

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kabupaten Rokan Hulu merupakan salah satu Kabupaten yang ada di Provinsi Riau dengan julukan Negeri Seribu Suluk yang sangat kental dengan keislaman dan memiliki keanekaragaman seni, kesenian-keseniannya dapat mempertahankan budaya dan adat istiadat yang beragam corak pada setiap Kecamatan yang ada di Rokan Hulu dan mengangkat norma daerah Rokan Hulu khususnya Pasir Pengaraian [1]

Rumah sakit adalah salah satu bentuk fasilitas pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan secara paripurna yang menyediakan berbagai macam pelayanan, antara lain pelayanan rawat inap, rawat jalan, dan gawat darurat [2]. Keberadaan rumah sakit sangat penting dalam sistem pelayanan kesehatan karena berperan sebagai rujukan utama dalam penanganan kasus-kasus medis yang tidak dapat ditangani di fasilitas kesehatan tingkat pertama. Selain itu, rumah sakit juga dilengkapi dengan tenaga medis profesional, peralatan medis canggih, serta sistem manajemen pelayanan yang mendukung proses diagnosis, pengobatan, dan rehabilitasi pasien secara menyeluruh.

Rumah sakit sebagai salah satu penyelenggara pelayanan kesehatan paripurna memiliki tugas yang lebih berat karena tidak hanya melayani penyakit umum saja, sehingga harus mampu memberikan pelayanan yang lebih berkualitas untuk memenuhi tuntutan masyarakat secara luas [3].

Rumah Sakit Umum Daerah (RSUD) Rokan Hulu, sebagai salah satu institusi layanan publik yang vital, menjadi salah satu objek yang sering mendapat tanggapan dari masyarakat, khususnya para pasien. Ulasan-ulasan ini mencerminkan persepsi dan kepuasan pasien terhadap kualitas pelayanan rumah sakit, mulai dari pelayanan medis, administrasi, hingga fasilitas yang disediakan.

Saat ini masyarakat dengan lebih mudah mendapatkan informasi, termasuk informasi tentang layanan kesehatan melalui *platform* digital. Salah satu platform yang populer untuk mencari informasi ini yaitu *Google Maps*. Calon pasien dapat membaca ulasan untuk mengetahui kelebihan dan kekurangan layanan rumah sakit sebelum memutuskan untuk berobat.

Namun, tidak semua ulasan di *Google Maps* bersifat objektif. Ada kalanya ulasan tersebut mengandung sentimen positif atau negatif yang dapat mempengaruhi persepsi calon pasien terhadap layanan rumah sakit. Ulasan Positif adalah pendapat atau komentar yang menunjukkan kepuasan, apresiasi, atau pengalaman baik dari seseorang terhadap suatu produk, layanan, atau pengalaman tertentu. Sedangkan ulasan negatif adalah pendapat atau komentar yang menyatakan ketidakpuasan, kritik, atau pengalaman buruk terhadap suatu produk, layanan, atau pengalaman. Oleh karena itu, diperlukan analisis sentimen dan emosi untuk mengidentifikasi dan memahami sentimen yang terkandung dalam ulasan-ulasan tersebut.

Analisis sentimen adalah analisis pengolahan bahasa alami, komputasi *linguistic* [4]. Salah satu metode yang bisa digunakan adalah metode *Logistic Regression*. *Logistic regression* adalah teknik analisis data dalam statistika yang

dirancang untuk mengetahui hubungan antar variabel, dimana variabel respon bersifat kategorikal, baik nominal maupun ordinal, dan variabel penjelas bersifat kategoris atau kontinu [5].

Logistic Regression Merupakan metode *supervised machine learning* yang dapat digunakan untuk menganalisis data serta mendeskripsikan antara satu atau lebih variabel prediksi dengan satu variabel respon. Variabel respon dari *Logistic Regression* hanya bernilai antara 0 dan 1 sehingga akan menghasilkan *class sentiment* positif dan negatif dengan batasan diantara keduanya adalah nilai 0,5 [6].

Beberapa penelitian telah melakukan penelitian tentang analisis sentimen dengan metode *Logistic Regression*. Sebelum penelitian dilakukan, penulis telah melakukan survei terhadap beberapa penelitian sebelumnya. Seperti yang dilakukan penelitian oleh Kelvin, Jepri Banjarnahor, dkk tahun 2022 dengan judul analisis perbandingan sentimen corona virus disease 2019 (COVID19) pada twitter menggunakan metode *Logistic Regression* dan *support vector machine (SVM)*. Dari hasil penelitian dapat disimpulkan Dari analisis yang telah dilakukan dengan tahapan yang ada dapat disimpulkan dari penelitian ini adalah nilai akurasi metode *Support Vector Machine (SVM)* lebih unggul dibandingkan metode *Logistic Regression* dalam melakukan klasifikasi terhadap sentimen Twitter tentang Tweet kasus *CoronaVirus Disease-2019 (COVID-19)* dengan nilai akurasi sebesar 91,15% [7].

