

## Analisis Sentimen *Review* Pelanggan Pada Layanan Indihome Menggunakan Algoritma *Naïve Bayes Classifier*

Arganata Alif Fani<sup>1)</sup> Pantjawati Sudarmaningtyas<sup>2)</sup> Vivine Nurcahyawati<sup>3)</sup>

Program Studi/Jurusan Sistem Informasi

Universitas Dinamika

Jl. Raya Kedung Baruk 98 Surabaya, 60298

Email: 1)[15410100053@dinamika.ac.id](mailto:15410100053@dinamika.ac.id), 2)[pantja@dinamika.ac.id](mailto:pantja@dinamika.ac.id), 3)[vivine@dinamika.ac.id](mailto:vivine@dinamika.ac.id)

**Abstract:** PT. Telkom Indonesia as a customer-oriented company certainly must understand and know how important customer assessment of services, agencies or organizations. at PT. Telkom Indonesia has a problem that is Telkom Indonesia does not know about the opinions or customer reviews of IndiHome services on websites and social media. At present Telkom Indonesia only knows customer reviews or opinions through complaints directly or via Customer Service (CS) telephone. Based on the existing problems, the solution is obtained in the form of an application that can display customer reviews whether positive or negative. Customer reviews can be processed words / sentences using the text mining method and words / sentences can be classified as positive and negative classes using the naïve Bayes method. From the results of the sentiment analysis research it can be concluded that the sentiment analysis in this study can display a visualization of the results of the sentiment analysis classification in the form of pie charts and produce a sentiment analysis classification as evidenced by 100% function testing, testing calculations with different values with the same classification, and 100% accuracy testing with predictive value / actual value is True Negative (TN).

**keywords:** Sentiment Analysis, Telkom Indonesia, Naïve Bayes, Text Mining.

### 1. PENDAHULUAN

#### Latar Belakang Penelitian

PT. Telekomunikasi Indonesia (Persero) Tbk, (Telkom Indonesia) memiliki salah satu layanan yang paling banyak digunakan oleh pelanggan yaitu IndiHome. Dikutip dari *website* IndiHome (2019) IndiHome adalah layanan digital berupa internet rumah, telepon rumah dan TV interaktif (UseeTV) yang menggunakan teknologi fiber optik. Dikutip dari *website detik.com* (2019) tercatat ada 5 juta pelanggan hingga awal desember 2018.

Saat ini Telkom Indonesia hanya mengetahui *review* atau pendapat pelanggan melalui komplain secara langsung atau melalui telepon *Customer Service* (CS). Sedangkan, pada Telkom Indonesia terdapat permasalahan yaitu Telkom Indonesia belum mengetahui tentang pendapat atau *review* pelanggan terhadap layanan IndiHome pada media sosial. Menurut situs layanan manajemen konten (*content management*) *HootSuite*, data tren internet dan media sosial 2019 di Indonesia sebagai berikut: 1.) Total Populasi (jumlah penduduk): 268,2 juta (naik 1% dari tahun 2018), 2.) Pengguna *Mobile* (*Smartphone*): 355,5 juta (naik 50 % dari tahun 2018), 3.)

Pengguna Internet: 150 juta (naik 12% dari tahun 2018), 4.) Pengguna Media Sosial Aktif: 150 juta (naik 13% dari tahun 2018), 5.) Pengguna Media Sosial *Mobile*: 130 juta (naik 8% dari tahun 2018). Menurut *HootSuite* 4 platform media sosial yang digunakan di Indonesia dengan angka di atas 80% sebagai berikut: 1. *youtube*: 88% (naik 45% dari tahun 2018), 2. *whatsapp*: 83% (naik 43% dari tahun 2018), 3. *facebook*: 81% (naik 40% dari tahun 2018), 4. *instagram*: 80% (naik 42% dari tahun 2018).

Untuk dapat mengoptimalkan sumber data mengenai pendapat pelanggan, Telkom Indonesia bisa mendapatkannya melalui internet dan media sosial. Internet dan media sosial adalah media dimana pelanggan dapat mengungkapkan pendapatnya mengenai layanan dan instansi yang digunakan secara jujur. Berdasarkan latar belakang dan temuan hasil dari aktivitas observasi dan wawancara didapatkan bahwa dibutuhkan suatu aplikasi untuk dapat melakukan analisis sentimen pada *review* atau pendapat pelanggan yang tertulis dalam halaman-halaman baik di media sosial maupun *website* berita. Analisis sentimen yang dilakukan akan menghasilkan sebuah informasi tentang bagaimana *review* pelanggan terhadap

layanan perusahaan. Analisis sentimen dengan hasil klasifikasi *class* positif dan negatif pada *review* pelanggan akan menjadi sebuah informasi yang berguna bagi perusahaan khususnya sebagai bahan tindak lanjut untuk kebijakan peningkatan layanan pelanggan. Visualisasi dari hasil analisis sentimen akan dituangkan dalam bentuk grafik pada *website* sehingga pihak perusahaan dapat membaca hasil analisis sentimen yang telah dilakukan.

### **Rumusan Masalah Penelitian**

Berdasarkan permasalahan di atas maka dapat dirumuskan bahwa masalah yang ada yaitu, bagaimana melakukan analisis sentimen pada *review* pelanggan terhadap layanan IndiHome pada Telkom Indonesia.

### **Batasan Masalah Penelitian**

Batasan masalah yang digunakan dalam melakukan analisis ini adalah:

1. Data sekunder yang digunakan dalam penelitian ini adalah media sosial (Twitter) dan website (mediakonsumen.com dan detik.com).
2. Data simulasi menggunakan dataset pada awal bulan januari tahun 2017 sampai akhir bulan mei tahun 2019, kemudian data simulasi dibagi menjadi dua bagian yaitu data training dan data testing.
3. Kamus kata-kata positif, negatif dan netral menggunakan kamus Liu (2012).
4. Dalam melakukan text mining, penelitian ini hanya menggunakan tahapan preprocessing task pada text mining.

## **2. LANDASAN TEORI**

### **Analisis Sentimen**

Menurut Liu (2012), analisis sentimen, juga disebut penambangan opini, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Ini mewakili ruang masalah yang besar. Ada juga banyak nama dan tugas yang sedikit berbeda, misalnya, analisis sentimen, penambangan opini, ekstraksi pendapat, penambangan sentimen, analisis subjektivitas, analisis pengaruh, analisis emosi, penambangan *review*, dll.

