

Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode *Lexicon Based* dan *Naïve Bayes Classifier*

Aqwam Rosadi Kardian dan Dede Gustiana

Manajemen Informatika, STMIK Jakarta STI&K
Jl. BRI No. 17, Radio Dalam, Keb. Baru, Jakarta Selatan
E-mail : aqwam@jak-stik.ac.id, gustianadede@gmail.com

Abstrak

Badan Penyelenggara Jaminan Sosial atau di singkat dengan BPJS adalah Lembaga yang dibentuk untuk menyelenggarakan program jaminan sosial di Indonesia menurut Undang-Undang No 40 tahun 2004 dan Undang-Undang No 24 tahun 2011 tentang sistem jaminan sosial nasional (SJSN). Berdasarkan Undang-Undang No 24 tahun 2011, BPJS akan menggantikan sejumlah Lembaga jaminan sosial yang ada di Indonesia yaitu Lembaga asuransi jaminan kesehatan PT. Akses menjadi BPJS kesehatan dan Lembaga jaminan sosial ketenagakerjaan PT. Jamsostek menjadi BPJS ketenagakerjaan. Proses klasifikasi analisis sentimen pada penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasi data dan *Lexicon Based* sebagai penentu sentimen Positif, Negatif dan Netral. Tahap awal analisis sentimen adalah pre-processing yang terdiri dari *Filtering*, *Case Folding*, Perbaikan Kata Tidak Baku, Mengubah Kata Bernegasi, *Stopword removal*, dan penghapusan spasi berlebih. Hasil analisis divisualisasikan dalam bentuk tabel, histogram, diagram lingkaran, diagram batang dan wordcloud, kemudian ditampilkan di situs web sebagai media visualisasi. Berdasarkan hasil ujianalisis menghasilkan akurasi sistem pada analisis sentimen menggunakan Algoritma Naive Bayes Classifier terhadap tweet dengan *confusion matrix* sebesar 71 %. dan 29 % sebagai kesalahan sistem dalam membaca data.

Kata Kunci: Analisis Sentimen, *Lexicon Based*, *Naive Bayes Classifier*, *Confusion Matrix*

Pendahuluan

Badan Penyelenggara Jaminan Sosial atau disingkat dengan BPJS adalah Lembaga yang dibentuk untuk menyelenggarakan program jaminan sosial di Indonesia menurut Undang-Undang No 40 tahun 2004 dan Undang-Undang No 24 tahun 2011 tentang sistem jaminan sosial nasional (SJSN). Berdasarkan Undang-Undang No 24 tahun 2011, BPJS akan menggantikan sejumlah Lembaga jaminan sosial yang ada di Indonesia yaitu Lembaga asuransi jaminan kesehatan PT Akses menjadi BPJS kesehatan dan Lembaga jaminan sosial ketenagakerjaan PT. Jamsostek menjadi BPJS ketenagakerjaan. Transporansi PT. Akses dan PT Jamsostek menjadi BPJS dilakukan secara bertahap. Pada awal 2014 PT. Akses akan menjadi BPJS kesehatan, selanjutnya pada tahun 2015 giliran PT. Jamsostek menjadi BPJS Ketenagakerjaan.[1] Media sosial adalah sebuah media online, media sosial bermanfaat bagi para pengguna dengan berbagai cara meliputi mengenai blog, jejaring sosial, wiki, forum dan dunia virtual.

Media sosial adalah sebuah kelompok aplikasi

yang berbasis internet yang dibangun berbasis teknologi web yang memungkinkan pertukaran *user-generated content* adapun beberapa contoh media sosial yang sedang berkembang saat ini yaitu *Twitter* [2].

Analisis sentimen disebut juga dengan *opinion mining* (penambangan opini) yaitu proses untuk mengekstrak suatu opini atau pendapat dari dokumen untuk topik tertentu. Analisis sentimen dilakukan untuk mengetahui suatu kecenderungan opini seseorang terhadap sebuah peristiwa atau masalah, apakah cenderung positif, negatif dan netral. Teknik yang digunakan yaitu *text mining*. *Text mining* merupakan sebuah teknik yang digunakan untuk mengekstraksi informasi yang berguna dari data teks yang tidak terstruktur. *Text mining* mengekstraksi kata kunci atau mengekstraksi pendapat dan ulasan analisis teks sehingga dapat mendukung untuk memahami pendapat masyarakat dalam data *text*. [3]

Analisis Sentimen, *Text Mining* dan *Twitter*

Sentiment analysis atau analisis sentimen dalam bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut dapat dikategorikan sebagai sentimen positif, sentimen negatif dan sentimen netral [4].