Penelitian lainnya yang telah diteliti oleh Jeffry, dkk pada tahun 2023 dengan judul analisis perilaku pelanggan menggunakan metode *Ensemble Logistic*

Regression. Dari hasil penelitian dapat disimpulkan perbandingan akurasi dari tiga metode yang berbeda dalam melakukan klasifikasi dengan menggunakan model *Logistic Regression* tanpa menggunakan teknik *ensemble* memiliki tingkat akurasi sekitar 0.59, menunjukkan performa model dasar tanpa peningkatan tambahan. Namun, dengan menerapkan teknik *ensemble*, baik dengan menggunakan metode *boosting* maupun *bagging*, akurasi model meningkat secara signifikan. Model yang ditingkatkan dengan teknik *boosting* mencapai akurasi sekitar 0.76, sementara model dengan teknik *bagging* mencapai akurasi sekitar 0.75. Teknik *boosting* dan *bagging* telah terbukti efektif dalam meningkatkan performa model *Logistic Regression*, dan pilihan antara keduanya dapat disesuaikan dengan kebutuhan khusus dan karakteristik dataset yang digunakan [8].

Dari pemaparan latar belakang diatas, penulis tertarik untuk melakukan penelitian dengan judul " Analisis Sentimen Menggunakan *Logistic Regresion* Untuk Klasifikasi Ulasan Pasien Terhadap Pelayan Rumah Sakit Umum Daerah " sebagai upaya untuk mengembangkan sistem klasifikasi sentimen berbasis kecerdasan buatan yang dapat memberikan gambaran umum terhadap kepuasan masyarakat, serta mendukung pengambilan keputusan yang lebih cepat dan tepat oleh manajemen rumah sakit.

1.2 Rumusan Masalah

Dari latar belakang yang telah dikemukakan, adapun rumusan masalah dalam penelitian ini adalah bagaimana merancang dan membangun sebuah sistem analisis sentimen dengan menggunakan *Logistic Regression* pada ulasan pasien terhadap Pelayanan Rumah Sakit Umum Daerah (RSUD)?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah untuk merancang dan membangun sebuah sistem analisis sentimen dengan menggunakan metode *Logistic Regression* pada ulasan pasien terhadap Pelayanan Rumah Sakit Umum Daerah (RSUD).

1.4 Batasan Masalah

Agar pengembangan sistem sentimen analisis metode *Logistic Regression* ini dapat berjalan secara terarah dan terfokus, maka dibuat beberapa batasan masalah sebagai berikut :

1. Data yang digunakan dalam penelitian ini terbatas pada ulasan berbentuk teks dari pasien terhadap pelayanan di Rumah Sakit Umum Daerah (RSUD) yang diperoleh melalui *platform digital* seperti *Google Maps*.
2. Penelitian menggunakan algoritma *Logistic Regression*.
3. Sentimen yang diklasifikasikan dibatasi pada tiga kategori, yaitu positif, negatif, dan netral.
4. Aplikasi yang dibangun menggunakan bahasa pemrograman Python.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Manfaat Penelitian Bagi Pihak RSUD

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analisis sentimen, yang dapat membantu pihak manajemen Rumah Sakit Umum Daerah (RSUD) dalam mengevaluasi dan meningkatkan kualitas pelayanan berdasarkan persepsi dan

pengalaman pasien yang diungkapkan melalui ulasan teks secara otomatis dan efisien.

2. Manfaat Penelitian Bagi Mahasiswa

Penelitian ini dapat dijadikan bahan pembelajaran dan referensi bagi mahasiswa dan akademisi dalam memahami implementasi analisis sentimen dengan pendekatan klasifikasi statistik, serta penerapannya dalam dunia nyata, khususnya di sektor kesehatan. Penelitian ini juga dapat mendorong mahasiswa untuk mengembangkan keterampilan teknis dalam pengolahan data teks dan pemodelan klasifikasi.

1.6 Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan pada penelitian ini. Teori-teori yang berhubungan seperti pengertian klasifikasi sentimen, algoritma *Logistic Regression*, tahapan *preprocessing* teks, serta tinjauan pustaka dari penelitian-penelitian sebelumnya yang relevan.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi tahapan-tahapan dalam pengumpulan data, perancangan sistem perumusan masalah dan analisa

BAB 4 ANALISA DAN PERANCANGAN

Bab ini berisi analisa dan perancangan sentiment analisis terhadap pelayanan RSUD Rokan Hulu dengan menggunakan metode *Logistic Regression*.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dari analisa dan perancangan dan pengujian pada aplikasi yang berhasil dibangun.

BAB 6 PENUTUP

Bab ini berisi rangkuman dari hasil penelitian yang telah dilakukan dan saran-saran untuk pengembangan aplikasi atau penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1 Machine Learning

Metode *machine learning* merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru cara manusia belajar dan secara bertahap dapat meningkatkan akurasi. Semakin bagus algoritma *machine learning* yang digunakan maka akan semakin baik pula keputusan yang keluar [9].

Machine learning adalah istilah yang digunakan untuk merujuk pada cabang ilmu komputer yang mempelajari metode desain algoritma yang mampu mempelajari atau beradaptasi dengan pola data tanpa diprogram secara eksplisit. *Machine learning* memiliki beberapa metode komputasi yang dapat meningkatkan kinerja dengan memanfaatkan pengetahuan yang didapat dari pengalaman saat belajar [10].

