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

**KOMPETENSI PEMROSESAN DATA BESAR DAN MANAJEMEN BISNIS**

**SKRIPSI**



**FATHIYARIZQ MAHENDRA PUTRA**

**NIM. 1608561008**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS UDAYANA**

**BUKIT JIMBARAN**

**2020**

# SURAT PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini menyatakan bahwa naskah Skripsi dengan judul:

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

Nama : Fathiyarizq Mahendra Putra

NIM : 1608561008

Program Studi : Teknik Informatika

E-mail : fathiyarizq.mahendra@gmail.com

Nomor telp/HP : 08873906394

Alamat : Jl Jayagiri XXI D No 1 Denpasar Timur, Kota Denpasar

Belum pernah dipublikasikan dalam dokumen skripsi, jurnal nasional maupun internasional atau dalam prosiding manapun, dan tidak sedang atau akan diajukan untuk publikasi di jurnal atau prosiding manapun. Apabila di kemudian hari terbukti terdapat pelanggaran kaidah – kaidah akademik pada karya ilmiah saya, maka saya bersedia menanggung sanksi-sanksi yang dijatuhkan karena kesalahan tersebut, sebagaimana diatur oleh Peraturan Menteri Pendidikan Nasional Nomor 17 Tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Demikian Surat Pernyataan ini saya buat dengan sesungguhnya untuk dapat dipergunakan bilamana diperlukan.

Bukit Jimbaran, 2020

Yang membuat pernyataan,

(Fathiyarizq Mahendra Putra)

NIM. 1608561008

# LEMBAR PENGESAHAN TUGAS AKHIR

Judul : Sebaran Lokasi Pengguna Twitter Berdasarkan Sentimen Terhadap Opini Pemindahan Ibukota Republik Indonesia

Kompetensi : Pemrosesan Data Besar dan Manajemen Bisnis

Nama : Fathiyarizq Mahendra Putra

NIM : 1608561008

Tanggal Seminar : 17 April 2020

Disetujui oleh :

|  |  |
| --- | --- |
| Pembimbing I | Penguji I |
|  |  |
| Drs. I Wayan Santiyasa, M.Si | Dr. Ir. I Ketut Gede Suhartana., S.Kom., M.Kom |
| NIP. 196704141992031002 | NIP. 197201102008121001 |

|  |  |
| --- | --- |
| Pembimbing II | Penguji II |
|  |  |
| Ida Bagus Gede Dwidasmara, S.Kom., M.Cs. | I Komang Ari Mogi, S.Kom., M.Kom. |
| NIP. 198503152010121007 | NIP. 198409242008011007 |

|  |  |
| --- | --- |
|  | Penguji III |
|  |  |
|  | Cokorda Rai Adi Pramartha,ST.MM.PhD |
|  | NIP. 197806212006041002 |

|  |
| --- |
| Mengetahui, |
| Koordinator Program Studi Teknik Informatika |
| FMIPA UNUD |
|  |
| Dr. Ir. I Ketut Gede Suhartana., S.Kom., M.Kom |
| NIP. 197201102008121001 |

# KATA PENGANTAR

Penelitian dengan judul Penerapan Metode Sebaran Lokasi Pengguna Twitter Berdasarkan Sentimen Terhadap Opini Pemindahan Ibukota Republik Indonesia ini disusun dalam rangkaian kegiatan pelaksanaan Tugas Akhir di Program Studi Teknik Informatika FMIPA UNUD. Proposal ini disusun dengan harapan dapat menjadi pedoman dan arahan dalam melaksanakan penelitian di atas.

Sehubungan dengan telah diselesaikannya penelitian ini, maka diucapkan terimakasih dan penghargaan kepada berbagai pihak yang telah membantu pengusul, antara lain :

1. Bapak Drs. I Wayan Santiyasa,M.Si. selaku pembimbing yang telah membantu memberikan materi dan arahan serta pencerahan kepada penulis.
2. Bapak Ida Bagus Gede Dwidasmara, S.Kom., M.Cs.. selaku pembimbing yang telah bersedia mengkritisi, memeriksa dan menyempurnakan penelitian ini.
3. Bapak dan Ibu dosen pengajar di Jurusan Ilmu Komputer Fakultas MIPA Universitas Udayana yang telah meluangkan waktu untuk memberikan saran dan masukan dalam menyempurnakan penelitian ini.
4. Rekan-rekan di Program Studi Teknik Informatika yang telah memberikan dukungan moral dalam penyelesaian penelitian tugas akhir ini.
5. Keluarga, kerabat serta semua pihak yang turut serta dalam memberikan dukungan semangat dan motivasi sehingga tugas akhir ini dapat terselesaikan.

Disadari pula bahwa sudah tentu tugas akhir ini masih memiliki kelemahan dan kekurangan. Memperhatikan hal ini, maka masukan dan saran-saran untuk penyempurnaan penelitian tugas akhir ini sangat diharapkan.

Bukit Jimbaran, Januari 2020

Penulis

Judul : Sebaran Lokasi Pengguna Twitter Berdasarkan Sentimen Terhadap Opini Pemindahan Ibukota Republik Indonesia

Nama : Fathiyarizq Mahendra Putra

NIM : 1608561008

Pembimbing : 1. Drs. I Wayan Santiyasa,M.Si.

2. Ida Bagus Gede Dwidasmara, S.Kom., M.Cs

# 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 85:15 didapatkan hasil akurasi sebesar 85,29%, dengan rata-rata akurasi yang didapatkan keseluruhan sebesar 78,51%. Pengujian juga dilakukan dengan mengganti kernel linear dengan kernel Polynomial dimana hasil dari pergantian kernel tersebut dapat mempengaruhi kinerja dari sistem, pada hal ini didapatkan akurasi tertinggi sebesar 45,8% dengan rata rata dari keseluruhan pengujian sebesar 40,6%. 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. Dalam penelitian ini data username dari pengguna twitter dapat di konversi menjadi data lokasi sesuai lokasi pengguna atau lokasi dimana tweet itu dibuat, dimana menghasilkan data dalam bentuk alamat, dan kode lokasi (Geolocation) juga mampu menvisualisasikan data lokasi ke dalam 3 bentuk yakni peta marker, peta cluster, dan HeatMap.

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

Title : Distribution of Twitter User Locations Based on Sentiments Against Opinions of the Capital of the Republic of Indonesia

Name : Fathiyarizq Mahendra Putra

NIM : 1608561008

Supervisor : 1. Drs. I Wayan Santiyasa,M.Si.

Co-Supervisor : 2. Ida Bagus Gede Dwidasmara, S.Kom., M.Cs

# ABSTRACT

The relocation of the capital of the Republic of Indonesia is the hottest issue at the moment, efforts to relocate the capital of Indonesia began in 2019 during the presidency of Joko Widodo. In addition to general reasons, namely socioeconomic considerations, and political considerations. Determination of new locations in Penajam Paser Utara. Related to the government's decision to relocate this location, the pros and cons of the people of Indonesia, the majority of Twitter social media users. In this study, trying to find the distribution of the location of Twitter users to the relocation of the Capital of the Republic of Indonesia from Twitter users who provide opinions, as well as to classify sentiments using the SVM (Support Vector Machine) method, what is done in this study is to find data, preprocessing, sentiment list, testing and evaluation, location search, and visualization to map form. From the results of testing with a linear kernel, the results obtained from test data and training data can improve the performance of the system being built. In this case the results obtained by using a ratio of test data and training data of 85:15 obtained an accuracy of 85.29%, with an average accuracy obtained overall of 78.51%. Testing is also done by replacing the linear kernel with a Polynomial kernel where the results of changing the kernel can affect the performance of the system, in this case the highest accuracy of 45.8% with an average of 40.6% of the overall test. the use of linear kernels has an important role in the classification of these sentiments, in testing the polynomial kernels. a higher yield is obtained from the polynomial kernel which is approved for the class label used. the use of linear kernels has an important role in the classification of these sentiments, in testing the polynomial kernel. linear kernels get higher accuracy results than polynomial kernels that affect the class label used. In this study, also processing username data from Twitter users who can be converted into location data according to the user's location or the location where the tweet was made, which produces data in the form of an address, and the location code (Geolocation) is also able to visualize location data into 3 forms namely marker maps, cluster maps, and HeatMap.

*Keywords: Sentiments, SVM, Location distribution, Location Map, Classification.*

# DAFTAR ISI

[SURAT PERNYATAAN KEASLIAN KARYA ILMIAH ii](#_Toc40698424)

[LEMBAR PENGESAHAN TUGAS AKHIR iii](#_Toc40698425)

[KATA PENGANTAR iv](#_Toc40698426)

[ABSTRAK v](#_Toc40698427)

[ABSTRACT vi](#_Toc40698428)

[DAFTAR ISI vii](#_Toc40698429)

[DAFTAR TABEL xi](#_Toc40698430)

[DAFTAR GAMBAR xii](#_Toc40698431)

[DAFTAR LAMPIRAN xiii](#_Toc40698432)

[BAB I xiv](#_Toc40698433)

[PENDAHULUAN xiv](#_Toc40698434)

[1.1 Latar Belakang xiv](#_Toc40698435)

[1.2 Rumusan Masalah 3](#_Toc40698436)

[1.3 Tujuan Penelitian 3](#_Toc40698437)

[1.4 Batasan Masalah 4](#_Toc40698438)

[1.5 Manfaat Penelitian 4](#_Toc40698439)

[1.6 Metode Penelitian 5](#_Toc40698440)

[1.6.1 Desain Penelitian 5](#_Toc40698441)

[1.6.2 Data Penelitian 6](#_Toc40698442)

[1.6.3. Pengelolaan Data Awal 6](#_Toc40698443)

[1.6.4. Metode yang Digunakan 6](#_Toc40698444)

[BAB II 7](#_Toc40698445)

[TINJAUAN PUSTAKA 7](#_Toc40698446)

[2.1. Tinjauan Empiris 7](#_Toc40698447)

[2.2 Tinjauan Teoritis 17](#_Toc40698448)

[2.2.1 Klasifikasi Sentimen 17](#_Toc40698449)

[2.2.2 *Twitterscraper* 17](#_Toc40698450)

[2.2.3 *Preprocessing* 17](#_Toc40698451)

[*2.2.4* *Term Frequence – Inverse Document Frequency (TF-IDF)* 18](#_Toc40698452)

[2.2.5 *Confusion Matrix* 19](#_Toc40698453)

[2.2.6 *K-Fold Cross Validation* 20](#_Toc40698454)

[2.2.7 *Support Vector Machine* 21](#_Toc40698455)

[2.2.8 *Geocoding* 22](#_Toc40698456)

[2.2.9 *MapFolium* 22](#_Toc40698457)

[2.2.10 Metode *Prototype* 23](#_Toc40698458)

[BAB III 25](#_Toc40698459)

[ANALISIS DAN PERANCANGAN 25](#_Toc40698460)

[3.1. Pengumpulan Data 25](#_Toc40698461)

[3.2 Alur Penelitian 25](#_Toc40698462)

[3.3 Pre Processing 26](#_Toc40698463)

[3.3.1 Case Folding 27](#_Toc40698464)

[3.3.2 Tokenization 28](#_Toc40698465)

[3.3.3 Cleansing 28](#_Toc40698466)

[3.3.4 Stopword Removal 29](#_Toc40698467)

[3.3.5 Stemming 30](#_Toc40698468)

[3. 4 Pembobotan Kata (TF-IDF) 30](#_Toc40698469)

[3.5 Proses SVM (Support Vector Machine) 30](#_Toc40698470)

[3.6. Tahap Pengujian dan Evaluasi 31](#_Toc40698471)

[3.7 Proses Pemetaan Lokasi 32](#_Toc40698472)

[3.8 Perancangan Desain Antarmuka 33](#_Toc40698473)

[3.8.1 Perancangan Struktur Menu 34](#_Toc40698474)

[3.8.2 Perancangan Tampilan 35](#_Toc40698475)

[1. Tampilan Halaman utama Web Penelitian 35](#_Toc40698476)

[2. Tampilan Summary 36](#_Toc40698477)

[3. Tampilan Background 37](#_Toc40698478)

[4. Tampilan Method 37](#_Toc40698479)

[BAB IV 39](#_Toc40698480)

[HASIL DAN PEMBAHASAN 39](#_Toc40698481)

[4.1 Gambaran Umum Sistem 39](#_Toc40698482)

[4.2 Implementasi Sistem 40](#_Toc40698483)

[4.3 Implementasi Preprocessing 41](#_Toc40698484)

[4.4. Implementasi Pembobotan Kata (TF-IDF) 43](#_Toc40698485)

[4.5. Implementasi Klasifikasi dengan Support Vector Machine 44](#_Toc40698486)

[4.6 Hasil Pengujian 45](#_Toc40698487)

[4.7 Pengaruh Parameter Akurasi 47](#_Toc40698488)

[4.7.1 Pengaruh Terhadap Data Latih dan data Uji 47](#_Toc40698489)

[4.7.2 Pengaruh Terhadap Kernel Polynomial 50](#_Toc40698490)

[4.8 Implementasi Pemetaan Sebaran Lokasi Pengguna Twitter 53](#_Toc40698491)

[4.8.1 Implementasi Pencarian Lokasi Bedasarkan Username 53](#_Toc40698492)

[4.8.2 Implementasi Konversi Lokasi Menjadi Geocode 54](#_Toc40698493)

[4.8.3 Visualisasi Peta Sebaran Lokasi Menggunakan MapFolium 55](#_Toc40698494)

[4.8.4 Klusterisasi Lokasi Pengguna Menggunakan FastMarkerCluster 57](#_Toc40698495)

[4.8.5 Visualisasi Aktivitas Tweet Dengan HeatMap 61](#_Toc40698496)

[4.9 Implementasi Antarmuka 64](#_Toc40698497)

[4.9.1 Implementasi Halaman Utama 64](#_Toc40698498)

[4.9.2 Implementasi Halaman Summary 65](#_Toc40698499)

[4.9.3 Implementasi Halaman Background 66](#_Toc40698500)

[4.9.4 Implementasi Halaman Method 66](#_Toc40698501)

[BAB V 68](#_Toc40698502)

[KESIMPULAN DAN SARAN 68](#_Toc40698503)

[5.1 Kesimpulan 68](#_Toc40698504)

[5.2 Saran 69](#_Toc40698505)

[DAFTAR PUSTAKA 70](#_Toc40698506)

[LAMPIRAN 73](#_Toc40698507)

# DAFTAR TABEL

[Tabel 2.1 Penelitian Terkait 7](#_Toc41851297)

[Tabel 2.2 Confusion Matrix 19](#_Toc41851298)

[Tabel 3.1 Case Folding ….....................................................................................27](#_Toc41851299)

[Tabel 3.2 Tokenization 28](#_Toc41851300)

[Tabel 3.3 Cleansing 29](#_Toc41851301)

[Tabel 3.4 Stopword Removal 29](#_Toc41851302)

[Tabel 3.5 Stemming 30](#_Toc41851303)

[Tabel 3.6 Pengujian data 32](#_Toc41851304)

[Tabel 4.1 Pelabelan Kategori Sentimen …............................................................39](#_Toc41851305)

[Tabel 4.2 Tabel Spesifikasi Sistem 40](#_Toc41851306)

[Tabel 4.3 Penggalan Kode Query Twitterscraper 40](#_Toc41851307)

[Tabel 4.4 Penggalan Kode Proses Case Folding 41](#_Toc41851308)

[Tabel 4.5 Penggalan Kode Proses Tokenizing 41](#_Toc41851309)

[Tabel 4.6 Penggalan Kode Proses Cleansing 41](#_Toc41851310)

[Tabel 4.7 Penggalan kode proses Stopword Removing 42](#_Toc41851311)

[Tabel 4.8 Penggalan kode proses Stemming 42](#_Toc41851312)