Sedangkan pokok masalah yang digunakan dalam penelitian ini adalah: dataset dikumpulkan dari tweet berbahasa Indonesia sebanyak 1.261 tweet, data tweet diambil pada tanggal 29 Oktober 2020 .

1. Tweet yang digunakan adalah tweet yang mengandung kata “BPJS”, dalam bentuk teks Bahasa Indonesia.
2. Tweet diklasifikasi ke dalam 3 kategori sentimen, yaitu kategori positif, negatif dan netral dengan menggunakan metode *Lexicon Based* dan *Naive Bayes Classifier* .
3. Hasil analisis sentimen akan diaplikasikan berbasis website dengan menggunakan bahasa pemrograman R dan Rstudio sebagai text editor.

Penelitian bertujuan untuk mengetahui sentiment pengguna twitter di Indonesia tentang kepuasan masyarakat pada BPJS. Tweet yang memiliki kecenderungan positif, negatif dan netral tersebut diharapkan dapat membantu pihak BPJS dalam melihat pendapat pengguna twitter di Indonesia mengenai pelayanan pada BPJS, dan mengklasifikasikan tweet kedalam sentimen positif, negatif, dan netral menggunakan metode *lexicon base* dan *naïve bayes classifier*, lalu divisualisasikan dalam bentuk diagram dan menghitung akurasi dari metode *naïve bayes classifier*.

Text Mining

Text mining memiliki definisi menambang data yang berupa teks di mana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen. *Text mining* merupakan proses penemuan akan informasi atau trend baru yang sebelumnya tidak terungkap dengan memproses dan menganalisis data dalam jumlah besar [5].

Twitter

Twitter adalah sebuah situs web yang dimiliki dan di operasikan oleh twitter inc, yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan pengguna untuk mengirim dan membaca

pesan tweets. Mikroblog adalah salah satu jenis alat komunikasi online dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan atau melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu.[2]



Gambar 1: Logo Twitter [2]

Naïve Bayes Classifier

Naïve Bayes Classifier adalah teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema bayes (Aturan Bayes). Dengan asumsi independensi (ketidaktergantungan) yang kuat. Dalam *naïve bayes* model yang digunakan adalah model fitur independen[6].

Naïve Bayes Classifier adalah salah satu metode yang populer digunakan untuk keperluan data mining karena kemudahan penggunaannya serta waktu pemrosesannya yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi. Performa *naïve bayes* yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keindependenan atribut (tidak ada kaitan antar atribut). Asumsi keindependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keindependenan tersebut dilanggar performa pengklasifikasian *naïve bayes* cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris. Dari definisi diatas dapat diambil kesimpulan bahwa *naïve bayes* adalah sebuah teknik klasifikasi probabilistik yang berdasarkan teorema bayes yang menggunakan asumsi independen atribut dalam proses pengklasifikasian. Dalam sebuah aturan yang mudah, sebuah klasifikasi *naïve bayes* diasumsikan bahwa ada atau tidaknya ciri tertentu dari sebuah kelas tidak ada hubungan dengan ciri dari kelas lainnya. Dalam penelitian ini data uji adalah dokumen tweets, ada dua tahap dalam klasifikasi dokumen. Tahap pertama adalah pelatihan dokumen yang sudah dikenal kategori. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Pada *Naïve Bayes Classifier* setiap dokumen diwakili oleh pasangan atribut “ $X_1, X_2, X_3 \dots X_n$ ” dimana X_1 adalah kata pertama, X_2 adalah kata

kedua dan seterusnya. Sementara V adalah kumpulan kategori tweet. Pada saat klasifikasi, algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang di uji (V_{NB}), dimana persamaannya :

$$V_{NB} = \frac{Argmax}{V_j V} = \frac{P(x_1, x_2, x_3, \dots, x_n)P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad (1)$$

Pada Variabel $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis :

$$V_{NB} = \frac{Argmax}{V_j V} = P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad (2)$$

Pada persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{NB} = \frac{Argmax}{V_j V} = \prod_{i=1}^n (P(X_i | V_j) P(V_j)) \quad (3)$$

Keterangan:

V_j = Kategori jawaban $j = 1, 2, 3, \dots, n$.

