**ANALISIS REAKSI PUBLIK TERHADAP PEMINDAHAN IBU KOTA NEGARA MENGGUNAKAN METODE *LEXICON BASED* DAN *KNN***

**TUGAS AKHIR**

Sebagai syarat untuk memperoleh gelar sarjana S-1 di Program Studi Informatika, Jurusan

Informatika, Fakultas Teknik Industri, Universitas Pembangunan Nasional “Veteran” Yogyakarta



Disusun oleh:

Rifqi Maulana

123200128

**PROGRAM STUDI INFORMATIKA**

**JURUSAN INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” YOGYAKARTA**

**2024**

**HALAMAN PENGESAHAN PEMBIMBING**

**ANALISIS REAKSI PUBLIK TERHADAP PEMINDAHAN IBU KOTA NEGARA MENGGUNAKAN METODE *LEXICON BASED* DAN *KNN***

Disusun Oleh:

Rifqi Maulana

123200128



Telah di

periksa

dan

disetujui

oleh pembimbing

pada tanggal : .................................

Menyetujui,

Pembimbing

Rifki Indra Perwira, S.Kom., M.Eng.

NIP 19830708 202121 1001

Mengetahui,

Koordinator Program Studi

Wilis Kaswidjanti, S.Si., M.Kom NIDN. 0513047601

**DAFTAR ISI**

COVER ..................................................................................................................................... 1

HALAMAN PENGESAHAN ................................................................................................... 2

DAFTAR ISI ............................................................................................................................. 3

BAB 1 PENDAHULUAN ......................................................................................................... 1

1.1 Latar Belakang ............................................................................................................ 1

1.2 Rumusan Masalah ....................................................................................................... 3

1.3 Batasan Masalah .......................................................................................................... 3

1.4 Tujuan Penelitian ......................................................................................................... 3

1.5 Manfaat Penelitian....................................................................................................... 3

1.6 Metodologi Penelitian dan Pengembangan Sistem ..................................................... 3

1.6.1 Metode Pengumpulan Data .................................................................................. 4

1.6.2 Metode Pengembangan Sistem ............................................................................ 4

1.7 Sistematika Penulisan .................................................................................................. 4

BAB II TINJAUAN LITERATUR ............................................................................................ 6

2.1 Analisis Sentimen ........................................................................................................ 6

2.2 Analisis Penerapan *K-Nearest Neighbor* …................................................................. 6

2.3 *X (Twitter)……*............................................................................................................. 6

2.4 *Scraping* ...................................................................................................................... 7

2.5 *Pre-processing* …….................................................................................................... 7

2.5.1 *Cleansing* dan *Case Folding* .............................................................................. 4

2.5.2 *Tokenization* ......................................................................................................... 4

2.5.3 *Stopwords Removal* ............................................................................................. 4

2.5.4 *Stemming*.............................................................................................................. 4

2.6 *Lexicon Based* …………………………………………............................................. 8

2.7 *K-Nearest Neighbor (KNN)* ......................................................................................... 9

2.8 Validasi dan Pengujian .............................................................................................. 12

2.9 Studi Pustaka ............................................................................................................. 14

BAB III METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM ..................... 16

3.1 Metodologi Penelitian ............................................................................................... 16

3.1.1 Metode Analisis Data ......................................................................................... 17

3.1.1.1 Analisis Masalah........................................................................................... 17

3.1.1.2 Pengumpulan Data........................................................................................ 17

3.1.1.3 *Pre-processing* Data................................................................................... 17

3.1.1.4 Pelabelan Data............................................................................................ 17

3.1.1.5 Pembobotan *Term-Invers Document Frequency* (TF-IDF) ........................ 17

3.1.1.6 Klasifikasi Model dengan *K-Nearest Neighbor (KNN)*............................... 17

3.1.1.6 Pengujian..................................................................................................... 17

3.1.2 Metode Pengembangan Sistem ........................................................................... 30

3.1.2.1 *Requirement Analysis*................................................................................... 17

3.1.2.2 *System and Software Design*........................................................................ 17

3.1.2.3 *Implementation*............................................................................................. 17

3.1.2.4 *System Testing*............................................................................................... 17

**DAFTAR GAMBAR**

Gambar 3. 1 Tahapan Penelitian .............................................................................................. 16

Gambar 3. 2 Flowchart Preprocessing ..................................................................................... 18

Gambar 3. 3 Flowchart *Cleansing* dan *Case Folding* .............................................................. 19

Gambar 3. 4 *Flowchart Tokenizing* .......................................................................................... 20

Gambar 3. 5 *Flowchart Stopword Removal* ............................................................................. 21

Gambar 3. 6 *Flowchart Stemming* ........................................................................................... 22 Gambar 3. 7 *Flowchart* TF-IDF .............................................................................................. 23

Gambar 3. 8 *Flowchart* Pelatihan SVM .................................................................................. 27

Gambar 3. 9 Arsitektur Sistem ................................................................................................. 31

Gambar 3. 10 *Flowchart* Perancangan Proses ......................................................................... 32

Gambar 3. 11 Rancangan Halaman List Rekomendasi Coffee Shop....................................... 33

Gambar 3. 12 Rancangan halaman Detail Ulasan ................................................................... 33

Gambar 3. 13 Rancangan Halaman Informasi Coffee Shop .................................................... 34

Gambar 3. 14 Rancangan Halaman Login ............................................................................... 34

Gambar 3. 15 Rancangan Halaman Dashboard ....................................................................... 35

Gambar 3. 16 Rancangan Halaman Data *Coffee Shop* ............................................................ 35

Gambar 3. 17 Rancangan Halaman Fitur Tambah Data *Coffee Shop* ...................................... 36

Gambar 3. 18 Rancangan Halaman Dataset ............................................................................ 36

Gambar 3. 19 Rancangan Halaman Uji Klasifikasi ................................................................. 37

Gambar 3. 20 Rancangan Halaman Data Hasil Pengujian ...................................................... 37 Gambar 3. 21 Rancangan Halaman Model SVM .................................................................... 38

**DAFTAR TABEL**

Tabel 2.1 *Confusion Matrix* .................................................................................................... 12 Tabel 2.2 Studi Pustaka ........................................................................................................... 14

Tabel 3.1 Contoh Penggunaan *Cleansing* dan *Case Folding*………………………………... 19

Tabel 3.2 Contoh Penggunaan *Tokenizing* .............................................................................. 20

Tabel 3.3 Contoh Penggunaan *Stopword Removal* ................................................................ 21

Tabel 3.4 Contoh Penggunaan *Stemming* ............................................................................... 22

Tabel 3.5 Contoh Hasil Pelabelan Data .................................................................................. 23

Tabel 3.6 Dokumen ................................................................................................................. 23

Tabel 3.8 Contoh Hasil Perhitungan TF .................................................................................. 24

Tabel 3.9 Contoh Hasil Perhitungan Bobot IDF ..................................................................... 25

Tabel 3.10 Contoh Hasil Perhitungan TF-IDF ........................................................................ 25

Tabel 3.11 Lanjutan Contoh Hasil Perhitungan TF-IDF ......................................................... 26

Tabel 3.12 Nilai pada Matriks K ............................................................................................. 27

Tabel 3.13 Nilai pada Matriks Hessian ................................................................................... 28

Tabel 3.14 Hasil Perhitungan 𝐸𝑖 ............................................................................................. 28

Tabel 3.15 Hasil Perhitungan 𝛿𝛼𝑖 ........................................................................................... 28

Tabel 3.16 Hasil Perhitungan 𝛼𝑖 baru ..................................................................................... 28

Tabel 3.17 Hasil Perhitungan 𝐸𝑖 iterani ke-2 .......................................................................... 29

Tabel 3.18 Hasil Perhitungan 𝛿𝛼𝑖 Iterasi ke-2 ........................................................................ 29

Tabel 3.19 Hasil Perhitungan 𝛼𝑖 baru ..................................................................................... 29

Tabel 3.20 Rancangan *Confusion Matrix* ................................................................................ 30

Tabel 3.21 Rancangan *K-Fold Cross Validation* ..................................................................... 30

Tabel 3.22 Kebutuhan *Hardware* ............................................................................................ 31

Tabel 3.23 Kebutuhan *Software* .............................................................................................. 31

Tabel 3.24 Rancangan Pengujian *Black Box* ........................................................................... 38

Tabel 3.25 Lanjutan Rancangan Pengujian *Black Box* ........................................................... 39

**BAB 1**

**PENDAHULUAN**

**1.1 Latar Belakang**

Keputusan Presiden Indonesia untuk memindahkan ibu kota negara ke luar Pulau Jawa menjadi salah satu proyek strategis yang tertuang dalam Rencana Pembangunan Jangka Menengah Nasional Tahun Anggaran 2020-2024. Pada 26 Agustus 2019, Presiden yang pada saat itu sedang menjabat mengumumkan ibu kota negara baru ini akan dibangun tepatnya di Kabupaten Penajam Paser Utara dan sebagian Kutai Kartanegara, Kalimantan Timur (Hadi, 2020). Pemindahan ibu kota negara Indonesia merupakan salah satu proyek mega infrastruktur yang paling signifikan dalam beberapa dekade terakhir. Keputusan berani ini tidak hanya memicu perdebatan di kalangan para pengambil kebijakan, tetapi juga memicu beragam reaksi dari masyarakat luas.

Transisi menuju ibu kota baru negara Indonesia merupakan hal yang sangat sensitif sehingga banyak dibicarakan di media sosial, tidak terkecuali pada media sosial *X* (sebelumnya dikenal dengan nama *Twitter*). *X* merupakan media sosial yang seringkali menjadi pusat *trending* mengenai isu di dunia baik itu skala nasional maupun internasional yang dijadikan warganet sebagai media untuk menyuarakan opini terkait sentimen terhadap apa pun yang terkini diperbincangkan di jejaring sosial yang begitu kompleks (Sandi et al., 2023). Sosial media *X* menjadi salah satu platform yang sering digunakan oleh masyarakat Indonesia untuk mengekspresikan pendapat dan respons terhadap peristiwa-peristiwa penting seperti pemindahan ibu kota negara baru. Indonesia menempati posisi ke-5 di dunia sebagai pengguna media sosial *X* terbanyak menurut data dari *Statista* pada bulan Januari 2023 (*Statista*, 2023). *Statista* merupakan salah satu situs *web* yang menyediakan data statistik yang dikenal di seluruh dunia.

Pada penelitian ini akan dilakukan analisa sentimen publik mengenai pemindahan ibu kota negara Indonesia di media sosial *X*. Teknik yang akan digunakan adalah *web scraping* untuk mengumpulkan data teks dari media sosial *X*. Penelitian ini akan dapat mengakses data teks yang mencakup berbagai macam opini, komentar, dan persepsi dari pengguna *X* terkait topik pemindahan ibu kota negara. Hal ini akan memungkinkan penulis untuk memiliki isi *dataset* dalam bentuk teks dan memadai untuk dilakukan analisis sentimen menggunakan metode klasifikasi *K-Nearest Neighbors (KNN)* dengan pelabelan sentimen *Lexicon Based*.

Metode *KNN* memberikan hasil yang kurang baik pada proses klasifikasi data karena terdapat fitur *noise*, namun terdapat beberapa penelitian yang menyebutkan bahwa performansi pada metode *machine learning* dapat menghasilkan performa yang baik ketika dikombinasikan dengan ekstraksi fitur dan seleksi fitur yang tepat. Penggabungan tersebut menawarkan pendekatan yang memadai dalam memproses dan menganalisis data teks untuk mengekstraksi informasi sentimen. Penerapan metode ini membuat kita dapat mengidentifikasi pandangan, sikap, dan respon yang diberikan publik terkait dengan topik pemindahan ibu kota negara tersebut.

Berdasarkan paparan sebelumnya, solusi yang penulis gunakan adalah metode klasifikasi *K-Nearest Neighbors* dengan pelabelan sentimen *Lexicon Based* serta ekstraksi fitur TF-IDF. Tahapan awal pada penelitian ini adalah *data preprocessing* bertujuan agar data yang nanti diproses lebih terstruktur, selanjutnya proses pelabelan dan ekstraksi fitur menggunakan metode TF-IDF dan leksikon, lalu terakhir yaitu tahap klasifikasi. Penulis memilih beberapa teknologi yang telah disebutkan sebelumnyakarena penulis sudah melakukan tinjauan pustaka dari beberapa metode sejenis sebelumnya. Penelitian yang dilakukan oleh Setyo Adji Pratomo, dkk pada tahun 2021, dengan judul “Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF dan *Lexicon* Pada Ulasan Film Menggunakan Metode *KNN*” Berkesimpulan bahwa penggabungan fitur ekstraksi TF-IDF dengan leksikon *SentiWordNet* memiliki hasil akurasi yang tidak lebih tinggi dibandingkan dengan hanya menggunakan fitur ekstraksi TF-IDF yaitu 73.31% dan penggunaan fitur seleksi *Information Gain* (*IG*) dengan *threshold* yang tepat mampu mengoptimasi hasil performansi. Lalu, penelitian yang dilakukan oleh Azhar pada tahun 2018, dengan judul “Analisis Kinerja Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Sentimen Analisis dengan Pendekatan Lexicon di Media Twitter”. *Dataset* yang digunakan merupakan data dari sosial media X (Twitter) denga­­­n menggunakan *Twitter API*. Proses *Natural Language Processing* yang digunakan adalah *case folding*, *filtering*, *tokenizing*, normalisasi, *stopwords*, dan *stemming*. Hasilnya nilai *KNN* pada *k*=5 dengan tingkat akurasi mencapai 77%. Terakhir, penelitian yang dilakukan oleh Muhammad Rayhan Elfansyah, dkk pada tahun 2024, dengan judul “Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naïve Bayes Terhadap Analisis Sentimen Pada Pengguna E-Wallet Aplikasi Dana Menggunakan Fitur Ekstraksi TF-IDF” berkesimpulan bahwa metode *KNN* dan *Naïve Bayes* memiliki akurasi yang berbeda berdasarkan sumber label data. Pada data yang diberi label model *Lexicon*, akurasi *KNN* mencapai 78% dan *Naïve Bayes* 74%.

Berdasarkan latar belakang yang telah dipaparkan, penulis berharap penelitian ini dapat memberikan kontribusi dalam memahami bagaimana reaksi publik, khususnya warganet di *X* (*Twitter*), terhadap pemindahan ibu kota negara Indonesia, apakah lebih cenderung positif, netral, atau negatif. Hal ini dilakukan dengan mengintegrasikan data teks yang telah dibersihkan untuk diproses lebih lanjut oleh teknik pelabelan sentimen berbasis leksikon yang telah dikustomisasi, serta mencari nilai *k* yang optimal sebagai variabel penting dalam optimasi metode klasifikasi *KNN,* tidak lupa juga menyertakan ekstraksi fitur TF-IDF untuk memperoleh prediksi sentimen dengan akurasi tinggi. Selain itu, penulis berharap penelitian ini dapat memberikan kontribusi kecil dalam memahami dinamika sosial yang muncul saat terjadi perubahan besar, seperti pemindahan ibu kota. Hasil penelitian ini diharapkan dapat menjadi referensi bagi penelitian serupa di masa depan.