[Tabel 4.9 Penggalan Kode proses ekstrasi fitur TF-IDF 43](#_Toc41851313)

[Tabel 4.10 Penggalan kode klasifikasi dengan Support Vector Machine 44](#_Toc41851314)

[Tabel 4. 11 Skala Pengujian 46](#_Toc41851315)

[Tabel 4. 12 Tabel Hasil Kuesioner 47](#_Toc41851316)

[Tabel 4. 13 Penggalan kode pencarian lokasi bedasarkan username 54](#_Toc41851317)

[Tabel 4.14 Implementasi Konversi Lokasi Menjadi Geocode 55](#_Toc41851318)

[Tabel 4.15 Visualiasi peta sebaran lokasi menggunakan MapFolium 56](#_Toc41851319)

[Tabel 4.16 Penggalan kode Klusterisasi Lokasi Menggunakan FastMarkerCluster 57](#_Toc41851320)

[Tabel 4.17 Hasil Sebaran Lokasi Bedasarkan Cluster Wilayah 59](#_Toc41851321)

[Tabel 4.18 Visualiasi Aktivitas Pengguna menggunakan HeatMap 61](#_Toc41851322)

# DAFTAR GAMBAR

[Gambar 3.1 Alur Umum Sistem 26](#_Toc41851404)

[Gambar 3.2 Alir Proses Pre-Processing 27](#_Toc41851405)

[Gambar 3.3 Alir Proses SVM 31](#_Toc41851406)

[Gambar 3.4 Alir Proses Pemetaan Lokasi 33](#_Toc41851407)

[Gambar 3.5 Alir Proses Perancangan Desain Antarmuka 33](#_Toc41851408)

[Gambar 3.6 Struktur Menu 34](#_Toc41851409)

[Gambar 3.7 Rancangan Tampilan Halaman Utama 35](#_Toc41851410)

[Gambar 3.8 Rancangan Tampilan Halaman Summary 36](#_Toc41851411)

[Gambar 3.9 Rancangan Tampilan Halaman Background 37](#_Toc41851412)

[Gambar 3.10 Rancangan Tampilan Halaman Method 38](#_Toc41851413)

[Gambar 4.1Grafik pengaruh akurasi terhadap data latih dan data uji …............. 47](#_Toc41851414)

[Gambar 4.2 Grafik Pengaruh Nilai Precision terhadap data uji dan data latih 48](#_Toc41851415)

[Gambar 4.3 Grafik Pengaruh Nilai Recall terhadap data uji dan data latih 49](#_Toc41851416)

[Gambar 4.4 Pengaruh nilai F1-Score terhadap data latih dan data uji 50](#_Toc41851417)

[Gambar 4.5 Perbandingan Akurasi Kernel Linear dan Kernel Polynomial 50](#_Toc41851418)

[Gambar 4.6 Perbandingan Nilai Presisi Terhadap Pergantian Kernel 51](#_Toc41851419)

[Gambar 4.7 Perbandingan Nilai Recall Terhadap Pergantian Kernel 52](#_Toc41851420)

[Gambar 4.8 Perbandingan Nilai F1-Score Terhadap Pergantian Kernel 52](#_Toc41851421)

[Gambar 4. 9 Hasil Visualiasis Peta Marker 57](#_Toc41851422)

[Gambar 4.10 Hasil Viusalisasi Peta FastMarkerCluster 59](#_Toc41851423)

[Gambar 4.11 Hasil Visualisasi Peta Heatmap dengan Label Positif 62](#_Toc41851424)

[Gambar 4.12 Hasil Visualisasi Peta Heatmap dengan Label Netral 62](#_Toc41851425)

[Gambar 4.13 Hasil Visualisasi Peta HeatMap dengan Label Negatif 63](#_Toc41851426)

[Gambar 4.14 Implementasi Halaman Utama 64](#_Toc41851427)

[Gambar 4.15 Halaman Summary 65](#_Toc41851428)

[Gambar 4.16 Implementasi Halaman Background 66](#_Toc41851429)

[Gambar 4.17 Implementasi Halaman Method 67](#_Toc41851430)

# DAFTAR LAMPIRAN

Lampiran A Confusion Matrix Pengujian Data Menggunakan Kernel Linear

Lampiran B Confusion Matrix Pengujian Data Menggunakan Kernel Polinomial

Lampiran C Partisipan Pengujian

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

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 peercabangan dari Analisis Sentimen 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, negatif dan netral, 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).*

## 1.2 Rumusan Masalah

Penelitian ini mengangkat 3 (tiga) buah rumusan yang menjadi pokok permasalahan dalam implementasi metode Pemetaan Sebaran Lokasi Pengguna Twitter Terhadap Opini Pemindahan Ibukota Republik Indonesia dan *SVM (Support Vector Machine)* untuk Klasifikasi Sentimen. Rumusan masalah tersebut adalah:

1. Bagaimana melakukan pemetaan dan memberikan analisis sebaran lokasi pengguna twitter yang mengangkat opini tersebut?
2. Bagaimana metode *SVM (Support Vector Machine)* memberikan klasifikasi sentimen pengguna twitter terhadap opini pemindahan ibukota Republik Indonesia?
3. Bagaimana metode yang digunakan mampu memberikan akurasi yang dihasilkan?

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

* 1. Melakukan pemetaan lokasi untuk mengetahui sebaran lokasi / daerah pengguna twitter.
  2. Mengimplementasikan metode SVM (*Support Vector Machine*) Sebagai klasifikasi sentimen opini pengguna twitter terhadap opini pemindahan ibukota Republik Indonesia
  3. Menjadi informasi yang dapat membantu pemerintah untuk mengetahui opini masyarakat pengguna *twitter* serta persebaran lokasi pengguna tersebut*.*
  4. Mengetahui akurasi yang dihasilkan dari pendekatan yakni SVM (Support Vector Machine) yang digunakan.

## 1.4 Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut :

* 1. Data yang digunakan berasal dari *twitter* dengan bantuan *API* dari *twitterscraper*, berupa nama pengguna, *username*, id pengguna, teks dari *tweet*, banyaknya *retweet* dan *likes*, serta lokasi pengguna yang di cantumkan pada akun *twitter*.
  2. Sistem yang dibangun hanyalah mampu mengklasifikasikan sebuah opini masyarakat terhadap pemindahan Ibukota Republik Indonesia saat ini yang bersifat positif atau negatif, dengan menggunakan metode *SVM* (*Support Vector Machine*).
  3. Sistem yang dibangun memberikan informasi pola sebaran lokasi pengguna twitter bedasarkan opini dalam penelitian ini
  4. Opini yang digunakan sebagai dataset hanya berbahasa Indonesia.
  5. Sistem yang akan dikembangkan berbasis web dan menggunakan jaringan offline / local.

## 1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat untuk beberapa komponen, antara lain:

* 1. Masyarakat dan pemerintah: penelitian ini dapat memudahkan masyarakat dan pemerintah dalam mengetahui opini pemindahan lokasi ibukota republik indonesia serta sebaran lokasi dari opini tersebut.
  2. Keilmuan: penelitian ini dapat menjadi referensi pada penelitian lain yang memiliki karakteristik *SVM* (*Support Vector Machine*)*.*

## 1.6 Metode Penelitian

Metode penelitian adalah langkah yang dimiliki dan dilakukan oleh peneliti dalam rangka mengumpulkan informasi atau data serta melakukan investigasi pada data yang diperoleh tersebut. Bagian ini menjelaskan secara sistematis penyusunan tugas akhir. Adapun bahasan yang akan dijelaskan dalam penelitian ini yaitu desain penelitian, data penelitian, dan metode yang digunakan.

## 1.6.1 Desain Penelitian

Penelitian yang dilakukan ini bertujuan untuk mengetahui klasifikasi sentimen tweet pada Twitter tentang pemindahan ibukota baru Republik Indonesia. Metode pengumpulan data yang dilakukan dalam penelitian ini adalah pengumpulan data sekunder yang didapat dari *scraping* menggunakan *twitterscraper*. Dalam proses klasifikasi menggunakan data berupa teks yaitu tweet, umumnya dilakukan dengan langkah-langkah sebagai berikut yaitu text preprocessing, ekstraksi fitur, dan proses klasifikasi.

Langkah text preprocessing yang dilakukan meliputi tahap Case *Folding*, tahap *Cleansing*, tahap *Tokenizing*, tahap *Stopword Removing* dan tahap *Stemming*.

Ekstraksi fitur dilakukan berselingan dengan preprocessing dimana fitur-fitur yang akan diambil yaitu Term Frequency-Inverse Document Frequency (TF-IDF), tahap pencarian lokasi bedasarkan username, konversi lokasi menjadi *geocode.*

Setelah mengekstrak semua fitur tersebut baru dilanjutkan dengan proses klasifikasi mengimplementasikan metode SVM (*Support Vector Machine*) serta melakukan visualisasi sebaran lokasi pengguna ke dalam bentuk Geographic Information System (SIG). Pada penelitian juga akan memperhatikan akurasi yang dihasilkan menggunakan pedekatan SVM (*Support Vector Machine*).

## 1.6.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil menggunakan Twitter API, dataset total berjumlah 481 dimana 215 tweet bersentimen positif, 216 tweet bersentimen negatif dan 50 tweet bersentimen netral. Dataset tweet yang didapat berupa tabel dengan ekstensi file CSV (*Comma Separated Value*).

## 1.6.3. Pengelolaan Data Awal

Proses yang dilakukan setelah medapatkan data terkait dengan penelitian yaitu pengolahan data sehingga data tersebut diubah dan menjadi informasi yang berguna. Data- data yang diperoleh tersebut, selanjutnya akan diolah sesuai dengan kebutuan sistem. Untuk dapat melakukan proses perhitungan selanjutnya, maka data tweet yang telah didapatkan dilakukan proses awal atau disebut sebagai text preprocessing.

## 1.6.4. Metode yang Digunakan

Pada sistem klasifikasi sentimen ini menggunakan metode SVM (*Support Vector Machine*) dan dihitung menghitung akurasi tertinggi yang didapatkan oleh sistem. Serta dalam sistem ini juga menvisualiasikan data sebaran lokasi menggunakan GIS (Geographic Information System).

# BAB II

# TINJAUAN PUSTAKA

## 2.1. Tinjauan Empiris

Tabel 2.1 Penelitian Terkait

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Judul | Penulis / Tahun | Metode | Hasil |
| 1 | *Twitter Sentiment Classification using Distant Supervision* | (Go, et al., 2009) | *SVM (Support Vector Machine)*. | *SVM (Support Vector Machine)* memberikan hasil dengan tingkat keakuratan hingga 82,2% |
| 2 | *Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features* | (Rofiqoh, et al., 2017) | *SVM (Support Vector Machine)*. | *Support Vector Machine* tanpa menggunakan *Lexicon Based Features* menghasilkan tingkat akurasi sebesar 84%. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | *Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm* | (ABDOULI, et al., 2017) | *Naive Bayes Algorithm, Map Folium* | menghasilkan data secara grafik yang diklasifikasikan di peta Maroko dengan menggunakan koordinat yang diekstrak dari dari data tweet mereka |

1. “*Twitter Sentiment Classification using Distant Supervision*”. (Go, et al., 2009)

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%.

1. “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada *Twitter* Dengan Metode *Support Vector Machine* dan *Lexicon Based Features*”. (Rofiqoh, et al., 2017)

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%.

1. “*Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm*” (ABDOULI, et al., 2017)

Dalam penelitiannya, tentang opini yang ada di masyarakat Negara 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.

1. “Sentiment Analysis Menggunakan Support Vector Machine (SVM)” (Nomleni, 2015)

Dalam penelitiannya, tentang klasifikasi sentimen dan keluhan masyarakat Kota Surabaya mengenai program/kebijakan Pemerintah Kota Surabaya pada media sosial facebook dan twitter sapawarga. Pengujian menghasilkan rata rata akurasi sebesar 80% dengan akurasi tertinggi sebesar 84.4086%.

## Tinjauan Teoritis

### Klasifikasi Sentimen

Klasifikasi sentimen merupakan bidang studi yang mengklasifikasi pendapat, sentimen, penilaian, evaluasi, sikap, dan emosi seseorang terkait suatu topik, layanan, produk, individu, organisasi, atau kegiatan tertentu (Liu, 2012). Analisis sentimen dilakukan untuk menentukan apakah opini atau komentar terhadap suatu permasalahan, memiliki kecenderungan positif atau negatif dan dapat dijadikan sebagai acuan dalam meningkatkan suatu pelayanan, ataupun meningkatkan kualitas produk.

### 2.2.2 *Twitterscraper*

*Twitterscraper* merupakan sebuah skrip sederhana untuk mengeruk data *Tweet* dengan menggunakan paket *Python* *request* untuk mengambil konten dan *Beautifulsoup4* untuk mem-parsing konten yang diambil. (Taspinar, 2017) Keunggulan menggunakan twitterscraper dibanding dengan menggunakan API Twitter adalah dari sisi pencarian data tidak dibatasi waktu / riwayat tweet. sedang menggunakan API Twitter hanya dibatasi 7 hari sebelumnya.

### 2.2.3 *Preprocessing*

*Preprocessing* adalah tahapan dimana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen*, Preprocessing* dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, et al., 2012). tahapan ini terdiri atas :

1. Case Folding

*Case folding* merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang dapat diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (pembatas)

1. Cleansing

*Cleansing* adalah proses yang dilakukan untuk membersihkan data teks dengan menghilangkan data yang tidak konsisten atau tidak relevan. Misalnya komponen khas di twitter seperti URL (Uniform Resource Locator), tanda retweet (RT) dan username

1. Tokenization

*Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), tanda kutip (“), tanda kurung (()), spasi dan karakter angka yang ada pada kata tersebut

1. Stopword Removal

*Stopword* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen, tahap dimana kata-kata umum yang tidak memiliki makna atau keterkaitan dengan sentiment analisis dihilangkan

1. Stemming

*Stemming* adalah proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi bentuk kata dasarnya, dimana untuk setiap kata yang memiliki imbuhan akan diubah menjadi kata dasar.

### ***Term Frequence – Inverse Document Frequency (TF-IDF)***

Pembobotan dilakukan untuk mendapatkan nilai dari kata/term yang berhasil diekstrak. Metode yang paling umum digunakan untuk melakukan pembobotan terhadap term adalah pembobotan TF-IDF. Metode ini banyak diterapkan dalam pencarian teks (text retrieval) dan pemrosesan teks (*text preprocessing*) (Riany , et al., 2016) Metode TF-IDF ini merupakan dibagi menjadi integrasi antar *term frequency* (TF), dan *inverse document frequency* (IDF). Metode TF-IDF dapat dirumuskan sebagai berikut :

(1)

Dimana:

= kata ke i

d = dokumen

TF(, d) = jumlah kemunculan kata pada dokumen d

IDF() = Invers Document Frequency dari kata .

Hasil dari perkalian term frequency dan inverse document frequency yang tiap nilainya didapatkan dengan persamaan (1). Pembobotan dihitung untuk tiap kata, Term Frequency menunjukkan bahwa semakin tinggi kemunculan kata pada sebuah dokumen maka semakin penting kata tersebut mewakili dokumen. sedangkan jika kata muncul pada seluruh dokumen maka kata tersebut bersifat umum atau tidak mewakili dokumen manapun.

(2)

Dimana:

= kata ke i

|D| = jumlah seluruh dokumen

IDF() = jumlah dokumen yang mengandung kata

### 2.2.5 *Confusion Matrix*

*Confusion Matrix* atau matriks kebingungan adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi (atau "classifier") pada serangkaian data uji yang nilai sebenarnya diketahui (Nomleni, 2015). Tabel berikut menunjukkan matriks kebingungan untuk classifier dua kelas.

