# PENGARUH STOPWORD TERHADAP PERFORMA ANALISIS SENTIMEN PADA ULASAN GOOGLE MAPS MENGGUNAKAN METODE

# NAÏVE BAYES CLASSIFIER

**Tugas Akhir**

Tugas Akhir ini sebagai salah satu syarat untuk memperoleh gelar sarjana

Informatika Universitas Pembangunan Nasional “Veteran” Yogyakarta



Disusun Oleh :

Ilham Utomo Aji

123160123

**PROGRAM STUDI INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2020**

# HALAMAN

**PENGESAHAN PEMBIMBING**

**PENGARUH STOPWORD TERHADAP PERFORMA ANALISIS SENTIMEN PADA ULASAN GOOGLE MAPS MENGGUNAKAN METODE**

**NAÏVE BAYES CLASSIFIER**



Disusun oleh:

Ilham Utomo Aji 123160123

Telah diperiksa dan disetujui oleh pembimbing untuk diseminarkan pada tanggal: 26 Desember 2020

|  |  |
| --- | --- |
| Menyetujui,  Pembimbing I  Oliver Samuel Simanjuntak, S.Kom., M.Eng.  NIK. 2 8305 11 0300 1 | Menyetujui,  Pembimbing II  Mangaras Yanu Florestiyanto, S.T., M.Eng.  NIK. 2 8201 13 0425 1 |

Mengetahui,

Koordinator Tugas Akhir

Dr. Awang Hendrianto Pratomo, S.T., M.T.

NIP. 1977 07 25 2005 01 1001

# DAFTAR ISI

[Halaman Judul i](#_Toc60489345)

[HALAMAN Pengesahan ii](#_Toc60489346)

[DAFTAR ISI iii](#_Toc60489347)

[DAFTAR GAMBAR vi](#_Toc60489348)I

[DAFTAR TABEL viii](#_Toc60489349)I

[BAB I Pendahuluan 1](#_Toc60489350)

[1.1 Latar Belakang 1](#_Toc60489351)

[1.2 Rumusan Masalah 3](#_Toc60489352)

[1.3 Batasan Masalah 3](#_Toc60489353)

[1.4 Tujuan Penelitian 3](#_Toc60489354)

[1.5 Manfaat Penelitian 3](#_Toc60489355)

[1.6 Metode Penelitian 4](#_Toc60489356)

[1.6.1 Metode Pengumpulan Data 4](#_Toc60489357)

[1.6.2 Metode Pengembangan Sistem 4](#_Toc60489358)

[1.7 Sistematika Penelitian 6](#_Toc60489359)

[BAB II Tinjauan pustaka 8](#_Toc60489360)

[2.1 Ulasan *Google Maps* 8](#_Toc60489361)

[2.2 Analisis Sentimen dan *Text Mining* 8](#_Toc60489362)

[2.3 *Web Scraping* 9](#_Toc60489363)

[2.4 *Text Processing* 10](#_Toc60489364)

[2.4.1 *Cleansing* 10](#_Toc60489365)

[2.4.2 *Slangword* 11](#_Toc60489366)

[2.4.3 *Stopword* *Removal* 11](#_Toc60489367)

[2.4.4 *Stemming* 11](#_Toc60489368)

[2.4.5 *Tokenizing* 12](#_Toc60489369)

[2.4.6 Ekstraksi Fitur 12](#_Toc60489370)

[2.4.6.1 *Term Frequency (TF)* 12](#_Toc60489371)

[2.4.6.2 *TF-IDF* 12](#_Toc60489372)

[2.5 *Naïve Bayes Clasiffier* 13](#_Toc60489373)

[2.6 Pengujian 15](#_Toc60489374)

[2.7 Studi Pustaka *(State Of Art)* 16](#_Toc60489375)

[BAB III Metodologi penelitian dan pengembangan sistem 20](#_Toc60489376)

[3.1 Metodologi Penelitian 20](#_Toc60489377)

[3.2 Analisis Pengumpulan Data 22](#_Toc60489378)

[3.2.1 Studi Pustaka 22](#_Toc60489379)

[3.2.2 *Scraping Data* 22](#_Toc60489380)

[3.3 Analisis Kebutuhan 23](#_Toc60489381)

[3.3.1 Analisis Masalah 23](#_Toc60489382)

[3.3.2 Analisis Kebutuhan Sistem 23](#_Toc60489383)

[3.3.2.1 Kebutuhan Data Input 23](#_Toc60489384)

[3.3.2.2 Kebutuhan Proses 24](#_Toc60489385)

[3.3.2.3 Kebutuhan Data Output 25](#_Toc60489386)

[3.3.2.4 Kebutuhan Non Fungsional 25](#_Toc60489387)

[3.3.2.5 Analisis *Text Preporcessing* 26](#_Toc60489388)

[3.3.2.6 Pelabelan Manual 26](#_Toc60489389)

[3.3.2.7 Analisis Pelatihan Klasifikasi *Term Frequency* (TF) 27](#_Toc60489390)

[3.3.2.8 Analisis Penerapan Algoritma *Naïve Bayes Classifier* 28](#_Toc60489391)

[3.4 Proses Desain 32](#_Toc60489392)

[3.4.1 Perancangan Sistem 32](#_Toc60489393)

[3.4.2 Perancangan Proses 33](#_Toc60489394)

[3.4.2.1 *Data Flow Diagram* Level 0 33](#_Toc60489395)

[3.4.2.2 *Data Flow Diagram* Level 1 34](#_Toc60489396)

[3.4.2.4 *Flowchart* Analisis Sentimen 36](#_Toc60489397)

[3.4.2.5 *Flowchart Preprocessing* Dengan *Stopword* 37](#_Toc60489398)

[3.4.2.6 *Flowchart Preprocessing* Tanpa *Stopword* 38](#_Toc60489399)

[3.4.2.7 *Flowchart Celansing* 40](#_Toc60489400)

[3.4.2.8 *Flowchart* *Slangword* 41](#_Toc60489401)

[3.4.2.9 *Flowchart* *Stopword Removal* 43](#_Toc60489402)

[3.4.2.10 *Flowchart Stemming* 44](#_Toc60489403)

[3.4.2.11 *Tokenizing* 45](#_Toc60489404)

[3.4.2.12 *Flowchart Naïve Bayes Classifier* 46](#_Toc60489405)

[3.4.3 Perancangan Basis Data 48](#_Toc60489406)

[3.4.3.1 *Entity Relationship Diagram* (ERD) 48](#_Toc60489407)

[3.4.3.2 Relasi Antar Tabel (RAT) 49](#_Toc60489408)

[3.4.3.3 Struktur Tabel 50](#_Toc60489409)

[3.4.4 Perancanagan Antarmuka 53](#_Toc60489410)

[3.4 Rancangan Pengujian 62](#_Toc60489411)

[DAFTAR PUSTAKA 64](#_Toc60489412)

# DAFTAR GAMBAR

[Gambar 1.1 Metode *Waterffal* (Kadir 2003) 6](#_Toc60505445)

[Gambar 3.1 Metode Penelitian 21](#_Toc60505468)

[Gambar 3.2 Arsitektur Sistem 33](#_Toc60505494)

[Gambar 3.3 *Data Flow Diagram* Level 0 34](#_Toc60505497)

[Gambar 3.4 *Data Flow Diagram* Level 1 35](#_Toc60505499)

[Gambar 3.5 *Flowchart* Sistem 36](#_Toc60505501)

[Gambar 3.6 *Flowchart Preprocessing* Dengan *Stopword* 37](#_Toc60505503)

[Gambar 3.7 *Flowchart* *Prerocessing* Tanpa *Stopword* 39](#_Toc60505506)

[Gambar 3.8 *Flwochart* *Cleansing* 40](#_Toc60505509)

[Gambar 3.9 *Flowchart Slangword* 42](#_Toc60505512)

[Gambar 3.10 *Flowchart* *Stopwrod Removal* 43](#_Toc60505515)

[Gambar 3.11 *Flowchart Stemming* 44](#_Toc60505518)

[Gambar 3.12 *Flowchart Tokenizing* 45](#_Toc60505521)

[Gambar 3.13 *Flowchart Training Naïve Bayes* 46](#_Toc60505524)

[Gambar 3.14 *Flowchart klasifikasi Naïve Bayes* 47](#_Toc60505525)

[Gambar 3.15 Rancangan ERD 49](#_Toc60505528)

[Gambar 3.16 Rancangan RAT 49](#_Toc60505530)

[Gambar 3.17 Rancangan Halaman *Login* 53](#_Toc60505541)

[Gambar 3.18 Rancangan Halaman *Dashboard* 54](#_Toc60505542)

[Gambar 3.19 Rancangan Halaman Data Kamus 54](#_Toc60505543)

[Gambar 3.20 Rancangan Halaman Tambah Data Kamus 55](#_Toc60505544)

[Gambar 3.21 Rancangan Halaman Data *Stopword* 55](#_Toc60505545)

[Gambar 3.22 Rancangan Halaman Tambah Data *Stopword* 56](#_Toc60505546)

[Gambar 3.23 Rancangan Halaman Edit Data *Stopword* 56](#_Toc60505547)

[Gambar 3.24 Rancangan Halaman Data *Slangword* 57](#_Toc60505548)

[Gambar 3.25 Rancangan Halaman Data *Latih* 57](#_Toc60505549)

[Gambar 3.26 Rancangan Halaman Tambah Data *Latih* 58](#_Toc60505550)

[Gambar 3.27 Rancangan Halaman Data Uji 58](#_Toc60505551)

[Gambar 3.28 Rancangan Halaman Data Admin 59](#_Toc60505552)

[Gambar 3.29 Rancangan Halaman Analisis 60](#_Toc60505553)

[Gambar 3.30 Rancangan Halaman Hasil Analisis 60](#_Toc60505554)

[Gambar 3.31 Rancangan Halaman Grafik Analisis 61](#_Toc60505556)

[Gambar 3.32 Rancangan Halaman Grafik Analisis 61](#_Toc60505557)

# DAFTAR TABEL

[Tabel 2.1 *Confusion* *Matrix* 15](#_Toc60505972)

[Tabel 2.2 Studi Pustaka 18](#_Toc60505974)

[Tabel 2.3 Lanjutan Tabel Studi Pustaka 19](#_Toc60505975)

[Tabel 3.1 Contoh Data Hasil *Scraping* 23](#_Toc60505982)

[Tabel 3.2 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak 25](#_Toc60505990)

[Tabel 3.3 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak 26](#_Toc60505991)

[Tabel 3.4 Contoh Data Beserta Label 26](#_Toc60505994)

[Tabel 3.5 Data Ulasan Setelah Proses *Text Processing* 27](#_Toc60505996)

[Tabel 3.6 *Term Frequency* Data Sentimen 27](#_Toc60505997)

[Tabel 3.7 Lanjutan *Term Frequency* Data Sentimen 28](#_Toc60505998)

[Tabel 3.8. Dokuem Kategori Positif 29](#_Toc60506000)

[Tebel 3.9 Dokumen Kategori Negatif 30](#_Toc60506001)

[Tabel 3.10 Contoh Penerapan *Preprocessing* Dengan *Stopword* 38](#_Toc60506014)

[Tabel 3.11 Contoh Penerapan *Preprocessing* Tanpa *Stopword* 40](#_Toc60506017)

[Tabel 3.12 Contoh Penerapan *Cleansing* 41](#_Toc60506020)

[Tabel 3.13 Contoh Penerapan *Slangword* 43](#_Toc60506023)

[Tabel 3.14 Contoh Penerapan *Stopword Removal* 44](#_Toc60506026)

[Tabel 3.15 Contoh Penerapan *Stemming* 45](#_Toc60506029)

[Tabel 3.16 Contoh Penerapan *Tokenizing* 46](#_Toc60506032)

[Tabel 3.17 Struktur Tabel Data Admin 50](#_Toc60506042)

[Tabel 3.18 Struktur Tabel Data *Stopword* 50](#_Toc60506043)

[Tabel 3.19 Struktur Tabel Data Kata Dasar 51](#_Toc60506044)

[Tabel 3.20 Struktur Tabel Data *Slangword* 51](#_Toc60506045)

[Tabel 3.21 Struktur Tabel Data Kategori 51](#_Toc60506046)

[Tabel 3.22. Struktur Tabel Hasil analisis 52](#_Toc60506047)

[Tabel 3.23 Struktur Tabel Data Latih 52](#_Toc60506048)

[Tabel 3.24 Struktur Tabel Data Uji 52](#_Toc60506049)

[Tabel 3.25 Tabel *Confusion Matrix* 62](#_Toc60506069)

[Tabel 3.26 Tabel *Confusion Matrix* 62](#_Toc60506070)

[Tabel 3.27 Rancangan Pengujian *Black Box Testing* 63](#_Toc60506071)

# BAB I

**PENDAHULUAN**

## Latar Belakang

*Google maps* adalah layanan aplikasi peta online yang disediakan oleh perusahaan *google*. *Google maps* mempunyai fitur ulasan yang dapat dipergunakan untuk menuliskan opini mengenai sebuah tempat (Wilianto et al., 2017). Ulasan pada *google maps* dapat dimanfaatkan untuk mengetahui sentimen opini pengunjung terhadap sebuah tempat (Haq et al., 2020). Salah satu pemanfatan data ulasan pada *google maps* adalah dengan melakukan analisis sentimen.

Analisis Sentimen merupakan proses memahami, mengesktrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu opini kalimat. Dari analisis sentimen dapat diketahui pendapat tentang suatu topik tersebut bersifat positif atau negatif (Wang et al., 2018). Namun sebagian besar data ulasan *google maps* merupakan data yang masih kotor dan mengandung banyak *noise* di dalamnya, sehingga sebelum dilakukan analisis sentimen, dataset ulasan terlebih dahulu dilakukan *text processing*. Peran *text processing* data sangat penting karena dapat mempengaruhi keakuratan data (Symeonidis et al., 2018). *Text processing* bertujuan untuk mempersiapkan dokumen teks yang tidak terstruktur menjadi data terstruktur yang siap digunakan untuk proses selanjutnya, salah satu tahapan *text processing* adalah *stopword* (Antinasari et al., 2017)*.*

*Stopword* merupakan proses pengambilan kata-kata yang dianggap penting atau mempunyai makna saja, dan membuang kata-kata yang cukup umum dan sering muncul. Penelitian sebelumnya oleh Saif (2018) telah dilakukan penelitian tentang perngaruh *stopword* terhadap klasifikasi sentimen pada dokumen bahasa Inggris. Hasil penelitian menunjukan bahwa penggunaan *stopword* berdampak negatif menurunkan performa klasifikasi sentimen Twitter, mengurangi runag fitur 65%, dan menurunkan ketersebaran data hingga 0,37%. Penelitian lain oleh Jayashree & Murty (2014) telah dilakukan penelitan tentang pengaruh *stopword* terhadap kinerja metode *naïve bayes* untuk klasifikasi dokumen bahasa Kanada. Hasil pengujian dengan *k-fold cross* *validaton* menunjukan penggunaan *stopword* bepengaruh meningkatkan akurasi, dengan rata-rata akurasi 61% tanpa menggunakan *stopword* dan 69% dengan menggunakan *stopword*. Namun, apakah penggunaan *stopword* juga dapat berpengaruh terhadap performa analisis sentimen dalam dokumen bahasa Indonesia, mengingat suatu kata memiliki arti yang berbeda dalam bahasa yang digunakan. Lebih lanjut penelitian ini akan menganalisa pengaruh *stopword* terhadap performa analisis sentimen, hal ini betujuan untuk mengetahui signifikansi tahapan *stopword* terhadap analisis sentimen.

Penelitian yang berkaitan analisis sentimen pernah dilakukan oleh Mutia (2017) terhadap *review* restoran dengan menggunakan metode *Naïve Bayes Classifier*, pada penelitian ini mengguakan dataset sejumlah 200 dibagi menjadi 100 data positif dan 100 data negatif. Pengujian dilakukan dengan *k-fold cross validation* dan *confusion matrix*, dengan akurasi mencapai 86.50%.

Penelitian ini akan menggunakan metode *Naïve Bayes*. Pemilihan penggunaan metode *Naïve Bayes* karena metode *Naïve Bayes* merupakan teknik *machine learning* yang efisien untuk klasifikasi teks dengan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian., selain itu *Naïve Bayes* merupakan metode *machine learning* yang mimiliki tingkat akurasi dan performasi yang baik dalam klasifikasi teks (Suryadi, 2017). Metode *naïve bayes* juga dapat diaplikasikan pada berbagai keadaan sehingga tidak akan mengalami kegagalan secara keseluruhan pada hasilnya (Yunitasari et al., 2019). Penelitian oleh Buntoro (2017), pada penelitian ini dilakuakn analisis sentimen pada Twitter dengan menggunakan 300 dataset. Pengujian dilakukan dengan *k-fold cross validation* dan *confusin matrix*, metode *naïve bayes* menunjukan akurasi yang lebih baik dengan rata-rata akurasi 95% sedangkan metode *support vector machine* rata-rata akurasi 90%. Namun penerapan metode *naïve bayes classifier* pada sebuah data terkadang juga menyebabkan *missclassification* jika pada data uji tidak ditemukan pada data latih, sehingga hasil probabilitas menjadi bernilai 0 (*zero*) dan menyebabkan *error* pada proses klasifikasi. Kekurangan pada *naïve bayes classifier* dapat diminimaisir dengan menggunakan metode *laplacian* (Listiowarni, 2019)*,* sehingga nilai probabilitas tidak berniali 0 *(Zero)*.

Berdasarkan paparan diatas, penelitian ini menitikberatkan pada implementasi model untuk menganalisa pengaruh *stopword* terhadap performa analisis sentimen menggunakan metode *naïve bayes classifier*. Hal ini dilakukan untuk mengetahui signifikasinsi tahapan *stopword* terhadap analisis sentimen pada ulasan *google maps* dalam dokumen bahasa Indonesia*,* serta mengolah data ulasan agar dapat memberikan informasi yang lebih bermanfaat, yaitu dengan melakukan analisis sentimen terhadap setiap ulasan untuk melihat sentimen opini mengenai sebuah tempat.

## Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat diambil rumusan masalah sebagai berikut:

1. Bagaimana melakakukan analisis sentimen dengan *stopword* dan tanpa *stopword* terhadap opini dari ulasan *google maps* menggunakan metode *Naïve Bayes Classifier*?
2. Bagaimana pengaruh *stopword* terhadap performa analisis analisis sentimen menggunakan metode *Naïve Bayes Classifier*?

## Batasan Masalah

Pada penelitian ini, permasalahan dibatasi sebagai berikut:

1. Data penelitian didapatkan dari ulasan *google maps*.
2. Data penelitian yang digunakan berupa ulasan berbahasa Indonesia.
3. Untuk analisis sentimen hanya menggunakan 2 kelas, yaitu negatif dan positif.
4. Label diberikan hanya salah satu yang paling dominan, bukan multilabel.

## Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut:

1. Menerapkan pemrosesan awal dengan *stopword* dan tanpa *stopword* pada analisis sentimen ulasan *google maps* menggunakan metode *Naïve Bayes Classifier*.
2. Menganalisis pengaruh *stopword* terhadap performa analisis sentimen menggunakan metode *Naïve Bayes Classifier*.

## Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah dapat mengetahui pengaruh *stopword* terhadap performa analisis sentimen pada ulasan *google maps* menggunakan metode *naïve bayes classifier*, serta menganalisa sentimen opini terhadap sebuah tempat pada ulasan *google maps*, sehingga dapat diketaui bagaimana sentimen opini terhadap sebuah tempat.

## Metode Penelitian

Metodologi penelitian yang digunakan untuk pengumpulan data pada penelitian ini adalah metode penelitian kualitatif. Metode penelitian kualitatif menggunakan data berupa teori-teori dan memanfaatkan teori yang ada untuk menjadi dasar penguat argumentasi.

## 1.6.1 Metode Pengumpulan Data

Tahapan pengumpulan data pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan cara mencari dan menghimpun data atau sumber-sumber pustaka yang dapat mendukung penelitian serta memberikan informasi untuk menyelesaikan permasalahan pada penelitian ini. Studi kepustakaan yang digunakan bersumber dari buku, jurnal, artikel dan paper yang berkaitan pada penelitian ini.

1. Pengumpulan Data

Data yang akan digunakan bersumber dari ulasan *google maps* dengan menggunakan teknik *web scraping*. Selanjutnya dilakukan proses pemberian label secara manual terhadap data yang sudah dikumpulkan.

## 1.6.2 Metode Pengembangan Sistem

Metodologi pengembangan sistem yang digunakan adalah metode *prototype*. Penggunaan metode *prototype* dikarenakan metode ini lebih menekankan pada komunikasi antara pengembang dan pelanggan sehingga kebutuhan pengguna dapat diterjemahkan dalam bentuk model dan hasil yang diharapkan dapat tercapai. Ada beberapa tahapan yang dilakukan pada metode *prototype* yaitu:

1. *Communication*

Pada tahapan ini dilakukan proses komunikasi antara pengembang dan pelanggan mengenai tujuan dibuatnya suatu perangkat lunak. Selain itu, akan dilakukan analisis serta identifikasi kebutuhan apa saja yang akan diperlukan selama proses pembuatan perangkat lunak

1. *Qiuck Plan* dan *Modeling Quick Design*

Pada bagian ini akan membuat perencanaan dan pemodelan secara cepat berdasarkan kebutuhan yang telah diidentifikasi pada tahap sebelumnya. Perencanaan yang akan dilakukan dapat berupa pembagian tugas dan rencana kerja. Sedangkan dari sisi pemodelan akan membuat desain model yang merepresentasikan aspek-aspek yang diinginkan dari pelanggan.

1. *Construction* *of* *Prototype*

Pada proses *contruction* *of* *prototype*, pengembang akan mulai membuat program (perangkat lunak) berdasarkan rencana dan model yang telah dibuat sebelumnya. Pada tahap ini juga termasuk testing dari perangkat lunak yang telah berhasil dibuat.

1. *Deployment Delivery & Feedback*

Setelah perangkat lunak berhasil dibuat dan telah melalui testing, proses selanjutnya yaitu tahap *deployment delivery & feedback*. Perangkat lunak tadi akan disampaikan kepada pelanggan untuk dilakukan pengecekan. Jika mendapatkan *feedback* baik, maka perangkat lunak akan diterima oleh pelanggan dan pembuatan perangkat lunak telah selesai. Apabila *feedback* yang diberikan kurang baik atau tidak sesuai dengan pelanggan, maka kembali ke tahap communication dengan membahas perbaikan yang perlu dilakukan oleh pengembang.