Machine learning merupakan bidang lain dari ilmu komputer yang merancang sebuah algoritma agar memungkinkan sebuah komputer untuk belajar melalui data sehingga sering dikatakan sebagai *learn from data*. Secara garis besar *machine learning* dibagi menjadi 3 yaitu [11] :

- a. *Supervised Learning* adalah algoritma *machine learning* yang proses pembelajarannya dibawah pengawasan. Yang termasuk dalam *supervised learning* diantaranya adalah *classification* dan *regression*.
- b. *Unsupervised Learning* adalah algoritma *machine learning* yang proses pembelajarannya tanpa pengawasan. Yang termasuk dalam

unsupervised learning diantaranya adalah *clustering* dan *dimensionality reduction*.

- c. *Reinforcement Learning* adalah algoritma *machine learning* yang bisa membuat *agent software* mesin bekerja dan bekerja secara otomatis untuk menentukan perilaku ideal sehingga dapat memaksimalkan kinerja algoritma. Yang termasuk dalam Reinforcement Learning diantaranya adalah *real-time decision*, *robot navigation*, *learning tasks*, *skills acquisition* dan *game AI*.

2.2 Kecerdasan buatan (Artificial Intelligence/AI)

Kecerdasan buatan (*Artificial Intelligence/AI*) mengacu pada kemampuan mesin untuk meniru kecerdasan manusia dalam mempelajari, berpikir, dan mengambil keputusan, ini melibatkan penggunaan algoritma dan teknik komputasi yang kompleks untuk mengolah data, mengenali pola, dan membuat prediksi atau tindakan yang cerdas, contoh penggunaannya adalah *machine learning*, *neural networks*, *natural language processing*, *computer vision*, dan robotika, kecerdasan buatan ini bertujuan untuk meningkatkan efisiensi dan kualitas hidup melalui teknologi cerdas. [12].

Kecerdasan buatan (AI) adalah bidang ilmu komputer yang menggunakan simbol daripada angka untuk menyampaikan informasi dan menggunakan metode heuristik atau sejumlah aturan untuk memproses data [13]. Salah satu bidang sains dan teknik yang paling baru adalah kecerdasan buatan, atau AI penelitian tentang kecerdasan buatan dimulai secara resmi setelah Perang Dunia II, dan nama AI pertama kali digunakan pada tahun 1956 [14].

Semakin banyak pengembangan *hardware* dan *software* yang menggunakan teknik kecerdasan buatan menunjukkan betapa berkembangnya kecerdasan buatan saat ini di bidang ilmu komputer [15]. Kemampuan mesin untuk meniru kecerdasan manusia dalam hal berpikir, belajar, dan membuat keputusan disebut kecerdasan buatan (AI). Ini mencakup penggunaan teknik komputasi yang kompleks dan algoritma untuk mengolah data, mengenali pola, dan membuat prediksi atau tindakan cerdas [16].

2.3 Feature Extraction (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan algoritma yang digunakan untuk melakukan pembobotan kata dalam suatu teks. Dalam algoritma ini nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* akan dihitung berdasarkan dokumen yang telah ditentukan pada korpus [17].

TF-IDF atau dapat dikatakan “*Term Frequency – Inverse Document Frequency*” merupakan suatu teknik yang digunakan untuk mengukur kata pada dokumen dengan menghitung bobot setiap kata pada suatu dokumen. Teknik ini banyak digunakan pada *Information Retrieval* dan *Text Mining*. Terdapat beberapa tahap yang akan dilakukan untuk melakukan TF-IDF ini, yang diantaranya yaitu [18] :

2.3.1 Term Frequency (TF)

Term Frequency (TF) berfungsi untuk mengukur jumlah kemunculan suatu kata atau term di dalam dokumen berdasarkan kata kunci yang telah ditentukan

sebelumnya. TF memiliki perhitungan yang mana dapat dilihat pada persamaan 1 seperti berikut :

$$tf_{t,d} = \frac{n_{t,d}}{(Total\ number\ of\ term\ in\ document)} \dots \dots \dots (1)$$

Keterangan dari $tf_{t,d}$ yaitu untuk menghitung kata atau term tertentu dari suatu kata kunci. Lalu untuk keterangan dari d berisi suatu dokumen, dimana dokumen tersebut berisi data yang akan dicocokkan berdasarkan kata kunci. Dimana dalam perhitungan tersebut untuk mencari TF akan diperlukan menghitung banyaknya dari suatu kata yang muncul dalam sebuah dokumen lalu dibagi dengan total dari kata yang ada di dalam dokumen tersebut.

2.3.2 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) digunakan untuk menghitung jumlah dokumen yang ada di dalam suatu korpus atau koleksi. Berikut ini persamaan dari IDF yang dapat dilihat pada persamaan 2 sebagai berikut:

$$idf_a = \log \frac{number\ of\ document}{(Total\ number\ of\ term\ in\ document)} \dots \dots \dots (2)$$

Dalam persamaan untuk mencari nilai idf_d menggunakan logarithm, dimana pada perhitungan ini akan menghitung jumlah dokumen yang ada di dalam korpus dan akan dibagi dengan jumlah dokumen yang mengandung term tertentu.