**1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang sebelumnya, rumusan masalah yang didapatkan penulis mencakup bagaimana analisis sentimen menggunakan metode klasifikasi *K-Nearest Neighbors* dengan pelabelan sentimen *Lexicon Based* dapat membantu dalam memahami pandangan dan sikap masyarakat atau publik terhadap keputusan Presiden terkait pemindahan ibu kota negara Indonesia serta bagaimana optimasi metode tersebut untuk mendapatkan nilai akurasi yang tinggi?

**1.3 Batasan Masalah**

Agar masalah yang diteliti menghasilkan sasaran yang jelas, maka dibuatlah batasan masalah untuk menghindari adanya perluasan pembahasan kedepannya sebagai berikut:

1. Data penelitian yang digunakan merupakan data yang dihasilkan dari *scraping* pada unggahan teks di media sosial *X* (*Twitter*), mengenai sentimen publik atas keputusan Presiden terkait pemindahan ibu kota negara.
2. *Scraping* terhadap unggahan teks terbaru pada media sosial *X* (*Twitter*) dilakukan pada tanggal 21 Agustus 2024 dengan kata kunci: ikn, pemindahan ibu kota, ibu kota nusantara, ibu kota baru, dan ibu kota pindah.
3. *Scraping* dilakukan dengan kueri seperti “*since:2023-01-01*”, “*lang:id*”, serta fitur “*Latest*” pada kolom pencarian teks yang terdapat di media sosial *X* (*Twitter*)
4. Dataset yang digunakan dalam penelitian memakai format *.csv*.
5. *Preprocessing* data mencakup pembersihan duplikasi baris data, pembersihan teks, integrasi pembakuan kata dari kata slang menjadi kata formal, integrasi *stopwords*, dan *stemming*.
6. Kategori sentimen dibagi menjadi 3 yaitu positif, negatif, serta netral.
7. Kategori sentimen teks ditentukan oleh skor yang dihasilkan dari proses integrasi leksikon yang mana teks yang memiliki skor sentimen lebih dari 0 akan bernilai positif, kurang dari 0 akan bernilai negatif, dan sama dengan 0 akan bernilai netral.
8. Pelabelan sentimen dilakukan dengan pendekatan *Lexicon Based*.
9. Leksikon yang digunakan adalah 2 leksikon yang berbahasa Indonesia yaitu *InSet* (Indonesia Sentiment Lexicon) dari literatur yang disusun oleh Fajri Koto dan Gemala Y. Rahmaningtyas pada tahun 2017 dengan judul “InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs” serta leksikon *sentistrength\_id* dari literatur yang disusun oleh Wahid dan Azhari pada tahun 2016 dengan judul “Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity”. Leksikon *InSet* terdiri atas 3,609 kata positif dan 6,609 kata negatif dengan bobot antara -5 sampai +5.
10. Menggunakan fitur pembobotan kata TF-IDF.
11. Rasio perbandingan pembagian data *train* dan data *test* adalah 80:20.
12. Percobaan beberapa nilai *k* yakni *k1, k3, k5, k7,* dan *k9* untuk melatih model *K-Nearest Neighbors*.
13. Metode evaluasi model yang digunakan adalah *confusion matrix*, *accuracy, precision, recall,* serta *f1-score*.
14. Model *KNN* yang dilatih mencakup 2 leksikon yaitu *InSet* dan *sentistrength\_id* dengan masing-masing leksikon memiliki model *KNN* dari nilai *k* yang berbeda-beda (*k1, k3, k5, k7, k9)*.
    1. **Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah mengimplementasikan serta mengoptimasi metode klasifikasi *K-Nearest Neighbors* dengan pelabelan sentimen *Lexicon Based* dalam melakukan analisa sentimen masyarakat berdasarkan unggahan teks di media sosial *X* (*Twitter*) mengenai keputusan Presiden terkait pemindahan ibu kota negara.

* 1. **Manfaat Penelitian**

Penelitian ini diharapkan dapat bermanfaat untuk memberikan kontribusi kecil dalam pengembangan metodologi analisis sentimen berbasis teks, khususnya dalam penerapan algoritma klasifikasi *K-Nearest Neighbors* dengan pelabelan sentimen *Lexicon Based* dalam konteks bahasa Indonesia. Penelitian ini juga diharapkan tidak hanya bermanfaat secara akademis sebagai referensi bagi penelitian-penelitian selanjutnya yang menggunakan pendekatan serupa, penelitian ini juga mendukung pengembangan algoritma *Natural Language Processing* (*NLP*) untuk teks berbahasa Indonesia. Di samping itu, penelitian ini juga merupakan bagian dari pemenuhan salah satu syarat kelulusan strata satu (S1) Program Studi Informatika Fakultas Teknik Industri.

**1.6 Metodologi Penelitian dan Pengembangan Sistem**

Metode penelitian ini menggunakkan metode penelitian kuantitatif. Metode penelitian kuantitatif merupakan penelitian empiris dimana data dalam bentuk sesuatu yang dapat dihitung atau angka (Punch, 1988). Berikut merupakan tahapan-tahapan penelitian yang dilakukan:

## 1.6.1 Metode Pengumpulan Data

Penelitian ini menggunakan teknik *web scraping* untuk mengumpulkan data teks yang berkaitan dengan pemindahan ibu kota negara dari platform media sosial *X* (*Twitter*).

**1.6.2 Metode Pengembangan Sistem**

Metode *Waterfall* merupakan salah satu model *SDLC (Software Development Life Cycle)* yang sering digunakan dalam pengembangan sistem informasi atau perangkat lunak. Metode ini menggunakan pendekatan sistematis dan berurutan. Tahapan dalam model ini dimulai dari tahap perencanaan hingga tahap pengelolaan (*maintenance*) dan dilakukan secara bertahap (Abdul Wahid, 2020). Tahapan metode *waterfall* adalah sebagai berikut:

1. *Requirements Analysis and Definition*

Merupakan tahapan awal yang melibatkan identifikasi dan pemahaman yang mendalam terhadap kebutuhan. Tujuan utamanya yaitu mengumpulkan persyaratan fungsional dan non-fungsional yang nantinya akan menjadi dasar dari pengembangan sistem.

1. *System and Software Design*

Tahapan perancangan sistem ini mengalokasikan kebutuhan-kebutuhan sistem pada perangkat keras maupun perangkat lunak dengan membentuk arsitektur sistem secara keseluruhan.

1. *Implementation*

Pada tahap ini, perancangan perangkat lunak direalisasikan sebagai serangkaian program.

1. *System Testing*

Merupakan tahap pengujian terhadap sistem yang telah dibuat yang bertujuan untuk mengetahui apakah sistem yang dibuat udah siap digunakan atau belum.

## 1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

|  |  |
| --- | --- |
| BAB I | PENDAHULUAN  Pada bagian pendahuluan membahas mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, serta sistematika penulisan. |
| BAB II | TINJAUAN PUSTAKA  Tinjauan pustaka merupakan bagian yang memuat mengenai dasar teori yang digunakan untuk analisis serta perancangan sistem dan juga implementasi pada penelitian ini. Selain itu juga digunakan sebagai bahan referensi serta pondasi untuk memperkuat argumentasi pada penelitian ini. |
| BAB III | METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM  Pada bab ini membahas mengenai metodologi penelitian, analisis sistem dan perancangan sistem analisis sentimen. |
| BAB IV | HASIL, PENGUJIAN, DAN PEMBAHASAN  Bab ini menyajikan hasil dari penelitian yang berisi hasil implementasi dari perancangan yang telah dibuat pada bab sebelumnya. Selain itu berisi pengujian terhadap hasil penelitian beserta pembahasannya. |
| BAB V | KESIMPULAN DAN SARAN  Bab ini berisi kesimpulan dari hasil penelitian serta saran yang diajukan oleh penulis untuk pengembangan pada penelitian selanjutnya. |

**BAB II**

**TINJAUAN LITERATUR**

**2.1 Analisis Sentimen**

Analisis sentimen adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap dan emosi seseorang terhadap sebuah produk, organisasi, individu, masalah, peristiwa atau topik (Liu, 2012). Tujuan dari analisis sentimen yaitu untuk memahami opini, perasaan, serta pandangan yang terkandung pada teks atau data unstruktural lainnya. Pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian mengenai analisis sentimen berkembang pesat, serta kurang lebih 20-30 perusahaan di Amerika berfokus pada layanan analisis sentimen (Liu, 2012). Manfaat sentimen analisis dalam dunia usaha antara lain untuk melakukan pemantauan terhadap suatu produk. Secara cepat dapat digunakan sebagai alat bantu untuk melihat respon masyarakat terhadap suatu produk, sehingga dapat diambil langkah strategis berikutnya. Garis besar analisis sentimen itu sendiri bertujuan untuk mengekstrak atribut dan komponen dari beberapa komentar yang ada di media sosial dan sehingga dapat menentukan beberapa kelas positif, negatif dan netral (Permatasari et al., 2021).

Pada umumnya, sentimen analisis merupakan klasifikasi tetapi kenyataannya tidak semudah proses kualifikasi biasa karena terkait penggunaan bahasa, dimana terdapat ambigu dalam penggunaan kata serta perkembangan bahasa itu sendiri.

Menurut Liu (2012), analisis sentimen memiliki beberapa tahap untuk melakukan analisis sentimen, yaitu:

1. Level Dokumen

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

1. Level Kalimat

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

3. Level Aspek

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

* 1. **Analisis Penerapan *K-Nearest Neighbor***

*K-Nearest Neighbor* merupakan teknik yang digunakan untuk mengidentifikasi adanya persamaan antara data baru dan lama. Kemudian, teknik ini akan memasukkan data baru tersebut dalam kategori yang paling mirip dengan kategori yang telah ada sebelumnya. Dengan kata lain, *K-Nearest Neighbor* menyimpan seluruh data lama dan mengklasifikasikan *data point* baru berdasarkan kemiripan.

Proses *training* pada penelitian ini dipakai untuk menghasilkan model sentimen. Tahapan algoritma klasifikasi sentimen dengan *data testing* ataupun data baru. Pada proses *training* akan digunakan metode *Lexicon Based*.

* 1. ***X (Twitter)***

*X* adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh Twitter Inc., yang menawarkan jaringan sosial berupa *microblog* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan dengan sebutan *tweet* atau kicauan(Akbar et al., 2013).

*Microblog* adalah jenis alat komunikasi daring dengan manfaat agar pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu. *Tweet* atau kicauan adalah teks tulisan hingga 140 karakter (atau lebih jika berlangganan fitur khusus pada platform tersebut) yang ditampilkan pada halaman profil pengguna. *Tweet* bisa dilihat secara publik atau dapat dibatasi pengiriman pesan ke daftar pengguna lain tertentu saja. Pengguna dapat melihat *tweet* pengguna lain yang dikenal dengan sebutan pengikut atau *followers* (Ramadhon, 2020).

* 1. ***Scraping***

*Web scraping* merupakan sebuah teknik untuk mendapatkan informasi dari situs web secara otomatis tanpa harus menyalinnya secara manual (Ayani et al., 2019). Teknik *scraping* memungkinkan konten utama yang terdapat pada situs dapat diekstraksi, dihimpun, kemudian dapat diproses. *Web scraping* akan melakukan ekstraksi data pada *World Wide Web* (*www)*, lalu data yang didapat akan disimpan pada *file system* atau basis data yang nantinya bisa diambil kembali atau dianalisis (Setiawan et al., 2020). Cara kerja *web scraping* adalah dengan mengakses halaman pada *web*, menentukan data yang dalam halaman tersebut, melakukan ekstraksi, dan transformasi bila diperlukan, kemudian menyimpan data tersebut menjadi dataset terstruktur (Boeing, 2016).

* 1. ***Pre-processing***

Tahapan *pre-processing* merupakan tahapan awal untuk mempersiapkan dokumen agar lebih mudah untuk diproses, tahapan *pre-processing* sebelum proses klasterisasi meliputi *cleansing, case folding*, *tokenizing*, *filtering*, dan *stemming* (Amalia et al., 2018). Berdasarkan hal tersebut, mengubah data yang sebelumnya tidak terstruktur memerlukan proses pengubahan menjadi data yang terstruktur untuk diproses pada langkah berikutnya. Berikut adalah beberapa tahapan dari *pre-processing*:

## 2.5.1 *Cleansing* dan *Case Folding*

*Cleansing* merupakan proses pembersihan data yang bertujuan untuk menghilangkan elemen-elemen yang tidak dibutuhkan pada sebuah dokumen. *Cleansing* bertujuan untuk memperbaiki kualitas data. Pada penelitian ini tahapan *cleansing* berfungsi untuk menghilangkan *mentions*, *hashtag*, *url* dan *uri*, tanda baca, *emoji*, serta menghilangkan angka-angka.

*Case folding* merupakan tahapan yang bertujuan untuk mengubah semua huruf yang terdapat pada dokumen menjadi huruf kecil. Huruf ‘a’ sampai ‘z’ yang akan diterima, apabila pada dokumen terdapat karakter selain huruf maka akan dihilangkan dan dianggap *delimiter*. Tahap *case folding* akan menghasilkan kalimat yang sudah rapi dan akan memudahkan proses selanjutnya.

## 2.5.2 *Tokenization*

*Tokenization* atau pemisahan rangkaian *term* merupakan tahapan pemisahan teks menjadi potongan-potongan berupa huruf, kata, atau kalimat menjadi kata yang tidak terhubung. Data *tweets* pada *X (Twitter)* yang masuk pada tahap *tokenization* akan diubah menjadi potongan-potongan kata. Pada umumnya, karakter spasi membedakan atau mengidentifikasi setiap kata satu sama lain. Sehingga, proses *tokenization* pada dokumen bergantung pada karakter spasi.

## 2.5.3 *Stopwords Removal*

Penghapusan *stopwords* bertujuan untuk menghilangkan kata yang dianggap tidak memiliki makna atau tidak relevan yang terdapat pada dokumen. Contoh *stopwords* yang ada pada Bahasa Indonesia seperti “yang”, “dan”, “di”, “itu”, “adapun”, “agak” dan sebagainya. Kata-kata umum tersebut tidak mempunyai nilai pada sebuah dokumen, sehingga kata-kata yang termasuk dalam kamus *stopwords* dihilangkan sehingga ukuran data juga akan berkurang.

