



**APLIKASI ANALISIS SENTIMEN PADA FEEDBACK PASIEN
MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER
(STUDI KASUS : RUMAH SAKIT UMUM SILOAM)**

TUGAS AKHIR

Ulfah Awaliah Mursholihah
41516129116

UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020



**APLIKASI ANALISIS SENTIMEN PADA FEEDBACK PASIEN
MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER
(STUDI KASUS : RUMAH SAKIT UMUM SILOAM)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Ulfah Awaliah Mursholihah
41516120116

UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41516120116

Nama : Ulfah Awaliah Mursholihah

Judul Tugas Akhir : Aplikasi Analisis Sentimen Feedback Pasien Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam)

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 11 Januari 2021



Ulfah Awaliah Mursholihah



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini

Nama Mahasiswa	Ulfah Awaliah Mursholihah
NIM	41516120116
Judul Tugas Akhir	Aplikasi Analisis Sentimen pada Feedback Pasien Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam)

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Noneksklusif (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 11 Januari 2021



Ulfah Awaliah Mursholihah

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Ulfah Awaliah Mursholihah
 NIM : 41516120116
 Judul Tugas Akhir : Aplikasi Analisis Sentimen Pada Feedback Pasien Menggunakan Metode Naïve Bayes Classifier
 (Studi Kasus : Rumah Sakit Umum Siloam)

Menyatakan bahwa Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal :		
	ISSN :		
2	Kertas Kerja, Merupakan material hasil penelitian sebagai kelengkapan Artikel Jurnal. Terdiri dari (minimal 4)	Literatur Review	[]
		Hasil analisa & perancangan aplikasi	[]
		Source code	[]
		Data set	[]
		Tahapan eksperimen	[]
		Hasil eksperimen seluruhnya	[]
3	HAKI Disubmit / Terdaftar	HKI	Diajukan
		Paten	Tercatat
		No & Tanggal Permohonan :	
		No & Tanggal Pencatatan :	

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 11 Januari 2021




Ulfah Awaliah Mursholihah

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120116
Nama : Ulfah Awaliah Mursholihah
Judul Tugas Akhir : Aplikasi Analisis Sentimen Pada Feedback Pasien
Menggunakan Metode Naïve Bayes Classifier (Studi
Kasus : Rumah Sakit Umum Siloam)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Februari 2021



(Drs. Achmad Kodar, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120116
Nama : Ulfah Awaliah Mursholihah
Judul Tugas Akhir : Aplikasi Analisis Sentimen Pada Feedback Pasien Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Februari 2021



(Dr. Ida Nurhaida, M.T)



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120116
Nama : Ulfah Awaliah Mursholihah
Judul Tugas Akhir : Aplikasi Analisis Sentimen Pada Feedback Pasien
Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Februari 2021



(Diky Firdaus, S.Kom, MM)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41516120116
Nama : Ulfah Awaliah Mursholihah
Judul Tugas Akhir : Aplikasi Analisis Sentimen Pada Feedback Pasien
Menggunakan Metode Naive Bayes Classifier
(Studi Kasus Rumah Sakit Umum Siloam)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Februari 2021

Menyetujui,



(Dr. Mujiono Sadikin, MT, CISA, CGEIT)
Dosen Pembimbing

Mengetahui,



(Diky Firdaus, S.Kom, MM)
Koord. Tugas Akhir Teknik Informatika



(Desi Ramayanti, S.Kom, MT)
Ka. Prodi Teknik Informatika

UNIVERSITAS
MERCU BUANA

ABSTRAK

Nama : Ulfah Awaliah Mursholihah
NIM : 41516120116
Pembimbing TA : Drs. Mujiono Sadikin, S.T, M. T,
Judul : Aplikasi Analisis Sentimen Pada Feedback Pasien
Menggunakan Metode Naïve Bayes Classifier
(Studi Kasus : Rumah Sakit Umum Siloam)

Proses analisa opini pada umpan balik dari pelayanan di Rumah Sakit Umum Siloam dilakukan secara manual pada tiap-tiap feedback inpatient oleh QR Department. Pada peningkatan jumlah data feedback, proses analisa membutuhkan waktu yang lebih lama. Dalam hal ini, diperlukan sistem untuk memahami feedback secara otomatis sehingga dapat mengklasifikasikan umpan balik dalam kategori “Complain” atau “Compliment” dengan tingkat akurasi yang baik. Analisis sentimen merupakan proses untuk memahami dan mengelompokkan emosi yang terdapat dalam tulisan menggunakan teknik analisa teks (opinion mining). Sistem analisis sentimen menggabungkan natural language processing dan teknik machine learning. Sistem analisis sentimen memerlukan algoritma untuk menentukan model klasifikasi teks seperti; naive bayes classifier, KNN, Decision tree dll. Metode Naive Bayes Classifier merupakan metode klasifikasi yang berakar pada teorema bayes, yang memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode Naive Bayes Classifier menghasilkan akurasi yang tinggi untuk mengambil keputusan dari ulasan pelanggan.

Aplikasi analisis sentimen pada *feedback inpatient* menggunakan metode naive bayes classifier dari 392 data feedback menggunakan bahasa pemrograman R dengan visualisasi data R shiny membuat data lebih menarik dengan hasil klasifikasi model pada TYPE Compliment sebanyak 301 data dan TYPE Complain sebanyak 56 data. Pada tampilan wordcloud didapatkan 5 kata yang paling menonjol diantaranya ; baik, pelayanan, terimakasih, siloam, ditingkatkan. Aplikasi analisis sentimen menggunakan metode naive bayes classifier dengan model klasifikasi sebanyak 392 record mendapatkan hasil akurasi model 91.1 %. Pemanfaatan model klasifikasi pada data uji sebanyak 95 data, mendapatkan jumlah klasifikasi prediksi masing - masing, 1 Complain dari data aktual sebanyak 5 data dan mendapatkan 90 Compliment dari data aktual sebanyak 90 data. Berdasarkan hal tersebut, metode naive bayes classifier mendapatkan hasil akurasi sebanyak 95.8 % dengan data kesalahan 4 data dari total 95 data uji. Berdasarkan hasil analisa tersebut, sebagian besar pasien yang pernah dirawat di Rumah Sakit Umum Siloam Karawaci merasa puas terhadap pelayanannya

Kata kunci :

Analisis sentimen, Naive Bayes Classifier, Feedback pasien., R.

ABSTRACT

Name : Ulfah Awaliah Mursholihah
Student Number : 41516120116
Counsellor : Drs. Mujiono Sadikin, S.T, M.T
Title : Analysis Sentiment Application in Patient Feedback with Naïve Bayes Classifier Method (Case Study : Rumah Sakit Umum Siloam)

The process of analyzing opinions on feedback from services at RSU Siloam is done manually for each inpatient feedback by the Quality Risk Department. In increasing the amount of data feedback, the analysis process takes more time. In this case, a system is needed to understand feedback automatically so that it can classify feedback into the "Complain" or "Compliment" category with a good level of accuracy. Sentiment analysis is a process to understand and classify the emotions contained in text using opinion mining techniques. The sentiment analysis system combines natural language processing and machine learning techniques. Sentiment analysis systems require algorithms to determine text classification models such as; naive bayes classifier, KNN, Decision tree etc. The Naive Bayes Classifier method is a classification method rooted in the Bayes theorem, which predicts future opportunities based on past experiences. The Naive Bayes Classifier method provides high accuracy for making decisions from customer reviews.

Sentiment analysis application on inpatient feedback using the naive bayes classifier method from 392 feedback data using the R programming language with shiny data visualization to make the data more attractive with the classification results of the model in TYPE Compliment as many as 301 data and TYPE Complain as many as 56 data. In the wordcloud display, there are 5 most prominent words including; baik, pelayanan, terimakasih, siloam, ditingkatkan. Sentiment analysis application using the Naive Bayes classifier method with a classification model of 392 records, getting the model accuracy 91.1%. Utilization of the classification model in the test data as many as 95 data, getting the number of predictive classifications for each, 1 Complain from the actual data as many as 5 data and getting 90 Compliments from the actual data as many as 90 data. Based on this, the Naive Bayes Classifier method obtained an accuracy of 95.8% with 4 error data out of a total of 95 test data. Based on the results of this analysis, most of the patients who have been treated at RSU Siloam Karawaci are satisfied with the service

Keywords :

Sentiment analysis, Naive Bayes Classifier, Patient feedback., R.

