melakukan Stemming mencari kata dasar dari kalimat yang ada, dengan hasil sebagai berikut :

Tabel 2.5 Stemming

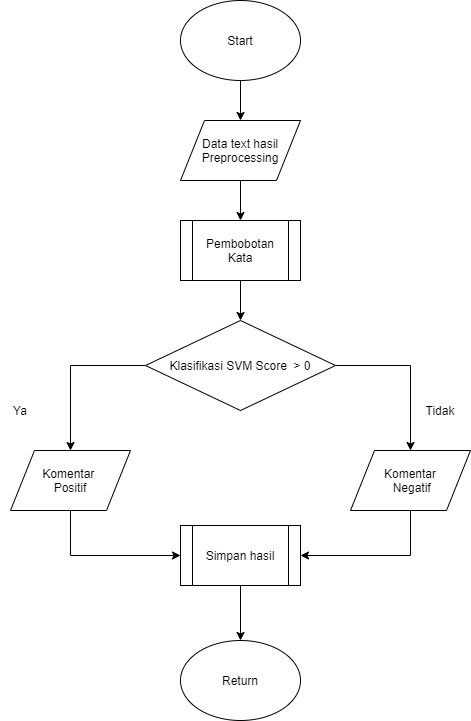
|  |  |  |  |
| --- | --- | --- | --- |
| Contoh Tweet | Hasil | | |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah | Visi  sampai  Cukup  Bisa  paham  Tapi  Lebih  Setuju  Sama | Sampai  Kita  Bukan  Negara  Kaya  Banyak  Hutang  Buat  Pindah | Ibu  Kota  Hasil  Jual  Aset  hutang |

## 2. 4 Pembobotan Kata (TF-IDF)

Dataset yang sudah di preprocessing sebelumnya kemudian diproses ­­kembali kedalam bentuk format bilangan biner sehingga dapat dikenal oleh sistem.

## 2.5 Proses SVM (Support Vector Machine)

data kata yang telah melalui proses preprocessing sebelumnya akan dihitung pembobotannya dengan metode TF-IDF, setelah pembobotan dilakukan proses klasifikasi apakah kata atau term tersebut bernilai = 0 maka kata tersebut dimasukkan dalam komentar negatif tetapi apabila kata tersebut = 1 maka kata tersebut dimasukkan dalam komentar positif.



Gambar 2.3 Alir Proses SVM

Proses analisis menggunakan SVM dimulai mengubah text menjadi data vektor. Vektor dalam penelitian ini memiliki dua komponen yaitu dimensi (word id) dan bobot. Bobot ini adalah nilai tf-idf, tujuan dari model ruang vektor digunakan untuk memberikan setiap kata dalam dokumen sebuah ID (dimensi) dan sebuah bobot berdasarkan seberapa penting keberadaannya dalam dokumen (posisi dokumen dalam dimensi tersebut).

## 2.6. Tahap Pengujian dan Evaluasi

Pengujian sistem berfungsi untuk mengetahui kinerja dari sistem itu sendiri dalam melakukan tugas yaitu mengklasifikasi sentimen. Dimana pada penelitian ini digunakan data opini yang diperoleh dari media sosial sebagai data testing yang diberikan label sesuai dengan jenis klasifikasi secara manual. Setelah itu data melewati tahapan preprocessing terlebih dahulu, setelah melewati tahap tersebut akan dilakukan pembobotan kata, lalu fitur tersebut akan disimpan dalam bentuk vektor dan melalui tahapan klasifikasi.

Scenario pengujian yang akan dilakukan pada penelitian ini adalah menghitung pengaruh parameter akurasi terhadap pembagian data latih dan data uji dengan pembagian sebagai berikut.