## 2.5.4 *Stemming*

*Stemming* merupakan proses yang bertujuan untuk menemukan kata dasar dari sebuah kata dengan menghapus imbuhan (*affixes*), termasuk awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), serta kombinasi dari awalan dan akhiran (*confixes*) pada kata turunan. Tujuan utama dari proses *stemming* ini adalah mengubah bentuk kata menjadi kata dasar sesuai dengan Bahasa Indonesia yang baik dan benar. Menurut algoritma Nazief & Adriani memiliki tahap-tahap sebagai berikut:

1. Mencari kata yang akan distem pada kamus, apabila ditemukan maka diasumsikan bahwa kata tersebut merupakan kata akar (*root word)*, jika terbukti maka proses diberhentikan pada tahap pertama.
2. Menghilangkan *inflection suffixes*, sebuah kata yang mengandung *inflection suffixes* yaitu apabila memiliki imbuhan “-lah”, ”-ku”, ”-kah”, ”-mu”, atau “-nya”, apabila kata berupa *particles* atau dalam kata lain yang mengandung “-lah”, ”-kah”, atau “-pun”, maka langkah untuk menghilangkan *inflection suffixes* diulangi agar dapat menghilangkan *possessive pronouns*, yang termasuk imbuhan *possessive pronouns* adalah “-ku”, “-mu”, atau “-nya”.
3. Menghapus *derivation suffixes* atau imbuhan turunan, yang termasuk pada kata imbuhan turunan adalah “-i”, ”-an”, atau “-kan”.
4. Menghapus *derivational prefix* atau imbuhan yang berada pada awal kata, yang dimaksud dengan *derivational prefix* adalah “be-”, ”di-“, ”ke-“, ”me-“, ”pe-“, ”se-“, dan “te-“.
5. Apabila 4 langkah tersebut telah dilakukan tetapi belum berhasil menemukan kata dasar maka algoritma ini akan melakukan analisis apakah kata tersebut termasuk ke dalam tabel ambiguitas kolom terakhir.
6. Apabila belum berhasil maka algoritma akan dikembalikan pada kata aslinya. *Stemming* pada *phyton* dilakukan melalui kelas *StemmerFactory* yaitu sebuah kelas yang terdapat pada *library* yang bernama “*Sastrawi”* dan kompatibel dengan *input* berbahasa indonesia, yang dimana *library* ini akan lebih dulu di-*import* sebelum meng-*import* kelas *StemmerFactory*.

**2.6 *Lexicon Based***

Metode *Lexicon Based* merupakan metode yang sederhana, layak, dan praktis untuk analisis sentimen dari data media sosial. Data yang cocok dengan metode *Lexicon Based* yaitu data kuesioner, data *Twitter*, data *Facebook*, atau media sosial lainnya yang berupa opini pelanggan tentang suatu produk atau pelayanan jasa (Matulatuwa et al., 2017).

*Lexicon Based* didasarkan pada asumsi bahwa orientasi sentimen kontekstual adalah jumlah dari orientasi sentimen setiap kata atau frasa. Metode ini dapat digunakan untuk melakukan ekstraksi sentimen dari blog dengan mengombinasikan *lexical knowledge* dan klasifikasi teks. Metode *Lexicon* dapat dibuat secara manual atau diperluas secara otomatis dari *seed of words* (Matulatuwa et al., 2017).

Kamus adalah komponen penting dalam sistem yang menggunakan *Lexicon Based*. Kamus digunakan dalam proses normalisasi kalimat dan ekstraksi kata kunci. Berikut adalah contoh kamus dan isinya (Matulatuwa et al., 2017).

1. *Positive keywords:* baik, bagus, bisa, *ok*, cepat, akurat, aman, senang.
2. *Negative keywords:* acuh, ambigu, bodoh, gagal, abnormal, susah, lambat.
3. *Negation keywords:* seharusnya, bukan, tidak.

(Hu et al., 2004 dikutip Liu, 2012) mengusulkan algoritma berbasis *Lexicon* untuk aspek penentuan orientasi sentimen. Orientasi sentimen kalimat itu ditentukan dengan menjumlahkan nilai orientasi semua kata sentimen pada kalimat. Sebuah kata positif diberi nilai sentimen dari +1 dan kata negatif diberi nilai sentimen dari -1. Kata negasi dan kata-kata yang bertentangan misalnya, “tetapi” dan “namun” juga dipertimbangkan. Ada empat langkah dalam menentukan orientasi sentimen berdasarkan pendekatan *Lexicon* yaitu (Ramadhon, 2020):

1. Tandai kata yang mengandung sentimen untuk setiap kalimat yang berisi satu atau lebih kata sentimen, langkah ini menandai semua kata dan frasa dalam sentimen kalimat. Setiap kata positif diberikan skor sentimen +1 dan setiap kata negatif diberikan skor sentimen -1. Contoh “Kualitas baju ini tidak baik [+1], tapi daya tahannya lama [+1]”. Berdasarkan contoh tersebut, kata “baik” bernilai +1 dan tahan lama bernilai +1 karena termasuk kata positif.
2. Terapkan sentimen *shifter*. *Shifter* adalah kata-kata dan frasa yang dapat mengubah orientasi sentimen. Ada beberapa jenis *shifter* kata negasi seperti “tidak”, “tidak pernah”, dan “tidak ada” adalah jenis yang paling umum. Berdasarkan hal tersebut, kalimat menjadi “Kualitas baju ini tidak baik [-1], tapi daya tahannya lama [+1]” karena kata negasi “tidak”.
3. Berlakukan *Agregat*. Langkah ini berlaku fungsi agregasi pendapat dengan skor sentimen yang dihasilkan untuk menentukan orientasi akhir dari sentimen.

|  |  |  |
| --- | --- | --- |
| Keterangan: | | |
| *w I SO* |  | : Jumlah (*score*) fitur ke *i* setiap sentimen (positif dan negatif) |
| *dis (wi,f)* |  | : Jumlah fitur yang berhasil ditemukan |

|  |
| --- |
| **2.7 *K-Nearest Neighbor (KNN)*** |

Algoritma *KNN* digunakan untuk klasifikasi objek baru berdasarkan atribut dan data *training* atau data latih dimana hasil dari data uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *KNN*. Algoritma *KNN* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari data uji yang baru (Krisandi et al., 2013). Data latih akan dibangun dengan memperhatikan keseimbangan dokumen satu sama lain. Adapun algoritma *KNN* dapat dijelaskan (Kurniawan, 2012) sebagai berikut:

1. Menentukan parameter nilai *k* = jumlah tetangga terdekat.
2. Hitung jarak antara data yang akan dievaluasi (data uji) dengan semua data latih yang telah dibangun. Penghitungan jarak dilakukan dengan rumus berikut:

|  |  |  |
| --- | --- | --- |
| Keterangan: | | |
|  |  | : Sampel Data |
|  |  | : Data Uji |
| *i* |  | : Variabel Data |
| *d* |  | : Jarak |
| *p* |  | : Dimensi |

1. Mengurutkan jarak yang terbentuk (urut naik) dan menentukan jarak terdekat sampai urutan ke-*k*
2. Pasangkan kategori atau kelas yang bersesuaian.
3. Cari jumlah terbanyak dari tetangga terdekat tersebut. Kemudian, tetapkan kategori tersebut sebagai kategori dari data yang dicari.

## 2.8 Validasi dan Pengujian

Validasi dan pengujian sangat diperlukan untuk menilai kinerja dari sebuah sistem. Pada penelitian ini validasi yang akan digunakan yaitu metode *k-fold cross validation*. Pengujian ini digunakan untuk melakukan validasi suatu model dengan pengujian sebanyak *K*. Pada *k-fold cross validation* ini data pelatihan dibagi menjadi *K* bagian yang setara, yang dikenal sebagai “*fold*”. Selama validasi berlangsung, satu *fold* digunakan sebagai set data validasi, sedangkan sisa datanya digunakan sebagai data latih. Proses ini akan diulangi sebanyak *K* kali, untuk memastikan setiap *fold* menjadi data validasi satu kali yang terbaik. *K-fold cross validation* merupakan salah satu metode yang dapat memeriksa *overfitting* atau terlalu cocok dengan data latih pada suatu model.

*Confusion matrix* merupakan suatu proses yang digunakan untuk menganalisa keakuratan dari model klasifikasi yang dibuat untuk mengidentifikasi data dengan kelas yang berbeda (Afrillia et al., 2022). Pengujian dengan *confusion matrix* ini digunakan untuk menghitung nilai *true positive*, *false positive*, *true negative*, serta *false negative* yang nantinya dapat digunakan untuk pengukuran nilai akurasi, presisi, serta *recall*. Dengan melakukan pengukuran Tingkat akurasi maka dapat mengetahui seberapa baik performa model klasifikasi tersebut. Bentuk dari *confusion matrix* adalah tabel dengan empat kombinasi yang berbeda antara nilai prediksi dan nilai aktual. Metode *confusion matrix* mempunyai empat kemungkinan yang merepresentasikan hasil dari proses klasifikasi, yaitu TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*).

Berikut merupakan tabel *confusion matrix*:

# **Tabel 2. 1 *Confusion Matrix***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | ***Predicted*** | |  |
| **-1 *(Negative)*** | **+1 *(Positive)*** | **0 *(Neutral)*** |
| ***Actual*** | -1 *(Negative)* | TN | FP | FL2 |
| +1 *(Positive)* | FN | TP | FP2 |
| 0 (*Neutral*) | FN2 | FP2 | TL |

Keterangan :

|  |  |
| --- | --- |
| TN (*True Negative*) | : Prediksi benar bernilai negatif |
| TP (*True Positive*) | : Prediksi benar bernilai positif |
| TL (*True Neutral*) | : Prediksi benar bernilai netral |
| FN (*False Negative*) | : Negatif terprediksi positif |
| FP (*False Positive*) | : Positif terprediksi negatif |
| FL (False Neutral) | : Positif terprediksi netral |

FN2 (*False Negative* 2) : Netral terprediksi negatif

FP2 (*False Positive* 2) : Netral terprediksi positif

FL2 (*False Neutral* 2) : Negatif terprediksi netral

Tabel *confusion matrix* di atas dapat digunakan dalam perhitungan *performance matrix* yang bertujuan untuk mengukur model yang digunakan untuk dapat memperoleh nilai *accuracy*, *recall*, dan *precision*.

## 1. *Accuracy*

Akurasi merupakan nilai yang menunjukan kedekatan antar nilai prediksi dan nilai aktual. Perhitungan akurasi dengan cara membagi jumlah data yang akan diklasifikasi secara tepat dengan totalsampel data *testing* yang diuji. Berikut merupakan rumus perhitungan nilai akurasi:

## 2. *Recall*

*Recall* merupakan perbandingan jumlah data yang dilakukan prediksi pada kelas positif yang benar dengan jumlah data yang diharapkan berada pada kelas positif. *Recall* dikatakan sebagai tingkat keberhasilan model dalam menemukan informasi. Berikut merupakan rumus perhitungan nilai *recall*:

## 3. *Precision*

*Precision* merupakan tingkat akurasi antar informasi yang diinginkan pengguna serta respon sistem. Nilai presisi menunjukkan data positif yang diklasifikasi dengan tepat kemudian dilakukan pembagian dengan jumlah data positif yang diklasifikasi. Berikut merupakan rumus perhitungan nilai presisi:

## 4. *F1-Score*

*F1-Score* merupakan perbandingan rata-rata *precision* dan *recall* yang telah dibobotkan. Berikut merupakan rumus perhitungan nilai *F1-Score*:

**2.9 Studi Pustaka**

Penelitian di bawah ini merupakan penelitian-penelitian yang telah dilakukan sebelumnya dan berkaitan dengan penelitian tugas akhir ini, sehingga menjadi referensi dalam penelitian ini.

**Tabel 2.2 Studi Pustaka**

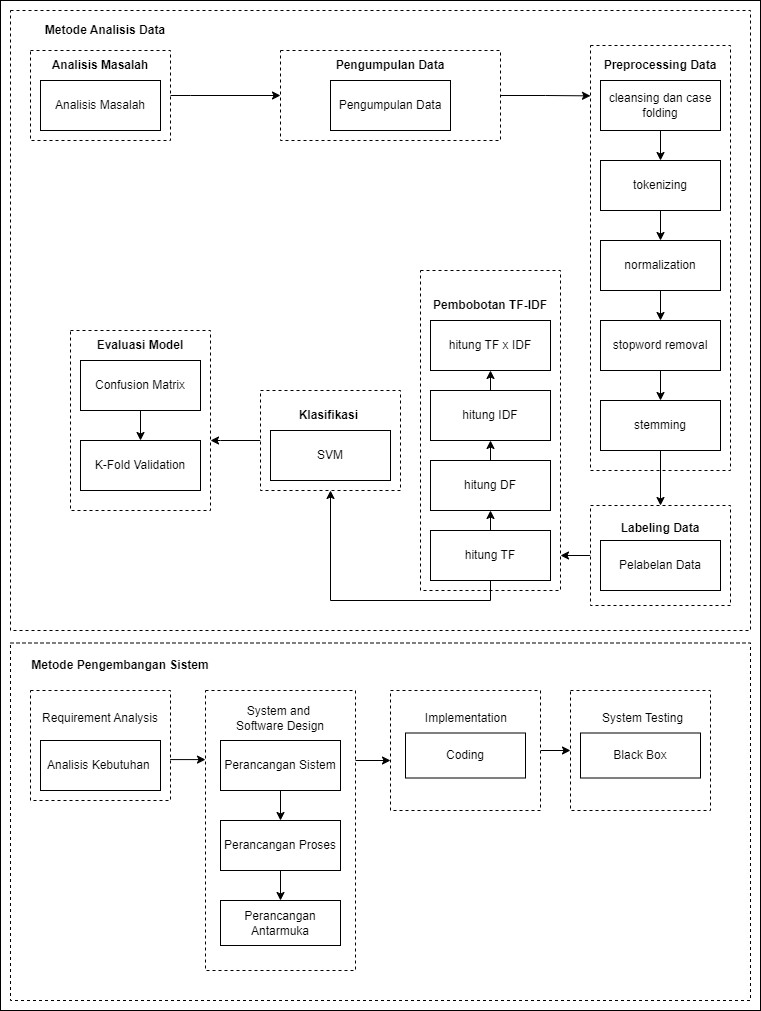
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Penulis** | **Judul** | **Metode** | **Hasil** |
| 1. | Lestari, 2022. | Analisis Sentimen Ibu Kota Negara Baru Indonesia Pada Twitter Menggunakan Algoritma Support Vector Machine (SVM) dan Seleksi Fitur Chi Square | SVM dan Chi Square |  |
| 2. | Saidah, Mayary, 2020. | Analisis  Sentimen Pengguna Twitter Terhadap Dompet  Elektronik Dengan Metode Lexicon Based  Dan K-Nearest Neighbor | Lexicon Based dan KNN |  |
| 3. | Habibah, Budianita, Fikry, Iskandar, 2023. | Analisis Sentimen Mengenai  Penggunaan E-Wallet Pada Google Play  Menggunakan Lexicon Based dan K-Nearest  Neighbor | Lexicon Based dan KNN |  |
| 4. | Fatra, Hayatin, Aditya, 2020. | Analisa Sentimen Tweet Berbahasa  Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota  Indonesia | Lexicon |  |
| 5. | Almaisah, 2022. | Analisis Sentimen  Masyarakat Terhadap Acara Televisi  Indonesia Pada Media Sosial Twitter  Menggunakan Pendekatan Algoritma  K-Nearest Neighbor (KNN) dan Lexicon  Based | Lexicon Based dan KNN |  |
| 6. | Aripiyanto, Tukino, Sufyan, Nandaputra, 2022. | Sentimen Analisis  Twitter Ibu Kota Negara Nusantara  Menggunakan Long Short-Term Memory dan  Lexicon Based | Long Short-Term Memory dan  Lexicon Based |  |
| 7. | Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, Fitri Nurapriani, 2023. | Analisis Sentimen  Relokasi Ibukota Nusantara Menggunakan  Algoritma Naïve Bayes dan KNN | Naïve Bayes dan KNN |  |
| 8. | Rachmadana Ismail, Bagus, Hakim, 2023. | Implementasi Lexicon  Based Untuk Analisis Sentimen Dalam  Mengetahui Trend Wisata Pantai Di  Yogyakarta Berdasarkan Data Twitter. | Lexicon Based |  |
| 9. | Dwiki, Putra, Juanita, 2021. | Analisis Sentimen Pada Ulasan Pengguna  Aplikasi Bibit Dan Bareksa Dengan  Algoritma KNN | KNN |  |
| 10. | Putriani, Umbara, Sabrina, 2022. | Analisis Sentimen pada Aplikasi  PeduliLindungi dengan Menggunakan Metode  Improved K-Nearest Neighbor dan Lexicon  Based | KNN dan Lexicon Based |  |