2.3.3 Term Frequency – Inverse Document Frequency (TF-IDF)

Untuk mendapatkan pembobotan pada setiap kata atau yang disebut juga TF-IDF maka diperlukan perhitungan seperti yang terlihat pada persamaan 3 seperti berikut ini :

$$tfidf_d = tf_{t,d} \times idf_d \dots \dots \dots (3)$$

Pada persamaan 3 tersebut menjelaskan bahwa pembobotan TF-IDF dihasilkan dari perkalian antara $tf_{t,d}$ dikali dengan idf_d . Sehingga hasil dari perkalian antara $tf_{t,d}$ dan idf_d mempresentasikan pembobotan pada setiap kata. Pembobotan setiap kata ini akan menghitung pembobotan dengan jumlah kata yang sama, sehingga hasil dari TF-IDF dapat digunakan tahapan Visualization dan modelling.

2.4 Text Preprocessing

Preprocessing menjadi tahap awal dalam klasifikasi teks untuk mempersiapkan data teks sebelum digunakan pada proses lainnya. Pada tahap ini akan mengubah data teks menjadi bentuk yang lebih baik sehingga menghasilkan informasi teks dengan kualitas yang baik dan siap digunakan pada proses selanjutnya. Pada penelitian ini teknik *preprocessing* yang digunakan meliputi *case folding*, normalisasi kata, *cleaning*, *stopword removal*, dan *stemming* [19].

Tahapan *preprocessing* dibagi menjadi beberapa tahapan yaitu [20] :

a. Case Folding

Case Folding merupakan tahapan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil (*lowercase*). Teks pesan yang diambil dari mesin pencari *Google Alerts* adalah huruf besar/kecil, sehingga perlu diubah menjadi huruf kecil agar mempermudah proses berikutnya.

b. Symbol Removal

Symbol Removal adalah penghapusan simbol digunakan untuk menghilangkan karakter khusus dalam suatu pesan/teks, seperti tanda baca, angka, dan karakter lain kecuali spasi.

c. *Konversi Slangword*

Konversi Slangword proses mengubah kata tidak baku ke dalam kata baku, Tahap ini dilakukan dengan menggunakan bantuan kamus slangword dan padanan dalam kata baku. Kata yang terdapat pada kamus *slangword* nantinya akan diubah menjadi kata baku. Contoh kata *slangword* yaitu baper, sotoy, galau, mager, dll.

d. *Stopword Removal*

Stopword Removal bisa disebut dengan *filtering*, yaitu untuk menghilangkan kata-kata yang tidak penting dalam proses klasifikasi dan penentuan alasan. Contohnya yang, tetapi, atau, ke, di, dengan, dan sebagainya.

e. *Stemming*

Stemming adalah proses pengubahan bentuk kata menjadi kata dasar atau tahap mencari *root* kata dari tiap kata hasil *filtering*.

f. *Tokenization*.

Tokenisasi berfungsi untuk memecah komentar menjadi satuan kata. Proses tokenisasi dilakukan dengan melihat setiap spasi yang ada dalam komentar maka berdasarkan spasi tersebut kata-kata dapat dipecah.

2.5 Ulasan

Para pengguna produk atau layanan dapat membagikan pengalaman mereka setelah menggunakan produk atau layanan melalui platform di internet. Ulasan ini biasanya berisi tentang kesan selama menggunakan produk atau layanan tertentu di suatu tempat. Dengan demikian, ulasan ini sering ditemukan sebagai kumpulan kata-kata positif dan negatif sehingga menghasilkan ulasan yang dapat dipahami [21].

Dengan menganalisis ulasan-ulasan yang diberikan oleh pengguna, baik itu berupa pujian, kritik, atau keluhan, dapat mengeksplorasi wawasan yang berharga tentang kelebihan dan kekurangan layanan, serta identifikasi area yang perlu perbaikan atau peningkatan [22].

2.6 Sentimen Analisis

Analisis sentimen adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan menggunakan teknik analisis teks. Sentimen analisis adalah proses penggunaan *text analytics* untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber [23].

Sentiment analysis adalah teknik untuk menganalisis pikiran, perasaan, dan penilaian pengguna berdasarkan informasi seperti opini tertulis. Hal ini dilakukan untuk menentukan apa yang diyakini pengguna terkait suatu topik atau subjek tertentu. *Sentiment analysis* memiliki keterkaitan yang erat dengan *Natural*

Language Processing (NLP), yang merupakan kemampuan mesin untuk memahami bahasa manusia dengan menggunakan algoritma komputer, matematika, dan linguistik komputasi [24].

Sentimen analisis adalah proses mendeteksi kecenderungan seseorang melalui tulisan atau teks atau mengekstrak data teks untuk mendapatkan informasi tentang sentimen yang biasanya dituliskan didalam media sosial baik bermakna positif, negatif atau netral. Proses analisis ini biasanya dilakukan secara otomatis melalui beberapa aplikasi atau flatform yang mendukung untuk dilakukannya analisis teks [25].