KATA PENGANTAR

Puji syukur kita panjatkan pada Allah SWT,
Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, penelitian ini tidak akan berjalan dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Mujiono Sadikin, MT. selaku dosen pembimbing tugas akhir yang telah meluangkan waktu, pengetahuan, dan kesabaran untuk memberikan bimbingan serta arahan kepada penulis dalam menulis menyelesaikan tugas akhir ini.
2. Bapak Leonard Goeirmanto, S.Kom, M.T. selaku Dosen Pembimbing Akademik.
3. Ibu Desi Ramayanti, S.Kom, MT selaku Ketua Program Studi Teknik Informatika.
4. Bapak Diky Firdaus, S.Kom, MM. selaku Koordinator Tugas Akhir Teknik Informatika.
5. Teman-teman seperjuangan Tugas Akhir Teknik Informatika Reguler 2, terima kasih untuk dukungan dan semangatnya dalam usaha untuk meraih gelar sarjana komputer, juga kepada banyak pihak yang telah berkontribusi dalam proses penulisan laporan tugas akhir ini yang tidak mungkin disebutkan satu per satu.

Akhir kata, penulis berharap penulis berharap semoga tugas akhir ini dapat bermanfaat bagi rekan-rekan mahasiswa dan pembaca sekalian.

Penulis menyadari bahwa dalam penulisan laporan tugas akhir ini masih jauh dari kesempurnaan. Karena itu, penulis memohon saran dan kritik yang membangun sehingga dapat berguna untuk pengembangan ilmu pengetahuan.

Jakarta, 11 Januari 2021

Ulfah Awaliah Mursholihah

DAFTAR ISI

HALAMAN SAMPUL.....	1
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
NASKAH JURNAL	xii
DAFTAR ISI.....	xiii
KERTAS KERJA.....	1
BAB 1. LITERATURE REVIEW	3
BAB 2. ANALISIS DAN PERANCANGAN.....	7
BAB 3. SOURCE CODE	11
BAB 4. DATASET.....	20
BAB 5. TAHAPAN EKSPERIMEN.....	27
BAB 6. HASIL SEMUA EKSPERIMEN.....	30
DAFTAR PUSTAKA.....	36
LAMPIRAN KORESPONDENSI.....	37
LAMPIRAN DOKUMEN HAKI.....	38

Aplikasi Analisis Sentimen Pada Feedback Pasien Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam) *

Ulfah Awaliah Mursholihah ^{1*}, Mujiono Sadikin ¹

¹ Fakultas Ilmu Komputer
Universitas Mercu Buana
Jakarta

*41516120116@student.mercubuana.ac.id

² Fakultas Ilmu Komputer
Universitas Mercu Buana
Jakarta

*mujiono@mercubuana.ac.id

Abstrak— Proses analisa opini pada umpan balik dari pelayanan di Rumah Sakit Umum Siloam dilakukan secara manual pada tiap-tiap feedback inpatient oleh QR Department. Pada peningkatan jumlah data feedback, proses analisa membutuhkan waktu yang lebih lama. Dalam hal ini, diperlukan sistem untuk memahami feedback secara otomatis sehingga dapat mengklasifikasikan umpan balik dalam kategori “Complain” atau “Compliment” dengan tingkat akurasi yang baik. Analisis sentimen merupakan proses untuk memahami dan mengelompokkan emosi yang terdapat dalam tulisan menggunakan teknik analisa teks (opinion mining). Sistem analisis sentimen menggabungkan natural language processing dan teknik machine learning. Sistem analisis sentimen memerlukan algoritma untuk menentukan model klasifikasi teks seperti; naive bayes classifier, KNN, Decision tree dll. Metode Naive Bayes Classifier merupakan metode klasifikasi yang berakar pada teorema bayes, yang memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode Naive Bayes Classifier menghasilkan akurasi yang tinggi untuk mengambil keputusan dari ulasan pelanggan.

Aplikasi analisis sentimen pada feedback inpatient menggunakan metode naive bayes classifier dari 392 data feedback menggunakan bahasa pemrograman R dengan visualisasi data shiny membuat data lebih menarik dengan hasil klasifikasi model pada TYPE Compliment sebanyak 301 data dan TYPE Complain sebanyak 56 data. Pada tampilan wordcloud didapatkan 5 kata yang paling menonjol diantaranya ; baik, pelayanan, terimakasih, siloam, ditingkatkan. Aplikasi analisis sentimen menggunakan metode naive bayes classifier dengan model klasifikasi sebanyak 392 record mendapatkan hasil akurasi model 91.1 %. Pemanfaatan model klasifikasi pada data uji sebanyak 95 data, mendapatkan jumlah klasifikasi prediksi masing - masing, 1 Complain dari data aktual sebanyak 5 data dan mendapatkan 90 Compliment dari data aktual sebanyak 90 data. Berdasarkan hal tersebut, metode naive bayes classifier mendapatkan hasil akurasi sebanyak 95.8 % dengan data kesalahan 4 data dari total 95 data uji. Berdasarkan hasil analisa tersebut, sebagian besar pasien yang pernah dirawat di Rumah Sakit Umum Siloam Karawaci merasa puas terhadap pelayanannya

Kata Kunci: Analisis sentimen, Naive Bayes Classifier, Feedback pasien., R.

PENDAHULUAN

. Feedback atau umpan balik dalam bisnis merupakan respon dari pelanggan kepada pemilik bisnis setelah berinteraksi dengan layanan maupun produk yang

ditawarkan oleh penjual atau pemilik usaha kepada pelanggan. Feedback menjadi komponen penting sebagai bahan riset awal untuk membuat sebuah perencanaan bisnis. Pada Rumah Sakit Umum Siloam, feedback patient digunakan sebagai sarana

peningkatan mutu dan kualitas. Umpan balik mengenai pelayanan dan fasilitas Rumah Sakit, disampaikan dalam bentuk Form Feedback Patient dan ulasan online yang disampaikan melalui pesan kepada customer care. Dengan adanya perkembangan data feedback yang semakin banyak, proses analisis menjadi lebih lama, sehingga diperlukan aplikasi untuk memudahkan proses analisa feedback.

Analisis sentimen adalah metode untuk menganalisis sebagian data untuk mengetahui emosi manusia (Dang dkk - 2010). Analisis sentiment dapat dikategorikan ke dalam tiga task, yaitu informative text detection, information extraction dan sentiment interestingness classification yang meliputi emotional, polarity dan identification. Dalam masalah ini, sentimen analisis menjadi metode yang tepat untuk membantu proses analisa feedback patient. Data diproses dan diklasifikasikan ke dalam dua kelas, yaitu sentimen positif dan negatif. Analisis sentimen memerlukan data latih dan data uji, sebagai bahan proses pembuatan machine learning. Data latih diperoleh dari dataset yang sudah dilakukan pre-processing data melalui tahap analisis data dengan menggunakan metode naïve bayes classifier.

Metode Naïve bayes classifier biasa digunakan untuk menyelesaikan masalah opinion mining. Naïve bayes digunakan untuk mengklasifikasikan opini ke dalam opini positif dan opini negatif dan berfungsi dengan baik sebagai metode pengklasifikasi teks. SM Kamaruzzaman dan Chowdury Mofizur Rahman serta Ashraf M Kibriya et.al telah menggunakan NBC

sebagai metode pengklasifikasi teks secara kualitatif dengan akurasi yang tinggi.