**BAB III**

**METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**

**3.1 Metodologi Penelitian**

Pada bagian ini akan dibahas mengenai metodologi penelitian serta pengembangan sistem yang akan dilakukan dalam penelitian ini. Dalam penelitian ini yaitu menentukan analisis sentimen *coffee shop* di Yogyakarta berdasarkan ulasan yang ada pada *google maps* yang nantinya akan dijadikan sebagai rekomendasi *coffee shop* di Yogyakarta dengan menggunakkan metode *Support Vector Machine* dan *TF-IDF*. metodologi penelitian merupakan sub bab yang menggambarkan alur kerja serta tahapan pada penelitian ini. Tahapan metodologi penelitian pada tugas akhir ini yaitu pengumpulan data, *text preprocessing*, pelabelan data, pembobotan kata (TF-IDF), pengujian model dengan *Support Vector Machine*, pengujian sistem, model terbaik dan analisis serta visualisasi hasil. Berikut merupakan tahapan pada metodologi penelitian ini dapat dilihat pada Gambar 3.1.



**Gambar 3. 1 Tahapan Penelitian**

## 3.1.1. Metode Analisis Data

**3.1.1.1 Analisis Masalah**

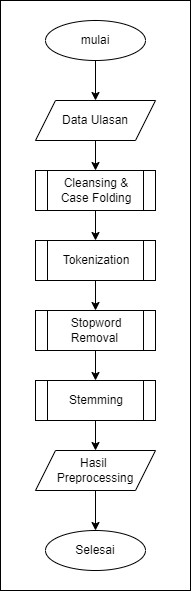
Analisis masalah merupakan Gambaran masalah dari penelitian tugas akhir Analisis Sentimen untuk Rekomendasi Tempat Kuliner di Yogyakarta menggunakkan metode *Support Vector Machine* dan TF-IDF. Penelitian ini dilakukan untuk merancang sistem yang dapat menghasilkan klasifikasi berdasarkan *review* atau ulasan *coffee shop* yang tedapat pada *google maps* yang bersentimen positif, negatif, ataupun netral. Pada sistem ini juga akan menampilkan *list* rekomendasi *coffee shop* yang ada di Yogyakarta.

**3.1.1.2 Pengumpulan Data**

Pada penelitian ini data yang dibutuhkan dikumpulkan dengan cara melakukan scraping data pada *google maps*. Tahapan scraping pada penelitian ini bertujuan untuk pengambilan data serta informasi secara spesifik dan juga otomatis dari *google maps*. Proses *scraping* pada google maps dilakukan menggunakkan *website Outscapper*. Data yang diambil secara spesifik yaitu data nama *coffee shop*, alamat *coffee shop*, *review text*, serta *rating coffee shop* yang berada di Yogyakarta. Data yang berhasil diambil dari tahapan scraping disimpan kedalam file berformat csv.

**3.1.1.3 Preprocessing Data**

Tahap *preprocessin*g merupakan serangkaian tahap yang sangat penting dalam klasifikasi. Tahap *preprocessing* bertujuan untuk membersihkan data, serta mempersiapkan data agar data yang akan digunakan tidak mengandung *noise* dan juga menghasilkan teks yang lebih jelas,singkat, serta padat tanpa mengurangi makna yang ada pada teks tersebut. Pada tahap preprocessing data terdapat beberapa tahapan seperti *cleansing*, *case folding*, *tokenizing*, *stopwords removal*, serta *stemming*. Berikut merupakan flowchart dari tahap preprocessing yang dapat dilihat pada Gambar 3.2.

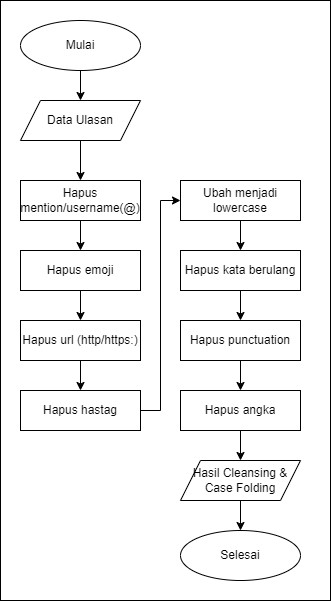


**Gambar 3. 2 *Flowchart Preprocessing***

Gambar 3.2 merupakan *flowchart* dari tahapan *preprocessing*, berdasarkan *flowchart* tersebut berikut merupakan penjelasan dari tahapan *preprocessing* yang dilakukan pada penelitian ini :

a. *Cleansing* dan *Case Folding*

*Cleansing* merupakan proses membersihkan data teks yang berupa tanda baca, link angka atau nomor, serta menghapus spasi untuk menghilangkan terjadinya *noise* pada data teks. *Case folding* merupakan tahap dimana data akan diubah menkadi huruf kecil. Berikut merupakan alur serta contoh hasil dari *cleansing* dan *case folding* dapat dilihat pada Gambar 3.3.



**Gambar 3. 3 *Flowchart* *Cleansing* dan *Case Folding***

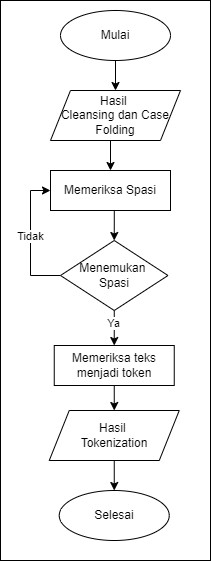
Berikut contoh hasil dari tahap *cleansing* dan *case folding* dapat dilihat pada Tabel 3.1.

## Tabel 3. 1 Contoh Penggunaan *Cleansing* dan *Case Folding*

|  |  |  |
| --- | --- | --- |
| **No.** | **Data *Review*** | **Hasil *Cleansing* dan *Case Folding*** |
| 1. | Kopi ENAK POL, tempat nyaman Untuk ngerjain tugas, pelayanan Ramah | kopi enak tempat nyaman untuk ngerjain tugas pelayanan ramah |
| 2. | Tempat nyaman, banyak tanaman, suasana asik, di Jogja Selatan. Yang paling PENTING kopinya cucok!!! | tempat nyaman banyak tanaman suasana asik di jogja selatan yang paling penting kopinya cucok |

b. *Tokenizing*

Setelah dilakukan tahapan *cleansing* dan *case folding*, tahap berikutnya yaitu tokenization Dimana pada tahap ini berfungsi untuk memisahkan kalimat pada teks menjadi satu kata berdasarkan spasi. Berikut merupakan alur serta contoh hasil *tokenizing* dapat dilihat pada Gambar 3.4.



**Gambar 3. 4 *Flowchart Tokenizing***

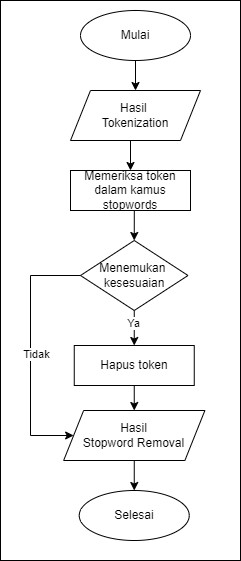
Berikut contoh hasil dari tahap *tokenizing* dapat dilihat pada Tabel 3.2.

**Tabel 3. 2 Contoh Penggunaan *Tokenizing***

|  |  |  |
| --- | --- | --- |
| **No.** | **Data *Review*** | **Hasil *Tokenizing*** |
| 1. | kopi enak tempat nyaman untuk ngerjain tugas pelayanan ramah | “kopi” “enak” “tempat” “nyaman” “untuk” “ngerjain” “tugas” “pelayanan” “ramah” |
| 2. | tempat nyaman banyak tanaman suasana asik di jogja selatan yang paling penting kopinya cucok | “tempat” “nyaman” “banyak” “tanaman” “suasana” “asik” “di”  “jogja” “selatan” “yang”  “paling” “penting” “kopinya” “cucok” |

## c. *Stopword Removal*

Tahapan ini merupakan tahapan yang berfungsi untuk menghilangkan kata yang tidak memiliki definisi yang jelas. apabila kata tersebut masuk dalam daftar *stopword* maka kata tersebut tidak bisa diproses, sedangkan apabila kata tidak masuk dalam daftar *stopword* maka akan masuk pada proses selanjutnya. Berikut merupakan alur dan contoh hasil dari tahap *stopword removal* dapat dilihat pada Gambar 3.5.



**Gambar 3. 5 *Flowchart Stopword Removal***

Berikut contoh hasil dari tahap *stopword removal* dapat dilihat pada Tabel 3.3.

**Tabel 3. 3 Contoh Penggunaan *Stopword Removal***

|  |  |  |
| --- | --- | --- |
| **No.** | **Data Review** | **Hasil *Stopword Removal*** |
| 1. | “kopi” “enak” “tempat” “nyaman” “untuk”  “ngerjain” “tugas” “pelayanan” “ramah” | “kopi” “enak” “tempat” “nyaman” “ngerjain” “tugas” “pelayanan” “ramah” |
| 2. | “tempat” “nyaman” “banyak” “tanaman” “suasana” “asik” “di” “jogja” “selatan” “yang” “paling” “penting” “kopinya” “cucok” | “tempat” “nyaman”  “banyak” “tanaman”  “suasana” “asik” “jogja” “selatan” “penting”  “kopinya” “cucok” |

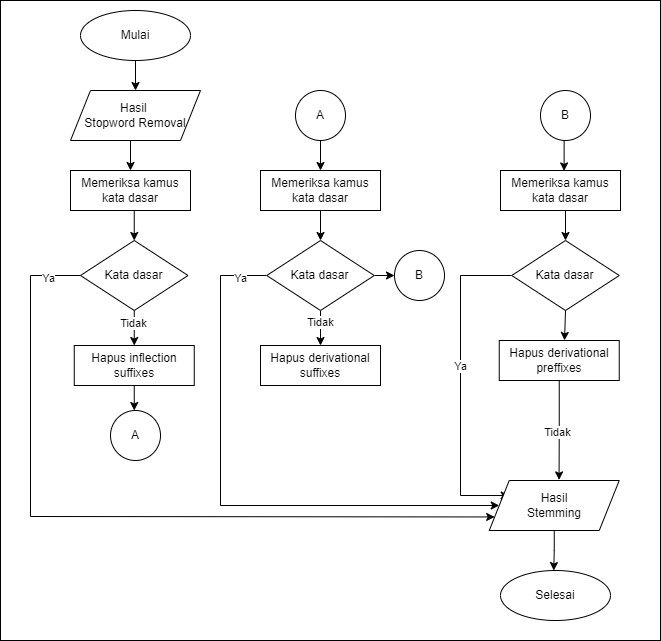
## d. *Stemming*

Tapah *stemming* merupakan tahap untuk pembentukan kata dasar. Pada tahapan ini menerapkan *library* dari sastrawi. Berikut merupakan alur dan contoh hasil pada tahap *stemming* dapat dilihat pada Gambar 3.6.

**Gambar 3.**

**6**

***Flowchart Stemming***



Berikut contoh hasil dari tahap *stemming* dapat dilihat pada Tabel 3.4.

**Tabel 3. 4 Contoh Penggunaan *Stemming***

|  |  |  |
| --- | --- | --- |
| **No.** | **Data *Review*** | **Hasil *Stemming*** |
| 1. | “kopi” “enak” “tempat” “nyaman”  “ngerjain” “tugas” “pelayanan” “ramah” | “kopi” “enak” “tempat”  “nyaman” “ngerjain”  “tugas” “pelayanan” “ramah” |
| 2. | “tempat” “nyaman” “banyak” “tanaman”  “suasana” “asik” “jogja” “selatan” “penting” “kopinya” “cucok” | “tempat” “nyaman”  “banyak” “tanaman”  “suasana” “asik” “jogja”  “selatan” “penting” “kopi” “cucok” |

**3.1.1.4 Pelabelan Data**

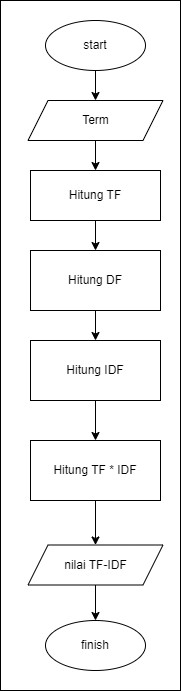
Data yang didapat dari proses *pre-processing* akan diberi label dengan menggunakkan metode *lexicon based* menjadi beberapa kelas sentimen yang terdiri dari kelas positif, negatif, serta netral berdasarkan ulasan. Label positif akan direpresentasikan dengan nilai 1, label negatif dengan nilai -1, dan label netral dengan label 0. Berikut merupakan contoh pelabelan data dapat dilihat pada Tabel 3.5 berikut :

# **Tabel 3. 5 Contoh Hasil Pelabelan Data**

|  |  |  |
| --- | --- | --- |
| **No.** | **Ulasan** | **Label** |
| 1. | Tempat nya asri sejuk dan tenang cocok kalau berkunjung pagi hari cobain best seller nya pasti ketagihan | 1 |
| 2. | Manual *brewing* bener bener gak enak sampe gak bisa diminum gak sesuai test noted yang tertera tapi harusnya minimal bisa diminum gak yang rasanya mengganggu banget | -1 |
| 3. | Tempat nongkrong ngopi pas di tiga terban jl solo pilih kopi dan snacknya lumayan di sini juga dapat food court dengan agam pilih makan tradisional jawa steak dll | 0 |

**3.1.1.5 Pembobotan *Term-Invers Document Frequency (*TF-IDF*)***

Pembobotan TF-IDF merupakan proses yang dilakukan apabila tahapan preprocessing telah dilakukan. Tujuan dari pembobotan kata dengan TF-IDF ini yaitu untuk memberikan bobot pada kata yang terdapat pada dokumen. Berikut merupakan *flowchart* proses perhitungan TF-IDF dapat dilihat pada Gambar 3.7.



