**IMPLEMENTASI ANALISIS *CLUSTERING* DAN SENTIMEN DATA TWITTER PADA OPINI WISATA PANTAI MENGGUNAKAN METODE *K-MEANS***

**SKRIP­SI**

Digunakan sebagai Syarat Maju Ujian Diploma IV

Politeknik Negeri Malang

**Oleh :  
RIZKI ANDI IRAWAN NIM. 1341180145**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**AGUSTUS 2017**

**IMPLEMENTASI ANALISIS *CLUSTERING* DAN SENTIMEN DATA TWITTER PADA OPINI WISATA PANTAI MENGGUNAKAN METODE *K-MEANS***

HAMAN JUDUL

**SKRIPSI**

Digunakan sebagai Syarat Maju Ujian Diploma IV

Politeknik Negeri Malang

**Oleh :  
RIZKI ANDI IRAWAN NIM. 1341180145**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**AGUSTUS 2017**

# HALAMAN PENGESAHAN

**IMPLEMENTASI ANALISIS *CLUSTERING* DAN SENTIMEN DATA TWITTER PADA OPINI WISATA PANTAI MENGGUNAKAN METODE *K-MEANS***

Disusun oleh:

RIZKI ANDI IRAWAN NIM. 1341180145

Skripsi ini telah diuji pada tanggal Agustus 2017

Disetujui oleh:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Penguji I | : | Dr. Eng. Rosa Andrie Asmara, S.T., M.MT  NIP. 19801010 200501 1 001 | ........................... |
|  |  |  |  |  |
| 2. | Penguji II | : | Ridwan Rismanto, SST., M.Kom  NIP. 19860318 201212 1 001 | ........................... |
|  |  |  |  |  |
| 3. | Pembimbing I | : | Yan Watequlis Syaifudin, S.T., M.MT  NIP. 19810105 200501 1 005 | ........................... |
|  |  |  |  |  |
| 4. | Pembimbing II | : | Dr.Eng. Faisal Rahutomo, S.T., M.Kom  NIP. 19771116 200501 1 008 | ........................... |

Mengetahui,

|  |  |
| --- | --- |
| Ketua Jurusan  Teknologi Informasi | Ketua Program Studi  Teknik Informatika |
| Rudy Ariyanto, S.T., M.Cs. | Ir. Deddy Kusbianto P., M.MKom. |
| NIP. 19711110 199903 1 002 | |  | | --- | | NIP. 19621128 198811 1 001 | |

# PERNYATAAN

Dengan ini saya menyatakan bahwa Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Ahli Madya/kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Malang, Agustus 2017

Rizki Andi Irawan

# ABSTRAK

**Irawan, Rizki Andi.** “Implementasi Analisis *Clustering* dan Sentimen Data Twitter pada Opini Wisata Pantai Menggunakan Metode *K-Means*”. **Pembimbing: (1) Yan Watequlis Syaifudin, S.T., M.MT. (2) Dr.Eng. Faisal Rahutomo, S.T., M.Kom.**

**Skripsi, Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang, 2017.**

Analisis sentimen atau *opinion mining* merupakan topik riset dari cabang penelitian *text mining*. Fokus dari analisis sentimen adalah melakukan analisis opini dari suatu dokumen teks, sehingga membantu usaha untuk melakukan riset pasar atas opini publik. Data opini diperoleh dari jejaring sosial Twitter dalam Bahasa Indonesia dengan topik suatu pantai. Klasifikasi opini diperlukan untuk memudahkan pengguna dalam melihat opini positif, negatif, ataupun netral. Algoritma yang digunakan dalam klasifikasi adalah *Support Vector Machine*.

Pada penelitian ini digunakan dataset dari sepuluh pantai yang ada di Indonesia sebanyak 500 *tweet*. Hasil akurasi dari klasifikasi menggunakan algoritma *Support Vector Machine* sebesar 74,39%. Selanjutnya data opini dari kuesioner ditambahkan untuk mengelompokkan pantai berdasarkan ketersediaan sumber daya, fasilitas, akses, kesiapan masyarakat, potensi pasar dan posisi pariwisata. Dalam proses pengelompokan data ini digunakan metode *K-Means*.

**Kata Kunci:** Analisis Sentimen, Wisata Pantai, *Support Vector Machine*, *K-Means Clustering*

.

.

# *ABSTRACT*

**Irawan, Rizki Andi.** “*Implementation of Clustering Analysis and Twitter Data Sentiment on Beach Opinion Using K-Means Method*”. ***Advisors*: (1) Yan Watequlis Syaifudin, S.T., M.MT. (2) Dr.Eng. Faisal Rahutomo, S.T., M.Kom.**

***Thesis, Informatics Engineering Study Program, Department of Information Technology, State Polytechnic of Malang, 2017.***

*Sentiments analysis or opinion mining is a research topic from text mining research branch. The focus of sentiment analysis is to conduct an opinion analysis of a text document, for helping to conduct market research on public opinion. The opinion data is obtained from the social network Twitter in Indonesian with the topic of a beach. The classification of opinion is needed to facilitate the user in viewing positive, negative, or neutral opinion. The algorithm used in the classification is the Support Vector Machine.*

*This study uses dataset from ten beaches in Indonesia as much as 500 tweets. The accuracy of classification using Support Vector Machine algorithm is 74.39%. Furthermore, the opinion data from the questionnaires were added to classify the coast based on the availability of resources, facilities, access, community readiness, market potential and tourism position. In grouping data process, it used K-Means method.*

**Keywords:** *Sentiment Analysis*, *Beach*, *Support Vector Machine*, *K-Means Clustering*

# KATA PENGANTAR

Puji Syukur Alhamdulillah penulis panjatkan kehadirat ALLAH SWT atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan laporan skripsi dengan judul “IMPLEMENTASI ANALISIS *CLUSTERING* DAN SENTIMEN DATA TWITTER PADA OPINI WISATA PANTAI MENGGUNAKAN METODE *K-MEANS*”. Laporan skripsi ini penulis susun sebagai persyaratan untuk menyelesaikan studi program Diploma IV Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang.

Penulis menyadari tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan laporan skripsi ini tidak akan dapat berjalan dengan baik. Untuk itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Bapak Rudy Ariyanto, ST., MCs. selaku Ketua Jurusan Teknologi Informasi Politeknik Negeri Malang.
2. Bapak Ir. Deddy Kusbianto P., M.MKom. selaku Ketua Program Studi Teknik Informatika Politeknik Negeri Malang.
3. Yan Watequlis Syaifudin, S.T., M.MT. selaku Dosen Pembimbing Skripsi.
4. Dr.Eng. Faisal Rahutomo, S.T., M.Kom. selaku Dosen Pembimbing Skripsi.
5. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Laporan Skripsi dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan akhir ini, masih banyak terdapat kekurangan dan kelemahan yang dimiliki baik sistematika penulisan maupun penggunaan bahasa. Untuk itu penulis mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan laporan. Semoga laporan ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, Agustus 2017

Rizki Andi Irawan

# DAFTAR ISI

Halaman

[HALAMAN PENGESAHAN iii](#_Toc492454111)

[PERNYATAAN iv](#_Toc492454112)

[ABSTRAK v](#_Toc492454113)

[*ABSTRACT* vi](#_Toc492454114)

[KATA PENGANTAR vii](#_Toc492454115)

[DAFTAR ISI viii](#_Toc492454116)

[DAFTAR GAMBAR x](#_Toc492454117)

[DAFTAR TABEL xi](#_Toc492454118)

[DAFTAR LAMPIRAN xii](#_Toc492454119)

[BAB I. PENDAHULUAN 1](#_Toc492454120)

[1.1 Latar Belakang 1](#_Toc492454121)

[1.2 Rumusan Masalah 2](#_Toc492454122)

[1.3 Tujuan 2](#_Toc492454123)

[1.4 Batasan Masalah 2](#_Toc492454124)

[1.5 Sistematika Penulisan 3](#_Toc492454125)

[BAB II. LANDASAN TEORI 5](#_Toc492454126)

[2.1 Analisis Sentimen 5](#_Toc492454127)

[2.2 Twitter 6](#_Toc492454128)

[2.3 *Support Vector Machine* 7](#_Toc492454129)

[2.3.1 *Kernel* *Trick* dan non linear *SVM* 8](#_Toc492454130)

[2.3.2 Gaussian Kernel 10](#_Toc492454131)

[2.4 K-Means Clustering 11](#_Toc492454132)

[2.5 PHP 12](#_Toc492454133)

[BAB III. METODOLOGI PENELITIAN 13](#_Toc492454134)

[3.1 Metode Pengumpulan Data 13](#_Toc492454135)

[3.2 Metode Pengolahan Data 13](#_Toc492454136)

[*3.2.1* *Preprocessing* 13](#_Toc492454137)

[3.3 Metode Pengembangan Sistem 14](#_Toc492454138)

[3.4 Data 16](#_Toc492454139)

[3.5 Metode Pengujian 20](#_Toc492454140)

[BAB IV. ANALISIS DAN PERANCANGAN 21](#_Toc492454141)

[4.1 Analisa Sistem 21](#_Toc492454142)

[4.1.1 Kebutuhan Perangkat Lunak 21](#_Toc492454143)

[4.1.2 Kebutuhan Perangkat Keras 21](#_Toc492454144)

[4.2 Analisa Algoritma 21](#_Toc492454145)

[4.2.1 Klasifikasi dengan *Support Vector Machine* 22](#_Toc492454146)

[4.2.2 Klasterisasi dengan *K-Means* 25](#_Toc492454147)

[4.3 Perancangan Sistem 30](#_Toc492454148)

[4.3.1 Use Case Diagram 30](#_Toc492454149)

[4.3.2 Flowchart 34](#_Toc492454150)

[4.3.3 Mockup 36](#_Toc492454151)

[BAB V. IMPLEMENTASI 41](#_Toc492454152)

[5.1 Implementasi Basis Data 41](#_Toc492454153)

[5.2 Implementasi Tampilan Antarmuka 42](#_Toc492454154)

[BAB VI. PENGUJIAN DAN PEMBAHASAN 47](#_Toc492454155)

[6.1 Uji Coba 47](#_Toc492454156)

[6.2 Pengujian Performa Fungsionalitas Sistem 47](#_Toc492454157)

[6.3 Pengujian Akurasi Sistem 48](#_Toc492454158)

[6.4 Analisa Hasil Clustering 51](#_Toc492454159)

[BAB VII. PENUTUP 53](#_Toc492454160)

[7.1 Kesimpulan 53](#_Toc492454161)

[7.2 Saran 53](#_Toc492454162)

[DAFTAR PUSTAKA 54](#_Toc492454163)

[LAMPIRAN 55](#_Toc492454164)

# DAFTAR GAMBAR

Halaman

[Gambar 2.1 *Hyperplane* terbaik 7](#_Toc492454165)

[Gambar 2.2 *Hyperplane* terbentuk 8](#_Toc492454166)

[Gambar 2.3 Pemetaan input space 9](#_Toc492454167)

[Gambar 3.1 Tahapan Preprocessing 13](#_Toc492454168)

[Gambar 3.2 Metode Waterfall 15](#_Toc492454169)

[Gambar 4.1 *Use Case Diagram* 30](#_Toc492454170)

[Gambar 4.2 *Flowchart* Klasifikasi *SVM* 34](#_Toc492454171)

[Gambar 4.3 Flowchart Clustering K-Means 35](#_Toc492454172)

[Gambar 4.2 Desain Form Menu Utama 36](#_Toc492454173)

[Gambar 4.5 Desain Form Dataset 37](#_Toc492454174)

[Gambar 4.6 Desain Form Kuesioner 38](#_Toc492454175)

[Gambar 4.7 Desain Form *Stopword* 38](#_Toc492454176)

[Gambar 4.8 Desain Form *Preprocessing* 39](#_Toc492454177)

[Gambar 4.9 Desain Form TF-IDF 39](#_Toc492454178)

[Gambar 4.10 Form Training 40](#_Toc492454179)

[Gambar 4.11 Form Testing 40](#_Toc492454180)

[Gambar 5.1 Struktur Tabel dataset 41](#_Toc492454181)

[Gambar 5.2 Struktur Tabel stopword 41](#_Toc492454182)

[Gambar 5.3 Struktur Tabel kuesioner 42](#_Toc492454183)

[Gambar 5.4 Implementasi Halaman Form Menu Utama 42](#_Toc492454184)

[Gambar 5.5 Implementasi Halaman Dataset 43](#_Toc492454185)

[Gambar 5.6 Implementasi Halaman Kuesioner 43](#_Toc492454186)

[Gambar 5.7 Implementasi Halaman Stopword 44](#_Toc492454187)

[Gambar 5.8 Implementasi Halaman Preprocessing 44](#_Toc492454188)

[Gambar 5.9 Implementasi Halaman TF-IDF 45](#_Toc492454189)

[Gambar 5.10 Implementasi Halaman Training 45](#_Toc492454190)

[Gambar 5.11 Implementasi Halaman Testing 46](#_Toc492454191)

[Gambar 6.1 Hasil Clustering Pantai 51](#_Toc492454192)

# DAFTAR TABEL

Halaman

[Tabel 2.1 *Kernel* yang Umum Digunakan 10](#_Toc492454193)

[Tabel 3.1 *Tweet* Pantai Anyer 17](#_Toc492454194)

[Tabel 3.2 Tweet Pantai Balekambang 17](#_Toc492454195)

[Tabel 3.3 Tweet Pantai Bantol 17](#_Toc492454196)

[Tabel 3.4 Tweet Pantai Goa Cina 18](#_Toc492454197)

[Tabel 3.5 Tweet Pantai Klayar 18](#_Toc492454198)

[Tabel 3.6 Tweet Pantai Pangandaran 18](#_Toc492454199)

[Tabel 3.7 Tweet Pantai Plengkung 19](#_Toc492454200)

[Tabel 3.8 Tweet Pantai Popoh 19](#_Toc492454201)

[Tabel 3.9 Tweet Pantai Sawarna 19](#_Toc492454202)

[Tabel 3.10 Tweet Pulau Sempu 20](#_Toc492454203)

[Tabel 4.6 Contoh Daftar Kalimat 22](#_Toc492454204)

[Tabel 4.7 Contoh *Term Frequency* 22](#_Toc492454205)

[Tabel 4.3 Kernel yang Umum Digunakan 23](#_Toc492454206)

[Tabel 4.4 Tabel Siswa dan Nilai 25](#_Toc492454207)

[Tabel 4.5 Pusat Awal *Cluster* 26](#_Toc492454208)

[Tabel 4.6 Jarak dengan Pusat *Centroid* 26](#_Toc492454209)

[Tabel 4.7 Pengelompokkan Data 27](#_Toc492454210)

[Tabel 4.8 Centroid Baru Iterasi ke 2 28](#_Toc492454211)

[Tabel 4.9 Jarak dengan Pusat *Centroid* Iterasi ke 2 28](#_Toc492454212)

[Tabel 4.10 Pengelompokkan Data Iterasi ke 2 29](#_Toc492454213)

[Tabel 4.11 Hasil *Clustering* Data Mahasiswa 29](#_Toc492454214)

[Tabel 4.12 Deskripsi *Use Case* 31](#_Toc492454215)

[Tabel 4.13 Skenario *Use Case* Melihat Dataset *Tweet* 31](#_Toc492454216)

[Tabel 4.14 Skenario *Use Case* Melihat Kuesioner 31](#_Toc492454217)

[Tabel 4.15 Skenario *Use Case* Melakukan *Stopword* 32](#_Toc492454218)

[Tabel 4.16 Skenario *Use Case* Melakukan *Preprocessing* 32](#_Toc492454219)

[Tabel 4.17 Skenario *Use Case* Melakukan Perhitungan *TF-IDF* 32](#_Toc492454220)

[Tabel 4.18 Skenario *Use Case* Melakukan *Training* dan *Testing* 33](#_Toc492454221)

[Tabel 4.19 Skenario *Use Case* Melakukan *Clustering* 33](#_Toc492454222)

[Tabel 6.1 Hasil Pengujian Fungsionalitas 47](#_Toc492454223)

[Tabel 6.2 Perbandingan Klasifikasi 48](#_Toc492454224)

[Tabel 6.3 Pengujian Akurasi Sistem 51](#_Toc492454225)

# DAFTAR LAMPIRAN

Lampiran 1 Kode Program Data Training

Lampiran 2 Kode Program Data Testing

Lampiran 3 Lembar Bimbingan I

Lampiran 4 Lembar Bimbingan II

Lampiran 5 Lembar Form Revisi Penguji I

Lampiran 6 Lembar Form Revisi Penguji II

# BAB I. PENDAHULUAN

## Latar Belakang

Pada era globalisasi, media sosial saat ini sudah sangat umum dan banyak digunakan untuk kepentingan masyarakat. Dalam implementasinya, media sosial lebih banyak digunakan untuk kegiatan jual beli, menyampaikan informasi, bahkan sebagai media untuk mengekspresikan diri. Pertumbuhan media sosial sangat cepat tidak hanya penggunaanya yang terus meningkat, namun semakin banyaknya media sosial yang ditawarkan melalui aplikasi mobile ataupun website. Salah satu media sosial yang banyak digunakan adalah *Twitter*. Kata yang terkandung dalam *Twitter* adalah bahasa alami manusia yang merupakan bahasa dengan struktur kompleks.

Indonesia yang dikenal sebagai negara maritim merupakan salah satu negara yang terkenal akan pariwisatanya. Banyak objek wisata yang berpotensi untuk menarik pendatang lokal maupun wisatawan. Salah satu objek wisata adalah pantai yang sangat banyak dijumpai hampir di seluruh wilayah Indonesia. Berkembangnya objek wisata pantai di Indonesia berawal karena banyaknya masyarakat yang menyampaikan opini tentang kunjungan wisata mereka dari mulut ke mulut, bahkan di era globalisasi sekarang opini menjadi lebih mudah disampaikan karena banyaknya media sosial salah satunya seperti *Twitter*.

Opini merupakan pendapat, pikiran, atau pendirian. Semakin bertambah banyak kalimat kritik dan saran pada media sosial sehingga dapat membentuk opini masyarakat. Opini ini dapat dijadikan masukan terhadap penilaian suatu objek. Salah satu objek yang sering dibahas dalam *twitter* adalah mengenai keadaan dan pelayanan di suatu tempat pariwisata yang nantinya akan menjadi opini terhadap masyara­­kat. Analisis sentimen atau opinion mining dapat digunakan untuk memperoleh gambaran umum persepsi masyarakat terhadap kualitas layanan, apakah cenderung positif, negatif atau netral. Opini biasanya bernilai positif atau negatif tetapi dapat dikategorisasikan juga menjadi baik, sangat baik, buruk, dan sangat buruk.

Wisata pantai adalah salah satu tempat wisata yang menjadi pendukung sektor perekononiam di Indonesia. Wisata pantai menjadi trend karena banyaknya opini masyarakat yang berkunjung. Berdasarkan trend tersebut, permasalahan seperti ketersediaan sumber daya, fasilitas, akses, kesiapan masyarakat, potensi pasar dan posisi pariwisata dapat dilihat dari opini masyarakat. Selanjutnya opini tersebut ditambahkan dengan hasil kuesioner dari beberapa aspek.