Penelitian yang sudah dilakukan oleh Ismail Sunni dan Dwi Hendratmo Widyantoro untuk menentukan Sentimen pada Opini Terhadap Tokoh Publik. Pada data sebanyak 2000 tweet yang dilabeli sebagai positif, negatif, netral menggunakan algoritma Naive Bayes. Untuk melihat karakter topik yang muncul menggunakan metode Tf-Idf mendapatkan hasil, bahwa dengan metode Naive Bayes dan menggabungkan 10 metode pra-proses akurasi yang didapatkan antara 69,4-72,8% . [2]

Penelitian yang dilakukan oleh Sipayung dkk., telah dirancang suatu sistem analisis sentimen dari komentar pelanggan terhadap produk dan layanan hotel dengan menggunakan metode Naïve Bayes Classifier yang menghasilkan 155 komentar kategori positif dan 20 komentar kategori negatif dari 175 data latih dan diperoleh pula nilai akurasi sebesar 75,42%. [4]

Penelitian yang dilakukan oleh Hanggara, Akhriza dan Husni, sentimen dapat diketahui dari opini konsumen atas produk yang disampaikan melalui website yang menjual produk tersebut. Beberapa aplikasi berbasis web untuk mengevaluasi sentimen konsumen sudah tersedia di internet. Analisis sentimen biasanya mengklasifikasi opini menjadi tiga kelas: Sentimen positif, negatif dan netral. Namun demikian aplikasi yang dapat melakukan proses pembuatan training dan testing set (dari dataset opini konsumen terhadap produk), menganalisis sentimen

konsumen sekaligus mengukur akurasi hasil analisis secara dinamis masih jarang tersedia. Penelitian ini bertujuan untuk mengembangkan aplikasi yang mampu melakukan analisis sentimen konsumen dengan fasilitas seperti yang disebutkan sebelumnya, dimana Naive Bayes Classifier digunakan sebagai metode klasifikasi. Hasil analisis pada testing set yang diperoleh dari Goodreads, Twitter, dan Tokopedia menunjukkan bahwa aplikasi yang dikembangkan mampu mengklasifikasikan opini ke dalam tiga kelas positif, negatif dan netral dengan tingkat akurasi rata-rata 70,43% - 88,03%.[5]

METODE PENELITIAN

Penelitian yang dilakukan adalah penelitian eksperimen yang termasuk dalam metode penelitian kuantitatif. Penelitian eksperimen adalah penelitian yang digunakan untuk mengetahui pengaruh variabel independen terhadap dependen dalam kondisi yang terkendali. Metode eksperimen yang digunakan adalah metode supervised learning, yaitu metode pembelajaran mesin yang membutuhkan data testing dan data training sehingga hasilnya berupa data output sesuai dengan proses training data yang telah dilakukan. Metode supervised learning yang digunakan untuk analisis sentimen adalah naïve bayes classifier.

Adapun tahapan dalam penelitian ini adalah sebagai berikut :



Gambar 1. Diagram Alir Penelitian

Studi Literatur didapat dari berbagai sumber baik jurnal, buku, dokumentasi, internet, pustaka dan segala referensi yang mendukung guna kebutuhan penelitian dalam bidang analisis sentimen. Pengumpulan data yang digunakan untuk eksperimen diperoleh dari formulir umpan balik pasien yang pernah dirawat inap di Rumah Sakit Umum Siloam. Metode pengolahan data yang digunakan adalah teknik analisa sentimen atau opinion mining, maka diperlukan data training dan data testing untuk membuat model machine learning. Kerangka konsep sistem dibuat untuk menggambarkan tahapan proses yang akan dilakukan oleh aplikasi analisis sentimen mulai dari tahap pengambilan data, preprocessing, klasifikasi dan penghitungan akurasi. Sedangkan implementasi sistem, merupakan tahapan penerapan kerangka konsep sistem sampai dengan data

divisualisasikan untuk mendapatkan hasil analisis dan kesimpulan.

Dalam proses analisis data, digunakan R-Programming sebagai bahasa programnya. R merupakan bahasa pemrograman tingkat tinggi yang serbaguna interaktif, berbasis S dan populer pada saat ini. R-Programming memiliki kelebihan yaitu fitur yang lengkap, dynamic, handal dan open source. Meski R-Programming pada saat ini sudah banyak digunakan oleh data scientist, namun penelitian menggunakan bahasa-R ini masih sedikit ditemukan di Indonesia.

Analisa Hasil Sentimen analisis ulasan secara mudah dapat dilakukan dengan melihat jumlah bintang yang diberikan oleh pembeli, tetapi jumlah bintang tidak dapat mewakili isi dari keseluruhan ulasan. Diperlukan melihat seluruh isi komentar ulasan untuk dapat mengetahui keseluruhan maksud ulasan. Sangat dimungkinkan untuk menganalisis ulasan secara manual dengan melihat satu persatu, namun apabila ulasan yang dimiliki banyak akan lebih cepat menggunakan sistem analisis sentimen [1].

Sentimen Analisis

Analisis Sentimen (Sentiment Analysis) disebut juga sebagai Opinion Mining adalah sebuah studi komputasional dalam bidang Natural Language Processing (NLP) yang bertujuan untuk menentukan nilai sentimen dalam sebuah dokumen teks baik berasal dari product review, movie review maupun political review. Nilai sentiment dapat berupa

pendapat, sikap maupun emosi terhadap entitas. Analisis sentiment dapat dikategorikan dalam aktivitas klasifikasi, yaitu mengelompokkan sentiment dalam sebuah data text ke dalam tiga kelas, yaitu positif, negatif atau netral.[11]

Naive Bayes Classifier

Learning yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, Naïve Bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah.[12]

2.1 Metode Pengumpulan Data

Pada penelitian ini, data yang digunakan diperoleh dari Umpan Balik atau Feedback Inpatient dari pasien maupun keluarga pasien yang dirawat inap di Rumah Sakit Umum Siloam. Pasien maupun keluarga pasien mengisi formulir Feedback InPatient dan diserahkan pada Quality and Risk Department. QR Department membuat laporan umpan balik setiap harinya untuk kemudian dianalisis sebagai bahan evaluasi kualitas pelayanan rumah sakit.

Data umpan balik yang diperoleh kemudian disimpan dengan file ekstensi .csv. Adapun data yang berhasil dikumpulkan adalah sebanyak 487 record dengan rincian sebagai berikut :

Tabel 1. Struktur Dataset Feedback Patient

Atribut	Tipe Data	Range Nilai
TGL DISCHARGE	Date	Date
NAMA PASIEN	Varchar	Varchar
MOBILE	Varchar	Varchar
MR	Integer	Integer
RUANG	Varchar	Varchar
CATEGORY	Varchar	Varchar
COMMENTS	Text	Text
TYPE	Varchar	Varchar

Keterangan Tabel :

Tgl Discharge : Tanggal feedback dibuat

Nama Pasien : Merupakan nama pasien yang pernah dirawat

Mobile : Nomor telepon pasien / keluarga

MR : Nomor rekam medik pasien

Ruang : Nama ruangan tempat pasien di rawat

Category : Kategori tujuan Feedback, contoh : FO,

Maintenance, Perawat, Dokter, Housekeeping

Comments : Teks umpan balik yang akan digunakan

sebagai bahan penelitian

Type : Label umpan balik yang berisi complain, dan compliment

Data yang diperoleh merupakan data terstruktur yang memudahkan dalam pembuatan model machine learning. Atribut dalam dataset tersebut tidak semua digunakan dalam penelitian ini. Atribut yang

digunakan dalam proses analisis sentimen adalah atribut “COMMENTS” dan “TYPE”.

Dalam proses analisis sentimen, atribut COMMENTS merupakan data teks yang dianalisis dan digunakan sebagai data latih dan data uji dengan perbandingan 8:2, 80% data digunakan sebagai data latih dan 20 % data digunakan sebagai data uji. Atribut TYPE merupakan atribut yang digunakan untuk melabeli umpan balik dengan label Compliment dan Complain. Label Compliment berisi ulasan bersifat positif, dan Complain berisi ulasan bersifat negatif.