### **Web crawler**

Menurut Shrivastava (2018), web crawler yang juga dikenal sebagai spider atau robot website adalah sebuah program yang secara otomatis melintasi sejumlah besar halaman website dengan mengikuti hyperlink, mengindeksnya dan menyimpan tautan halaman website yang dilalui untuk pengembangan penggunaan. Web crawler adalah program yang mengunduh laman website secara otomatis dari website raksasa dan besar. Bagian utama dari data yang ada di website adalah data tidak terstruktur yang tidak diatur dalam cara yang ditentukan dan tidak memiliki model data yang telah ditentukan. Banyak pencarian sedang dilakukan untuk merancang spider yang sempurna (web crawler) sehingga proses pencarian dan perayapan dapat ditingkatkan untuk sebagian besar. Mesin pencarian yang berbeda-beda menggunakan berbagai jenis web crawler misalnya Google menggunakan GoogleBot, web crawler Microsoft adalah Msnbot/Bingbot, mesin pencari Baidu menggunakan Baidu Spider.

### **Text mining**

Sedangkan, Menurut Feldman & Sanger (2007), text mining dapat didefinisikan secara luas sebagai proses intensif pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Dengan cara yang analog (serupa) dengan data mining, text mining berupaya mengekstraksi informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Dalam kasus text mining, bagaimanapun, sumber data adalah kumpulan dokumen, dan pola yang menarik yang tidak ditemukan di antara catatan basis data yang diformalkan tetapi dalam data tekstual yang tidak terstruktur dalam dokumen dalam koleksi ini. Langkah-langkah yang dilakukan dalam text mining adalah preprocessing task dan core mining operations:

### **Preprocessing task**

Menurut Feldman & Sanger (2007), preprocessing task adalah semua rutinitas, proses, dan metode yang diperlukan untuk menyiapkan data untuk operasi penemuan pengetahuan inti pada sistem text mining. Tugas-tugas ini biasanya berpusat pada kegiatan praproses (preprocessing) dan

kategorisasi sumber data (categorization activities). Tugas preprocessing umumnya mengubah informasi dari masing-masing sumber data asli ke dalam format kanonik (format dasar) sebelum menerapkan berbagai jenis metode ekstraksi fitur terhadap dokumen-dokumen ini untuk membuat kumpulan dokumen baru yang sepenuhnya diwakili oleh konsep. Jika memungkinkan, preprocessing dapat juga mengekstraksi atau menerapkan aturan untuk membuat prangko tanggal dokumen, atau melakukan keduanya. Kadang-kadang, tugas preprocessing bahkan dapat mencakup metode yang dirancang khusus yang digunakan dalam pengambilan awal data "mentah" yang sesuai dari sumber data asli yang berbeda.

**Case folding**

Menurut Langgeni (2010), case folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (dibatasi).

Tabel 1. Case Folding

Reply tweet @frikepriska pada tweet @TelkomCare pada tanggal 15 mei 2019	Hasil case folding
Indihome sangat tidak memuaskan !!!! Sangat mengecewakan !!!! Baru tgl 24 April pasang indihome, sejak tgl 4 mei sampai sekarang tgl 15 Mei masih tidak bisa digunakan !!	indihome sangat tidak memuaskan !!!! sangat mengecewakan !!!! baru tgl 24 april pasang indihome, sejak tgl 4 mei sampai sekarang tgl 15 mei masih tidak bisa digunakan !!

**Tokenizing**

Menurut Weiss , Indurkhya , Zhang , & Damerau (2005), tokenization adalah proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik(.), koma(,), spasi dan karakter angka yang ada pada kata.

Tabel 2. Tokenizing

Hasil case folding	Hasil tokenizing
indihome sangat tidak memuaskan !!!!	indihome
sangat mengecewakan !!!! baru tgl 24	sangat
april pasang indihome, sejak tgl 4 mei	tidak
sampai sekarang tgl 15 mei masih tidak	memuaskan
bisa digunakan !!	sangat
	mengecewakan
	baru
	tgl
	april
	pasang
	indihome
	sejak
	tgl
	mei
	sampai
	sekarang
	tgl
	mei
	masih
	tidak
	bisa
	digunakan

**Filtering**

Menurut Langgeni (2010), filtering adalah tahap mengambil kata penting dari hasil proses token. Bisa menggunakan algoritma stop list (kata yang tidak penting) atau word list (kata yang penting). Sedangkan, menurut Weiss , Indurkhya , Zhang , & Damerau (2009), stop list atau stop words atau stop word removal merupakan kosa kata yang bukan merupakan ciri (kata unik) dari suatu dokumen.

Tabel 3. Filtering

Hasil tokenizing	Hasil filtering
indihome	indihome
sangat	tidak
tidak	memuaskan
memuaskan	mengecewakan
sangat	baru
mengecewakan	tgl
baru	april
tgl	pasang
april	indihome
pasang	tgl
indihome	mei
sejak	tgl
tgl	mei
mei	tidak
sampai	bisa
sekarang	digunakan
tgl	
mei	
masih	
tidak	
bisa	
digunakan	

**Stemming**

Menurut Tala (2003), stemming adalah suatu proses yang menyediakan pemetaan variasi kata morfologis yang berbeda ke dalam kata dasar / umum (root). Sedangkan, menurut Gurusamy (2014), stemming adalah proses menggabungkan bentuk varian kata ke dalam representasi umum, bentuk batang. Misalnya, kata-kata: "mempresentasikan", "dipresentasikan", dan "berpresentasi" semua dapat direduksi menjadi representasi umum "presentasi". Ini adalah prosedur yang banyak digunakan dalam pemrosesan teks untuk Information Retrieval (IR) berdasarkan pada asumsi bahwa mengajukan pertanyaan dengan istilah penyajian menyiratkan minat pada dokumen yang berisi kata-kata presentasi dan disajikan.