Tabel 2.2 Confusion Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| Data Aktual | Data Prediksi | | |
| Negatif | Netral | Positif |
| Negatif | TNg | NgN | FNg |
| Netral | NNg | TN | NP |
| Positif | FP | PN | TP |

Pada penelitian ini entri dalam confusion matrix memiliki arti seperti berikut :

* + FNg adalah jumlah prediksi yang salah bahwa instance negatif
  + TNg adalah jumlah prediksi yang benar bahwa sebuah instance negatif.
  + NgN adalah jumlah kata negatif yang terprediksi netral
  + NNg adalah jumlah kata netral yang terprediksi negatif
  + TN adalah jumlah prediksi yang benar bahwa sebuah instance netral.
  + NP adalah jumlah kata netral yang terprediksi nositif
  + TP adalah jumlah prediksi yang benar bahwa sebuah instance positif
  + PN adalah jumlah kata positif yang terprediksi netral
  + FP adalah jumlah prediksi yang salah bahwa sebuah instance positif

Pada penelitian ini ukuran kinerja dari sistem yaitu akurasi, precision dan recall untuk mengukur seberapa tepat klasifikasi terhadap kelas. Recall adalah rasio jumlah data relevan yang telah ditemukan terhadap sebuah kelas yang telah diprediksi. Recall menggambarkan tingkat keberhasilan sistem dalam memanggil dokumen yang relevan. Sedangkan precision adalah rasio jumlah data relevan yang telah ditemukan terhadap data pada kelas tertentu dari dataset. Precision menggambarkan tingkat keefektifan dan ketepatan sistem dalam memanggil dokumen yang relevan dari seluruh dokumen yang diambil. Rumus perhitungan kinerja yang digunakan pada penelitian ini adalah sebagai berikut

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |
|  | (5) |
|  | (6) |

### 2.2.6 *K-Fold Cross Validation*

Cross-validation (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. K-fold crossvalidation adalah kasus khusus dari cross-validation di mana mengulangi set dataset k kali. Dalam setiap putaran, dataset akan dibagi menjadi bagian k: satu bagian digunakan untuk validasi, dan bagian yang tersisa digabung ke dalam subset pelatihan untuk evaluasi. 10-fold cross validation adalah salah satu k-fold cross validation yang direkomendasikan untuk pemilihan model terbaik (wibowo, 2017)

### 2.2.7 *Support Vector Machine*

SVM (Support Vector Machine) adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. Algoritma SVM (Support Vector Machine) asli diciptakan oleh Vladimir Vapnik (Cortes & Vapnik, 1995). SVM (Support Vector Machine) standar mengambil himpunan data input, dan memprediksi, untuk setiap masukan yang diberikan, kemungkinan masukan adalah anggota dari salah satu kelas dari dua kelas yang ada, yang membuat sebuah SVM (Support Vector Machine) sebagai penggolong nonprobabilistik linier biner.

Konsep klasifikasi dengan SVM (Support Vector Machine) adalah mencari hyperplane terbaik yang berfungsi sebagai pemisah dua kelas data (Cortes & Vapnik, 1995). SVM (Support Vector Machine) mampu bekerja pada dataset yang berdimensi tinggi dengan menggunakan kernel trik. SVM (Support Vector Machine) hanya menggunakan beberapa titik data terpilih yang berkontribusi (support vector) untuk membentuk model yang akan digunakan dalam proses klasifikasi (Rofiqoh, et al., 2017)

Persamaan Support Vector Machine:

(7)

atau

(8)

Keterangan:

𝑤 : parameter hyperplane yang dicari (garis yang tegak lurus antara garis hyperplane dan titik support vector)

𝑥 : titik data masukan Support Vector Machine

𝑎𝑖 : nilai bobot setiap titik data

𝐾(𝑥,𝑥𝑖) : fungsi kernel

𝑏 : parameter hyperplane yang dicari (nilai bias)

Untuk penelitian ini menggunakan kernel polynomial. Persamaannya:

(9)

Keterangan:

𝐾(𝑥,𝑦) : Nilai kernel dari data x dan data y

𝑥 : fitur data 1

𝑦 : fitur data 2

𝑐 : nilai konstanta

𝑑 : nilai degree

### 2.2.8 *Geocoding*

Geocoding adalah proses mengambil teks input, seperti alamat atau nama tempat, dan mengembalikan lokasi garis lintang / bujur di permukaan bumi untuk lokasi tersebut. Geocoding bergantung pada representasi komputer dari titik-titik alamat, jaringan jalan / jalan, bersama dengan batas pos dan administrasi.

Geocoding adalah tugas yang melibatkan banyak dataset dan proses, yang semuanya bekerja bersama. Geocoder dibuat dari dua komponen penting: dataset referensi dan algoritma geocoding. Masing-masing komponen ini terdiri dari sub-operasi dan sub-komponen. (Pelias, 2017).

### 2.2.9 *MapFolium*

MapFolium adalah perpustakaan Python yang memungkinkan memvisualisasikan data geospasial ke peta interaktif; ini menyediakan fasilitas untuk mengubah koordinat ke berbagai proyeksi peta. Visualisasi terjadi "inline" atau dalam lingkungan Python, menggunakan Notebook IPython dan hasilnya interaktif yang membuat perpustakaan ini sangat berguna untuk membangun dashboard (Folium, 2013).

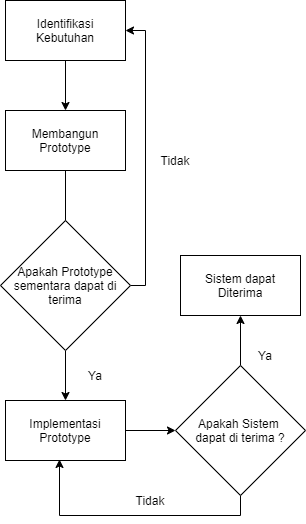
MapFolium dibangun berdasarkan data ekosistem Python dan pemetaan perpustakaan leaflet.js. Memanipulasi data dengan Python, lalu memvisualisasikannya di peta Leaflet melalui MapFolium.

MapFolium memungkinkan pengikatan data ke peta untuk visualisasi choropleth serta menyampaikan visualisasi vektor / raster / HTML yang kaya sebagai penanda di peta, selain itu memiliki sejumlah tileset bawaan dari OpenStreetMap, Mapbox, dan Stamen, dan mendukung tileset khusus dengan kunci-kunci Mapbox atau Cloudmade API. folium mendukung Image, Video, GeoJSON dan TopoJSON.

### 2.2.10 Metode *Prototype*

Prototype didefinisikan sebagai satu versi dari sebuah sistem potensial yang memberikan ide bagi para pengembang dan calon pengguna, bagaimana sistem akan berfungsi dalam bentuk yang telah selesai. (Jogianto, 2008) Dasar dari pemikiran ini adalah membuat prototipe secepat mungkin, lalu memperoleh umpan balik dari pengguna yang akan memungkinkan prototype tersebut diperbaiki kembali dengan sangat cepat. Semua rancangan diagaram atau model yang dibuat tidak diharuskan telah sempurna dan final dalam pendekatan prototype. Tujuan utama dari penyiapan rancangan adalah sebagai alat bantu dalam memberi gambaran sistem seperti materi dan menu yang perlu dimasukkan dalam prototype yang akan dikembangkan. Setelah rancangan terbentuk, dilanjutkan dengan mulai mengembangkan prototype (Malatista & Sediyono, 2012).

Metode prototype sesuai untuk menjelaskan kebutuhan pengguna secara lebih rinci karena pengguna sering mengalami kesulitan dalam penyampaian kebutuhannya secara detail tanpa melihat gambaran yang jelas. Untuk mengantisipasi agar proyek dapat berjalan sesuai dengan rencana, target waktu, dan biaya diawal, maka sebaiknya spesifikasi kebutuhan sistem harus sudah disepakati terlebih dahulu oleh pengembang dengan pengguna dalam hal ini klien. Adapun tahapan-tahapannya metode prototype digambarkan pada gambar 2.1.



Gambar 2.1 Tahapan Metode Prototype

Berikut ini adalah penjelasan dari tahapan – tahapan metode prototype yang terdapat pada Gambar 2.1 :

* 1. Identifikasi Kebutuhan

Di tahap ini pengembangan melakukan identifikasi software dan semua kebutuhan sistem yang akan dibuat.

* 1. Membangun Prototype

Pada tahapan ini adalah membuat perancangan sementara yang berfokus pada penyajian kepada pelanggan (misalnya dengan membuat input dan format output)

* 1. Evaluasi prototype

serta melakukan evaluasi untuk mengetahui apakah prototype sudah sesuai dengan tujuan yang harapkan

* 1. Implementasi prototype kedalam sistem

Pada tahap ini prototyping yang sudah disetujui akan diubah ke dalam bahasa pemrograman.

Di tahap ini dilakukan untuk menguji sistem perangkat lunak yang sudah dibuat

* 1. Sistem dapat diterima

Perangkat lunak yang sudah diuji siap digunakan.

# BAB III

# ANALISIS DAN PERANCANGAN

Pada bagian metodelogi penelitian ini menjelaskan gambaran langkah langkah yang akan dilakukan dalam menjalankan penelitian ini, langkah langkah tersebut meliputi pengumpulan data, alur metodologi penelitian, tahap preprocessing, tahap klasifikasi artikel dan tahap pengujian.

## 3.1. Pengumpulan Data

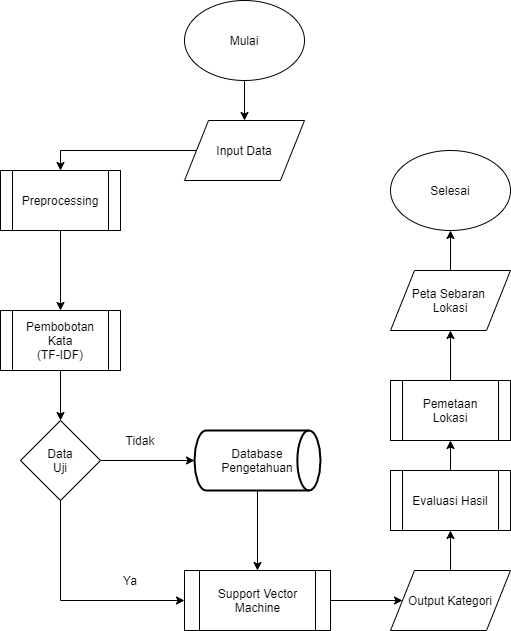
Pengumpulan data dalam penelitian ini dilakukan dengan pengumpulan data sekunder. Data diperoleh berdasarkan pencarian atas term objek pada tweet berbahasa Indonesia. Pengambilan data dilakukan berdasarkan pencarian tweet dengan kata kunci “pemindahan ibukota”, “ibu kota pindah” atau “ibukota baru” dari tanggal 1 Juni 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 tiga jenis yaitu positif untuk tweet yang memberikan respon setuju, masukan, dan dukungan, sentimen negative untuk tweet yang memberikan respon tidak setuju dari isu tersebut dan sentimen netral untuk tweet yang sifatnya tidak diantara negatif maupun positif. Data tweet yang akan dikumpulkan sebanyak 481 data tweet dengan proporsi sebanyak 44,6% dari data sebesar 215 tweet sebagai data bersentimen positif, proporsi data 44,9%, sebanyak 216 tweet sebagai data bersentimen negatif dan data bersentimen netral sebanyak 50 tweet atau 10,3% dari proporsi data. 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

## 3.2 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,

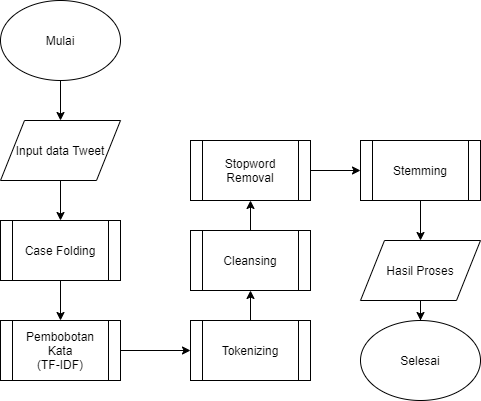
Selanjutnya akan dilakukan proses ekstraksi fitur dan lalu masuk ke tahap klasifikasi sentimen positif atau negatif, dan pada tahap akhir nantinya akan dilakukan evaluasi terhadap performa sistem dan metode yang digunakan serta visualisasi data kedalam bentuk peta



Gambar 3.1 Alur Umum Sistem

## 3.3 Pre Processing

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, et al., 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 3.2.



Gambar 3.2 Alir Proses Pre-Processing

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

### 3.3.1 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 3.1 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 |

### 3.3.2 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 3.2 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 |

### 3.3.3 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 3.3 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 |

### 3.3.4 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 3.4 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 |

### 3.3.5 Stemming

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

Tabel 3.5 Stemming

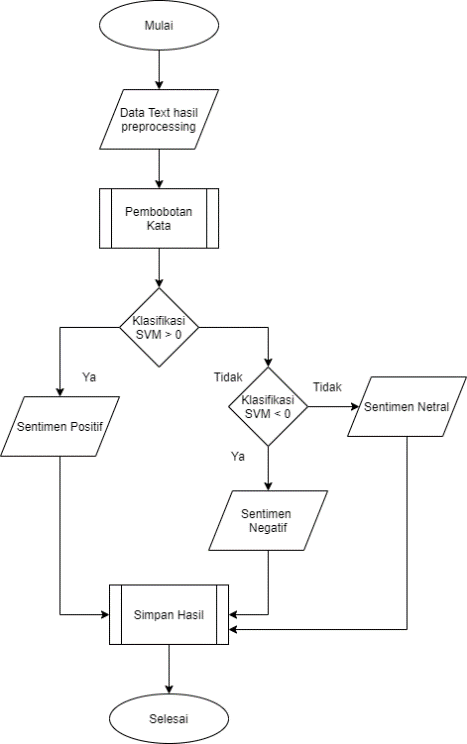
|  |  |  |  |
| --- | --- | --- | --- |
| Contoh Tweet | Hasil | | |
| Visi yg disampaikan @TsamaraDKI cukup bisa dipahami, tapi lebih setuju sama yg disampaikan @SherlyAnnavita . Kita ini bukan negara kaya, banyak hutang, buat apa pindah ibu kota hasil jual aset & ngutang lagi.  #ILCPerlukahIbuKotaPindah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

## 3. 4 Pembobotan Kata (TF-IDF)

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

## 3.5 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 = -1 maka kata tersebut dimasukkan dalam komentar negatif lalu apabila kata tersebut = 1 maka kata tersebut dimasukkan dalam komentar positif dan juga jika nilai term tersebut adalah = 0 maka akan dimasukkan kedalam komentar netral.



Gambar 3.3 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).

## 3.6. 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.

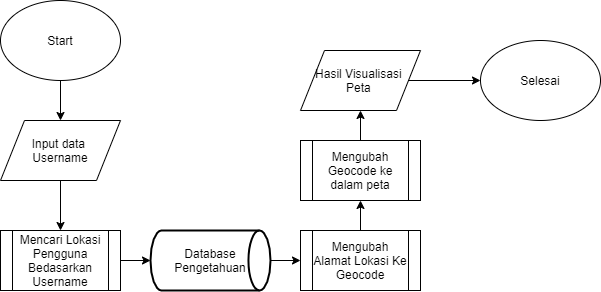
Tabel 3.6 Pengujian data

|  |  |  |  |
| --- | --- | --- | --- |
| Pengujian | Rasio Data | Data Latih | Data Uji |
| 1 | 50,1:49,1 | 241 (50,1%) | 240 (49,9%) |
| 2 | 66,7:33,3 | 321 (66,7%) | 112 (33,3%) |
| 3 | 75:25 | 361 (75%) | 120 (25%) |
| 4 | 80:20 | 385 (80%) | 96 (20%) |
| 5 | 83:17 | 401 (83%) | 80 (17%) |
| 6 | 85:15 | 413 (85%) | 68 (15%) |
| 7 | 87:13 | 421 (87%) | 60 (13%) |
| 8 | 90:10 | 428 (90%) | 53 (10%) |
| 9 | 91:9 | 433 (91%) | 48 (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.