Dimana dalam penelitian ini j_1 = kategori berita negatif dan j_2 = kategori berita positif

$P(x_i | V_j)$ = Probabilitas x_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_i | V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (4)$$

$$P(x_i | V_j) = \frac{nk+1}{n+kosakata}$$

Keterangan :

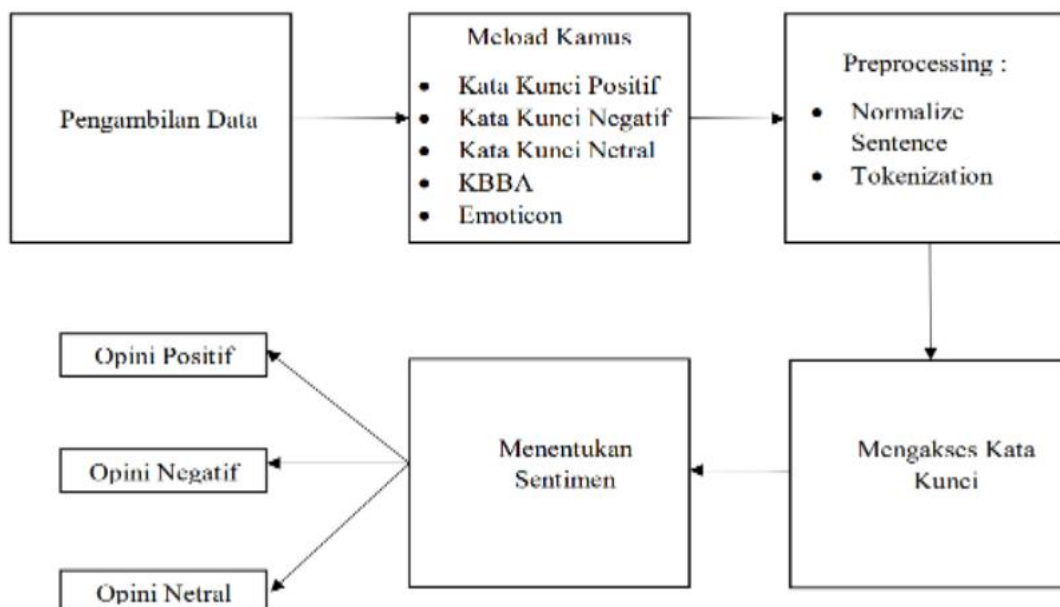
$|docs\ j|$ = jumlah dokumen setiap kategori j
 $|contoh|$ = jumlah dokumen dari semua kategori
 nk = jumlah frekuensi kemunculan setiap kata
 n = jumlah frekuensi kemunculan kata dari setiap kategori
 kosakata = jumlah semua kata dari semua kategori

Lexicon-Based

Lexicon-Based dalam analisis sentimen dikenal adanya metode berbasis *machine learning*, berbasis kamus (*lexicon-based*), atau gabungan keduanya. Secara umum menurut alur proses analisis sentimen menggunakan *lexicon based* dapat dilihat pada Gambar 2[3].

Metode ini disebut juga sebagai *lexical based approach*. *Lexical based approach* merupakan sebuah metode untuk melakukan analisis sentimen dengan menggunakan sebuah kamus sebagai sumber bahasa atau leksikal. Tahapan proses dari metode *lexicon based*.

1. *Data Acquisition* Pada proses ini dilakukan pengambilan data dari media sosial berbahasa indonesia. Hasil dari proses ini adalah daftar opini pembaca serta metadata seperti: nama user dan waktu.
2. *Load Dictionary* Pada proses ini dilakukan load kamus. Kamus adalah komponen penting dalam sistem yang menggunakan pendekatan *lexicon based*. Kamus digunakan dalam proses normalisasi kalimat dan ekstraksi kata kunci, dimana dalam penelitian ini merujuk pada kamus KBBI.