Tabel 4. Stemming

Hasil filtering	Hasil stemming
indihome	indihome
tidak	tidak
memuaskan	puas
mengecewakan	kecewa
baru	baru
tgl	tanggal
april	april
pasang	pasang
indihome	indihome
tgl	tanggal
mei	mei
tgl	tanggal
mei	mei
tidak	tidak
bisa	bisa
digunakan	guna

**Naïve bayes**

Menurut Saleh (2015), naïve bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan naïve bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Berikut adalah rumus dari *naïve bayes*:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Dimana:

*X*: Data dengan *class* yang belum diketahui

*H*: Hipotesis data merupakan suatu *class* spesifik

*P(H|X)*: Probabilitas hipotesis *H* berdasar kondisi *X* (*posteriori* probabilitas)

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

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

*P(X)*: Probabilitas *X*

Rumus dari *naïve bayes* disesuaikan sesuai dengan penelitian sebagai berikut:

$$P(Wk|Ci) = \frac{n_k + 1}{n + |Vocabulary|} \quad (2)$$

Dimana:

*P(Wk)*: Probabilitas (peluang) kemunculan kata

*Ci*: Kategori kelas

*P(Wk|Ci)*: Probabilitas kemunculan kata (*Wk*) pada suatu dokumen dengan kategori kelas (*Ci*)

*n<sub>k</sub>*: nilai kemunculan kata pada kategori *c<sub>i</sub>*

*n*: Jumlah keseluruhan kata pada kategori *c<sub>i</sub>*

*Vocabulary*: Jumlah keseluruhan kata

**Visualisasi**

Menurut Kosara (2007), kriteria berikut ini sebagai sekumpulan persyaratan minimal untuk visualisasi apa pun. Sisa ini akan berkonsentrasi pada visualisasi informasi, tetapi

kriteria ini berlaku sama untuk visualisasi ilmiah.

**Software Development Life Cycle (SDLC)**

Menurut Sukamto dan Shalahuddin (2013) mengemukakan bahwa Software Development Life Cycle (SDLC) atau sering disebut juga System Development Life Cycle adalah proses mengembangkan atau mengubah suatu sistem perangkat lunak dengan menggunakan model-model dan metodologi yang digunakan orang untuk mengembangkan sistem-sistem perangkat lunak sebelumnya, berdasarkan best practice atau cara-cara yang sudah teruji baik. Model waterfall sering juga disebut model sekuensi linear atau alur hidup klasik. Pengembangan sistem dikerjakan secara terurut mulai dari analisis, desain, pengkodean, dan pengujian.

**Confusion matrix**

Menurut Manning, Raghavan, & Schütze (2009) confusion matrix atau disebut juga matriks klasifikasi adalah suatu alat visual yang biasanya digunakan dalam supervised learning. Dikutip dari website Toward Data Science (2019) confusion matrix adalah ringkasan hasil prediksi dalam masalah klasifikasi. Prediksi yang benar dan salah dirangkum dalam sebuah tabel dengan nilainya dan dipecah oleh masing-masing kelas.

		Predictive Value	
		Positive (1)	Negative (0)
Actual values	Positive (1)	TP	FN
	Negative (0)	FP	TN

Gambar 1. Confusion Matrix

Dimana:

*True Positive (TP)*: Ketika prediksi positif dan ternyata hasilnya benar positif.

*True Negative (TN)*: Ketika prediksi negatif dan ternyata hasilnya benar negatif.

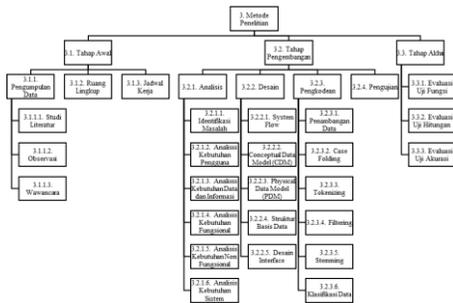
*False Positive (FP)*: Ketika prediksi positif dan ternyata hasilnya salah.

*False Negative (FN)*: ketika prediksi negatif dan hasilnya ternyata salah.

**3. METODE PENELITIAN**

Bab ketiga yang ditulis pada penelitian ini adalah metodologi penelitian. Pada bab ini menjelaskan tentang tahapan-tahapan yang digunakan untuk menyelesaikan penelitian. Menurut wahono (2019) tahapannya terdiri atas

tahap awal, tahap pengembangan dan tahap akhir.



Gambar 2. Metode Penelitian

**Studi Literatur**

Pada tahap awal metode penelitian, hal yang pertama kali dilakukan adalah melakukan studi literatur. Pada studi literatur, yang dilakukan adalah *review* terhadap penelitian-penelitian sebelumnya dalam bentuk Tugas Akhir (TA), Jurnal, dsb. Setelah melakukan *review* pada penelitian yang sudah ada, maka pada penelitian ini dapat ditentukan studi kasus dan tahapan-tahapan penyelesaian yang akan digunakan saat melakukan penelitian.

**Observasi**

Pada pengumpulan data, yang dilakukan pertama kali adalah melakukan observasi. Observasi dilakukan untuk mengamati objek yang digunakan sebagai sumber data penelitian. Objek yang diamati adalah *tweet* atau *reply* pada *twitter*, pelaporan pada *detik.com* dan postingan pada *mediakonsumen.com*. Pada beberapa objek yang diamati terdapat berbagai macam konteksnya, seperti macam-macam opini netizen atau warganet pada *twitter* atau *instagram* dan macam-macam postingan dari *website*.

**Wawancara**

Pada pengumpulan data, yang dilakukan selanjutnya adalah melakukan wawancara. Wawancara dilakukan dengan pegawai PT. Telkom Indonesia yang ada pada divisi *officer digital consumer*. Pelaksanaan wawancara berlokasi di PT. Telkom Indonesia Divisi Regional V Surabaya pada tanggal 02/07/2019. Pada wawancara ditemukan hasil bahwa Telkom Indonesia belum memiliki divisi atau aplikasi untuk menganalisis tentang *review* pelanggan terhadap IndiHome seperti apa saja

laporan yang sering muncul dari pelanggan dan bagaimana performa Telkom Indonesia dalam menangani laporan pelanggan.

**Ruang Lingkup**

Bagian ini menjelaskan tentang lingkup penelitian yang akan dibahas dan yang tidak dibahas. Penelitian ini membahas tentang analisis sentimen mengenai *review* pelanggan dengan konotasi positif dan negatif yang ada di media sosial dan *website*. Penelitian ini tidak membahas tentang langkah yang harus diambil perusahaan dalam mengembangkan bisnis setelah mendapatkan hasil dari analisis sentimen.