Adapun rincian penghitungan data latih dan data uji nya adalah sebagai berikut :

Data Latih = 80 % x 487 = 392 data

Data Uji = 20% x 487 = 95 data

Berikut merupakan dataset yang digunakan.

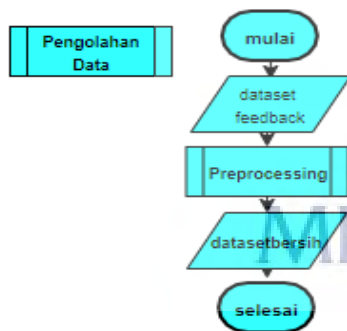
No	COMMENTS	TYPE
1	parkir motor jangan terlalu jauh dari RS	Complain
2	untuk pelayanan (staff) di maternity untuk lebih ramah lagi dalam melayani keluarga pasien	Complain
3	tenaga kebersihan untuk kamar inap masih agak kurang .	Complain
4	perawatnya lebih ramah lagi	Complain
5	pelayanan di gedung B kurang memuaskan	Complain
6	agar motor bisa masuk UGD (saya lari-lari bawa anak ke UGD karena motor tdk bisa masuk) kamar kelas	Complain

	1 untuk anak-anak tolong diadakan	
7	fasilitas kamar mandi yang perlu ditingkatkan	Complain
8-486
487	toiletnya kadang kurang wangi, bagi pasien yang cuek biasanya tidak disemprot lantainya supaya tidak berbau	Complain

Gambar 2. Dataset Feedback InPatient

2.2 Metode Pengolahan Data

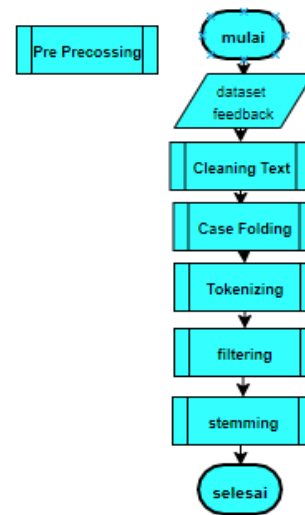
Data yang sudah dikumpulkan kemudian diolah sehingga mendapatkan model machine learning sebagai model klasifikasi analisis sentimen. Berikut merupakan diagram alir pengolahan data :



Gambar 3. Diagram Alir Pengolahan Data

Pre-Processing

Data yang digunakan dalam proses analisa tidak selamanya dalam kondisi ideal untuk diproses, maka pada tahap awal pengolahan data dilakukan pra-proses pada data mentah untuk menghindari data yang kurang sempurna.



Gambar 4. Tahap Pra Proses Data

Adapun tahapan – tahapan dalam preprocessing data adalah :

Cleaning text : Merupakan tahap dalam pembersihan data karakter selain huruf, seperti tanda baca, mention, hashtag, maupun URL.

Case Folding : Merupakan teknik untuk mengubah semua huruf dalam data menjadi huruf kecil, hanya huruf “a” sampai dengan “z”.

Tokenizing : merupakan tahap parsing atau pemotongan string input berdasarkan tiap kata penyusunnya.

Filtering : Tahap ini merupakan tahap pengambilan kata-kata penting dari hasil token. Pada tahapan ini, karena dataset berbahasa Indonesia, diperlukan stopwords (kata yang tidak deskriptif yang dapat dibuang) berbahasa Indonesia untuk memudahkan proses filtering data. Seperti kata berikut : “yang”, ”dan”, “di”, ”dari”, ”dengan” dan kata lainnya.

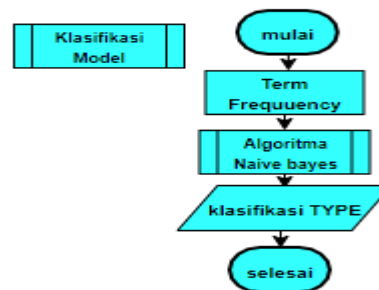
Stemming : Stemming merupakan tahap pencarian akar kata dari tiap kata hasil filtering. Pada proses ini, kata berbahasa Indonesia lebih sulit diterapkan karena bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen. Proses stemming dilakukan dalam upaya normalisasi data untuk mendapatkan akurasi analisis sentimen yang tinggi.

Setelah tahapan pra-proses selesai dilakukan, maka didapat dataset yang bersih kemudian disimpan untuk tahap pemodelan menggunakan metode klasifikasi naïve bayes classifier. Menurut Zhao (2012), sebuah term-document matrix menunjukkan hubungan antara term dan dokumen, dimana setiap baris berisi term dan setiap kolom untuk dokumen. Sebelum masuk pada tahap klasifikasi model, data yang sudah bersih dilakukan proses pembobotan data untuk mendapatkan frekuensi kemunculannya menggunakan Term Frequency-Inverse Document Frequency (TF - IDF).

Klasifikasi Model

Pada proses ini, data bersih yang diperoleh dari proses sebelumnya diberikan label atau TYPE kategori “Compliment” dan “Complain”. Pemberian label yang tepat, akan meningkatkan performa ketepatan klasifikasi. Proses pengklasifikasian data menggunakan algoritma naïve bayes classifier dilakukan dengan tiga tahapan utama, yaitu ; menghitung nilai prior, menghitung nilai likelihood, serta menghitung nilai posterior. Prior merupakan peluang kemunculan dari sebuah kategori yang digunakan pada proses klasifikasi. Nilai posterior

didapat dari hasil perkalian nilai prior dan nilai likelihood. Dibawah ini merupakan gambaran flowchart proses pengolahan Dataset feedback menggunakan algoritma NBC :



Gambar 5. Diagram Alir klasifikasi model menggunakan algoritma naïve bayes

Naive Bayes Classifier adalah salah satu pengklasifikasian statistik, dimana pengklasifikasian ini dapat memprediksi probabilitas keanggotaan kelas suatu data tuple yang akan masuk ke dalam kelas tertentu. Naive bayes classifier yang dikemukakan oleh ilmuwan inggris Thomas Bayes, dapat memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode ini sering digunakan dalam menyelesaikan masalah dalam bidang machine learning karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan yang sederhana.

Teorema bayes merupakan dasar aturan dari naive bayes classifier berikut teorema bayes akan disajikan pada persamaan :

$$P(H|X) = \frac{P(X|H)}{\sum_{i=1}^n P(H_i|X)} \cdot P(H)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H beredar kondisi X (posterior probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas x berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Teorema Bayes sering pula dikembangkan mengingat berlakunya hukum probabilitas total, menjadi seperti berikut:

$$P(H|X) = \frac{P(X|H)}{\sum_{i=1}^n P(H_i|X)} \cdot P(H)$$

Keterangan :

i : 1,2,3,...., n jumlah data Hipotesis (Prior Probabilitas)

dimana : $H_1 \cup H_2 \cup H_3 \dots \cup H_n = S$

S : Probabilitas total H

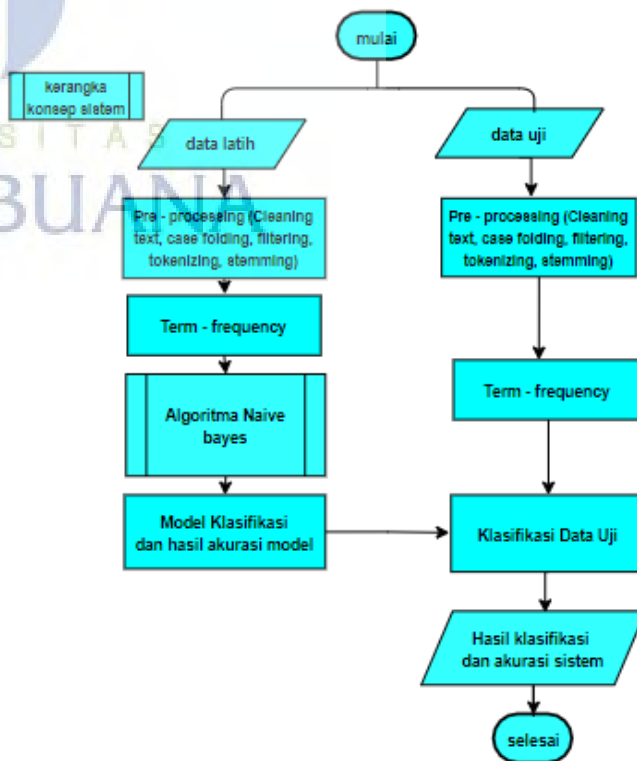
Untuk menjelaskan teorema naive bayes perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema bayes diatas disesuaikan sebagai berikut :

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence).