## Sistematika Penelitian

Sistematika penulisan yang digunakan dalam menyusun laporan penelitian ini adalah sebagai berikut :

**Bab I Pendahuluan**

Pada bagian ini membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

**Bab II Tinjauan Pustaka**

Tinjauan pustaka memuat tentang karya ilmiah yang sudah ada sebagai bahan referensi terkini dan pondasi dalam membangun sistem ini. Selain itu juga dalam bab ini memuat teori-teori yang digunakan dalam penelitian ini.

**Bab III Metodologi Penelitian dan Pengembangan Sistem**

Bab ini membahas tentang metode penelitian yang berisi tentang penjelasan mengenai langkah – langkah yang harus dilakukan untuk mencapai tujuan akhir serta pengembangan perangkat lunak hingga tahap testing fungsional.

**Bab IV Hasil, Pengujian, dan Pembahasan**

Pada bab ini memuat mengenai detail proses implementasi sistem beserta hal – hal yang mendukung dalam pembangunan sistem seperti proses instalasi dan konfigurasi software yang digunakan. Bab ini juga akan dilakukan pengujian sistem yang telah dibangun untuk mengevaluasi kinerja sistem, mengurangi adanya kesalahan, setta memastikan apakah sistem dapat digunakan dengan baik.

**Bab V Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan. Kesimpulan diambil berdasarkan hasil dan pembahasan pada bab sebelumnya. Selanjutnya, kekurangan yang ada pada sistem akan dituliskan pada bagian saran untuk pengembangan penelitian selanjutnya.

# BAB II

**TINJAUAN PUSTAKA**

## 2.1 Ulasan *Google Maps*

*Google maps* adalah *globe virtual online* yang disediakan oleh perusahaan *google*. *Google maps* juga mempunyai fitur ulasan sehingga penggunanya secara langsung dapat menambahkan ulasan pada tempat yang dikunjunginya (Wilianto et al., 2017). Ulasan *google maps* merupakan salah satu fitur yang dibuat oleh *google* untuk seluruh alamat yang tercantum dalam peta *google*. Biasanya ulasan *google* *maps* muncul saat pengguna mencari sebuah tempat atau setelah pengguna mengunjungi suatu tempat. Ulasan juga dapat dilihat saat pengguna mencari sebuah tempat. Ketika seseorang mencari destinasi yang dituju, ulasan dalam bentuk ulasan kata akan muncul. Disana pengguna dapat melihat sejauh mana orang-orang yang berkunjung ke tempat tersebut puas terhadap apa yang dilayani (Haq et al., 2020). Ulasan yang diisi oleh pengguna *google* *maps* dapat berupa pujian, kritik dan juga saran untuk lokasi yang sedang dikunjungi.

## 2.2 Analisis Sentimen dan *Text Mining*

Analisis sentimen merupakan salah satu ilmu yang melibatkan *text mining*, *natural language program*, dan *artificial intelegence*. Proses yang dilakukan oleh analisis sentimen untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat (Akbari et al., 2017).Analisis sentimen digunakan untuk memahami komentar/pendapat pengguna internet/sosial media dan menjelaskan bagaimana sebuah produk, brand, objek ataupun subjek diterima oleh mereka (Rizal, 2017). Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif.

Analisis sentimen bisa dianggap sebagai kombinasi dari *text mining* dengan *natural langague processing*. *Text mining* merupakan cabang ilmu dari *data mining*, dimana *data mining* merupakan proses unttuk mencari sebuah pola pada kumpulan data. Pada dasarnya proses kerja dari *text mining* banyak mengadopsi penelitian data mining, namun yang terjadi perbedaaan adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alamai yang tidak terstruktur, sedangkan dalam *data mining* pola yang diambil adalah dari data yang terstruktur.

*Text mining* adalah salah satu teknik yang digunakan untuk melakukan klasifikasi data, dimana *text mining* berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar (Ratniasih et al., 2017). Proses *text mining* pada penelitian ini diawali dengan melakukan *preprocessing* pada dokumen teks. Selanjutnya dokumen teks yang telah di *preprocessing* akan diubah menjadi data berupa angka. Proses terakhir pada *text mining* adalah evaluasi.

## 2.3 *Web Scraping*

*Web scraping* adalah proses untuk mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari *web scraping* adalah untuk mencari informasi tertentu dan kemudian mengumpulkannya dalam *web* yang baru. *Web scraping* berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Manfaat dari *web scraping* adalah agar informasi yang diambil lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu. Aplikasi *web scraping* hanya fokus pada cara untuk memperoleh data melalui pengambilan dan ekstraksi data dengan ukuran data yang berfarisai (A. Yani et al., 2019).

*Web scraping* dapat menggunakan teknik dengan bahasa pemrograman dan tidak dengan bahasa pemrograman, dimana *scraping* secara otomatis memiliki beberapa cara yaitu menggunakan *coding*, aplikasi, atau *extension browser*. Pada penelitian ini digunakan teknik dengan bahasa pemrograman (*coding*) yaitu dengan *library selenium*.

*Selenium* adalah *tools* untuk *automation browser*, singkatnya *browser* akan dijalankan secara otomatis melalui program yang dirancang. *Selenium* dapat berjalan secara otomatis dibutuhkan *WebDriver* sesuai dengan *browser* yang digunakan. Setiap *browser* memiliki *driver* masing-masing. *Driver* inilah yang akan digunakan oleh *selenium* untuk menjalankan *browser* secara otomatis.

## 2.4 *Text Processing*

*Text processing* merupakan tahap awal dari *text mining*. Pada *text mining* dibutuhkan beberapa tahapan untuk mengolah teks/dokumen agar menjadi lebih terstruktur. *Text* *processing* adalah tahapan dimana data teks yang dikumpulkan disiapkan agar menjadi data teks yang siap untuk dianalisis (Antinasari et al., 2017). Tahapan *preprocessing* yang digunakan pada penilitian ini yaitu:

## 2.4.1 *Cleansing*

Tahap *cleansing* merupakan tahapan yang dilakukan untuk menghilangkan *noise* atau elemen-elemen yang tidak diperlukan dalam suatu data atau kalimat. Tujuan dilakukannya *cleansing* adalah untuk meningkatkan dan memperbaiki kualitas data (Himawan et al., 2018). Tahapan ini dapat mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter* atau pembatas (Shafira, 2018). Selain itu, menghapus tanda baca atau karakter simbol seperti [! “ # $ % & ‘ () \* + , - . / : ; < = > ? @ [ ] \ ^ \_ ` { } | ~], menghapus karakter angka atau digit, dan menghapus spasi yang terdapat kemungkinan suatu kalimat memiliki spasi lebih dari satu.

## 2.4.2 *Slangword*

*Slangword* atau perbaikan kata tidak baku merupakan proses yang digunakan untuk mengubah kata-kata yang tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia. Pada perbaikan kata tidak baku terdapat bermacam-macam kata yang dianggap tidak baku, seperti penulisan dengan bahasa modern atau slang (ngga = tidak), penulisan dengan kaidah atau pedoman bahasa (kwalitas = kualitas), penulisan dengan kata singkat (tdk = tidak), dan penulisan dengan kata salah eja (jelsk = jelek). Pada penelitian ini digunakan perbaikan kata tidak baku untuk menangani permasalahan kata tidak baku hanya pada bahasa modern atau slang, penyingkatan kata, dan salah eja saja. Jika terdapat kata tidak baku makan akan diubah menjadi kata baku sesuai dengan kamus. Sehingga didapatkan hasil akhir berupa kalimat dengan menggunakan kata baku.

## 2.4.3 *Stopword* *Removal*

Tahapan *stopword removal* merupakan proses pengambilan kata-kata penting atau membuang kata-kata yang kurang penting dari token yang telah dihasilkan pada tahapan sebelumnya (Nurrohmat & SN, 2019). Pembuangan *stopword* tidak akan mengubah makna dan isi dari suatu data. Bebrapa contoh *stopword* seperti aku, juga, kamu, dan, itu, ini. Pada penelitian ini digunakan *stopword* dari *library* Sastrawi.

## 2.4.4 *Stemming*

Tahap *stemming* merupakan proses pemotongan imbuhan pada kata berimbuhan yang dilakukan dengan algoritma tertentu untuk mengembalikan suatu kata kebentuk dasarnya. Pada penelitian ini proses *stemming* yang dilakukan hanya akan menangani kata yang memiliki penambahan awalan dan akhiran. Hal ini dilakukan karena proses penambahan *infiks* dalam bahasa Indonesia jarang terjadi sehingga tidak ada pengaruh yang signifikan terhadap akurasi sistem (Mahmudy & Widodo, 2014).

Terdapat beberapa metode yang dapat digunakan untuk proses *stemming* pada dokumen berbahasa Indonesia salah satunya adalah algoritma Nazief dan Adriani. Algoritma ini bekerja atas dasar aturan-aturan yang mengelompokkan imbuhan yang diperbolehkan dan dilarang untuk digunakan. (Zulkifli et al., 2010).

Algoritma Nazief dan Adriani adalah algoritma stemming untuk bahasa Indonesia. Algoritma ini memiliki beberapa aturan morfologi untuk merubah suatu kata menjadi kata dasar yaitu dengan menghilangkan affiks (awalan, imbuhan dan lain-lain) dan kemudian dicocokkan dengan kamus kata dasar. Kamus kata dasar merupakan hal yang paling utama pada algoritma ini karena semakin banyak kamus kata dasarnya maka semakin tinggi pula akurasinya.

## 2.4.5 *Tokenizing*

Tahap *tokenizing* adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusunnya (Murnawan, 2017). Proses tokenizing biasanya berpatokan pada karakter whitespace seperti spasi, enter ataupun tab sebagai pemisah atau pemoting antar kata. Tiap kata yang akan dihasilkan pada proses ini akan disebut sebagai token. Stiap kata teridentifikasi dengan kata yang lain oleh karakter spasi, shingga proses *tokenizing* akan mengendalikan karakter spasi pada dokumen untuk pemisah kata.

## 2.4.6 Ekstraksi Fitur *Term Frequency (TF)*

Ekstraksi fitur atau pembobotan kata adalah suatu proses ekstraksi untuk mengidentifikasikan entitas-entitas yang dimaksud. Pada ekstraksi fitur bertujuan untuk memberikan bobot pada fitur-fitur kata berdasarkan frekuensi kemunculan kata, penelitian ini akan menggunakan ekstraksi fitur *Term Frequency (TF)*. *Term Frequency* *(TF)* merupakan suatu metode yang digunakan untuk mencari bobot suatu kata dalam dokumen kunci di setiap kategori dan mencari kata kunci yang hampir mirip dengan kategori yang tersedia, dimana kata akan disususn menjadi *vector* (Saputro et al., 2017). Jika semakin besar kemunculan suatu term dalam dokumen, maka akan memeberikan bobot atau nilai kesesuaian yang semakin besar.

## 2.5 *Naïve Bayes Clasiffier*

Metode *Naïve Bayes Classifier* adalah salah satu metode yang digunakan untuk keperluan data mining karena penggunaannya tidak terlalu rumit dan memiliki waktu pemrosesan yang relative cepat serta lebih mudah diimplementasikan dan memiliki efektifitas yang tinggi (Antinasari et al., 2017). Metode ini merupakan suatu metode *probabilistic* sederhana yang menerapkan Teorema Bayes. Dengan perumusan sebagai berikut:

……………………………………………….……. (2.1)

Keteranngan:

P(A|B) : Peluang kejadian A jika diketahui B

P(B|A) : Peluang kejadian B jika diketahui A

P(A) : Peluang kejadian A

P(B) : Peluang kejadian B

Model *Multinominal Naïve Bayes* dijelaskan oleh (Manning et al., 2009) bahwa probabilitas suatu dokumen d dalam kelas c dihitung dengan persamaan (2.2).

𝑃(𝑐|𝑑) ∝ 𝑃(𝑐) ∏1≤𝑘≤𝑛𝑑 𝑃 (𝑡𝑘 | 𝑐) .................................................................. (2.2)

Pada persamaan (2.2) untuk𝑃(𝑡𝑘|𝑐) merupakan sebuah probabilitas dari *term* *tk* yang terdapat didalam dokumen kelas c. Sedangkan keterangan untuk *nd* adalah jumlah token pada kelas d. Selanjutnya untuk pencarian kelas terbaik dari hasil klasifikasi *Naïve Bayes* dapat diperolah dari *Maximum* *a Posteriori* (MAP) Kelas Cmap (Manning et al., 2009) yang dinyatakan dalam persamaan (2.3).

𝐶𝑚𝑎𝑝 = 𝑎𝑟𝑔𝑚𝑎𝑥𝑐 ∈ 𝐶𝑃̂ (𝑐|𝑑) = 𝑎𝑟𝑔𝑚𝑎𝑥𝑐 ∈ 𝐶𝑃̂ (𝑐) ∏1≤𝑘≤𝑛𝑑 𝑃̂ (𝑡𝑘|𝑐) ….……………… (2.3)

Berdasarkan persamaan (2.3) nilai 𝑃̂ (𝑐) dan 𝑃(𝑡𝑘|𝑐) adalah nilai yang didapatkan dari data pelatihan dan nilai 𝑃̂ (𝑐) diperoleh dari persamaan (2.4) untuk fitur *term frequency.*

𝑃̂ (𝑐) = ………………………………………………………………………………….. (2.4)

Keterangan:

P(c) : Probabilitas prior kemunculan suatu dokumen didalam kelas c

Nc : Jumlah dokumen dalam kelas c

N’ : Jumlah seluruh dokumen

Menurut Manning, dkk (2009) perhitungan atau probabilitas kondisional frekuensi relative pada term di kelas c atau 𝑃̂(𝑡𝑘|𝑐) dapat dihitung dengan persamaan (2.5).

…………………………………………………… (2.5)

Keterangan:

Tct : Jumlah kemunculan suatu term dokumen kelas c

Tct` : Jumlah kemunculan semua term pada kelas c dalam data pelatihan n

## 2.6 *Laplacian Smoothing*

Metode *laplacian* *smoothing* merupakan metode untuk menghindari hasil klasifikasi yang bernilai 0 dikarenakan data uji tidak ditemukan pada data latih. Pada penelitian ini digunakan metode *laplacian* *smoothing*, yaitu merode yang paling sederhana dengan menambahkan angka 1, tapi metode laplacian smoothing memiliki performance yang cukup baik dibandingkan dengna metode smoothing lainnya (Listiowarni, 2019). Sehingga untuk mengantisipasi kemunculan nilai 0 pada persamaan (2.5) dilakukan penambahan nilai suatu yang biasanya disebut *laplance smoothing* seperti persamaan (2.6).

………………………………………………..……. (2.6)

Pada persamaan (2.6) nilai B` merupakan jumlah term dalam *vocabulary*.

## 2.6 Pengujian

Pengujian merupakan suatu langkah untuk mengetahui proses klasifikasi yang telah dilakukan. Pengujian dilakukan menggunakan *k-fold cross validation* dan *Confusion Matrix*.

## 2.6.1 *K-Fold Cross Validation*

*K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengukur rata-rata tingkat keberhasilan dari suatu model. Hasil tersebut dilakukan dengan melakukan perulangan untuk menguji model dengan data acak. Awalnya, dataset yang dimiliki akan dibagi sejumlah *n-fold* yang diinginkan sehingga membentuk kelompok secara acak. Lalu akan dilakukan perulangan terhadap data yang telah dibagi menjadi data uji dan data latih secara bergantikan (Buntoro, 2017). Sehingga setiap kelompok pernah menjadi data latih dan data uji. Berikut adalah ilustrasi dari proses *k-fold cross validation* pada Gambar 2.6.

#### E:\b04c27c5-7e3f-428a-9aa6-bb3ebcd3584c.pngGambar 3.2 *K-Fold Cross Validation*

## 2.6.2 *Confusion Matrix*

*Confusion matrix* merupakan metode yang sering digunakan untuk mengukur tingkat performa dari suatu model klasifikasi. *Confusion matrix* akan memberikan hasil perbandingan antara hasil klasifikasi yang dilakukan oleh model dengan hasil klasifikasi yang sebenarnya. Terdapat empat istilah yang merepresentasikan hasil dari proses klasifikasi pada *confusion matrix* yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Tabel *confusion matrix* dapat dilihat pada Tabel 2.4.

#### Tabel 2.1 *Confusion Matrix*

|  |  |  |
| --- | --- | --- |
| ***Predicted***  ***Values*** | ***Actual Values*** | |
| ***Positif*** | ***Negative*** |
| *Positive* | *True Positive* (TP) | *True Positive* (FP) |
| *Negative* | *False Negative* (FN) | *True Negative* (TN) |

Keterangan:

*True Positive* (TP) : Jumlah data positif yang diprediksi benar

*True Negative* (TN) : Jumlah data negatif yang diprediksi benar

*False Positive* (FP) : Jumlah data negatif yang diprediksi salah

*False Negative* (FN) : Jumlah data positif yang diprediksi salah

Berdasarkan tebel *confusion matrix* tersebut dapat menghitung *performance matrix* untuk mengukur model yang telah dibuat. Berikut perhitungan akurasi, presisi dan *recall* menurut (Han et al., 2014):

**a. Akurasi**

Akurasi merupakan nilai yang menggambarkan seberapa akurat sebuah model dalam mengklasifikasikan dengan benar. Nilai akurasi ini mempresentasikan nilai kedekatan antara nilai prediksi dan nilai yang sebenarnya. Persamaan untuk menghasilkan nilai akurasi yaitu dengan membandingkan nilai prediksi benar dengan keseluruhan data. Rumus menghitung nilai akurasi dapat dilihat pada persamaan 2.15.

*X* 100% ………………………………………… (2.7)

**b. Presisi**

Presisi merupakan proporsi jumlah dokumen teks yang relevan terkenali diantara semua dokumen teks yang dipilih oleh terpilih oleh sistem. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan jumlah data yang diklasifikasikan positif. Untuk menghitung presisi dapat dilihat pada persamaan 2.8.

%…………………………………………………(2.8)

**c. *Recall***

*Recall* merupakan perbandingan antara prediksi benar positif dengan keseluruhan data yang benar positif. Untuk menghitung nilai recall dapat menggunakan rumus yang dapat dilihat pada peprsamaan 2.18.

…………………………………………………... (2.9)

## 2.7 Studi Pustaka *(State Of Art)*

Keaslian penelitian yang dilakukan pada penelitihan terdahulu yang mempunyai karakteristik relatif sama dalam hal tema, objek, dan metode yang digunakan.

Penelitian yang dilakukan oleh (Utami, 2018) dengan judul penelitian Analisis Sentimen Review Kosmetik Bahasa Indonesia Menggunakan *Algoritma Naïve Bayes Classifier* dimana penelitian ini bertujuan untuk membuat keputusan dari *review* produk dengan algoritma *Naïve Bayes Classifier*. Pada penelitian ini dilakukan testing dengan data *review* sebanyak 100 untuk menentukan kelas dari *review* tersebut apakah positif atau negatif. *Algoritma Naïve Bayes Classifier* menghasilkan akurasi cukup bagus pada penelitian ini yaitu sebesar 80%. Sistem ini bekerja dengan baik dan lebih akurat apabila pada *review* tersebut terdapat *emoticon*. Kemudian dengan menggunakan *stopword* dan *filtering* pada tahap *text preprocessing* dapat mengurangi kata dan mengurangi perhitungan *Naïve Bayes Classifier,* sehingga hanya kata penting yang akan dihitung dan dapat meningkatkan akurasi.

Penelitian yang dilakuakan oleh (Ghag & Shah, 2016) dengan judul penelitian *Comparative Analysis of Effect of Stopwords Removal on Sentiment Classification* dimana penelitian ini bertujuan untuk menganalisa perbandingan pengaruh penghapusan *stopword* pada klasifikasi sentimen pada dokumen bahasa Inggris. Pada penelitian ini pengklasifikasi sentimen menunjukkan peningkatan dalam akurasi dari 50% menjadi 58,6% saat *stopword* dihapus, sedangkan pendekatan *ARTFSC, TFIDF* dan *RTFSC* bekerja hampir sama. Ini menunjukkan bahwa pendekatan menangani *stopwords* secara efisien pada saat klasifikasi menggunakan nilai batas netral. Tugas *preprocessing* terpisah untuk menghapus *stopwords* tidak diperlukan untuk *algoritma* ini.

Penelitian oleh (Listiowarni, 2019) dengan judul Implementasi *Naïve* *Bayessian* dengan *Laplacian* *Smoothing* untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan, pada penelitian ini menggunakan metode *laplacian* *smoothing*. Metode *laplacian smoothing* adalah metode untuk menghindari hasil klasifikasi menjadi bernilai 0 (zero) dikarenakan data uji tidak ditemukan pada data latih, untuk menghindari nilai 0 pada klasifikasi maka digunakan *metode* *laplacian*. Metode *laplacian* berhasil mengurangi kemungkinan kegagalan klasifikasi, dengan mencegah hasil klasifikasi bernilai 0 *(zero).* Hasil pengujian dengan *confusion matrix* dan *k-fold cross validation* sebanyak 5 kali, didapatkan nilai akurasi 92,11%, dan nilai *error* *rate* 7,02%.

Penelitian yang dilakukan oleh (Haddi et al., 2013) dengan judul penelitian *The Role of Text Pre-processing in Sentiment Analysi*s pada penelitian ini dilakukan analisis sentimen pada ulasan film online dengan dokumen bahasa Inggris. Pada penelitian ini menggunakan kombinasi metode pra-pemrosesan yang berbeda untuk mengurangi *noise* dalam teks selain menggunakan metode *chi-squared* untuk menghapus fitur tidak relevan yang tidak memengaruhi orientasinya. Hasil penelitian ini menunjukan bahwa penggunaan *stemming* dan penghapusan *stopword* meningkatkan akurasi analisis sentimen dalam semua situasi pada dokumen bahasa Inggris.