**Ruang Lingkup**

Pada bagian ini menjelaskan bagaimana melakukan penjadwalan pengerjaan penelitian dari tahap awal, tahap pengembangan, dan tahap akhir.

Tabel 5. Ruang Lingkup

No	Kegiatan	Bulan Ke-1				Bulan Ke-2				Bulan Ke-3				Bulan Ke-4				Bulan Ke-5			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Tahap Awal: Pengumpulan data (Studi Literatur, Observasi dan Wawancara). Ruang Lingkup dan Jadwal Kerja.	■	■	■	■																
	Tahap Pengembangan: Analisis, Desain, Pengkodean, dan Pengujian.					■	■	■	■	■	■	■	■	■	■	■	■				
3.	Tahap Akhir: Evaluasi Uji Fungsi, Evaluasi Uji Kinerja, dan Evaluasi Uji Akurasi.																	■	■	■	■

**Identifikasi Masalah**

Tahap identifikasi masalah digunakan sebagai dasar dalam melakukan penelitian. Identifikasi masalah yang dilakukan adalah dengan mempelajari studi literatur yang serupa dengan penelitian yang dikerjakan, melakukan observasi pada media sosial dan *website*, dan melakukan wawancara dengan pihak Telkom Indonesia sehingga didapatkan informasi yang berkaitan dengan penelitian.

Berdasarkan Studi Literatur, Observasi, dan Wawancara yang sudah dilakukan, Telkom Indonesia melakukan rapat tahunan untuk mengetahui seberapa besar *review* atau pendapat pada produk yang digunakan oleh pelanggan. Sedangkan dalam mengatasi *review* atau pendapat pelanggan, Telkom Indonesia secara langsung melalui *komplain* secara langsung atau melalui telepon

*Customer Service* (CS) dan media online *twitter*.

Dari masalah tersebut, Telkom Indonesia yang berada pada Divisi Regional V Jawa Timur membutuhkan sebuah aplikasi yang dapat membantu untuk mengetahui *review* atau pendapat pelanggan terhadap layanan IndiHome.

### **Analisis Kebutuhan Pengguna**

Analisis Kebutuhan Pengguna dilakukan dengan tujuan untuk mengetahui kebutuhan pengguna yang digunakan dan/atau dibutuhkan. Berdasarkan hasil Wawancara dan Identifikasi Masalah maka dapat dilakukan analisis kebutuhan pengguna untuk penelitian yang akan dibuat, pengguna tersebut dapat diidentifikasi, yaitu *Consumer Marketing*. *Consumer Marketing* bertugas untuk memberikan informasi mengenai pengguna yang menggunakan IndiHome.

### **Analisis Kebutuhan Data dan Informasi**

Analisis kebutuhan Data dan Informasi dilakukan dengan tujuan untuk mengetahui data dan informasi yang digunakan dan/atau dibutuhkan oleh pengguna sistem (perangkat lunak) yang akan dibuat. Selain itu juga untuk menganalisis *output* yang diperoleh dari pengguna tersebut.

### **Analisis Kebutuhan Fungsional**

Berikut adalah kebutuhan-kebutuhan fungsional yang nantinya ada di dalam aplikasi yang dibagi kedalam beberapa kategori.

### **Analisis Kebutuhan Non Fungsional**

Berikut adalah kebutuhan-kebutuhan non fungsional yang nantinya ada di dalam aplikasi yang dibagi kedalam beberapa kategori.

### **Analisis Kebutuhan Sistem**

#### **a. Kebutuhan Perangkat Lunak**

Kebutuhan perangkat lunak atau *software requirement* adalah komponen perangkat lunak yang digunakan untuk mengembangkan dan membangun aplikasi pada penelitian ini. Adapun perangkat lunak yang digunakan sebagai berikut:

1. Sistem operasi yang digunakan pada penelitian ini menggunakan sistem operasi *windows 10* karena performa yang lebih cepat daripada sistem operasi *windows*

yang lainnya dan lebih familier untuk digunakan.

2. *XAMPP* digunakan untuk membuat *web server localhost* pada penelitian ini karena penggunaan untuk menyimpan database pada folder lebih mudah.
3. Bahasa pemrograman yang digunakan pada penelitian ini yaitu *Hypertext Preprocessor* (PHP) karena PHP banyak digunakan untuk membuat aplikasi berbasis *website* yang membutuhkan fungsionalitas maksimal dengan minimal pengkodean.
4. Basis data yang digunakan pada penelitian ini adalah *My Structured Query Language* (MYSQL) karena basis data (*database*) ini bersifat *Open Source* sehingga dapat digunakan oleh siapapun tanpa dibebani biaya lisensi yang lumayan tinggi.
5. Alat pengkodean yang digunakan pada penelitian ini adalah *Visual Code* karena tidak dibebankan biaya.

#### **b. Kebutuhan Perangkat Keras**

Berdasarkan analisis kebutuhan perangkat lunak yang telah dibuat maka dapat dilanjutkan dengan analisis kebutuhan perangkat keras atau *hardware*. Kebutuhan perangkat keras merupakan komponen peralatan fisik yang membentuk sistem komputer terstruktur, serta perangkat keras lain yang mendukung komputer dalam menjalankan fungsinya. Kebutuhan perangkat keras yang dibutuhkan adalah sebagai berikut:

1. Pada penelitian ini menggunakan *Central Processor Unit* (CPU) *intel core i5-8250U*.
2. Pada penelitian ini menggunakan *Random Access Memory* (RAM) sebesar 8GB.
3. Pada penelitian ini menggunakan *Monitor* dengan resolusi 1920x1080 *pixel*.
4. Pada penelitian ini menggunakan *Graphic Processing Unit* (GPU) *NVIDIA GeForce MX150*.

### **Desain**

Tahapan selanjutnya yang dilakukan pada Tahap Pengembangan yaitu Desain. Tahap Desain bertujuan untuk mengetahui desain apa saja yang dibutuhkan pada penelitian ini. Desain yang dibutuhkan pada penelitian ini adalah *System Flow*, *Conceptual Data Model* (CDM), *Physical Data Model* (PDM), Struktur Basis Data (Desain *Database*), Desain *interface*.