Dalam penelitian ini, peneliti mengusulkan metode *Support Vector Machine* *(SVM)* dalam menganalisis opini yang telah disampaikan oleh beberapa pengguna twitter agar mendapatkan suatu informasi. Metode ini nantinya akan mengklasifikasikan semua opini tentang pariwisata kedalam kategori baik, buruk, maupun netral. Selanjutnya hasil opini tersebut ditambahkan kuesioner dari masyarakat untuk di c*luster.*

## Rumusan Masalah

Berdasarkan latar belakang masalah yang diuaraikan, maka dapat dirumuskan masalah sebagai berikut :

1. Bagaimana menganalisa sentimen pada sebuah *tweet* masyarakat mengenai suatu objek wisata pantai ?
2. Bagaimana mengembangkan suatu aplikasi untuk analisis sentimen dengan algoritma *Support Vector Machine* ?
3. Bagaimana cara menyajikan hasil *clustering* dari hasil opini masyarakat ?

## Tujuan

Tujuan yang ingin dicapai dalam pembuatan penelitian ini adalah :

1. Mengalaisa sentimen pada sebuah *tweet* masyarakat mengenai suatu objek wisata pantai di Indonesia.
2. Mengembangkan suatu aplikasi untuk analisis sentimen dengan algoritma *Support Vector Machine*
3. Menyajikan hasil clustering dari hasil opini masyarakat*.*

## Batasan Masalah

Berdasarkan dengan latar belakang dan perumusan masalah yang telah diuraikan, dibatasi hal-hal sebagai berikut :

1. *Tweet* yang dianalisis hanyalah *tweet* berbahasa Indonesia.
2. Objek yang digunakan hanyalah objek pariwisata pantai yang ada di Indonesia
3. Objek pantai yang akan dianalisa sebanyak 10 pantai.
4. Analisis sentimen dilakukan menggunakan algoritma *Support Vector Machine*.

## Sistematika Penulisan

Dalam penyusunan skripsi ini, pembahasan penulis disajikan dalam tujuh bab pokok bahasan, antara lain sebagai berikut.

|  |  |
| --- | --- |
| BAB I | PENDAHULUAN  Pada bab ini menerangkan tentang latar belakang, perumusan masalah, tujuan, batasan masalah dan sistematika penulisan |
| BAB II | LANDASAN TEORI  Pada bab ini berisikan teori-toeri yang relevan yang melengkapi latar belakang sekaligus memberi *review* tentang pustaka yang telah dibaca selama masa pencarian solusi terhadap masalah yang diangkat dalam tugas akhir. |
| BAB III | METODOLOGI PENELITIAN  Pada bab ini terdiri dari langkah-langkah yang akan membimbing penulis memilih metode, teknik, prosedur apa yang tepat dan *tools* apa yang akan digunakan sehingga setiap tahap penelitian dapat dilakukan dengan tepat. |
| BAB IV | ANALISIS DAN PERANCANGAN  Pada bab ini diuraikan dengan jelas sistem yang akan dibuat dan kebutuhan sistem yang meliputi kebutuhan fungsional dan kebutuhan non-fungsional. Rancangan sistem meliputi rancangan model, rancangan arsitektur sistem, rancangan proses, rancangan prosedural, rancangan data dan rancangan *user interface.* |
| BAB V | IMPLEMENTASI  Pada bab ini implementasi sistem dipaparkan secara detail sesuai rancangan dan komponen *(tools)* bahasa pemrograman yang dipakai. Implementasi rancangan proses dapat disertai dengan potongan kode pada proses yang dimaksud. |
| BAB VI | PENGUJIAN DAN PEMBAHASAN  Pada bab ini pengujian dipaparkan secara detail mengenai metode pengujian, tujuan pengujian, proses pengujian serta analisa pengujian. Sedangkan pada pembahasan dapat disajikan dalam bentuk teoritik baik secara kualitatif maupun kuantitatif. Penyajian hasil penelitian dapat diperjelas dengan menggunakan tabel, kurva, grafik atau bentuk lain yang dapat digunakan sesuain keperluan secraa lengkap dan jelas. |
| BAB VII | KESIMPULAN  Pada bab ini berisi uraian singkat dan jelas tentang hasil tugas akhir yang diperoleh sesuai dengan tujuan penelitian. Apabila diperlukan saran dapat digunakan untuk menyampaikan hal-hal yang dapat diperbaiki, dikembangkan atau dijadikan penelitian lebih lanjut. |

# BAB II. LANDASAN TEORI­­

## Analisis Sentimen

Analisis sentimen adalah bidang studi yang mengalaisis pendapat, sentiment, evaluasi, penilaian, sikap, dan emosi seseorang terhadap sebuah produk, organisasi, individu, masalah, peristiwa atau topik [1].

Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur/tingkat aspek dan menentukan apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas/aspek bersifat positif, negatif atau netral. Lebih lanjut sentiment analysis dapat menyatakan emosional sedih, gembira, atau marah.

Ekspresi atau sentiment mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda. Sebagai contoh, adalah hal yang baik untuk mengatakan alur film tidak terprediksi, tapi adalah hal yang tidak baik jika ‘tidak terprediksi’ dinyatakan pada kemudi dari kendaraan. Bahkan pada produk tertentu, kata-kata yang sama dapat menggambarkan makna kebalikan, contoh adalah hal yang buruk untuk waktu *start-up* pada kamera digital jika dinyatakan “lama”, namun jika” lama” dinyatakan pada usia baterai maka akan menjadi hal positif. Oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining. *Sentiment analysis* digunakan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang menuju ke opini positif atau negatif [2].

Analisis sentimen terdiri dari tiga level analisis yaitu :

1. Level Dokumen

Level dokumen menganalisis satu dokumen penuh dan mengklasifikasikan dokumen tersebut memiliki sentiment positif atau negative. Level analisis ini berasumsi bahwa keseluruhan dokumen hanya berisi opini tentang satu entitas saja. Level analisis ini tidak cocok diterapkan pada dokumen yang membandingkan lebih dari satu entitas.

1. Level Kalimat

Level kalimat menganalisis satu kalimat dan menentukan tiap kaliamat bernilai sentiment positif, negatif, atau netral. Sentimen netral berarti kalimat tersebut bukan opini.

1. Level Entitas dan Aspek

Level aspek tidak melakukan analisis pada konstruksi bahasa (dokumen, paragraph, kalimat, klausa, atau frase) melainkan langsung pada opini itu sendiri. Hal ini didasari bahwa opini terdiri dari sentiment (positif atau negative) dan target dari opini tersebut. Tujuan level analisis ini adalah untuk menemukan sentiment pada tiap aspek yang dibahas.

## Twitter

*Twitter* adalah layanan jejaring sosial dan mikroblog yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (*Tweet*). *Twitter* didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, *Twitter* telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan "pesan singkat dari Internet." Di *Twitter*, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler.

*Twitter* mengalami pertumbuhan yang pesat dan dengan cepat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif. Lonjakan penggunaan *Twitter* umumnya berlangsung saat terjadinya peristiwa-peristiwa populer. Pada awal 2013, pengguna *Twitter* mengirimkan lebih dari 340 juta kicauan per hari, dan *Twitter* menangani lebih dari 1,6 miliar permintaan pencarian per hari. Hal ini menyebabkan posisi *Twitter* naik ke peringkat kedua sebagai situs jejaring sosial yang paling sering dikunjungi di dunia, dari yang sebelumnya menempati peringkat dua puluh dua.

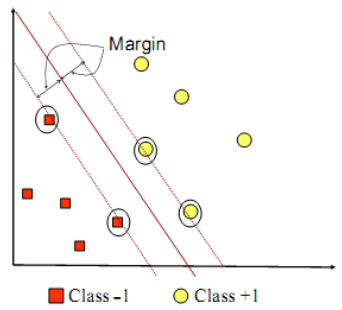
Tingginya popularitas *Twitter* menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai aspek, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat. *Twitter* juga dihadapkan pada berbagai masalah dan kontroversi seperti masalah keamanan dan privasi pengguna, gugatan hukum, dan penyensoran.

*Twitter* dimiliki dan dioperasikan oleh Twitter, Inc., yang berbasis di San Francisco, dengan kantor dan peladen tambahan terdapat di New York City, Boston, dan San Antonio. Hingga Mei 2015, Twitter telah memiliki lebih dari 500 juta pengguna, 302 juta di antaranya adalah pengguna aktif.

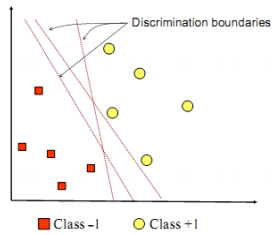
## *Support Vector Machine*

*Support Vector Machine* pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep- konsep unggulan dalam bidang pattern recognition [3]. *SVM* adalah algoritma *machine learning* yang bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input* *space*.

Gambar 2.1 memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah *class* : +1 dan –1. *Pattern* yang tergabung pada *class* –1 disimbolkan dengan warna merah (kotak), sedangkan *pattern* pada *class* +1, disimbolkan dengan warna kuning (lingkaran). Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut.



Gambar 2. *Hyperplane* terbaik

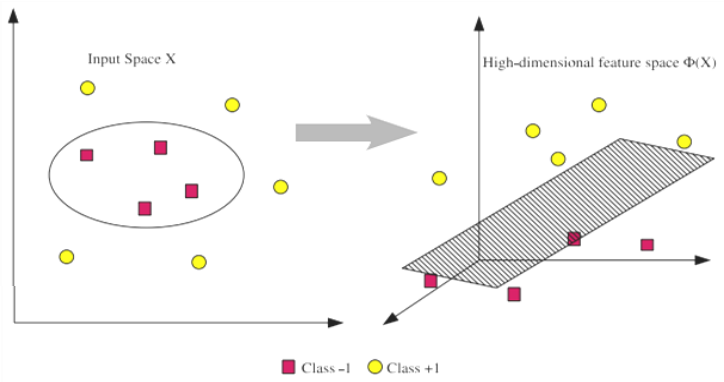


Gambar 2. *Hyperplane* terbentuk

*Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin* *hyperplane* tersebut. dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan pattern terdekat dari masing- masing *class*. *Pattern* yang paling dekat ini disebut sebagai support *vector*. Garis *solid* pada gambar menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah- tengah kedua *class*, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support* *vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada *SVM*.

### *Kernel* *Trick* dan non linear *SVM*

Pada umumnya masalah dalam domain dunia nyata (*real* *world* *problem*) jarang yang bersifat linear *separable*, kebanyakan bersifat non linear. Untuk menyelesaikan problem non linear, *SVM* dimodifikasi dengan memasukkan fungsi *Kernel*. Dalam non linear *SVM*, pertama- tama data x dipetakan oleh fungsi Φ(x) ke ruang vektor yang berdimensi lebih tinggi. Pada ruang vektor yang baru ini, *hyperplane* yang memisahkan kedua class tersebut dapat dikonstruksikan. Hal ini sejalan dengan teori *Cover* yang menyatakan “*Jika suatu transformasi bersifat non linear dan dimensi dari feature space cukup tinggi, maka data pada input space dapat dipetakan ke feature space yang baru, dimana patternpattern tersebut pada probabilitas tinggi dapat dipisahkan secara linear*”.



Gambar 2. Pemetaan input space

Ilustrasi dari konsep ini dapat dilihat pada gambar. Pada gambar di atas sisi kiri diperlihatkan data pada *class* kuning dan data pada class merah yang berada pada *input* *space* berdimensi dua tidak dapat dipisahkan secara linear. Selanjutnya gambar menunjukkan bahwa fungsi Φ memetakan tiap data pada *input* *space* tersebut ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), dimana kedua *class* dapat dipisahkan secara linear oleh sebuah *hyperplane*. Notasi matematika dari mapping ini adalah sebagai berikut.

|  |  |
| --- | --- |
|  | (1) |

Pemetaan ini dilakukan dengan menjaga topologi data, dalam artian dua data yang berjarak dekat pada input space akan berjarak dekat juga pada *feature* *space*, sebaliknya dua data yang berjarak jauh pada input space akan juga berjarak jauh pada *feature* *space*.

Proses pembelajaran pada *SVM* dalam menemukan titik- titik *support* *vector*, hanya bergantung pada *dot* *product* dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu :

|  |  |
| --- | --- |
|  | (2) |

Karena umumnya transformasi Φ ini tidak diketahui, dan sangat sulit untuk dipahami secara mudah, maka perhitungan *dot* *product* tersebut sesuai teori *Mercer* dapat digantikan dengan fungsi *kernel* yang mendefinisikan secara implicit transformasi Φ. Hal ini disebut sebagai *Kernel* *Trick*, yang dirumuskan:

|  |  |
| --- | --- |
|  | (3) |

*Kernel* *trick* memberikan berbagai kemudahan, karena dalam proses pembelajaran *SVM*, untuk menentukan *support* *vector*, kita hanya cukup mengetahui fungsi *kernel* yang dipakai, dan tidak perlu mengetahui wujud dari fungsi non linear Φ. Berbagai jenis fungsi kernel dikenal, sebagaimana dirangkumkan pada tabel 2.1.

Tabel 2. *Kernel* yang Umum Digunakan

|  |  |  |
| --- | --- | --- |
| **Jenis *Kernel*** | **Definisi** | |
| *Polynomial* |  | (4) |
| *Gaussian RBF* |  | (5) |
| *Sigmoid* |  | (6) |
| *Linear* |  | (7) |

Selanjutnya hasil klasifikasi dari data xr diperoleh dari persamaan berikut.

|  |  |
| --- | --- |
|  | (8)  (9)  (10) |

*Support* *vector* pada persamaan di atas dimaksudkan dengan subset dari training set yang terpilih sebagai *support* *vector*, dengan kata lain data xi yang berkorespondensi pada αi ≥ 0.

### Gaussian Kernel

Penggunaan kernel merupakan salah satu faktor dalam keberhasilan dari banyaknya algoritma klasifikasi untuk permukaan non linear. *Gaussian* *Kernel* adalah pilihan *kernel* yang menjanjikan. *Kernel* ini secara non linear memetakan sampel ke dalam ruang dimensi yang lebih tinggi, sehingga tidak seperti *kernel* linear, *kernel* ini dapt menangani kasus ketika hubungan antara label kelas dan atributnya tidak linear. Alasan kedua adalah pada *kernel* *Gaussian* *Kernel*, kompleksitas *hyperparameter*-nya lebih sedikit dibandingkan dengan *kernel* non linear lain seperti kernel polinomial dengan persamaan:

|  |  |
| --- | --- |
|  | (11) |

Dimana niali γ ≥ 0 adalah sebuah parameter yang mengontrol besarnya fleksibilitas persamaan *Gaussian* ini. Dapat dilihat bahwa data yang bersifat linear dapat di kernelisasikan, dengan syarat bahwa independensi data hanya menggunakan *dot* *products*.

## K-Means Clustering

*K-Means* merupakan metode pengklasteran secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Dengan *partitioning* secara iteratif, *K-Means* mampu meminimalkan rata-rata jarak setiap data ke klasternya. Dalam algoritam *K-Means*, setiap data harus termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke *cluster* yang lain. Pada dasarnya penggunaan algoritma *K-Means* dalam melakukan proses *clsutering* tergantung dari data yang ada dan ada konklusi yang ingin dicapai.

Algoritma *K-Means* pada awalnya mengambil sebagian dari banyaknya komponene dari populasi untuk dijadikan pusat *cluster* awal. Pada step ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap *cluster.* Posisi pusat *cluster* akan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap *cluster* dan terakhir akan terbentuk posisi *cluster* baru.

Algoritma *K-Means* pada dasarnya melakukan dua proses yakni proses pendeteksian lokasi pusat *cluster* dan proses pencarian anggota dari tiap-tiap *cluster.* Proses *clsutering* dimulai dengan mengidentifikasi data yang akan di*cluster* Xij (i=1,....,n; j=1,...,m) dengan n adalah jumlah data yang akan di*cluster* dan m jumlah variabel. Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas, Ckj (k=1,....,n; j=1,...,m). Kemudian dihitung jarak antara setiap data dengan pusat *cluster* ke-k (ck), diberi nama (dik), dapat digunakan formula *euclidean*. Suatu data akan menjadi anggota dari *cluster* ke-k bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lain. Proses dasar *algoritma K-Means* antara lain

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk. Tetapkan pusat *cluster.*
2. Hitung jarak setiap data ke pusat *cluster* menggunakan persamaan.

(2.1)

1. Kelompokkan data ke dalam *cluster* yang dengan jarak yang paling pendek menggunakan persamaan.

(2.2)

1. Hitung pusat cluster yang baru menggunakan persamaan

(2.3)

Dimana

(2.4)

P = banyaknya anggota *cluster* ke k.

1. Ulangi langkah b sampai d hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain [4].

## PHP

PHP merupakan suatu bahasa *scripting* khususnya digunakan untuk *web development.* Karena sifatnya yang *server side scripting,* maka untuk menjalankan PHP harus menggunakan *web server*. Pada umumnya, PHP lebih banyak digunakan bersama *file* HTML. Dengan menggunakan PHP, *website* yang dibuat akan menjadi *website powerful* yang dinamis dengan disertai databasenya. Selain itu, PHP juga dapat digunakan untuk berbagai macam *platform* sistem operasi dan mendukung berbagai macam *web server* [5].

# BAB III. METODOLOGI PENELITIAN

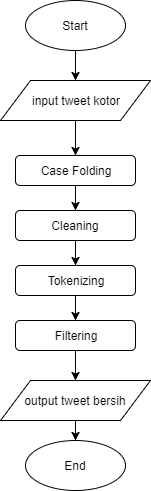
## Metode Pengumpulan Data

Pengambilan data dilakukan berdasarkan query atas term objek pada aplikasi yang terhubung menggunakan Twitter API. Hasil query berupa tweet kotor yang kemudian dijadikan input pada sistem untuk diolah lebih lanjut.

## Metode Pengolahan Data

### *Preprocessing*

Dalam penelitian ini, *preprocessing* dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten sehingga keluaran dari klasifikasi memiliki keakurasian yang tinggi. Tahapan dalam *preprocessing* ditunjukkan oleh flowchart pada Gambar 3.1.



Gambar 3.1 Tahapan Preprocessing

Berdasarkan pada Gambar 3.1 dijelaskan bahwa hasil *tweet* kotor yang didapat dari *Twitter* API diolah menggunakan *text preprocessing* guna mendapatkan *tweet* bersih untuk dilakukan proses klasifikasi. Tahapan pada *text preprocessing* yang dilakukan adalah:

1. *Case folding* dengan mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima.

Contoh : Mandi pakai LUX menambah wangi segar !!

Menjadi : mandi pakai lux menambah wangi segar

1. *Cleaning* adalah tahap dimana karakter selain huruf dihilangkan dan dianggap *delimiter*, dan menghapus juga *URL*, *mention* dan *hastag*.

Contoh :@x pakai LUX menambah wangi !!https://path.com/p/3pB4Qs

Menjadi : pakai LUX menambah wangi segar

1. Tahap *tokenizing / parsing* adalah tahap pemotongan string *input* berdasarkan tiap kata yang menyusunnya.

Contoh : mandi pakai lux menambah wangi segar

Menjadi : mandi | pakai | lux | menambah | wangi | segar

1. *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang. Contoh stopwords adalah “yang”, “dan”, “di”, “dari”, “dengan” dan seterusnya.