Pada penelitian yang dilakuakan oleh (Saif et al., 2018) berjudul *On Stopwords, Filteringand Data Sparsityf or Sentiment Analysis of Twitter*. Dalam penelitian ini telah dilakukan penelitian bagaimana penghapusan *stopword* dapat membantu atau menghambat efektivitas metode klasifikasi sentimen *Twitter*. Penelitian ini menerapkan enam metode identifikasi *stopword* yang berbeda ke data Twitter dari enam set data yang berbeda dan mengamati bagaimana menghapus *stopwords* mempengaruhi dua metode klasifikasi sentimen. Hasil penelitian menunjukkan bahwa menggunakan daftar *stopwords* yang telah disusun sebelumnya berdampak negatif terhadap kinerja pendekatan klasifikasi sentimen *Twitter*. Di sisi lain, pembuatan daftar *stopword* yang dinamis, dengan menghapus istilah-istilah yang jarang muncul hanya sekali dalam korpus, tampaknya menjadi metode yang optimal untuk mempertahankan kinerja klasifikasi tinggi dengan mempersempit pengurangan parsitas data dan secara substansial menyusutkan ruang fitur.

Penelitian oleh dengan judul Analisis Sentimen Pada *Review* Restoran Dengan Teks Bahasa Indonesia Menggunakan *Algoritma Naïve Bayes.* Pada penelitian ini dilakukan analisis sentimen pada *riview* restoran menggunakan *Algoritma* *Naïve* *Bayes*. Data yang digunakan pada penelitian ini berupa ulasan restoran dengan teks bahasa Indonesia yang berjumlah 100 *review* positif dan 100 *review* negatif. Preprocessing yang digunakan adalah *Genetic Algorithm*. Pengujian dilakukan dengan 10 *k-fold cross validation* dimana pengujian model dilakukan dengan membagi data secara acak ke dalam 10 bagian, dan *confusion matrix* untuk menampilkan perbandingan dari hasil akurasi model *naïve bayes* sebelum ditambahkan metode pemilihan fitur*.* Hasil pengujian menunjukan akurasi *naïve* *bayes* mencapai 86,50%, dan setelah menggunakan pemilihan fitur meningkat menjadi 90,50%.

Adapun beberapa penelitian yang telah dilakukan dan dijadikan referensi dalam penelitian ini dapat dilihat pada Tabel 2.2.

#### Tabel 2.2 Studi Pustaka

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Penulis** | **Judul** | **Intisari** |
| 1 | Kranti Vithal Ghag, Dr. Ketan Shah, 2016 | *Comparative Analysis of Effectof Stopwords Removal on Sentiment Classification* | Pengklasifikasi sentimen dengan *stopword* menunjukkan peningkatan dalam akurasi dari 50% menjadi 58,6% saat *stopword* dihapus, sedangkan pendekatan *ARTFSC, TFIDF* dan *RTFSC* bekerja hampir sama. Ini menunjukkan bahwa pendekatan menangani *stopwords* secara efisien pada saat klasifikasi menggunakan nilai batas netral. Tugas *preprocessing* terpisah untuk menghapus *stopwords* tidak diperlukan untuk algoritma ini. |
| 2 | Putri Dinda Utami, 2018 | Analisis Sentimen *Review* Kosmetik Bahasa Indonesia Menggunakan *Algoritma Naïve Bayes Classifier* | Sentimen analisis untuk membuat keputusan dari review produk dengan algoritma *Naïve Bayes Classifier* menghasilkan akurasi yang cukup bagus yaitu sebesar 80%. Pada penelitian ini melakukan testing sebanyak 100 *review* produk untuk menentukan kelas dari *review* tersebut apakah positif, atau negatif. Dengan data testing yang terdiri dari *emoticon*, kata positif, dan negatif. |
| 3 | HassanSaif, MiriamFernandez, YulanHe, HarithAlani, 2018 | *On Stopwords, Filtering and Data Sparsityf or Sentiment Analysis of Twitter* | Dalam penelitian ini telah dilakukan apakah menghapus *stopword* dapat membantu atau menghambat efektivitas metode klasifikasi sentimen Twitter. Hasil penelitian menunjukkan bahwa menggunakan daftar *stopwords* yang telah disusun sebelumnya berdampak negatif terhadap kinerja pendekatan klasifikasi sentimen Twitter. Di sisi lain, pembuatan daftar *stopword* yang dinamis, dengan menghapus istilah-istilah yang jarang muncul hanya sekali dalam korpus, tampaknya menjadi metode yang optimal untuk mempertahankan kinerja klasifikasi tinggi dengan mempersempit penguranganparsitas data dan secara substansial menyusutkan ruang fitur. |

#### 

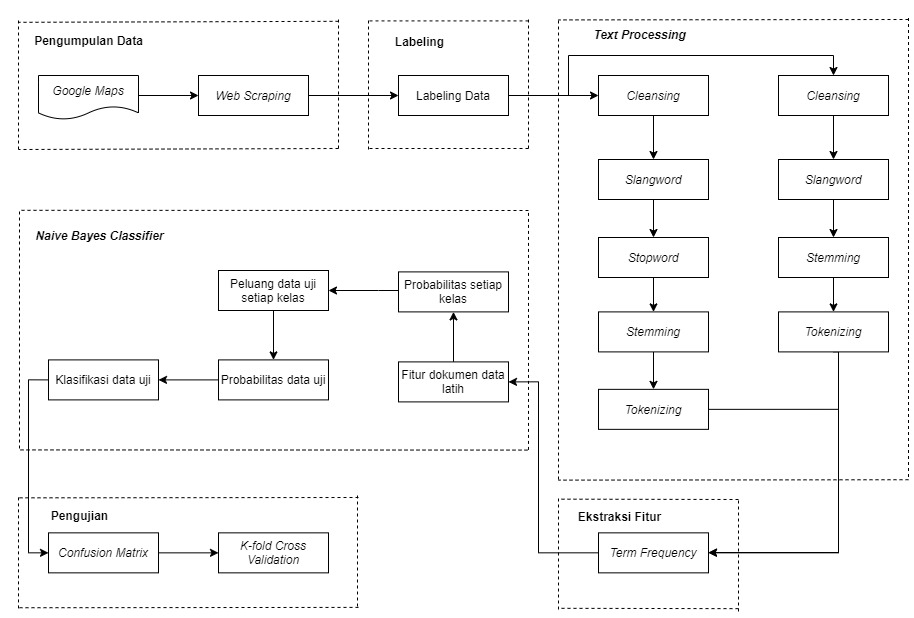
#### Tabel 2.3 Lanjutan Tabel Studi Pustaka

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Penulis** | **Judul** | **Intisari** |
| 4 | Indah Listiowarni, Nilam Ramadhani, 2019 | Implementasi *Naïve* *Bayessian* dengan *Laplacian* *Smoothing* untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan | Pada penelitian ini menggunakan metode *laplacian* *smoothing*. Metode *laplacian* berhasil mengurangi kemungkinan kegagalan klasifikasi, dengan mencegah hasil klasifikasi bernilai 0 *(zero).* Hasil pengujian dengan *confusion matrix* dan *k-fold cross validation* sebanyak 5 kali, didapatkan nilai akurasi 92,11%, dan nilai *error* *rate* 7,02%. |
| 5 | Emma Haddia, Xiaohui Liua, Yong Shib, 2013 | *The Role of Text Pre-processing in Sentiment Analysis* | Pada penelitian ini menggunakan kombinasi metode pra-pemrosesan yang berbeda untuk mengurangi *noise* dalam teks selain menggunakan metode *chi-squared* untuk menghapus fitur tidak relevan yang tidak memengaruhi orientasinya. Hasil penelitian ini menunjukan bahwa penggunaan stemming dan penghapusan *stopwords* meningkatkan akurasi analisis sentimen dalam semua situasi pada dokumen bahasa Inggris. |
| 6 | Harnani Mat Zin, Norwati Mustapha, Masrah Azrifah Azmi Murad, Nurfadhlina Mohd Sharef, 2017 | *The Effects of Pre-Processing Strategies in Sentiment Analysis of Online Movie Reviews* | Dalam penelitian ini, metode svm digunakan untuk mengklasifikasikan review. Performa pengklasifikasi diperiksa berdasarkan hasil presisi, *recall*, f-*measure*, dan akurasi. Dua fitur representasi yang berbeda digunakan yaitu term frekuensi dan term frekuensi dokumen inversi frekuensi. Hasil penelitian menunjukkan bahwa analisis sentimen dengan *preprocessing* memberikan pengaruh yang signifikan terhadap proses klasifikasi. |
| 7 | Dinda Ayu Muthia | Analisis Sentimen Pada *Review* Restoran Dengan Teks Bahasa Indonesia Menggunakan *Algoritma Naïve Bayes* | Pada penelitian ini dilakukan analisis sentimen pada *riview* restoran menggunakan *Algoritma* *Naïve* *Bayes*. Data yang digunakan pada penelitian ini berupa ulasan restoran dengan teks bahasa Indonesia yang berjumlah 100 *review* positif dan 100 *review* negatif. Pengujian dilakukan dengan 10 *k-fold cross validation* dan *confusion matrix.* Hasil pengujian menunjukan akurasi *naïve* *bayes* mencapai 86,50%, dan setelah menggunakan pemilihan fitur meningkat menjadi 90,50%. |

# BAB III

**METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**

## 3.1. Metodologi Penelitian

 Pada bagian ini akan membahas tentang metodologi penelitian yang digunakan pada penelitian ini. Metodologi penelitian berisi tahapan-tahapan perencanaan yang akan dilakukan selama proses pembuatan sistem pada penelitian ini. Metodologi penelitian bertujuan untuk menjadi bahan acuan dalam meminimalkan resiko kegagalan sehingga mencapai tujuan yang diinginkan. Adapun tahapan metodologi penelitian yang dilakukan dapat dilihat pada gambar 3.1.

#### Gambar 3.1 Metode Penelitian

Pada Gambar 3.1 diatas secara garis besar terdapat 6 tahapan yang akan dilakukan yaitu pengumpulan data, labeling, *text processing*, ekstraksi fitur dengan *term frequency*, algoritma *naïve bayes classifier*, serta validasi dan pengujian.

## 3.2. Analisis Pengumpulan Data

Data berperan penting pada suatau penelitian untuk diolah menjadi sebuah informasi yang dapat membantu dalam mengambil keputusan. Analisis pengumpulan data bertujuan untuk memperoleh informasi mengenai apa saja yang dibutuhkan untuk mendukung pembuatan sistem. Berikut adalah tahap pengumpulan data pada penelitian ini:

## 3.2.1. Studi Pustaka

Studi pustaka merupakan salah satu tahapan pengumpulan data yang digunakan untuk mempelajari berbagai sumber referensi dari buku, jurnal, dan sumber lain yang dapat dipercaya dan dipertanggung jawabkan.

Pada tahapan ini dilakukan pencarian informasi mengenai metode dan teknik analisis sentimen yang berkaitan dengan penelitian sebelumnya untuk pengumpulan informasi tentang masalah penelitian ini. Pengumpulan informasi dibutuhkan dalam mencari referensi dan mendapatkan informasi kelebihan dan kekurangan metode dan teknik analisis tersebut dan menjadi bahan evaluasi yang dapat mendukung pengembangan sistem pada penelitian ini.

## 3.2.2. *Scraping Data*

Penelitian ini menggunakan data sekunder yang berasal dari ulasan google maps dengan metode *web scraping*. Penggunaan metode *web scraping* dirasa efektif untuk memudahkan pengambilan data dari *google maps*. Hal ini dikarenakan metode ini mampu untuk mengambil data secara otomatis tanpa harus menyalin data tersebut satu persatu. *Web scraping* memiliki beberapa metode. Pada penelitian ini digunakan metode *extension browser* yang bernama data miner dari *browse google*. Contoh data ulasan hasil *scraping* dapat dilihat pada Tabel 3.1.

#### Tabel 3.1 Contoh Data Hasil *Scraping*

|  |  |
| --- | --- |
| **Doc** | **Ulasan** |
| 1 | Pelayanan sangat cepat dan ramah |
| 2 | Pelayanan ruang tunggu antrian kurang memadahi dan tempat cuci tangan airnya habis. |
| 3 | Tempat parkir motor luas pelayanan baik ramah tapi karena antrian banyak harus sabar menunggu |
| 4 | Pelayanan kurang cepat |
| 5 | Dinas Kependudukan dan Pencacatan Sipil. Pelayanannya cepat dan ramah |

## 3.2.3. Labeling

Labeling merupakan tahapan untuk memberikan label atau kategori pada setiap data. Semua data akan dikategorikan menjadi 2 kelas yaitu positif dan negatif. Untuk mendapatkan objektifitas pada labeling data latih dan data uji maka dilakukan pengklasifikasian kelas sentimen dengan membagikan kuisioner terhadap 7 orang responden terpilih. Berikut contoh hasil labeling dapat dilihat pada Tabel 3.2.

#### Tabel 3.2 Contoh Data Beserta Label

|  |  |  |
| --- | --- | --- |
| **No** | **Ulasan** | **Label Sentimen** |
| 1 | Pelayanan sangat cepat dan ramah | Positif |
| 2 | Pelayanan ruang tunggu antrian kurang memadahi dan tempat cuci tangan airnya habis. | Negatif |
| 3 | Tempat parkir motor luas pelayanan baik ramah tapi karena antrian banyak harus sabar menunggu | Positif |
| 4 | Pelayanan kurang cepat | Negatif |
| 5 | Dinas Kependudukan dan Pencacatan Sipil. Pelayanannya cepat dan ramah | Positif |

Berdasasarkan hasil dari proses labeling dapat diketahui rincian dari data set yang akan digunakan. Data yang didapatkan terdiri dari 110 ulasan yang berkategori positif dan 90 ulasan berkategori negatif. Jadi apabila diakumulasikan maka total dataset yang digunakan pada penelitian ini berjumlah 200 data. Rincian dari jumlah dataset pada penelitian ini dapat dilihat pada Tabel 3.3.

#### Tabel 3.3 Rincian Dataset

|  |  |  |
| --- | --- | --- |
| **No** | **Label Sentimen** | **Jumlah Data** |
| 1 | *Positive* | 110 |
| 2 | *Negative* | 90 |
| **Total Data** | | **200** |

## 3.2.4. *Text Preprocessing* Dengan *Stopword*

Flowchart yang akan ditampilkan merupakan proses seleksi data ulasan yang diinputkan user agar data yang akan digunakan menjadi lebih terstruktur sebelum diklasifikasikan nanti. *Preprocessing* pada penelitian ini terdiri dari beberapa tahapan, yaitu *cleansing, slangword, stopword removal, stemming*, dan *tokenizing*. Pemilihan tahap *preprocessing* bergantung pada kebutuhan setiap penelitian. *Flowchart* dari *preprocessing* dapat dilihat pada Gambar 3.2.

#### Gambar 3.2 *Flowchart Preprocessing* Dengan *Stopword*

Berikut ini adalah contoh hasil *preprocessing* terhadap sebuah ulasan melalui semua tahapan *preprocessing* dapat dilihat pada Tabel 3.4.

#### Tabel 3.4 Contoh Penerapan *Preprocessing* Dengan *Stopword*

|  |  |
| --- | --- |
| **Teks Awal** | **Hasil *Preprocessing* Dengan *Stopword*** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | [‘waktu’, ‘buka’, ‘jam’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘lumayan’, ‘ramai’, ‘layan’, ‘cepat’, ‘instruksi’] |

## 3.2.5. *Text Preprocessing* Tanpa *Stopword*

Alur proses dari *preprocessing* tanpa *stopword* dapat dilihat pada Gambar 3.3. Proses *preprocessing* terhadap data dilakukan untuk menghindari adanya data-data yang tidak konsisten. Tahapan penggunaan *stopword* merupakan proses yang dilakukan untuk menghapus kata yang tidak mempunyai makna atau arti pada kalimat.

Dalam penelitian ini akan dibandingkan hasil dari *preprocessing* dengan penggunaan *stopword* dengan tidak menggunakan *stopword*.

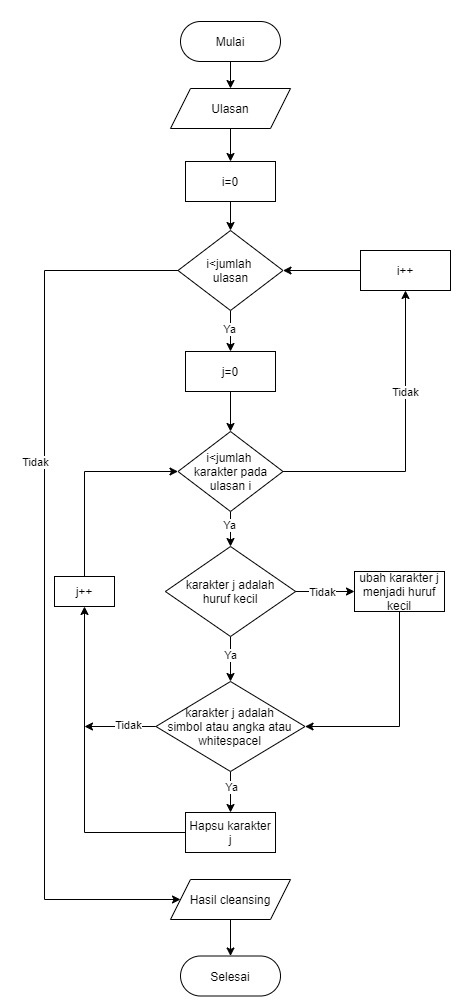
#### Gambar 3.3 *Text* *Prerocessing* Tanpa *Stopword*

Berikut ini adalah contoh hasil *preprocessing* terhadap sebuah ulasan tanpa melalui stopword dapat dilihat pada Tabel 3.5.

#### Tabel 3.5 Contoh Penerapan *Preprocessing* Tanpa *Stopword*

|  |  |
| --- | --- |
| **Teks Awal** | **Hasil *Preprocessing* Dengan *Stopword*** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | [‘datang’, ‘waktu’, ‘buka’, ‘jam’, ‘pagi’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘tidak’, ‘ramai’, ‘layan’, ‘juga’, ‘baik’, ‘cepat’, ‘dan’, ‘jelas’, ‘dalam’, ‘memberi’, ‘instruksi’] |

## 3.2.6. *Celansing*

*Cleansing* adalah proses mengubah semua huruf pada ulasan menjadi huruf kecil *(lower case),* menghapus tanda baca (karakter simbol), angka, dan *whitespace*. Pada *cleansing* ini akan menerima input berupa ulasan dengan inisiasi i yang menjadi jumlah ulasan dan j mewakili jumlah karakter pada ulasan. Prosesnya yaitu melakukan pengecekan ke setiap karakter dari awal hingga akhir. Jika karakter merupakan huruf kecil, maka karakter dilanjutkan ke kondisi selanjutnya. Juka menemukan karakter yang huruf besar maka dirubah terlebih dahulu karakter tersebut menjadi huruf kecil, lalu dilanjutkan ke kondisi selanjutnya. Kondisi selanjutnya melakukan pengecekan karakter kembali, juka karakter merupakan simbol (tanda baca) atau angka atau *whitespace*. *Flowchart* dari proses *cleansing* dapat dilihat pada Gambar 3.4.

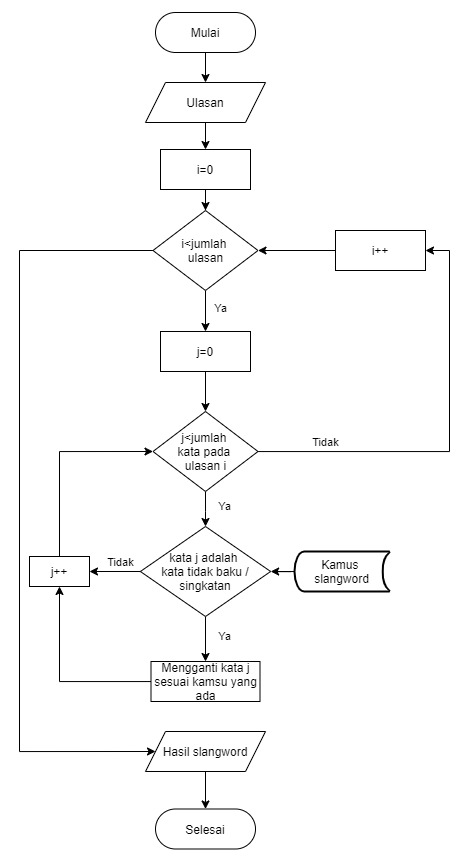
#### Gambar 3.4 Flwochart *Cleansing*

Jika karakter merupakan dalam kondisi tersebut maka karakter tersebut, maka akan dihapus. Tetapi jika bukan dalam kondisi tersebut, maka *flowchart* ini akan memberikan output berupa hasil *cleansing*. Berikut contoh penerapan cleansing terhadap ulasan *google maps* yang akan diproses dapat dilihat pada Tabel 3.6.

#### Tabel 3.6 Contoh Penerapan *Cleansing*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | datang waktu buka jam selesai jam pagi antrinya lumayan nggak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi |

## 3.2.7. *Slangword*

 *Slangword* adalah proses mengganti kata yang tidak baku atau singkatan dari hasil tahapan *preprocessing cleansing* sebelumnya menjadi kata baku atau aslinya. *Flowchart* dari proses *slangword* dapat dilihat pada Gambar 3.5.

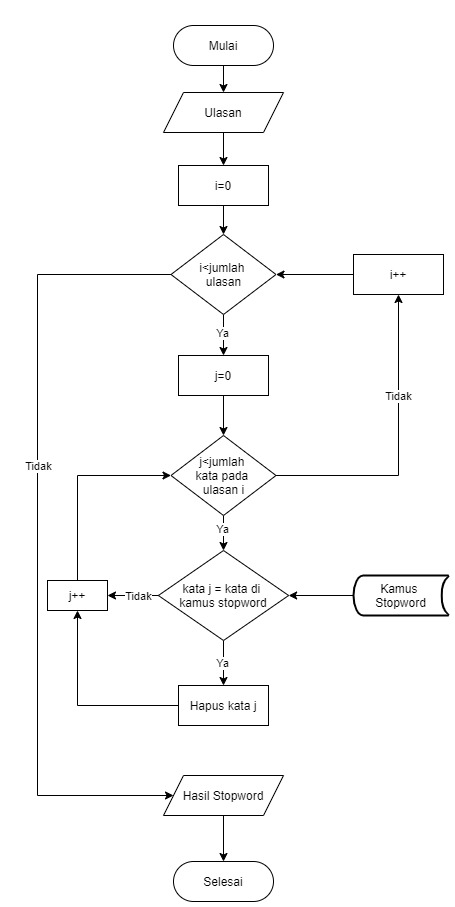
#### Gambar 3.5 *Flowchart Slangword*

Proses *slangword* ini akan menerima input berupa hasil proses sebelumnya yaitu *cleansing*. Dalam proses ini akan dilakukan pengecekan per kata dan menggunakan fungsi sinonim dengan mengganti kata yang sebelumnya sudah dibuat daftar kata sinonim dari kata tidak baku atau singkata yang telah disimpan di dalam *database*. Jika kata tersebut tidak baku atau singkatan seperti dalam kamus, maka akan diganti dengan sinonim dari kata pada kamus tersebut lalu disimpan. Jika kata tersebut baku atau bukan singkatan, maka kata tersebut akan disimpan. Apabila semua kata telah diperiksa, maka *flowchart* ini akan menampilkan *output* berupa *slangword*. Berikut contoh penerapan *slangword* terhadap ulasan *google maps* dapat dilihat pada Tabel 3.7.