**Pengkodean**

Tahapan selanjutnya yang dilakukan pada tahap pengembangan yaitu pengkodean. Tahap Pengkodean bertujuan untuk menjalankan proses pengkodean pada tahap yang dilakukan pada penelitian ini. Pengkodean yang dilakukan pada penelitian ini adalah pengkodean Penambangan Data, Ekstrasi Data, dan Klasifikasi Data.

**Pengujian**

Tahapan selanjutnya yang dilakukan pada tahap pengembangan yaitu pengujian. Tahap pengujian bertujuan untuk melakukan uji coba pada fungsi yang digunakan pada penelitian ini. pengujian yang dilakukan pada penelitian ini adalah pengujian fungsi penambangan data, ekstrasi data, dan klasifikasi data. Sumber data yang digunakan pada penelitian ini menggunakan *mediakonsumen.com*.

**Tahap Akhir**

Bagian ini menjelaskan tentang Tahap Akhir untuk menyelesaikan penelitian ini. Tahap Akhir meliputi 3 (Tiga) Subbab, yaitu Evaluasi Uji Fungsi, Evaluasi Uji Hitungan, dan Evaluasi Uji Akurasi.

**4. HASIL DAN PEMBAHASAN**

Bab keempat yang ditulis pada penelitian ini adalah hasil dan pembahasan. Pada bab ini akan dibahas hasil dan implementasi dari penelitian berdasarkan tahapan-tahapan yang sudah dilakukan sesuai dengan metode yang digunakan. Hasil dan implementasi dibagi menjadi 3 (tiga) subbab, yaitu: Implementasi, Pengujian, dan Evaluasi.

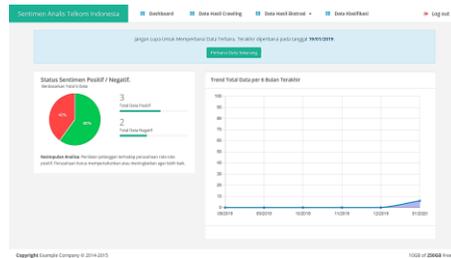
**Implementasi**

Bagian ini menjelaskan tentang implementasi yang dilakukan pada penelitian ini. Implementasi berisi tampilan hasil rancangan aplikasi yang disesuaikan dengan Kebutuhan Fungsional dan Kebutuhan Non Fungsional. Berikut adalah implementasi pada tampilan hasil rancangan aplikasi.

**Halaman Utama**

Setelah melakukan *login*, *Digital Consumer* dapat masuk ke dalam halaman utama dan melihat hasil visualisasi. Pada halaman utama menampilkan visualisasi dari hasil melakukan analisis sentimen sesuai

dengan tahapan dan metode yang digunakan. Halaman utama berisi data perbulan dalam bentuk grafik garis (*line chart*) dan kelas analisis sentimen positif atau negatif dalam bentuk grafik lingkaran (*pie chart*).



Gambar 3. Halaman Utama

**Halaman Hasil Data Crawling**

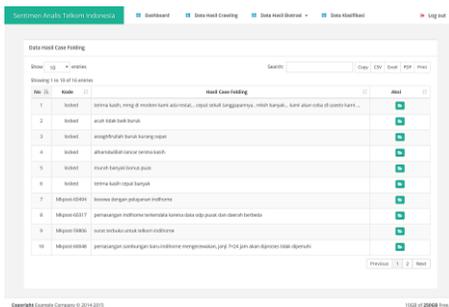
Halaman Hasil *Crawling* Data menampilkan hasil dari penambangan data. halaman ini berisi nama *author*, tanggal *posting*, *link*, sumber, imputan dan aksi. Pada saat menekan tombol aksi, halaman *pop up* detail dari hasil dari penambangan data akan ditampilkan.

No. ID	Author	Tanggal Posting	Link	Sumber	Imputan
1	Iskand	2020-01-01	Iskand	Iskand	terima kasih yang di berikan kepada saya, saya akan sangat berterima kasih, semoga sukses. Love dan love di semua sisi.
2	Iskand	2020-01-01	Iskand	Iskand	Acara akan lebih seru
3	Iskand	2020-01-01	Iskand	Iskand	Acara yang akan lebih seru
4	Iskand	2020-01-01	Iskand	Iskand	akan sangat seru dan seru
5	Iskand	2020-01-01	Iskand	Iskand	Acara yang akan lebih seru
6	Iskand	20-01-02	Iskand	Iskand	terima kasih yang di berikan
7	Hafid Jihan	20 Januari 2020	mediakonsumen.com	mediakonsumen.com	Seneca dengan Penerimaan Indonesia
8	Alik	24 Januari 2020	mediakonsumen.com	mediakonsumen.com	Peningkatan Indeks Kinerja Terlembaga karena Dampak Puncak dan Dasar Negara
9	Denny Pradiha	23 Januari 2020	mediakonsumen.com	mediakonsumen.com	Sangat Terlembaga untuk Terlembaga Indonesia
10	Rahm	21 Januari 2020	mediakonsumen.com	mediakonsumen.com	Peningkatan Sentimen dengan Indeks Kinerja Terlembaga, yang Peningkatan akan Peningkatan Terlembaga

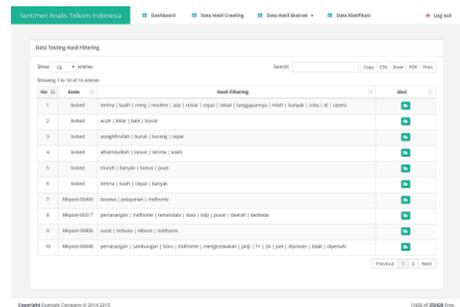
Gambar 4. Halaman Hasil Data Crawling

**Halaman Hasil Ekstrasi Data (Case Folding)**

Halaman hasil ekstrasi data (*case folding*) menampilkan hasil dari *case folding* dimana data inputan diubah menjadi *lowercase*. Halaman hasil ekstrasi data (*case folding*) berisikan kode, hasil *case folding*, dan aksi. Pada saat menekan tombol aksi, maka akan keluar halaman *pop up* detail dari hasil dari ekstrasi data (*case folding*).