**Gambar 3. 7** *Flowchart* TF-IDF

Berikut merupakan beberapa contoh data yang akan digunakan untuk perhitungan manual pada TF-IDF dapat dilihat pada Tabel 3.6.

# **Tabel 3. 6** Dokumen

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Dokumen** | **Sentimen** | **Label** |
| D1 | kopi enak makan enak tempat cocok nongkrong kota hening nugas | Positif | 1 |
| D2 | tempat asyik banget harga menu standar gak mahal | Netral | 0 |
| D3 | rasa masakan gak banget menu minuman tawarkan  pembayaran terlalu lama saat melakukan transaksi | Negatif | -1 |
| D4 | kopi enak makan enak layan oke strategis pandang tugu jogja | Positif | 1 |

1. Hitung *Term Frequency* (TF)

*Term frequency* bertujuan untuk menghitung kemunculan suatu *term* dalam suatu dokumen, semakin sering kemuculan term maka nilai kesesuaian yang dihasilkan bobot yang dihasilkan dari kata tersebut akan menjadi semakin besar. Pada tabel 3.5 Jumlah kemunculan pada kata tiap dokumen direpresentasikan sebagai D1 yaitu ulasan 1, D2 yaitu ulasan 2, D3 yaitu ulasan 3, dan D4 yaitu ulasan 4. Langkah awal dalam perhitungan TF-IDF yaitu menentukan TF, dengan persamaan sebagai berikut: 𝑡𝑓 = 1 + 𝑙𝑜𝑔10

Maka hasil perhitungan TF berdasarkan dokumen diatas dapat dilihat pada Tabel 3.7 berikut :

# **Tabel 3. 7** Contoh Hasil Perhitungan TF

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Term** |  | **TF** | |  |
| **D1** | **D2** | **D3** | **D4** |
| 1. | asyik | 0 | 1 | 0 | 0 |
| 2. | banget | 0 | 1 | 1 | 0 |
| 3. | cocok | 1 | 0 | 0 | 0 |
| 4. | enak | 2 | 0 | 0 | 2 |
| 5. | gak | 0 | 1 | 1 | 0 |
| 6. | harga | 0 | 1 | 0 | 0 |
| 7. | hening | 1 | 0 | 0 | 0 |
| 8. | jogja | 0 | 0 | 0 | 1 |
| 9. | kopi | 1 | 0 | 0 | 1 |
| 10. | kota | 1 | 0 | 0 | 0 |
| 11. | lama | 0 | 0 | 1 | 0 |
| 12. | layan | 0 | 0 | 0 | 1 |
| 13. | mahal | 0 | 1 | 0 | 0 |
| 14. | makan | 1 | 0 | 0 | 1 |
| 15. | masakan | 0 | 0 | 1 | 0 |
| 16. | melakukan | 0 | 0 | 1 | 0 |
| 17. | menu | 0 | 1 | 1 | 0 |
| 18. | minuman | 0 | 0 | 1 | 0 |
| 19. | nongkrong | 1 | 0 | 0 | 0 |
| 20. | nugas | 1 | 0 | 0 | 0 |
| 21. | oke | 0 | 0 | 0 | 1 |
| 22. | pandang | 0 | 0 | 0 | 1 |
| 23. | pembayaran | 0 | 0 | 1 | 0 |
| 24. | rasa | 0 | 0 | 1 | 0 |
| 25. | saat | 0 | 0 | 1 | 0 |
| 26. | standar | 0 | 1 | 0 | 0 |
| 27. | strategis | 0 | 0 | 0 | 1 |
| 28. | tawarkan | 0 | 0 | 1 | 0 |
| 29. | tempat | 1 | 1 | 0 | 0 |
| 30. | transaksi | 0 | 0 | 1 | 0 |
| 31. | tugu | 0 | 0 | 0 | 1 |

1. Hitung Nilai *Inverse Document Frequency* (IDF)

Apabila sudah dilakukan perhitungan *term frequency* (TF), maka tahap selanjutnya yaitu menghitung (IDF) atau menghitung banyaknya term pada seluruh dokumen.

Perhitungan IDF didapat dari (*log* (N/DF)), Dimana N adalah jumlah dokumen.

Perhitungan IDF dapat dilakukan melaluipersamaan sebagai berikut:

𝑁

𝑖𝑑𝑓𝑡 = 𝑙𝑜𝑔 𝑑𝑓𝑡.

Berikut merupakan hasil perhitungan IDF dapat dilihat pada Tabel 3.9.

**Tabel 3. 8** Contoh Hasil Perhitungan Bobot IDF

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Term** | **Term Frequency** | | | | **DF** | **Nd/df(t)** | **IDF** |
| **D1** | **D2** | **D3** | **D4** |
| 1. | asyik | 0 | 1 | 0 | 0 | 1 | 2 | 1,916 |
| 2. | banget | 0 | 1 | 1 | 0 | 2 | 1,2 | 1,511 |
| 3. | cocok | 1 | 0 | 0 | 0 | 1 | 2 | 1,916 |
| 4. | enak | 2 | 0 | 0 | 2 | 2 | 1,5 | 1,511 |
| 5. | gak | 0 | 1 | 1 | 0 | 2 | 3 | 1,511 |
| 6. | harga | 0 | 1 | 0 | 0 | 1 | 3 | 1,916 |
| 7. | hening | 1 | 0 | 0 | 0 | 1 | 3 | 1,916 |
| 8. | jogja | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |
| 9. | kopi | 1 | 0 | 0 | 1 | 2 | 3 | 1,511 |
| 10. | kota | 1 | 0 | 0 | 0 | 1 | 3 | 1,916 |
| 11. | lama | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 12. | layan | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |
| 13. | mahal | 0 | 1 | 0 | 0 | 1 | 3 | 1,916 |
| 14. | makan | 1 | 0 | 0 | 1 | 2 | 3 | 1,511 |
| 15. | masakan | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 16. | melakukan | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 17. | menu | 0 | 1 | 1 | 0 | 2 | 3 | 1,511 |
| 18. | minuman | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 19. | nongkrong | 1 | 0 | 0 | 0 | 1 | 3 | 1,916 |
| 20. | nugas | 1 | 0 | 0 | 0 | 1 | 3 | 1,916 |
| 21. | oke | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |
| 22. | pandang | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |
| 23. | pembayaran | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 24. | rasa | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 25. | saat | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 26. | standar | 0 | 1 | 0 | 0 | 1 | 3 | 1,916 |
| 27. | strategis | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |
| 28. | tawarkan | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 29. | tempat | 1 | 1 | 0 | 0 | 2 | 3 | 1,511 |
| 30. | transaksi | 0 | 0 | 1 | 0 | 1 | 3 | 1,916 |
| 31. | tugu | 0 | 0 | 0 | 1 | 1 | 3 | 1,916 |

1. Hitung nilai TF-IDF

Setelah didapatkan nilai TF, DF, dan IDF maka langkah selanjutnya yaitu menghitung nilai TF-IDF, melalui persamaan berikut :

𝑊𝑑𝑡 = 𝑡𝑓𝑑𝑡. 𝑖𝑑𝑓𝑡

Dengan W merupakan hasil dari perhitungan TF dikalikan dengan IDF (TF\*IDF)

Maka, hasil perhitungan TF-IDF dari dokumen diatas dapat dilihat pada Tabel 3.10.

**Tabel 3. 9** Contoh Hasil Perhitungan TF-IDF

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Term** | **Term Frequency** | | | | **DF** | **IDF** |  |  | **W** |  |
| **D1** | **D2** | **D3** | **D4** | **D1** | **D2** | **D3** | **D4** |
| 1. | asyik | 0 | 1 | 0 | 0 | 1 | 1,916 | 0 | 1,916 | 0 | 0 |
| 2. | banget | 0 | 1 | 1 | 0 | 2 | 1,511 | 0 | 1,511 | 1,511 | 0 |
| 3. | cocok | 1 | 0 | 0 | 0 | 1 | 1,916 | 1,916 | 0 | 0 | 0 |
| 4. | enak | 2 | 0 | 0 | 2 | 2 | 1,511 | 3,022 | 0 | 0 | 3,022 |
| 5. | gak | 0 | 1 | 1 | 0 | 2 | 1,511 | 0 | 1,511 | 1,511 | 0 |
| 6. | harga | 0 | 1 | 0 | 0 | 1 | 1,916 | 0 | 1,916 | 0 | 0 |

**Tabel 3. 10 Lanjutan Contoh Hasil Perhitungan TF-IDF**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Term** | **Term Frequency** | | | | **DF** | **IDF** | **W** | | |  |
| **D1** | **D2** | **D3** | **D4** | **D1** | **D2** | **D3** | **D4** |
| 7. | hening | 1 | 0 | 0 | 0 | 1 | 1,916 | 1,916 | 0 | 0 | 0 |
| 8. | jogja | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |
| 9. | kopi | 1 | 0 | 0 | 1 | 2 | 1,511 | 1,511 | 0 | 0 | 1,511 |
| 10. | kota | 1 | 0 | 0 | 0 | 1 | 1,916 | 1,916 | 0 | 0 | 0 |
| 11. | lama | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 12. | layan | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |
| 13. | mahal | 0 | 1 | 0 | 0 | 1 | 1,916 | 0 | 1,916 | 0 | 0 |
| 14. | makan | 1 | 0 | 0 | 1 | 2 | 1,511 | 1,511 | 0 | 0 | 1,511 |
| 15. | masakan | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 16. | melakukan | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 17. | menu | 0 | 1 | 1 | 0 | 2 | 1,511 | 0 | 1,511 | 1,511 | 0 |
| 18. | minuman | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 19. | nongkrong | 1 | 0 | 0 | 0 | 1 | 1,916 | 1,916 | 0 | 0 | 0 |
| 20. | nugas | 1 | 0 | 0 | 0 | 1 | 1,916 | 1,916 | 0 | 0 | 0 |
| 21. | oke | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |
| 22. | pandang | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |
| 23. | pembayaran | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 24. | rasa | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 25. | saat | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 26. | standar | 0 | 1 | 0 | 0 | 1 | 1,916 | 0 | 1,916 | 0 | 0 |
| 27. | strategis | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |
| 28. | tawarkan | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 29. | tempat | 1 | 1 | 0 | 0 | 2 | 1,511 | 1,511 | 1,511 | 0 | 0 |
| 30. | transaksi | 0 | 0 | 1 | 0 | 1 | 1,916 | 0 | 0 | 1,916 | 0 |
| 31. | tugu | 0 | 0 | 0 | 1 | 1 | 1,916 | 0 | 0 | 0 | 1,916 |

Berdasarkan hasil perhitungan TF-IDF selanjutnya dituliskan dalam bentuk notasi *vector* yang akan digunakan untuk melakukan *training* dengan menggunakkan metode SVM.

X1 = [0, 0, 4.790, 5.036, 0, 0, 4.790, 0, 2.518, 4.790, 0, 0, 0, 2.518, 0, 0, 0, 0, 4.790, 4.790, 0, 0, 0, 0, 0, 0, 0, 0, 2.518, 0, 0]

X2 = [1.916, 1.510, 0, 0, 1.510, 1.916, 0, 0, 0, 0, 0, 0, 1.916 , 0, 0, 1.510, 0, 0, 0, 0, 0, 0 , 0, 0, 1.916, 0, 0, 1.510, 0]

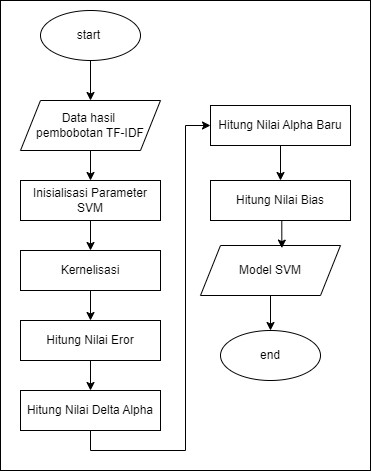
X3 = [0, 1.510, 0, 0, 1.510, 0, 0, 0, 0, 0, 1.916, 0, 0, 0, 1.916, 1.916, 1.510, 1.916, 0, 0,

0, 0, 1.916, 1.916, 1.916, 0, 0, 1.916, 0, 1.916, 0]

X4 = [0, 0, 0, 3.021, 0, 0, 0, 1.916, 1.510, 0, 0, 1.916, 0, 1.510, 0, 0, 0, 0, 0, 0, 1.916, 1.916, 0, 0, 0, 0, 1.916, 0, 0, 0, 1.916]

## 3.1.1.6 Klasifikasi Model dengan *Support Vector Machine (SVM)*

Tahap klasifikasi dengan model SVM ini bertujuan untuk memperoleh hasil akhir penelitian ini dengan mencari nilai klasifikasi sentimen berdasarkan data ulasan dari google maps. Alur tahapan SVM dapat dilihat pada Gambar 3.8 berikut :



**Gambar 3. 8 *Flowchart* Pelatihan SVM**

Metode yang digunakan dalam perhitungan ini yaitu *sequential* SVM dengan kernel linear. Berikut merupakan alur proses SVM:

1. Melakukan inisialisasi pada parameter (Vijayakumar, 1998), :
   1. 𝜆 = 0,5
   2. 𝛾 = 0,001
   3. ε = 0,0001
   4. 𝐶 = 1
   5. α = 0
2. Proses setelah inisialisasi parameter yaitu mencari nilai kernel pada matriks K beserta matriks *hessian* yang berukuran n x n yaitu sebanyak data training.

Berikut merupakan contoh perhitungannya.

𝑘 (𝑥𝑖. 𝑥𝑗) = (𝑥𝑖. 𝑥𝑗 + 1)2

𝑘 (𝑥𝑖. 𝑥𝑗) = ((0 x 1,916) + (0 x 1,511) + (1,916 x 0) + (3,022 x 0) + (0 x 1,511)

+ ( 0 x 1,916) + ( 1,916 x 0) + (0 x 0) + (0 x 0) + (0 x 1,916) + (1,511 x 0) + (0 x 0) + (1,916 x 0) + (1,916 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (0 x 0) + (1,511 x 0) + (0 x 0) + (0 x 0 + 1)2

𝑘 (𝑥𝑖. 𝑥𝑗) = 10,78

Perhitungan tersebut terus dilakukan pada dokumen lainnya sehingga memperoleh hasil matriks K seperti pada Tabel 3.11.