#### Tabel 3.7 Contoh Penerapan *Slangword*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| datang waktu buka jam selesai jam pagi antrinya lumayan nggak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi | datang waktu buka jam selesai jam pagi antrinya lumayan tidak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi |

## 3.2.8. *Stopword*

*Stopword* merupakan proses menghapus kata-kata yang sering muncul dan tidak memiliki makna pada kalimat. *Flowchart* dari proses *stopword* dapat dilihat pada Gambar 3.6.

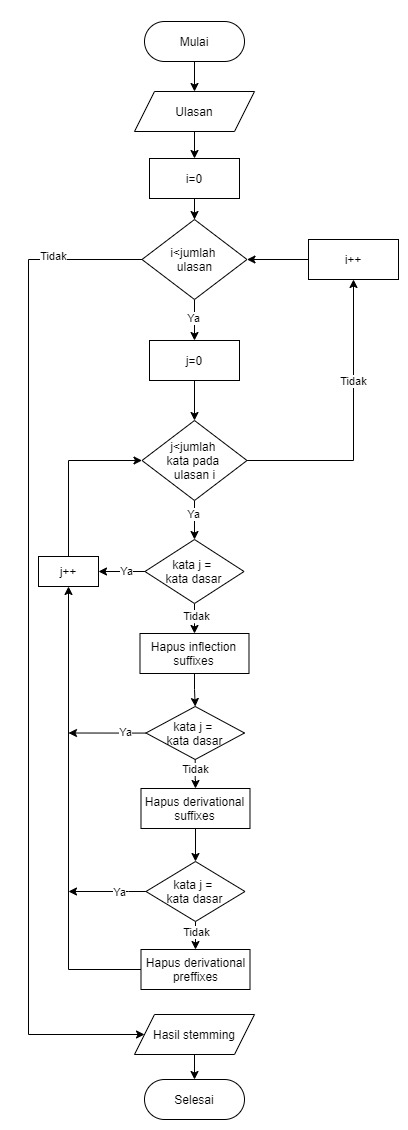
#### Gambar 3.6 Flowchart *Stopword*

Proses *stopword* akan menerima input berupa hasil proses sebelumnya yaitu *cleansing*. Dalam proses ini akan dilakukan pengecekan per kata dan menghapus kata yang tidak memiliki makna pada kalimat atau ulasan. Penghapusan *stopword* dengan pengecekan menggunakan kamu *stopword*. Jika kata tersebut ada dalam kamus, maka akan dihapus. Jika kata tersebut tidak ada dalam kamus, maka akan disimpan. Apabila semua kata telah diperiksa, maka *flowchart* ini akan memberikan output berupa hasil *stopword*. Berikut contoh penerapan *stopword* terhadap ulasan *google maps* yang akan diporses dapat dilihat pada Tabel 3.8.

#### Tabel 3.8 Contoh Penerapan *Stopword*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| datang waktu buka jam selesai jam pagi antrinya lumayan tidak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi | waktu buka jam selesai jam pagi antrinya lumayan ramai pelayanannya cepat instruksi |

## 3.2.9. *Stemming*

*Stemming* adalah proses menghilangkan imbuhan dari sebuah kata atau pembentukan kata dasar. *Flwochart* dari proses stemming dapat dilihat pada Gambar 3.7.

#### Gambar 3.7 *Flowchart Stemming*

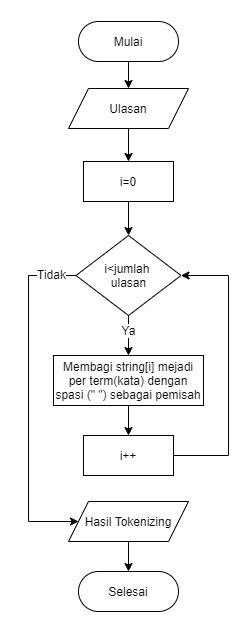
Proses *stemming* akan menerima input berupa hasil proses sebelumnya yaitu *stopword*. Dalam proses ini akan dilakukan pengecekan per kata dan mencari kata dasar. Jika kata bukan kata dasar, maka akan dihapus *inflection suffixes*, *derivational suffixes*, dan atau *derivational suffixes* lalu disimpan. Jika kata tersebut adalah kata dasar, maka akan disimpan.

Apabila semua kata telah diperiksa, maka *flowchart* ini akan memberikan output berupa hasil *stemming*. Berikut adalah contoh penerapan *stemming* terhadap ulasan *google maps* yang akan diproses dapat dilihat pada Tabel 3.9.

#### Tabel 3.9 Contoh Penerapan *Stemming*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| waktu buka jam selesai jam pagi antrinya lumayan ramai pelayanannya cepat instruksi | waktu buka jam selesai jam pagi antri lumayan ramai layan cepat instruksi |

## 3.2.10. *Tokenizing*

*Tokenizing* merupakan proses pemisahan kata dari bentuk string pada kalimat atau hasil dari tahapan *preprocessing* *stemming* sebelumnya. Proses ini akan memotong data menjadi per kata spasi (“ ”) sebagai pemisahnya. *Flowchart* dari proses *tokenizing* dapat dilihat pada Gambar 3.8.

#### Gambar 3.8 *Flowchart Tokenizing*

Berikut contoh penerapan *tokenizing* terhadap ulasan *google maps* yang akan diporses dapat dilihat pada Tabel 3.10.

#### Tabel 3.10 Contoh Penerapan *Tokenizing*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| waktu buka jam selesai jam pagi antri lumayan ramai layan cepat instruksi | [‘waktu’, ‘buka’, ‘jam’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘lumayan’, ‘ramai’, ‘layan’, ‘cepat’, ‘instruksi’] |

## 3.2.11. *Term Frequency* (TF) Dengan *Stopword*

Analisis pelatihan klasifikasi dalam penelitian ini dilakukan dengan *Term Frequency*. Adapun langkah dalam tahapan ini adalah menghitung *Term Frequency*. *TF* (*Term Frequency*) merupakan frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan. Penggunaan *Term Frequency* untuk menghitung kemunculan suatu *term* dalam suatu *corpus* berdasarkan bobot suatu *term* t pada dokumen d. Apabila semakin besar jumlah kemunculan sebuah *term* dalam dokumen maka akan semakin besar pula nilai bobotnya. Dalam penelitian ini jenis *term* yang digunakan adalah *TF* murni yang mana nilai *TF* diberikan berdasarkan jumlah kemunculan suatu term di dokumen. Berikut ini contoh data ulasan yang telah melalui tahapan *preprocessing*, kemudian akan dihitung kemunculan kata atau *Term* *Frequencynya*:

#### Tabel 3.11 Data Ulasan Setelah Proses *Text Processing*

|  |  |  |  |
| --- | --- | --- | --- |
| **Doc** | **Ulasan** | **Label Sentimen** | **Keterangan** |
| 1 | layan cepat ramah | Positif | Data latih |
| 2 | layan ruang tunggu antri kurang madah cuci tangan air habis | Negatif | Data latih |
| 3 | parkir motor luas layan baik ramah antri sabar tunggu | Positif | Data latih |
| 4 | layan kurang cepat | Negatif | Data latih |
| 5 | dinas kependudukan cacat sipil layan cepat ramah | ? | Data uji |

Ulasan 1: layan cepat ramah

Ulasan 2: layan ruang tunggu antri kurang madah cuci tangan air habis

Ulasan 3: parkir motor luas layan baik ramah antri sabar tunggu

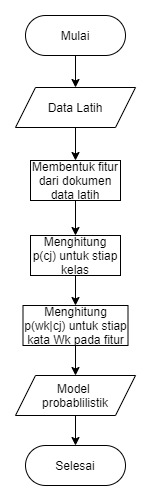
Ulasan 4: layan kurang cepat

Maka *Term* *Frequency* dari 4 ulasan diatas dapat dilihat pada Tabel 3.12.

#### Tabel 3.12 *Term Frequency* Data Sentimen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | | | |
| **D1** | **D2** | **D3** | **D4** |
| 1 | **layan** | 1 | 1 | 1 | 1 |
| 2 | **cepat** | 1 | 0 | 0 | 1 |
| 3 | **ramah** | 1 | 0 | 1 | 0 |
| 4 | **ruang** | 0 | 1 | 0 | 0 |
| 5 | **tunggu** | 0 | 1 | 1 | 0 |
| 6 | **antri** | 0 | 1 | 1 | 0 |
| 7 | **kurang** | 0 | 1 | 0 | 1 |
| 8 | **mudah** | 0 | 1 | 0 | 0 |
| 9 | **cuci** | 0 | 1 | 0 | 0 |
| 10 | **tangan** | 0 | 1 | 0 | 0 |
| 11 | **air** | 0 | 1 | 0 | 0 |
| 12 | **habis** | 0 | 1 | 0 | 0 |
| 13 | **parkir** | 0 | 0 | 1 | 0 |
| 14 | **motor** | 0 | 0 | 1 | 0 |
| 15 | **luas** | 0 | 0 | 1 | 0 |
| 16 | **baik** | 0 | 0 | 1 | 0 |
| 17 | **Sabar** | 0 | 0 | 1 | 0 |

## 3.2.12. *Naïve Bayes Classifier*

 Alur proses analisis dengan metode *naïve bayes* dibagi menjadi dua tahapan yaitu *training naïve bayes* dan klasifikasi *naïve bayes*. *Flowchart training naïve bayes* dapat dilihat pada gambar 3.9.

#### Gambar 3.9 *Flowchart Training Naïve Bayes*

Proses *naïve bayes classifier* dimulai dengan menghitung probabilitas pada setiap kata dalam ulasan. Selanjutnya dilakukan perhitungan nilai probabilitas prior yaitu nilai kemungkinan munculnya kelas tertentu dalam beberapa data. Langkah-langkah dalam tahapan ini adalah:

1. Menghitung nilai peluang

Menghitung nilai peluang dari kelas tertentu 𝑷 ̂ (𝒄) dengan menggunakan rumus pada persamaan 2.4. Ulasan kelas positif berjumlah 2, yaitu: D1, D3.

𝑷 ̂ (𝒑𝒐𝒔𝒊𝒕𝒊𝒇) **=**  = = **0.5**

Sedangkan kelas ulasan negatif berjumlah 2, yaitu: D2, D4.

𝑷 ̂ (𝒏𝒆𝒈𝒂𝒕𝒊𝒇) **=**  = = **0.5**

1. Menghitung probabilitas

Menghitung probabilitas kondisional frekuensi relative *term* t pada dokumen kelas c atau 𝑷 ̂ (𝒕𝒌|𝒄) yang dihitung dengan persamaan 2.6. Kemudian dokumen berdasarkan kategori positif.

#### Tabel 3.13. Dokuem Kategori Positif

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | |
| **D1** | **D3** |
| 1 | layan | 1 | 1 |
| 2 | cepat | 1 | 0 |
| 3 | ramah | 1 | 1 |
| 4 | ruang | 0 | 0 |
| 5 | tunggu | 0 | 1 |
| 6 | antri | 0 | 1 |
| 7 | kurang | 0 | 0 |
| 8 | mudah | 0 | 0 |
| 9 | cuci | 0 | 0 |
| 10 | tangan | 0 | 0 |
| 11 | air | 0 | 0 |
| 12 | habis | 0 | 0 |
| 13 | parkir | 0 | 1 |
| 14 | motor | 0 | 1 |
| 15 | luas | 0 | 1 |
| 16 | baik | 0 | 1 |
| 17 | sabar | 0 | 1 |

Kemudian dihitung P(layan|Positif), P(cepat|Positif), P(ramah|Positif), P(tunggu|Positif), P(antri|Positif), P(parkir|Positif), P(motor|Positif), P(luas|Positif), P(sabar|Positif). Pada dokumen class positif terdapat 12 kata.

P (Positif) = 0.5

P (layan|Positif) = 0,1034

P (cepat|Positif) = 0,6896

P (ramah|Positif) = 0,1034

P (tunggu|Positif) = 0,6896

P (antri|Positif) = 0,6896

P (parkir|Positif) = 0,6896

P (motor|Positif) = 0,6896

P (luas|Positif) = 0,6896

P (sabar|Positif) = 0,6896

Kemudian dihitung dokumen berdasarkan kategori negatif.

#### Tebel 3.14 Dokumen Kategori Negatif

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | |
| **D2** | **D4** |
| 1 | layan | 1 | 1 |
| 2 | cepat | 0 | 1 |
| 3 | ramah | 0 | 0 |
| 4 | ruang | 1 | 0 |
| 5 | tunggu | 1 | 0 |
| 6 | antri | 1 | 0 |
| 7 | kurang | 1 | 1 |
| 8 | mudah | 1 | 0 |
| 9 | cuci | 1 | 0 |
| 10 | tangan | 1 | 0 |
| 11 | air | 1 | 0 |
| 12 | habis | 1 | 0 |
| 13 | parkir | 0 | 0 |
| 14 | motor | 0 | 0 |
| 15 | luas | 0 | 0 |
| 16 | baik | 0 | 0 |
| 17 | sabar | 0 | 0 |

Kemudian dihitung P(layan|Negatif), P(cepat|Negatif), P(ruang|Negatif), P(tunggu|Negatif), P(antri|Negatif), P(mudah|Negatif), P(cuci|Negatif), P(tangan|Negatif), P(air|Negatif), P(habis|Negatif). Pada dokumen class negatif terdapat 13 kata.

P (Negatif) = 0.5

P(layan|Negatif) = = 0,1

P(cepat|Negatif) = = 0,0666

P(ruang|Negatif) = = 0,0666

P(tunggu|Negatif) = = 0,0666

P(antri|Negatif) = = 0,0666

P(kurang|Negatif) = = 0,1

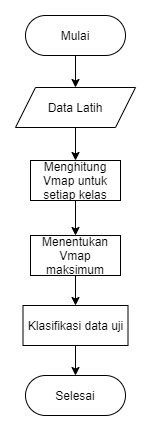
P(mudah|Negatif) = = 0,0666

P(cuci|Negatif) = = 0,0666

P(tangan|Negatif) = = 0,0666

P(air|Negatif) = = 0,0666

P(habis|Negatif) = = 0,0666

Selanjutnya proses yang kedua yaitu klasifikasi *naïve bayes* yaitu sebagai berikut:

#### Gambar 3.10 *Flowchart klasifikasi Naïve Bayes*

Klasifikasi ulasan baru yang belum diketahui nilai sentimennnya dalam proses pengujian dilakukan langkah-langkah sebagai berikut. Misalnya dimasukkan ulalsan baru, yaitu:

Dokumen 5: ‘dinas kependudukan cacat sipil layan cepat ramah’. Proses untuk mengklasifikasikannya dengan cara menghitung probabilitas komentar tersebut dalam kelas positif atau negatif, menggunakan hasil dari proses pelatihan. Maka hasil *preprocessing* dari ulasan pada dokumen 5 tersebut adalah kata ‘dinas’, ‘kependudukan’, ‘catat’, ’sipil’, ‘layan’, ‘cepat’, ‘ramah’. Setelah dilakukan proses preprocessing yaitu mengambi nilai peluang term tersebut dari kelas positif dan negatif. Langkah berikutnya adalah dengan menggunakan persamaan 2.5. Hasil probabilitas pada class positif dan negatif pada dokumen 5:

P(Positif) = P(Positif) x P(dinas|Positif) x P(kependudukan|Positif) x P(catat|Positif) x P(sipil|Positif) x P(layan|Positif) x P(cepat|Positif) x P(ramah|Positif)

= 0.5 x 1 x 1 x 1 x 1 x 0.1034 x 0.0689 x 0.1034

= 0.00036

P(Negatif) = P(Negatif) x P(dinas|Negatif) x P(kependudukan|Negatif) x P(catat|Negatif) x P(sipil|Negatif) x P(layan|Negatif) x P(cepat|Negatif) x P(ramah|Negatif)

= 0.5 x 1 x 1 x 1 x 1 x 0.1 x 0.0666 x 0.0333

= 0.00011

Sehingga diperoleh P(Positif) sebesar 0.00036, P(Negatif) sebesar 0.00011. Karena P(Positif) lebih besar dari P(Negatif) maka dapat disimpulkan bahwa dokumen 5 tersebut dikategorikan sebagai dokumen Positif.

## 3.2.13. *Term Frequency* (TF) Tanpa *Stopword*

Berikut ini contoh data ulasan yang telah melalui tahapan *preprocessing* tanpa *stopword*, kemudian akan dihitung kemunculan kata atau *Term* *Frequencynya*:

#### Tabel 3.15 Data Ulasan Setelah Proses *Text Processing* Tanpa *Stopword*

|  |  |  |  |
| --- | --- | --- | --- |
| **Doc** | **Ulasan** | **Label Sentimen** | **Keterangan** |
| 1 | layan sangat cepat dan ramah | Positif | Data latih |
| 2 | layan ruang tunggu antri kurang madahi dan tempat cuci tangan air habis | Negatif | Data latih |
| 3 | tempat parkir motor luas layan baik ramah tapi karena antri banyak harus sabar tunggu | Positif | Data latih |
| 4 | layan kurang cepat | Negatif | Data latih |
| 5 | dinas kependudukan dan cacat sipil layan cepat dan ramah | ? | Data uji |

Ulasan 1: layan sangat cepat dan ramah

Ulasan 2: layan ruang tunggu antri kurang madahi dan tempat cuci tangan air habis

Ulasan 3: tempat parkir motor luas layan baik ramah tapi karena antri banyak arus sabar tunggu

Ulasan 4: layan kurang cepat

Maka *Term* *Frequency* dari 4 ulasan diatas dapat dilihat pada Tabel 3.16.

#### Tabel 3.16 *Term Frequency* Data Sentimen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | | | |
| **D1** | **D2** | **D3** | **D4** |
| 1 | **air** | 0 | 1 | 0 | 0 |
| 2 | **antri** | 0 | 1 | 1 | 0 |
| 3 | **baik** | 0 | 0 | 1 | 0 |
| 4 | **banyak** | 0 | 0 | 1 | 0 |
| 5 | **cepat** | 1 | 0 | 0 | 1 |
| 6 | **cuci** | 0 | 1 | 0 | 0 |
| 7 | **dan** | 1 | 1 | 0 | 0 |
| 8 | **habis** | 0 | 1 | 0 | 0 |
| 9 | **harus** | 0 | 0 | 1 | 0 |
| 10 | **karena** | 0 | 0 | 1 | 0 |
| 11 | **kurang** | 0 | 1 | 0 | 1 |
| 12 | **layan** | 1 | 1 | 1 | 1 |
| 13 | **luas** | 0 | 0 | 1 | 0 |
| 14 | **madahi** | 0 | 1 | 0 | 0 |
| 15 | **motor** | 0 | 0 | 1 | 0 |
| 16 | **parkir** | 0 | 0 | 1 | 0 |
| 17 | **ramah** | 1 | 0 | 1 | 0 |
| 18 | **ruang** | 0 | 1 | 0 | 0 |
| 19 | **sabar** | 0 | 0 | 1 | 0 |
| 20 | **sangat** | 1 | 0 | 0 | 0 |
| 21 | **tangan** | 0 | 1 | 0 | 0 |
| 22 | **tapi** | 0 | 0 | 1 | 0 |
| 23 | **tempat** | 0 | 1 | 1 | 0 |
| 24 | **tunggu** | 0 | 1 | 1 | 0 |

## 3.2.14. *Naïve Bayes Classifier*

Proses *naïve bayes classifier* dimulai dengan menghitung probabilitas pada setiap kata dalam ulasan. Selanjutnya dilakukan perhitungan nilai probabilitas prior yaitu nilai kemungkinan munculnya kelas tertentu dalam beberapa data. Langkah-langkah dalam tahapan ini adalah:

1. Menghitung nilai peluang

Menghitung nilai peluang dari kelas tertentu 𝑷 ̂ (𝒄) dengan menggunakan rumus pada persamaan 2.4. Ulasan kelas positif berjumlah 2, yaitu: D1, D3.

𝑷 ̂ (𝒑𝒐𝒔𝒊𝒕𝒊𝒇) **=**  = = **0.5**

Sedangkan kelas ulasan negatif berjumlah 2, yaitu: D2, D4.

𝑷 ̂ (𝒏𝒆𝒈𝒂𝒕𝒊𝒇) **=**  = = **0.5**

1. Menghitung probabilitas

Menghitung probabilitas kondisional frekuensi relative *term* t pada dokumen kelas c atau 𝑷 ̂ (𝒕𝒌|𝒄) yang dihitung dengan persamaan 2.6. Kemudian dokumen berdasarkan kategori positif.

#### Tabel 3.17 Dokumen Kategori Positif

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | |
| **D1** | **D3** |
| 1 | **air** | 0 | 0 |
| 2 | **antri** | 0 | 1 |
| 3 | **baik** | 0 | 1 |
| 4 | **banyak** | 0 | 1 |
| 5 | **cepat** | 1 | 0 |
| 6 | **cuci** | 0 | 0 |
| 7 | **dan** | 1 | 0 |
| 8 | **habis** | 0 | 0 |
| 9 | **harus** | 0 | 1 |
| 10 | **karena** | 0 | 1 |
| 11 | **kurang** | 0 | 0 |
| 12 | **layan** | 1 | 1 |
| 13 | **luas** | 0 | 1 |
| 14 | **madahi** | 0 | 0 |
| 15 | **motor** | 0 | 1 |
| 16 | **parkir** | 0 | 1 |
| 17 | **ramah** | 1 | 1 |
| 18 | **ruang** | 0 | 0 |
| 19 | **sabar** | 0 | 1 |
| 20 | **sangat** | 1 | 0 |
| 21 | **tangan** | 0 | 0 |
| 22 | **tapi** | 0 | 1 |
| 23 | **tempat** | 0 | 1 |
| 24 | **tunggu** | 0 | 1 |

Kemudian dihitung P(layan|Positif), P(antri|Positif), P(baik|Positif), P(banyak|Positif), P(cepat|Positif), P(dan|Positif), P(harus|Positif), P(karena|Positif), P(layan|Positif), P(luas|Positif) , P(motor|Positif) , P(parkir|Positif) , P(ramah|Positif) , P(sabar|Positif) , P(sangat|Positif) , P(tapi|Positif) , P(tempat|Positif) , P(tunggu|Positif). Pada dokumen class positif terdapat 18 kata.