Gambar 5. Halaman hasil ekstrasi data (case folding)



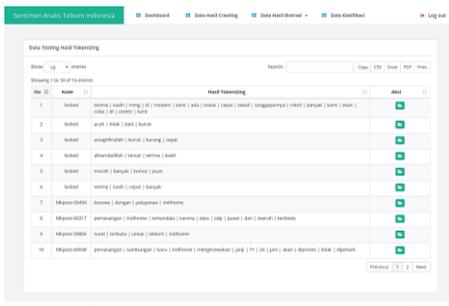
Gambar 7. Halaman hasil ekstrasi data (filtering)

**Halaman Hasil Ekstrasi (Tokenizing)**

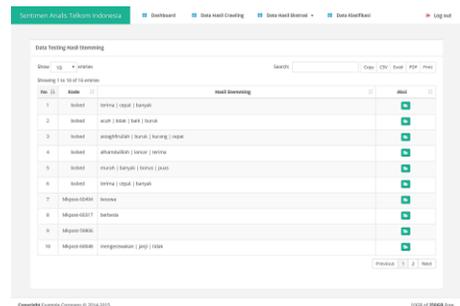
Halaman hasil ekstrasi data (*tokenizing*) menampilkan hasil dari *tokenizing* dimana data inputan dipotong-potong. Halaman hasil ekstrasi data (*tokenizing*) berisikan kode, hasil *tokenizing*, dan aksi. Pada saat menekan tombol aksi, maka akan keluar halaman *pop up* detail dari hasil dari ekstrasi data (*tokenizing*).

**Halaman Hasil Ekstrasi (Stemming)**

Halaman hasil ekstrasi data (*stemming*) menampilkan hasil dari *stemming* dimana data inputan dikembalikan ke kata dasar. Halaman hasil ekstrasi data (*stemming*) berisikan kode, hasil *stemming*, dan aksi. Pada saat menekan tombol aksi, maka akan keluar halaman *pop up* detail dari hasil dari ekstrasi data (*stemming*).



Gambar 6. Halaman hasil ekstrasi data (tokenizing)



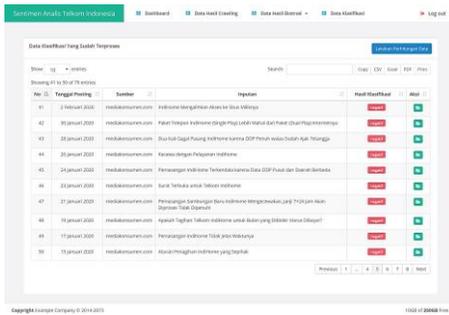
Gambar 8. Halaman hasil ekstrasi data (stemming)

**Halaman Hasil Ekstrasi (Filtering)**

Halaman hasil ekstrasi data (*filtering*) menampilkan hasil dari *filtering* dimana data inputan dihilangkan kata-kata imbuhan dsb. Halaman hasil ekstrasi data (*filtering*) berisikan kode, hasil *filtering*, dan aksi. Pada saat menekan tombol aksi, maka akan keluar halaman *pop up* detail dari hasil dari ekstrasi data (*filtering*).

**Halaman Hasil Klasifikasi Data**

Halaman hasil klasifikasi data menampilkan hasil dari klasifikasi data. Halaman hasil klasifikasi berisikan tanggal posting, sumber, inputan, hasil klasifikasi, dan aksi. Pada halaman hasil klasifikasi data terdapat tombol 'lakukan perhitungan data' untuk mengubah halaman hasil klasifikasi data menjadi halaman perhitungan klasifikasi data.



Gambar 9. Halaman klasifikasi data

**Pengujian**

Bagian ini menjelaskan tentang pengujian yang dilakukan pada penelitian. Pengujian yang dilakukan terdiri dari pengujian fungsi, pengujian perhitungan, dan pengujian akurasi. Pengujian menggunakan sumber yang sama, yaitu *mediakonsumen.com*, *twitter.com*, dan *detik.com*. data yang digunakan pada pengujian dengan sumber *mediakonsumen.com* adalah ‘Pemasangan Indihome Tidak Jelas Waktunya’ dengan link ‘<https://mediakonsumen.com/2020/01/17/surat-pembaca/pemasangan-indihome-tidak-jelas-waktunya>’, sedangkan pada *twitter.com* menggunakan data ‘ni knapa yak... tiba” mati @IndiHome’ dengan link ‘<https://twitter.com/IndiHome/status/1224542985477509120>’ dan *detik.com* menggunakan data ‘Tanggapan Lama, Kecewa Layanan Customer Service Indihome’ dengan link ‘<https://news.detik.com/suara-pembaca/d-4860295/tanggapan-lama-kecewa-layanan-customer-service-indihome>’. Berikut adalah hasil pengujian yang ada pada penelitian ini.

**Pengujian Fungsi**

Bagian ini menjelaskan tentang pengujian fungsi yang digunakan pada penelitian ini. Pengujian fungsi yang dilakukan terdiri dari 4 (empat) pengujian, yaitu Pengujian Fungsi Mengambil Data/*Crawling* Data, Pengujian Fungsi Mengekstrasi Data, Pengujian Fungsi Mengklasifikasi Data, dan Pengujian Fungsi Memvisualisasikan Data.

Tabel 6. Pengujian Fungsi

No.	Fungsi	Skenario	Hasil Yang Diharapkan	Hasil Akhir
3.	Mengklasifikasi Data.	Menekan tombol ‘klasifikasi data’ untuk melakukan proses klasifikasi data.	Aplikasi dapat melakukan klasifikasi data dari hasil ekstrasi data dengan sumber <i>mediakonsumen.com</i> dan dapat menyimpan data.	Data hasil ekstrasi data berhasil diklasifikasi dan disimpan. Bukti ada pada gambar 4.16. Halaman Hasil Klasifikasi Data.

**Pengujian Perhitungan**

Bagian ini menjelaskan tentang pengujian perhitungan yang digunakan pada penelitian ini. Pengujian perhitungan yang dilakukan terdiri dari 2 (dua) pengujian. Pengujian perhitungan yang dilakukan adalah pengujian perhitungan aplikasi dan pengujian perhitungan manual.

**Pengujian Perhitungan Aplikasi**

Bagian ini menjelaskan tentang pengujian perhitungan pada aplikasi. proses perhitungan pada data *training* positif dan negatif, hasil perhitungan data *testing* ditampilkan pada halaman hasil perhitungan *testing*. Hasil perhitungan *testing* berisikan hasil *stemming*, nilai positif, nilai negatif, dan kesimpulan.