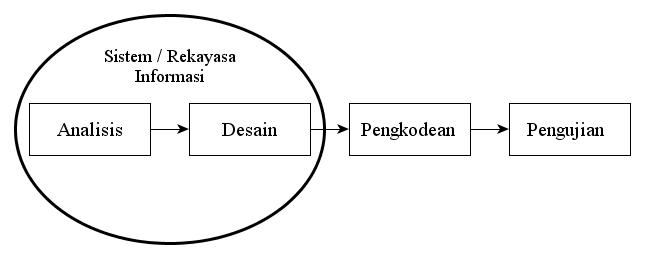
Contoh : mandi | pakai | lux | menambah | wangi | segar | Senang

Menjadi : mandi | pakai | lux | wangi | segar | Senang

## Metode Pengembangan Sistem

Pada penelitian ini, penulis menggunakan metode *Waterfall.* Metode *Waterfall* merupakan suatu proses pengembangan perangkat lunak berurutan, dimana kemajuan dipandang sebagai terus mengalir kebawah (seperti air terjun) melewati fase-fase perencanaan, pemodelan, implementasi (konstruksi) dan pengujian [6].

Metode *Waterfall* sering juga disebut model sekuensial linier *(sequential linier)* atau alur hidup klasik *(classic life cycle)*. Metode ini menyediakan pendekatan alur hidup perangkat secara sekuensial atau terurut dimulai dari analisis, desain, pengodean, pengujian dan tahap pendukung. Gambar 3.2 merupakan pengembangan perangkat lunak metode *Waterfall*.



Gambar 3.2 Metode Waterfall

Dalam pengembangan sistem menggunakan metode *Waterfall* memiliki beberapa tahapan yang runtut, yaitu sebagai berikut.

1. Analisis Kebutuhan

Tahap ini merupakan tahap awal dalam pembentukan aplikasi untuk mengimplementasikan sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means* dengan menetapkan berbagai kebutuhan yang diperlukan. Dalam menganalisa kebutuhan perangkat, sistem ataupun data yang diperlukan diantaranya sebagai berikut.

1. Laptop dengan spesifikasi
   * Processor : Intel Core i5
   * RAM : 2.00 GB
   * System Type : 64-bit Operating System
   * Harddisk : 500 GB
   * Display : 1366\*768 *pixel*
   * Sistem Operasi : Windows 10
2. XAMPP v3.2.2
3. PHP v.7.1.2
4. *Browser* Google Chrome
5. Desain

Tahapan ini merupakan tahap desain aplikasi dalam mengimplementasikan sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means*.

1. Implementasi

Tahapan ini merupakan tahapan dimana setelah melakukan desain program yang ditejemahkan kedalam kode program dengan menggunakan bahasa pemrograman PHP. Pada tahapan ini nantinya akan dilakukan pengujian terhadap sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means*.

1. Pengujian

Tahapan ini merupakan tahapan dimana program yang sudah selesai diuji dari segi fungsionalitas dan memastikan bahwa semua bagian sudah diuji. Hal ini dilakukanm untuk meminimalisir kesalahan dan memastikan keluaran atau *output* yang dihasilkan sesuai dengan yang diinginkan

1. Pemeliharaan

Tahapan ini merupakan tahapan terakhir dimana setelah semua proses tahapan dilakukan pengoperasian aplikasi dan akan dilakukan proses pemeliharaan seperti penyesuaian dan perubahan. Karena tidak menutup kemungkinan sebuah perangkat lunak mengalami perubahan ketika sudah digunakan oleh *user*. Hal ini terjadi bisa karena adanya kesalahan yang tidak terdeteksi pada saat pengujian dan perangkat lunak harus beradaptasi dengan lingkungan baru. Sedangkan pemeliharaan ini berguna apabila perangkat lunak yang dibangun diperlukan perubahan-perubahan sesuai dengan keinginan *user* namun tidak membuat perangkat lunak baru hanya mengulangi proses pengembangan.

## Data

Komponen Data yang digunakan pada penelitian ini adalah *tweet* berbahasa Indonesia yang ditulis oleh para pengguna layanan *Twitter*. *Tweet* yang dikumpulkan merupakan *tweet* yang berisi tentang 10 pantai yang ada di Indonesia. Penulis mengumpulkan masing-masing 50 *tweet* untuk tiap pantai sehingga total *tweet* yang digunakan sebagai data berjumlah 500.

Berdasarkan pengambilan data yang dilakukan maka berikut ini merupakan beberapa contoh hasil *tweet* yang digunakan.

Tabel 3. *Tweet* Pantai Anyer

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Para pejuang yg kurang piknik (with Veri, Langgeng, and 4 others at Pantai Florida Anyer) [pic] — <https://path.com/p/3jendF> |
| 2. | Pantai Anyer, baginda? Tampan banget sih, hamba suka deh. |
| 3. | Paparazzi @ Anyer Beach - Wisata Pantai Anyer https://www.instagram.com/p/BSOGzG2DqMc/ |
| 4. | Refresh otak penat wkwk (at Pantai Anyer (Anyer Beach)) — <https://path.com/p/6kQmX> |
| 5. | Mencari seseorang yg "hilang" kmrn siang (at Pantai legon prima anyer) — <https://path.com/p/4CDTzz> |

Pada Tabel 3.1 menampilkan hasil *tweet* yang berisi tentang Pantai Anyer

Tabel 3. Tweet Pantai Balekambang

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Pagi di Pantai Balekambang, min @infomalang Cerah #infocuaca :) (at Pantai Balekambang) [pic] ? https://path.com/p/2kFKvC |
| 2. | Pantai Balekambang adalah salahsatu keindahan wisatanya yg keren banget #GunungKawiGegerDiMalang pic.twitter.com/8WwEplxySV |
| 3. | 8.000 Umat Hindu Tumplek Blek di Pantai Balekambang http://dlvr.it/NjsYTDÿ #MalangRaya #malangvoice |
| 4. | Cuacanya cerah.. dah berapa tahun ga kesini yes (at Pantai Balekambang) ? https://path.com/p/3ksTUI |
| 5. | Minggu konvoi lagi rencana konvoi diarahkan ke pantai Balekambang sekalian merayakan HUT Bupati Malang.. \*6 |

Pada Tabel 3.2 menampilkan hasil *tweet* yang berisi tentang Pantai Balekambang

Tabel 3. Tweet Pantai Bantol

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Pantai Bantol, Surga Lobster di Malang Selatan https://www.shortir.com/pantai-bantol-surga-lobster-di-malang-selatan …pic.twitter.com/ZTgfvKmJ75 |
| 2. | Bantol, Pantai dengan Sejuta Keindahan di Malang Selatan <http://ift.tt/1RCoQkF> |
| 3. | Pantai bantol malang selatan siiiip deh pic.twitter.com/L2ik5Ia7sY |
| 4. | Meskipun akses lumayan sulit, tapi suguhan alamnya mempesona. Pantai Bantol di Ds Banjarejo, Kec Donomulyo Kab Malang #wisatamalang #pantai |
| 5. | @pakdekarwo1950 @ainissfm @e100ss pantai bantol akses jalanya blom aspal, jalan lintas selatan juga blom selesai, |

Pada Tabel 3.3 menampilkan hasil *tweet* yang berisi tentang Pantai Bantol

Tabel 3. Tweet Pantai Goa Cina

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Motret Cantik! #iPhone7 #JOErney #iphoneonly @ Pantai Goa Cina https://www.instagram.com/p/BSgg0vmldY1/ |
| 2. | Alhamdulillah Keturutan Vitamin Sea-nya #vitaminsea @ Pantai Goa Cina https://www.instagram.com/p/BSMrj0yAPEt/ |
| 3. | Rileksss , @ Pantai goa cina - malang https://www.instagram.com/p/BQ1wTxXhAJy/ |
| 4. | akibat miskin sinyal (at Pantai Goa Cina) |
| 5. | Bersyukur atas apa yang diciptakan tuhan #dwiright #pantaigoacina…https://www.instagram.com/p/BSjBA-JjtIQVgyXBRUPFM3te1Qz9i4V2L9CZfc0/ … |

Pada Tabel 3.4 menampilkan hasil *tweet* yang berisi tentang Pantai Goa Cina

Tabel 3. Tweet Pantai Klayar

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Indonesia itu indah lhooo... (at Pantai Klayar) [pic] |
| 2. | Sudah hitam gausah setengah setengah wkw (at Pantai Klayar) — https://path.com/p/PxKFF |
| 3. | Pesona Keindahan Pantai Klayar Pacitan http://masyudha.com/pesona-keindahan-pantai-klayar-pacitan/ …pic.twitter.com/JtyhkrtQ8q |
| 4. | Pantai Klayar; Primadona Tersembunyi di Pacitan Jawa Timur - Traveling |
| 5. | Ke pacitan jg kak. Ke pantai klayar. Bagus bgttt, gak bakalan nyesel deh https://twitter.com/PerempuanThicka/status/843485897777532928 … |

Pada Tabel 3.5 menampilkan hasil *tweet* yang berisi tentang Pantai Klayar

Tabel 3.6 Tweet Pantai Pangandaran

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Holiday yang menyenangkan (at Pantai Barat Pangandaran) — https://path.com/p/2OLgyM |
| 2. | Fak this is so beautiful @ Pantai Pangandaran https://www.instagram.com/p/BSiFKIQDaHP/ |
| 3. | Ini di pantai Pangandaran ya? apa di Batukaras? "...." pic.twitter.com/50ybdoEIOF |
| 4. | Persahabatan dibalut keindahan Pantai Pangandaran, Jawa Barat. #IMS\_WISATABAHARI pic.twitter.com/3l8WkX8IKT |
| 5. | Panasnya gak kuat aslian :( (at Pantai Barat Pangandaran) — https://path.com/p/4brZkp |

Pada Tabel 3.6 menampilkan hasil *tweet* yang berisi tentang Pantai Pangandaran

Tabel 3.7 Tweet Pantai Plengkung

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Ombak pantai Plengkung, pulau G-Land, Banyuwangi, termasuk 7 terbaik dunia. Ombak bisa setinggi 6m, panjang 2km dalam 7 susun gelomb #dfact |
| 2. | provinsi ini memiliki pantai yang ombak nya merupakan salah satu yang terbaik didunia yaitu pantai plengkung #GunungKawiMovie |
| 3. | Freeee ~~ (at Pantai Plengkung / G-Land) — https://path.com/p/3OmDsZ |
| 4. | klo profil pict iya kawah ijen, klo sampul itu di pantai plengkung mz, masuk ke alas purwo |
| 5. | Pantai Plengkung menjadi slh 1 "surga" bagi peselancar krn Pantai Plengkung mempunyai gulungan ombak hingga setingi 6 meter #JatimTersenyum |

Pada Tabel 3.7 menampilkan hasil *tweet* yang berisi tentang Pantai Plengkung

Tabel 3.8 Tweet Pantai Popoh

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Gak indah saiki,. (with Lusi Kartika at Pantai Indah Popoh) — https://path.com/p/1qipHw |
| 2. | Akhirnya lihat laut juga :3 (at Pantai Indah Popoh) — https://path.com/p/1nR0a3 |
| 3. | Indahnya pantai popoh tulungagung jawa timur....pantai selatan lautan lepas pic.twitter.com/0ZctA30OK4 |
| 4. | Sejumlah Nelayan Pantai Popoh Waspadai Gelombang Tinggi Di Tengah Laut. / …pic.twitter.com/3uYwVNU84L |
| 5. | indahnya pantai popoh tulung agung http://fb.me/7WushIDtr |

Pada Tabel 3.8 menampilkan hasil *tweet* yang berisi tentang Pantai Popoh

Tabel 3.9 Tweet Pantai Sawarna

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Surga Tersembunyi di pantai Sawarna http://banten.travel/eksotisme-pantai-sawarna-yang-sayang-terlewatkan/ …pic.twitter.com/QaMtNr4wTz |
| 2. | Lokasi Semen Merah Putih ini dekat sekali dg wisata Pantai Sawarna, Bayah, Lebak, Banten. Aktifitasnya tak indahkan lingkungan. Cc @jokowi pic.twitter.com/G8gRDQfMEC |
| 3. | Saya menambahkan video ke playlist @YouTube http://ln.is/www.youtube.com/OHQ81 … PKS TV - Inilah Keindahan Potensi Wisata Pantai Sawarna Yang Belum |
| 4. | Sisi lain dari ciptaan Tuhan, ternyata masih tetap menunjukkan keindahan. @ Pantai Sawarna Lebak… https://www.instagram.com/p/BR5lrx4DYug/ |
| 5. | Bro&SisTEA, coba liburan ke objek wisata Desa Sawarna, yuk. Desa pesisir di Kabupaten Lebak, Banten, ini punya pantai yg masih #asli & indahpic.twitter.com/RIBg8WQSsz |

Pada Tabel 3.9 menampilkan hasil *tweet* yang berisi tentang Pantai Sawarna

Tabel 3.10 Tweet Pulau Sempu

|  |  |
| --- | --- |
| **No** | **Tweet** |
| 1. | Kamu itu pulau rahasia ku @ Sempu Island https://www.instagram.com/p/BRUGJ1rAxDI/ |
| 2. | Semua pasti butuh proses untuk akhir yang indah. #latepos #sempu… (w/ Reza & Indra at Cagar Alam Pulau Sempu) — https://path.com/p/E8kuM |
| 3. | SURGA ITU BERNAMA SEGARA ANAKAN Tak asing bagi gue mendengar nama Pulau Sempu. Pulau yang eksis… https://www.instagram.com/p/BRk4vY7BmZk/ |
| 4. | Paling enak, abis berenang di laut minum es degan..mak nyess (at Pulau Sempu) [pic] — https://path.com/p/2Wspes |
| 5. | Enak ya jadi Adam & Hawa, gapernah bahas Mantan. @ Pulau Sempu Malang https://www.instagram.com/p/BRull4eh7et/ |

Pada Tabel 3.10 menampilkan hasil *tweet* yang berisi tentang Pulau Sempu

## Metode Pengujian

Pengujian yang dilakukan untuk menguji sistem dilakukan dengan 2 langkah pengujian, yaitu pengujian fungsionalitas dan pengujian akurasi sistem. Pengujian bertujuan untuk memastikan bahwa sistem yang dibangun telah berjalan sesuai dengan yang digunakan.

1. Pengujian Fungsionalitas

Uji coba fungsionalitas dikatakan berhasil apabila fungsi yang ada pada sistem sesuai dengan yang diharapkan pengguna.

1. Pengujian Akurasi Sistem

Dalam pengujian ini, dilakukan untuk mengetahui keakuratan aplikasi dalam menampilkan data hasil pencarian oleh pengguna.

# BAB IV. ANALISIS DAN PERANCANGAN

## Analisa Sistem

Analisa sistem merupakan suatu penjabaran mengenai komponen-komponen penyusunan sistem dalam penelitian ini baik perangkat lunak maupun perangkat keras. Serta gambaran umum sistem yang akan berjalan.

### Kebutuhan Perangkat Lunak

Untuk dapat melakukan perancangan dan menjalankan sistem Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means perlu memperhatikan kebutuhan perangkat lunak yaitu diantaranya sebagai berikut.

Sistem Operasi Windows 10

PHP sebagai bahasa pemrograman

*Web server* menggunakan *Apache* milik XAMPP v.3.2.2

*Database* menggunakan MySQL

Browser *Google Chrome* untuk menjalankan aplikasi

### Kebutuhan Perangkat Keras

Spesifikasi perangkat keras yang digunakan untuk melakukan pengujian sistem Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means adalah sebagai berikut.

* + Processor : Intel Core i5
  + RAM : 2.00 GB
  + System Type : 64-bit Operating System
  + Harddisk : 500 GB
  + Display : 1366\*768 *pixel*
  + OS : Windows 10

## Analisa Algoritma

Analisa algoritma merupakan suatu penjabaran mengenai algoritma-algortima penyusunan sistem dalam penelitian ini.

### Klasifikasi dengan *Support Vector Machine*

Proses klasifikasi menggunakan SVM dimulai dengan mengubah text menjadi data vector. Vector dalam penelitian ini memiliki dua komponen yaitu dimensi (word id) dan bobot. Bobot ini sering dikombinasikan ke dalam sebuah nilai tf-idf (*Term Frequency*).

*Term Frequency* adalah frekuensi dari kemunculan sebuah *term* dalam suatu dokumen. *Term* dapat berupa kata, frase atau unit hasil *indexing* lainnya yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata tersebut diberikan sebuah indikator. Berikut merupakan contoh pengolahan text hingga menjadi *term frequency* pada Tabel 4.1 dan 4.2.

Tabel 4. Contoh Daftar Kalimat

|  |  |
| --- | --- |
| **DAFTAR KALIMAT** | |
| **1** | Saya suka balekambang |
| **2** | Pantai popoh bagus |
| **3** | Pantai balekambang kabupaten malang |
| **4** | Jalan popoh buruk |
| **5** | Pantai anyer menarik |

Tabel 4. Contoh *Term Frequency*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | d1 | d2 | d3 | d4 | d5 | df | idf | tf-idf | | | | |
| d1 | d2 | d3 | d4 | d5 |
| saya | 1 | 0 | 0 | 0 | 0 | 1 | 0.6990 | 0.6990 | 0 | 0 | 0 | 0 |
| suka | 1 | 0 | 0 | 0 | 0 | 1 | 0.6990 | 0.6990 | 0 | 0 | 0 | 0 |
| balekambang | 1 | 0 | 1 | 0 | 0 | 2 | 0.3979 | 0.3979 | 0 | 0.3979 | 0 | 0 |
| pantai | 0 | 1 | 1 | 0 | 1 | 3 | 0.2218 | 0 | 0.2218 | 0.2218 | 0 | 0.2218 |
| popoh | 0 | 1 | 0 | 1 | 0 | 2 | 0.3979 | 0 | 0.3979 | 0 | 0.3979 | 0 |
| bagus | 0 | 1 | 0 | 0 | 0 | 1 | 0.6990 | 0 | 0.6990 | 0 | 0 | 0 |
| kabupaten | 0 | 0 | 1 | 0 | 0 | 1 | 0.6990 | 0 | 0 | 0.6990 | 0 | 0 |
| malang | 0 | 0 | 1 | 0 | 0 | 1 | 0.6990 | 0 | 0 | 0.6990 | 0 | 0 |
| jalan | 0 | 0 | 0 | 1 | 0 | 1 | 0.6990 | 0 | 0 | 0 | 0.6990 | 0 |
| buruk | 0 | 0 | 0 | 1 | 0 | 1 | 0.6990 | 0 | 0 | 0 | 0.6990 | 0 |
| anyer | 0 | 0 | 0 | 0 | 1 | 1 | 0.6990 | 0 | 0 | 0 | 0 | 0.6990 |
| menarik | 0 | 0 | 0 | 0 | 1 | 1 | 0.6990 | 0 | 0 | 0 | 0 | 0.6990 |

Selanjutnya adalah klasifikasi menggunakan SVM dimana dari keseluruhan dataset diambil secara acak oleh sistem sebagai training sebanyak 70 persen. Setelah algoritma berhasil melakukan training pada dataset, proses testing dapat dilakukan. Dataset yang digunakan sebagai testing berjumlah 30 persen dari dataset seluruhnya. Proses *training* digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai pedoman dalam klasifikasi dengan data *testing* atau data yang berbeda. Berikut adalah algoritma klasifikasi sentimen menggunakan *SVM*.

1. Proses Training
2. Menentukan *kernel* yang digunakan

*SVM* memiliki beberapa pilihan *kernel* yang dapat digunakan sesuai dengan kebutuhan data dan klasifikasi. Pilihan *kernel* akan mendukung akurasi dari hasil klasifikasi yang dilakukan. Berikut adalah beberapa pilihan *kernel* yang disediakan oleh *SVM*.