P (Positif) = 0.5

P(antri|Positif) = 0,0465

P(baik|Positif) = 0,0465

P(banyak|Positif) = 0,0465

P(cepat|Positif) = 0,0465

P(dan|Positif) = 0,0465

P(harus|Positif) = 0,0465

P(karena|Positif) = 0,0465

P(layan|Positif) = 0,0697

P(luas|Positif) = 0,0465

P(motor|Positif) = 0,0465

P(parkir|Positif) = 0,0465

P(ramah|Positif) = 0,0697

P(sabar|Positif) = 0,0465

P(sangat|Positif) = 0,0465

P(tapi|Positif) = 0,0465

P(tempat|Positif) = 0,0465

P(tunggu|Positif) = 0,0465

Kemudian dihitung dokumen berdasarkan kategori negatif.

#### Tabel 3.18 Dokumen Kategori Negatif

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Doc** | ***Term Frequency*** | |
| **D2** | **D4** |
| 1 | **air** | 1 | 0 |
| 2 | **antri** | 1 | 0 |
| 3 | **baik** | 0 | 0 |
| 4 | **banyak** | 0 | 0 |
| 5 | **cepat** | 0 | 1 |
| 6 | **cuci** | 1 | 0 |
| 7 | **dan** | 1 | 0 |
| 8 | **habis** | 1 | 0 |
| 9 | **harus** | 0 | 0 |
| 10 | **karena** | 0 | 0 |
| 11 | **kurang** | 1 | 1 |
| 12 | **layan** | 1 | 1 |
| 13 | **luas** | 0 | 0 |
| 14 | **madahi** | 1 | 0 |
| 15 | **motor** | 0 | 0 |
| 16 | **parkir** | 0 | 0 |
| 17 | **ramah** | 0 | 0 |
| 18 | **ruang** | 1 | 0 |
| 19 | **sabar** | 0 | 0 |
| 20 | **sangat** | 0 | 0 |
| 21 | **tangan** | 1 | 0 |
| 22 | **tapi** | 0 | 0 |
| 23 | **tempat** | 1 | 0 |
| 24 | **tunggu** | 1 | 0 |

Kemudian dihitung P(air|Negatif), P(antri|Negatif), P(cepat|Negatif), P(cuci|Negatif), P(dan|Negatif), P(habis|Negatif), P(kurang|Negatif), P(layan|Negatif), P(madahi|Negatif), P(ruang|Negatif), P(tangan|Negatif), P(tunggu|Negatif), P(ruang|Negatif). Pada dokumen class negatif terdapat 13 kata.

P (Negatif) = 0.5

P(air|Negatif) = 0,0512

P(antri|Negatif) = 0,0512

P(cepat|Negatif) = 0,0512

P(cuci|Negatif) = 0,0512

P(dan|Negatif) = 0,0512

P(habis|Negatif) = 0,0512

P(kurang|Negatif) = 0,0769

P(layan|Negatif) = 0,0769

P(madahi|Negatif) = 0,0512

P(ruang|Negatif) = 0,0512

P(tangan|Negatif) = 0,0512

P(tunggu|Negatif) = 0,0512

P(ruang|Negatif) = 0,0512

Klasifikasi ulasan baru yang belum diketahui nilai sentimennnya dalam proses pengujian dilakukan langkah-langkah sebagai berikut. Misalnya dimasukkan ulalsan baru, yaitu:

Dokumen 5: ‘dinas kependudukan dan cacat sipil layan cepat dan ramah’. Proses untuk mengklasifikasikannya dengan cara menghitung probabilitas komentar tersebut dalam kelas positif atau negatif, menggunakan hasil dari proses pelatihan. Setelah dilakukan proses preprocessing yaitu mengambi nilai peluang term tersebut dari kelas positif dan negatif. Langkah berikutnya adalah dengan menggunakan persamaan 2.5. Hasil probabilitas pada class positif dan negatif pada dokumen 5:

P(Positif) = P(Positif) x P(dinas|Positif) x P(kependudukan|Positif) x P(dan|Positif) x P(catat|Positif) x P(sipil|Positif) x P(layan|Positif) x P(cepat|Positif) x P(dan|Positif) x P(ramah|Positif)

= 0.5 x 1 x 1 x 0,0465 x 1 x 1 x 0,0697x 0,0465 x 0,0465 x 0,0697

= 2,4488-7

P(Negatif) = P(Negatif) x P(dinas|Negatif) x P(kependudukan|Negatif) x P(dan|Negatif) x P(catat|Negatif) x P(sipil|Negatif) x P(layan|Negatif) x P(cepat|Negatif) x P(dan|Negatif) x P(ramah|Negatif)

= 0.5 x 1 x 1 x 0,0512 x 1 x 1 x 0,0769 x 0,0512 x 0,0512 x 0,0256

= 1,3300-7

Sehingga diperoleh P(Positif) sebesar 2,4488-7, P(Negatif) sebesar 1,3300-7. Karena P(Positif) lebih besar dari P(Negatif) maka dapat disimpulkan bahwa dokumen 5 tersebut dikategorikan sebagai dokumen Positif.

## 3.2.15. *Laplacian Smoothing*

Metode *laplacian* *smoothing* merupakan metode untuk menghindari hasil klasifikasi yang bernilai 0 dikarenakan data uji tidak ditemukan pada data latih. Pada penelitian ini digunakan metode *laplacian* *smoothing*, yaitu merode yang paling sederhana dengan menambahkan angka 1.

P(Positif) = P(Positif) x P(dinas|Positif) x P(kependudukan|Positif) x P(dan|Positif) x P(catat|Positif) x P(sipil|Positif) x P(layan|Positif) x P(cepat|Positif) x P(dan|Positif) x P(ramah|Positif)

= 0.5 x 1 x 1 x 0,0465 x 1 x 1 x 0,0697x 0,0465 x 0,0465 x 0,0697

= 2,4488-7

Pada contoh perhitungan probabilitas positif diatas dimana kata dinas, kependudukan, catat, sipil, tidak terdapat pada data latih sehingga ditambahkan nilai 1 untuk menghindari nilai 0.

P(Negatif) = P(Negatif) x P(dinas|Negatif) x P(kependudukan|Negatif) x P(dan|Negatif) x P(catat|Negatif) x P(sipil|Negatif) x P(layan|Negatif) x P(cepat|Negatif) x P(dan|Negatif) x P(ramah|Negatif)

= 0.5 x 1 x 1 x 0,0512 x 1 x 1 x 0,0769 x 0,0512 x 0,0512 x 0,0256

= 1,3300-7

Pada contoh perhitungan probabilitas negatif diatas dimana kata dinas, kependudukan, catat, sipil, tidak terdapat pada data latih sehingga ditambahkan nilai 1 untuk menghindari nilai 0.

## 3.3. Metodologi Pengembangan Sistem

Pada bagian ini akan membahas tentang metodologi pengembangan sistem yang akan digunakan sebagai bahan acuan dalam proses pembuatan penelitian Pengaruh Stopword Terhadap Performa Analisis Sentimen Pada Ulasan Google Maps Menggunakan Metode Naïve Bayes Calssifier.

Metodologi penelitian yang digunakan untuk pengumpulan data pada penelitian ini adalah metode penelitian kualitatif. Metode penelitian kualitatif menggunakan data berupa teori-teori dan memanfaatkan teori yang ada untuk menjadi dasar penguat argumentasi. Data penelitian yang akan diambil berupa ulasan dari google maps.

Metode pengembangan sistem pada penelitian ini mengguanakan metode prototype. Pada tahapan pengumpulan kebutuhan dilakukan dengan pengumpulan data yang bersumber dari studi pustaka, web scraping, dan analisis kebutuhan sistem yang berupa kebutuhan fungsional dan kebutuhan non fungsional.

Tahapan selanjutnya yaitu proses desain dilakukan perancangan sistem seperti perancangan arsitektur, perancangan proses, perancangan basis data, dan perancangan antarmuka. Tahapan membangun sistem berdasarkan perancangan antarmuka yang telah dibuat dengan penulisan program.

#### Gambar 3.11 Metodologi Pengembangan Sistem

## 3.3.1. Analisis Kebutuhan Sistem

Tujuan dari analisis kebutuhan sistem adalah untuk mendeskripsikan dan memberikan detail apa yang seharusnya dilakukan oleh sistem. Ada beberapa analisis yang dibutuhkan dalam menyelesaikan beberapa permasalahan.

## a. Kebutuhan Data Input

Data input merupakan data-data yang digunakan sebagai masukan pada sistem yang dibangun. Data tersebut kemudian disimpan pada basis data. Data input yang digunakan adalah:

1. Data Ulasan *Google Maps*

Dalam penelitian ini data ulasan diperoleh dari ulasan *google maps* yang diambil dengan cara *scraping* mulai tahun Januari 2017 hingga November 2020. Data ulasan tersebut nantinya akan dibagi menjadi dua bagian yaitu data latih dan data uji.

1. Data *Stopword*

Data *stopword* adalah kumpulan kata umum yang sering digunakan pada suatu bahasa namun kata-kata tersebut tidak berpengaruh dalam klasifikasi teks.

1. Data *Slangword*

Data *slangword* adalah kumpulan kata dalam ulasan yang tidak baku menjadi kata baku sesuai Bahasa Indonesia yang baik dan benar.

1. Data Kata Dasar

Data kata dasar diperoleh dari kamus data Sastrawi. Daftar kata dasar meliputi id kata dasar, id kata, dan status kata dasar.

## b. Kebutuhan Proses

Proses adalah sebuah keadaan ketika program sedang dieksekusi. Berikut merupakan kebutuhan proses yang dilakukan oleh sistem yaitu:

1. Proses *Text Preprocessing*

Pemrosesan awal teks ini memiliki tujuan untuk mempersiapkan teks sebelum diolah lebih lanjut dapat dimengerti oleh sistem. Tujuan utama proses ini adalah mendapatkan data yang sudah siap diproses oleh sistem.

1. Proses Ekstraksi Fitur

Tujuan dari proses ekstraksi fitur untuk menghitung banyaknya kemunculan suatu term atau kata dalam sebuah dokumen.

1. Proses Latih dan Uji Data

Tujuan melakukan proses data latih untuk mencari kata kunci atau *keyword* serta probabilitasnya yang nantinya akan digunakan dalam proses pengujian. Proses data uji merupakan pengklasifikasian terhadap data yang belum diketahui kategorinya.

## c. Kebutuhan Data Output

Kebutuhan data *output* adalah suatu kebutuhan data keluaran hasil dari proses yang dilakukan oleh sistem. Data keluaran yang dihasilkan oleh sistem ini sebagai berikut:

1. Data ulasan *google maps* hasil proses *preprocessing* yang sudah siap untuk diolah dan diklasifikasikan sesuai opini sentimennya.
2. Hasil pengujian terhadap sistem yang telah dibangun dalam melakukan analisis sentimen yaitu dengan *confusion matrix* dan *k-fold cross validation*.

## d. Kebutuhan Non Fungsional

Kebutuhan non-fungsional merupakan kebutuhan yang meliputi properti perilaku yang dimiliki oleh sistem. Kebutuhan ini melibatkan beberapa perangkat yang mendukung sistem, yaitu perangkat keras *(hardware)* dan kebutuhan perangkat lunak *(software).*

1. Perangkat Keras *(Hardware)*

Perangkat keras yang dibutuhkan untuk mengembangkan sistem pada penelitian ini dengan spesifikasi yang dapat dilihat pada Tabel 3.19.

#### Tabel 3.19 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak

|  |  |  |
| --- | --- | --- |
| **No** | **Perangkat Keras** | **Keterangan** |
| 1 | *Processor* | AMD A8-6410 |
| 2 | *RAM* | 4GB DDR4 |
| 3 | *Storage* | Hardisk 500GB |
| 4 | *Graphic* | AMD Radeo Graphics |
| 5 | Perangkat Input dan Output | *Keyboard, Mouse, dan Monitor* |

1. Perangkat Lunak *(Software)*

Perangkat lunak yang dibutuhkan untuk mengembangkan sistem pada penelitian ini dapat dilihat pada Tabel 3.20.

#### Tabel 3.20 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak

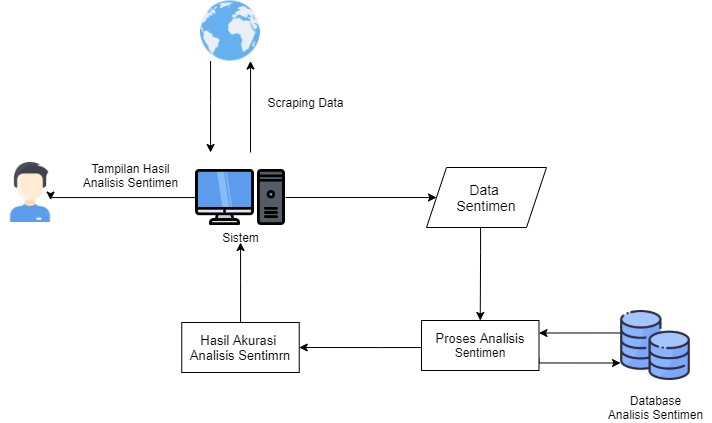
|  |  |  |
| --- | --- | --- |
| **No** | **Perangkat Lunak** | **Keterangan** |
| 1 | Operating System | Windows 10 64 bit |
| 2 | Bahasa Python | Bahasa permrograman |
| 3 | Bahasa PHP | Bahasa pemrograman |
| 4 | Sublime Text | Perangkat lunak *code editor* |
| 5 | XAMPP | Perangkat lunak sebagai server lokal |
| 6 | MySQL | Tempat penyimpanan database |
| 7 | Google Chrome | *Web browser* |
| 8 | *Draw.io* | Perangkat lunak untuk membuat desain diagram |
| 9 | *Figma* | Perangkat lunak untuk rancangan *interface* |

## 3.4. Proses Desain

Proses desain adalah tahapan yang membahas proses pembuatan rancangan untuk disajikan kepada pengguna berdasarkan hasil komunikasi antara pengembang dan pengguna sebelumnya. Perancangan yang dibuat bertujuan untuk memberikan gambaran dan mempermudah proses pembuatan sistem yang nantinya akan dibangun. Pada proses desain ini akan dibuat rancangan sistem dan rancangan pengujian.

## 3.4.1 Perancangan Sistem

Rancangan sistem yang akan dibuat terdiri dari admin, pengguna, *database*, model *Naïve Bayes Classifier*, dan proses klasifikasi. Terdapat dua aktor pada sistem ini, yaitu admin dan user. Admin merupakan orang tertentu yang akan melakukan pencarian sentimen dengan memilih sesuai dinas yang akan dimasukan pada sistem dan memiliki hak akses khusus pada sistem ini. Data yang berasal dari *web scraping* yaitu data ulasan dari pengguna *google maps* yang menulisakan ulasannya pada aplikasi *google maps*, data tersebut selanjutnya akan masuk ke tahap proses analisis sentimen. Pada tahapan ini data tersebut akan diproses *text preprocessing* untuk membersihkan data serta penyeragaman kata sehingga kata tersebut siap untuk di ekstraksi ke tahap berikutnya yaitu proses klasifikasi. Salah satu tahapan *text preprocessing* adalah *stopword removal*. Pada sistem ini akan dilakukan perbandingan penggunaan *stopword removal* dan *tanpa stopword removal*, yang bertujuan untuk mengetahui pengaruh *stopword removal* terhadap performa analisis sentimen. Selanjutnya sistem akan melakukan proses klasifikasi berdasarkan model *Naïve Bayes Classifier* yang telah melalui tahap *training* dan *testing* sebelumnya, pada proses ini menghasilkan hasil klasifikasi.

Data yang telah diklasifikasi kemudian akan digunakan untuk pembobotan dalam penentuan hasil klasifikasi. Sistem juga akan menyimpan data yang telah diolah ke *database* maupun memberikan output berupa data yang berasal dari *database* berdasarkan proses yang dilakukan. Sedangkan user merupakan semua pengguna yang secara umum dapat menggunakan sistem ini. *User* dapat melihat data informasi dinas dan hasil klasifikasi. Ilustrasi arsitektur sistem dapat dilihat pada Gambar 3.12.

#### Gambar 3.12 Arsitektur Sistem

## 3.4.2 Perancangan Proses

Pada tahap ini akan menggambarkan alur proses yang akan digambarkan dengan *flowchart* yang menjelaskan semua proses yang ada pada sistem.

## a. *Data Flow Diagram* Level 0

Diagram konteks atau yang biasanya disebut dengan DFD level 0 adalah suatu gambaran lingkaran besar yang akan mewakili proses dari sistem yang akan dibangun ini. DFD level 0 dapat dilihat pada Gambar 3.3.

Pada Gambar 3.13 *Data Flow Diagram* level 0 dibawah ini terdapat satu proses yaitu penerapan analisis sentimen pada ulasan *google maps* yang terdiri dari dua entitas yaitu admin dan *web scraping* ulasan *google maps*. Pada entitas admin terdapat aliran data keluar dan aliran data masuk yaitu data login, data latih, data stopword, data slangword, data uji, data kamus, dan hasil analisis.

#### E:\dfd0.jpg Gambar 3.13 *Data Flow Diagram* Level 0

## b. *Data Flow Diagram* Level 1

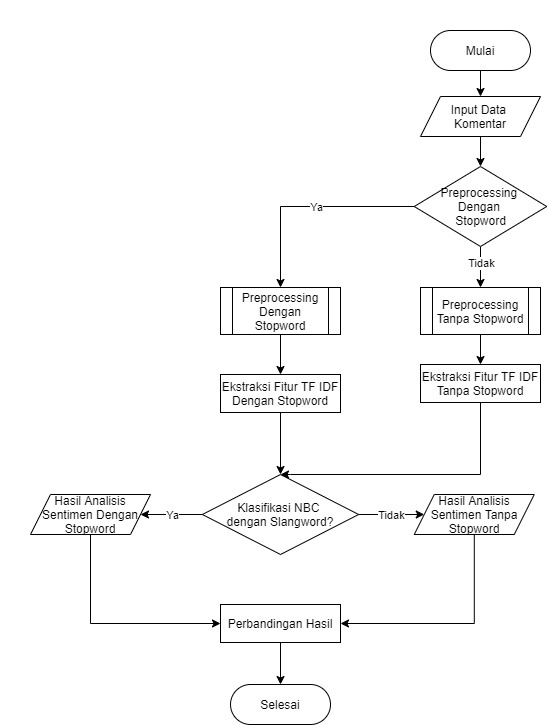
Pada *Data Flow Diagram* level 1 ini menguraikan suatu proses yang ada dalam diagram DFD level 0. Dalam proses layanan admin hanya untuk mengelola data-data yang terdapat pada data *store*.

Beberapa proses yang terdapat pada DFD level 1 ini yaitu kelola data ulasan, kelola data latih, kelola data kata dasar, kelola data admin, kelol data *stopword*, kelola data *slangword*, kelola hasil analisis, dan halaman analisis. *Data Flow Diagram* level 1 untuk sistem ini dapat dilihat pada gambar 3.14.

#### E:\dfd1.jpgGambar 3.14 *Data Flow Diagram* Level 1

Pada Gambar 3.14 *Data Flow Doagram* level 1 ini menjabarkan tentang arus aliran data yang dilakukan oleh admin. DFD level 1 terdiri dari beberapa proses untuk menggambarkan aliran data serta interaksi antara admin dan sistem. Beberapa proses yang terdapat pada DFD level 1 ini yaitu login, kelola data latih, kelola data uji, kelola data *stopword*, kelola data *slangword*, kelola data kata dasar, kelola data kategori, data hasil analisis, dan halaman analisis. Selanjutnya sistem dapat memproses dan menghubungkan semua data tersebut dengan *database* yang tersedia baik itu penambahan, penghapusan, atau perubahan data.

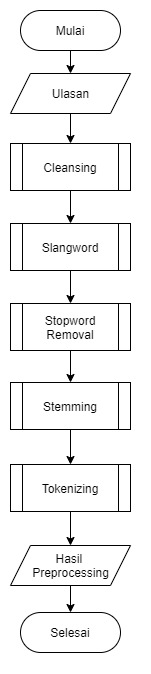
## c. *Flowchart* Analisis Sentimen

 *Flowchart* merupakan simbol-simbol tertentu dan memiliki penjelasan yang berbeda pada suatu diagram. Simbol-simbol tersebut dapat menggambarkan tentang urutan proses secara detail. Tahapan pada perancangan proses sangat dibutuhkan dalam membangun sebuah sistem yang bertujuan menggambarkan arus data saat perancangan sistem. Pada *flowchart* proses keseluruhan merupakan gambaran alur dari keseluruhan sistem ini. *Flowchart* dari proses keseluruhan dapat dilihat pada Gambar 3.15.

#### Gambar 3.15 *Flowchart* Sistem

Proses ini akan dimulai dari memasukan teks ulasan pada halaman klasifikasi. Kemudian dilakukan *text preprocessing*, yang terdiri dari tahapan-tahapan dalam pembersihan dan penyeleksian data. Hasil *preprocessing* tersebut kemudian akan digunakan untuk proses klasifikasi untuk menentukan sentimen positif atau sentimen negatif. Output dari sistem ini adalah hasil klasifikasi ulasan yaitu sentimen positif atau sentimen negatif.

## d *Flowchart Preprocessing* Dengan *Stopword*

 Flowchart yang akan ditampilkan merupakan proses seleksi data ulasan yang diinputkan user agar data yang akan digunakan menjadi lebih terstruktur sebelum diklasifikasikan nanti. *Preprocessing* pada penelitian ini terdiri dari beberapa tahapan, yaitu *cleansing, slangword, stopword removal, stemming*, dan *tokenizing*. Pemilihan tahap *preprocessing* bergantung pada kebutuhan setiap penelitian. *Flowchart* dari *preprocessing* dapat dilihat pada Gambar 3.16.