## 3.7 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.

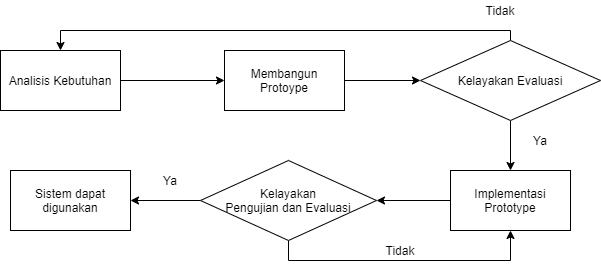


Gambar 3.4 Alir Proses Pemetaan Lokasi

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.*

## 3.8 Perancangan Desain Antarmuka

Pada tahap ini perancangan Antarmuka meliputi perancangan struktur menu dan perancangan tampilan, pada tahapan perancangan ini, penulis menggunakan pendekatan metode prototype. Dimana penulis membuat gambar rancangan dalam bentuk berupa wireframe dari sistem yang akan dibangun, meliputi perancangan struktur menu yang ada dalam sistem, serta perancangan tampilan awal sisitem. Langkah-langkah yang dilakukan penulis dalam merancang sistem ini adalah sebagai berikut :



Gambar 3.5 Alir Proses Perancangan Desain Antarmuka

1. Mengidentifikasi Kebutuhan

Pada tahapan ini penulis melakukan dentifikasi software dan semua kebutuhan yang akan dibuat, serta mencari referensi dari sistem yang sudah ada sebelumnya.

1. Mengembangkan Kebutuhan kedalam bentuk prototype serta evaluasi

Pada tahapan ini penulis membangun prototyping dengan membuat perancangan sementara yang berfokus pada penyajian kepada pengguna (misalnya dengan membuat input dan format output), lalu selanjutnya melakukan evaluasi mengenai prototype yang telah dibangun

1. Implementasi kedalam bentuk sistem

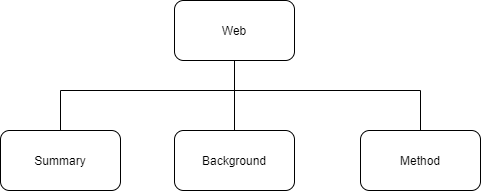
Pada tahap ini prototype yang sudah disetujui akan diubah ke dalam bahasa pemrograman

1. Pengujian sistem dan evaluasi

Pada tahap ini penulis melakukan pengujian sistem dengan beberapa skenario yang telah ditentukan, dan juga melakukan pengujian terhadap pengguna apakah program dapat berjalan sesuai tujuannya

### 3.8.1 Perancangan Struktur Menu

Pembuatan web sebaran lokasi pengguna twitter berdasarkan sentimen terhadap opini pemindahan ibukota republik indonesia menggunakan struktur menu pada Gambar 3.5



Gambar 3.6 Struktur Menu

Ada tiga hal yang ditampilkan dalam menu website ini, yaitu halaman Summary, Background, dan Method. Halaman Summary berisi tentang kesimpulan proses dari pengerjaan penelitian ini, meliputi peta sebaran lokasi, chart akumulasi dari data sentimen seperti persentase sentimen di beberapa wilayah, dan juga terdapat 10 tweet populer bedasarkan jumlah like dan retweet yang diberikan. Halaman Background menampilkan latar belakang penelitian yang dilakukan oleh penulis, Sedangkan Halaman Method menampilkan metode penelitian apa saja yang dilakukan pada penelitian ini, meliputi pengumpulan data, preprocessing, pembobotan kata, klasifikasi menggunakan SVM, dan juga melakukan pencarian lokasi bedasarkan username dari pengguna tersebut serta melakukan visualisasi kedalam bentuk peta sebaran lokasi.

### 3.8.2 Perancangan Tampilan

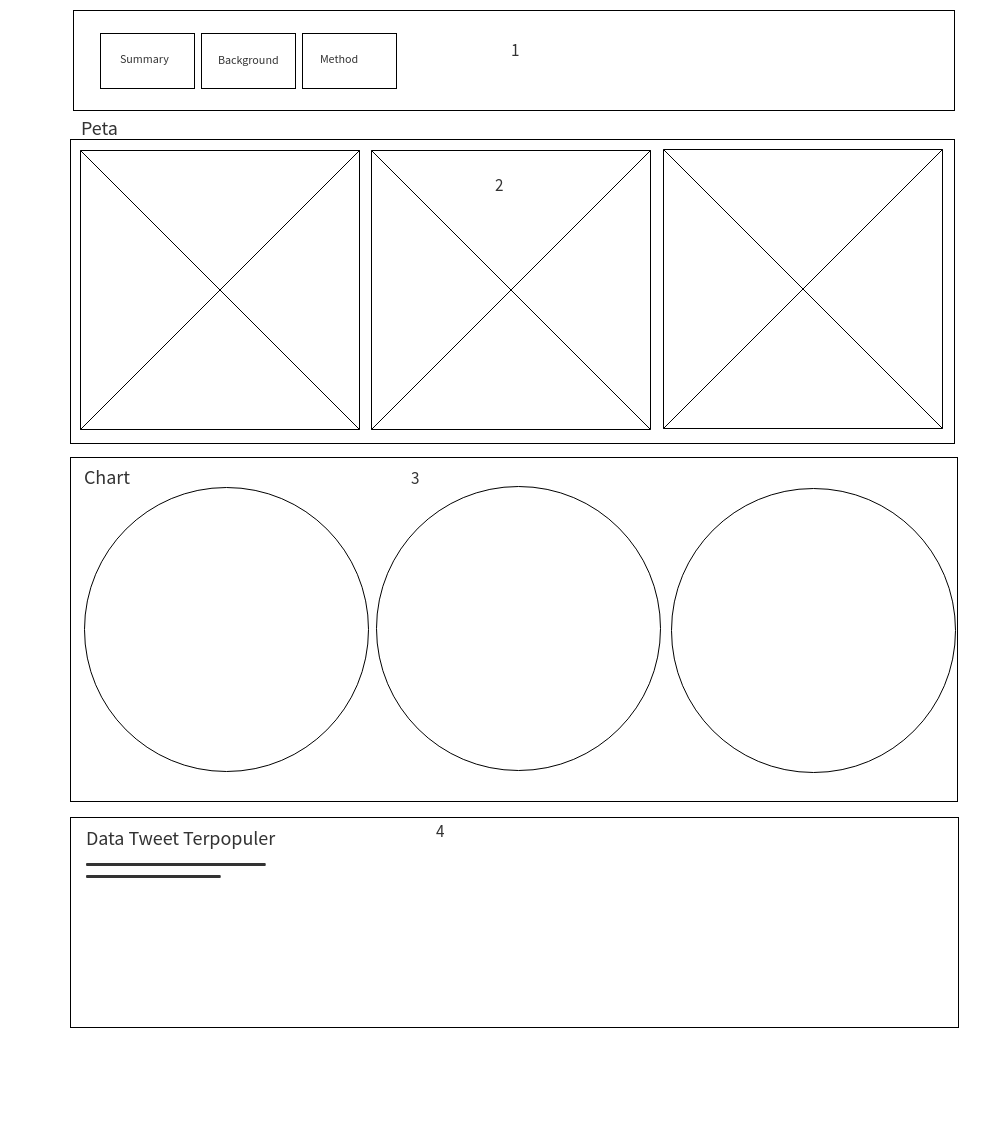
### 1. Tampilan Halaman utama Web Penelitian

Rancangan tampilan halaman utama dibagi menjadi dua bagian, di bagian atas terdapat header yang berisi menu summary, menu background, dan menu method, sedangkan di bagian tengah terdapat area tampilan/konten yang tersaji didalamnya. Yang digambarkan pada bagian Gambar 3.6.



Gambar 3.7 Rancangan Tampilan Halaman Utama

### 2. Tampilan Summary



Gambar 3.8 Rancangan Tampilan Halaman Summary

Rancangan tampilan Summary pada Gambar 3.7 terdapat 4 bagian, di bagian atas terdapat header yang berisi menu summary, menu background, dan menu method, di bagian tengah (nomor 2) terdapat area tampilan/konten yang berisi hasil dari visualiasi dalam bentuk peta, sedang di bawahnya (nomor 3) terdapat informasi persentase dan jumlah sentimen yang telah di proses ke dalam bentuk chart. Untuk data tweet terpopuler terdapat pada bagian paling bawah (nomor 4).

### 3. Tampilan Background



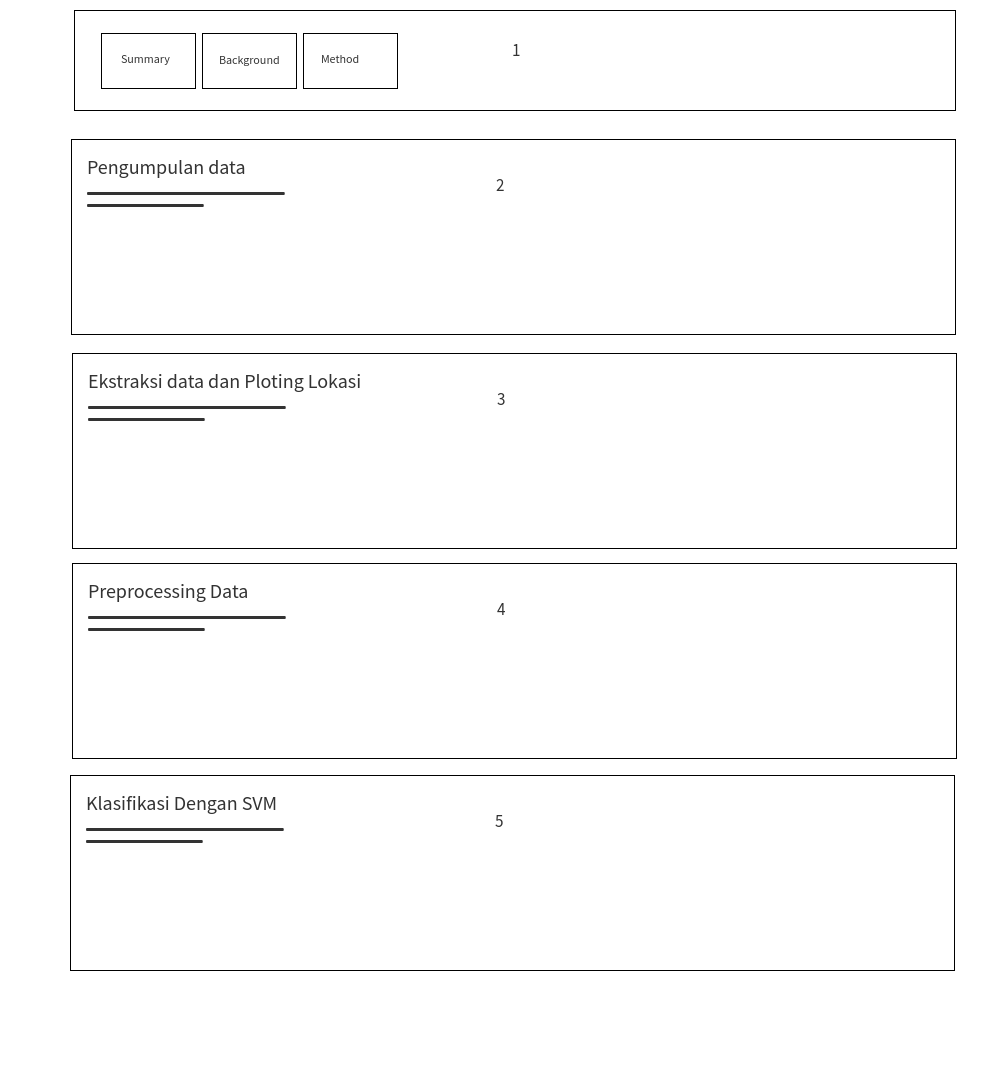
Gambar 3.9 Rancangan Tampilan Halaman Background

Pada Gambar 3.8 rancangan tampilan background dibagi menjadi dua bagian, di bagian atas terdapat header yang berisi menu bar, sedangkan di bagian tengah terdapat area tampilan/konten berupa deskripsi latar belakang, rumusan masalah dan tujuan dari penelitian yang diangkat.

### 4. Tampilan Method

Pada menu tampilan method, dijelaskan bagaimana cara penelitian ini dilakukan, bagaimana cara mengumpulkan data tweet menggunakan twitterscraper, lalu bagaimana data yang didapat diproses text tweet disederhanakan ke dalam bentuk text preprocessing, lalu cara melakukan klasifikasi svm, dan juga cara melakukan proses ekstraksi data user dan plotting kedalam bentuk peta sebaran.

Seperti pada Gambar 3.9, pada tampilan method terdapat 5 bagian, di bagian atas terdapat header yang berisi menu bar dari summary, menu background, dan menu method, di bagian tengah (nomor 2) terdapat area tampilan/konten yang berisi cara pengumpulan data tweet menggunakan twitterscraper, sedang di bawahnya (nomor 3) terdapat informasi ekstraksi data username kedalam bentuk pencarian lokasi tweet pengguna, mengubah data lokasi ke dalam bentuk geocoding, dan melakukan plotting kedalam visualisasi peta. Dijelaskan juga cara bagaimana melakukan preprocessing data text sentimen seperti case folding, tokenizing, cleansing, stopword removal, dan stemming (nomor 4), Untuk klasifikasi tweet terdapat pada bagian paling bawah (nomor 5).



Gambar 3.10 Rancangan Tampilan Halaman Method

# BAB IV

# HASIL DAN PEMBAHASAN

## 4.1 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 481 data dari berbagai akun dengan topik yang membahas pemindahan ibukota Republik Indonesia, dan data tersebut sudah dilabeli sentimen positif, negatif dan netral, untuk pelabelan sentimen peneliti bekerja sama dengan lulusan dari Fakultas Ilmu Budaya Universitas Udayana untuk memberikan pelabel sentimen yang sesuai dengan nilai/konteks dari pernyataan yang telah dikumpulkan sebelumnya. dalam pelabelan, variabel target akan mengubah data kategori tipe string dalam set data menjadi nilai numerik yang dapat dipahami model yang didapat dilihat pada Tabel 4.1.

Tabel 4.1 Pelabelan Kategori Sentimen

|  |  |
| --- | --- |
| Kategori Sentimen | Label yang diubah |
| Negatif | -1 |
| Netral | 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)*.

## 4.2 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 4.2.

Tabel 4.2 Tabel Spesifikasi Sistem

|  |  |
| --- | --- |
| Perangkat Keras / Perangkat Lunak | Spesifikasi |
| Sistem Operasi | Windows 10 |
| Tipe Sistem | 64-bit |
| Prosesor | Intel(R) Core(TM) i7-7700HQ CPU @ 2.6 GHz |
| Memori (RAM) | 8,00 GB |
| Harddisk | 1 TB |
| Perangkat Internet | Modem Broadband atau WLAN dengan Kecepatan 14 Mbps |
| Perangkat Lunak | * IDE Spyder 3 (Bahasa Pemograman Python) * Microsoft Excel 2016 |

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 Juni 2019 sampai tanggal 31 Desember 2019 dapat dilihat pada Tabel 4.3 :

Tabel 4.3 Penggalan Kode Query Twitterscraper

twitterscraper 'ibukota baru or pemindahan ibu kota' since : 2019-06-01 until : 2019-12-31 -o "ibukota\_tweets.csv"

## 4.3 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 *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.

Tabel 4.4 Penggalan Kode Proses Case Folding

case\_folding = df\_pd['text'].str.lower()

Tabel 4.5 Penggalan Kode Proses Tokenizing

nltk.tokenize.word\_tokenize(kalimat['text'])

Proses *Case Folding* dan juga *tokenizing* pada program dapat dilihat pada tabel 4.4 dan tabel 4.5. Pada proses *tokenizing* hasil token dari setiap tweet akan disimpan pada *list words*

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*.