Gambar 2: Alur Proses Analisis Sentimen dengan Lexicon Based

Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi pada set data pengujian yang nilainya telah diketahui. *Confusion matrix* sendiri relatif sederhana dan mudah untuk mengerti, tetapi terminologi yang terkait dapat membingungkan apabila belum memahami konsep dari confusion matrix[5].

Tabel 1: Model *Confusion Matrix*

		True Values	
		+	-
Prediction	+	TP	FP
	-	FN	TN

Berdasarkan Tabel 1, maka :

1. True Positives (TP) adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif
2. False Positives (FP) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai positif
3. False Negatives (FN) adalah jumlah record data positif yang diklasifikasikan sebagai nilai positif
4. True Negatives (TN) adalah jumlah record data negatif yang diklasifikasikan sebagai nilai negative

Nilai yang dihasilkan melalui metode *Confusion Matrix* adalah berupa evaluasi sebagai berikut :

1. *Accuracy*, presentase jumlah record data yang diklasifikasikan (prediksi) secara benar oleh algoritma Rumus : $(TP + TN) / \text{Total data} = \text{Accuracy}$
2. *Misclassification (Error) Rate*, presentase jumlah record data yang diklasifikasikan (prediksi secara salah oleh algoritma. Rumus : $(FP + FN) / \text{Total data} = \text{Misclassification Rate}$.

Metode Penelitian

Pada metode penelitian melalui tahapan :

1. Perencanaan
Tahapan perencanaan dilakukan dengan merencanakan pembuatan sistem termasuk mengumpulkan data dari Twitter yang diperlukan dan berkaitan dengan analisis sentimen.
2. Analisis
Tahap analisis dilakukan dengan analisis sistem, analisis masalah, analisis persyaratan fungsional dan analisis persyaratan non-fungsional.

3. Desain

Fase desain dilakukan dengan membuat desain aplikasi dalam lima bentuk Unified Modeling Language (UML), yaitu Use Case Diagram, Activity Diagram, Sequence Diagram, Class Diagram menggunakan perangkat lunak DIA Diagram dan kemudian membuat desain antarmuka.

4. Implementasi dan pengujian

Tahap implementasi dan pengujian akan dilakukan dengan pengujian pada sistem, serta memastikan program dapat berjalan seperti yang diharapkan.

Perancangan awal penelitian digunakan untuk memberikan gambaran tentang analisis sentiment yang akan dibuat yang diawali dengan membuat perancangan analisis sentimen, pengumpulan data, preprocessing, perancangan klasifikasi menggunakan metode Lexicon Based dan Naïve Bayes Classifier dari perancangan dengan menggunakan UML, rancangan visualisasi, dan rancangan tampilan [7].

Pengumpulan

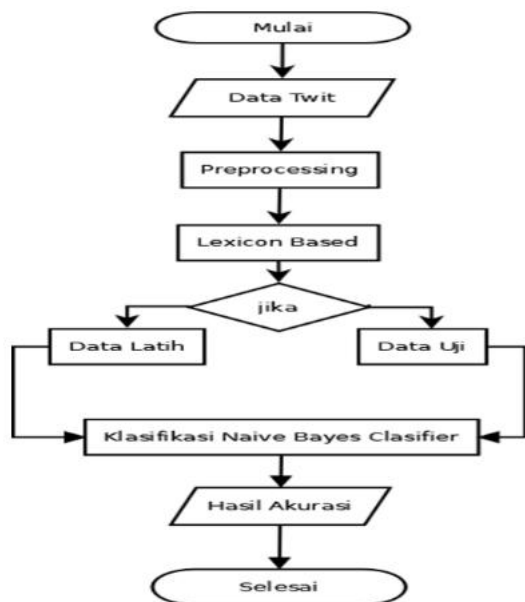
Data Data tweet diambil dari media twitter dengan menggunakan twitter API. Data yang digunakan merupakan data sentimen masyarakat Indonesia yang memberikan tweet BPJS. Data yang diapat pada pengambilan data ini sebanyak 1.261 tweet, data ini akan digunakan sebagai data sebagai analisis sentimen.