#### Gambar 3.16 *Flowchart Preprocessing* Dengan *Stopword*

Adapun tahapan *preprocessing* dengan *stopword removal* dalam penelitian ini diantaranya:

1. *Cleansing* merupakan tahapan yang dilakukan untuk menghilangkan noise atau elemen-elemen yang tidak diperlukan dalam suatu data atau kalimat.
2. *Slangword* merupakan tahapan mengubah kata-kata yang tidak baku, singkatan dan kesalahan ketik yang sering terjadi menjaddi baku atau kata yang seharusnya.
3. *Stopword* merupakan merupakan proses pengambilan kata-kata penting atau membuang kata-kata yang kurang penting dari token yang telah dihasilkan pada tahapan sebelumnya.
4. *Stemming* merupaakan tahapan proses pemotongan imbuhan pada kata berimbuhan yang dilakukan dengan algoritma tertentu untuk mengembalikan suatu kata kebentuk dasarnya.
5. *Tokenizing* adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusunnya.

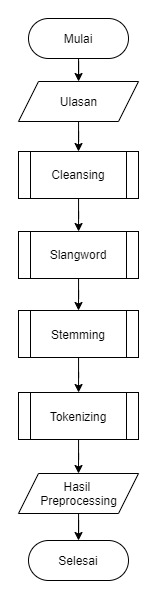
Berikut ini adalah contoh hasil *preprocessing* terhadap sebuah ulasan melalui semua tahapan *preprocessing* dapat dilihat pada Tabel 3.21.

#### Tabel 3.21 Contoh Penerapan *Preprocessing* Dengan *Stopword*

|  |  |
| --- | --- |
| **Teks Awal** | **Hasil *Preprocessing* Dengan *Stopword*** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | [‘waktu’, ‘buka’, ‘jam’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘lumayan’, ‘ramai’, ‘layan’, ‘cepat’, ‘instruksi’] |

## e *Flowchart Preprocessing* Tanpa *Stopword*

Alur proses dari *preprocessing* tanpa *stopword* dapat dilihat pada Gambar 3.17. Proses *preprocessing* terhadap data dilakukan untuk menghindari adanya data-data yang tidak konsisten. Tahapan penggunaan *stopword* merupakan proses yang dilakukan untuk menghapus kata yang tidak mempunyai makna atau arti pada kalimat.

Dalam penelitian ini akan dibandingkan hasil dari *preprocessing* dengan penggunaan *stopword* *removal* dengan tidak menggunakan *stopword* *removal*.

#### Gambar 3.17 *Flowchart* *Prerocessing* Tanpa *Stopword*

Adapun tahapan *preprocessing* tanpa melakukan penggunaan *stopword* dalam penelitian ini diantaranya:

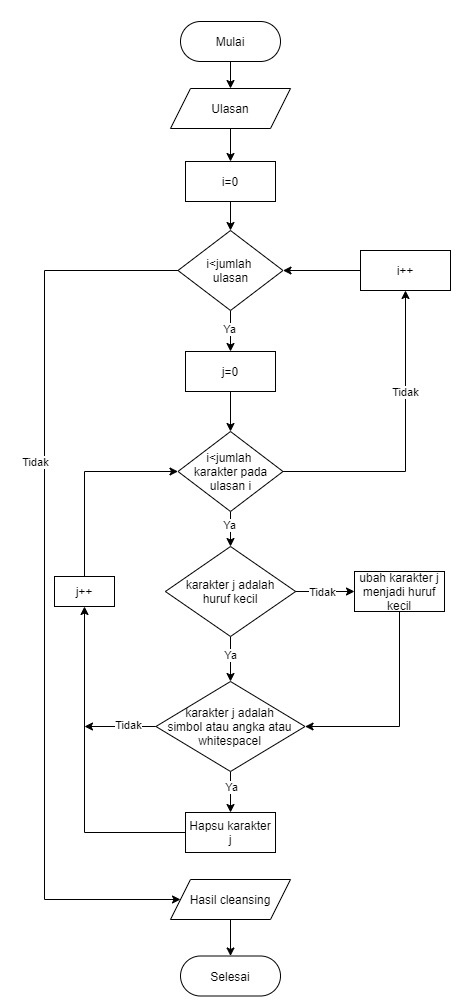
1. *Cleansing* merupakan tahapan yang dilakukan untuk menghilangkan *noise* atau elemen-elemen yang tidak diperlukan dalam suatu data atau kalimat.
2. *Slangword* merupakan tahapan mengubah kata-kata yang tidak baku, singkatan dan kesalahan ketik yang sering terjadi menjaddi baku atau kata yang seharusnya.
3. *Stemming* merupaakan tahapan proses pemotongan imbuhan pada kata berimbuhan yang dilakukan dengan algoritma tertentu untuk mengembalikan suatu kata kebentuk dasarnya.
4. *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

Berikut ini adalah contoh hasil *preprocessing* terhadap sebuah ulasan tanpa melalui stopword dapat dilihat pada Tabel 3.22.

#### Tabel 3.22 Contoh Penerapan *Preprocessing* Tanpa *Stopword*

|  |  |
| --- | --- |
| **Teks Awal** | **Hasil *Preprocessing* Dengan *Stopword*** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | [‘datang’, ‘waktu’, ‘buka’, ‘jam’, ‘pagi’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘tidak’, ‘ramai’, ‘layan’, ‘juga’, ‘baik’, ‘cepat’, ‘dan’, ‘jelas’, ‘dalam’, ‘memberi’, ‘instruksi’] |

## f. *Flowchart Celansing*

*Cleansing* adalah proses mengubah semua huruf pada ulasan menjadi huruf kecil *(lower case),* menghapus tanda baca (karakter simbol), angka, dan *whitespace*. *Flowchart* dari proses *cleansing* dapat dilihat pada Gambar 3.18.

#### Gambar 3.18 Flwochart *Cleansing*

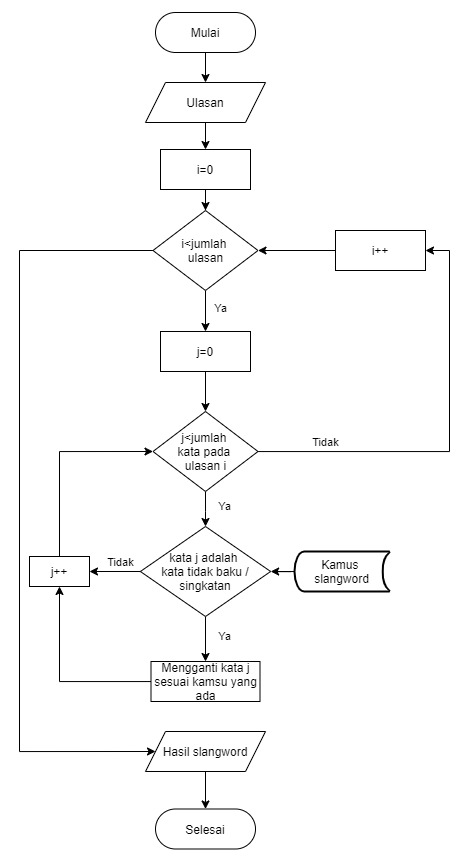
Pada *cleansing* ini akan menerima input berupa ulasan dengan inisiasi i yang menjadi jumlah ulasan dan j mewakili jumlah karakter pada ulasan. Prosesnya yaitu melakukan pengecekan ke setiap karakter dari awal hingga akhir. Jika karakter merupakan huruf kecil, maka karakter dilanjutkan ke kondisi selanjutnya. Juka menemukan karakter yang huruf besar maka dirubah terlebih dahulu karakter tersebut menjadi huruf kecil, lalu dilanjutkan ke kondisi selanjutnya. Kondisi selanjutnya melakukan pengecekan karakter kembali, juka karakter merupakan simbol (tanda baca) atau angka atau *whitespace*.

Jika karakter merupakan dalam kondisi tersebut maka karakter tersebut, maka akan dihapus. Tetapi jika bukan dalam kondisi tersebut, maka *flowchart* ini akan memberikan output berupa hasil *cleansing*. Berikut contoh penerapan cleansing terhadap ulasan *google maps* yang akan diproses dapat dilihat pada Tabel 3.23.

#### Tabel 3.23 Contoh Penerapan *Cleansing*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| Datang waktu buka jam 8 selesai jam 10 pagi. Antrinya lumayan nggak ramai. Pelayanannya juga baik, cepat, dan jelas dalam memberi instruksi. | datang waktu buka jam selesai jam pagi antrinya lumayan nggak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi |

## g. Flowchart *Slangword*

 *Slangword* adalah proses mengganti kata yang tidak baku atau singkatan dari hasil tahapan *preprocessing cleansing* sebelumnya menjadi kata baku atau aslinya. *Flowchart* dari proses *slangword* dapat dilihat pada Gambar 3.19.

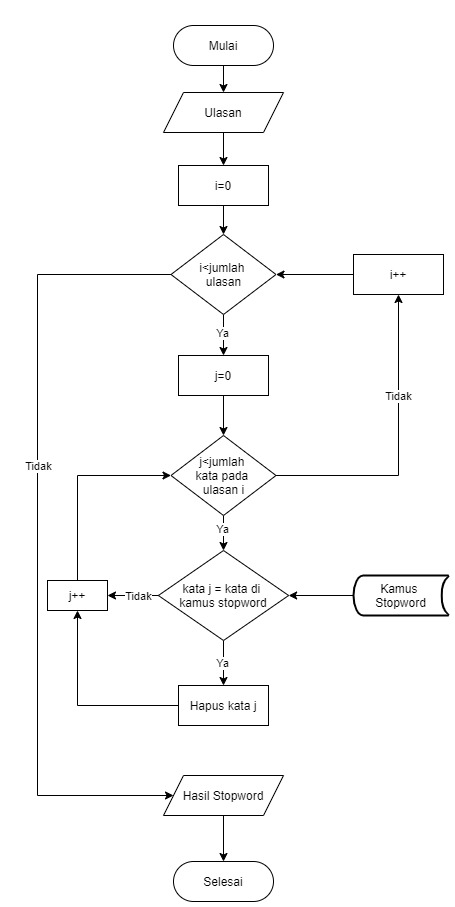
#### Gambar 3.19 *Flowchart Slangword*

Proses *slangword* ini akan menerima input berupa hasil proses sebelumnya yaitu *cleansing*. Dalam proses ini akan dilakukan pengecekan per kata dan menggunakan fungsi sinonim dengan mengganti kata yang sebelumnya sudah dibuat daftar kata sinonim dari kata tidak baku atau singkata yang telah disimpan di dalam *database*. Jika kata tersebut tidak baku atau singkatan seperti dalam kamus, maka akan diganti dengan sinonim dari kata pada kamus tersebut lalu disimpan. Jika kata tersebut baku atau bukan singkatan, maka kata tersebut akan disimpan. Apabila semua kata telah diperiksa, maka *flowchart* ini akan menampilkan *output* berupa *slangword*. Berikut contoh penerapan *slangword* terhadap ulasan *google maps* dapat dilihat pada Tabel 3.24.

#### Tabel 3.24 Contoh Penerapan *Slangword*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| datang waktu buka jam selesai jam pagi antrinya lumayan nggak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi | datang waktu buka jam selesai jam pagi antrinya lumayan tidak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi |

## h. Flowchart *Stopword Removal*

*Stopword removal* merupakan proses menghapus kata-kata yang sering muncul dan tidak memiliki makna pada kalimat. *Flowchart* dari proses *stopword* dapat dilihat pada Gambar 3.20.

#### Gambar 3.20 Flowchart *Stopwrod Removal*

Proses *stopword* akan menerima input berupa hasil proses sebelumnya yaitu *cleansing*. Dalam proses ini akan dilakukan pengecekan per kata dan menghapus kata yang tidak memiliki makna pada kalimat atau ulasan. Penghapusan *stopword* dengan pengecekan menggunakan kamu *stopword*. Jika kata tersebut ada dalam kamus, maka akan dihapus. Jika kata tersebut tidak ada dalam kamus, maka akan disimpan. Apabila semua kata telah diperiksa, maka *flowchart* ini akan memberikan output berupa hasil *stopword*. Berikut contoh penerapan *stopword* terhadap ulasan *google maps* yang akan diporses dapat dilihat pada Tabel 3.25.

#### Tabel 3.25 Contoh Penerapan *Stopword Removal*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| datang waktu buka jam selesai jam pagi antrinya lumayan tidak ramai pelayanannya juga baik cepat dan jelas dalam memberi instruksi | waktu buka jam selesai jam pagi antrinya lumayan ramai pelayanannya cepat instruksi |

## i. *Flowchart Stemming*

*Stemming* adalah proses menghilangkan imbuhan dari sebuah kata atau pembentukan kata dasar. *Flwochart* dari proses stemming dapat dilihat pada Gambar 3.21.

#### Gambar 3.21 *Flowchart Stemming*

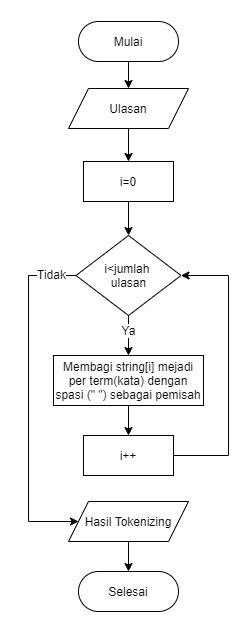
Proses *stemming* akan menerima input berupa hasil proses sebelumnya yaitu *stopword*. Dalam proses ini akan dilakukan pengecekan per kata dan mencari kata dasar. Jika kata bukan kata dasar, maka akan dihapus *inflection suffixes*, *derivational suffixes*, dan atau *derivational suffixes* lalu disimpan. Jika kata tersebut adalah kata dasar, maka akan disimpan.

Apabila semua kata telah diperiksa, maka *flowchart* ini akan memberikan output berupa hasil *stemming*. Berikut adalah contoh penerapan *stemming* terhadap ulasan *google maps* yang akan diproses dapat dilihat pada Tabel 3.26.

#### Tabel 3.26 Contoh Penerapan *Stemming*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| waktu buka jam selesai jam pagi antrinya lumayan ramai pelayanannya cepat instruksi | waktu buka jam selesai jam pagi antri lumayan ramai layan cepat instruksi |

## j. *Tokenizing*

*Tokenizing* merupakan proses pemisahan kata dari bentuk string pada kalimat atau hasil dari tahapan *preprocessing* *stemming* sebelumnya. *Flowchart* dari proses *tokenizing* dapat dilihat pada Gambar 3.22.

#### Gambar 3.22 *Flowchart Tokenizing*

Berikut contoh penerapan *tokenizing* terhadap ulasan *google maps* yang akan diporses dapat dilihat pada Tabel 3.27.

#### Tabel 3.27 Contoh Penerapan *Tokenizing*

|  |  |
| --- | --- |
| **Sebelum** | **Sesudah** |
| waktu buka jam selesai jam pagi antri lumayan ramai layan cepat instruksi | [‘waktu’, ‘buka’, ‘jam’, ‘selesai’, ‘jam’, ‘pagi’, ‘antri’, ‘lumayan’, ‘ramai’, ‘layan’, ‘cepat’, ‘instruksi’] |

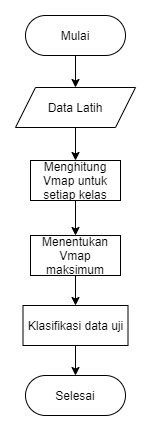
## k. *Flowchart Naïve Bayes Classifier*

Alur proses analisis dengan metode *naïve bayes* dibagi menjadi dua tahapan yaitu *training naïve bayes* dan klasifikasi *naïve bayes*. *Flowchart training naïve bayes* dapat dilihat pada gambar 3.23.

#### Gambar 3.23 *Flowchart Training Naïve Bayes*

Adapun tahapan proses *training naïve bayes* adalah sebagai berikut:

1. Memasukan data latih yang telah diberi label sentimen secara manual pada *database*.
2. Selanjutnya membentuk fitur yaitu dimana ulasan (dokumen) diklasifikasikan pada kelas yang telah ditentukan (kelas postif atau kelas negatif).
3. Kemudian setelah dibentuk fitur dengan kemunculan dari data latih. Selanjutnya menghitung probabilitas p(ci) dari setiap kelas dengan rumus.
4. Setelah mendapatkan probabilitas dari setiap kelas, selanjutnya menghitung untuk menentukan probabilitas setiap fitur pada setiap kelas dengan rumus probabilitas.
5. Maka akan dihasilkan output model probabilistik.
6. Proses berhenti.

Selanjutnya proses yang kedua yaitu klasifikasi *naïve bayes* yaitu sebagai berikut:

#### Gambar 3.24 *Flowchart klasifikasi Naïve Bayes*

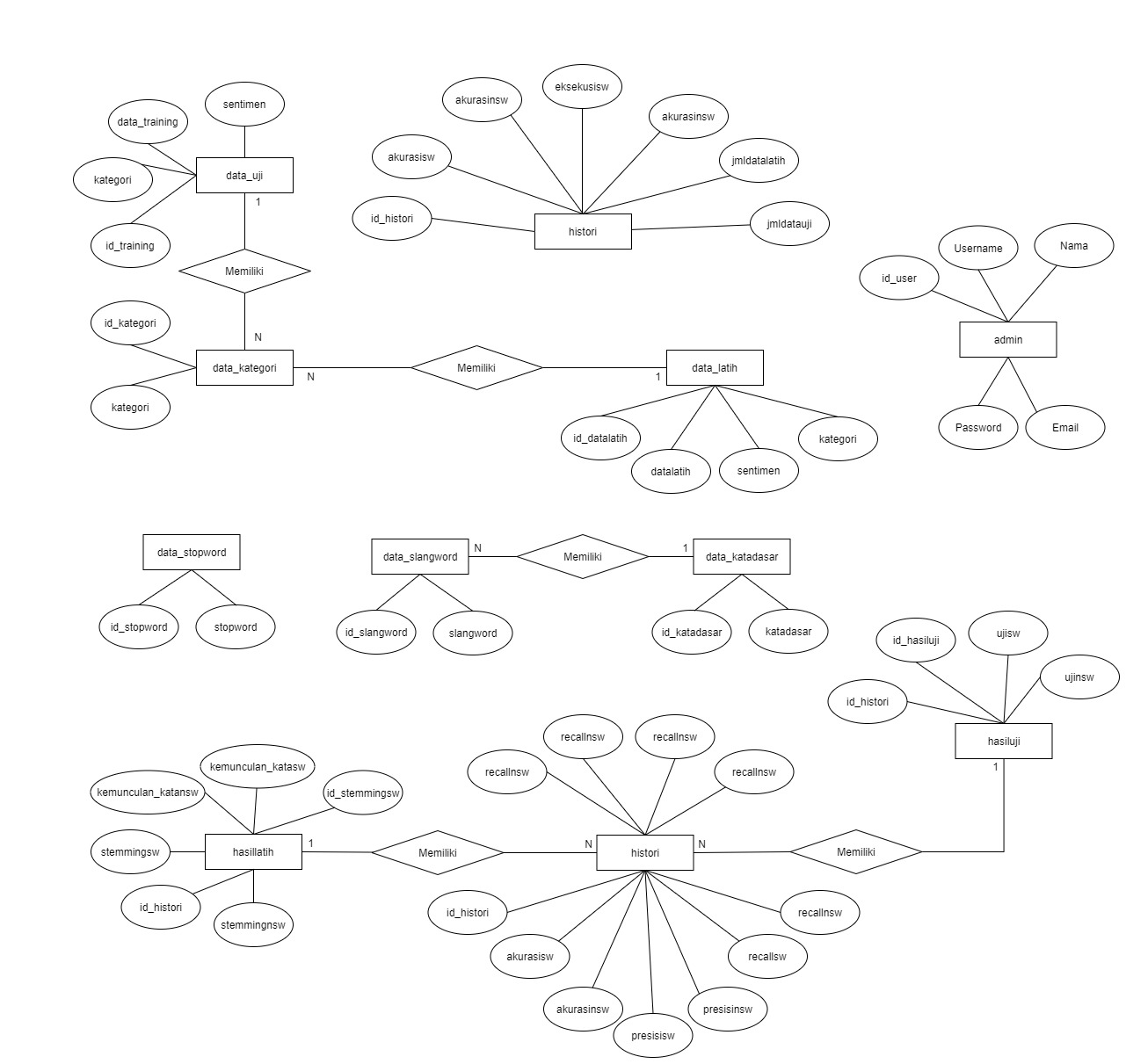
Adapun tahapan proses untuk klasifikasi *naïve bayes* yaitu sebagai berikut:

1. Menentukan beberapa data yang akan digunakan sebagai data uji.
2. Tahapan proses selanjutnya menghitung Vmap. Vmap adalah perhitungan yang digunakan untuk menentukan peluang atau probabilitas data uji dari masing-masing kelas berdasarkan proses *learning* pada klasifikasi *naïve bayes*.
3. Selanjutnya menentukan Vmap maksimum. Dari perhitungan Vmap sebelumnya akan di dapatkan nilai nilai Vmap positif dan Vmap negatif. Perhitungan yang memiliki hasil nilai Vmap tertinggi maka itu yang akan dipilih.
4. Tahap selanjutnya klasifikasi data uji. Setelah Vmap yang memiliki nilai maksimum, dapat disimpilkan ulasan tersebut diklasifikasikan ke dalam sentimen positif atau negatif
5. Proses berhenti.

## 3.4.3 Perancangan Basis Data

Dalam Penelitian ini membutuhkan *database* sebagai tempat penyimpanan seluruh data yang dibutuhkan dan akan mempermudah pengembang dalam memvisualisasikan data-data yang perlu untuk ditampilkan di dalam sistem. Pada *database* yang bernama “dbnbc” ini didalamnya terdapat enam tabel, yaitu data\_user, data\_stopword, data\_kategori, data\_slangword, data\_kamus, dan data\_training. Basis data pada sistem ini menggunakan *database* *MySql* dan *Server HTTP Apache*.

## a. *Entity Relationship Diagram* (ERD)

 *Entity Relationship Diagram* (ERD) adalah suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data mempunyai hubungan antar relasi. Proses ini akan memodelkan struktur data dan hubungan antar data. Terdapat enam buah entitas pada ERD data\_user, data\_stopword, data\_kategori, data\_slangword, data\_kamus, dan data\_training. Rancangan ERD pada penelitian ini dapat dilihat pada Gambar 3.25.