Tabel 4.3 Kernel yang Umum Digunakan

|  |  |  |
| --- | --- | --- |
| **Jenis *Kernel*** | **Definisi** | |
| *Polynomial* |  |  |
| *Gaussian* *RBF* |  |  |
| *Sigmoid* |  |  |
| *Linear* |  |  |

1. Menghitung matriks *kernel* K dengan *kernel* *Gaussian* *RBF*

*Kernel* *Gaussian* *RBF*

Mengubah fungsi menjadi

Fungsi matriks *kernel* K

Menghitung matriks *kernel* K

Untuk memudahkan perhitungan, persamaan diubah menjadi

1. Menghitung nilai melalui persamaan *Lagrange*
2. Menghitung nilai *hyperplane*

Menghitung nilai w

Menghitung nilai b

1. Proses *Testing*
   1. Melakukan Klasifikasi

Klasifikasi dilakukan dengan menggunakan pedoman persamaan *hyperplane* yang telah didapatkan. Klasifikasi dilakukan dengan cara substistusi nilai data testing ke dalam persamaan *hyperplane*. Apabila hasil lebih dari atau kurang dari persamaan, data akan masuk ke dalam kategori tertentu.

### Klasterisasi dengan *K-Means*

Contoh klasterisasi dengan *K-Means* penulis dapat bersumber dari jurnal Fenty Eka M.Agustin dengan studi kasus SMP Negeri 101 Jakarta. Berikut adalah langkah langkah perhitungan dengan metode *K-Means* [4]:

Percobaan dilakukan dengan menggunakan parameter-parameter berikut:

Jumlah Cluster : 4

Jumlah Data : 12

Jumlah Atribut : 4 (IPA, MTK, B. Indonesia, dan B.Inggris)

Tabel 4.4 Tabel Siswa dan Nilai

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NIS | NAMA | RATA-RATA | | | |
| BIND | BING | IPA | MTK |
| 10943 | ARIS APRISALY | 8.1 | 7.3 | 7.76 | 7.42 |
| 0944 | ARMANDA D | 8.17 | 7.33 | 7.64 | 7.36 |
| 0945 | AYU INDAH L | 8.5 | 7.45 | 7.88 | 7.59 |
| 10904 | AZIZAH LARAS | 8.65 | 8.42 | 8.18 | 8.13 |
| 10835 | CINDY AULIA | 8.08 | 7.37 | 8.02 | 7.54 |
| 10800 | DENY OKTA | 7.27 | 8 | 8 | 8 |
| 11012 | DHEA AMEILYA | 8.23 | 7.33 | 7.28 | 7.69 |
| 11092 | ESA SALSA Z | 8.52 | 7.82 | 8.22 | 7.81 |
| 10913 | FARRAS JIHAN A | 8.43 | 7.85 | 7.9 | 7.87 |
| 10876 | FAUZAN HANIF | 8.15 | 7.58 | 7.88 | 7.47 |
| 11020 | GAYUSMAN P | 8.12 | 7.42 | 7.55 | 7.35 |

Iterasi ke-1

1. Penentuan pusat awal *cluster*. Untuk penentuan awal diasumsikan dengan menggunakan nilai terendah pada masing-masing mata pelajaran

Tabel 4.5 Pusat Awal *Cluster*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data 6 | 7.28 | 8 | 8 | 8 |
| Data 1 | 8.1 | 7.3 | 7.758 | 7.42 |
| Data 7 | 8.23 | 7.33 | 7.28 | 7.69 |
| Data 12 | 8.3 | 7.42 | 7.658 | 7.34 |

1. Perhitungan jarak pusat *cluster* Untuk mengukur jarak antara data dengan pusat *cluster* digunakan Euclidian *distance*, kemudian akan didapatkan matriks jarak sebagai berikut :

Rumus *Euclidian distance* :

Cij : Pusat Cluster

Ckj : Data

Sebagai contoh, perhitungan jarak dari data ke-1 terhadap pusat cluster adalah:

C1=

C2=

C3=

C4=

Dan seterusnya dilanjutkan untuk data ke 2, ...N Kemudian akan didapatkan matrix jarak sebagai berikut :

D1=

Iterasi ke-1

Tabel 4.6 Jarak dengan Pusat *Centroid*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | C1 | C2 | C3 | C4 |
| Aris Aprisaly | 1,26 | 0 | 0,57 | 0,26 |
| Armanda Dwicahya | 1,34 | 0,15 | 0,5 | 0,16 |
| Atu Indah Lestari | 1,41 | 0,48 | 0,67 | 0,39 |
| Azizah Laraswati | 1,46 | 1,5 | 1,53 | 1,42 |
| Cindy Aulia | 1,13 | 0,3 | 0,77 | 0,47 |
| Deny Oktaviano | 0 | 1,26 | 1,41 | 1,4 |
| Dhea Ameila | 1,41 | 0,57 | 0 | 0,53 |
| Esa salsa zenita | 1,29 | 0,9 | 1,1 | 0,86 |
| Faras Jihan Afifah | 1,18 | 0,8 | 0,85 | 0,73 |
| Fauzan Hanif | 1,12 | 0,31 | 0,69 | 0,34 |
| Gayusman Putrata | 1,3 | 0,25 | 0,46 | 0,21 |
| Govinda Pangestu | 1,4 | 0,26 | 0,53 | 0 |

Setiap kolom pada matrix menunjukkan nilai jarak data terhadap pusat cluster. Baris pertama pada matrix menunjukkan nilai jarak data terhadap titik pusat cluster pertama, baris kedua pada matrix menunjukkan nilai jarak data terhadap titik pusat cluster kedua dan seterusnya.

1. Pengelompokkan data Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat. Berikut ini akan ditampilkan data matriks pengelompokkan group, nilai 1 berarti data tersebut berada dalam group.

Tabel 4.7 Pengelompokkan Data

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| G1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| C1 |  |  |  |  |  | 1 |  |  |  |  |  |  |
| C2 | 1 | 1 |  |  | 1 |  |  |  | 1 | 1 |  |  |
| C3 |  |  |  |  |  |  | 1 |  |  |  |  |  |
| C4 |  |  | 1 | 1 |  |  |  | 1 |  |  | 1 | 1 |

Keterangan :

* Jika nilai D1 lebih dekat jaraknya dari nilai hasil perhitungan data *cluster* ke-1 (C1) atau data *cluster* ke-2 (C2) atau data *cluster* ke-3 (C3) atau atau data *cluster* ke-4 (C4) maka G1 bernilai 0 dan tidak termasuk grup atau kelompok *cluster* baru.
* Jika nilai D1 lebih jauh jaraknya dari nilai hasil perhitungan data *cluster* ke-1 (C1) atau data *cluster* ke-2 (C2) atau data *cluster* ke-3 (C3) atau atau data *cluster* ke-4 (C4) maka G1 bernilai 1 dan termasuk grup atau kelompok *cluster* baru.
* Berdasarkan matriks data perhitungan, didapat :

C1 data 6 C2 data 1,2,5,9,10

C3 data 7 C4 data 3,4,8,11,12

1. Penentuan pusat cluster baru Setelah diketahui anggota tiap-tiap cluster kemudian pusat cluster baru dihitung berdasarkan data anggota tiap-tiap cluster sesuai dengan rumus pusat anggota cluster. Sehingga didapatkan perhitungan sebagai berikut :

Karena C1 memiliki 1 anggota yaitu data 6 maka perhitungan *cluster* baru menjadi:

C1 =

Lakukan langkah sama pada C2, C3, dan C4. Nilai *cluster* yang baru adalah:

Cluster baru untuk iterasi ke-2

Tabel 4.8 Centroid Baru Iterasi ke 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C1 | 7.27 | 8 | 8 | 8 |
| C2 | 8.19 | 7.49 | 7.84 | 7.53 |
| C3 | 8.23 | 7.33 | 7.28 | 7.69 |
| C4 | 8.42 | 7.7 | 7.9 | 7.64 |

Iterasi ke-2

1. Ulangi langkah ke 2 (dua) hingga posisi data tidak mengalami perubahan

Tabel 4.9 Jarak dengan Pusat *Centroid* Iterasi ke 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| D2 | C1 | C2 | C3 | C4 |
| 1 | 1.256 | 0.248 | 0.567 | 0.58 |
| 2 | 1.337 | 0.303 | 0.495 | 0.59 |
| 3 | 1.413 | 0.303 | 0.672 | 0.27 |
| 4 | 1.459 | 1.249 | 1.536 | 0.94 |
| 5 | 1.128 | 0.239 | 0.767 | 0.5 |
| 6 | 0 | 1.162 | 1.409 | 1.24 |
| 7 | 1.409 | 0.603 | 0.005 | 0.74 |
| 8 | 1.293 | 0.662 | 1.1 | 0.39 |
| 9 | 1.185 | 0.557 | 0.853 | 0.27 |
| 10 | 1.117 | 0.127 | 0.689 | 0.34 |
| 11 | 1.297 | 0.354 | 0.457 | 0.61 |
| 12 | 1.397 | 0.293 | 0.526 | 0.49 |

Tabel 4.10 Pengelompokkan Data Iterasi ke 2

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| G2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| C1 |  |  |  |  |  | 1 |  |  |  |  |  |  |
| C2 | 1 | 1 |  |  | 1 |  |  |  |  | 1 | 1 | 1 |
| C3 |  |  |  |  |  |  | 1 |  |  |  |  |  |
| C4 |  |  | 1 | 1 |  |  |  | 1 | 1 |  |  |  |

C1 data 6 C2 data 1,2,5,10,11,12

C3 data 7 C4 data 3,4,8,9

Iterasi akan terus dilakukan hingga nilai G pada tiap iterasi memiliki nilai yang sama. Pada saat nilai G yang sama telah terjadi, maka clustering telah mencapai stabil dan konvergen. Dan pada proses penghitungan manual ini, nilai G yang konvergen terjadi iterasi ke-4, nilai G3=G4. G3 dan G4

Tabel 4.11 Hasil *Clustering* Data Mahasiswa

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| C1 |  |  |  |  |  | 1 |  |  |  |  |  |  |
| C2 | 1 | 1 | 1 |  | 1 |  |  |  |  | 1 | 1 | 1 |
| C3 |  |  |  |  |  |  | 1 |  |  |  |  |  |
| C4 |  |  |  | 1 |  |  |  | 1 | 1 |  |  |  |

C1 data 6 C3 data 7

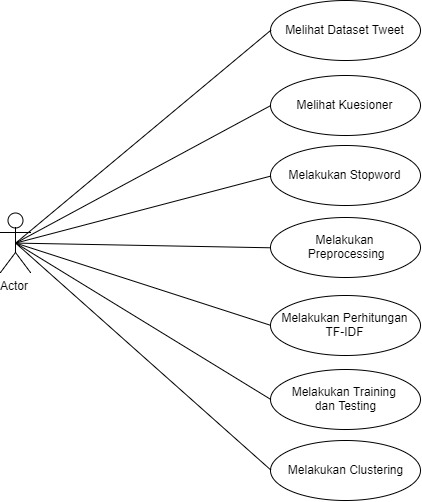
C2 data 1,2,3,5,10,11,12 C4 data 3,4,8,9

## Perancangan Sistem

Perancangan sistem merupakan suatu desain sistem sebagai penggambaran, perencanaan dan pembuatan sketsa dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi. Rancangan akan dibagi menjadi 4 yaitu *use case diagram*, *flowchart* dan perancangan antar muka (*interface*) yang ditampilkan ke dalam bentuk *mockup.*

### Use Case Diagram

*Use* *case* merupakan diagram yang terdiri dari aktor dan proses yang dapat dilakukan oleh aktor tersebut. *Use* *case* dibuat berdasarkan layanan yang dimiliki oleh sistem. *Use* *case* memiliki hak akses untuk melakukan beberapa layanan, yaitu melihat dataset *tweet*, melihat kuesioner, melakukan stopword, melakukan preprocessing, melakukan perhitungan TF-IDF, melakukan training dan testing, dan melakukan clustering



Gambar 4. *Use Case Diagram*

#### **4.3.1.1 Definisi Use Case**

Tabel 4.12 Deskripsi *Use Case*

|  |  |  |
| --- | --- | --- |
| **No** | **Use Case** | **Deskripsi** |
| 1 | Melihat dataset *tweet* | Melihat *tweet* kotor yang telah diambil sebelumnya melalui *Twitter* API |
| 2 | Melihat kuesioner | Melihat kuesioner dari pengguna serta dapat memasukkan data baru |
| 3 | Melakukan *stopword* | Melihat data *stopword* |
| 4 | Melakukan *preprocessing* | Melakukan pembersihan kata-kata dari *tweet* kotor |
| 5 | Melakukan perhitungan *TF-IDF* | Melihat hasil perhitungan *TF-IDF* oleh sistem |
| 6 | Melakukan training dan testing | Membagi 70% data sebagai training dan 30% sebagai testing klasifikasi |
| 7 | Melakukan clustering | Melakukan klasterisasi dari hasil kuesioner setiap pantai |

#### **4.3.1.2 Skenario Use Case**

Skenario use case menggambarkan alur penggunaan sistem dimana setiap scenario digambarkan dari sudut pandang aktor, seseorang atau piranti yang berinteraksi dengan perangkat lunak dalam berbagai cara.

Tabel 4.13 Skenario *Use Case* Melihat Dataset *Tweet*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melihat Dataset *Tweet* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman dataset | Melakukan *load* dataset dari *database* |

Tabel 4.14 Skenario *Use Case* Melihat Kuesioner

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melihat Kuesioner |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman kuesioner | Mengisi data-data kuesioner |
|  | Melihat hasil kuesioner |

Tabel 4.15 Skenario *Use Case* Melakukan *Stopword*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melakukan *Stopword* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman *stopword* | Menampilkan daftar kata-kata *stopword* dari *databse* |
|  | Sistem menghapus kata-kata *stopword* yang ada pada dataset |

Tabel 4.16 Skenario *Use Case* Melakukan *Preprocessing*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melakukan *Preprocessing* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman *preprocessing* | Menampilkan hasil pembersihan kata-kata yang dilakukan sistem |

Tabel 4.17 Skenario *Use Case* Melakukan Perhitungan *TF-IDF*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melakukan Perhitungan *TF-IDF* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman *TF-IDF* | Menampilkan hasil pemecahan setiap kata pada tiap tweet |
|  | Melakukan pembobotan pada setiap kata menggunakan TF-IDF |
| Memilih tombol Simpan | Menyimpan struktur perhitungan T*F-IDF* kedalam *Excel* |

Tabel 4.18 Skenario *Use Case* Melakukan *Training* dan *Testing*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melakukan *Training* dan *Testing* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman *Training* dan *Testing* | Menampilkan 70% dataset yang digunakan sebagai data *training* |
|  | Menampilkan 30% dataset yang digunakan sebagai data *testing* |
|  | Sistem melakukan klasifikasi dari data *testing* |
|  | Menampilkan tingkat akurasi sistem |

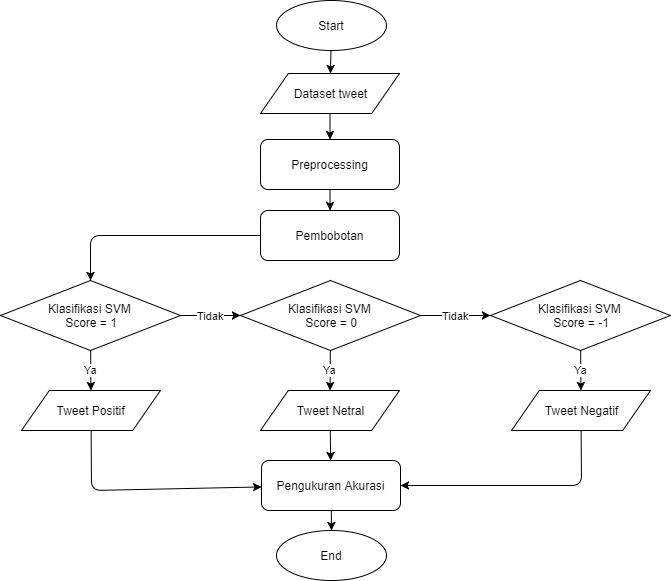
Tabel 4.19 Skenario *Use Case* Melakukan *Clustering*

|  |  |
| --- | --- |
| **Identifikasi** | |
| **Nama** | Melakukan *Clustering* |
| **Aktor** | Pengguna |
| **Skenario Utama** | |
| **Aksi Aktor** | **Reaksi Sistem** |
| Membuka halaman *clustering* | Menampilkan halaman clustering |
| Memasukkan jumlah *cluster* | Melakukan perhitungan sesuai dengan jumlah *cluster* dari pengguna |

### Flowchart

Pada sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means* terdapat *flowchart* yang menjelaskan alur metode *Support Vector Machine (SVM)* dan *K-Means* *Clustering* yang dapat dilihat pada gambar dibawah ini.

* + - 1. **Flowchart Klasifikasi SVM**



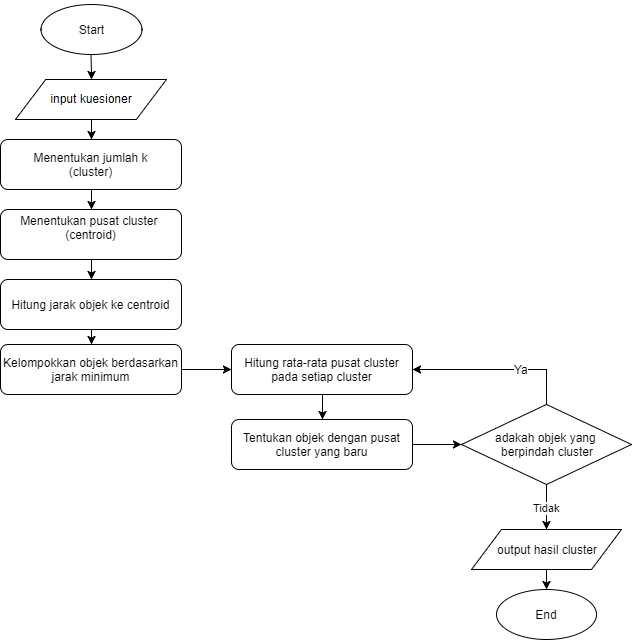
Gambar 4.2 *Flowchart* Klasifikasi *SVM*

Berdasarkan flowchart pada gambar 4.2 menjelaskan bahwa :

SVM adalah metode yang sangat cocok digunakan untuk masalah klasifikasi, tetapi SVM hanya dapat mengklasifikasikan tentang titik dalam ruang. Untuk tujuan ini model ruang *vector* digunakan untuk memberikan setiap kata dalam dokumen sebuah ID (dimensi) dan sebuah bobot berdasarkan seberapa penting keberadaan *term* dalam dokumen (posisi dokumen dalam dimensi itu). SVM mencoba untuk menemukan garis yang terbaik untuk membagi dua kelas, dan kemudian mengklasifikasikan dokumen uji berdasarkan di sisi mana dari garis tersebut akan muncul.

SVM akan memisahkan *vector* berdasarkan garis terbaik yang memiliki margin terbesar diantaranya dan contoh titik pelatihan terdekat di kedua sisinya. Oleh karena itu, *vector* contoh (data t*raining*) berperan besar dalam menentukan margin tersebut adalah yang paling dekat dengan *dividing lines*. Sehingga SVM dapat memberikan keputusan fungsi (kelas atau bukan kelas) untuk classifier.