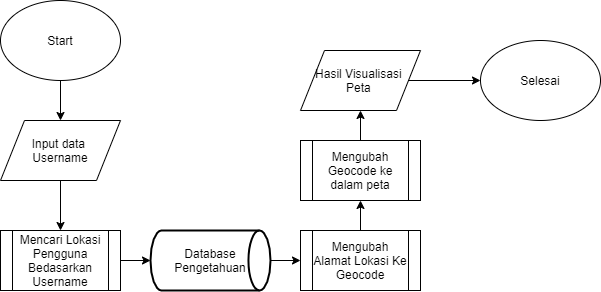
Tabel 2.6 Pengujian data

|  |  |  |
| --- | --- | --- |
| Rasio Data | Data Latih | Data Uji |
| 50:50 | 169 (50%) | 169 (50%) |
| 67:34 | 226 (67%) | 112 (34%) |
| 75:25 | 254 (75%) | 84 (25%) |
| 80:20 | 271 (80%) | 67 (20%) |
| 83:17 | 282 (83%) | 56 (17%) |
| 85:15 | 290 (85%) | 48 (15%) |
| 87:13 | 296 (87%) | 42 (13%) |
| 90:10 | 301 (90%) | 37 (10%) |
| 91:9 | 305 (91%) | 33 (9%) |

Pada setiap iterasi performa dihitung berdasarkan nilai akurasi, precision dan recall dengan menggunakan confusion matriks dengan rumus yang dijelaskan pada persamaan (3) untuk perhitungan akurasi, persamaan (4) untuk perhitungan precision, persamaan (5) untuk perhitungan recall dan persamaan (6) F-1 Score. Dalam penelitian ini juga melakukan pengujian terhadap pergantian kernel dari linear ke polynomial.

## 2.7 Proses Pemetaan Lokasi

Proses ini adalah pencarian dan pemetaan lokasi dari data tweet pengguna. Dimana dimulai dari pencarian lokasi pengguna bedasarkan nama pengguna (username), pengkonversian lokasi kedalam bentuk *geolocation*, lalu konversi geolocation kedalam peta sebaran lokasi.



Gambar 2.4 Alir Proses Pemetaan Lokasi

Setelah melakukan tahap pencarian lokasi, masuk ke tahap pemetaan lokasi menggunakan MapFolium dimana data *Geolocation* akan diolah menjadi data titik koordinat yang pada penelitian ini akan dibentuk menjadi 3 jenis peta yang terdiri dari peta titik koordinat, peta *FastMarkerCluster*, dan *HeatMap.*

1. **Result and Discussion**

## 3.1 Gambaran Umum Sistem

Sistem klasifikasi sentimen pada tweet twitter menggunakan metode *SVM (Support Vector Machine)* telah dibangun. Selain untuk menganalisis dan menguji algoritma yang digunakan untuk melakukan klasifikasi nilai sentimen pada data twitter tentang peristiwa pemindahan Ibukota Baru. Pada implementasinya sistem ini dapat digunakan oleh pihak akademisi ataupun pemerintah.

Pada penelitian ini penulis menggunakan data tweet dari Twitter yang dihimpun dengan cara *scraping* menggunakan *twitterscraper* yang nantinya data yang didapatkan berupa data *tabular* dengan ekstensi *CSV* (*Comma Separated Value*). Data tersebut berjumlah 339 data dari berbagai akun dengan topik yang membahas pemindahan ibukota Republik Indonesia, dan data tersebut sudah dilabeli sentimen positif dan negatif dengan bantuan pakar pada penelitian ini. dalam pelabelan variabel target akan mengubah data kategori tipe string dalam set data menjadi nilai numerik yang dapat dipahami model.

Tabel 3.1 Pelabelan Kategori Sentimen

|  |  |
| --- | --- |
| Kategori Sentimen | Label yang diubah |
| Negatif | 0 |
| Positif | 1 |

Pengujian algoritma pada penelitian ini untuk mengetahui seberapa baik algoritma dapat mengklasifikasikan sentimen tweet dari pengguna. Dengan menggunakan *SVM (Support Vector Machine)*.