#### Gambar 3.25 Rancangan ERD

## b. Relasi Antar Tabel (RAT)

Relasi Antar Tabel (RAT) akan mempresentasikan hubungan antar tabel yang berfungsi untuk mengatur operasi suatu basis data. Rancangan RAT pada penelitian ini dapat dilihat pada Gambar 3.26.

#### E:\rat.PNGGambar 3.26 Rancangan RAT

## c. Struktur Tabel

Basis data pada penelitian ini terdiri dari beberapa tabel yang masing-masing tabel memiliki struktur tabel dapat dilihat sebagai berikut:

1. Tabel data\_admin

Pada tabel data admin mempunyai 5 buah atribut yaitu id\_user, username, nama, email, password dimana id\_user sebagai *primary key*. Tabel data admin merupakan tabel yang menyimpan semua data yang berhubungan dengan admin sebagai hak akses untuk masuk ke sistem. Struktur tabel data user dapat dilihat pada Tabel 3.28.

#### Tabel 3.28 Struktur Tabel data\_admin

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_user | int(4) | *Primary key, Auto Increment* | Id user |
| 2 | *username* | varchar(40) | None | *Username* user |
| 3 | nama | varchar(40) | None | Nama user |
| 4 | email | varchar(40) | None | Email user |
| 5 | *password* | Int(8) | None | Password user |

1. Tabel data\_stopword

Pada tabel data *stopwrod* mempunyai 2 buah atribut yaitu id\_stopword dan stopword dimana id\_stopword sebagai *primary key.* Tabel data stopword merupakan tabel yang menyimpan semua data yang berhubungan dengan data *stopword*. Struktur tabel data *stopword* dapat dilihat pada Tabel 3.29.

#### Tabel 3.29 Struktur Tabel data\_stopword

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama Field** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | Id\_*stopword* | int(4) | *Primary key, Auto Increment* | Id *stopword* |
| 2 | stopword | varchar(40) | None | Data *stopword* |

1. Tabel data\_katadasar

Pada tabel data kata dasar mempunyai 2 buah atribut yaitu id\_katadasar dan katadasar dimana id\_katadasar sebagai *primary key*. Tabel data kata dasar merupakan tabel yang menyimpan semua data yang berhubungan dengan data kata dasar. Struktur tabel data katadasar dapat dilihat pada Tabel 3.30.

#### Tabel 3.30 Struktur Tabel data\_katadasar

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama Field** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_katadasar | int(4) | *Primary key, Auto Increment* | Id kata dasar |
| 2 | katadasar | varchar(40) | None | Data kata dasar |

1. Tabel data\_slangword

Pada tabel data *slangword* mempunyai 2 buah atribut yaitu id\_slangword dan slangword dimana id\_slangword sebagai *primary key*. Tabel data kamus merupakan tabel yang menyimpan semua data yang berhubungan dengan data *slangword*. Struktur tabel data *slangword* dapat dilihat pada Tabel 3.31.

#### Tabel 3.31 Struktur Tabel data\_slangword

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_*slangword* | int(4) | *Primary key, Auto Increment* | Id *slangword* |
| 2 | *slangword* | varchar(40) | None | Data *slangword* |

1. Tabel data\_kategori

Pada tabel data kategori mempunyai 2 buah atribut yaitu id\_kategori dan kategori dimana id\_kategori sebagai *primary key*. Tabel data kategori merupakan tabel yang menyimpan semua data yang berhubungan dengan data kategori. Struktur tabel data kategori dapat dilihat pada Tabel 3.32.

#### Tabel 3.32 Struktur Tabel data\_kategori

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_kategori | int(4) | *Primary key, Auto Increment* | Id kategori |
| 2 | kategori | varchar(40) | None | Data kategori |

1. Tabel histori

Pada tabel hasil analisis mempunyai 7 atribut yaitu id\_histori, akurasisw, akurasinsw, eksekusisw, eksekusinsw, jmldatalatih, jmldatauji, dan id\_histori sebagai *primary key*. Tabel hasil analisis merupakan tabel yang menyimpan semua data yang berhubungan dengan data kategori. Struktur tabel data hasil analisis dapat dilihat pada Tabel 2.33.

#### Tabel 3.34. Struktur Tabel histori

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_histori | int(5) | *Primary key, Auto Increment* | Id histori |
| 2 | akurasisw | float | None | Akurasi stopword |
| 3 | akurasinsw | float | None | Akurasi non stopword |
| 4 | presisi | float | None | Presisi stopword |
| 5 | presisinsw | float | None | Presisi non stopword |
| 6 | recall | float | None | Recall stopword |
| 7 | recallnsw | float | None | Recall non stopword |
| 8 | eksekusi | float | None | Eksekusi stopword |
| 9 | eksekusinsw | float | None | Eksekusi non stopword |
| 10 | jmldatalatih | int(5) | None | Jumlah data latih |
| 11 | jmldatauji | int(5) | None | Jumlah data uji |

1. Tabel data\_latih

Pada tabel data latih mempunyai 4 buah atribut yaitu id\_datalatih, datalatih, sentimen, dan kategori dimana id\_datalatih sebagai *primary key*. Tabel data latih merupakan tabel yang menyimpan semua data yang berhubungan dengan data latih. Struktur tabel data latih dapat dilihat pada Tabel 3.35.

#### Tabel 3.35 Struktur Tabel data\_latih

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_datalatih | int(8) | *Primary key, Auto Increment* | Id data latih |
| 2 | datalatih | varchar(400) | None | Data latih |
| 3 | sentimen | varchar(40) | None | Kelas sentimen |
| 4 | kategori | varchar(40) | None | Kelas kategori |

1. Tabel data\_uji

Pada tabel data uji mempunyai 4 buah atribut yaitu id\_datauji, datauji, sentimen, dan kategori dimana id\_datauji sebagai *primary key*. Tabel data uji merupakan tabel yang menyimpan semua data yang berhubungan dengan data uji. Struktur tabel data uji dapat dilihat pada Tabel 3.36.

#### Tabel 3.36 Struktur Tabel data\_uji

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_datauji | int(8) | *Primary key, Auto Increment* | Id data uji |
| 2 | datauji | varchar(400) | None | Data uji |
| 3 | sentimen | varchar(40) | None | Kelas sentimen |
| 4 | kategori | varchar(40) | None | Kelas kategori |

9. Tabel hasillatih

Pada tabel data latih mempunyai 6 buah atribut yaitu id\_stemmingsw, stemmingsw, stemmingnsw, kemunculan\_katasw, kemunculan\_katansw, id\_histori, dimana id\_stemmingsw sebagai *primary key*. Tabel hasillatih merupakan tabel yang menyimpan semua data yang berhubungan dengan hasil latih. Struktur tabel hasil latih dapat dilihat pada Tabel 3.36.

#### Tabel 3.36 Struktur Tabel hasillatih

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_stemmingsw | int(8) | *Primary key, Auto Increment* | Id data uji |
| 2 | stemmingsw | text | None | Hasil processing |
| 3 | stemmingnsw | text | None | Hasil processing |
| 4 | kemunculan\_katasw | int(5) | None | Kemunculan kata |
| 5 | kemunculan\_katansw | int(5) | None | Kemunculan kata |
| 6 | id\_histori | int(5) | None | Id histori |

10. Tabel hasiluji

Pada tabel data uji mempunyai 4 buah atribut yaitu id\_hasiluji, ujinsw, ujinsw, id\_histori dimana id\_hasiluji sebagai *primary key*. Tabel hasiluji merupakan tabel yang menyimpan semua data yang berhubungan dengan hasil uji. Struktur tabel hasil uji dapat dilihat pada Tabel 3.36.

#### Tabel 3.36 Struktur Tabel hasiluji

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama *Field*** | **Tipe Data** | ***Constraint*** | **Keterangan** |
| 1 | id\_datauji | int(8) | *Primary key, Auto Increment* | Id data uji |
| 2 | ujisw | text | None | Hasil processing |
| 3 | ujinsw | text | None | Hasil processing |
| 4 | id\_histori | int(40) | None | Id histori |

## 3.4.4 Perancanagan Antarmuka

Perancangan antarmuka atau *user interface* adalah model mekanisme komunikasi antara pengguna dengan sistem. Antarmuka atau *user interface* merupakan bentuk sistem dalam tampilan grafis yang dapat menerima informasi dan memberikan informasi pada pengguna dalam bentuk tampilan yang interaktif. Secara garis besar, user interface pada sistem ini akan dibagi menjadi dua aktor, yaitu pengguna dan admin.

Berikut merupakan perancangan *user interface* pada aktor pengguna yang terdiri dari beberapa bagian, diantaranya adalah sebagai berikut:

1. Rancangan Halaman Masuk *(Login)*

Pada halaman login ini memvisualisasikan mengenai halaman pengunjung untuk *login* kedalam sistem. Pengguna memasukan *username* dan *password* dengan benar. Rancangan halaman masuk dapat dilihat pada gambar 3.27.

#### Gambar 3.27 Rancangan Halaman *Login*

1. Rancangan Halaman *Dashboard*

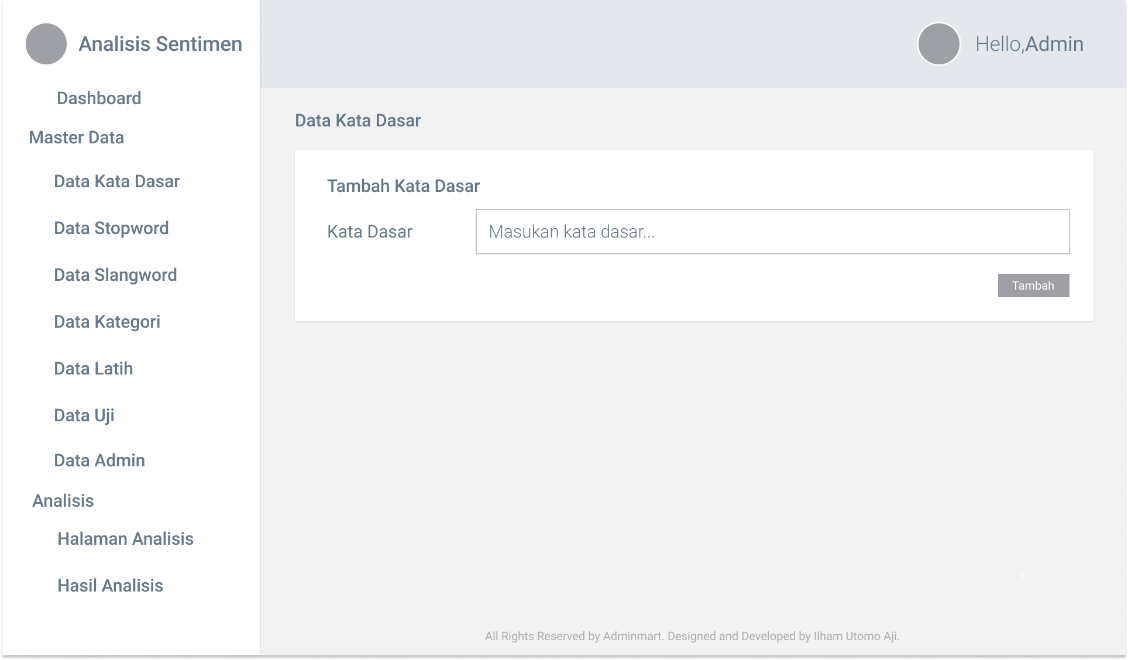
Halaman *dashboard* merupakan halaman yang akan tampil pertama kali saat diakses oleh pengguna. Rancanagan halaman *dashboard* dapat dilihat pada Gambar 3.28.

#### E:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Dashboard.pngGambar 3.28 Rancangan Halaman *Dashboard*

1. Rancangan Halaman Data Kata Dasar

Rancangan halaman data kata dasar berfungsi untuk mengolah data kata dasar yang digunakan dalam proses *stemming*. Pengguna dapat menambah, mengedit, dan menghapus data kata dasar jika terdapat perubahan dan penambahan data sewaktu-waktu. Untuk melihat rancangan halaman data kamus, penambahan, dan mengedit data dapat dilihat pada Gambar 3.29, dan Gambar 3.30.

#### E:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Data Kata Dasar.pngGambar 3.29 Rancangan Halaman Data Kata Dasar



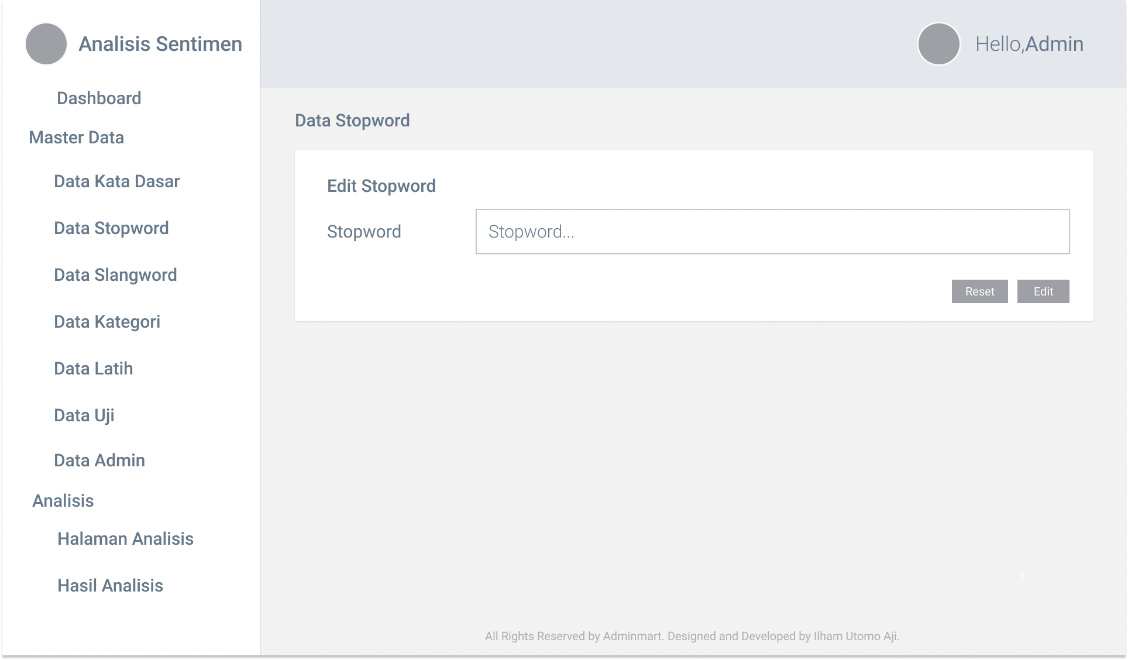
#### Gambar 3.30 Rancangan Halaman Tambah Data Kata Dasar

1. Rancangan Halaman Data *Stopword*

Rancangan halaman data *stopword* berfungsi untuk mengolah data *stopword* yang digunakan dalam proses *preprocessing*. Pengguna dapat menambah, mengedit, dan menghapus data *stopword* jika terdapat perubahan dan penambahan data sewaktu-waktu. Untuk melihat rancangan halaman data *stopword* dapat dilihat pada Gambar 3.31, Gambar 3.32, dan Gambar 3.33.

#### E:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Tambah Stopword.pngE:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Data Stopword.pngGambar 3.31 Rancangan Halaman Data *Stopword*

#### Gambar 3.32 Rancangan Halaman Tambah Data *Stopword*



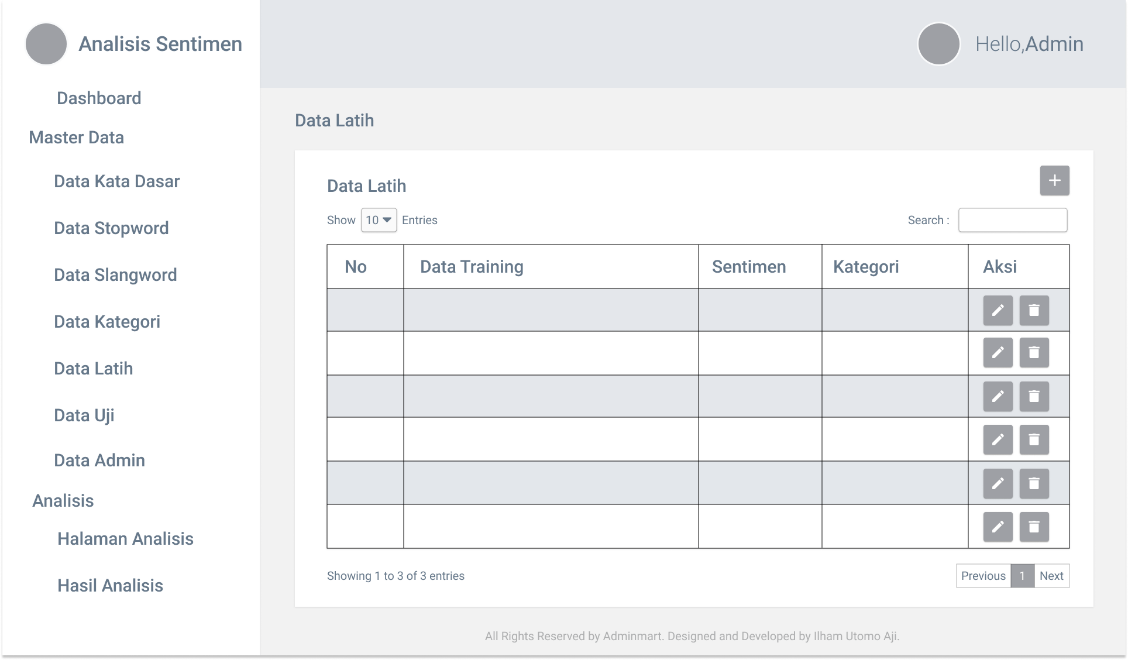
#### Gambar 3.33 Rancangan Halaman Edit Data *Stopword*

1. Rancangan Halaman Data *Slangword*

Rancangan halaman data *slangword* berfungsi untuk mengolah data *slangword* yang digunakan dalam proses *preprocessing*. Pengguna dapat menambah, mengedit, danmenghapus data *slangword* jika terdapat perubahan dan penambahan data sewaktu waktu. Untuk melihat rancangan halaman data *slangword* dapat dilihat pada Gambar 3.34.

#### E:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Data Slangword.pngGambar 3.34 Rancangan Halaman Data *Slangword*

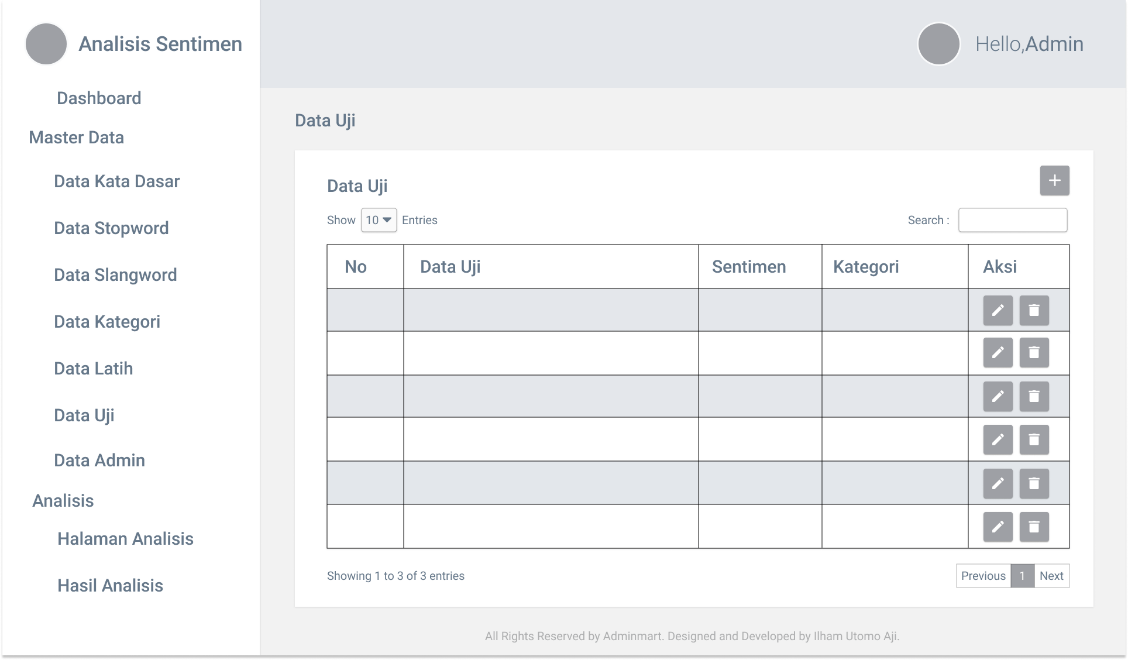
1. Rancangan Halaman Data *Latih*

Pada halaman data *training* merupakan halaman yang berisi kumpulan data opini masyarakat dari *google maps* tentang Dinas Kependudukan dan Catatan Sipil. Pengguna dapat menambah, mengedit, dan menghapus data apabila terdapat perubahan data sewaktu-waktu. Rancangan halaman data *training* dapat dilihat pada Gambar 3.35, dan Gambar 3.36.