Analisis sentimen memiliki 4 jenis yaitu [26] :

1. Analisis sentimen bertingkat, dimana digunakan untuk mengartikan pemberian bintang dalam ulasan. 5 bintang sangat positif dan 1 bintang sangat negatif.
2. Deteksi Emosi, dimana digunakan untuk mendeteksi emosi seperti rasa Bahagia, marah, sedih, dan frustasi.
3. Berbasis aspek, dimana digunakan untuk mengetahui aspek atau fitur yang memiliki sentiment positif, negatif atau netral
4. Multibahasa, dimana mendeteksi bahasa di dalam teks sesuai dengan bahasa yang dipilih.

2.7 *Logistic Regresion*

Regresi logistik (*Logistic Regression*) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi. Variabel dikotomi biasanya hanya terdiri atas dua nilai yang mewakili

kemunculan atau tidak adanya suatu kejadian yang biasanya diberi angka 0 atau 1 [27].

Model logistik regresi adalah suatu model statistik yang digunakan untuk mengetahui pengaruh variabel prediktor (X) terhadap variabel respon (Y) dengan variabel responnya berupa data dikotomi yaitu bernilai 1 menyatakan bahwa variabel respon memiliki kriteria yang ditentukan dan 0 menyatakan bahwa variabel respon tidak memiliki kriteria yang ditentukan [28].

Algoritma *Logistic Regression* adalah menjelaskan variabel dependen dan variabel independen untuk menghubungkan satu atau lebih variabel bebas dengan variabel terkait yang berisi data berkode 0 dan 1, benar atau salah. Metode ini merupakan metode jenis model klasifikasi, Rumus algoritma *Logistic Regression* memiliki persamaan sebagai berikut [29] :

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1X \dots \dots \dots (4)$$

Keterangan :

$\ln$ = Logaritma natural

B_0 = Konstanta

B_1 = Koefisien masing-masing variabel

x = Variabel independen

p = Probabilitas logistik yang dirumuskan sebagai berikut

$$p = \frac{\exp(B_0 + B_1X)}{1 + \exp(B_0 + B_1X)} = \frac{e^{B_0+B_1X}}{1 + e^{B_0+B_1X}} \dots \dots \dots (5)$$

Keterangan :

exp atau e = Fungsi *exponent*

2.8 Klasifikasi

Klasifikasi merupakan suatu proses untuk menemukan fitur yang dapat membedakan atau mendeskripsikan kedalam kategori kelas data. Klasifikasi bertujuan untuk memberikan perkiraan suatu objek yang belum diketahui labelnya kedalam suatu kelas [30].

Klasifikasi merupakan suatu teknik yang digunakan untuk mendapatkan model yang menjelaskan kelas data untuk melakukan prediksi kelas yang belum diketahui pada suatu objek yang diamati. Teknik ini melakukan manipulasi data dengan memberikan klasifikasi data baru untuk mendapatkan suatu aturan. Analisis Diskriminan dan Regresi Logistik merupakan metode klasifikasi yang biasanya digunakan dalam statistika, akan tetapi semakin berkembangnya teknologi maka dibutuhkan suatu metode analisis yang multifungsi untuk memberikan informasi dari suatu data yang sangat besar sehingga menjadi pengetahuan yang terorganisir. Pada machine learning metode klasifikasi yang sering digunakan diantaranya adalah *Classification and Regression Trees (CART)*, *Random Forest*, *Naïve Bayes*, *Support Vector Machines (SVM)*, dan sebagainya [31].

Beberapa langkah yang digunakan dalam pengklasifikasian data diantaranya langkah training, dimana serangkaian data latih dilakukan analisis dengan tujuan untuk mendapatkan model klasifikasi, dan langkah klasifikasi, yaitu langkah yang digunakan untuk mengidentifikasi model klasifikasi, dimana model yang telah didapatkan digunakan untuk memprediksi kelas untuk data testing[32].

2.9 Python

Python adalah bahasa pemrograman yang populer dan mudah untuk dipelajari. Python banyak digunakan dalam pengembangan perangkat lunak, kecerdasan buatan, pengembangan *web*, *machine learning*, dan analisis data. Python menyiapkan berbagai *Library*, seperti *NumPy* untuk komputasi numerik dan *Pandas* untuk analisis data sehingga memudahkan seseorang untuk melakukan tugas tertentu dengan cepat dan efisien [33].

Python adalah bahasa pemrograman yang fleksibel dan sederhana yang didefinisikan dalam dokumen-dokumennya sebagai. Python adalah sebuah bahasa pemrograman dinamis yang sering digunakan dalam pengembangan aplikasi pada berbagai domain. Hal ini memungkinkan suatu program ditulis dalam beberapa pendekatan sekaligus. Misalnya, antar muka grafis dibuat dalam bentuk orientasi objek, sedangkan pemrosesan dalam bentuk fungsional atau prosedural. Bahasa pemrograman Python memiliki beberapa fitur yang dapat digunakan oleh pengembang perangkat lunak [34].

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python juga didukung oleh komunitas yang besar. Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional [35].