Gambar 10. Pengujian perhitungan aplikasi

**Pengujian Perhitungan Manual**

Bagian ini menjelaskan tentang pengujian perhitungan secara manual. Setelah data *training* positif dan data *training* negatif dihitung, dilakukan perhitungan pada tiap data *testing*. data *testing* yang digunakan menggunakan sumber yang sama pada perhitungan aplikasi. Berikut adalah proses perhitungan data *testing*.

**Dokumen 41** = ‘Pemasangan Indihome Tidak Jelas Waktunya’ dengan hasil *stemming* = ‘tidak | jelas’

Perhitungan berdasarkan hasil probabilitas positif dengan dokumen uji

$$P(\text{positif}) = P(\text{positif}) + P(\text{tidak} | \text{positif}) + P(\text{jelas} | \text{positif})$$

$$= 0,075 + 0,0238095238095238 + 0,0238095238095238$$

$$= 0,1226190476190476$$

Perhitungan berdasarkan hasil probabilitas negatif dengan dokumen uji

$$\begin{aligned}
 P(\text{negatif}) &= P(\text{negatif}) + P(\text{tidak} | \text{negatif}) + P(\text{jelas} | \text{negatif}) \\
 &= 0,925 + 0,1038961038961039 + 0,012987012987013 \\
 &= 1,041883116883117
 \end{aligned}$$

Dari hasil perhitungan di atas dapat diketahui bahwa dokumen 41 ada dalam kelas negatif, karena nilai perhitungan dokumen uji pada kelas negatif lebih besar daripada nilai perhitungan dokumen uji pada kelas positif.

**Pengujian Akurasi**

Bagian ini menjelaskan tentang pengujian akurasi yang digunakan pada penelitian ini. pengujian akurasi pada penelitian ini menggunakan metode *confusion matrix*. Pengujian akurasi menggunakan hasil dari perhitungan aplikasi dan perhitungan manual.

Tabel 7. Pengujian akurasi

No. Dokumen	Teks Data testing	Hasil Perhitungan Klasifikasi Aplikasi	Hasil Perhitungan klasifikasi Manual	Actual Value / Predictive Value
41	Pemasangan Indihome Tidak Jelas Waktunya	Negatif	Negatif	TN (True Negative)
42	ni knapa yak... tiba" mati @IndiHome	Negatif	Negatif	TN (True Negative)
43	Tanggapan Lama, Kecewa Layanan Customer Service IndiHome	Negatif	Negatif	TN (True Negative)

Setelah dibuat tabel data pengujian akurasi data *testing*, dilakukan perhitungan akurasi dengan menggunakan rumus *confusion matrix*. Berikut adalah proses perhitungan akurasi.

$$\begin{aligned}
 Akurasi &= \frac{TP + TN}{TP + TN + FP + FN} \\
 Akurasi &= \frac{0 + 3}{0 + 3 + 0 + 0} \\
 Akurasi &= 1 \times 100\% = 100\%
 \end{aligned}$$

**Evaluasi**

Bagian ini menjelaskan tentang evaluasi yang dilakukan pada penelitian. Evaluasi yang dilakukan terdiri dari Evaluasi Hasil Uji Fungsi, Evaluasi Uji Perhitungan, dan Evaluasi Uji Akurasi. Berikut adalah hasil pengujian yang ada pada penelitian.

**Evaluasi Hasil Uji Fungsi**

Bagian ini membahas mengenai evaluasi berdasarkan pada pengujian fungsi yang sudah dilakukan. Pengujian fungsi yang dilakukan berupa pengujian fungsi

penambahan data, ekstrasi data, klasifikasi data, dan visualisasi data. Pengujian pada penambahan data dilakukan sebanyak 3 (tiga) kali, pengujian pada ekstrasi data dilakukan sebanyak 3 (tiga) kali, pengujian pada klasifikasi data dilakukan sebanyak 3 (tiga) kali, dan pengujian pada visualisasi data dilakukan sebanyak 1 (satu) kali. Dari 3 pengujian penambahan data, 3 pengujian berhasil dilakukan dengan tingkat keberhasilan 100%. Dari 3 pengujian ekstrasi data, 3 pengujian berhasil dilakukan dengan tingkat keberhasilan 100%. Dari 3 pengujian klasifikasi data, 3 pengujian berhasil dilakukan dengan tingkat keberhasilan 100%. Dari 1 pengujian visualisasi data, 1 pengujian berhasil dilakukan dengan tingkat keberhasilan 100%. Tingkat keberhasilan 100% berdasarkan pada jumlah pengujian yang dilakukan dibagi dengan jumlah keberhasilan pengujian.

**Evaluasi Hasil Uji Hitungan**

Bagian ini membahas mengenai evaluasi berdasarkan pada pengujian perhitungan yang sudah dilakukan. Pengujian perhitungan yang dilakukan berupa pengujian perhitungan aplikasi dan pengujian perhitungan manual. Berdasarkan pengujian perhitungan aplikasi dan pengujian perhitungan manual didapatkan hasil berupa nilai yang berbeda dengan nilai yang sama.

Tabel 8. Evaluasi hasil uji hitungan

sumber	Data	Perhitungan aplikasi		Perhitungan manual	
		positif	negatif	positif	negatif
<a href="https://mediakonsumen.com/2020/01/17/surat-pembaca-pemasangan-indihome-tidak-jelas-waktunya">https://mediakonsumen.com/2020/01/17/surat-pembaca-pemasangan-indihome-tidak-jelas-waktunya</a>	Pemasangan Indihome Tidak Jelas Waktunya	0.111 14457 83132 53	1.0197 368421 052633	0.122 61904 76190 476	1.0418 831168 83117
<a href="https://twitter.com/IndiHome/status/1224542985477509120">https://twitter.com/IndiHome/status/1224542985477509120</a>	ni knapa yak... tiba" mati @IndiHome	0.087 04819 27710 8433	0.9671 052631 578948	0.098 80952 38095 238	0.9769 480519 480519
<a href="https://news.detik.com/suara-pembaca/d-4860295/tanggapan-lama-kecewa-layanan-customer-service-indihome">https://news.detik.com/suara-pembaca/d-4860295/tanggapan-lama-kecewa-layanan-customer-service-indihome</a>	Tanggapan Lama, Kecewa Layanan Customer Service IndiHome	0.099 09638 55421 6867	0.9968 421052 63158	0.122 61904 76190 476	1.0159 090909 09091