* + - 1. **Flowchart K-Means Clustering**



Gambar 4.3 Flowchart Clustering K-Means

Berdasarkan *flowchart* pada gambar 4.3 menjelasakan bahwa :

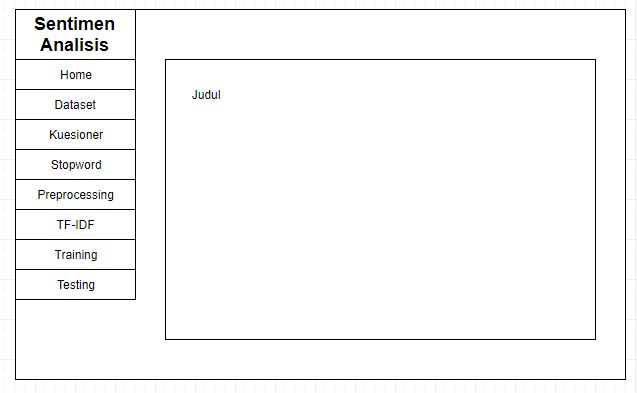
* + - * 1. Memasukkan data kuesioner yang telah diisikan masyarakat
        2. Menentukan jumlah k atau jumlah *cluster*
        3. Menentukan pusat *cluster* sesuai dengan jumlah *cluster*
        4. Htung jarak objek ke pusat *cluster* atau *centroid* menggunakan rumus *euclidean distance*
        5. Mengelompokkan data sesuai *cluster* berdasarkan jarak minimun
        6. Hitung rata-rata setiap *cluster* untuk mendapatkan pusat *cluster* yang baru
        7. Htung jarak objek ke pusat *cluster* atau centroid yang baru dan mengelompokkan berdasarkan jarak minimum
        8. Jika masih ada objek yang berpindah *cluster* maka hitung rata-rata setiap *cluster* untuk mendapatkan pusat *cluster* yang baru
        9. Jika sudah tidak ada objek yang berpindah maka perhitungan selesai dan akan tampil hasil proses *clustering.*

### Mockup

Dalam perancangan sebuah sistem diperlukan *mockup* atau rancangan *interface* untuk memudahkan penyusunan antarmuka yang disediakan bagi pengguna aplikasi. Berikut *mockup* sistem Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means.

1. Halaman Form Utama (Home)

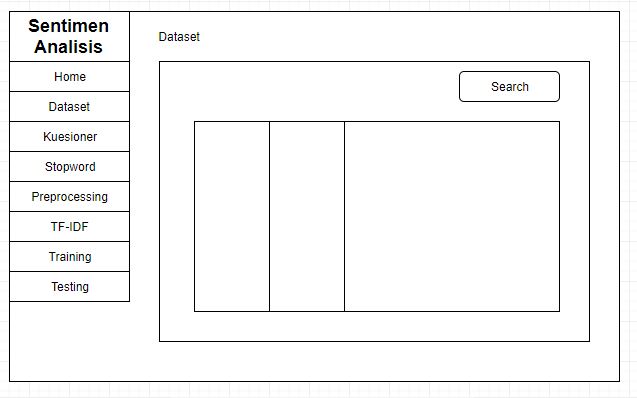
Halaman form utama adalah tampilan awal aplikasi saat pertama dibuka dan untuk memilih proses selanjutnya. Komponen dalam halaman form utama berupa judul, dan button untuk beralih ke menu lain dalam sistem. Berikut desain tampilan *interface* dari halaman menu utama pada Gambar 4.4.



Gambar 4. Desain Form Menu Utama

1. Halaman Dataset

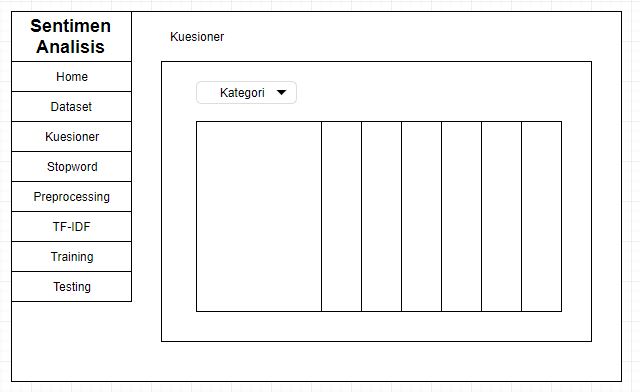
Pada halaman dataset digunakan untuk menampilkan seluruh data *tweet* yang telah disimpan dan juga penilaian secara manual yang dilakukan oleh 3 orang*.* Komponen dalam form ini berupa menu navigasi ke form lain. Berikut desain tampilan halaman data komentar terdapat pada Gambar 4.5.



Gambar 4.5 Desain Form Dataset

1. Halaman Kuesioner

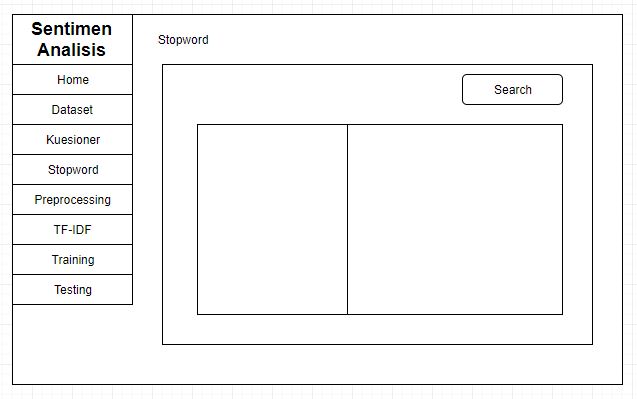
Halaman kuesioner digunakan untuk memasukkan data kuesioner baru serta menampilkan data hasil kuesioner. Komponen dalam form ini adalah menu radio button untuk memilih nilai kuesioner, menu navigasi ke form lain dan dropdown untuk memilih kategori pantai yang akan diisikan pada form kuesioner. Berikut desain tampilan halaman kuesioner terdapat pada Gambar 4.6.



Gambar 4.6 Desain Form Kuesioner

1. Halaman *Stopword*

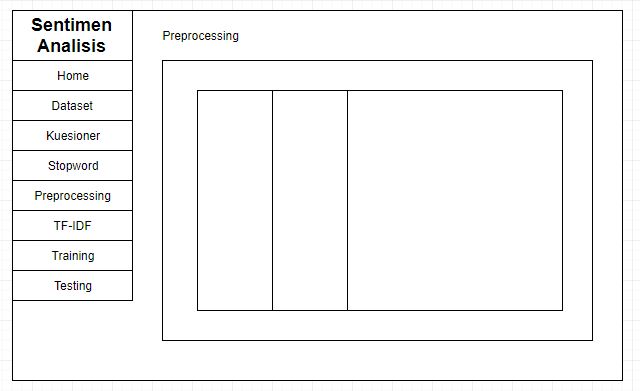
Halaman *stopword* digunakan untuk menunjukkan daftar kata-kata apa saja yang nantinya akan dihapus pada proses preprocessing. Komponen dalam form ini adalah menu navigasi ke form lain. Berikut desain tampilan halaman *stopword* terdapat pada Gambar 4.7.



Gambar 4.7 Desain Form *Stopword*

1. Halaman Preprocessing

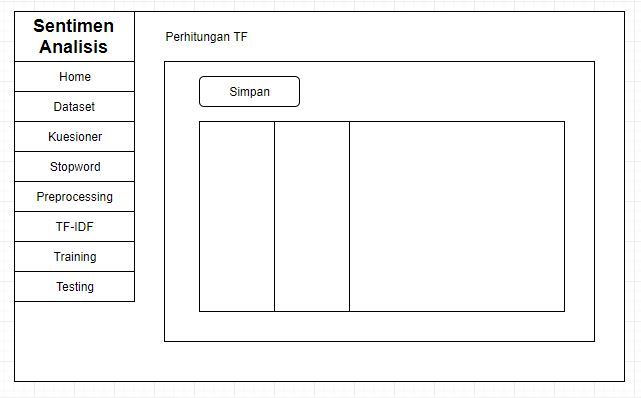
Halaman *preprocessing* digunakan untuk menunjukan hasil dari *preprocessing* sehingga didapat *tweet* yang bersih. Komponen dalam form ini adalah menu navigasi ke form lain dan textarea untuk menunjukkan pembersihan data. Berikut desain tampilan halaman *preprocessing* terdapat pada Gambar 4.8.



Gambar 4.8 Desain Form *Preprocessing*

1. Halaman TF-IDF

Halaman form TF-IDF adalah tampilan untuk menunjukan perhitungan dari kumpulan kalimat yang sudah dibuat oleh pengguna*.* Komponen dalam form ini berupa menu navigasi ke form lain. Berikut desain tampilan halaman TF-IDF terdapat pada Gambar 4.9.



Gambar 4.9 Desain Form TF-IDF

1. Halaman Training

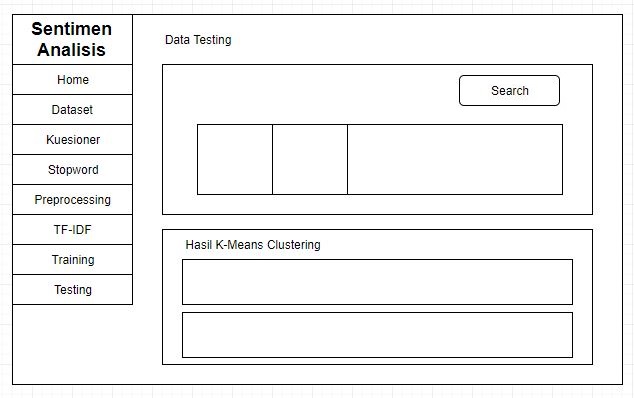
Pada halaman Training digunakan untuk melihat data mana saja yang digunakan sebagai pelatihan klasifikasi SVM. Komponen dalam form ini adalah menu navigasi ke form lain. Berikut desain tampilan halaman klasifikasi terdapat pada Gambar 4.10.



Gambar 4.10 Form Training

1. Halaman Testing

Pada halaman Testing digunakan untuk menunjukkan hasil klasifikasi yang dilakukan sistem dan juga hasil K-Means Clustering dari kuesioner. Komponen dalam form ini adalah menu navigasi ke form lain dan tombol proses untuk memproses klasterisasinya. Berikut desain tampilan halaman klasifikasi terdapat pada Gambar 4.11.



Gambar 4.11 Form Testing

# BAB V. IMPLEMENTASI

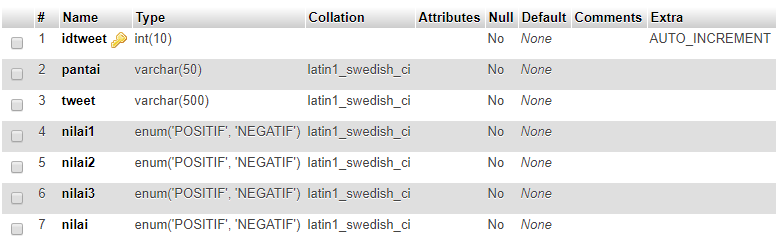
Pada bab ini akan akan membahas implementasi sistem yang telah dibangun menggunakan bahasa pemrograman PHP. Dibawah ini merupakan langkah - langkah penggunaan sistem *sentiment analysis* yang telah dibuat.

## Implementasi Basis Data

Implementasi sistem merupakan proses bagaimana sistem berjalan berdasarkan analisa dan perancangan yang dilakukan sebelumnya. Dimana sistem dapat melakukan klasterisasi analisis sentimen data *twitter* terhadap opini wisata pantai. *Database* dari sistem aplikasi ini dibangun menggunakan sistem *database* MySql sebagai tempat penyimpanan data. Terdapat 3 tabel dalam *database* ini antara lain:

1. Tabel dataset

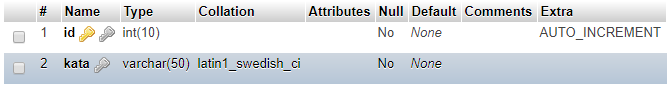
Tabel ini berfungsi untuk menampung seluruh data tweet mentah yang digunakan dalam proses klasterisasi. Struktur table dataset terdapat pada Gambar 5.1.



Gambar 5. Struktur Tabel dataset

1. Tabel stopword

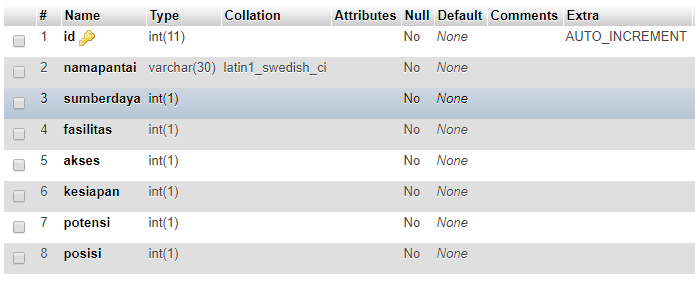
Tabel ini untuk menyimpan data berisi kata-kata yang diabaikan/dihapuskan pada pemrosesan selanjutnya. Struktur tabel stopword terdapat pada Gambar 5.2.



Gambar 5. Struktur Tabel stopword

1. Tabel kuesioner

Tabel ini untuk menyimpan hasil kuesioner dari penilaian-penilaian masyarakat. Struktur tabel kuesioner terdapat pada Gambar 5.3.



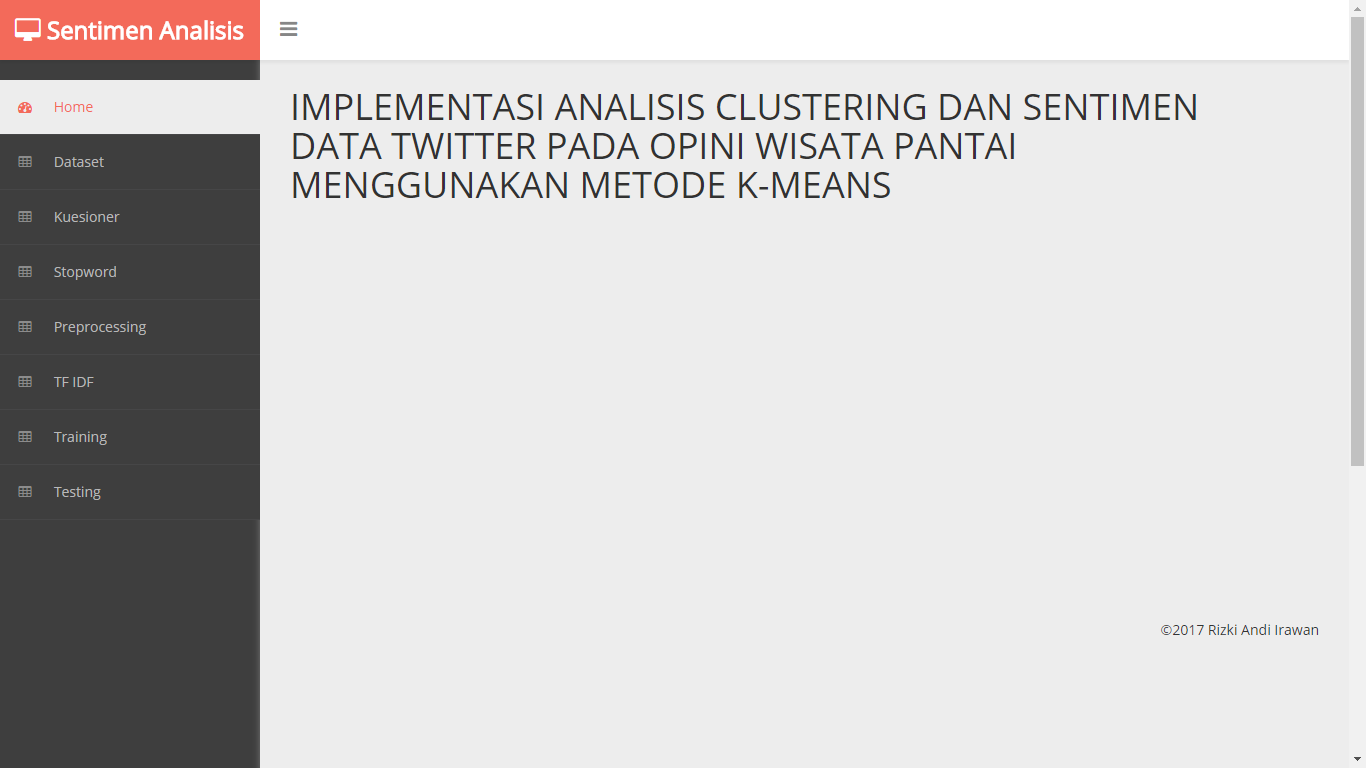
Gambar 5.3 Struktur Tabel kuesioner

## Implementasi Tampilan Antarmuka

Pembuatan tampilan antarmuka (*interface*) Sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means* ini memiliki bagian - bagian seperti berikut:

1. Halaman Form Utama (*Home*)

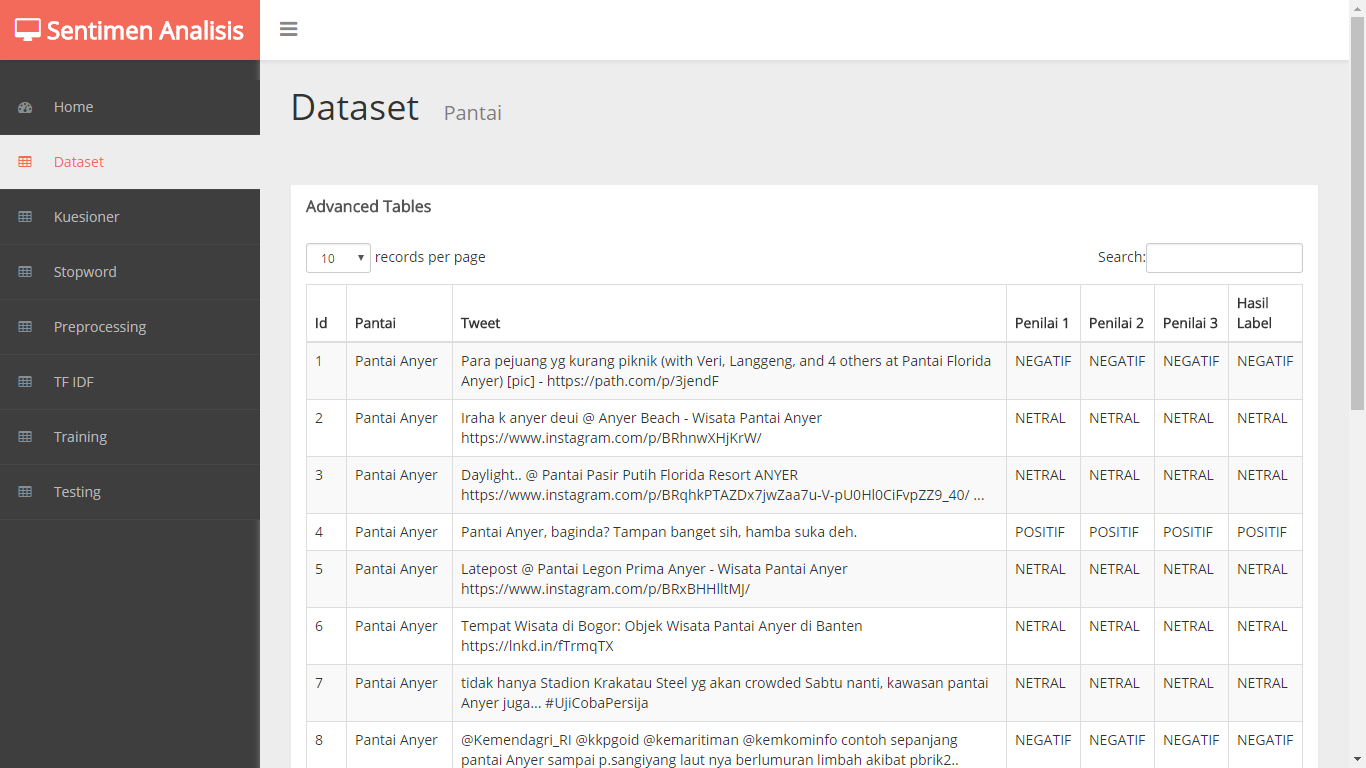
Halaman yang pertama kali ditampilkan oleh sistem adalah halaman menu utama. Pengguna dapat mengetahui judul penelitian yang sedang dilakukan. Berikut tampilan halaman menu utama pada Gambar 5.4.