2.10 Google Collab

Google Colab adalah sebuah IDE untuk pemrograman Python dimana pemrosesan akan dilakukan oleh server Google yang memiliki perangkat keras dengan performa yang tinggi. Dari sisi perangkat lunak, *Google Colab* telah menyediakan hampir sebagian besar pustaka (*library*) yang dibutuhkan. Dalam penelitian ini pustaka yang dibutuhkan adalah *Keras*, *TensorFlow*, *NumPy*, *Pandas*, dan pendukung lainnya, misalnya untuk pembuatan grafik lewat *Matplotlib*. Bahkan seluruh versi disediakan, misalnya *TensorFlow* versi 1.x maupun 2.x tersedia, begitu juga versi Python yang mulai dari versi 2.x hingga 3.x [36].

Google Collab adalah *platform* yang memungkinkan pengguna untuk membuat dan berbagi kode secara *online* dengan mudah dan gratis dengan menggunakan bahasa pemrograman python yang menyediakan fitur built in *interpreter* [37].

Google Colaboratory, juga dikenal sebagai *Colab*, adalah *platform* komputasi awan gratis yang menyediakan lingkungan *notebook* tanpa *server* interaktif untuk menulis, berbagi, dan mengeksekusi kode. *Colab* beroperasi pada *notebook Jupyter*, yang kompatibel dengan bahasa pemrograman populer seperti Python, R, dan Julia. *Colab* menawarkan sumber daya komputasi yang kuat, termasuk akses ke mesin virtual jarak jauh berbasis *Linux* dengan unit pemrosesan grafis (GPU) yang canggih hingga 12 jam. Hal ini menjadikannya platform yang nyaman untuk berbagai aplikasi, termasuk tugas pembelajaran mendalam dan pembelajaran mesin [38].

2.11 Penelitian Terdahulu

Untuk penelitian terdahulu dapat dilihat pada tabel berikut :

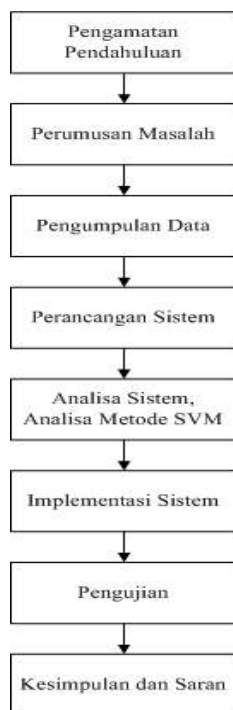
Tabel 2.1 Penelitian Terdahulu

No	Nama Peneliti (Tahun)	Judul Penelitian	Metode	Kesimpulan
1	Fajar & Putri (2021)	Klasifikasi Sentimen Ulasan Pasien terhadap Pelayanan RS Menggunakan <i>Logistic Regression</i>	<i>Logistic Regression</i>	Metode Logistic Regression mampu mengklasifikasikan sentimen dengan akurasi sebesar 85%.
2	Wahyuni (2022)	Analisis Sentimen Komentar Pasien di Media Sosial Menggunakan <i>Logistic Regression</i>	<i>Logistic Regression</i>	<i>Logistic Regression</i> efektif dalam mengolah data komentar teks untuk identifikasi sentimen.
3	Nugroho et al. (2022)	Penerapan <i>Logistic Regression</i> untuk Analisis Sentimen Layanan Publik Bidang Kesehatan	<i>Logistic Regression</i>	Model menunjukkan hasil yang stabil dan akurat pada data ulasan layanan kesehatan.
4	Amelia & Rudi (2023)	Klasifikasi Sentimen Ulasan Pasien terhadap RSUD X dengan <i>Logistic Regression</i>	<i>Logistic Regression</i>	<i>Logistic Regression</i> memberikan hasil yang baik dengan preprocessing yang optimal.
5	Yuliana & Hendra (2024)	Analisis Sentimen Pasien Menggunakan Logistic Regression pada Feedback Online	<i>Logistic Regression</i>	Model mudah diimplementasikan dan memberikan hasil klasifikasi yang dapat diandalkan.

BAB 3

METODOLOGI PENELITIAN

Tahapan penelitian yang akan dilakukan dan penyelesaian masalah terhadap Penerapan *text mining Logistic Regresion* untuk klasifikasi sentimen ulasan pasien terhadap pelayan Rumah Sakit Umum Daerah (RSUD). Adapun tahapan metodologi yang dilakukan selama penelitian dapat dilihat pada Gambar 3.1 yang mana merupakan proses yang dimulai dari studi literatur hingga diperoleh kesimpulan.



Gambar 3.1 Tahapan Metodologi Penelitian

3.1 Pengamatan Pendahuluan

Pengamatan pendahuluan dilakukan sebagai langkah awal untuk memahami kondisi nyata terkait persepsi pasien terhadap pelayanan di Rumah Sakit Umum Daerah (RSUD). Tahapan ini bertujuan untuk memperoleh gambaran awal mengenai bentuk, sumber, dan karakteristik data yang akan digunakan dalam penelitian. Adapun rincian tahapan pengamatan pendahuluan adalah sebagai berikut :

1. Identifikasi Sumber Ulasan

Peneliti mengamati dan mengidentifikasi platform digital tempat pasien memberikan ulasan terhadap pelayanan RSUD pada google maps.