2.3 Kerangka Konsep Sistem

Kerangka konsep system dari penelitian ini digambarkan pada gambar 6.



Gambar 6. Kerangka konsep sistem

Adapun tahapan – tahapan proses aplikasi analisis sentimen berdasarkan gambar tersebut adalah proses pengumpulan data, pembagian data latih dan data uji, pre processing, sampai dengan perhitungan klasifikasi dan akurasi. Gambaran output system yang diinginkan adalah prediksi klasifikasi kategori (Compliment , Complain) hasil dari analisis sentimen serta akurasi yang didapatkan menggunakan algoritma Naïvbayes classifier dalam bentuk persentase.

Bahasa pemrograman R

Dalam proses analisis sentimen digunakan bahasa pemrograman R. Bahasa R merupakan suatu fasilitas perangkat lunak terpadu untuk memanipulasi data, simulasi, kalkulasi dan peragaan grafik. R memiliki kemampuan menganalisis data dengan efektif dan dilengkapi dengan operator pengolahan array dan matriks. R dibuat oleh Ross Ihaka dan Robert Gentleman pada tahun 1995 sebagai implementasi dari bahasa program S.

Dalam proses analisis sentimen, pada pemrograman R dibutuhkan beberapa library untuk mendukung proses analisis. Library yang digunakan dalam proses pemodelan analisis sentimen pada R adalah sebagai berikut :

library(tm) : Merupakan package dalam R yang biasa digunakan dalam NLP untuk proses preprocessing, impor data, dan penanganan korpus serta manajemen metadata.

library(NLP) : Digunakan untuk menyederhanakan pemrosesan awal

library(textclean) : Digunakan dalam proses pembersihan dan normalisasi data

library(stringr) : berfokus pada fungsi manipulasi string

library(caret) : Kependekan dari Classification And Regression Training. Digunakan untuk merampingkan proses pelatihan model pada masalah klasifikasi dan regresi

library(dplyr) : Digunakan pada saat eksplorasi data untuk memanipulasi data

library(tau) : digunakan untuk proses cleaning data

library(parallel) : Dukungan untuk pemrosesan paralel pada R

library(katadasaR) : Paket untuk stemming bahasa Indonesia pada R

library(stopwords) : Menyediakan berbagai sumber stopword untuk digunakan dalam analisis teks dan pemrosesan bahasa alami

library(e1071) : Paket dukungan untuk analisis data dalam bidang klasifikasi

library(naivebayes) : Digunakan dalam pengklasifikasian data menggunakan metode naive bayes.

R Shiny

Shiny merupakan sebuah paket dalam R yang mengizinkan penggunaannya membangun web apps yang interaktif. Shiny menggabungkan antara

kekuatan komputasi statistika R dan interaksinya dengan web modern. Visualisasi dari hasil analisis sentimen ditampilkan pada Shiny yang merupakan salah satu framework dalam R untuk membangun aplikasi berbasis website. Untuk menampilkan shiny maka harus digunakan Bahasa pemrograman R[3].

HASIL DAN PEMBAHASAN

Dari rangkaian proses penelitian yang dilakukan, didapatkan data bersih dari hasil *pre-processing*, model *machine learning* dengan algoritma *naïve bayes classifier*, dan sistem aplikasi analisis sentimen.

Berikut merupakan rincian data yang digunakan :

```
Rows: 487
Columns: 3
$ X      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12...
$ text   <chr> "parkir motor jangan terlalu jauh dar...
$ datanya.TYPER <chr> "Complain", "Complain", "Complain", "..."
```

Gambar 6. Rangkuman Dataset yang digunakan

```
Complain Compliment
92          395
```

```
Complain Compliment
0.1889117 0.8110883
```

Gambar 7. Rincian kelas kategori

Berdasarkan keterangan diatas, dataset yang digunakan memiliki jumlah tipe kategori *Complain* sebanyak 92 data, dan tipe kategori *Compliment* sebanyak 395 data. dengan data akurasi acuan 0.1889117 untuk tipe kategori *Complain*, dan 0.8110883 untuk tipe kategori *Compliment*.

3.1 Preprocessing

Pra-Proses pada tahap awal penelitian menghasilkan data bersih dengan rincian sebagai berikut :

a. *Cleaning Text*

Pada proses ini, dataset yang masih terdapat tanda baca, karakter, angka maupun simbol dihilangkan.

b. *Case Folding*

Pada proses ini, semua huruf besar dalam data diganti menjadi huruf kecil “a” sampai “z”.

Tabel 2. Proses Case Folding

Sebelum	Sesudah
[1] parkir motor jangan terlalu jauh dari RS	[1] parkir motor jangan terlalu jauh dari rs
[2] untuk pelayanan staff di maternity untuk lebih ramah lagi dalam melayani keluarga pasien	[2] untuk pelayanan staff di maternity untuk lebih ramah lagi dalam melayani keluarga pasien
[3] tenaga kebersihan untuk kamar inap masih agak kurang .	[3] tenaga kebersihan untuk kamar inap masih agak kurang
[4] perawatnya lebih ramah lagi	[4] perawatnya lebih ramah lagi
[5] pelayanan di gedung B kurang memuaskan	[5] pelayanan di gedung b kurang memuaskan
[6] agar motor bisa masuk UGD (saya lari-lari bawa anak ke UGD karena motor tdk bisa masuk) kamar kelas 1 untuk anak-anak tolong diadakan	[6] agar motor bisa masuk ugd (saya lari-lari bawa anak ke ugd karena motor tdk bisa masuk) kamar kelas untuk anak-anak tolong diadakan
[7] fasilitas kamar mandi yang perlu ditingkatkan	[7] fasilitas kamar mandi yang perlu ditingkatkan
[8] mohon diperbaiki layanan administrasi rawat inap bila istirahat ada yang gantinya supaya proses kepulangan bisa cepat dan pasien baru yg akan menempati kamar dapat juga segera dipakai	[8] mohon diperbaiki layanan administrasi rawat inap, bila istirahat ada yang gantinya supaya proses kepulangan bisa cepat dan pasien baru yg akan menempati kamar dapat juga segera dipakai
[9] fasilitas ac diruang pendaftaran dan ruang tunggu	[9] fasilitas ac diruang pendaftaran dan ruang tunggu
[10] mohon ditingkatkan lagi untuk pelayanan rs baik dari sisi administrasi perawat maupun dokter . jangan biarkan pasien ranap	[10] mohon ditingkatkan lagi untuk pelayanan rs baik dari sisi administrasi perawat maupun dokter . jangan biarkan pasien ranap

[9] fasilitas AC diruang pendaftaran dan ruang tunggu	menunggu lama untuk penggantian infus dll
[10] mohon ditingkatkan lagi untuk pelayanan RS baik dari sisi administrasi jangan biarkan pasien ranap menunggu lama untuk penggantian infus dll	

[3] tenaga kebersihan untuk kamar inap masih agak kurang	[5] pelayanan gedung b memuaskan
[4] perawatnya lebih ramah lagi	[6] motor masuk ugd lari lari bawa anak ugd motor tdk masuk kamar kelas anak anak tolong diadakan
[5] pelayanan di gedung b kurang memuaskan	[7] fasilitas kamar mandi ditingkatkan
[6] agar motor bisa masuk ugd (saya lari-lari bawa anak ke ugd karena motor tdk bisa masuk) kamar kelas untuk anak-anak tolong diadakan	[8] mohon diperbaiki layanan administrasi rawat inap istirahat gantinya proses kepulangan cepat pasien yg menempati kamar dipakai
[7] fasilitas kamar mandi yang perlu ditingkatkan	[9] fasilitas ac di ruang pendaftaran ruang tunggu
[8] mohon diperbaiki layanan administrasi rawat inap bila istirahat ada yang gantinya supaya proses kepulangan bisa cepat dan pasien baru yg akan menempati kamar dapat juga segera dipakai	[10] mohon ditingkatkan pelayanan rs sisi administrasi perawat dokter biarkan pasien ranap menunggu penggantian infus dll
[9] fasilitas ac di ruang pendaftaran dan ruang tunggu	
[10] mohon ditingkatkan lagi untuk pelayanan rs baik dari sisi administrasi perawat maupun dokter . jangan biarkan pasien ranap menunggu	

c. *Tokenizing*

Pada tahap ini, pemotongan string input berdasarkan tiap kata yang menyusunnya .