Gambar 5.4 Implementasi Halaman Form Menu Utama

1. Halaman Dataset

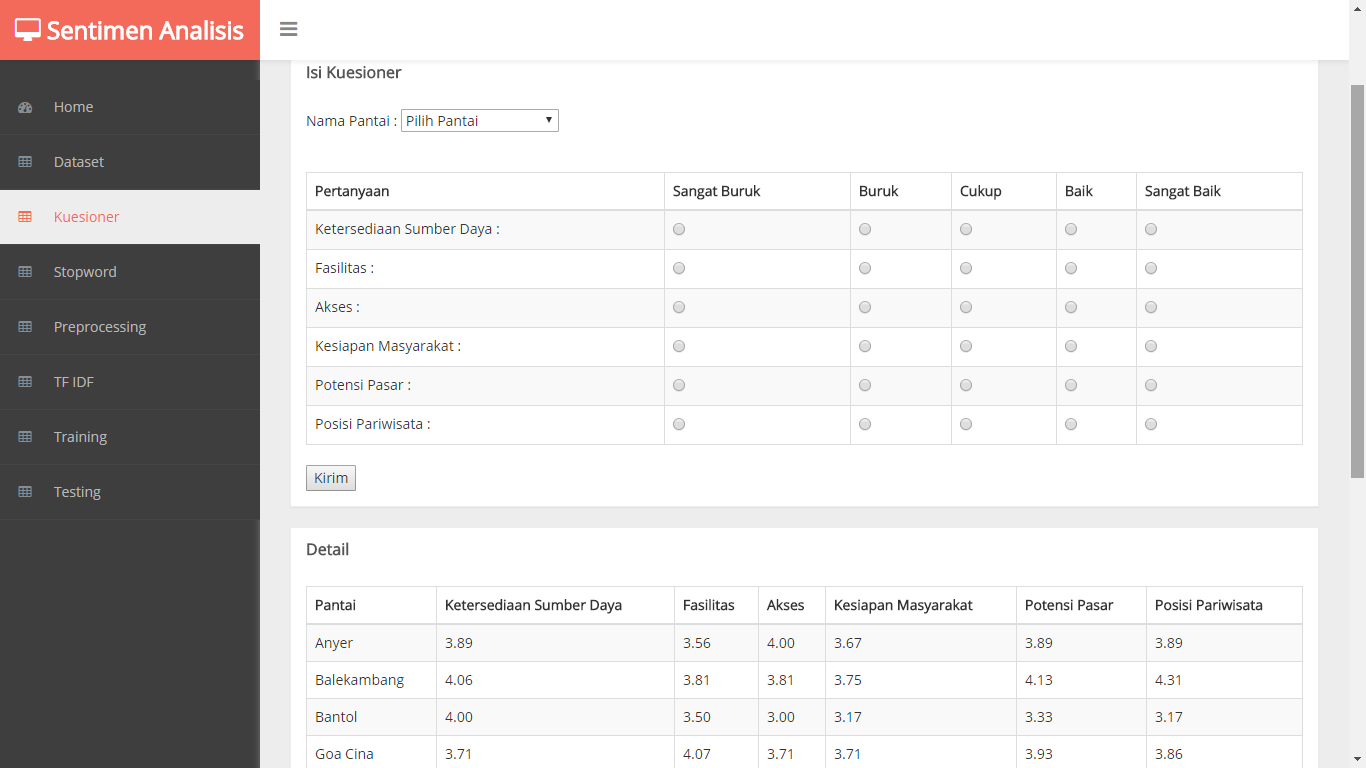
Halaman dataset menampilkan seluruh data *tweet* yang telah disimpan dan juga penilaian secara manual yang dilakukan oleh 3 orang. Berikut tampilan halaman dataset pada Gambar 5.5.



Gambar 5.5 Implementasi Halaman Dataset

1. Halaman Kuesioner

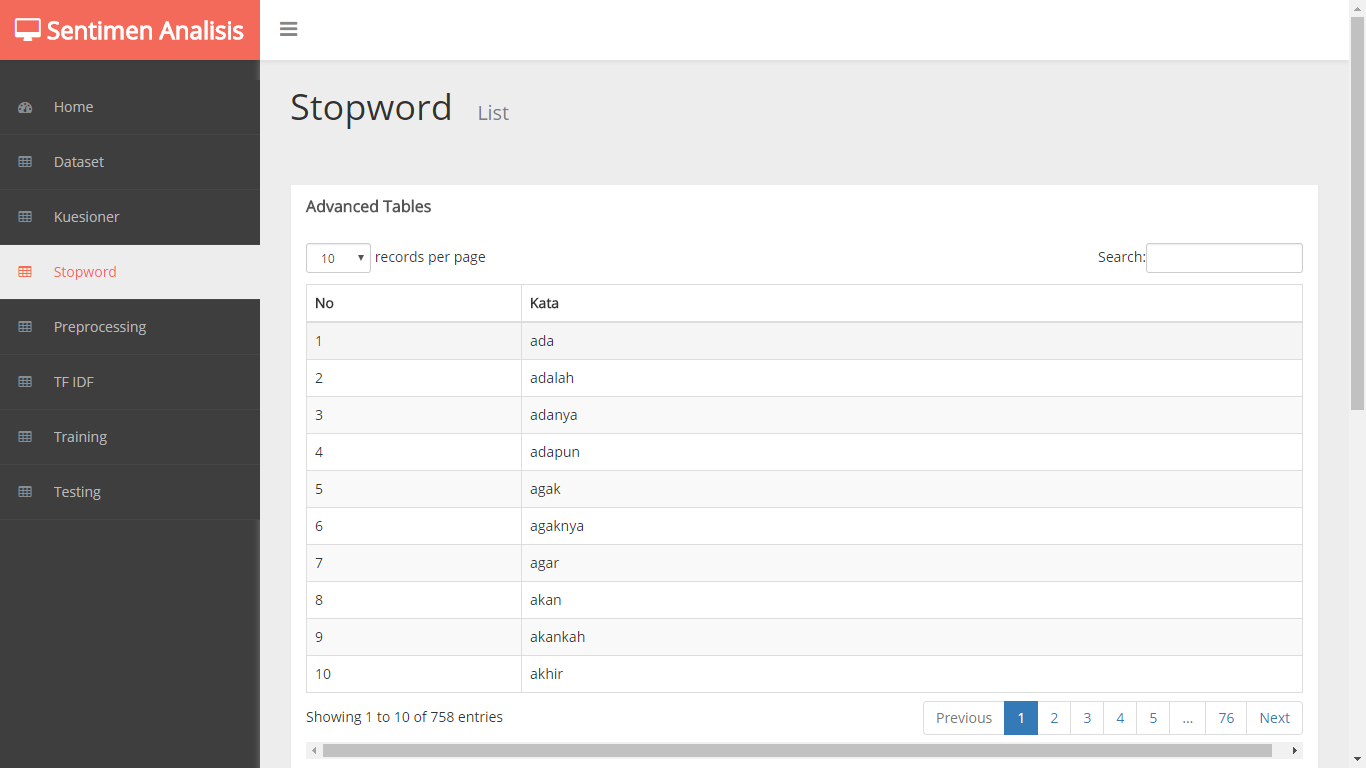
Halaman stopword digunakan untuk memasukkan data kuesioner baru serta menampilkan data hasil kuesioner. Berikut tampilan halaman Stopword pada Gambar 5.6.



Gambar 5.6 Implementasi Halaman Kuesioner

1. Halaman Stopword

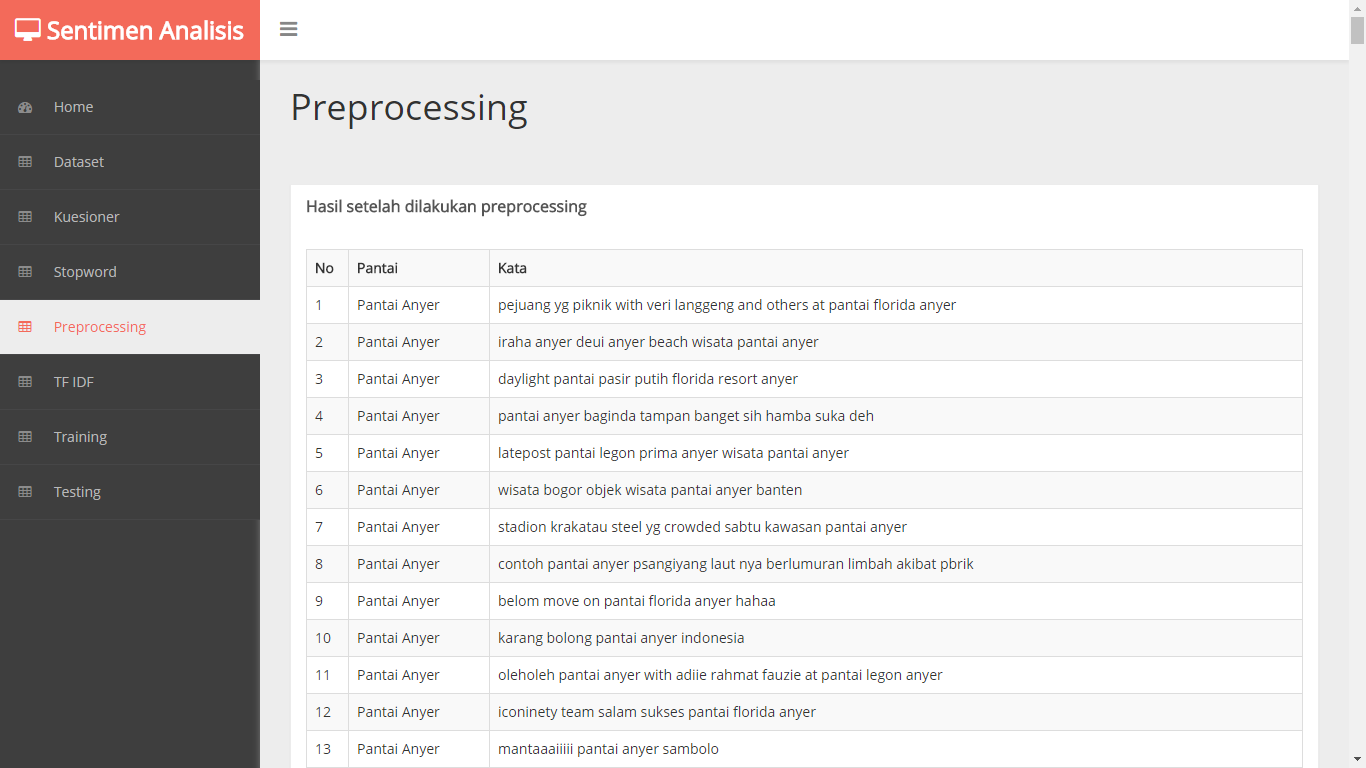
Halaman stopword digunakan untuk menunjukkan daftar kata-kata apa saja yang nantinya akan dihapus pada proses preprocessing. Berikut tampilan halaman Stopword pada Gambar 5.7.



Gambar 5.7 Implementasi Halaman Stopword

1. Halaman Preprocessing

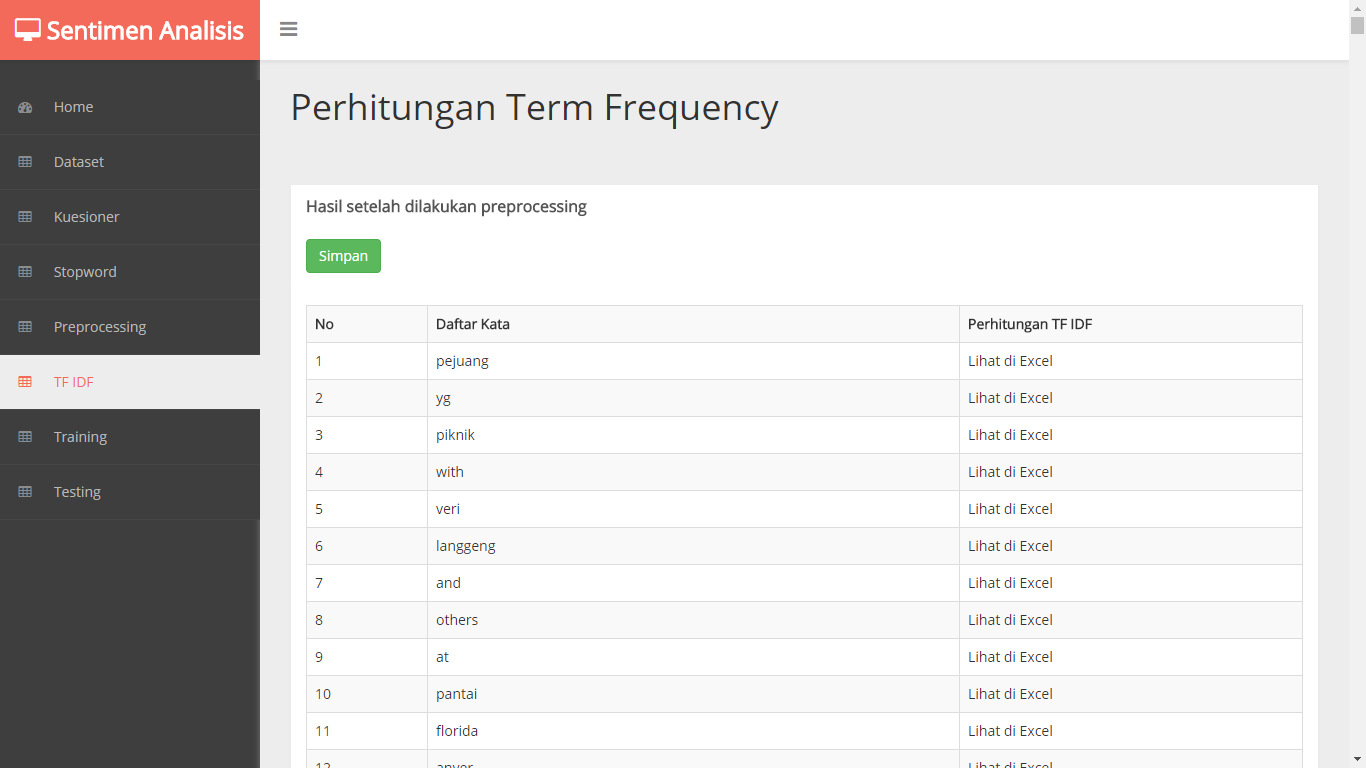
Halaman preprocessing digunakan untuk menunjukan hasil dari preprocessing (pembersihan kalimat) sehingga didapat *tweet* yang bersih. Berikut tampilan halaman preprocessing pada Gambar 5.8.



Gambar 5.8 Implementasi Halaman Preprocessing

1. Halaman TF-IDF

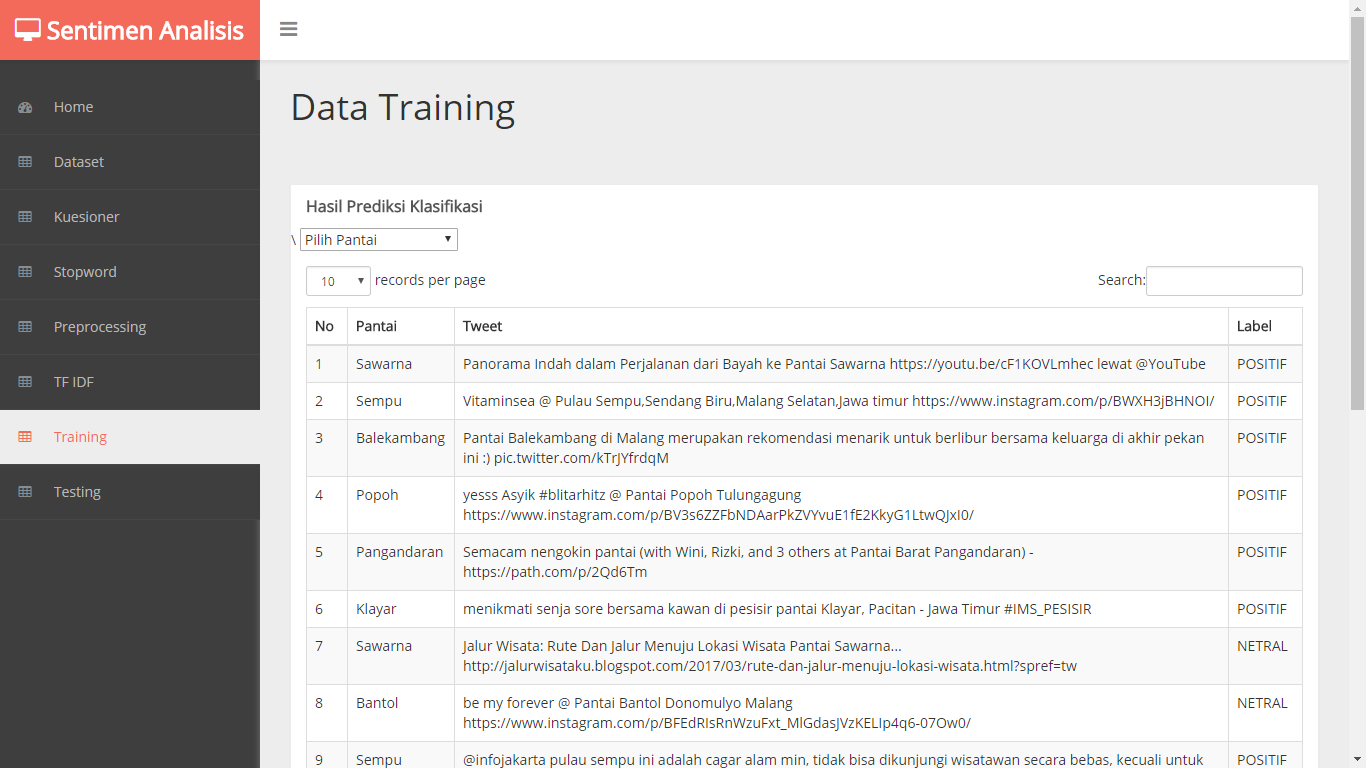
Halaman form TF-IDF adalah untuk menampilkan daftar kata yang telah dipecah dari setiap kalimat. Terdapat tombol simpan untuk menyimpan kedalam excel. Berikut tampilan halaman TF-IDF pada Gambar 5.9.