2. Klasifikasi Umum Sentimen

Dari hasil observasi awal, peneliti mencoba mengelompokkan ulasan berdasarkan sentimen secara manual (positif, negatif, netral) untuk memahami distribusi sentimen secara umum.

3. Analisis Karakteristik Data Teks

Peneliti menganalisis struktur teks, panjang ulasan, jenis kata yang sering muncul, serta gaya bahasa yang digunakan oleh pasien, guna menentukan pendekatan *preprocessing* yang sesuai pada tahap berikutnya.

3.2 Pengumpulan Data

Pengumpulan data adalah tahapan-tahapan yang bertujuan dalam memperoleh data-data informasi yang berhubungan dengan penelitian ini, pada

tahapan pengumpulan data ini juga berguna untuk mengumpulkan semua kebutuhan data yang akan diproses nantinya menggunakan metode “Metode *Logistic Regresion*”. Pengumpulan data tersebut dilakukan dengan beberapa cara, antara lain :

1. Observasi

Pengumpulan data melalui observasi dalam penelitian ini dimaksudkan sebagai proses pengamatan langsung terhadap ulasan atau komentar pasien yang tersedia pada google maps. Observasi ini bertujuan untuk memperoleh data primer berupa persepsi, pengalaman, serta pendapat pengguna layanan kesehatan yang disampaikan secara spontan dalam bentuk teks. Proses observasi dilakukan dengan langkah-langkah sebagai berikut :

- a. Mengakses halaman profil RSUD pada Google Maps.
- b. Mengidentifikasi dan menyalin komentar-komentar atau ulasan yang ditinggalkan oleh pasien atau pengunjung.
- c. Menyaring komentar yang relevan, yakni komentar yang berisi opini, keluhan, atau pujian terkait pelayanan rumah sakit seperti pelayanan dokter, perawat, administrasi, kebersihan, waktu tunggu, dan fasilitas.
- d. Mencatat tanggal ulasan dan identitas pengguna (jika tersedia) untuk keperluan analisis temporal atau klasifikasi sumber data.

Dengan pendekatan observasi ini, peneliti dapat memperoleh gambaran autentik dan aktual mengenai kualitas pelayanan dari sudut

pandang pengguna layanan, yang menjadi dasar dalam proses analisis sentimen menggunakan metode yang telah ditentukan.

2. Wawancara

Pengumpulan data melalui wawancara bertujuan untuk memperoleh informasi tambahan secara langsung dari narasumber yang berkompeten, terutama yang berkaitan dengan pengelolaan pelayanan pasien dan tanggapan terhadap ulasan pasien di Rumah Sakit Umum Daerah (RSUD). Wawancara ini dilakukan sebagai pelengkap dari data observasi, dengan harapan dapat memberikan pemahaman yang lebih mendalam mengenai bagaimana rumah sakit merespons dan menindaklanjuti ulasan atau komentar dari pasien. Wawancara ini dilakukan dengan Humas RSUD Rokan Hulu.

3. Studi Pustaka

Pengumpulan data melalui studi pustaka bertujuan untuk memperoleh landasan teoritis dan referensi ilmiah yang mendukung pelaksanaan penelitian. Studi pustaka dilakukan dengan cara menelaah berbagai sumber tertulis yang relevan, seperti :

- a. Jurnal ilmiah nasional
- b. Buku-buku yang membahas *text mining*, analisis sentimen, dan *Logistic Regression*

3.3 Perancangan Sistem

Setelah tahapan analisa selesai dilakukan, maka tahapan selanjutnya adalah perancangan system, tahapan perancangan sistem terdiri dari :

3.3.1 Perancangan Arsitektur Sistem

Perancangan arsitektur sistem dilakukan untuk menggambarkan alur proses dari sistem klasifikasi sentimen yang akan dibangun, mulai dari proses input data hingga keluaran berupa hasil klasifikasi. Arsitektur ini dirancang agar memudahkan dalam implementasi serta memisahkan setiap tahap pengolahan secara modular.

Komponen utama dalam arsitektur sistem arsitektur sistem klasifikasi sentimen secara umum terdiri dari beberapa komponen utama sebagai berikut:

1. Input Data

Data yang digunakan berupa ulasan atau komentar pasien yang diambil dari Google Maps.

2. Preprocessing (Pra-pemrosesan Teks)

Tahap ini bertujuan untuk membersihkan dan menyiapkan data teks sebelum dianalisis. Langkah-langkahnya meliputi:

- a. Tokenisasi : Memecah kalimat menjadi kata.
- b. Case Folding : Mengubah semua huruf menjadi huruf kecil.
- c. Stopword Removal : Menghapus kata-kata umum yang tidak memiliki nilai informasi penting.
- d. Stemming : Mengembalikan kata ke bentuk dasarnya.

3. Feature Extraction (Ekstraksi Fitur)

Setelah data dibersihkan, dilakukan proses untuk mengubah teks menjadi representasi numerik menggunakan metode seperti TF-IDF (*Term Frequency-Inverse Document Frequency*)

4. Model Klasifikasi (Sentiment Classifier)

Modul utama yang digunakan untuk memproses fitur dan menghasilkan prediksi sentimen. Algoritma yang digunakan bisa berupa Logistic Regression.