Tabel 4.6 Penggalan Kode Proses Cleansing

def identify\_tokens(row):

text = row['text']

tokens = nltk.tokenize.word\_tokenize(text)

# taken only words (not punctuation)

token\_words = [w for w in tokens if w.isalpha()]

return token\_words

df\_pd['words'] = df\_pd.apply(identify\_tokens, axis=1)

print(df\_pd['words'])

Pada table 4.6 merupakan *Source code* proses cleansing untuk menghilangkan username, hastag (#) dan URL pada tweet jika kata tersebut mengandung simbol tersebut atau merupakan URL maka kata tersebut akan dihilangkan.

Tahapan selanjutnya yaitu tahapan *Stopword Removing* untuk menghilangkan semua kata-kata yang merupakan *Stopword* pada Bahasa Indonesia seperti pada tabel 4.7.

Tabel 4.7 Penggalan kode proses Stopword Removing

def remove\_stops(row):

my\_list = row['words']

stop = [stopword.remove(str(word)) for word in my\_list]

return (stop)

df\_pd['stop\_words'] = df\_pd.apply(remove\_stops, axis=1)

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

Tabel 4.8 Penggalan kode proses Stemming

def stem\_list(row):

my\_list = row['stop\_words']

stemmed\_list = [stemmer.stem(str(word)) for word in my\_list]

return (stemmed\_list)

df\_pd['stemmed\_words'] = df\_pd.apply(stem\_list, axis=1)

Pada table 4.8 merupakan Source code proses Stemming untuk mengurai bentuk dari suatu kata menjadi bentuk kata dasarnya.

## 4.4. 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). Penggalan code perhitungan TFIDF dapat dilihat pada tabel 4.9

Tabel 4.9 Penggalan Kode proses ekstrasi fitur TF-IDF

def create\_bag\_of\_words(X):

from sklearn.feature\_extraction.text import CountVectorizer

print ('Creating bag of words...')

vectorizer = CountVectorizer(analyzer = "word", \

tokenizer = None, \

preprocessor = None, \

stop\_words = None, \

ngram\_range = (1,2), \

max\_features = 10000

)

train\_data\_features = vectorizer.fit\_transform(X)

train\_data\_features = train\_data\_features.toarray()

from sklearn.feature\_extraction.text import TfidfTransformer

tfidf = TfidfTransformer()

tfidf\_features = tfidf.fit\_transform(train\_data\_features).toarray()

vocab = vectorizer.get\_feature\_names()

return vectorizer, vocab, train\_data\_features, tfidf\_features, tfidf

vectorizer, vocab, train\_data\_features, tfidf\_features, tfidf = \

create\_bag\_of\_words(X\_train)

bag\_dictionary = pd.DataFrame()

bag\_dictionary['ngram'] = vocab

bag\_dictionary['count'] = train\_data\_features[0]

bag\_dictionary['tfidf\_features'] = tfidf\_features[0]

bag\_dictionary.sort\_values(by=['count'], ascending=False, inplace=True)

## 4.5. Implementasi Klasifikasi dengan Support Vector Machine

Proses selanjutnya setelah mendapatkan semua fitur yang dibutuhkan adalah proses klasifikasi menggunakan algoritma SVM. Tahap ini akan menerapkan algoritma SVM yang telah dipaparkan pada subbab sebelumnya.

Tabel 4.10 Penggalan kode klasifikasi dengan Support Vector Machine

X = np.array(df\_pd['processed'])

y = np.array(df\_pd['label'])

kf=KFold(n\_splits=3, random\_state=42, shuffle=False)

print(kf) #buat tau Kfold dan parameter defaultnya

i=1 #ini gapenting, cuma buat nandain fold nya.

for train\_index, test\_index in kf.split(X):

print("Fold ", i)

print("TRAIN :", train\_index, "TEST :", test\_index)

X\_train=X[train\_index]

X\_test=X[test\_index]

y\_train=y[train\_index]

y\_test=y[test\_index]

i+=1

print("shape x\_train :", X\_train.shape)

print("shape x\_test :", X\_test.shape))

svm = Pipeline([('vect', CountVectorizer()),

('tfidf', TfidfTransformer()),

('clf', svm.SVC(C=1.0, kernel='linear', degree=2, gamma='auto')),

])

svm.fit(X\_train, y\_train)

y\_pred = svm.predict(X\_test)

actual = y\_test

predicted = y\_pred

results = confusion\_matrix(actual, predicted)

print ('Confusion Matrix :')

print(results)

print(classification\_report(y\_test, y\_pred))

print('SVM Accuracy : ',accuracy\_score(y\_pred, y\_test)\*100)

## 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 dimana hasil tersebut dapat dilihat pada subbab 4.7.1.

Pengujian yang dilakukan meliputi pengujian terhadap rasio data latih, dan pengaruh terhadap pergantian kernel dari linear ke polynomial yang telah dilakukan pada subbab 4.8.1.

Lalu kami telah melakukan evaluasi antarmuka dari sistem yang dibangun dengan menyebarkan kuisioner secara online kepada 15 audiens namun hanya ada 12 responden bersedia memberikan hasil pengujiannya, Responden berasal dari pedesaan sebanyak 42% (5 orang) dan 58% (7 orang) berasal dari perkotaan, rentang umur dari dari responden yakni 19-26 tahun. Dalam kuesioner ini menggunakan skala 1-5 yang dijelaskan pada Tabel 4.11, pada evaluasi dengan audiens ini, kami membuat terlebih dahulu prototype berupa halaman web yang dibangun sesuai dengan rancang antarmuka yang kami usulkan pada Bab 3, Subbab 3.8. dari prototype yang dibangun itulah para peserta/audiens dapat menguji antarmuka dari rancangan yang dibangun.

Tabel 4. 11 Skala Pengujian

|  |  |
| --- | --- |
| Skala | Keterangan |
| 1 | Sangat Buruk |
| 2 | Buruk |
| 3 | Netral |
| 4 | Baik |
| 5 | Sangat Baik |

Untuk pertanyaan yang diajukan dalam pengujian kuesioner ini adalah sebagai berikut:

1. Apakah Informasi yang disediakan oleh aplikasi mudah di mengerti ?
2. Apakah Konten yang ditampikan (Chart,Peta,Gambar,Borang/Form) dapat dipahami oleh pengguna
3. Apakah Aplikasi dapat memberikan hasil / output yang diinginkan ?
4. Apakah Aplikasi Nyaman digunakan ?
5. Secara keseluruhan apakah penggunaan aplikasi ini memuaskan ?
6. Apakah Aplikasi ini sesuai dengan kebutuhan ?
7. Apakah aplikasi dapat mudah dipelajari ?
8. Apakah Aplikasi mempunyai kemampuan dan fungsi sesuai yang diharapkan ?
9. Apakah Aplikasi Nyaman digunakan menggunakan Perangkat Mobile/Handheld (Bagi yang Menggunakan Perangkat Mobile/Handheld) ?

Hasil dari kuesioner yang dilakukan oleh responden, didapat nilai sebagai berikut :

Tabel 4. 12 Tabel Hasil Kuesioner

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Statistik | Pertanyaan | | | | | | | | | Total |
| P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
| Mean | 4 | 4.3 | 4.08 | 3.83 | 4.25 | 4.3 | 4 | 4.16 | 3.5 | 4.05 |
| Median | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Maximum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 4.8 |
| Minimum | 3 | 4 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2.8 |
| Mode | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Dari Pengujian yang dilakukan dengan memberikan kuesioner mengenai antarmuka ini dapat disimpulkan bahwa rata-rata peserta menyatakan bahwa antarmuka yang dibangun memberikan nilai rerata sebesar 4.05 dengan predikat Baik, yang artinya tampilan tersebut dapat diterima oleh para peserta pengujian.

## 4.7 Pengaruh Parameter Akurasi

### 4.7.1 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 3.2 menggunakan kernel linear, proses eksperimen ini membandingkan hasil dari Akurasi, Recall, Presisi, F1-Score

Gambar 4.1Grafik pengaruh akurasi terhadap data latih dan data uji

Dari hasil tersebut dapat dilihat pada Gambar 4.1, bahwa jumlah rasio data latih yang digunakan dapat mempengaruhi akurasi sentimen yang didapatkan, dimana pada percobaan ke-6 dengan rasio data latih sebesar 85:15 didapatkan hasil akurasi sebesar 85.29%, dengan rata rata akurasi yang didapatkan keseluruhan percobaan sebesar 78.51%

Gambar 4.2 Grafik Pengaruh Nilai Precision terhadap data uji dan data latih

Pada gambar 4.2. untuk data yang terprediksi sesuai dengan label aktual untuk hasil akurasi, didapatkan bahwa pada rasio data latih sebesar 85:15 (nomor 6), presisi yang didapatkan sebesar 80% untuk presisi negatif dan 88% untuk presisi positif, yang artinya percobaan dengan rasio sebesar 85:15 tersebut lebih banyak data yang terprediksi dengan benar sesuai dengan label aktualnya. namun untuk presisi netral didapatkan hasil 0% untuk rata rata dari nilai presisi yang dimiliki yaitu masing masing 73.3% untuk label negatif, 3,6% untuk label netral, dan 82,4% untuk label positif. dikarenakan jumlah data uji yang digunakan cenderung sedikit.

Gambar 4.3 Grafik Pengaruh Nilai Recall terhadap data uji dan data latih

Pada gambar 4.3, untuk sensitifitas/recall terhadap data atau kemampuan sistem sesuai dengan tingkat kebeneran untuk memanggil dokumen yang relevan. Dimana didapatkan hasil pada rasio data latih sebesar 85:15 (nomor 6), nilai recall yang didapat sebesar 80% untuk recall negatif dan 93% untuk recall positif sedang recall netral hanya mendapatkan 0%, nilai recall (nilai sensitifitas) digunakan untuk menghitung berapa banyak entitas yang relevan dengan benar atau dalam hal ini disebut false negative yakni entitas terlewat atau yang terprediksi salah dengan label.

Lalu selanjutnya pada penelitian ini juga menghitung f1-score yaitu nilai untuk menghitung kinerja dari nilai precision dan recall dari sentimen yang dimiliki, grafik pengaruh nilai F1-Score dapat dilihat pada Gambar 4.4.

Gambar 4.4 Pengaruh nilai F1-Score terhadap data latih dan data uji

Pada Gambar 4.4, nilai F1-Score tertinggi didapatkan pada rasio 85:15 (nomor 6) yakni sebesar 80% untuk F1-Score dengan label negatif dan 90% untuk F1-Score dengan label 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.

Gambar 4.5 Perbandingan Akurasi Kernel Linear dan Kernel Polynomial

Pada Gambar 4.5, nilai akurasi tertinggi yang didapat sebesar 45.8% pada skenario uji dengan data rasio 91:9 (nomor 9) dengan rata-rata nilai akurasi sebesar 40.6%, dari nilai akurasi ini kernel polynomial cenderung lebih kecil daripada hasil pengujian sebelumnya.

Selanjutnya, pada pengujian hasil presisi kernel, hasil presisi yang didapatkan kernel polynomial cenderung lebih rendah dibanding kernel linear. Pada Gambar 4.6, pada penggunaan kernel polynomial presisi negatif tertinggi berada di rasio data 91:9 (nomor 9) sebesar 46% dengan rata-rata presisi negatif hanya 40,88%, yang berarti hanya 40,88% data terprediksi dengan benar sesuai dengan label aktualnya yaitu label negatif yang di prediksi menggunakan kernel polynomial.

Gambar 4.6 Perbandingan Nilai Presisi Terhadap Pergantian Kernel

Untuk hasil recall yang didapatkan, penggunaan kernel polynomial pada penelitian ini menghasilkan nilai 100% persen pada recall negatif masing masing pengujian rasio data latih dan data uji sedangkan untuk recall positif dan recall netral menghasilkan nilai 0 pada masing masing pengujian, grafik hasil dari setiap pengujian dapat dilihat pada Gambar 4.7, untuk penggunaan kernel polynomial terhadap data atau kemampuan sistem sesuai dengan tingkat kebeneran untuk memanggil dokumen yang relevan pada penelitian ini mampu memprediksi label negatif sebesar 100%, sedang untuk kemampuan sistem memprediksi label positif dan netral yang relevan pada penelitian ini hanya menghasilkan nilai 0.

Gambar 4.7 Perbandingan Nilai Recall Terhadap Pergantian Kernel

Lalu selanjutnya pada penelitian ini juga menghitung f1-score yaitu nilai untuk menghitung kinerja dari nilai precision dan recall dari sentimen yang dimiliki. Pada kernel polynomial nilai f1-score tertinggi berada pada rasio 91:9 dengan label negatif sebesar 63%, untuk hasil dari perbandingan tersebut dapat dilihat pada Gambar 4.8

Gambar 4.8 Perbandingan Nilai F1-Score Terhadap Pergantian Kernel

Dari penelitian ini, dari akurasi yang didapat mengunakan kernel polynomial mendapatkan hasil rata rata akurasi 40.6% sedangkan kernel linear mendapatkan hasil rata rata akurasi sebesar 78.51%.

Pengujian nilai presisi kernel polynomial hanya mampu memberikan rata rata nilai sebesar 40,88% pada label yang diprediksi negatif, sedangkan label positif dan netral dari pengujian menggunakan kernel polynomial mendapatkan nilai 0, akan tetapi pada pengujian menggunakan kernel linear mendapatkan nilai rata-rata presisi dari masing masing label sebesar 73.3% untuk label negatif, 3,6% untuk label netral, dan 82,4% untuk label positif, yang berarti kemampuan kernel polynomial untuk memprediksi sesuai labelnya lebih rendah dari kernel linear.

Untuk pengujian nilai recall, kernel polynomial memberikan hasil 100% pada label negatif, ini terjadi dikarenakan banyaknya label yang relevan dengan label negatif, namun untuk label positif dan negatif hanya mampu memberikan hasil 0%, sedangkan penggunaan kernel linear mendapatkan nilai rata-rata 77,2% untuk label negatif, 0,7% untuk label netral, dan 85,6% untuk label positif.

Pada pengujian F1-Score kernel polynomial, hanya memberikan rata-rata hasil untuk label negatif sebesar 57,8% sedang untuk kernel linear memberikan rata -rata hasil untuk setiap label adalah 75,2% untuk label negatif, 1.2% untuk label netral dan 83.7% untuk label positif.

Dari hasil pengujian tersebut, maka dapat disimpulkan bahwa kernel linear baik dari kernel polynomial karena pengujian tersebut hasil akurasi, presisi, recall dan F1-Score dari kernel linear memberikan presentase lebih besar daripada pengujian menggunakan kernel polynomial

Selain itu hasil presentase yang didapatkan juga bergantung dari banyaknya data yang dimiliki, karena semakin banyaknya data maka nilai akurasi yang didapat cenderung lebih besar.