Gambar 5.9 Implementasi Halaman TF-IDF

1. Halaman Training

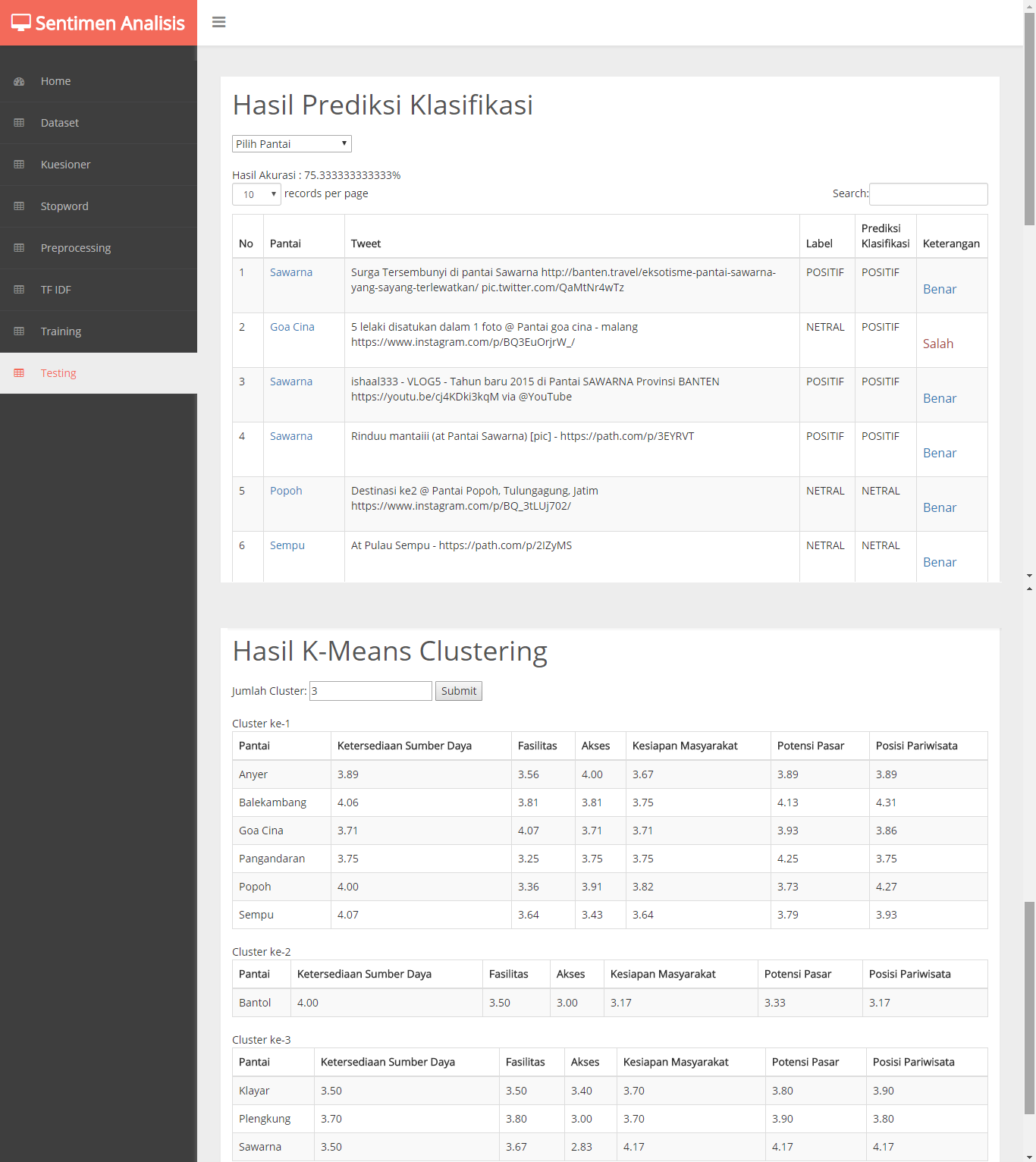
Pada halaman Training digunakan untuk melihat data mana saja yang digunakan sebagai pelatihan untuk klasifikasi SVM. Berikut tampilan halaman Trainingpada Gambar 5.10.



Gambar 5.10 Implementasi Halaman Training

1. Halaman Testing

Pada halaman Testing digunakan untuk menunjukkan hasil klasifikasi yang dilakukan sistem dan juga hasil K-Means Clustering dari kuesioner.. Berikut tampilan halaman Testingpada Gambar 5.11.



Gambar 5.11 Implementasi Halaman Testing

# BAB VI. PENGUJIAN DAN PEMBAHASAN

Setelah melakukan proses klasifikasi menggunakan algoritma *Support Vector Machine*, maka untuk melihat apakah hasil dari klasifikasi data tweet terhadap opini wisata pantai telah berhasil akan dilakukan suatu pengujian.

Pengujian dilakukan berdasarkan spesifikasi sistem dan pengujian ketepatan klasterisasi. Pengujian spesifikasi sistem yang dilakukan meliputi pengujian kesesuaian proses tiap tahap yang dilakukan oleh sistem. Sedangkan pengujian ketepatan klasterisasi meliputi ketepatan sistem dalam mengklasterisasikan sentimen pada tweet yang digunakan.

## Uji Coba

Pengujian merupakan cara atau teknik untuk menguji perangkat lunak, mempunyai mekanisme untuk menentukan data uji yang dapat menguji perangkat lunak secara lengkap dan mempunyai kemungkinan tinggi untuk menemukan kesalahan. Berikut ini merupakan pengujian yang dilakukan di dalam sistem Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means*.

## Pengujian Performa Fungsionalitas Sistem

Untuk pengujian performa fungsionalitas sistem dilakukan dengan cara menjalankan setiap fitur dalam aplikasi dan melihat kesesuaian hasil yang terjadi dengan hasil yang diharapkan. Berikut hasil pengujian fungsionalitas dari sistem pada Tabel 6.1.

Tabel 6. Hasil Pengujian Fungsionalitas

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Pola Pengujian** | **Hasil Pengujian** | **Status Pengujian** |
| 1 | Pada menu Dataset | Tabel data berisi data – data tweet yang telah disimpan pada *database* | Berhasil |
| 2 | Pada menu Kuesioner | Sistem dapat memasukkan data baru dan melihat hasil kuesioner | Berhasil |
| 3 | Pada menu Stopword | Tabel data berisi data – data stopword yang disimpan pada *database* | Berhasil |
| 4 | Pada menu *Preprocessing* | Sistem dapat melakukan proses pembersihan data tweet | Berhasil |
| 5 | Pada menu TF-IDF | TF IDF menampilkan hasil dari perhitungan dataset yang telah melalui *preprocessing* | Berhasil |
| 6 | Pada menu *Training* | *Training* menampilkan data untuk proses *training* | Berhasil |
| 7 | Pada menu *Testing* | *Testing* menampilkan data *testing* klasifikasi dan *cluster* dari kuesioner | Berhasil |

## Pengujian Akurasi Sistem

Pengujian akurasi sistem dilakukan untuk mengetahui tingkat kebenaran ssistem dalam mengklasifikasikan data *tweet* yang telah dijadikan data uji. Untuk mengetahui keberhasilan sistem dalam proses klasifikasi maka akan dibandingkan proses klasifikasi yang dilakukan secara manual dan klasifikasi yang dilakukan oleh sistem. Perbandingan tersebut akan dihitung tingkat kebenarannya menggunakan rumus *Accuracy*. Semakin tinggi nilai *accuracy* yang didapatkan sistem pada proses hasil klasifikasi menunjukkan sistem telah bekerja dengan baik pada proses klasifikasi. Sebaliknya apabila hasil nilai *accuracy* rendah menunjukkan sistem telah gagal dalam melakukan proses klasifikasi. Berikut merupakan beberapa contoh perbandingan hasil klasifikasi yang dilakukan secara manual dengan klasifikasi yang dilakukan oleh sistem pada Tabel 6.2.

Tabel 6.2 Perbandingan Klasifikasi

|  |  |  |  |
| --- | --- | --- | --- |
| **Tweet** | **Manual** | **Sistem** | **Hasil** |
| Pantai Plengkung (G-Land), Gulungan Ombak Legendaris,https://gpswisataindonesia.info/2015/03/pantai-plengkung-g-land-gulungan-ombak-legendaris/ | Positif | Positif | Benar |
| @rdi971 @KongJaing2 @ainatitta aq kmaren rebutan sobat rdi sm @PamanFCR yg mo ikut JJE ke pantai florida anyer...#ToTwit | Netral | Positif | Salah |
| 1.Pantai Batu Karas Pantai Batu Karas merupakan pantai yang terletak tidak jauh dari pantai Pangandaran, pantai... http://fb.me/8xnfUPwZf | Netral | Netral | Benar |
| Malam minggu kemarin, Pantai Klayar jadi tempat bermalam kami. pic.twitter.com/tmwXjUOR4m | Netral | Positif | Salah |
| Lokasi: Pantai Klayar, Kab. Pacitan, Jawa Timur Pantai Klayar menawarkan eksotisme dan fenomena alam yang cukup... http://fb.me/zERHUXLG | Positif | Netral | Salah |
| PANTAI BANTOL location : Kedung Celeng, Banjarjo, Donomulyo, Malang | Netral | Netral | Benar |
| pantai gading mbek, kulon e plengkung gading https://twitter.com/mufqiS/status/840218721926840321 | Netral | Netral | Benar |
| Pantai Anyer, baginda? Tampan banget sih, hamba suka deh. | Positif | Positif | Benar |
| Pantai Goa Cina. Kapan? Besok weekend panjang lo. https://www.instagram.com/p/BR8XCc0j4Rm/ | Netral | Netral | Benar |
| Secercah harapan Sabtu malamnya Di pesisir pantai goa cina Dalam satu atap tenda diskusi menarik terselenggara... http://fb.me/8J81WD9Vc | Positif | Positif | Benar |
| Indahnya....AKU dan PANTAI @ Tanjung Layar, Sawarna Beach https://www.instagram.com/p/BSLHpeeBu6b/ | Positif | Positif | Benar |
| G-Land alias Pantai Plengkung di Taman Nasional Alas Purwo Banyuwangi, Spot surfing terbaik ASIA Tenggara. gland #pantaiplengkung pic.twitter.com/IeDpS6lTVX | Positif | Positif | Benar |
| Paket Wisata Pulau Sempu - Bromo 2D1N https://www.bukalapak.com/p/tiket-voucher/travel-hiburan/7p9lja-jual-paket-wisata-pulau-sempu-bromo-2d1n | Netral | Positif | Salah |
| Piknik tipis-tipis @ Pantai Bantol https://www.instagram.com/p/BOlqVkhgkTg/ | Positif | Positif | Benar |
| Bahagia itu sederhana #bigfamily @ Pantai Goa Cina https://www.instagram.com/p/BQZtYdygffn/ | Positif | Positif | Benar |

Dari percobaan data tweets seperti ditunjukkan oleh tabel 6.2 merupakan perbandingan klasifikasi yang dilakukan secara manual dengan yang dilakukan oleh sistem. Penilaian akan bernilai benar jika kedua perbandingan memiliki hasil klasifikasi yang sama sedangkan akan bernilai salah jika kedua perbandingan memiliki hasil klasifikasi yang berbeda. Setelah didapatkan hasil penilaian (bernilai benar atau salah) maka akan dimasukkan pada rumus *accuracy* sehingga diketahui sistem memiliki akurasi yang tinggi atau rendah.

Sistem ini dalam pembagian data antara data pembelajaran dan data uji dilakukan secara acak oleh sistem, dari 500 total data tweet diambil 70% untuk digunakan sebagai data pembelajaran dan 30% sebagai data uji. Setiap pengujian akan memberikan hasil *accuracy* yang berbeda karena data pembelajaran dan data uji akan berbeda dalam sekali proses klasifikasi. Maka dari itu akan dilakukan proses klasifikasi sebanyak 5 (lima) kali proses klasifikasi, dimana setiap proses ke-1 sampai proses ke-5 menggunakan data pembelajaran dan data uji yang berbeda-beda yang dipilih acak oleh sistem seperti yang ditunjukkan pada tabel 6.3. Pengujian akurasi sistem dilakukan dengan cara menghitung nilai dari *accuracy*.

Rumus untuk menghitung nilai *accuracy* sebagai berikut.

|  |  |  |  |
| --- | --- | --- | --- |
| Keterangan | | | |
| v | : | Jumlah data benar |
| n | : | Jumlah dokumen |

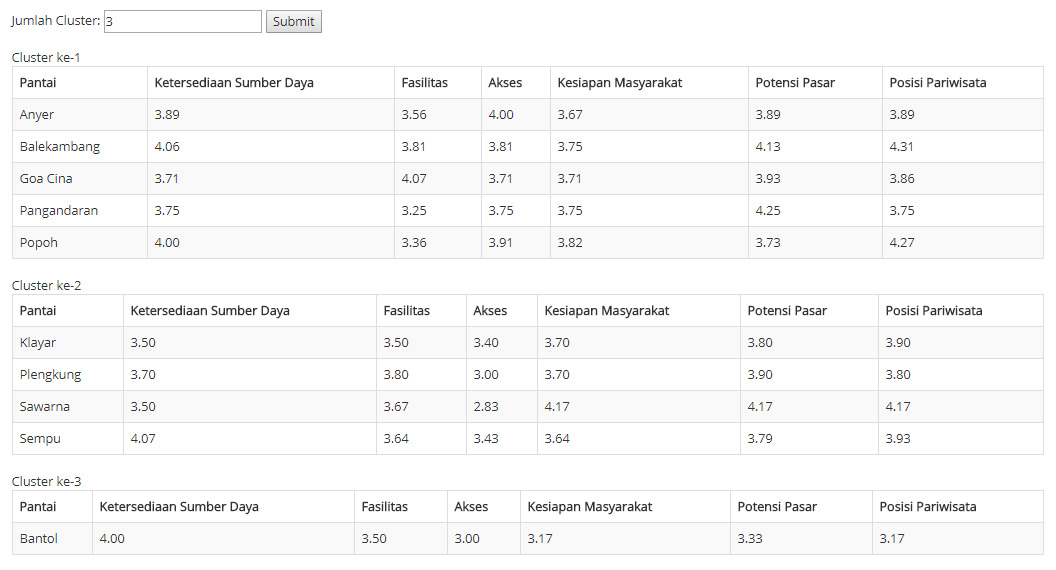
Tabel 6.3 Pengujian Akurasi Sistem

|  |  |
| --- | --- |
| **Pengujian** | ***Accuracy*** |
| **1** | 76,66 % |
| **2** | 78,66 % |
| **3** | 69,33 % |
| **4** | 72,00 % |
| **5** | 75,33 % |
| **Rata- rata** | **74,39 %** |

Dari percobaan seperti ditunjukkan oleh tabel 6.3 nilai rata-rata *accuracy* yang didapat oleh sistem adalah 74,39%, sedangkan nilai terendah pada proses ke-3 yaitu 69,33% dan nilai tertinggi didapatkan pada proses ke-2 yaitu 78,66%. Sehingga membuktikan algoritma *Support Vector Machine*  memberikan hasil yang baik dalam mengklasifikasi sebuah sentimen.

## Analisa Hasil Clustering

Analisa hasil clustering dilakukan untuk mengetahui potensi ketersediaan sumber daya, fasilitas, akses, kesiapan masyarakat, potensi pasar, posisi pariwisata yang ada pada pantai di Indonesia.



Gambar 6.1 Hasil Clustering Pantai

Berdasarkan Gambar 6.1 telah dilakukan cluster menjadi tiga cluster, dari hasil cluster yang telah dilakukan menampilkan bahwa hasil pada cluster pertama Pantai Anyer, Balekambang, Goa Cina, Pangandaran, dan Popoh menunjukan bahwa pantai tersebut memiliki nilai tiap kriteria yang tinggi. Pada cluster kedua Pantai Klayar, Plengkung, Sawarna, dan Sempu menunjukan pantai tersebut memiliki nilai yang lebih rendah dibandingkan cluster pertama. Pada cluster ketiga Pantai Bantol menunjukan bahwa pantai tersebut memiliki nilai yang rendah dibandingkan cluster pertama dan kedua. Berdasarkan hasil cluster tersebut bahwa cluster pertam­a dengan nilai tiap kriteria yang tinggi dapat dijadikan sebagai destinasi tujuan wisata.

# BAB VII. PENUTUP

## Kesimpulan

Adapun kesimpulan yang dapat diambil dari hasil penelitian yang dilakukan mengenai Implementasi Analisis *Clustering* Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode *K-Means* adalah sebagai berikut.

1. Implementasi *Support Vector Machine (SVM)* pada Analisis Sentimen bekerja dengan baik.
2. Algoritma *SVM* memberikan hasil dengan rata-rata *accuracy* yang didapatkan oleh sistem sebesar 74,39%
3. Hasil *K-Means* *clustering* dipengaruhi dari nilai titik pusat *cluster (centroid)* yang dan jumlah data yang digunakan. Selain itu perbedaan pengambilan data pusat *cluster* awal yang digunakan juga akan mempengaruhi hasil akhir pengelompokkan.

## Saran

Berdasarkan penelitian, ada beberapa hal yang disarankan antara lain sebagai berikut.

1. Aplikasi dapat dijalankan secara realtime guna mendapatkan hasil data yang selalu terbaru setiap waktunya.
2. Pada penelitian ini sistem tidak memiliki fitur *stemming* yang digunakan untuk mencari kata dasar/baku suatu kata, untuk penelitan selanjutnya diharapkan sistem memiliki fitur *stemming* pada proses *preprocessing*.

# DAFTAR PUSTAKA

[1] Liu, B. *Sentiment Analysis and Opining Mining*. Morgan & Claypool Publishers, 2012

[2] Pang, Bo and Lee, L, Vaithyanathan, S. 2002. “Sentiment Classification Using Machine Learning Techniques”. Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing (EMNLP-02). USA, 2002.

[3] Feldman, R, Sanger, J,. The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. New York : Cambridge University Press, 2007.

[4] Fenty Eka, dkk., “Implementasi Algoritma K-Means untuk Menetukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus : SMA Negeri 101 Jakarta)”, Jurnal Teknik Infomatika, 2015.