5. Output (Hasil Klasifikasi)

Hasil akhir dari sistem berupa label sentimen (misalnya: positif, negatif, atau netral) yang dapat digunakan oleh pihak rumah sakit untuk mengevaluasi kualitas pelayanan.

3.3.2 Perancangan Antar Muka (Interface)

Dalam mempermudah komunikasi antara sistem dengan pengguna, maka perlu di rancang antar muka (*interface*), dalam perancangan *interface* hal terpenting yang harus dilakukan ialah bagaimana menciptakan tampilan yang baik dan mudah dimengerti oleh pengguna.

3.4 Analisa Sistem

Tahapan selanjutnya adalah melakukan analisa metode sistem dari penelitian skripsi ini, adapun tahapan analisa dalam penelitian Skripsi ini adalah sebagai berikut:

3.4.1 Analisa Sistem Lama

Analisis sistem lama merupakan langkah penting dalam proses pengembangan atau peningkatan aplikasi, fungsi utama dari analisis sistem lama adalah untuk memahami kondisi saat ini dari sistem yang ada sebelum melakukan perubahan atau pengembangan baru, dengan melakukan analisis sistem lama secara menyeluruh, organisasi dapat membuat keputusan yang lebih baik mengenai bagaimana cara mengembangkan atau memperbarui sistem yang ada, hal ini juga membantu dalam meminimalkan risiko dan memastikan bahwa sistem baru atau yang ditingkatkan dapat memenuhi kebutuhan bisnis dan pengguna dengan lebih efektif.

3.4.2 Analisa Fungsi Sistem Aplikasi

Setelah melakukan tahapan analisis terhadap algoritma *Logistic Regression* maka selanjutnya adalah analisa fungsional sistem yang akan dibangun, adapun tahapan-tahapan analisis fungsional yaitu dalam pembuatan *usecase diagram*, *activity diagram*, *Sequence Diagram* dan *class diagram*.

3.4.3 Analisa Algoritma *Logistic Regression*

Algoritma *Logistic Regression* merupakan salah satu metode klasifikasi yang digunakan untuk memprediksi nilai kategorikal (diskrit), seperti sentimen positif, negatif, atau netral berdasarkan input fitur numerik. Dalam konteks penelitian ini, *Logistic Regression* digunakan untuk mengklasifikasikan ulasan pasien terhadap pelayanan rumah sakit berdasarkan data teks yang telah diproses dan dikonversi menjadi representasi numerik.

3.4.4 Analisa Sistem Baru

Analisis sistem baru adalah langkah penting dalam pengembangan atau implementasi aplikasi baru, fungsi utama dari analisis sistem baru adalah untuk memastikan bahwa sistem yang akan dikembangkan atau diimplementasikan sesuai dengan kebutuhan dan tujuan organisasi, dengan melakukan analisis sistem baru yang komprehensif, organisasi dapat memastikan bahwa sistem yang dikembangkan atau diimplementasikan akan memenuhi kebutuhan bisnis, meningkatkan efisiensi operasional, dan memberikan nilai tambah yang signifikan bagi organisasi dan pengguna.

3.5 Implementasi Sistem Aplikasi

Beberapa komponen pendukung yang memiliki peran yang sangat penting dalam implementasi sistem diantaranya adalah perangkat keras (*hardware*) dan perangkat lunak (*software*) adapun spesifikasi dari perangkat keras (*hardware*) dan perangkat lunak (*software*) yang di gunakan sebagai berikut :

1. Perangkat keras (*hardware*), antara lain :

Prosesor : Intel Core i5

Memory (RAM) : 4 GB RAM

System type : 64-bit Operating System

Harddisk : 500 GB

2. Perangkat Lunak (*Software*), antara lain :

Sistem operasi : Windows 11

3.6 Pengujian Aplikasi

Pengujian sistem ini dilakukan dengan cara menggunakan *Black Box*. Dalam Pengujian *Black Box* ini berfokus pada perangkat lunak untuk

mendapatkan hasil yang baik, apabila terjadi *error* atau tidak sesuai dengan tujuan yang diharapkan maka dilakukan penganalisaan sistem kembali hingga tidak ditemukan *error*. *Black box testing* merupakan pengujian kualitas perangkat lunak yang berfokus pada fungsionalitas perangkat lunak, *pengujian black box testing* bertujuan untuk menemukan fungsi yang tidak benar kesalahan antarmuka kesalahan pada struktur data, kesalahan performansi, kesalahan inisialisasi dan terminasi, dalam pengujian *black box testing* digunakan alat untuk pengumpulan data yang disebut dengan *user acceptance test*, dokumen ini terdiri deskripsi indikator dari prosedur-prosedur pengujian fungsionalitas dari perangkat lunak.

3.7 Kesimpulan dan Saran

Tahapan terakhir adalah menarik kesimpulan dari hasil penelitian yang didapatkan penerapan *Logistic Regresion* untuk klasifikasi sentimen ulasan pasien terhadap pelayan Rumah Sakit Umum Daerah, pada tahapan ini juga berisikan saran peneliti bagi pembaca untuk melakukan pengembangan terhadap penelitian ini kedepannya.