Alur Sistem

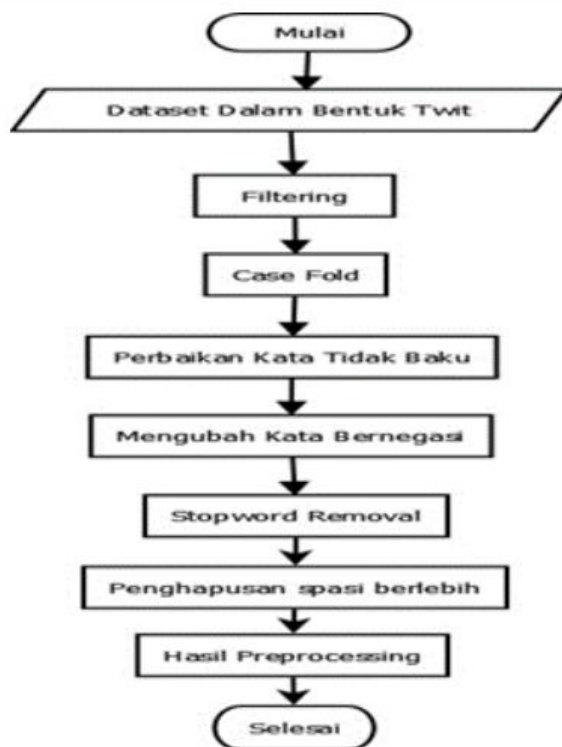
Analisis sistem bertujuan untuk menganalisa sistem akan bekerja, berawal dari input data dari sistem, proses kerja sistem, dan output dari sistem tersebut, lihat Gambar 3.

Preprocessing

Tahap preprocessing data merupakan bagian yang penting dalam melakukan analisis sentimen. Pre processing data bertujuan untuk transformasi data ke dalam suatu format agar dapat lebih mudah dipahami, serta menghilangkan kata yang tidak memiliki makna, menyeragamkan bentuk kata dan mengurangi volume kosa kata, sehingga data tersebut dapat diolah menjadi data yang berkualitas, lihat Gambar 4.



Gambar 3: Diagram Alur Sistem



Gambar 4: Diagram Alur Preprocessing

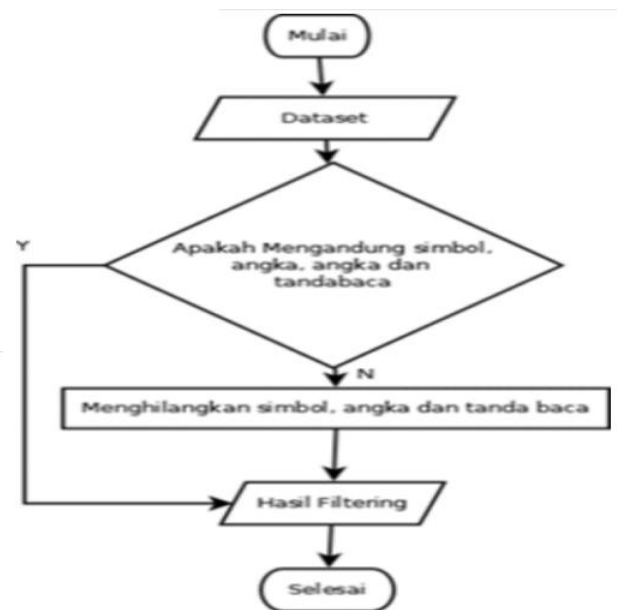
Data Preprocessing

Preprocessing bertujuan untuk mengubah data menjadi format yang lebih baik, sehingga data tersebut dapat diolah menjadi data yang berkualitas. Preprocessing dalam penelitian ini, dilakukan dengan membersihkan dan menormalkan data teks dalam bentuk tweet dari komponennya yang tidak diperlukan.

Pada tahap preprocessing terdiri dari 5 tahap yang dilakukan diantaranya :

1. Filtering

Tahap filtering adalah tahap menghilangkan tanda baca, simbol, dan angka. Alur menghilangkan simbol, angka, dan tanda baca (lihat Gambar 5).



Gambar 5: Diagram Alur Filtering

Tabel 2 adalah contoh hasil dari tahap preprocessing filtering.