**Evaluasi Hasil Uji Akurasi**

Bagian ini membahas mengenai evaluasi berdasarkan pada pengujian perhitungan akurasi yang sudah dilakukan. Pengujian perhitungan akurasi yang dilakukan berupa pengujian perhitungan menggunakan metode *confusion matrix*. Pengujian perhitungan akurasi menggunakan sumber data

*mediakonsumen.com*, *twitter.com*, dan *detik.com*. Dari tabel data pengujian akurasi data *testing* didapatkan bahwa didapatkan *Actual Value / Predictive Value* berupa 3 *True Negative* (TN). *Actual Value / Predictive Value* didapatkan dari hasil perhitungan aplikasi dan perhitungan manual. Berdasarkan perhitungan dengan uji akurasi didapatkan bahwa hasil akurasi perhitungan analisis sentimen yang didapatkan sebesar 100%.

## 5. PENUTUP

### KESIMPULAN

Berdasarkan dari hasil dan pembahasan yang ada pada penelitian ini, maka didapatkan kesimpulan sebagai berikut:

1. Analisis sentimen pada penelitian ini dapat menampilkan visualisasi dari hasil klasifikasi analisis sentimen dalam bentuk diagram lingkaran (pie chart).
2. Menghasilkan klasifikasi analisis sentimen yang dibuktikan dengan pengujian fungsi sebesar 100%, pengujian perhitungan dengan nilai yang berbeda dengan klasifikasi yang sama, dan pengujian akurasi sebesar 100% dengan predictive value/actual value adalah True Negative (TN).

### SARAN

Adapun saran-saran yang dapat menjadi masukan untuk penelitian-penelitian yang akan datang antara lain adalah:

1. Menambah data yang dilakukan pengujian untuk penelitian yang akan datang.
2. Mengembangkan dan menambah penggunaan data *training* dan data *testing* dari sumber yang berbeda untuk penelitian yang akan datang.
3. Mengembangkan algoritma atau metode lain untuk melakukan pengolahan data pada penelitian yang akan datang.

### DAFTAR PUSTAKA

- Assauri, MBA, P. S. (2012). *Strategic Marketing: Sustaining Lifetime Customer Value*. Depok: PT. RAJAGRAFINDO PERSADA.
- Azhar, S. (2004). *Sistem Informasi Manajemen*. Bandung: Lingga Jaya.
- Baj-Rogowska, A. (2017). *Sentiment Analysis of Facebook Posts: theUbercase*. The 8th IEEE International Conference on

Intelligent Computing and Information Systems (ICICIS 2017), 1.

- Dragut, E., Fang, F., Sistla, P., & Yu, C. (2009). *Proceedings of the VLDB Endowment* 2(1):349-360. Stop Word and Related Problems in Web Interface Integration.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.
- Gurusamy, V. (2014). *Preprocessing Techniques for Text Mining*, 4.
- Indonesia, P. T. (2019, 4 29). Apa Itu IndiHome? Retrieved from IndiHome: <https://indihome.co.id/pusat-bantuan/kenali-indihome/apa-itu-indihome>
- Indonesia, P. T. (2019, 4 29). Profil dan Riwayat Singkat. Retrieved from Telkom Indonesia: [https://www.telkom.co.id/servlet/tk/about/id\\_ID/stocklanding/profil-dan-riwayat-singkat.html](https://www.telkom.co.id/servlet/tk/about/id_ID/stocklanding/profil-dan-riwayat-singkat.html)
- Kadhim, A. I., Cheah, Y.-N., & Ahamed, N. H. (2014). 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology. Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering, 2.
- Kosara, R. (2007). 2007 11th International Conference Information Visualization (IV '07). Visualization Criticism - The Missing Link Between Information Visualization and Art, 2.
- Langgeni, D. P. (2010). Seminar Nasional Informatika 2010 (semnasIF 2010) ISSN: 1979-2328 . CLUSTERING ARTIKEL BERITA BERBAHASA INDONESIA MENGGUNAKAN UNSUPERVISED FEATURE SELECTION , 2.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. In B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction To Information Retrieval*. Cambridge: Cambridge University Press.
- Rahman, A. F. (2019, 5 6). Akhir 2018, IndiHome Tembus 5 Juta Pelanggan. Retrieved from [inet.detik.com](http://inet.detik.com): <https://inet.detik.com/telecommunicat>

- ion/d-4340753/akhir-2018-indihome-tembus-5-juta-pelanggan
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes. *Citec Journal*, Vol. 2, No. 3, Mei 2015 – Juli 2015, 4.
- Saragih, M. H. (2017). Sentiment Analysis of Customer Engagement on Social Media in Transport Online. 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2.
- Septiani, I., & Kasmirudin. (2014). *Jurnal Online Mahasiswa (JOM) Bidang Ilmu Sosial dan Ilmu Politik* Vol 1, No 1. PENGARUH FAKTOR RESPON KONSUMEN TERHADAP CITRA BISNIS JASA KARAOKE KELUARGA INUL VIZTA (Studi Evaluasi Pemasaran PT. Vizta Pratama Pekanbaru), 6.
- Sharma, P. (2019, Juli 27). Decoding the Confusion Matrix. Retrieved from *Toward Data Science*: <https://towardsdatascience.com/decoding-the-confusion-matrix-bb4801decbb>
- Shrivastava, V. (2018). *ISSN : 2248-9622* Vol. 8, Issue 11 (Part -I) Nov 2018, pp 01-08 . A Methodical Study of Web Crawler , 1.
- Srividhya, V. (2010). *International Journal of Computer Science and Application* Issue 2010 ISSN 0974-0767. Evaluating PreprocessingTechniques in Text Categorization , 2.
- Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia, 7.
- Wahono, R. S. (2019). *Research Methodology*.
- Watson, K. M. (2013). The Impact Of Natural Language Processing-Based. *ECIS 2013 Completed Research*, 4.
- Weiss , S. M., Indurkha , N., Zhang , T., & Damerau, F. J. (2005). *Text Mining : Predictive Methods for Analyzing Unstructured Information*. New York: Springer.