## 3.2 Implementasi Sistem

Sistem diimplementasikan menggunakan bahasa pemrogramman Python dengan versi 3.7 serta menggunakan IDE Spyder 3. Penelitian ini juga menggunakan software Microsoft Excel 2016 dalam menyimpan dataset. Sistem ini dijalankan dengan menggunakan Sistem Operasi Windows 10 64-bit yang mempunyai spesifikasi Intel(R) Core(TM) i7-7700HQ CPU @ 2.6 GHz dan memory RAM sebesar 8,00 GB serta menggunakan koneksi internet dengan kecepatan 14 Mbps, yang akan dijelaskan pada tabel dibawah

Dalam pengambilan data yang digunakan package Twitterscraper yang merupakan salah satu package Python. agar dapat melakukan scraping data dari Twitter, harus ditentukan terlebih dahulu query untuk Twitterscraper.

Query untuk mencari data tweets yang mengandung kata “pemindahan ibukota”, “ibu kota pindah” atau “ibukota baru” dari tanggal 1 Januari 2019 sampai tanggal 31 Desember 2019 :

Tabel 3.2 Penggalan Kode Query Twitterscraper

twitterscraper 'ibukota baru or pemindahan ibu kota' since : 2019-01-01 until : 2019-12-31 -o "ibukota\_tweets.csv"

**3.3 Implementasi Preprocessing**

Sebelum masuk ke tahap proses klasifikasi sentimen pada tweet, tahap yang harus dilalui yaitu proses text preprocessing. Tahap text preprocessing yaitu tahap dimana perbaikan teks tweet dari teks yang sulit dikenali oleh sistem menjadi teks yang mudah untuk diproses klasifikasi sentimennya. Tahap ini terdapat 5 tahap preprocessing yakni tahap Case Folding, Cleansing, Tokenizing, Stopword Removing dan Stemming.

Tahap Awal yang dilakukan adalah Case Folding merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z”. tahap selanjutnya yaitu tokenization yaitu tahapan dimana tweet dipisahkan menjadi token-token berdasarkan katanya. Setelah itu masuk ketahap Cleansing yakni proses yang dilakukan untuk membersihkan data teks berupa komponen khas di twitter seperti URL (Uniform Resource Locator), dan username. Tahapan selanjutnya yaitu tahapan Stopword Removing untuk menghilangkan semua kata-kata yang merupakan Stopword pada Bahasa Indonesia, Setelah melakukan Stopword Removing tahap selanjutnya yakni Stemming yakni mengurai berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya, dimana untuk setiap kata yang memiliki imbuhan akan diubah menjadi kata dasar

**3.4. Implementasi Pembobotan Kata (TF-IDF)**

Pada tahap ini akan melakukan ekstraksi fitur, fitur yang diekstraksi pada proses ini adalah Term Frequency-Inverse Document Frequency (TF-IDF).

**3.5. Implementasi Klasifikasi dengan Support Vector Machine**

Proses selanjutnya setelah mendapatkan semua fitur yang dibutuhkan adalah proses klasifikasi menggunakan algoritma SVM.

## 4.6 Hasil Pengujian

Pengujian pertama dilakukan terhadap Algoritma SVM dengan menggunakan K-Fold Cross Validation menghasilkan 10 buah kombinasi data latih dan data uji untuk masing masing kelas sentimen. Masing-masing kombinasi data latih dan data uji dimasukkan dalam proses SVM.

Pengujian yang dilakukan meliputi pengujian terhadap rasio data latih, dan pengaruh terhadap pergantian kernel dari linear ke polynomial.

## 3.7 Pengaruh Parameter Akurasi

### 3.7.1 Pengaruh Terhadap Data Latih dan data Uji

Eksperimen yang dilakukan menggunakan algoritma SVM dengan tujuan membandingkan rasio data latih dan data uji yang telah ditentukan pada tabel 2.2, proses eksperimen ini membandingkan hasil dari Akurasi, Recall, Presisi, F1-Score

Gambar 3.1Grafik pengaruh terhadap data latih dan data uji

Dari hasil tersebut dapat dilihat bahwa akurasi tertinggi yang didapat sebesar 73.2 % dengan rasio 67:34 dengan nilai rata rata akurasi sebesar 68.01 %. Untuk rasio 67:34 akurasi yang didapatkan sebesar 73.2%, presisi negatif sebesar 81%, presisi positif sebesar 62%, nilai recall dengan masing masing 74% untuk recall negatif dan 71% untuk recall positif, nilai F1-Score yang didapat memiliki nilai sebesar 78% untuk F1-Score Negatif dan 67% F1-Score Positif.