#### E:\Kuliah\Tugas Akhir\Laporan Tugas Akhir\Gambar\Tambah Training.pngGambar 3.35 Rancangan Halaman Data Latih

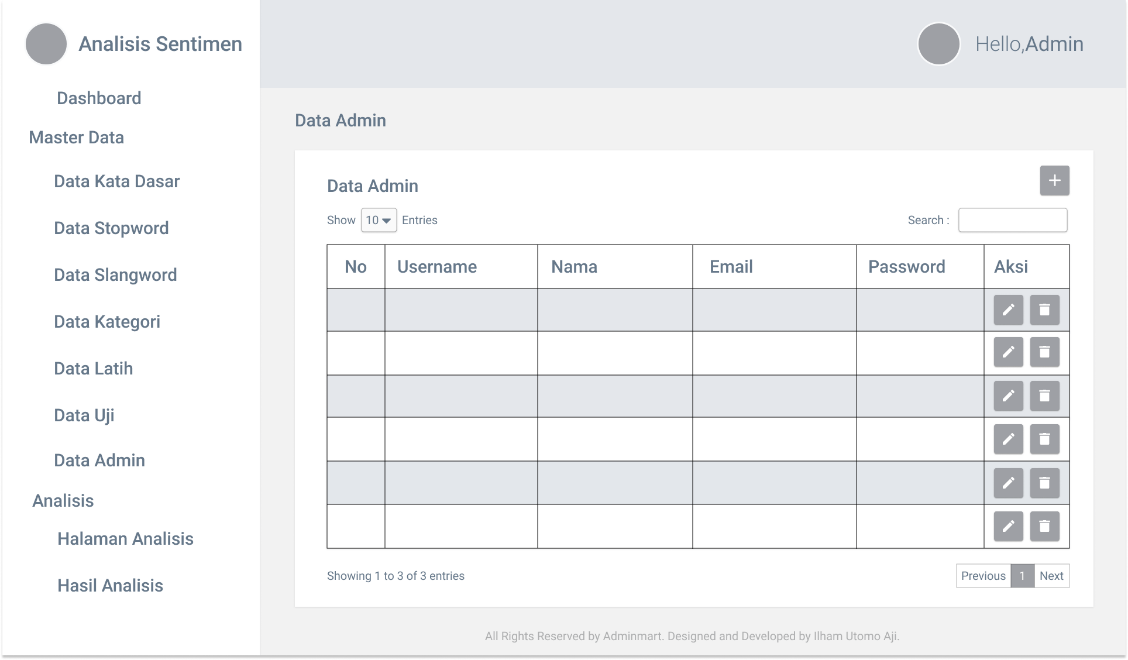
#### Gambar 3.36 Rancangan Halaman Tambah Data Latih

1. Rancangan Halaman Data Uji

Pada halaman data uji merupakan halaman yang berisi kumpulan data opini masyarakat dari *google maps* tentang Dinas Kependudukan dan Catatan Sipil. Pengguna dapat menambah, mengedit, dan menghapus data apabila terdapat perubahan data sewaktu-waktu. Rancangan halaman data uji dapat dilihat pada Gambar 3.37.

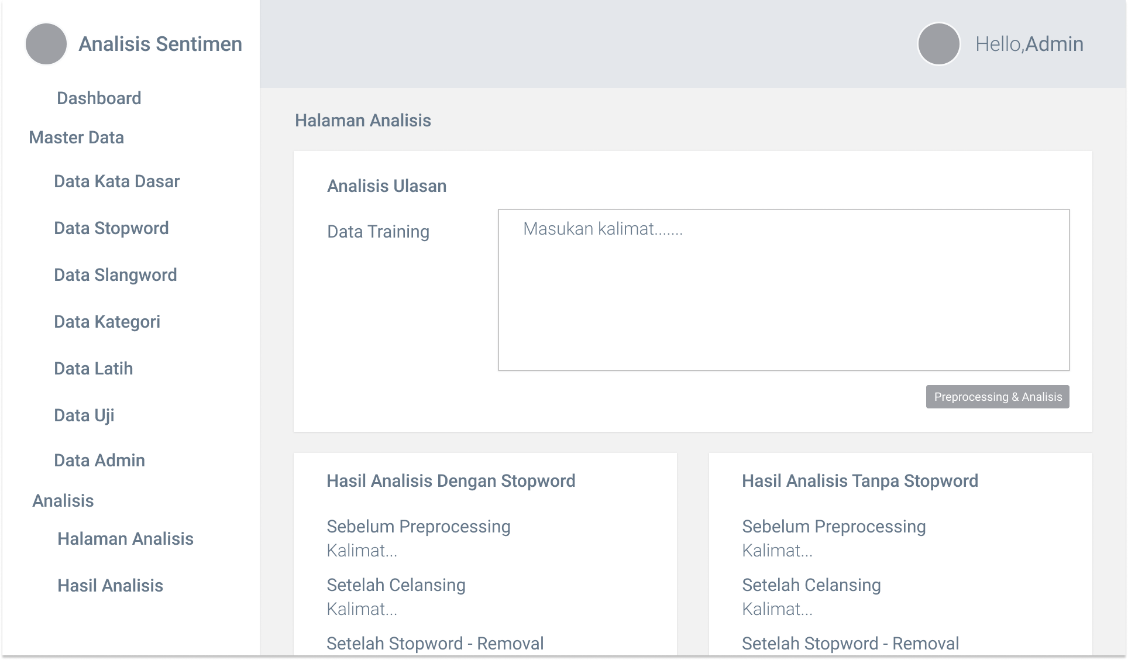
#### Gambar 3.37 Rancangan Halaman Data Uji

1. Rancangan Halaman Data Admin

Rancangan halaman data admin merupakan halaman yang akan menampilkan data admin yang sudah terdaftar ke dalam sistem. Rancangan halaman data admin dapat dilihat pada Gambar 3.38.

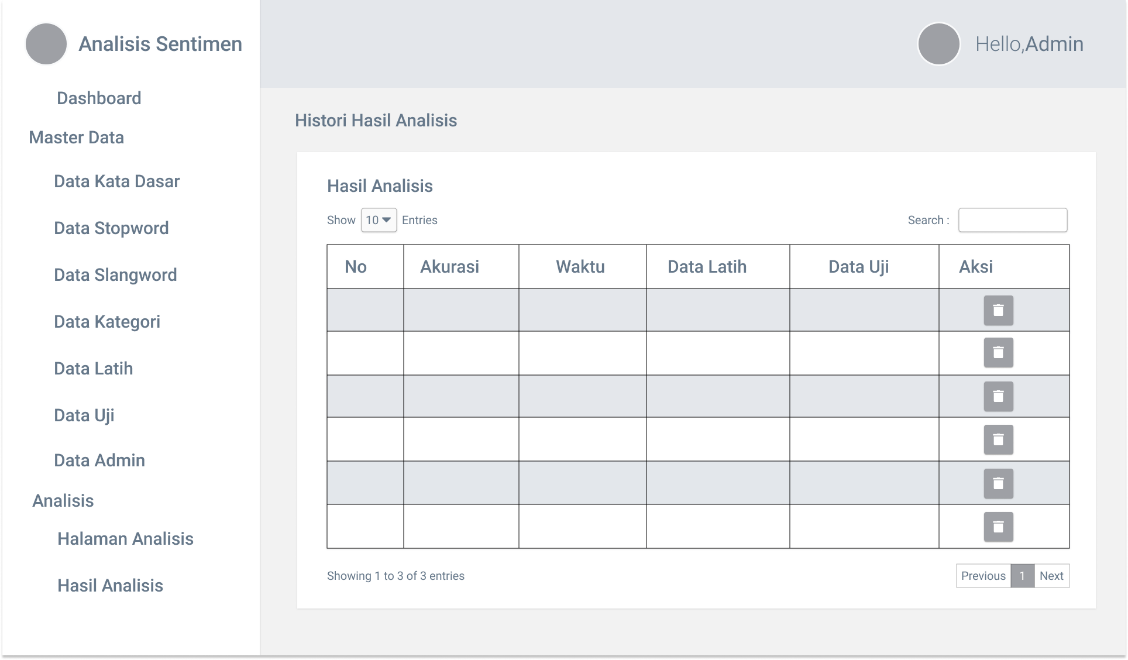
#### Gambar 3.38 Rancangan Halaman Data Admin

1. Rancangan Halaman Analisis

Pada rancangan halaman analisis, pengguna dapat menginputkan data atau ulasan yang ingin diketahui hasil analisis sentimennya. Selanjutnya sistem akan memproses ulasan tersebut dan akan di tampilkan hasil sentimen dengan metode *naïve bayes classifier* serta perbandingan dengan *stopword* maupun tanpa *stopword* pada proses *preprocessing*. Rancangan halaman analisis dapat dilihat pada gambar 3.39.

#### Gambar 3.39 Rancangan Halaman Analisis

1. Rancangan Halaman Hasil Analisis

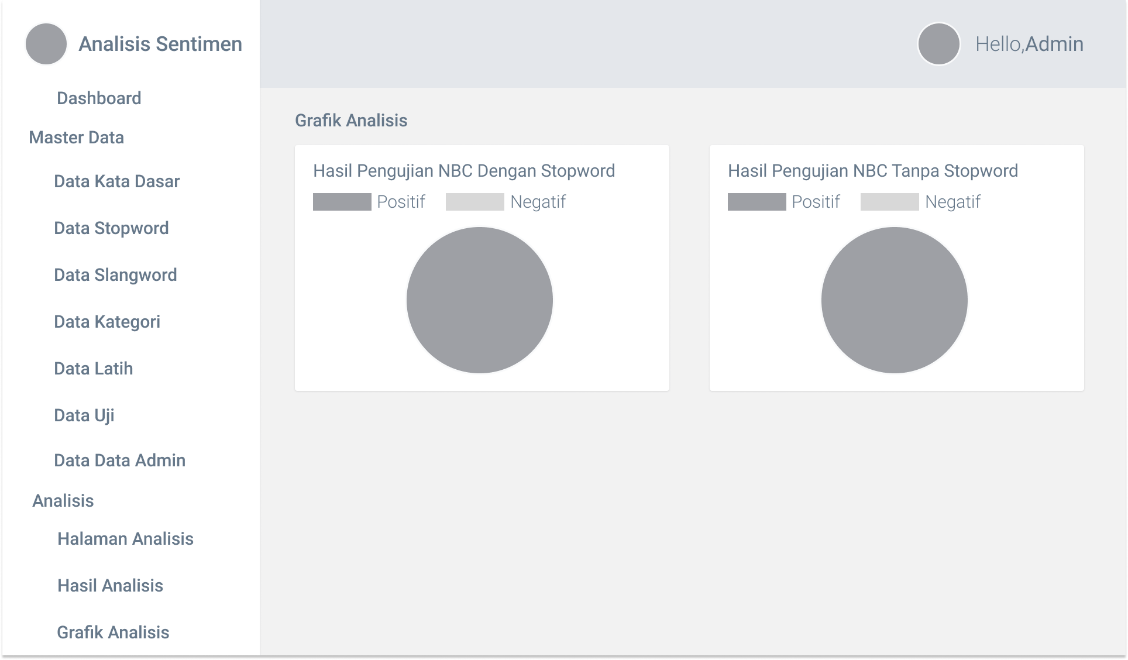
Perancangan halaman hasil analisis adalah sebuah halaman untuk melihat data hasil analisis. Pada halaman ini juga dapat melihat hasil dari analisis sentimen dengan perbandingan menggunakan *stopword* dan tanpa menggunakan *stopword* dengan metode *naïve bayes*. Rancangan halaman hasil analisis dapat dilihat pada Gambar 3.40.

#### Gambar 3.40 Rancangan Halaman Hasil Analisis

1. Rancangan Halaman Grafik Analisis

Dalam rancangan halaman grafik analisis, pengguna dapat melihat hasil akurasi, *presisi*, *recall*, menggunakan metode *naïve bayes classifier* serta perbandingan dengan *stopword* maupun tanpa *stopword*. Rancangan grafik analisis dapat dilihat pada gambar 3.41 dan gambar 3.42.

#### Gambar 3.41 Rancangan Halaman Grafik Analisis



#### Gambar 3.42 Rancangan Halaman Grafik Analisis

## 3.5 Rancangan Pengujian

Pada bagian ini akan membahas tentang perancangan pengujian. Penelitian ini akan melakukan dua pengujian yaitu oengujian terhadap model yang telah dibuat dan pengujian terhadap sistem.

#### 3.5.1 Pengujian Model

Pengujian model merupakan proses validasi dan pengujian dari dataset pada algoritma *naïve bayes*. Tahap ini bertujuan untuk mengukur kinerja dari suatu sistem yang telah dibangun. Proses ini bisa disebut juga dengan data *training* dan *testing*. Bagian training yaitu untuk melatih sistem dengan dataset yang telah diberi label sebagai bahan ajar akan menghasilkan suatu model. Model yang dihasilkan ini akan digunakan untuk melakukan klasifikasi pada tahap testing dengan memberi data oada model tanpa adanya label. Pada testing ini akan mengukur seberapa baik model yang telah dihasilkan. Metode pengujian yang akan digunakan pada penelitin ini yaitu *k-fold cross* *validation*, dan *confusion matrix*.

Metode *k-fold cross* validation bertujuan untuk melakukan validasi dari siatu model dengan melakukan pengujian sebanyak K, dimana pada penelitian ini menggunakan K=5. Sedangkan *confusion* *matrix* untuk mengukur tingkat akurasi, presisi, dan *recall* dari suatu model berdasarkan prediksi yang dihasilkan oleh sistem. Parameter perhitungan *confusion* *matrix* terdiri dari 4 hal yaitu *True* *Positive* (TP), *True* *Negative* (TN), *False* *Positive* (FP), *False* *Negative* (FN). Berikut rancangan pengujian menggunakan *k-fold cross validation*, dan *confusion matrix* dapat dilihat pada Tabel 3.26 dan Tabel 3.27.

#### Tabel 3.25 Tabel *Confusion Matrix*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Fold** | **Preprocessing** | ***Confusion Matrix*** | | | |
| **TP** | **FP** | **FN** | **TN** |
| 1 | Dengan *stopword* |  |  |  |  |
| Tanpa *stopword* |  |  |  |  |
| 2 | Dengan *stopword* |  |  |  |  |
| Tanpa *stopword* |  |  |  |  |
| 3 | Dengan *stopword* |  |  |  |  |
| Tanpa *stopword* |  |  |  |  |
| 4 | Dengan *stopword* |  |  |  |  |
| Tanpa *stopword* |  |  |  |  |
| 5 | Dengan *stopword* |  |  |  |  |
| Tanpa *stopword* |  |  |  |  |
| **Rata-rata** | |  |  |  |  |

#### Tabel 3.26 Tabel *k-fold cross validation*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fold** | **Preprocessing** | **Akurasi** | **Presisi** | ***Recall*** |
| 1 | Dengan *stopword* |  |  |  |
| Tanpa *stopword* |  |  |  |
| 2 | Dengan *stopword* |  |  |  |
| Tanpa *stopword* |  |  |  |
| 3 | Dengan *stopword* |  |  |  |
| Tanpa *stopword* |  |  |  |
| 4 | Dengan *stopword* |  |  |  |
| Tanpa *stopword* |  |  |  |
| 5 | Dengan *stopword* |  |  |  |
| Tanpa *stopword* |  |  |  |
| **Rata-rata** | |  |  |  |

## 3.4.2 Pengujian Sistem

Pada bagian ini akan dibuat suatu rancangan pengujian sistem. Pembuatan dari rancangan pengujian bertujuan untuk mempersiapkan konsep pengujian yang akan dilakukan untuk mengukur kinerja dari suatu sistem yang telah dibangun. Metode pengujian yang akan digunakan pada penelitian ini yaitu *black box testing*.

Pengujuan dengan *black box testing* bertujuan untuk mengukur kinarja dari suatu perangkat lunak yang telah dibangun. Metode pengujian secara *black box* lebih menekankan pada bagian fungsionalitas dari suatu perangkat lunak sehingga dapat memastikan setiap fungsi yang ada berjalan dengan semestinya. Berikut adalah detail rancangan pengujian dengan menggunakan black box dapat dilihat pada Tabel.

#### Tabel 3.27 Rancangan Pengujian *Black Box Testing*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aktor** | **Halaman** | **Detail Pengujian** | **Pengujian** | |
| **Berhasil** | **Gagal** |
| Admin | Login | Melakukan proses login |  |  |
| Fungsi session |  |  |
| Dashboard | Menampilkan jumlah data |  |  |
| Data Akun | Menampilkan data akun |  |  |
| Menambahkan data akun |  |  |
| Mengedit data akun |  |  |
| Menghapus data akun |  |  |
| Data kamus | Menampilkan data kamus |  |  |
| Menambahkan data kamus |  |  |
| Mengedit data kamus |  |  |
| Menghapus data kamus |  |  |
| Data Stopword | Menampilkan data stopword |  |  |
| Menambahkan data stopword |  |  |
| Mengedit data stopword |  |  |
| Menghapus data stopword |  |  |
| Data Slangword | Menampilkan data slangword |  |  |
| Menambahkan data slangword |  |  |
| Mengedit data slangword |  |  |
| Menghapus data slangword |  |  |
| Data Training | Menampilkan data training |  |  |
| Menambahkan data training |  |  |
| Mengedit data training |  |  |
| Menghapus data training |  |  |
| Data Kategori | Menampilkan data kategori |  |  |
| Menambahkan data kategori |  |  |
| Mengedit data kategori |  |  |
| Menghapus data kategori |  |  |
| Preprocessing | Melakukan proses preprocessing |  |  |
|  | Menampilkan hasil preprocessing |  |  |
| Uji coba analissis | Melakukan proses analisis |  |  |
| Menampilkan hasil proses analisis |  |  |
| Data Pengujian | Menampilkan data hasil pengujian |  |  |
| Menampilkan detail hasil pengujian |  |  |
| Model NBC | Menampilkan akurasi |  |  |
| Halaman Pengunjung | Berpindah ke halaman user |  |  |
| Pengaturan Akun Admin | Menampilkan data admin |  |  |
| Menanmbahkan data admin |  |  |
| Mengedit data admin |  |  |
| Menghapus data admin |  |  |
| Keluar | Fungsi logout |  |  |
| User | Dashboard | Input data analisis |  |  |
| Menampilkan hasil analisis |  |  |

# DAFTAR PUSTAKA

A. Yani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *7*(4), 257. https://doi.org/10.26418/justin.v7i4.30930

Akbari, M. I. H. A. D., Novianty, A., & Casi, S. (2017). Analisis Sentimen Menggunakan Metode Learning Vector Quantization Sentiment Analysis Using Learning Vector Quantization Method. *E-Proceeding of Engineering*, *4*(2), 2283–2292. https://openlibrary.telkomuniversity.ac.id/pustaka/files/135356/jurnal\_eproc/analisis-sentimen-menggunakan-metode-learning-vector-quantization.pdf

Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(12), 1733–1741.

Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, *1*(1), 32–41. https://www.researchgate.net/profile/Ghulam\_Buntoro/publication/316617194\_Analisis\_Sentimen\_Calon\_Gubernur\_DKI\_Jakarta\_2017\_Di\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf

Ghag, K. V., & Shah, K. (2016). Comparative analysis of effect of stopwords removal on sentiment classification. *IEEE International Conference on Computer Communication and Control, IC4 2015*, 2–7. https://doi.org/10.1109/IC4.2015.7375527

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

Han, J., Kamber, M., & Pei, J. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. https://doi.org/10.1109/ICMIRA.2013.45

Haq, F. U., Rachmat, H., Pascasarja, S., & Padjadjaran, U. (2020). *Penggunaan Google Review Sebagai Penilaian Kepuasan*. *2*(1), 10–12.

Himawan, H., Kaswidjanti, W., Sentimen, A., Sosial, M., & Based, L. (2018). *Metode Lexicon Based Dan Support Vector Machine Untuk Menganalisis Sentimen Pada Media Sosial Sebagai Rekomendasi Oleh-Oleh Favorit*. *2018*(November), 235–244.

Listiowarni, I. (2019). Implementasi Naïve Bayessian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *8*(2), 124. https://doi.org/10.32736/sisfokom.v8i2.652

Mahmudy, W. F., & Widodo, A. W. (2014). Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier yang Dimodifikasi. *Tekno*, *21*, 1–10.

Manning, C., Prabhakar Raghavan, & Hinrich, S. (2009). An Introduction to Information Retrieval. In *Introduction, An Retrieval, Information* (Issue c).

Murnawan, M. (2017). Pemanfaatan Analisis Sentimen Untuk Pemeringkatan Popularitas Tujuan Wisata. *Jurnal Penelitian Pos Dan Informatika*, *7*(2), 109. https://doi.org/10.17933/jppi.2017.070203

Nurrohmat, M. A., & SN, A. (2019). Sentiment Analysis of Novel Review Using Long Short-Term Memory Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(3), 209. https://doi.org/10.22146/ijccs.41236

Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017). Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat. *Majalah Ilmiah Teknologi Elektro*, *16*(3), 13. https://doi.org/10.24843/mite.2017.v16i03p03

Rizal, M. (2017). *Analisis Sentimen Pengguna Twitter Terhadap Objek Pariwisata di Indonesia Menggunakan Algoritma Pengolahan Deep Natural Langague dari IBM Insight Untuk Twitter*.

Saif, H., Fernandez, M., He, Y., & Alani, H. (2018). On stopwords, filtering and data sparsity for sentiment analysis of twitter. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2018*, *i*, 810–817.

Saputro, P. H., Aristin, M., & Tyas, Dy. L. (2017). Berdasarkan Lirik Menggunakan Metode Tf-. *Jurnal Teknoloi Informatika Dan Terapan*, *4*(1), 45–50.

Shafira, T. (2018). *Implementasi Convolutional Neural Networks Untuk Klasifikasi Citra Tomat Menggunakan KERAS: Vols. 10(2), 1–1*.

Suryadi. (2017). No Title. *Analisis Sentimen Riview Hotel Menggunakan Algoritme Naive Bayes Classifier Dan Pendekatan Lexicon Based*.

Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Systems with Applications*, *110*, 298–310. https://doi.org/10.1016/j.eswa.2018.06.022

Utami, P. D. (2018). Analisis Sentimen Review Kosmetik Bahasa Indonesia Menggunakan Algoritma Naïve Bayes. *Departemen Teknik Informatika & Komputer, Politeknik Negeri Jakarta Jl. Prof. Dr. G.A Siwabessy, Kampus Baru UI Depok, Indonesia, 16424*.

Wang, Y., Kim, K. T., Lee, B. J., & Youn, H. Y. (2018). Word clustering based on POS feature for efficient twitter sentiment analysis. *Human-Centric Computing and Information Sciences*, *8*(1). https://doi.org/10.1186/s13673-018-0140-y

Wilianto, L., Hendro Pudjiantoro, T., & Rakhmat Umbara, F. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naïve Bayes Classifier Studi Kasus Jawa Barat. *Prosiding SNATIF*, *0*(0), 439–448.

Yunitasari, Y., Musdholifah, A., & Sari, A. K. (2019). Sarcasm Detection For Sentiment Analysis in Indonesian Tweets. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(1), 53. https://doi.org/10.22146/ijccs.41136

Zulkifli, Wibowo, A. T., & Septiana, G. (2010). *Pembobotan Fitur pada Peringkasan Teks Bahasa Indonesia Menggunakan Algoritme Genetika*. *2*(2), 6481–6489.