# **Tabel 3. 11** Nilai pada Matriks K

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dokumen | X1 | X2 | X3 | X4 |
| X1 | 1248,85 | 10,78 | 1,00 | 215,96 |
| X2 | 10,78 | 615,99 | 61,59 | 1,00 |
| X3 | 1,00 | 61,59 | 1672,59 | 1,00 |
| X4 | 215,96 | 1,00 | 1,00 | 1348,99 |

Untuk memperoleh nilai matriks hessian, perlu dilakukan perhitungan dengan proses perhitungan sebagai berikut

𝐷𝑖𝑗 = 𝑦𝑖𝑦𝑗(𝐾(𝑥𝑖. 𝑥𝑗) + λ2)

𝐷12 = (1)(0)((10,78) + 0,52))

𝐷12 = (0) (10,78 + 0,25)

𝐷12 = 0

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan hasil matriks hessian yang terdapat pada Tabel 3.12.

**Tabel 3. 12** Nilai pada Matriks Hessian

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dokumen** | **X1** | **X2** | **X3** | **X4** |
| X1 | 1249,096 | 0 | -1,25 | 216,210 |
| X2 | 0 | 0 | 0 | 0 |
| X3 | -1,25 | 0 | 1672,840 | -1,25 |
| X4 | 216,210 | 0 | -1,25 | 1349,238986 |

3. Apabila sudah didapatkan nilai pada matriks hessian, proses selanjutnya yaitu mencari nilai eror (𝐸𝑖). Berikut merupakan persamaan yang digunakan yaitu 𝐸𝑖 = ∑ 𝑛𝑗 = 1 𝛼𝑖 𝐷𝑖𝑗

𝐸𝑖= (0 x 1249,096) + (0 x 0) + (0 x (-1,25)) + (0 x 216,210) = 0

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan hasil nilai 𝐸𝑖 yang terdapat pada Tabel 3.13.

# **Tabel 3. 13** Hasil Perhitungan 𝐸𝑖

|  |  |
| --- | --- |
| **Dokumen** | 𝑬𝒊 |
| D1 | 0 |
| D2 | 0 |
| D3 | 0 |
| D4 | 0 |

4. Proses selanjutnya yaitu mencari nilai 𝛿𝛼𝑖 melalui persamaan

𝛿𝛼𝑖 = min{max[γ(1−𝐸𝑖),𝛼𝑖], 𝐶 − 𝛼𝑖. Berikut merupakan contoh perhitungannya: 𝛿𝛼𝑖 = min{max[0,001(1-0),0]1-0} = 0,001.

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan nilai 𝛿𝛼𝑖 seperti pada Tabel 3.14.

# **Tabel 3. 14** Hasil Perhitungan 𝛿𝛼𝑖

|  |  |  |
| --- | --- | --- |
| **Dokumen** |  | 𝛿𝜶𝒊 |
| D1 |  | 0,001 |
| D2 |  | 0,001 |
| D3 |  | 0,001 |
| D4 |  | 0,001 |

5. Tahap selanjutnya yaitu mencari nilai baru dengan menggunakkan persamaan

𝛼𝑖 = 𝛼𝑖 + δ𝛼𝑖 = 0 + 0,001

= 0,001

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan nilai 𝛼𝑖 seperti pada Tabel 3.15.

# **Tabel 3. 15** Hasil Perhitungan 𝛼𝑖 baru

|  |  |
| --- | --- |
| **Dokumen** | 𝜶𝒊 **baru** |
| D1 | 0,001 |
| D2 | 0,001 |
| D3 | 0,001 |
| D4 | 0,001 |

6. Pada tahap ini yaitu mengulang perhitungan pada tahap ke-3 sampai tahap ke5 hingga mencapai batas maksimal iterasi. Berikut merupakan contoh perhitungan iterasi ke-2.

𝐸𝑖 = (0,001 x 1249,096)+(0,001 x 0)+(0,001 x (-1,25))+(0,001 x 216,210) = 1,4640

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan hasil nilai 𝐸𝑖 yang terdapat pada Tabel 3.16.

# **Tabel 3. 16** Hasil Perhitungan 𝐸𝑖 iterani ke-2

|  |  |
| --- | --- |
| **Dokumen** | 𝑬𝒊 |
| D1 | 1,4641 |
| D2 | 0 |
| D3 | 1,6703 |
| D4 | 1,5642 |

7. Tahapan berikutnya yaitu menghitung 𝛿𝛼𝑖 pada iterasi ke-2. Berikut merupakan contoh perhitungan 𝛿𝛼𝑖 pada iterasi ke-2

𝛿𝛼𝑖 = min{max[𝛾(1-𝐸𝑖),𝛼𝑖],𝐶 − 𝛼𝑖*}*

𝛿𝛼𝑖 = min{max[0,001(1-1,4641),0,001],1-0,001}

= 0,001

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan hasil nilai 𝛿𝛼𝑖 yang terdapat pada Tabel 3.17.

# **Tabel 3. 17** Hasil Perhitungan 𝛿𝛼𝑖 Iterasi ke-2

|  |  |
| --- | --- |
| **Dokumen** | 𝛿𝜶𝒊 |
| D1 | -0,001000464 |
| D2 | -0,000999 |
| D3 | -0,00100067 |
| D4 | -0,001000564 |

𝛼𝑖 = 𝛼𝑖 + 𝛿𝛼𝑖

= 0,001 + -0,001000464

= -0,0000004641

Ulangi perhitungan tersebut pada dokumen lainnya sehingga mendapatkan hasil nilai 𝛿αi yang terdapat pada Tabel 3.19

# **Tabel 3. 18** Hasil Perhitungan 𝛼𝑖 baru

|  |  |
| --- | --- |
| **Dokumen** | 𝜶𝒊 **baru** |
| D1 | -0,001000464 |
| D2 | -0,000999 |
| D3 | -0,00100067 |
| D4 | -0,001000564 |

8. Tahapan selanjutnya adalah perhitungan untuk mencari nilai bias dengan persamaan 𝑏 = −((𝑤. 𝑥−1) + (𝑤. 𝑥+1)). 𝑥−1 masuk pada golongan kelas negatif yang dimana memiliki nilai alpha paling besar, sedangkan 𝑥+1 masuk pada golongan kelas positif yang Dimana memiliki nilai alpha paling besar.

𝑤. 𝑥−1 = (1 𝑥 (−0,001000464)𝑥 1 + (0 𝑥 (−0,000999) 𝑥 61,59)

+ (−1 𝑥 (−0,00100067)𝑥 1671,59) + (1𝑥 (−0,001000564)𝑥 1)

= 1,672534942

𝑤. 𝑥+1 = (1 𝑥 (−0,001000464)𝑥 1248,85 + (0 𝑥 (−0,000999) 𝑥 10,78)

+ (−1 𝑥 (−0,00100067)𝑥 1) + (1𝑥 (−0,001000564)𝑥 215,96)

= -1,464505898

Dengan diperolehnya kedua nilai tersebut, maka untuk mendapatkan nilai bias dapat dilakukan dengan perhitungan sebagai berikut:

𝑏 = −  (1,672534942 − (−1,464505898))

= -1,56852042

## 3.1.1.7 Pengujian

Pengujian pada sistem dilakukan dengan menggunakkan *k-fold cross validation* serta *confusion matrix* yang bertujuan untuk mengetahui performa model yang diterapkan. Pesngujian dengan *confusion matrix* ini digunakan untuk mengukur Tingkat akurasi, presisi, serta *recall* dari suatu model. Rancangan pengujian *confusion matrix* dapat dilihat pada Tabel 3.19.

**Tabel 3. 19 Rancangan *Confusion Matrix***

|  |  |  |  |
| --- | --- | --- | --- |
| ***Actual*** |  | ***Classified*** |  |
| ***Negative*** | ***Positive*** | ***Neutral*** |
| *Negative* | *TN (True Negative)* | *FP (False Positive)* | *FL2 (False Neutral 2)* |
| *Positive* | *FN (False Negative)* | *TP (True Positive)* | *FP2(False Positive 2)* |
| *Neutral* | *FN2(False Negative 2)* | *FP2(False Positive 2)* | *TL (True Neutral)* |

Selanjutnya pengujian *k-fold cross validation* bertujuan untuk melakukan validasi dari suatu model dengan melakukan pengujian sebanyak K, Dimana pada penelitian ini menggunakkan K sebanyak 5. Berikut merupakan tabel rancangan *k-fold cross validation* dapat dilihat pada Tabel 3.20.

**Tabel 3. 20 Rancangan *K-Fold Cross Validation***

|  |  |  |  |
| --- | --- | --- | --- |
| **Fold** |  | **SVM** |  |
| **Akurasi** | **Presisi** | **Recall** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |

## 3.1.2. Metode Pengembangan Sistem

**3.1.2.1 *Requirement analysis***

Analisis kebutuhan sistem merupakan proses analisis terhadap kebutuhan-kebutuhan yang digunakan pada penelitian ini yang meliputi kebutuhan fungsional dan kebutuhan non-fungsional.

1. **Kebutuhan Fungsional**

Adapun kebutuhan-kebutuhan yang termasuk kebutuhan fungsional dari sistem yang akan dibangun adalah sebagai berikut:

* 1. Sistem mampu melakukan proses *preprocessing* terhadap data ulasan yang dikumpulkan
  2. Sistem mampu menampilkan hasil klasifikasi sentimen menggunakkan metode *Support Vector Machine*
  3. Sistem mampu melakukan penginputan kalimat untuk pengujian sistem
  4. Sistem mampu menunjukkan nilai akurasi dari hasil akurasi perhitungan terhadap data yang diproses.
  5. Sistem mampu menampilkan list rekomendasi *coffee shop*
  6. Sistem mampu melakukan proses *login* dan *logout* untuk admin

1. **Kebutuhan Non-Fungsional**

Dalam pengimplementasian sistem yang akan dibangun agar sistem sesuai dengan apa yang telah dirancang maka dibutuhkan beberapa perangkat keras atau *hardware*, perangkat lunak atau *software*. Berikut merupakan kebutuhan non-fungsional untuk merancang sistem ini:

* 1. Kebutuhan *Hardware*

Berikut merupakan spesifikasi *hardware* yang digunakan dalam pembuatan sistem pada penelitian ini terdapat pada Tabel 3.21 berikut :

**Tabel 3. 21 Kebutuhan *Hardware***

|  |  |  |
| --- | --- | --- |
| No. | *Hardware* | Keterangan |
| 1. | *Processor* | 11th Gen Intel(R) Core(TM) i5-1135G7 @  2.40GHz 2.42 GHz |
| 2. | RAM | 16,0 GB |
| 3. | Koneksi internet | WIFI |
| 4. | Perangkat *input* dan *output* | *Keyboard*, dan *Mouse* |
| 5. | Koneksi Internet | *WIFI* |

* 1. Kebutuhan *Software*

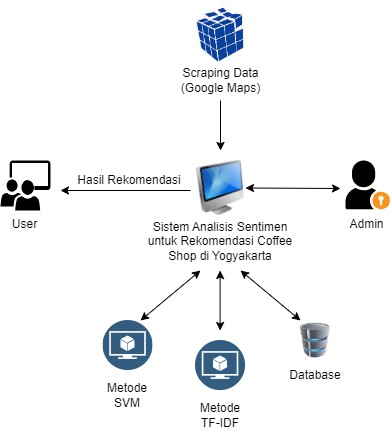
Berikut merupakan spesifikasi *software* yang digunakan dalam pembuatan sistem pada penelitian ini terdapat pada Tabel 3.22 berikut :

**Tabel 3. 22 Kebutuhan *Software***

|  |  |  |
| --- | --- | --- |
| No. | *Software* | Keterangan |
| 1. | *Operating system* | Windows 11, 64 bit |
| 2. | *Google Colab* | Code Editor Online |
| 3. | *Web Browser* | *Google Chrome, Microsoft Edge* |
| 4. | *Draw.io* | *Software* untuk membuat *design* diagram |
| 5. | Bahasa Pemrograman | *Python*, PHP |
| 6. | IDE | *Visual Studio Code* |
| 7. | Figma | *Design Sistem* |

**3.1.2.2 *System and Software Design* A. Perancangan Arsitektur**

Pada bagian perancangan sistem yang akan dibangun pada penlitian ini dibuat untuk menganalisis sentimen ulasan serta merekomendasikan *coffee shop* di Yogyakarta. Rancangan arsitektur sistem pada penelitian ini dapat dilihat pada Gambar 3.9 berikut:

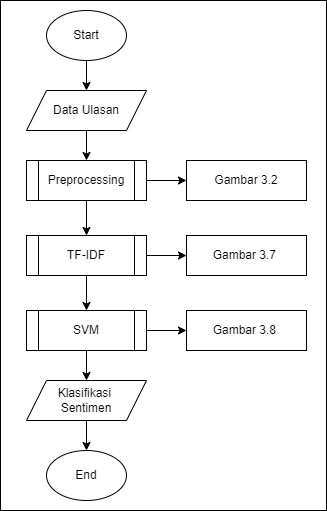


**Gambar 3. 9 Arsitektur Sistem**

Arsitektur yang akan dibangun terdiri dari *admin* dan *user*. Dimana *admin* akan meng*input*kan data pada sistem, data tersebut didapatkan dari proses *scraping* dari *google maps*. Kemudian data tersebut akan masuk pada tahap *preprocessing* yang bertujuan agar data yang telah disimpan dapat dipahami oleh sistem. Tujuan utama dari tahap preprocessing yaitu untuk menghasilkan data yang siap diproses oleh sistem serta membersihkan dan juga penyeragaman kata yang nantinya akan di ekstrasi pada tahap berikutnya. Tahap berikutnya yaitu tahap ekstrasi fitur dimana tahapan ini bertujuan untuk menghitung kemunculan term pada sebuah dokumen. Selanjutnya terks akan diklasifikasikan dengan menggunakkan Metode *Support Vector Machine* yang akan menghasilkan data dengan kelas positif, negatif, dan juga netral. Data yang telah diklasifikasikan kemudian akan digunakan untuk pembobotan, dan nantinya akan ditampilkan rekomendasi *coffee shop* berdasarkan analisis sentimen.