## 4.8 Implementasi Pemetaan Sebaran Lokasi Pengguna Twitter

### 4.8.1 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. Penggalan *code* pencarian lokasi menggunakan data username

Tabel 4. 13 Penggalan kode pencarian lokasi bedasarkan username

#inisialisasi target url yang dituju

base\_url = u'https://twitter.com/'

#membaca data csv yang berisi username

df\_pd = pd.read\_csv("ibukota\_new\_processed.csv",encoding = 'utf-8')

user = df\_pd['username']

user\_df = pd.DataFrame(user,columns=['username'])

user\_unique=user\_df['username'].unique()

#melakukan pencarian lokasi bedasarkan username

location = []

for user in user\_unique:

url = str(base\_url) + str(user)

r = requests.get(url)

soup = BeautifulSoup(r.text,'html.parser')

loc = soup.find('span',class\_='ProfileHeaderCard-locationText')

if hasattr(loc, 'text') == True:

location.append(''.join(loc.text.split()))

#mencetak hasil pencarian lokasi

print(location)

### 4.8.2 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. Penggalan code Konversi Lokasi Menjadi Geocode

Tabel 4.14 Implementasi Konversi Lokasi Menjadi Geocode

#mengimport library geopy

from geopy.geocoders import Nominatim

from geopy.extra.rate\_limiter import RateLimiter

#menamai locator atau aplikasi mencari geolokasi

locator = Nominatim(user\_agent="myGeocoder")

df = pd.read\_csv("location.csv")

df.head()

print(df)

## 1 - fungsi untuk menunda antar panggilan geocoding

geocode = RateLimiter(locator.geocode, min\_delay\_seconds=1)

## 2- - membuat kolom lokasi

df['location'] = df['Address'].apply(geocode)

print(df['location'])

## 3 - membuat bujur, lintang, dan ketinggian dari kolom lokasi (menghasilkan tupel)

df['point'] = df['location'].apply(lambda loc: tuple(loc.point) if loc else None)

### 4.8.3 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

Tabel 4.15 Visualiasi peta sebaran lokasi menggunakan MapFolium

map1 = folium.Map(

location=[-2.548926, 118.0148634],

#mengatur tampilan peta

tiles='cartodbpositron',

#mengatur zoom pada peta

zoom\_start=3,

)

def color(sentiment):

if sentiment in ['positif']:

col = 'green'

else: col='red'

return col

for lat,lan,text,sentiment in zip(df['latitude'],df['longitude'],df['text'],df['sentiment']):

folium.Marker(location=[lat,lan],popup = text,

icon= folium.Icon(color=color(sentiment))).add\_to(map1)

# Save the file created above

print(map1.save('point.html'))



Gambar 4. 9 Hasil Visualiasis Peta Marker

Pada bagian ini, seperti yang terlihat pada Gambar 4.9 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.

### 4.8.4 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.

Tabel 4.16 Penggalan kode Klusterisasi Lokasi Menggunakan FastMarkerCluster

map = folium.Map(

#mengatur latitude dan longitude

location=[-2.548926, 118.0148634],

#mengatur tampilan peta

tiles='cartodbpositron',

#mengatur zoom pada peta

zoom\_start=3,

)

def color(sentiment):

if sentiment in ['positif']:

col = 'green'

else:

col='red'

return col

marker\_cluster = MarkerCluster(location=[lat,lan]).add\_to(map)

for lat,lan,text,sentiment in zip(df['latitude'],df['longitude'],df['text'],df['sentiment']):

# Marker() takes location coordinates

# as a list as an argument

folium.CircleMarker(location=[lat,lan],

radius=9,

popup = text,

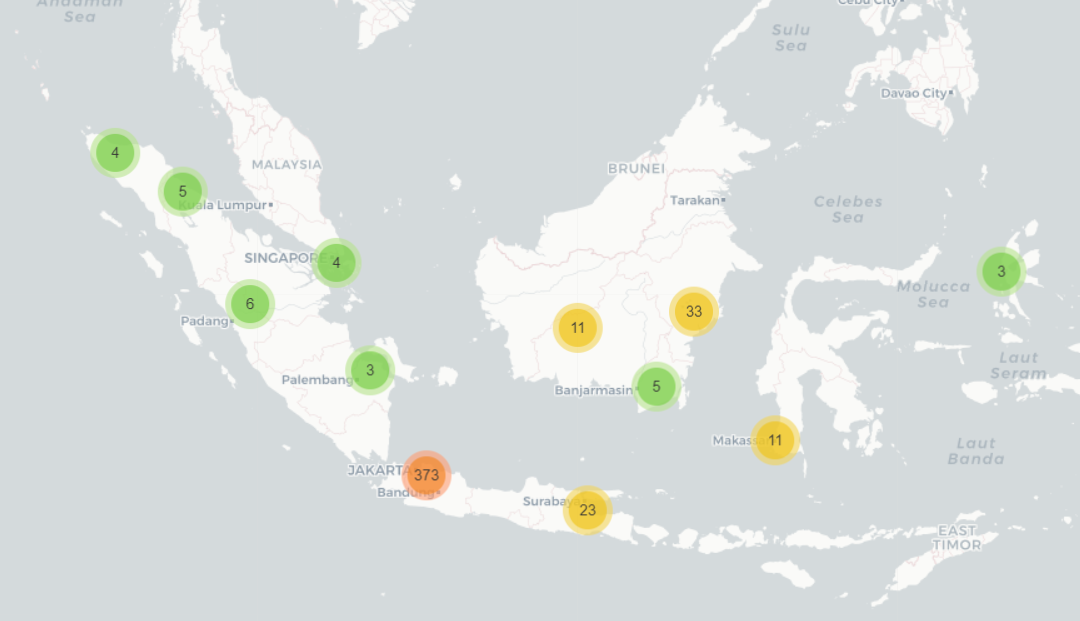
fill\_color = color(sentiment),

color = "gray",

fill\_opacity = 0.9).add\_to(marker\_cluster)

f = 'map\_cluster.html'

map.save(f)



Gambar 4.10 Hasil Viusalisasi Peta FastMarkerCluster

Pada Gambar 4.10 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 DKI Jakarta dengan jumlah tweet sebanyak 256, dimana tweet dengan sentimen negatif sebesar 110 tweet, 25 tweet bersentimen netral, dan 121 tweet bersentimen positif. Untuk data lebih lengkapnya dapat dilihat pada tabel 4.15.

Tabel 4.17 Hasil Sebaran Lokasi Bedasarkan Cluster Wilayah

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lokasi | Total | Negatif | Netral | Positif | % Negatif | % Netral | % Positif |
| Aceh | 4 | 3 | 0 | 1 | 75 | 0 | 25 |
| Bali | 3 | 3 | 0 | 0 | 100 | 0 | 0 |
| Bangka Belitung | 1 | 0 | 0 | 1 | 0 | 0 | 100 |
| Banten | 13 | 6 | 2 | 5 | 46.154 | 15.385 | 38.462 |
| DI Yogyakarta | 17 | 9 | 1 | 7 | 52.941 | 5.8824 | 41.176 |
| DKI Jakarta | 256 | 110 | 25 | 121 | 42.969 | 9.7656 | 47.266 |
| Jawa Barat | 67 | 31 | 7 | 29 | 46.269 | 10.448 | 43.284 |
| Jawa Tengah | 15 | 4 | 4 | 7 | 26.667 | 26.667 | 46.667 |
| Jawa Timur | 18 | 8 | 3 | 7 | 44.444 | 16.667 | 38.889 |
| Kalimantan Barat | 4 | 0 | 0 | 4 | 0 | 0 | 100 |
| Kalimantan Selatan | 5 | 3 | 0 | 2 | 60 | 0 | 40 |
| Kalimantan Tengah | 7 | 6 | 0 | 1 | 85.714 | 0 | 14.286 |
| Kalimantan Timur | 31 | 17 | 4 | 10 | 54.839 | 12.903 | 32.258 |
| Kalimantan Utara | 2 | 0 | 0 | 2 | 0 | 0 | 100 |
| Kepulauan Riau | 7 | 5 | 1 | 1 | 71.429 | 14.286 | 14.286 |
| Lampung | 7 | 2 | 0 | 5 | 28.571 | 0 | 71.429 |
| Maluku Utara | 3 | 0 | 0 | 3 | 0 | 0 | 100 |
| Sulawesi Selatan | 11 | 6 | 2 | 3 | 54.545 | 18.182 | 27.273 |
| Sumatera Barat | 3 | 1 | 0 | 2 | 33.333 | 0 | 66.667 |
| Sumatera Selatan | 2 | 1 | 0 | 1 | 50 | 0 | 50 |
| Sumatera Utara | 5 | 1 | 1 | 3 | 20 | 20 | 60 |

Pada tabel 4.15, dijelaskan bahwa pengguna twitter dengan jumlah sentimen negatif terbanyak yakni berada di cluster wilayah DKI Jakarta dengan jumlah sentimen negatif sebanyak 110 tweet. untuk jumlah pengguna twitter dengan sentimen positif terbanyak berada di daerah DKI Jakarta dengan jumlah 121 tweet lalu untuk cluster daerah dengan sentimen netral terbanyak juga dipegang oleh DKI Jakarta, dengan jumlah tweet sebanyak 25 tweet. untuk daerah dengan presentase sentimen negatif terbesar berada di daerah Aceh sebesar 75% dari keseluruhan data sentimen yang berada pada daerah tersebut, Bali sebanyak 100%, Kalimantan Tengah dengan 85,7%, dan Kepulauan Riau dengan 71,4%, lalu untuk daerah dengan presentase sentimen positif terbesar berada di daerah Bangka Belitung sebesar 100% dari keseluruhan data sentimen di daerah tersebut, Lampung sebesar 71,4%, Maluku Utara dengan 100%, dan Sumatera barat sebesar 66,6%, untuk daerah dengan presentase sentimen netral terbesar berada pada daerah Jawa Tengah sebesar 26,6% dari keseluruhan data, disusul dengan Sumatera Utara sebesar 20%, Sulawesi Selatan sebesar 18,1%, dan Jawa Timur dengan presentase sebesar 16,6% dari keseluruhan data sentimen di daerah tersebut.

### 4.8.5 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.

Tabel 4.18 Visualiasi Aktivitas Pengguna menggunakan HeatMap

map = folium.Map(

#mengatur latitude dan longitude

location=[-2.548926, 118.0148634],

#mengatur tampilan peta

tiles='cartodbpositron',

#mengatur zoom pada peta

zoom\_start=3,

)

heat\_df = df[df['sentiment'].str.match('positif')]

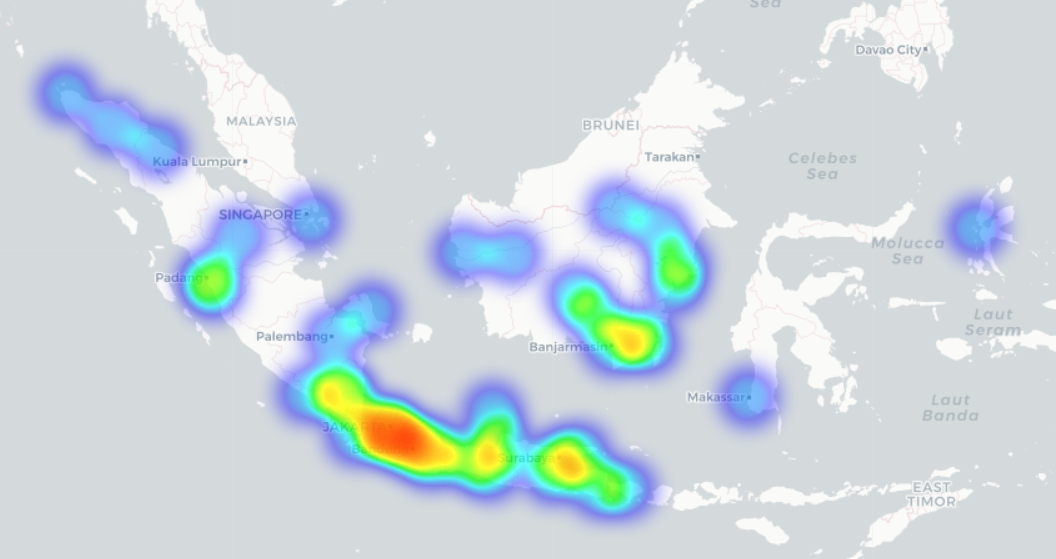
heat\_df = df[['latitude', 'longitude']]

heat\_data = [[row['latitude'],row['longitude']] for index, row in heat\_df.iterrows()]

HeatMap(heat\_data).add\_to(map)

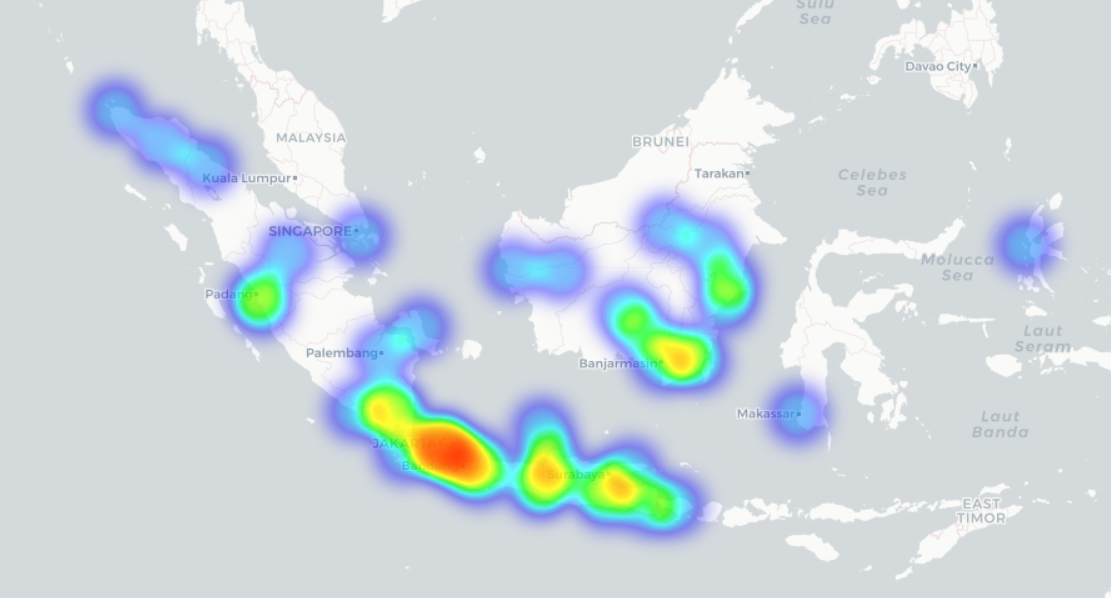
f = 'map\_heatmap.html'

map.save(f)



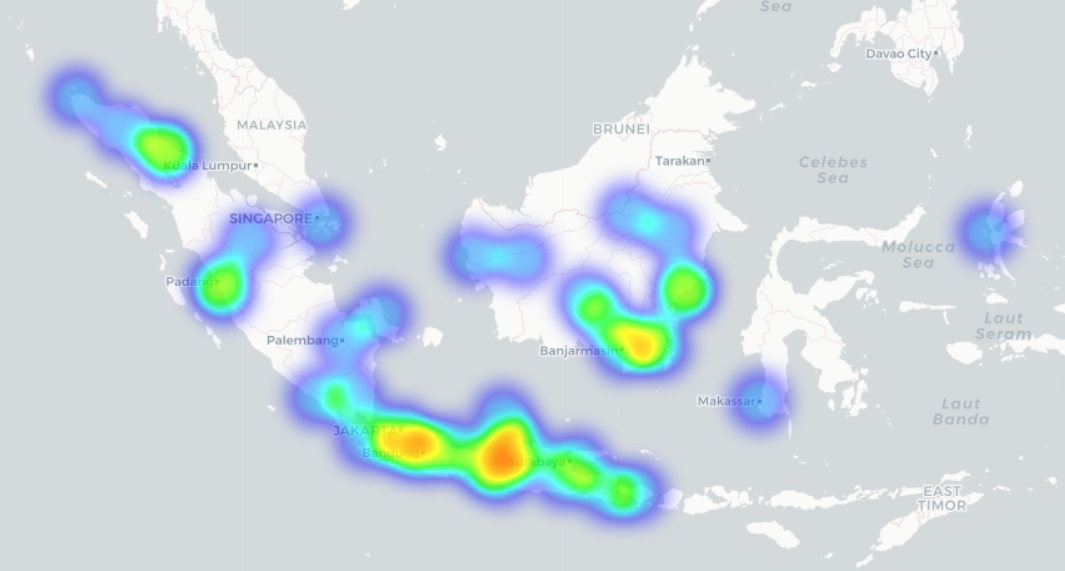
Gambar 4.11 Hasil Visualisasi Peta Heatmap dengan Label Positif

Pada gambar Gambar 4.11 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 tweet dengan sentimen positif terbanyak terdapat di pulau jawa, dengan epicentrum berada di bagian barat Pulau Jawa yang ditandai dengan semakin terangnya warna merah di daerah tersebut.