### 4.7.2 Pengaruh Terhadap Kernel Polynomial

Penelitian yang dilakukan selanjutnya adalah menggunakan algoritma SVM dengan tujuan membandingkan pengaruh hasil dari pergantian kernel dari kernel linear ke polinomial proses eksperimen ini membandingkan hasil dari Akurasi, Recall, Presisi, F1-Score. Pada gambar dibawah merupakan hasil dari kernel polynomial. Dimana Polynomial mendapatkan nilai akurasi tertinggi yang didapat sebesar 62.5% pada rasio data latih 67:34 dengan nilai rata-rata akurasi sebesar 46.48 %. Untuk rasio 67:34 dengan rincian hasil yaitu akurasi yang didapatkan sebesar 62.5%, presisi negatif sebesar 62%, presisi positif sebesar 0%, nilai recall dengan masing masing 100% untuk recall negatif dan 0% untuk recall positif, nilai F1-Score yang didapat memiliki nilai sebesar 77% untuk F1-Score Negatif dan 0% F1-Score Positif.

Gambar 2. 2 Grafik hasil akurasi dari Kernel Polynomial

Setelah mendapatkan hasil dari kernel polynomial, maka pada tahap ini akan membandingkan hasil dari eksperimen pada subbab 3.7.1 dengan hasil dari Akurasi, Recall, Presisi, F1-Score yang dimiliki polinomial

Gambar 3.3 Grafik hasil akurasi dari kernel Polynomial

Gambar 3.4 Grafik hasil Perbandingan Presisi dari Kernel

Gambar 3. 5 Grafik hasil perbandingan Presisi dari Kernel

Gambar 3. 6 Grafik hasil Perbandingan F-1 Score dari Kernel

Dari hasil tersebut dapat dilihat bahwa kernel linear lebih baik dari kernel polynomial. Kernel linear lebih baik dikarenakan kelas label yang digunakan pada kasus ini hanya sebanyak 2 label yakni label negatif dan label positif yang telah di ubah nilainya menjadi 0 untuk label negatif dan 1 untuk label positif. Sedang kernel polynomial lebih baik ketika label kelasnya lebih dari 2.

**3.8 Implementasi Pemetaan Sebaran Lokasi Pengguna Twitter**

**3.8.1 Implementasi Pencarian Lokasi Bedasarkan Username**

Pada bagian ini sistem akan melakukan pencarian lokasi menggunakan data username yang didapatkan pada hasil scraping menggunakan twitterscraper. Dalam proses pencarian lokasi ini dibutuhkan koneksi internet.

**3.8.2 Implementasi Konversi Lokasi Menjadi Geocode**

Setelah melakukan pencarian lokasi, sistem akan melakukan konversi data lokasi yang didapatkan sebelumnya menjadi Geocode berupa latitude dan longitude. Dalam proses ini juga dibutuhkan library yaitu geopy yang berfungsi sebagai locator (mencari lokasi) bedasarkan data csv yang akan diproses serta memerlukan koneksi internet dalam melakukannya.

**3.8.3 Visualisasi Peta Sebaran Lokasi Menggunakan MapFolium**

Pada tahapan ini adalah membuat pemetaan marker menggunakan MapFolium, yang nantinya akan menghasilkan titik lokasi pengguna twitter dari sentimen yang ada. Dimana pada tahap ini akan memunculkan tag bewarna “Merah” sebagai representasi dari sentimen negatif, dan tag bewarna “Hijau” sebagai representasi dari sentimen positif