1. **Perancangan Proses**

Rancangan proses digambarkan melalui *flowchart* yang merupakan langkah-langkah atau tahapan tindakan terhadap sistem yang akan dibuat. Perancangan proses direpresentasikan melalui *flowchart*. Berikut merupakan *flowchart* dari perancangan proses dapat dilihat pada Gambar 3.10 berikut :



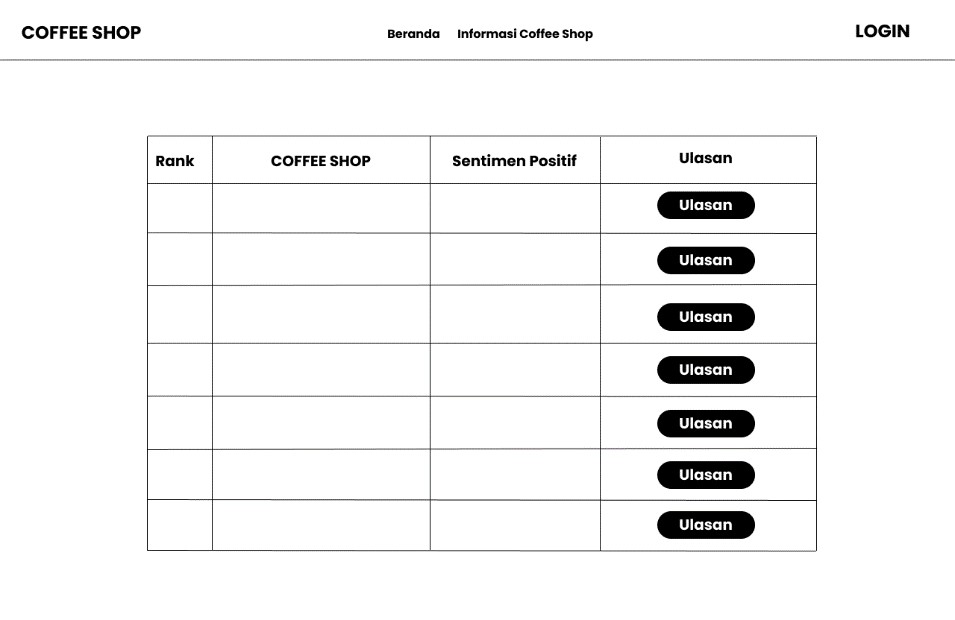
**Gambar 3. 10 *Flowchart* Perancangan Proses**

1. **Perancangan *User Interface* (Antarmuka)**

Rancangan *user interface* merupakan suatu proses *design* yang berfokus pada model komunikasi antara pengguna dengan sistem. Rancangan *user interface* mencakup beberapa aspek seperti, tata letak, dan navigasi yang bertujuan untuk memastikan pengguna dapat menggunakkan sistem dengan mudah dan efektif. Pada sistem ini memiliki dua aktor yaitu, *admin* serta *user*. Berikut merupakan gambaran rancangan *user interface* yang terdiri dari beberapa bagian adalah sebagai berikut:

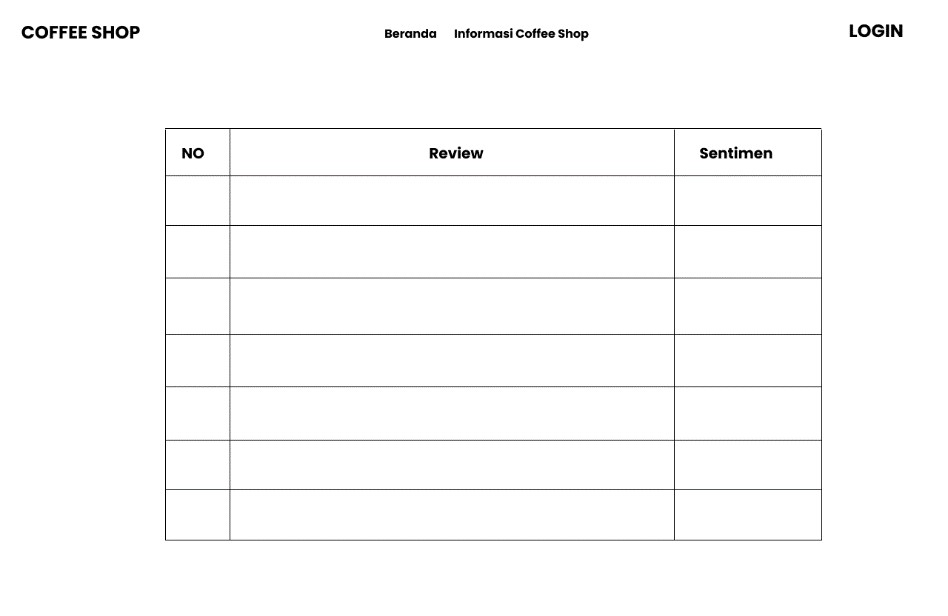
1. Rancangan Halaman List Rekomendasi *Coffee Shop*

Halaman ini merupakan halaman yang akan tampil saat user pertama kali mengakses sistem. Dimana pada halaman ini terdapat list rekomendasi beberapa *coffee shop* di Yogyakarta. Rancangan halaman utama pengunjung dapat dilihat pada Gambar 3.11 berikut :



## Gambar 3. 11 Rancangan Halaman List Rekomendasi Coffee Shop

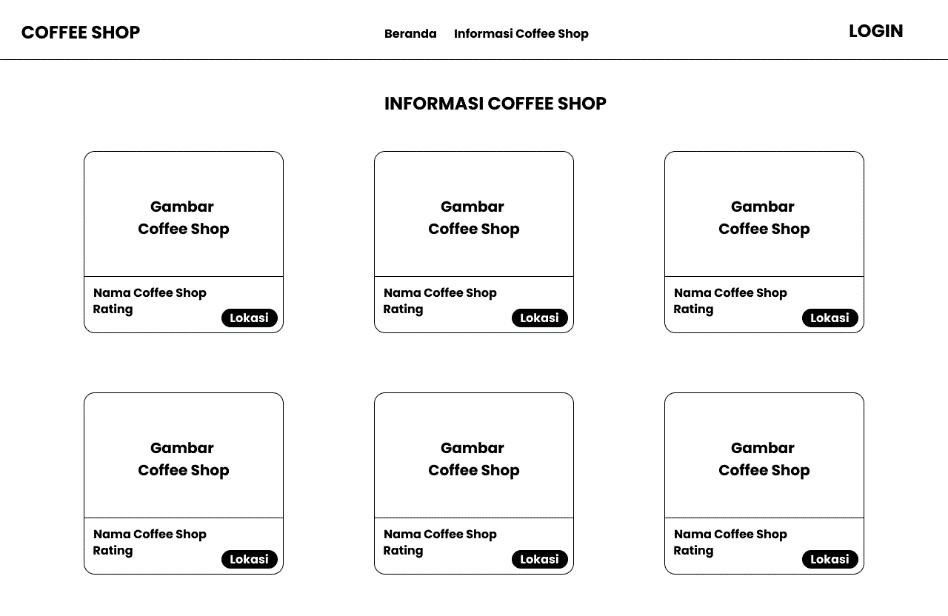
Pada fitur ulasan user dapat melihat detai review beserta sentimen dari *coffee shop* tersebut. Rancangan halaman detail ulasan dapat dilihat pada Gambar 3.12 berikut :



## Gambar 3. 12 Rancangan halaman Detail Ulasan

2. Rancangan Halaman Informasi *Coffee Shop*

Halaman ini merupakan halaman yang menampilkan mengenai beberapa coffee shop yang ada di Yogyakarta. Rancangan halaman informasi coffee shop dapat dilihat pada Gambar 3.13 berikut :



## Gambar 3. 13 Rancangan Halaman Informasi Coffee Shop

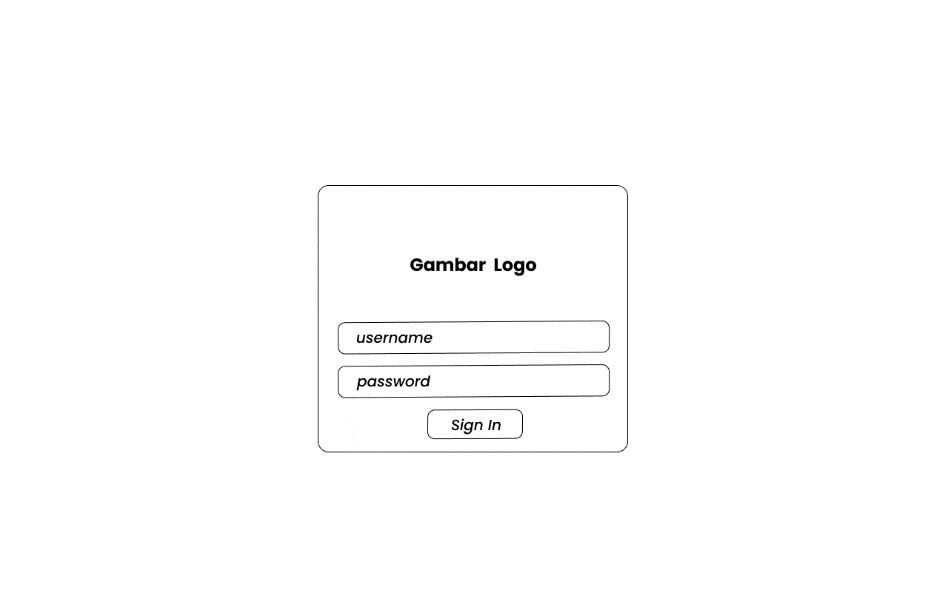
1. Rancangan Halaman *Login Admin*

Halaman *login admin* merupakan halaman yang akan tampil apabila masuk pada sistem sebagai *admin*. Pada halaman ini *admin* harus menginputkan *username* serta *password* agar bisa masuk ke dalam sistem. Rancangan halaman *login admin* dapat dilihat pada Gambar 3.14 berikut :

**Gambar 3.**

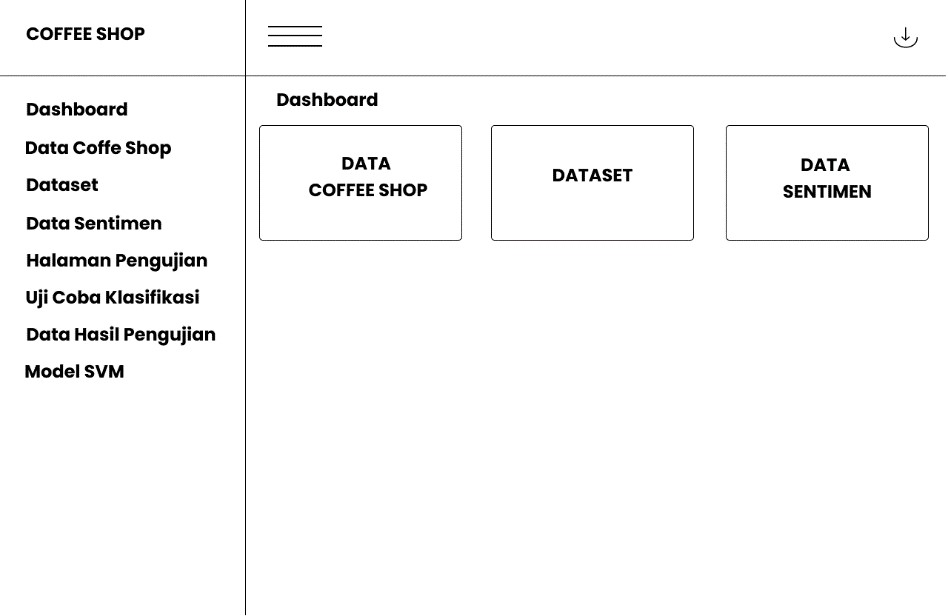
**14**

**Rancangan Halaman Login**



1. Rancangan Halaman *Dashboard Admin*

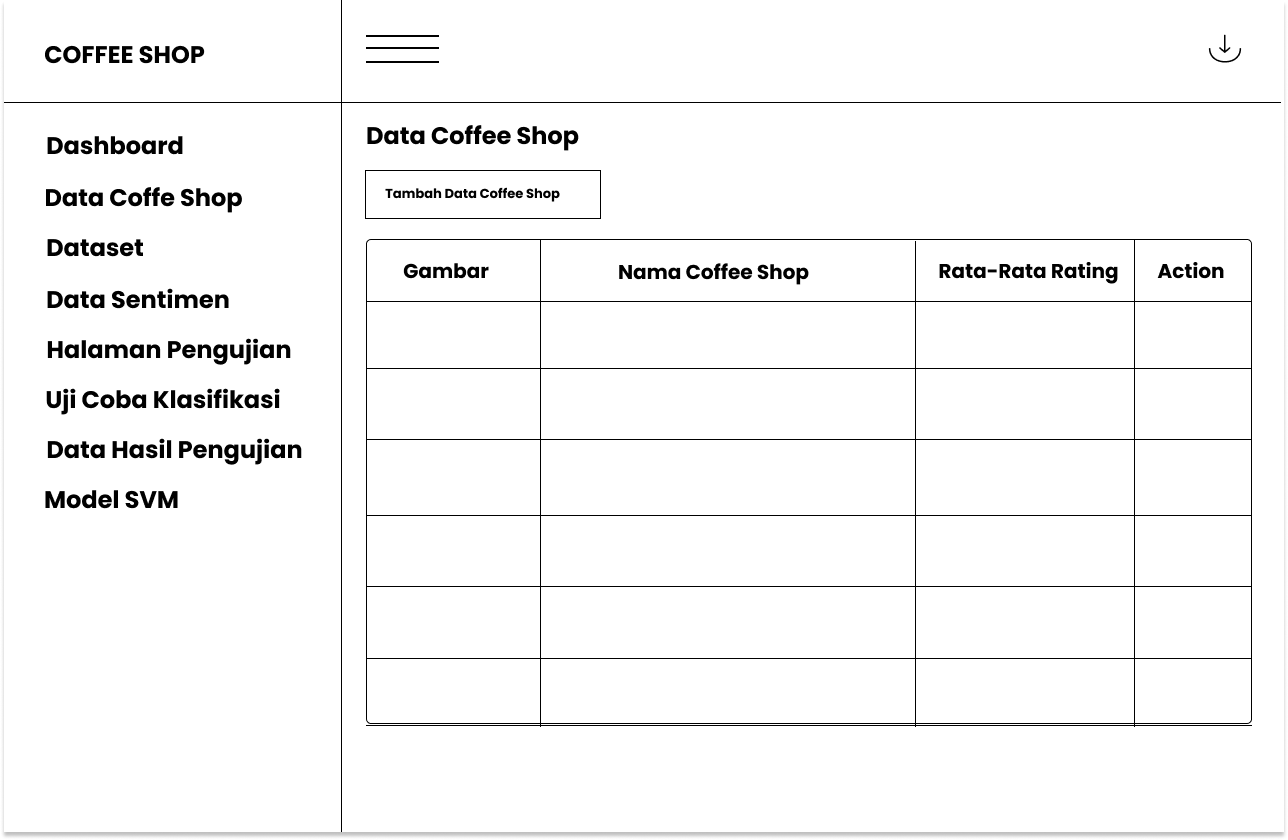
Halaman *dashboard admin* ini merupakan halaman yang pertama kali tampil apabila berhasil *login* sebagai *admin*. Dimana pada halaman dashboard ini menampilkan jumlah data coffee shop, dataset, serta data sentimen. Pada halaman dashboard ini memiliki beberapa menu diantaranya menu data coffee shop, dataset, data sentimen, halaman pengujian serta halaman model SVM, yang dimana pada setiap menu tersebut memiliki fungsi masing-masing. Rancangan halaman *dashboard* dapat dilihat pada Gambar 3.15 berikut :