Gambar 4.12 Hasil Visualisasi Peta Heatmap dengan Label Netral

Untuk aktivitas tweet dengan sentimen netral pada penelitian ini, mengacu pada Gambar 4.12, menunjukkan untuk daerah barat Pulau Jawa masih menunjukkan aktivitas tweet terbanyak, akan tetapi pada bagian Tengah Pulau Jawa sudah mulai menunjukkan bahwa ada peningkatan aktivitas tweet bersentimen netral di daerah tersebut.



Gambar 4.13 Hasil Visualisasi Peta HeatMap dengan Label Negatif

Selanjutnya untuk daerah bersentimen negatif pada Gambar 4.13, aktivitas tweet dari pengguna tersebut telah berbeda aktivitas sebarannya, dimana sebarannya sudah tidak lagi terpusat namun sudah menyebar ke seluruh daerah dibuktikan pada Gambar 4.13 bahwa tidak ada lagi daerah dengan warna oranye terang seperti pada Gambar 4.12 dan Gambar 4.11, selain itu untuk daerah Pulau Sumatera tepatnya di daerah Aceh menunjukkan adanya aktivitas tweet bersentimen negatif dimana pada Gambar 4.12 dan Gambar 4.11 sebelumnya tidak menunjukkan aktivitas tweet di daerah tersebut yang digambarkan dengan warna hijau muda.

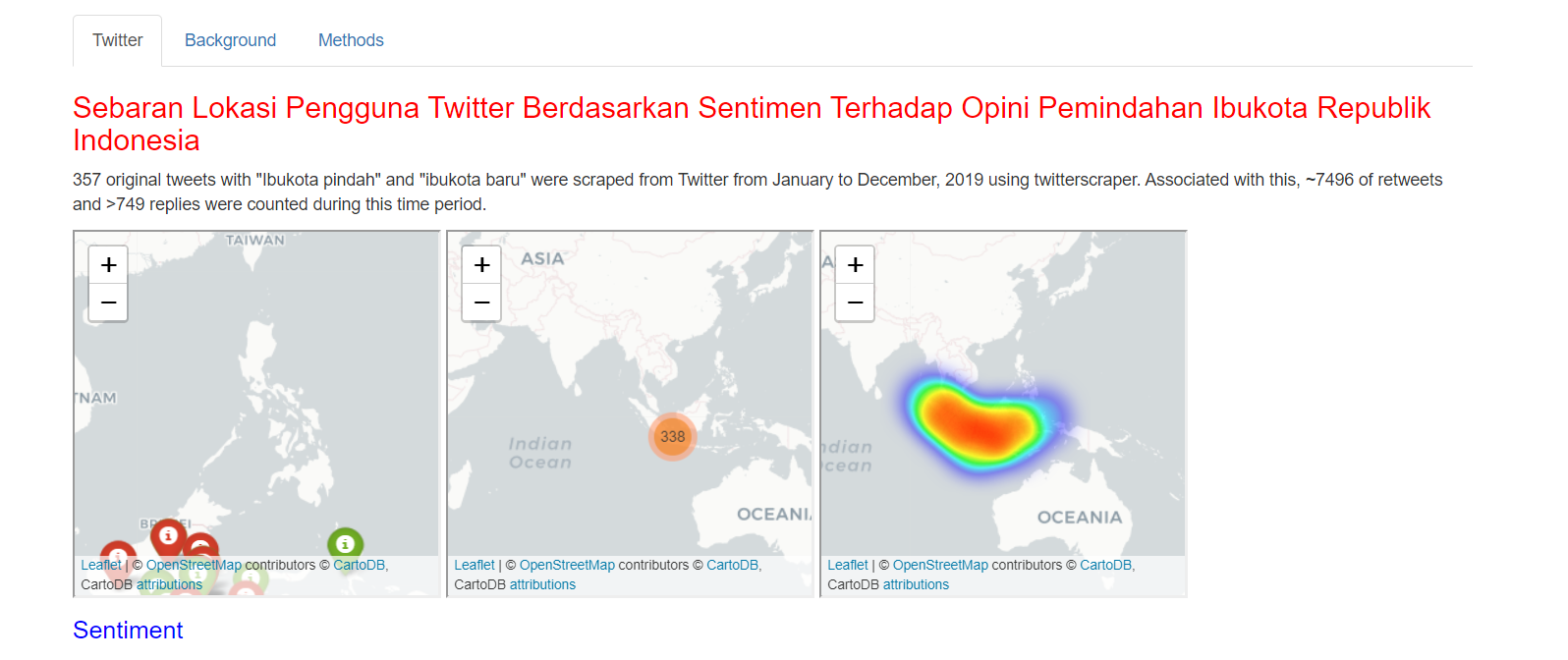
Jadi pada bagian Visualisasi dalam bentuk Heatmap ini menunjukkan bahwa daerah epicentrum frekuensi tweet terbanyak pada label positif dan netral berada pada daerah barat dari Pulau Jawa. Sedangkan untuk tweet dengan label negatif, dimana frekuensi tweet tersebut tidak lagi terpusat di daerah barat dari Pulau Jawa namun sudah tersebar ke berbagai daerah dan juga aktivitas dari label negatif juga bertambah di daerah Pulau Sumatera.

Fungsi dari HeatMap ini adalah menvisualisasikan kepadatan aktivitas pengguna twitter yang membuat opini mengenai sentimen, dan juga memberikan gambaran mengenai sebaran aktivitas pengguna twitter yang beropini mengenai pemindahan ibukota.

## 4.9 Implementasi Antarmuka

Implementasi antarmuka meliputi setiap halaman program yang dibuat beserta pengkodean dalam bentuk file program. Berikut adalah implementasi antarmuka Website Penelitian Sebaran Lokasi Pengguna Twitter Berdasarkan Sentimen Terhadap Opini Pemindahan Ibukota Republik Indonesia :

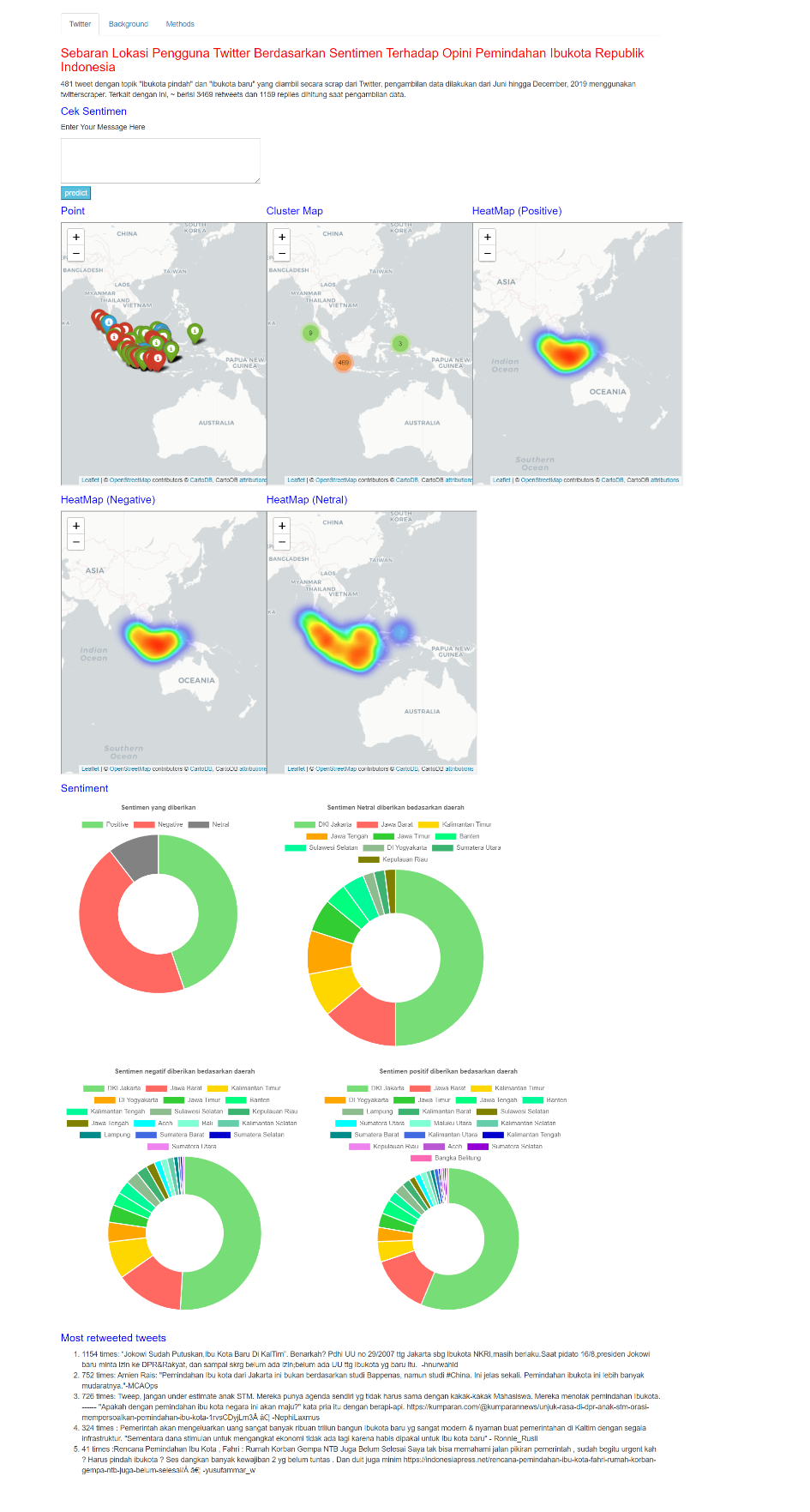
### 4.9.1 Implementasi Halaman Utama



Gambar 4.14 Implementasi Halaman Utama

Pada tampilan halaman utama berisi dua bagian, di bagian atas terdapat header yang berisi menu bar summary, menu background, dan menu method, sedangkan di bagian tengah terdapat area tampilan/konten yang tersaji didalamnya. Pada bagian ini terdapat judul dari penelitian yang dilakukan dan informasi singkat didalamnya.

### 4.9.2 Implementasi Halaman Summary



Gambar 4.15 Halaman Summary

Pada tampilan Summary terdapat 4 bagian, di bagian atas terdapat header yang berisi menu bar summary, menu background, dan menu method, di bagian tengah terdapat area tampilan/konten yang berisi hasil dari visualiasi dalam bentuk peta, pada bagian ini peta yang di tampilkan sebanyak tiga buah yakni, peta marker positif dan negatif. Peta cluster wilayah. Dan peta heatmap aktivitas tweet pada penelitian ini. sedang di bawahnya terdapat informasi persentase sentimen yang ada dimana jumlah sentimen negatif sebanyak 44,9%, jumlah sentimen positif sebesar 44,6% dan 10,3 jumlah sentimen netral dari jumlah data yang ada. Lalu terdapat juga jumlah sentimen yang tersebar di berbagai daerah dimana dibagi menjadi dua yaitu jumlah sentimen negatif yang diberikan bedasarkan daerah dan yaitu jumlah sentimen positif yang diberikan bedasarkan daerah. Untuk jumlah sentimen telah di proses ke dalam bentuk chart. Untuk data tweet terpopuler terdapat pada bagian paling bawah.

### 4.9.3 Implementasi Halaman Background



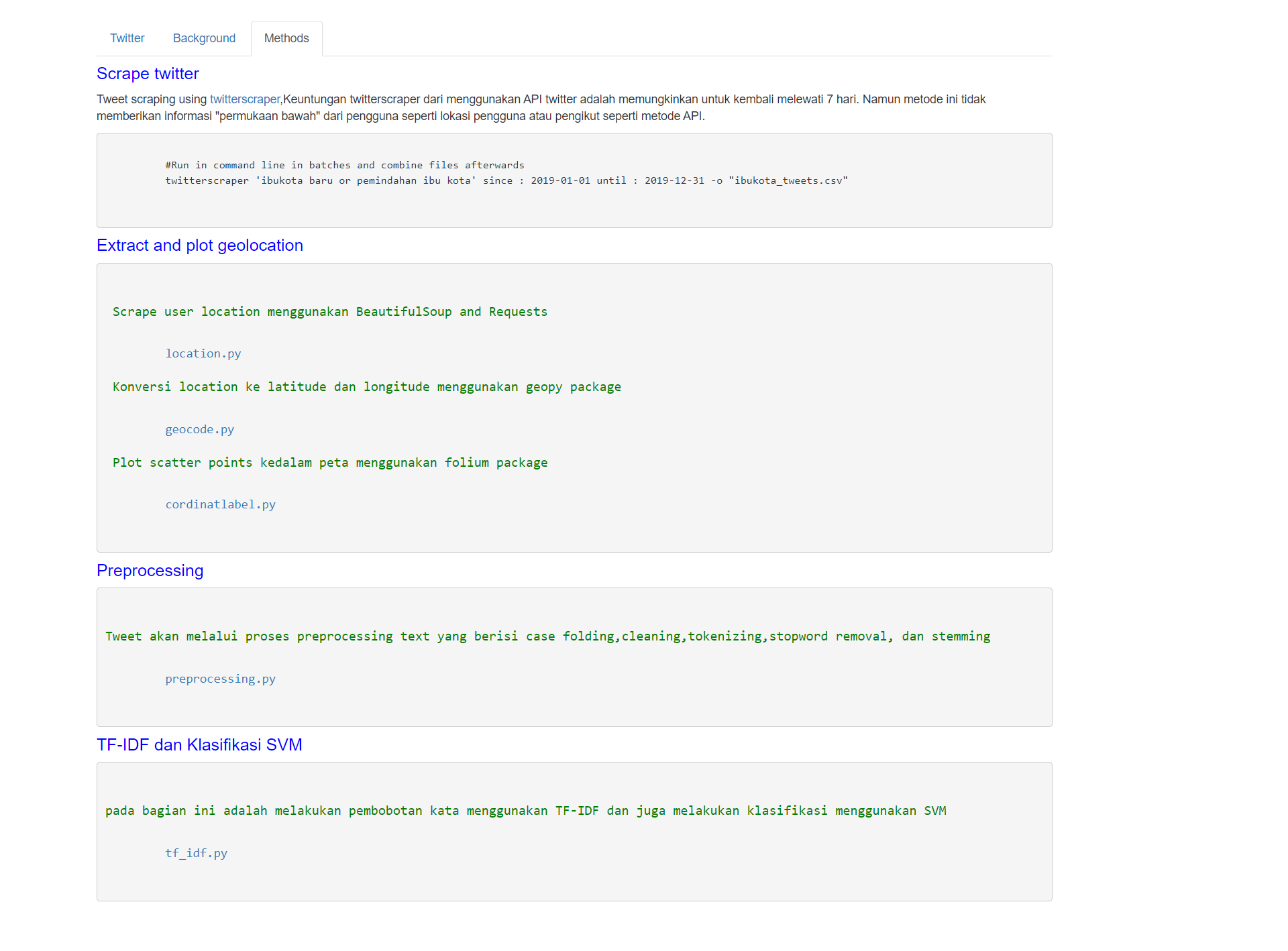
Gambar 4.16 Implementasi Halaman Background

Pada tampilan background dibagi menjadi dua bagian, di bagian atas terdapat header yang berisi menu bar, sedangkan di bagian tengah terdapat area tampilan/konten berupa deskripsi latar belakang, rumusan masalah dan tujuan dari penelitian yang diangkat. Pada bagian ini dijelaskan juga rangkuman mengenai latar belakang masalah, rumusan masalah, tujuan penelitian dan manfaat dari penelitian berikut.