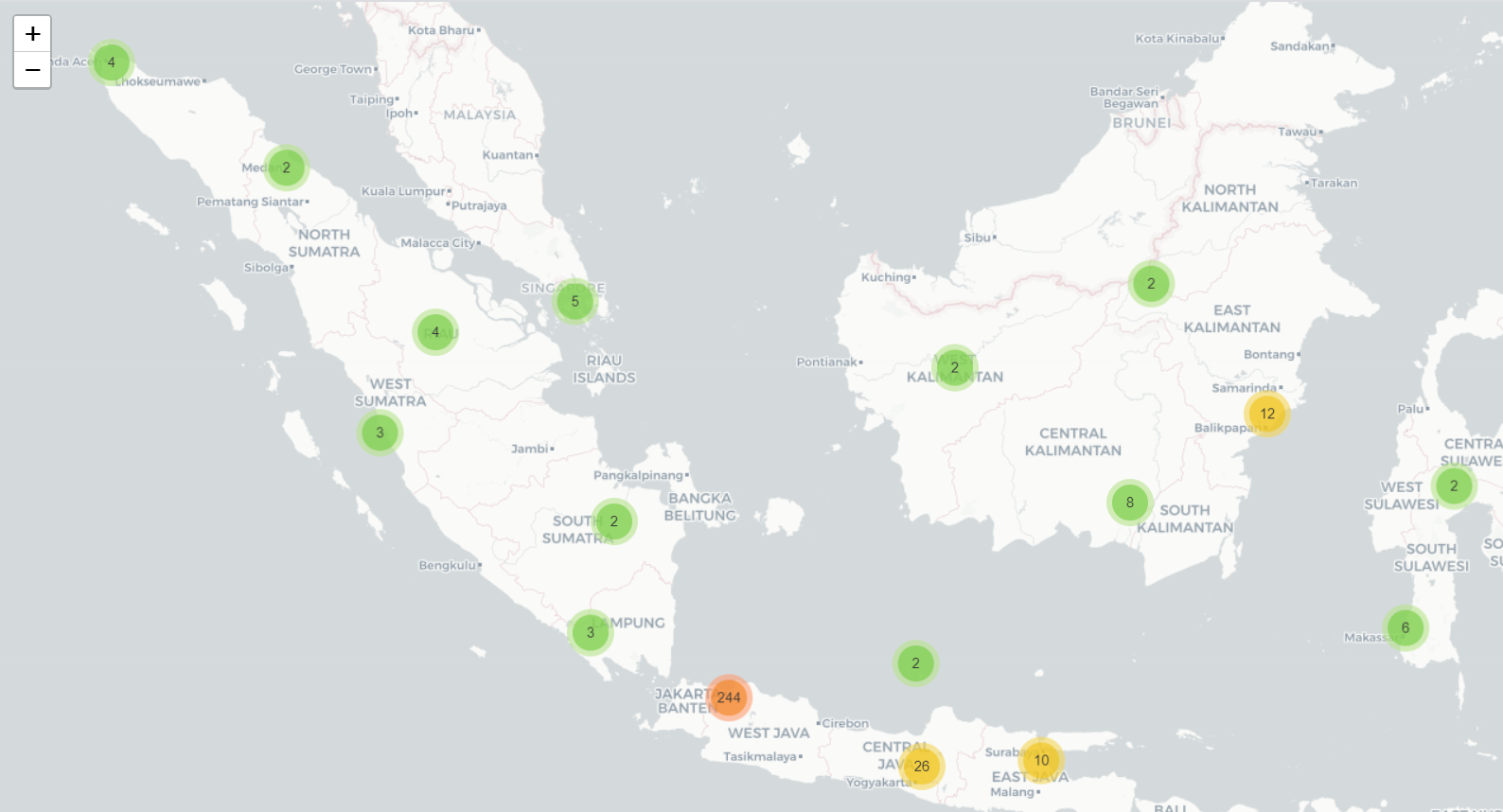


Gambar 3. 7 Hasil Visualiasis Peta Marker

Pada bagian ini hasil yang diberikan dari visualisasi peta marker ini sebagian besar datanya tumpang tindih, dikarenakan beberapa titik lokasi yang didapat dari sistem memiliki kordinat bujur dan lintang yang sama, maka hasil visualisasi yang didapat hanya terlihat sebagian.

### 3.8.4 Klusterisasi Lokasi Pengguna Menggunakan FastMarkerCluster

Selanjutnya pada tahap ini adalah membuat kelompok cluster lokasi menggunakan MapFolium, yang nantinya akan menghasilkan titik lokasi pengguna twitter dari sentimen yang telah dikelompokkan bedasarkan kecocokan data kordinat yang ada. Dimana pada tahap ini akan memunculkan tag yang berisi angka banyaknya sentimen atau data yang memiliki kesamaan lokasi.