## Gambar 3. 15 Rancangan Halaman Dashboard

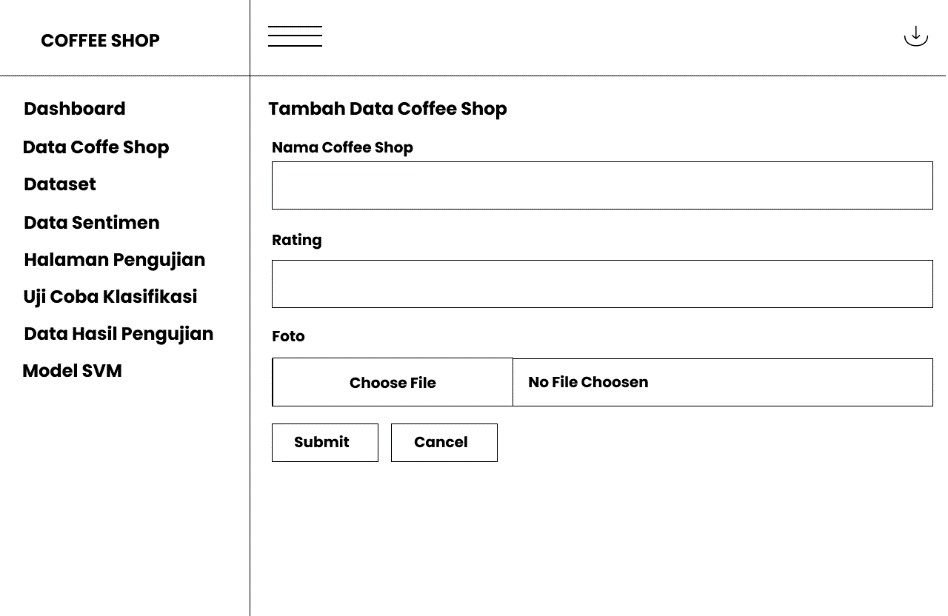
5. Rancangan Halaman Data *Coffee Shop*

Pada halaman ini menampilkan data *coffee shop* yang digunakan pada sistem ini dalam bentuk tabel. Pada halaman data coffee shop ini menampilkan data berupa gambar, nama, rating dari coffee shop. Halaman ini memiliki fitur edit data, tambah data, serta hapus data untuk mengolah data. Rancangan halaman data *coffee shop* dapat dilihat pada Gambar 3.16 berikut :



## Gambar 3. 16 Rancangan Halaman Data *Coffee Shop*

Pada halaman ini juga terdapat opsi tambah data yang dapat diakses oleh admin. Pada halaman tambah data ini berisi form nama *coffee shop*, *rating*, serta foto *coffee shop*, rancangan halaman tambah data *coffee shop* dapat dilihat pada Gambar 3.17 berikut :



## Gambar 3. 17 Rancangan Halaman Fitur Tambah Data *Coffee Shop*

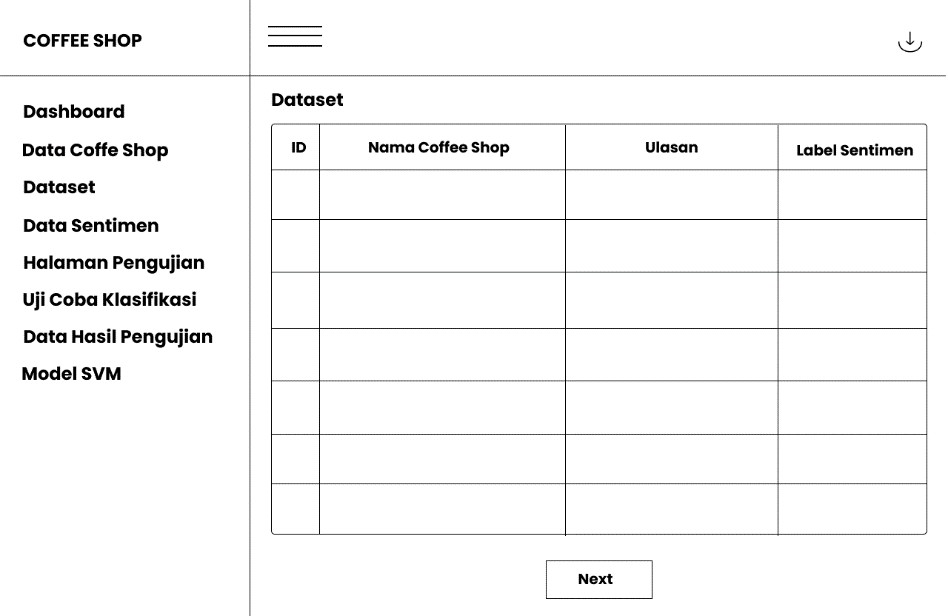
1. Rancangan Halaman Data Sentimen

Rancangan pada halaman ini merupakan halaman yang menampilkan ulasan yang akan digunakan dataset dalam model SVM. Rancangan halaman dataset dapat dilihat pada Gambar 3.18 berikut :

**Gambar 3.**

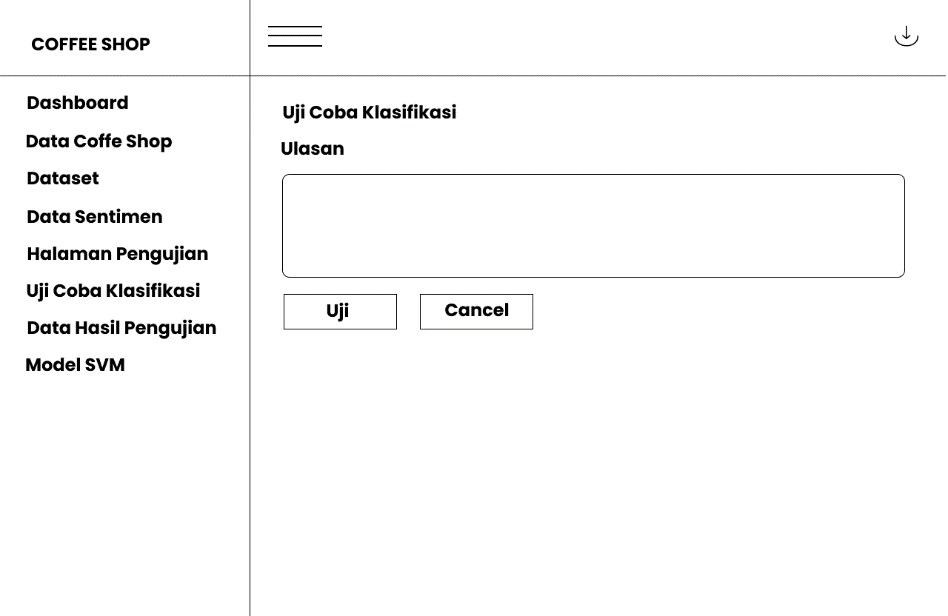
**18**

**Rancangan Halaman Dataset**



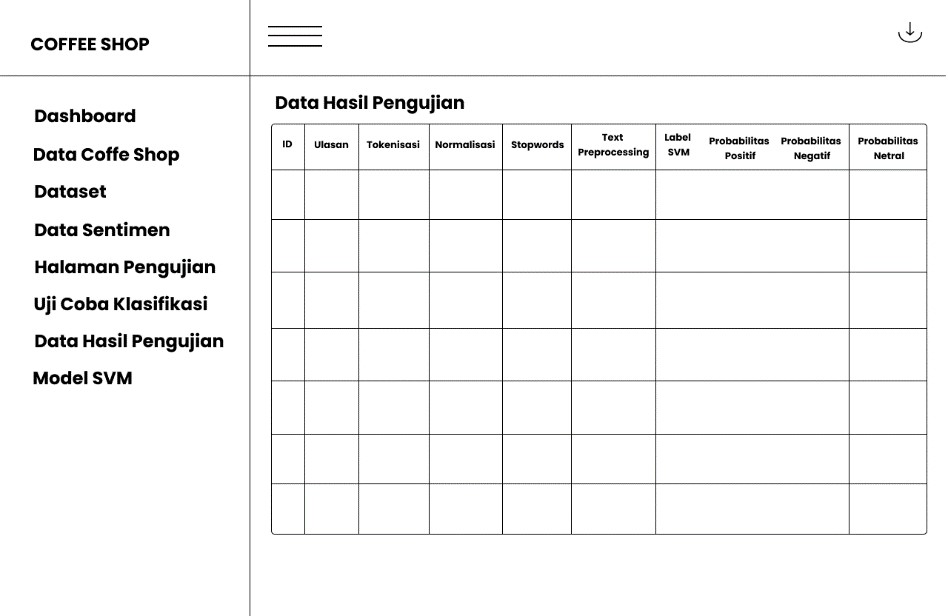
1. Rancangan Halaman Pengujian

Rancangan halaman pengujian ini merupakan halaman yang akan memproses data, serta menampilkan data yang telah diuji. Pada halaman pengujian ini admin dapat menginputkan kalimat yang nantinya akan diuji oleh sistem serta mendapatkan hasil prediksi sentiment kalimat yang telah diinputkan. Rancangan halaman uji klasifikasi dapat dilihat pada Gambar 3.19 berikut :



## Gambar 3. 19 Rancangan Halaman Uji Klasifikasi

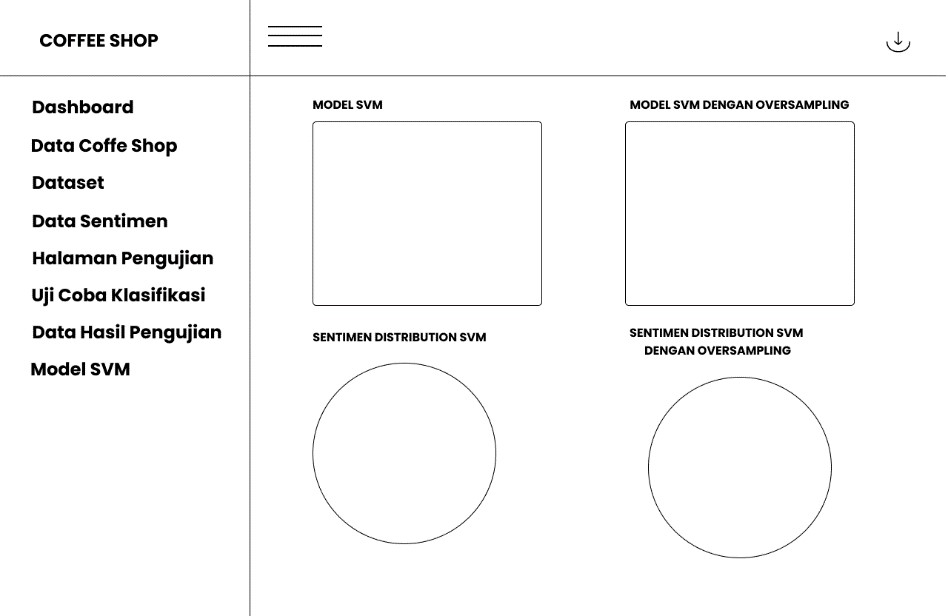
Hasil pengujian sebelumnya dapat dilihat pada fitur Data Hasil Pengujian, Dimana pada halaman ini berisikan proses ulasan berdasarkan kalimat yang telah diiputkan sebelumnya. Rancangan halaman data hasil pengujian dapat dilihat pada Gambar 3.20 berikut :



## Gambar 3. 20 Rancangan Halaman Data Hasil Pengujian

8. Rancangan Halaman Model SVM

Rancangan halaman ini nantinya akan menampilkan mengenai hasil akurasi, presisi, dan recall dari model SVM ke dalam bentuk pie chart serta diagram. Pada halaman model SVM ini juga ditampilkan hasil pemodelan SVM dengan oversampling. Rancangan halaman model SVM dapat dilihat pada Gambar 3.21 berikut :



**Gambar 3. 21 Rancangan Halaman Model SVM**

## 3.1.2.3 *Implementation*

Pada tahap *implementation*, merupakan tahapan dimana perancangan sistem dan *design* yang telah dibuat akan direalisasikan menjadi kode program yang akan menghasilkan suatu sistem. Implementasi sistem analisis sentimen dibuat dan dirancang sesuai dengam metode yang digunakan berdasarkan tahapan analisis sentimen untuk rekomendasi *coffee shop* di Yogyakarta.

## 3.1.2.4 *System Testing*

Pada tahap ini dibuat suatu rancangan pengujian. Rancangan pengujian bertujuan untuk mempersiapkan konsep uji yang akan dilakukan untuk mengevaluasi tingkat keberhasilan proses klasifikasi yang telah dijalankan. Metode yang digunakan pada penelitian ini adalah *black box testing*.

Pengujian dengan metode *black box testing* bertujuan untuk menilai kinerja sistem yang telah dikembangkan. Serta pada pengujian *black box* ini berfokus pada fungsionalitas sistem, sehingga memastikan bahwa setiap fungsi berjalan sesuai dengan yang diharapkan. Detail mengenai rancangan pengujian dengan metode *black box testing* dapat dilihat pada Tabel 3.22 berikut :

# **Tabel 3. 23 Rancangan Pengujian *Black Box***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aktor** | **Halaman** | **Detail Pengujian** |  | **Nilai** |
| **Berhasil** | **Tidak Berhasil** |
| Admin | *Login* | Melakukan proses login |  |  |
| *Dashboard* | Menampilkan jumlah tiap data |  |  |
| Data *Coffee Shop* | Menampilkan data *coffee shop* |  |  |
|  | Menambahkan data *coffee shop* |  |  |
|  | Mengedit data *coffee shop* |  |  |
|  | Menghapus data *coffee shop* |  |  |
| Dataset | Menampilkan dataset |  |  |
| Data Sentimen | Menampilkan data sentimen |  |  |
| Halaman Model SVM | Menampilkan grafik model SVM |  |  |

**Tabel 3. 24 Lanjutan Rancangan Pengujian *Black Box***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aktor** | **Halaman** | **Detail Pengujian** |  | **Nilai** |
| **Berhasil** | **Tidak Berhasil** |
| Admin | *Logout* | Fungsi *Logout* |  |  |
| User | Rekomendasi *Coffee Shop* | Menampilkan list rekomendasi *coffee shop* |  |  |
| Informasi *Coffee Shop* | Menampilkan informasi  alamat *coffee shop* |  |  |