### 4.9.4 Implementasi Halaman Method

Pada menu tampilan method ini, dijelaskan bagaimana cara penelitian ini dilakukan, bagaimana cara mengumpulkan data tweet menggunakan twitterscraper, lalu bagaimana data yang didapat diproses text tweet disederhanakan ke dalam bentuk text preprocessing, lalu cara melakukan klasifikasi svm, dan juga cara melakukan proses ekstraksi data user dan plotting kedalam bentuk peta sebaran.

Seperti pada gambar dibawah, pada tampilan method terdapat 5 bagian, di bagian atas terdapat header yang berisi menu bar dari summary, menu background, dan menu method, di bagian tengah terdapat area tampilan/konten yang berisi cara pengumpulan data tweet menggunakan twitterscraper dan dijelaskan mengapa memilih twitterscraper, sedang di bawahnya terdapat informasi ekstraksi data username kedalam bentuk pencarian lokasi tweet pengguna, mengubah data lokasi ke dalam bentuk geocoding, dan melakukan plotting kedalam visualisasi peta. cara bagaimana melakukan preprocessing data text sentimen seperti case folding, tokenizing, cleansing, stopword removal, dan stemming juga dijelaskan di bagian ini. Untuk pembobotan kata dan klasifikasi tweet terdapat pada bagian paling bawah.



Gambar 4.17 Implementasi Halaman Method

# BAB V

# KESIMPULAN DAN SARAN

## 5.1 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 85:15 didapatkan hasil akurasi sebesar 85.29%, dengan rata rata akurasi yang didapatkan keseluruhan percobaan sebesar 78.51%
      2. Pengujian juga dilakukan dengan mengganti kernel linear dengan kernel Polinomial juga dapat mempengaruhi kinerja dari sistem, pada hal ini didapatkan nilai akurasi tertinggi sebesar 45.8% pada skenario uji dengan data rasio 91:9 dengan rata-rata nilai akurasi sebesar 40.6%
      3. Dari hasil pengujian tersebut, maka dapat disimpulkan bahwa kernel linear baik dari kernel polynomial karena pengujian tersebut hasil akurasi, presisi, recall dan F1-Score dari kernel linear memberikan presentase lebih besar daripada pengujian menggunakan kernel polynomial.
      4. Dalam penelitian ini data username dari pengguna twitter dapat di konversi menjadi data lokasi sesuai lokasi pengguna atau lokasi dimana tweet itu dibuat, dimana menghasilkan data dalam bentuk alamat, dan kode lokasi (Geolocation) dan pada penelitian ini mampu menvisualisasikan data lokasi ke dalam 3 bentuk yakni peta marker, peta cluster dam HeatMap.
      5. Wilayah cluster dengan tweet terbanyak berada di Wilayah DKI Jakarta dengan jumlah tweet sebanyak 256, berisi tweet dengan sentimen negatif sebesar 110 tweet, 25 tweet bersentimen netral dan 121 tweet bersentimen positif. untuk daerah dengan presentase sentimen negatif terbesar berada di daerah Aceh (75%), Bali (100%), Kalimantan Tengah (85,7%), dan Kepulauan Riau (71,4%), daerah dengan presentase sentimen positif terbesar berada di daerah Bangka Belitung (100%), Lampung (71,4%), Maluku Utara (100%), dan Sumatera barat (66,6%), untuk daerah dengan presentase sentimen netral terbesar berada pada daerah Jawa Tengah (26,6%) dari keseluruhan data, disusul dengan Sumatera Utara (20%), Sulawesi Selatan (18,1%), dan Jawa Timur (16,6%).
      6. Visualisasi dalam bentuk Heatmap menunjukkan bahwa daerah epicentrum frekuensi tweet terbanyak pada label positif dan netral berada pada daerah barat dari Pulau Jawa. Sedangkan untuk tweet dengan label negatif, dimana frekuensi tweet tersebut tidak lagi terpusat di daerah barat dari Pulau Jawa namun sudah tersebar ke berbagai daerah dan juga aktivitas dari label negatif juga bertambah di daerah Pulau Sumatera.

## 5.2 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

# DAFTAR PUSTAKA

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),* Volume Vol. 15, pp. 191-200.

Bappenas, 2019. *https://www.bappenas.go.id/.* [Online]   
Available at: https://www.bappenas.go.id/files/diskusi-ikn-2/Analisis%20Hal%20Keamanan%20Pemindahan%20Ibu%20Kota\_kastaf-v3.pdf  
[Diakses 15 September 2019].

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

Folium, 2013. *Folium Documentation.* [Online]   
Available at: https://python-visualization.github.io/folium/  
[Diakses 8 February 2020].

Go, A., Bhayani , R. & Huang, L., 2009. Twitter Sentiment Classification using Distant Supervision. *CS224N Project Report, Stanford,* p. 2009.

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

Hutasoit, W., 2018. Analisa Pemindahan Ibukota Negara. *Dedikasi,* pp. 109-128.

Jogianto, H., 2008. *Sistem Teknologi Informasi.* Yogyakarta: Penerbit Andi.

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

Kominfo, 2019. *https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\_satker.* [Online]   
Available at: https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\_satker

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

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

Malatista, B. R. & Sediyono, E., 2012. Model Pembelajaran Matematika untuk Siswa Kelas IV SDLB Penyandang Tunarungu dan Wicara dengan Metode Komtal Berbantuan Komputer. *Jurnal Teknik Informatika dan Sistem Informasi,* 7(1), pp. 7-26.

Media, K., 2019. *kompas.com.* [Online]   
Available at: https://money.kompas.com/read/2019/05/09/184859926/kepala-bappenas-pemindahan-ibu-kota-masuk-rpjmn-2020-2024

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 .

Pelias, 2017. *Pelias.io.* [Online]   
Available at: https://pelias.io/  
[Diakses 8 February 2020].

Pressman, R. S., 2012. Pendekatan Praktisi Rekayasa Perangkat Lunak. *Yogyakarta: Andi Offset.*

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

Rofiqoh, U., Perdana, R. S. & Fauzi , M. A., 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,* p. 964X.

Taspinar, A. a. S. L., 2017. *Twitterscraper 0.2. 7: Python Package Index.* s.l.:s.n.

Twitter, 2019. *https://about.twitter.com/id/company.html.* [Online]   
Available at: https://about.twitter.com/id/company.html

wibowo, a., 2017. *Binus University Graduate Program.* [Online]   
Available at: https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/  
[Diakses 8 February 2020].

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

# 

# LAMPIRAN

**Lampiran A**

**Confusion Matrix Pengujian Data Menggunakan Kernel Linear**

1. Confusion Matrix Pengujian Data Rasio (50,1:49,1) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix - 1 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 76 | 0 | 28 | 0.68 | 0.73 | 0.71 | 104 |
| Netral | 8 | 1 | 6 | 0.33 | 0.07 | 0.11 | 15 |
| Positif | 27 | 2 | 92 | 0.73 | 0.76 | 0.74 | 121 |

1. Confusion Matrix Pengujian Data Rasio (66,7:33,3) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix - 2 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 46 | 0 | 14 | 0.66 | 0.77 | 0.71 | 60 |
| Netral | 7 | 0 | 3 | 0 | 0 | 0 | 10 |
| Positif | 17 | 0 | 73 | 0.81 | 0.81 | 0.81 | 90 |

1. Confusion Matrix Pengujian Data Rasio (75:25) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix - 3 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 39 | 0 | 11 | 0.65 | 0.78 | 0.71 | 50 |
| Netral | 6 | 0 | 1 | 0 | 0 | 0 | 7 |
| Positif | 15 | 0 | 48 | 0.8 | 0.76 | 0.78 | 63 |

1. Confusion Matrix Pengujian Data Rasio (80:20) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix - 4 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 33 | 0 | 9 | 0.75 | 0.79 | 0.77 | 42 |
| Netral | 4 | 0 | 0 | 0 | 0 | 0 | 4 |
| Positif | 7 | 0 | 43 | 0.83 | 0.86 | 0.84 | 50 |

1. Confusion Matrix Pengujian Data Rasio (87:13) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix - 5 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 24 | 0 | 7 | 0.75 | 0.77 | 0.76 | 31 |
| Netral | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| Positif | 5 | 0 | 41 | 0.85 | 0.89 | 0.87 | 46 |

1. Confusion Matrix Pengujian Data Rasio (85:15) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix – 6 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | Support |
| Negatif | 20 | 0 | 5 | 0.8 | 0.8 | 0.8 | 25 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 3 | 0 | 38 | 0.88 | 0.93 | 0.9 | 41 |

1. Confusion Matrix Pengujian Data Rasio (87:13) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix – 7 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | Support |
| Negatif | 17 | 0 | 5 | 0.77 | 0.77 | 0.77 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 3 | 0 | 33 | 0.87 | 0.92 | 0.89 | 36 |

1. Confusion Matrix Pengujian Data Rasio (90:10) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix – 8 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | Support |
| Negatif | 17 | 0 | 5 | 0.77 | 0.77 | 0.77 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 3 | 0 | 26 | 0.84 | 0.9 | 0.87 | 29 |

1. Confusion Matrix Pengujian Data Rasio (91:9) Menggunakan Kernel Linear

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix – 9 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | Support |
| Negatif | 17 | 0 | 5 | 0.77 | 0.77 | 0.77 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 3 | 0 | 21 | 0.81 | 0.88 | 0.84 | 24 |

**Lampiran B**

**Confusion Matrix Pengujian Data Menggunakan Kernel Polinomial**

1. Confusion Matrix Pengujian Data Rasio (50,1:49,1) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 1 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | Support |
| Negatif | 104 | 0 | 0 | 0.43 | 1 | 0.6 | 104 |
| Netral | 15 | 0 | 0 | 0 | 0 | 0 | 15 |
| Positif | 121 | 0 | 0 | 0 | 0 | 0 | 121 |

1. Confusion Matrix Pengujian Data Rasio (66,7:33,3) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 2 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 60 | 0 | 0 | 0.38 | 1 | 0.55 | 60 |
| Netral | 10 | 0 | 0 | 0 | 0 | 0 | 10 |
| Positif | 90 | 0 | 0 | 0 | 0 | 0 | 90 |

1. Confusion Matrix Pengujian Data Rasio (75:25) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 3 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 50 | 0 | 0 | 0.42 | 1 | 0.59 | 50 |
| Netral | 7 | 0 | 0 | 0 | 0 | 0 | 7 |
| Positif | 63 | 0 | 0 | 0 | 0 | 0 | 63 |

1. Confusion Matrix Pengujian Data Rasio (80:20) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 4 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 42 | 0 | 0 | 0.44 | 1 | 0.61 | 42 |
| Netral | 4 | 0 | 0 | 0 | 0 | 0 | 4 |
| Positif | 50 | 0 | 0 | 0 | 0 | 0 | 50 |

1. Confusion Matrix Pengujian Data Rasio (87:13) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 5 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 31 | 0 | 0 | 0.39 | 1 | 0.56 | 31 |
| Netral | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| Positif | 46 | 0 | 0 | 0 | 0 | 0 | 46 |

1. Confusion Matrix Pengujian Data Rasio (85:15) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 6 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 25 | 0 | 0 | 0.37 | 1 | 0.54 | 25 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 41 | 0 | 0 | 0 | 0 | 0 | 41 |

1. Confusion Matrix Pengujian Data Rasio (87:13) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 7 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 22 | 0 | 0 | 0.37 | 1 | 0.54 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 36 | 0 | 0 | 0 | 0 | 0 | 36 |

1. Confusion Matrix Pengujian Data Rasio (90:10) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 8 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 22 | 0 | 0 | 0.42 | 1 | 0.59 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 29 | 0 | 0 | 0 | 0 | 0 | 29 |

1. Confusion Matrix Pengujian Data Rasio (91:9) Menggunakan Kernel Polynomial

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrix 9 | | | | | | | |
|  | Negatif | Netral | Positif | precision | recall | f1-score | support |
| Negatif | 22 | 0 | 0 | 0.46 | 1 | 0.63 | 22 |
| Netral | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Positif | 24 | 0 | 0 | 0 | 0 | 0 | 24 |

**LAMPIRAN C**

**PARTISIPAN PENGUJIAN DAN HASIL KUESIONER**

1. Partisipan Pengujian

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Nama | Jenis Kelamin | Umur | Daerah Tinggal |
| 1 | I Made Wardana | Laki-Laki | 21 | Perkotaan |
| 2 | I Made Pegi Kurnia Amerta | Laki-Laki | 22 | Pedesaan |
| 3 | Luh Putu Eka Nadya Wati | Perempuan | 19 | Perkotaan |
| 4 | Ni Putu Novia Ardiyanti | Perempuan | 19 | Pedesaan |
| 5 | Gede Pawitradi | Laki-Laki | 22 | Pedesaan |
| 6 | I Made Tangkas Wahyu Kencana Yuda | Laki-Laki | 22 | Pedesaan |
| 7 | I Made Wardana | Laki-Laki | 22 | Perkotaan |
| 8 | Irianto Liko Koten | Laki-Laki | 21 | Perkotaan |
| 9 | Muhammad Afif Ubaidillah | Laki-Laki | 26 | Perkotaan |
| 10 | Ade Wahyu Andriyanto | Laki-Laki | 21 | Perkotaan |
| 11 | Devin Reness | Perempuan | 22 | Pedesaan |
| 12 | Desak Putu Sekar Merta Putri | Laki-Laki | 22 | Perkotaan |
| 13 | I Gede Angga Purnajiwa Arimbawa | Laki-Laki | 21 | Perkotaan |

1. Hasil Kuesioner

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Pertanyaan | Skala | | | | | Total |
| Sangat Buruk | Buruk | Netral | Baik | Sangat Baik |
| Apakah Informasi yang disediakan oleh aplikasi mudah di mengerti | 0 | 0 | 1 | 10 | 1 | **12** |
| Apakah Konten yang ditampikan (Chart,Peta,Gambar,Borang/Form) dapat dipahami oleh pengguna | 0 | 0 | 0 | 8 | 4 | **12** |
| Apakah Aplikasi dapat memberikan hasil / output yang diinginkan ? | 0 | 0 | 3 | 5 | 4 | **12** |
| Apakah Aplikasi Nyaman digunakan ? | 0 | 0 | 3 | 8 | 1 | **12** |
| Secara keseluruhan apakah penggunaan aplikasi ini memuaskan | 0 | 0 | 1 | 7 | 4 | **12** |
| Apakah Aplikasi ini sesuai dengan kebutuhan | 0 | 0 | 1 | 6 | 5 | **12** |
| Apakah aplikasi dapat mudah dipelajari | 0 | 1 | 1 | 7 | 3 | **12** |
| Apakah Aplikasi mempunyai kemampuan dan fungsi sesuai yang diharapkan ? | 0 | 0 | 2 | 6 | 4 | **12** |
| Apakah Aplikasi Nyaman digunakan menggunakan Perangkat Mobile/Handheld (Bagi yang Menggunakan Perangkat Mobile/Handheld) | 0 | 1 | 4 | 7 | 0 | **12** |
| **Total Masukan** | **0** | **2** | **16** | **64** | **26** | **108** |