Gambar 4.8 Hasil Viusalisasi Peta FastMarkerCluster

Pada gambar diatas terlihat angka yang merepresentasikan jumlah data dari kelompok lokasi yang telah di klusterisasi. Pada penelitian ini lokasi cluster dengan jumlah tweet terbanyak berada di Wilayah jabodetabek dengan jumlah tweet sebanyak 225, dimana tweet dengan sentimen negatif sebesar 131 tweet, dan 89 tweet bersentimen positif.

Tabel 3.2 Hasil Sebaran Lokasi Bedasarkan Cluster Wilayah

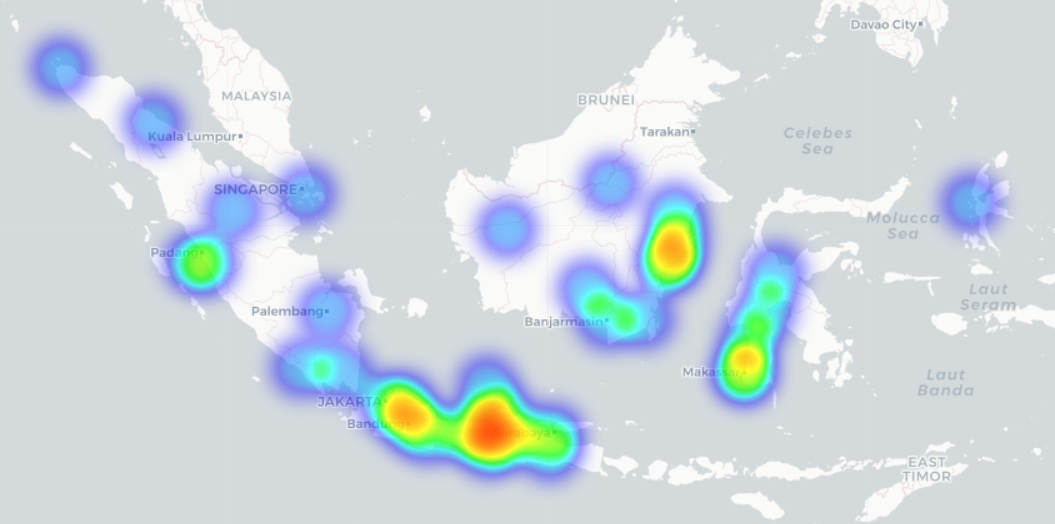
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lokasi | Total data | Positif | Negatif | %Positif | %Negatif |
| Aceh | 4 | 0 | 4 | 0 | 100 |
| Jabodetabek | 220 | 89 | 131 | 40.45455 | 59.54545 |
| Jawa Barat | 23 | 8 | 15 | 34.78261 | 65.21739 |
| Jawa Tengah | 14 | 5 | 9 | 35.71429 | 64.28571 |
| Jawa Timur | 10 | 5 | 5 | 50 | 50 |
| Kalimantan Barat | 2 | 0 | 2 | 0 | 100 |
| Kalimantan Selatan | 3 | 2 | 1 | 66.66667 | 33.33333 |
| Kalimantan Tengah | 5 | 5 | 0 | 100 | 0 |
| Kalimantan Timur | 13 | 8 | 5 | 61.53846 | 38.46154 |
| Kalimantan Utara | 1 | 0 | 1 | 0 | 100 |
| Kepulauan Riau | 5 | 4 | 1 | 80 | 20 |
| Lampung | 3 | 2 | 1 | 66.66667 | 33.33333 |
| Maluku Utara | 1 | 1 | 0 | 100 | 0 |
| Riau | 4 | 3 | 1 | 75 | 25 |
| Sulawesi Selatan | 7 | 2 | 5 | 28.57143 | 71.42857 |
| Sulawesi Tengah | 1 | 1 | 0 | 100 | 0 |
| Sumatera Barat | 3 | 1 | 2 | 33.33333 | 66.66667 |
| Sumatera Selatan | 2 | 1 | 1 | 50 | 50 |
| Sumatera Utara | 2 | 1 | 1 | 50 | 50 |
| Yogyakarta | 14 | 5 | 9 | 35.71429 | 64.28571 |