Contoh:

```
[[487]]
[1] tidak ada, sudah cukup baik dan cepat dalam pelayanan untuk pasien-nya.
> token[[487]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 71
```

Gambar 7. Tokenizing

Berdasarkan data diatas terdapat 7 metadata penting penyusun kalimat, dan terdapat 71 karakter yang digunakan pada data ke- 487.

d. *Filtering*

Pada proses ini, digunakan data external stopword.txt sebagai pedoman daftar kata stopword Bahasa Indonesia.

Tabel 3. Feedback sebelum di filtering dan sesudah di filtering

Sebelum	Sesudah
[1] parkir motor jangan terlalu jauh dari rs	[1] parkir motor rs
[2] untuk pelayanan (staff) di maternity untuk lebih ramah lagi dalam melayani keluarga pasien	[2] pelayanan staff maternity ramah melayani keluarga pasien
[3] tenaga kebersihan kamar inap	[3] tenaga kebersihan kamar inap
[4] perawatnya ramah	[4] perawatnya ramah

lama untuk penggantian infus dll	
----------------------------------	--

e. *Stemming*

Proses terakhir dalam tahapan *Preprocessing data* adalah tahap *Stemming* atau tahap penyederhanaan kata – kata menjadi kata dasar. Berikut merupakan hasil proses stemming .

Tabel 4. Tahap Stemming

Sebelum	Sesudah
[1] parkir motor rs	[1] parkir motor rs
[2] pelayanan staff maternity ramah melayani keluarga pasien	[2] pelayanan staff maternity ramah melayani keluarga pasien
[3] tenaga kebersihan kamar inap	[3] tenaga kebersihan kamar inap
[4] perawatnya ramah	[4] perawatnya ramah
[5] pelayanan gedung b memuaskan	[5] pelayanan gedung b memuaskan
[6] motor masuk ugd lari lari bawa anak ugd motor tdk masuk kamar kelas anak anak tolong diadakan	[6] motor masuk ugd lari lari bawa anak ugd motor tdk masuk kamar kelas anak anak tolong diadakan
[7] fasilitas kamar mandi ditingkatkan	[7] fasilitas kamar mandi ditingkatkan
[8] mohon diperbaiki layanan administrasi rawat inap istirahat gantinya proses kepulangan cepat pasien yg menempati kamar dipakai	[8] mohon diperbaiki layanan administrasi rawat inap istirahat gantinya proses kepulangan cepat dan pasien yg menempati kamar dipakai
[9] fasilitas ac di ruang pendaftaran	[9] fasilitas ac di ruang pendaftaran
[10] mohon ditingkatkan pelayanan rs sisi administrasi perawat dokter biarkan pasien ranap menunggu penggantian infus dll	[10] mohon ditingkatkan pelayanan rs sisi administrasi perawat dokter biarkan pasien ranap menunggu penggantian infus dll

ruang tunggu	
[10] mohon ditingkatkan pelayanan rs sisi administrasi perawat dokter biarkan pasien ranap menunggu penggantian infus dll	

f. Pembobotan

Pembobotan kata bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. *Term Frequency* (TF) merupakan jumlah kemunculan atau frekuensi kata pada suatu dokumen. Sedangkan *Inverse Document Frequency* (IDF) bertujuan untuk mengetahui apakah term yang dicari cocok dengan kata kunci yang diinginkan. Term yang sering muncul memberikan pengaruh yang kecil dalam menentukan keterkaitan kata kunci dokumen.

TF-IDF dihitung dengan menggunakan persamaan seperti berikut [15]:

$$W_{j,i} = \frac{n_{j,i}}{\sum_k n_{k,i}} \cdot \log_2 \frac{D}{d_j}$$

dimana:

$W_{j,i}$ = Pembobotan TF-IDF untuk term ke j pada dokumen ke i.

$n_{j,i}$ = Jumlah kemunculan term ke j pada dokumen ke i.

$\sum_k n_{k,i}$ = Jumlah kemunculan seluruh term pada dokumen ke i.

D = Banyaknya dokumen yang dibangkitkan.

d_j = Banyaknya dokumen yang mengandung term ke j.

Pada proses pembobotan kata didapatkan hasil sebagai berikut :

```

> dtm<-DocumentTermMatrix(corpus.clean)
> inspect(dtm[1:10,1:20])
<<DocumentTermMatrix (documents: 10, terms: 20):
Non-/sparse entries: 26/174
Sparsity : 87%
Maximal term length: 11
weighting : term frequency (tf)
Sample :
  Terms
Docs 1 2 3 4 5 6 7 8 9
1 0 0 0 0 0 0 0 0 0
10 0 0 0 0 0 0 0 0 0
2 0 0 0 0 0 0 0 0 0
3 0 0 0 0 0 0 0 0 0
4 0 0 0 0 0 0 0 0 0
5 0 0 0 0 0 0 0 0 0
6 0 0 0 0 0 0 0 0 0
7 1 0 0 0 0 0 0 0 0
8 0 0 0 0 0 0 0 0 0
9 1 0 0 2 0 0 0 0 0

  Terms
Docs mudah mudahan pelayanan siloam
1 0 0 0 0 0
10 0 0 0 0 0
2 0 0 0 1 1
3 0 0 0 0 0
4 1 1 0 0 0
5 0 0 0 0 0
6 1 1 0 0 0
7 0 0 0 0 0
8 0 0 0 0 0
9 0 0 1 1 1

```

Gambar 8. Contoh Pembobotan Kata

3.2 Model Machine Learning dengan Algoritma Naive Bayes Classifier

Dataset yang sudah melalui *preprocessing* kemudian disimpan dan dibagi menjadi data training dan data testing. Data training dan data testing yang digunakan sudah dilabeli kelas Compliment dan kelas Complain. Kemudian data tersebut dihitung *term frequency* menggunakan pembobotan TF-IDF untuk menilai bobot relevansi *term* dari sebuah data terhadap keseluruhan dokumen yang ada. Proses pemodelan menggunakan algoritma *naive bayes classifier* dilakukan menggunakan pemrograman R dengan memanfaatkan beberapa library.

Masuk pada tahap klasifikasi menggunakan metode NBC, Dataset yang digunakan adalah 487 record dengan jumlah *feedback Complain* sebanyak 92 *record* dan *feedback Compliment* sebanyak 395 *record*. Pembagian data latih dan data uji sebanyak 392 *record* [1:392] dan 95 *record* [393:487]. Untuk membuat *classifiernya* diperlukan fitur yang sesuai dan relevan sebagai kamus dalam menentukan kelas kategori yang tepat untuk proses analisa selanjutnya, apakah *feedback* termasuk *Complain* ataupun *Compliment*. Berikut ini adalah hasil klasifikasi menggunakan metode NBC :

df.datanya.TYPE.1.392.	prediksitrain	text
1 Complain	Complain	parkir motor jangan lalu jauh dari rs
2 Complain	Compliment	untuk ayan staff di maternity untuk lebih ramah lagi dalam l...
3 Complain	Complain	tenaga bersih untuk kamar inap masih agak kurang
4 Complain	Compliment	awat lebih ramah lagi
5 Complain	Compliment	ayan di gedung b kurang muas
6 Complain	Complain	agar motor bisa masuk ugd saya larilari bawa anak ke ugd k...
7 Complain	Complain	fasilitas kamar mandi yang perlu tingkat
8 Complain	Complain	mohon baik layan administrasi rawat inap bila istirahat ada ...
9 Complain	Compliment	fasilitas ac ruang daftar dan ruang tunggu
10 Complain	Complain	mohon tingkat lagi untuk ayan rs baik dari sisi administrasi ...