[5] Priyanto Hidayatullah dan Jauhari Khairul Kawistara, “Pemrograman WEB”, Bandung : Informatika, 2014

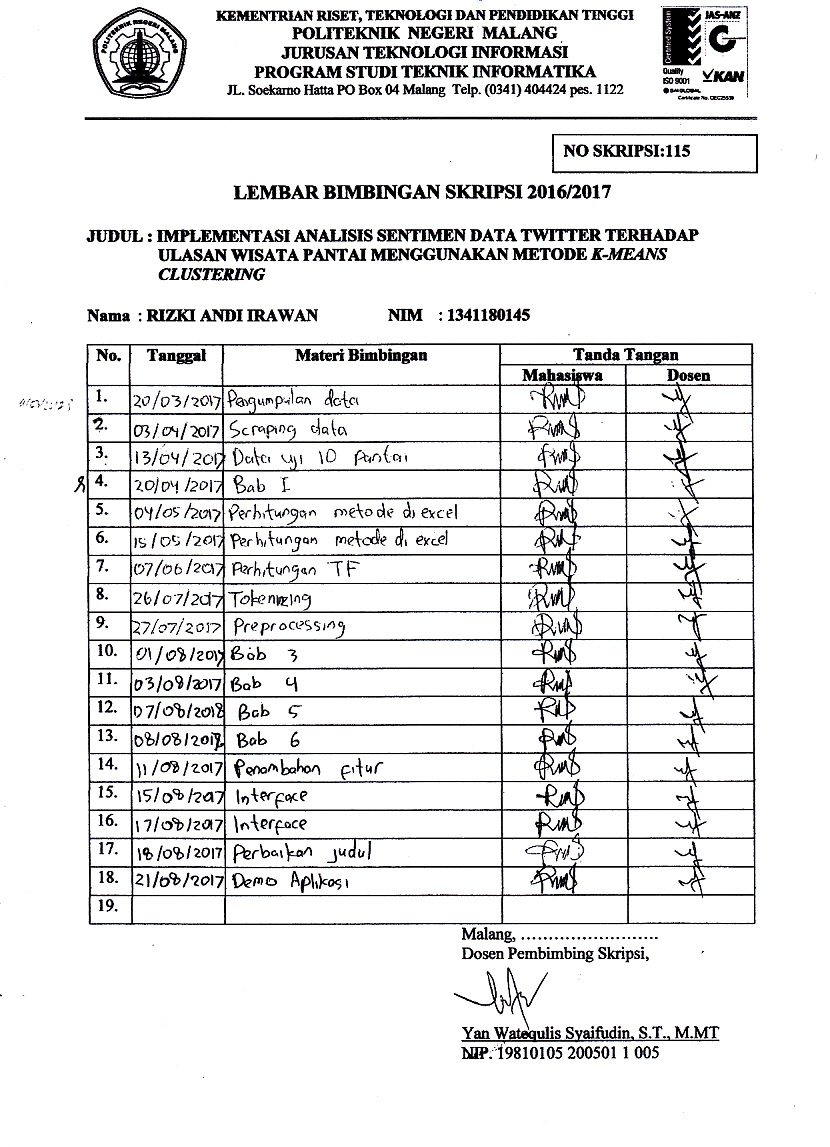
[6] Pressman, Roger S. *Software Engineering: A Practitioner’s*. McGraw-Hill Publishers, 2009

# LAMPIRAN

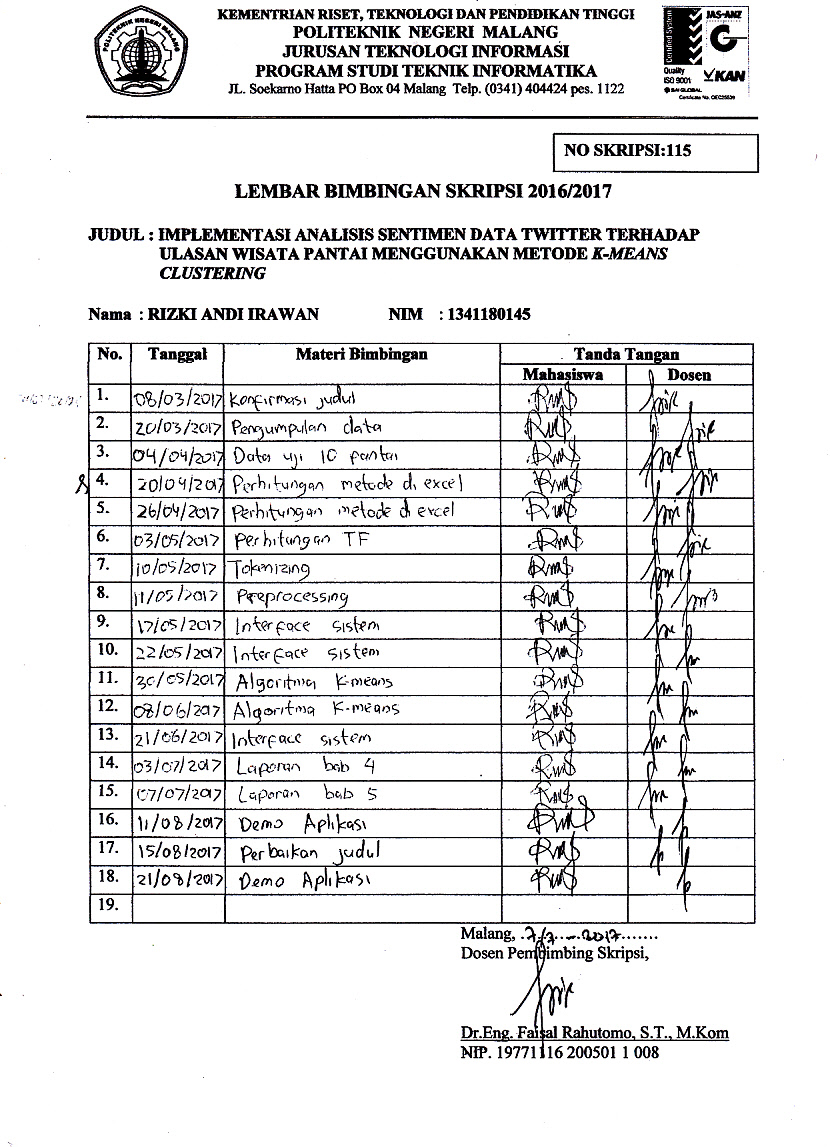
|  |
| --- |
| **Lampiran 1 Kode Program Data Training** |
| <?php  $vectorizer = new TokenCountVectorizer(new WordTokenizer());  $tfIdfTransformer = new TfIdfTransformer();  function preproses($samples) {    $samples = array\_map('strtolower', $samples);  $samples = preg\_replace("/[0-9]/", "",$samples);  $samples = preg\_replace("/\n/", " ", $samples); //enter  $samples = preg\_replace("/\s\s+/", " ", $samples); //  $samples = preg\_replace("/#(\\w+)/i", "", $samples);  $samples = preg\_replace("(https?://[^\\s]+)", "", $samples);  $samples = preg\_replace("(http?://[^\\s]+)", "", $samples);  $samples = preg\_replace("/@([^\\s]+)/i", "", $samples);  $samples = preg\_replace("/pic.([^\\s]+)/i", "", $samples);  $samples = preg\_replace("/@|pic/", "", $samples);  $samples = preg\_replace('/[^\sa-z0-9]+/S', '', $samples);    return $samples;  }  function hilangkata($samples) {  //koneksi ke db  $link = mysqli\_connect("localhost", "root", "", "skripsi");  $hasil = array();  //hilangkan stopword  $stopword = array();  $q= mysqli\_query($link, "SELECT kata FROM stopword");  while($h=mysqli\_fetch\_array($q , MYSQLI\_ASSOC))  {  $stopword[]=str\_replace("\r", '', $h['kata']);  }  foreach ($samples as $teks) {  $teks = preg\_replace('/\b(' . implode('|', $stopword) . '|[a-z])\b/', '', $teks);  $teks = preg\_replace('/\s+/S', ' ', $teks);  $hasil[] = trim($teks);  }  return $hasil;  }    $link = mysqli\_connect("localhost", "root", "", "skripsi");  $isitweet = array();  $datalatih = array();  $labelin = array();  $pantai = array();  $q= mysqli\_query($link, "SELECT \* FROM dataset");  while($h=mysqli\_fetch\_array($q , MYSQLI\_ASSOC))  {  $isitweet[]=$h['tweet'];  $datalatih[]=$h['tweet'];  $labelin[] =$h['nilai'];  $pantai[]=$h['pantai'];  }  $datalatih2 = preproses($datalatih);  $datalatih3 = hilangkata($datalatih2);  //print\_r($datalatih3);  $vectorizer->fit($datalatih3);  $vectorizer->transform($datalatih3);  $tfIdfTransformer->fit($datalatih3);  $tfIdfTransformer->transform($datalatih3);    $dataset = new ArrayDataset($isitweet, $datalatih3, $labelin, $pantai);  //print\_r($dataset);  $jumlahtesting = 0.3;  $randomSplit = new RandomSplit($dataset, $jumlahtesting);  $printpantai = $randomSplit->getTrainPantai();  $printtweet = $randomSplit->getTrainTweets();  $printlabel = $randomSplit->getTrainLabels();  $classifier = new SVC(Kernel::RBF, 10000);  $classifier->train($randomSplit->getTrainSamples(), $randomSplit->getTrainLabels());  $predictedLabels = $classifier->predict($randomSplit->getTestSamples());  ?>  <thead>  <tr>  <th>No</th>  <th>Pantai</th>  <th>Tweet</th>  <th>Label</th>  </tr>  </thead>  <tbody>  <?php foreach($printpantai as $key => $value) { ?>  <tr>  <!--urutan nomor otomatis +1 karena mulai awal 0 -->  <td><?= $key+1 ?></td>  <td><?= $printpantai[$key] ?></td>  <td><?= $printtweet[$key] ?></td>  <td><?= $printlabel[$key] ?></td>  </tr>  <?php |

|  |
| --- |
| **Lampiran 2 Kode Program Data Testing** |
| <?php  $vectorizer = new TokenCountVectorizer(new WordTokenizer());  $tfIdfTransformer = new TfIdfTransformer();  function preproses($samples) {    $samples = array\_map('strtolower', $samples);  $samples = preg\_replace("/[0-9]/", "",$samples);  $samples = preg\_replace("/\n/", " ", $samples); //enter  $samples = preg\_replace("/\s\s+/", " ", $samples); //  $samples = preg\_replace("/#(\\w+)/i", "", $samples);  $samples = preg\_replace("(https?://[^\\s]+)", "", $samples);  $samples = preg\_replace("(http?://[^\\s]+)", "", $samples);  $samples = preg\_replace("/@([^\\s]+)/i", "", $samples);  $samples = preg\_replace("/pic.([^\\s]+)/i", "", $samples);  $samples = preg\_replace("/@|pic/", "", $samples);  $samples = preg\_replace('/[^\sa-z0-9]+/S', '', $samples);    return $samples;  }  function hilangkata($samples) {  //koneksi ke db  $link = mysqli\_connect("localhost", "root", "", "skripsi");  $hasil = array();  //hilangkan stopword  $stopword = array();  $q= mysqli\_query($link, "SELECT kata FROM stopword");  while($h=mysqli\_fetch\_array($q , MYSQLI\_ASSOC))  {  $stopword[]=str\_replace("\r", '', $h['kata']);  }  foreach ($samples as $teks) {  $teks = preg\_replace('/\b(' . implode('|', $stopword) . '|[a-z])\b/', '', $teks);  $teks = preg\_replace('/\s+/S', ' ', $teks);  $hasil[] = trim($teks);  }  return $hasil;  }    $link = mysqli\_connect("localhost", "root", "", "skripsi");  $isitweet = array();  $datalatih = array();  $labelin = array();  $pantai = array();  $q= mysqli\_query($link, "SELECT \* FROM dataset");  while($h=mysqli\_fetch\_array($q , MYSQLI\_ASSOC))  {  $isitweet[]=$h['tweet'];  $datalatih[]=$h['tweet'];  $labelin[] =$h['nilai'];  $pantai[]=$h['pantai'];  }  $datalatih2 = preproses($datalatih);  $datalatih3 = hilangkata($datalatih2);  //print\_r($datalatih3);  $vectorizer->fit($datalatih3);  $vectorizer->transform($datalatih3);  $tfIdfTransformer->fit($datalatih3);  $tfIdfTransformer->transform($datalatih3);  $dataset = new ArrayDataset($isitweet, $datalatih3, $labelin, $pantai);  //print\_r($dataset);  $jumlahtesting = 0.3  ;  $randomSplit = new RandomSplit($dataset, $jumlahtesting);  $printpantai = $randomSplit->getTestPantai();  $printtweet = $randomSplit->getTestTweets();  $printlabel = $randomSplit->getTestLabels();  $classifier = new SVC(Kernel::RBF, 10000);  $classifier->train($randomSplit->getTrainSamples(), $randomSplit->getTrainLabels());  $predictedLabels = $classifier->predict($randomSplit->getTestSamples());  //echo 'Accuracy: '.Accuracy::score($randomSplit->getTestLabels(), $predictedLabels);  ?>  <thead>  <tr>  <th>No</th>  <th>Pantai</th>  <th>Tweet</th>  <th>Label</th>  <th>Prediksi Klasifikasi</th>  <th>Keterangan</th>  </tr>  </thead>  <tbody>  <?php  foreach($printpantai as $key => $value) {  if (!isset($dataChart[$value])) {  $dataChart[$value] = array(  "positif" => 0,  "negatif" => 0,  "netral" => 0  );}  switch ($predictedLabels[$key]) {  case 'POSITIF':  $dataChart[$value]["positif"]++;  break;  case 'NEGATIF':  $dataChart[$value]["negatif"]++;  break;  case 'NETRAL':  $dataChart[$value]["netral"]++;  break;}  ?> |

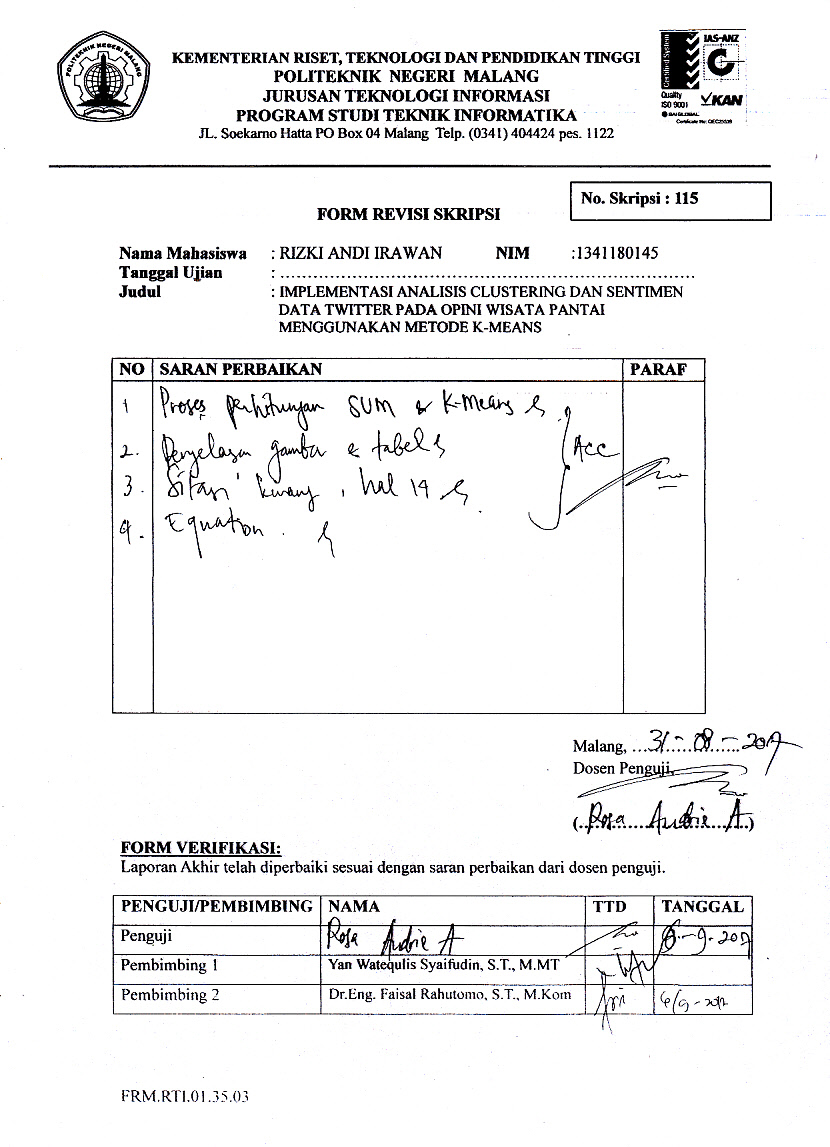
**Lampiran 3 Lembar Bimbingan I**

****

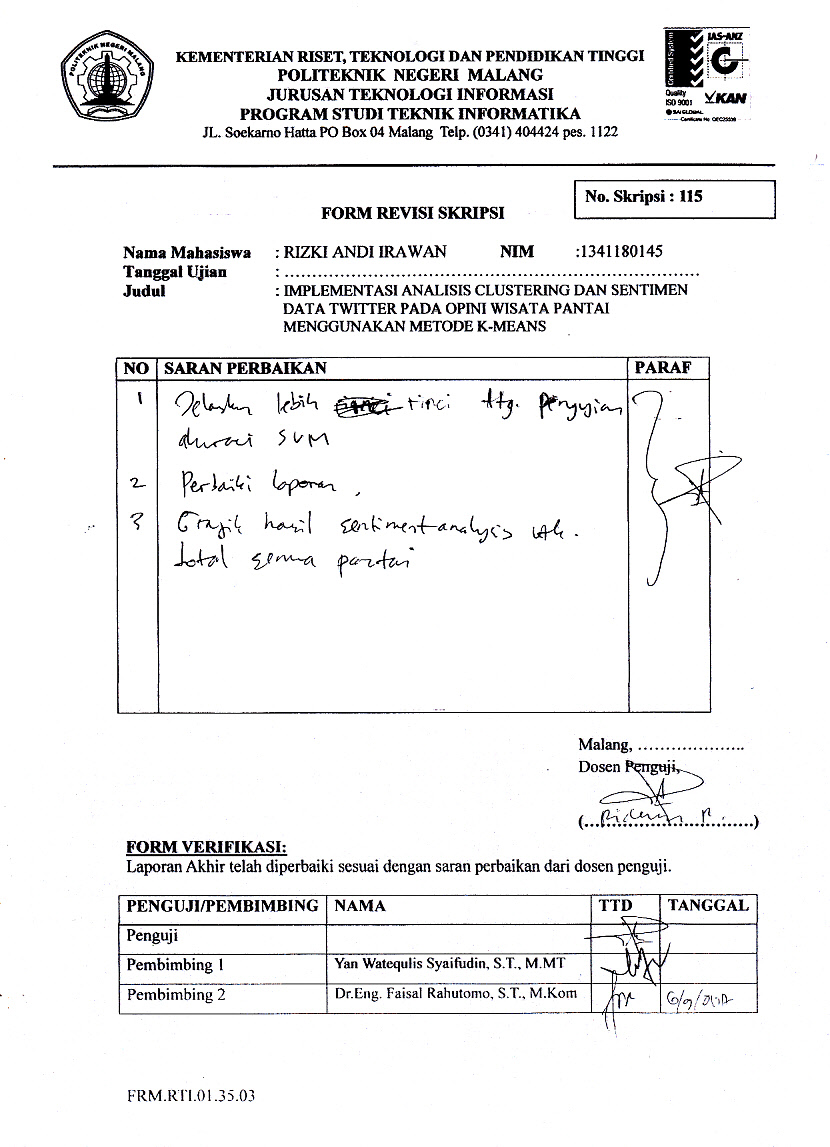
**Lampiran 4 Lembar Bimbingan II**

****

**Lampiran 5 Lembar Form Revisi Penguji I**

****

**Lampiran 6 Lembar Form Revisi Penguji II**

****

**PROFIL PENULIS**

|  |  |  |
| --- | --- | --- |
| **IDENTITAS PRIBADI** | | |
| Nama | Rizki Andi Irawan | E:\Gambar\pas foto\DSC_3372 jas biru.png |
| Umur | 22 |
| Tempat, Tanggal Lahir | Malang, 24 Januari 1995 |
| Jenis Kelamin | Laki-Laki |
| Kewarganegaraan | Indonesia |
| Alamat | Jl. L.A. Sucipto 350 Blimbing, Malang |
| Email | [rizkiandiirawan@gmail.com](mailto:rizkiandiirawan@gmail.com) |
| No Handphone | 089505900986 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **RIWAYAT PENDIDIKAN** | | | | |
| **Nama** | **Tempat** | **Mulai** | **Selesai** | **Jurusan** |
| Sekolah Dasar Muhammadiyah 8 Malang | Malang, Jawa Timur | 2001 | 2007 |  |
| SMPN 8 Malang | Malang, Jawa Timur | 2007 | 2010 |  |
| SMKN 5 Malang | Malang, Jawa Timur | 2010 | 2013 | Teknik Komputer dan Jaringan |
| Perguruan Tinggi : Politeknik Negeri Malang | Malang, Jawa Timur | 2013 | 2017 | Diploma IV Teknik Informatika |