Pada tabel 3.2, dijelaskan bahwa pengguna twitter dengan jumlah sentimen negatif terbanyak yakni berada di cluster wilayah Jabodetabek dengan jumlah sentimen negatif sebanyak 131 tweet. untuk jumlah pengguna twitter dengan sentimen positif terbanyak berada di daerah Jabodetabek dengan jumlah 89 tweet. untuk daerah dengan presentase sentimen negatif terbesar berada di daerah Aceh, Kalimantan Barat, dan Kalimantan Utara dengan presentase sentimen sebesar 100%, lalu untuk daerah dengan presentase sentimen positif terbesar berada di daerah Kalimantan Tengah, Sulawesi Tengah, dan Maluku Utara dengan presentase sentimen sebesar 100%.

**3.8.5 Visualisasi Aktivitas Tweet Dengan HeatMap**

Analisis HeatMap adalah representasi grafis dari data yang memvisualisasikan kepadatan titik dalam suatu lapisan. Dimungkinkan untuk melakukan analisis Heatmap pada layer titik yang berada di Sistem File atau di Database.

Pada tahap ini bertujuan menvisualisasikan kepadatan aktivitas pengguna twitter yang membuat opini mengenai sentimen. Dimana dalam peta tersebut akan terdapat warna yang merepresentasikan banyaknya pengguna. Warna “biru” merepresentasikan titik aktivitas pengguna twitter yang membicarakan topik tersebut rendah, lalu warna “merah menyala” yang merepresentasikan merepresentasikan titik aktivitas pengguna twitter yang membicarakan topik tersebut sangat tinggi.



Gambar 3. 9 Hasil Visualisasi Peta Heatmap

Pada gambar diatas terlihat warna yang merepresentasikan kepadatan dari aktivitas pengguna twitter yang membuat opini dalam penelitian ini, dimana semakin terang warna tersebut menggambarkan titik paling aktivitas terpadat yang pada penelitian ini aktivitas terbanyak terdapat di pulau jawa.

1. **Kesimpulan**

Kesimpulan yang dapat diambil dari penelitian yang telah dilakukan adalah sebagai berikut :

* + - 1. Dengan melakukan pengujian dengan kernel linear terhadap data uji dan data latih diketahui bahwa jumlah data uji dan data latih mempengaruhi kinerja dari sistem yang dibangun. Pada hal ini didapatkan hasil dengan menggunakan rasio data uji dan data latih sebesar 67:34 didapatkan akurasi sebesar 73.2%, dan rata rata akurasi yang didapatkan dari keseluruhan pengujian yakni sebesar 68.01%
      2. Pengujian juga dilakukan dengan mengganti kernel linear dengan kernel Polinomial juga dapat mempengaruhi kinerja dari sistem, pada hal ini didapatkan akurasi tertinggi sebesar 62.5% dengan rata rata dari keseluruhan pengujian sebesar 46.4%
      3. Dengan label yang digunakan dalam sentimen ini sebanyak 2 label, maka penggunaan kernel Linear yang memiliki pengaruh lebih terhadap akurasi yang didapat. Karena pengunaan dari kernel polinomial bersifat nonlinear.
      4. Dalam penelitian ini juga mampu menvisualisasikan data lokasi ke dalam 3 bentuk yakni peta marker, peta cluster dam HeatMap. Didapatkan hasil bahwa wilayah cluster dengan tweet terbanyak berada di Wilayah jabodetabek dengan jumlah tweet sebanyak 225, berisi tweet dengan sentimen negatif sebesar 131 tweet, dan 89 tweet bersentimen positif. Untuk daerah dengan presentase sentimen negatif terbesar berada di daerah Aceh, Kalimantan Barat, dan Kalimantan Utara dengan presentase sentimen sebesar 100%, dan presentase sentimen positif terbesar berada di daerah Kalimantan Tengah, Sulawesi Tengah, dan Maluku Utara dengan presentase sentimen sebesar 100%.