Gambar 9. Klasifikasi feedback Prediksi dan Aktual

Model yang didapatkan dari hasil klasifikasi menggunakan metode naive bayes kemudian dilakukan perhitungan nilai *apper* pada data latih dan data uji untuk menilai ketepatan klasifikasinya. Berikut ini adalah hasil perhitungan klasifikasi feedback menggunakan naive bayes classifier :

Total Observations in Table: 392

predicted	actual		Row Total
	Complain	Compliment	
Complain	56 0.933 0.644	4 0.067 0.013	60 0.153
Compliment	31 0.093 0.356	301 0.907 0.987	332 0.847
Column Total	87 0.222	305 0.778	392

Gambar 10. Hasil Akurasi sistem Berdasarkan hasil klasifikasi model, dapat dihitung akurasi model *naive bayes classifier* menggunakan nilai *APPER*. Kesalahan klasifikasi pada model adalah sebanyak 35 data.

$$APPER = (35:392) \times 100\% = 8.9 \%$$

Dari hasil klasifikasi tersebut, perhitungan kesalahan model adalah sebesar 8.9 %, maka hasil akurasi model adalah sebesar 91.1 %.

3.3 Visualisasi Data Aplikasi Analisis Sentimen menggunakan R Shiny

R Shiny digunakan untuk visualisasi data. Tampilan hasil analisis data dibagi menjadi beberapa tampilan, diantaranya :

Input dataset

Pada halaman ini, dataset yang telah disiapkan di input dan ditampilkan pada aplikasi menggunakan r shiny. Dataset awal yang terdiri dari 8 atribut dengan jumlah 487 record ditampilkan pada halaman awal menu tab Data.

TGL DISCHARGE	PATIENT NAME	MOBILE	MR	RUANG	CATEGORY	COMMENTS	TYPE	URGENSI
2020-01-04	DHESY MAHARANI	8964891908	885883	Medical & Surgical Lantai 2	Parkir	parkir motor jangan terlalu jauh dari RS	Complain	Low
2020-01-04	RESKI PUTRAMA RENTA	8132722459	896847	Maternity	Perawat Rawat Inap	untuk perawatan staff di maternity untuk lebih ramah lagi dalam melayani keluarga pasien	Complain	Low
2020-01-11	ENI INDRANA	8771819059	435838	Maternity	Housekeeping	tenaga kebersihan untuk kamar Inap masih agak kurang	Complain	Low
2020-01-11	RAMLIAN HATTAHUKU	689129		Lantai 6	Perawat Rawat Inap	perawatnya lebih ramah lagi	Complain	Low
2020-01-15	Hulu anjris	8190213800	848538	118 Podiatry	Pendaftaran Rawat Inap	pelayanan di gedung B kurang memuaskan	Complain	Low
2020-01-15	DUMALJI AGATHA SIKHARJAS	829530		Lantai 6	Facility / Maintenance	apgar motor bisa masuk RSUD (bisa terlanjur bawa anak ke RSUD karena motor tidak bisa masuk) kamar kelas 1 untuk anak-anak tidak dibedakan	Complain	Low
2020-01-15	ELAWATI	8151879030	806910	Lantai 6	Facility / Maintenance	fasilitas kamar mandi yang perlu ditingkatkan	Complain	Low
2020-01-15	MARLOD	8861699077	20881	Lantai 6	Perawat Rawat Inap	mohon diperbaiki layanan administrasi rawat inap. Disarankan ada yang garhwa supaya proses skrulingnya bisa cepat dan pasien bang ya	Complain	Low

Gambar 11. Halaman awal input data

Preprocessing

Pada halaman ini, data input yang sudah disimpan dalam sistem kemudian dibaca dan dilakukan tahap *preprocessing*, sama dengan proses yang dilakukan di awal pada model. Pada saat halaman ini diklik, maka otomatis data input yang dimasukkan sebelumnya diproses dan kemudian ditampilkan hasil proses nya dalam bentuk tabel. Dataset ditampilkan dalam bentuk tabel dengan 3 atribut, yaitu :x, text , datanya.type. Atribut x merupakan atribut penomoran data tabel. Atribut **text**, merupakan hasil *query* atribut COMMENTS, sedangkan **datanya.type** merupakan hasil *query* atribut TYPE. Pada tabel COMMENTS berisi teks *feedback* yang sudah bersih dan pada atribut TYPE, berisi prediksi kelas *Compliment* atau *Complain*. Namun karena ini adalah dataset yang ingin diklasifikasikan, maka Atribut TYPE merupakan klasifikasi manual sebagai acuan sistem untuk mengklasifikasikan *feedback* menggunakan *naive bayes classifier*.

text	datanya.TYPE
1 parkir motor rs	Complain
2 pelayanan staff maternity ramah melayani keluarga pasien	Complain
3 tenaga kebersihan kamar sup	Complain
4 perawatnya ramah	Complain
5 pelayanan gedung b memuaskan	Complain
6 motor masuk ugk lantai bawa anak ugk motor tidak masuk kamar kelas anaknak tolong dibedakan	Complain
7 fasilitas kamar mandi ditingkatkan	Complain
8 mohon diperbaiki layanan administrasi rawat inap diharapkan proses skrulingnya cepat dan pasien yg mempunyai kamar dipakai	Complain
9 fasilitas ac di ruang pendaftaran dan ruang tunggu	Complain
10 mohon ditingkatkan pelayanan rs sur administrasi perawat dokter bank pasien rawat inap sehingga penggantian infus di	Complain

Gambar 12. Halaman Tampilan Hasil Preprocessing

Term Frequency

Pada saat Tab Term Frequency diklik maka sistem melakukan proses pembobotan pada atribut COMMENTS dengan membagi -bagi kata dengan jumlah frekuensinya dan ditampilkan pada halaman ini dengan bentuk tabel.

word	freq
dan	51200
pelayanan	50780
terimakasih	27440
pasien	25970
ditingkatkan	24500
silom	20900
dokter	19190
bagus	18620
pelayananya	18150
lingkasan	15980
diperhatikan	14700
semoga	14700
perhatikan	14210
rumah	13720

Gambar 13. Halaman Tampilan Term Frequency

Pada tab ini, kata kata yang terdapat dalam *feedback* dihitung frekuensi kemunculannya dan dilakukan pembobotan kata.

Wordcloud

Pada tab ini, kata - kata yang sudah dilakukan pembobotan dan dihitung *term-frekuensinya* ditampilkan dalam bentuk *wordcloud*. Pada tampilan ini, kata yang memiliki nilai term tinggi terlihat lebih menonjol dibandingkan kata yang lainnya.



Gambar 14. Halaman Tampilan Word Cloud

Pada tampilan data menggunakan *wordcloud* didapatkan 5 kata paling menonjol, yaitu : baik, pelayanan, terimakasih, silom, ditingkatkan.