1. **Saran**

Untuk pengembangan lebih lanjut adapun saran yang dapat disampaikan antara lain:

1. Klasifikasi sentiment yang dibuat pada penelitian ini sangat bergantung pada pengetahuan yang diekstraksi dari data. Penelitian selanjutnya dapat menambahkan data lebih banyak lagi agar jenis kata kata lebih bervariasi lagi.
2. Dalam pencarian data yang dilakukan bersifat scraping data yang dimana hanya dapat mengambil data lawas, dalam penelitian selanjutnya diharapkan dapat menggunakan cara streaming data, dimana data yang didapat bisa lebih update
3. Untuk penelitian selanjutnya dapat menggunakan algoritma lain untuk menghasilkan nilai akurasi yang lebih tinggi

# **Reference**

[1] ABDOULI, A. E., HASSOUNI, L., & ANOUN, H. (2017). Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm . *International Journal of Computer Science and Information Security (IJCSIS), Vol. 15*, 191-200.

[2] Bappenas. (2019, September 24). *https://www.bappenas.go.id/*. Dipetik September 15, 2019, dari bappenas.go.id: https://www.bappenas.go.id/files/diskusi-ikn-2/Analisis%20Hal%20Keamanan%20Pemindahan%20Ibu%20Kota\_kastaf-v3.pdf

[3] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

[4] Folium. (2013). *Folium Documentation*. Dipetik February 8, 2020, dari https://python-visualization.github.io/folium/

[5] Go, A., Bhayani , R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *CS224N Project Report, Stanford*, 2009.

[6] Hemalatha, I., Varma, G., & Govardhan, A. (2012). Preprocessing the informal text for efficient sentiment analysis. *International Journal of Emerging Trends \& Technology in Computer Science (IJETTCS), 1*(2), 58--61.

[7] Hutasoit, W. (2018). Analisa Pemindahan Ibukota Negara. *Dedikasi*, 109-128.

[8] Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data transformation pada data mining. *Prosiding Konferensi nasional “Inovasi dalam Desain dan Teknologi”. IDeaTech*.

[9] Kominfo. (2019, September 5). *https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\_satker*. Diambil kembali dari Kementrian Komunikasi dan Informasi: https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\_satker

[10] Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 4065–4074.

[11] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1--167.

[12] Media, K. (2019). *kompas.com*. Diambil kembali dari Kompas Media: https://money.kompas.com/read/2019/05/09/184859926/kepala-bappenas-pemindahan-ibu-kota-masuk-rpjmn-2020-2024

[13] Nomleni, P. (2015). SENTIMENT ANALYSIS MENGGUNAKAN SUPPORT VECTOR MACHINE(SVM) . SURABAYA: PROGRAM PASCA SARJANA BIDANG KEAHLIAN TELEMATIKA (CIO) JURUSAN TEKNIK ELEKTRO FAKULTAS TEKNOLOGI INDUSTRI INSTITUT TEKNOLOGI SEPULUH NOPEMBER .

[14] Pelias. (2017). *Pelias.io*. Dipetik February 8, 2020, dari https://pelias.io/

[15] Pressman, R. S. (2012). Pendekatan Praktisi Rekayasa Perangkat Lunak. *Yogyakarta: Andi Offset*.

[16] Riany , J., Fajar, M., & Lukman, M. (2016). Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor. *SISFO Vol 6 No 1, 6*(1).

[17] Rofiqoh, U., Perdana, R., & Fauzi , M. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *urnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 964X.

[18] Taspinar, A. a. (2017). Twitterscraper 0.2. 7: Python Package Index.

[19] Twitter. (2019, September 5). *https://about.twitter.com/id/company.html*. Diambil kembali dari Twitter - About: https://about.twitter.com/id/company.html

[20] wibowo, a. (2017). *Binus University Graduate Program*. Dipetik February 8, 2020, dari https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/

[21] Wirawan, I., & Eksistyanto, I. (2015). Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel. *JUTI: Jurnal Ilmiah Teknologi Informasi, 13*(2), 182--189.