Naive bayes Classifier

Pada halaman ini data yang sudah diinput, dibersihkan, dan dilakukan *term frequency*, diklasifikasikan menggunakan metode naive bayes classifier. Data *feedback* dengan atribut *COMMENTS* terklasifikasi dalam kelas *Compliment* dan *Complain*. Hasil klasifikasi tersebut ditampilkan dalam bentuk tabel yang masing-masing *feedback*nya sudah terisi *TYPE*nya. Berikut merupakan tampilan hasil klasifikasi :

id	id datanya	TYPE	prediksi	text
1	1	Complain	Complain	partir motor jaganan terlalu jauh dari rs
2	2	Complain	Complain	untuk pelayanan staff di maternity untuk lebih ramah lagi dalam melayani keluarga pasien
3	3	Complain	Complain	tenaga kesehatan untuk kamar inap masih agak kurang
4	4	Complain	Complain	perawatnya sudah ramah lagi
5	5	Complain	Complain	pelayanan di gedung b kurang memuaskan
6	6	Complain	Complain	agar motor bisa masuk ugdi saya takutin bawa anak ke ugdi karena motor tdk bisa masuk kamar relas until anakanan toong diadkan
7	7	Complain	Complain	fasilitas kamar mandi yang perlu ditingkatkan
8	8	Complain	Complain	mohon dipertahankan layanan administrasi rawat inap bila terlahat ada yang ganggu supaya proses kepulangan bisa cepat dan pasien baru yg akan menempati kamar dapat juga segera dipaka
9	9	Complain	Complain	fasilitas ac di ruang pendaftaran dan ruang tunggu
10	10	Complain	Complain	mohon ditingkatkan lagi untuk pelayanan rs baik dari sisi administrasi perawat maupun dokter jangan biarkan pasien tetap menunggu ama untuk penggantian infus di

Gambar 15. Tampilan Klasifikasi

Hasil akurasi sistem

Model klasifikasi yang sudah dibuat digunakan sebagai acuan dalam klasifikasi *feedback* pada data uji. Berikut ini merupakan hasil akurasi klasifikasi menggunakan model data latih yang dihitung dari nilai *Appernya*. Pada data uji, hasil prediksi yang tidak sesuai adalah sebanyak 4 data.

$$APPER = (4 : 95) \times 100\% = 4.2 \%$$

Berdasarkan perhitungan diatas, tingkat kesalahan klasifikasi adalah 4.2 %, sehingga nilai ketepatan klasifikasi pada data latih adalah sebesar 95.8%.

Total observations in Table: 95			
predicted	actual		Row Total
	Complain	Compliment	
Complain	1	0	1
	1.000	0.000	0.011
Compliment	4	90	94
	0.043	0.957	0.989
Column Total	5	90	95
	0.053	0.947	

Gambar 16. Hasil akurasi data uji

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka didapat beberapa hasil penelitian, antara lain :

Pada tahap pembuatan model klasifikasi, dari 392 record dengan 2 kelas *Compliment* dan *Complain* didapatkan hasil akurasi yang cukup untuk membuat sistem yaitu 91.1 %. Hasil tersebut dinilai cukup karena lebih dari 80%. Untuk membuat model yang lebih baik, dibutuhkan data yang lebih berimbang antara kelas *Compliment* dan kelas *Complain* sehingga proses klasifikasi jadi lebih akurat.

Visualisasi menggunakan R shiny dengan menampilkan tabel data input, tabel data preprocessing, tabel term frequency, dan word cloud dari data *feedback* membuat data lebih menarik dan mudah dianalisis. Pada proses implementasi sistem, dengan data uji sebanyak 95 record didapatkan hasil klasifikasi sesuai yang diinginkan. hasil klasifikasi *feedback* dengan *TYPE Complain* adalah sebanyak 1 data, dari data aktual sebanyak 5 data dan pada *TYPE Compliment* sebanyak 90 data, dari data aktual sebanyak 90 data. Berdasarkan hal tersebut, tingkat akurasi model naive bayes classifier pada aplikasi analisis sentimen adalah sebanyak 95.8 % dengan jumlah kesalahan klasifikasi sebanyak 4 data dari total 95 data uji.

Penggunaan R shiny untuk visualisasi hasil klasifikasi analisis sentimen memudahkan dalam proses analisa hasil dan penarikan kesimpulan. Dari total data uji sebanyak 95 data, sebanyak 90 data *feedback*

menyatakan kepuasan terhadap pelayanan Rumah Sakit Umum Siloam Karawaci. Dari daftar kata dengan frekuensi kemunculan paling banyak diantaranya adalah : pelayanan, baik, terimakasih, siloam, ditingkatkan. Kata - kata tersebut merupakan kata yang berhubungan erat dengan kepuasan pasien. Dimana dapat disimpulkan bahwa sebagian besar pasien merasa puas dengan pelayanan yang baik dari siloam dan mengucapkan terimakasih dan berharap pelayanan yang lebih baik dapat ditingkatkan lagi.

Daftar Pustaka

- [1] Muktafin, E. H., Kusriani, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplora Informatika*, 10(1), 32–42. <https://doi.org/10.30864/eksplora.v10i1.390>
- [2] Sunni, I., & Widyantoro, D. H. (2012). Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro Dan Informatika*, 1(2), 200–206.
- [3] Permadi, V. A. (2020). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. *Jurnal Buana Informatika*, 11(2), 140. <https://doi.org/10.24002/jbi.v11i2.3769>
- [4] Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi*, 8(1), 958–965. <https://ejournal.unsri.ac.id/index.php/jsi/article/view/3250/1907>
- [5] Mahardhika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes. *Prosiding SINTAK 2018, 2015*, 409–413.
- [6] Feldman, R., and J. Sanger. 2007. *The Text Mining Handbook : Advanced Approaches in Analysing Unstructured Data*. New York : Cambridge University Press
- [7] Han, J., and M.Kamber. 2006. *Data Mining Concept and Techniques Second Edition*. San Frasisco : Elsevier.
- [8] Aggarwal, C., and C. Zhai. 2012. *Mining Text Data Chapter A Survey of Text Classification Algorithms*. London : Kluwer Academic Publisher.
- [9] Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. 7(1).
- [10] Reddy, V., S. R., Somayajulu, D. V. L. N., & Dani, A. R. (2011). Classification of Movie Reviews Using Complemented Naive Bayesian Classifier. *International Journal of Intelligent Computing Research*, 2(3), 148–153. <https://doi.org/10.20533/ijicr.2042.4655.2011.0019>
- [11] Setya Rintyarna, B. (n.d.). (2016). Sentiment Analysis pada Movie Review dengan Pendekatan Klasifikasi dalam Algoritma J.48. www.amazon.com
- [12] Muthia, D. A. (2017). Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 39–45. <https://doi.org/10.1515/HUMOR.2006.009>



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang berjudul “Aplikasi Analisis Sentimen Pada Feedback Pasien Menggunakan Metode Naïve Bayes Classifier (Studi Kasus : Rumah Sakit Umum Siloam)”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat / atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan beberapa bagian sebagai berikut :

1. Literature Review

Di dalam *literature review* ini ada beberapa hasil review jurnal yang terkait dengan penelitian yang dilakukan. Secara garis besar *literature review* yang dilakukan tentang, *machine learning*, *natural language processing*, analisis sentimen, *naive bayes classifier*, *R programming*, dan *R Shiny*.

2. Analisa dan Perancangan

Analisa dan perancangan berisi penjelasan beserta gambar dari tahapan – tahapan yang dilakukan dalam penelitian eksperimen. Pada tahapan ini dijelaskan bahwa penelitian dilakukan dengan mengolah dataset menjadi *data training* dan *data testing* , membuat model klasifikasi, menghitung akurasi model dan memvisualisasikan hasil analisa pada sistem aplikasi.

3. Source Code

Source code berisi bahasa pemrograman untuk mengolah dataset menggunakan bahasa pemrograman R pada Rstudio dan memanfaatkan R shiny sebagai framework aplikasi.

4. Dataset

Berisi penjelasan tentang asset yang digunakan dalam membuat Aplikasi analisis sentimen. Pada bagian ini dijelaskan struktur awal dataset, *treatment* yang dilakukan supaya dataset siap digunakan dan hasil dataset bersih.

5. Tahapan Eksperimen

Merupakan penjelasan tahapan-tahapan eksperimen pada penulisan ini yaitu dengan percobaan klasifikasi model pada *data training*.

6. Hasil Semua Eksperimen

Merupakan hasil yang ditampilkan pada aplikasi sesuai dengan metode eksperimen yang digunakan.