Tabel 2: Hasil Filtering

Contoh Data			
Data Uji		Tahapan	
<i>Input</i>	@dede BPJS dapat membantu orang yang tidak mampu mengeluarkan uang yang Besar	<i>Output</i>	BPJS dapat membantu orang yang tidak mampu mengeluarkan uang yang besar

2. Case Folding Pada tahapan ini dilakukan perubahan huruf kapital menjadi huruf kecil pada data tweet. Hanya huruf 'a' sampai dengan 'z' yang diterima. Langkah case folding dapat dilihat Gambar 6.



Gambar 6: Diagram Alur Case Folding

Tabel 3. adalah contoh hasil dari tahap pre-processing Case folding.

Tabel 3: Hasil Case Folding merubah huruf besar menjadi huruf kecil

Contoh Data			
Data Uji		Tahapan	
<i>Input</i>	BPJS dapat membantu orang yang tidak mampu mengeluarkan uang yang Besar	<i>Output</i>	bpjs dapat membantu orang yang tidak mampu mengeluarkan uang yang bes

3. Perbaikan Kata Tidak Baku

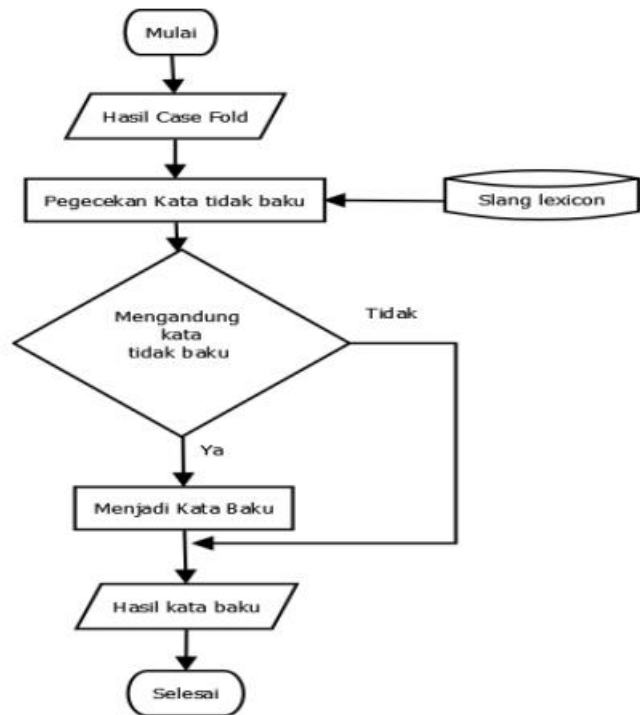
Pada proses ini mengubah kata yang tidak standar menjadi kata standar. Pada tahap ini menggunakan kamus untuk merubah kata non standar menjadi standar. Tabel 3 merupakan kamus kata yang tidak baku, jadi ketika nanti data ada yang menggunakan kata tidak baku maka sistem akan menyamakan dengan yang ada di tabel slang lexicon, misalnya ada kata “jdi” maka akan di baca oleh sistem dan disamakan dengan yang ada di dalam kamus lalu akan di betulkan dengan kata yang ada

pada tabel slang lexicon kata “jdi” tadi akan di berubah menjadi kata “jadi” (lihat Gambar 7).

Tabel 4: Slang *Lexicon*

No	Slang	Formal
1	met	selamat
2	yaa	ya
3	bgt	banget
4	brp	berapa
5	rb	ribu
6	lg	lagi
7	awaas	awas
8	tuuuh	tuh
9	kn	kan
10	jdi	jadi

Untuk Alur proses kata tidak baku dapat dilihat pada Gambar 7.



Gambar 7: Diagram Alur Kata Tidak Baku

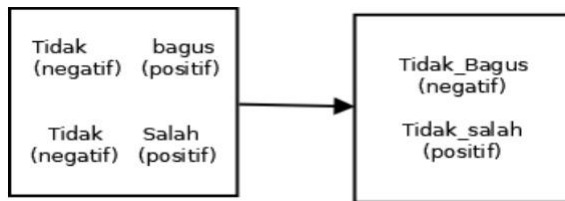
Tabel 5 adalah contoh hasil dari Hasil Perbaikan Kata Tidak Baku

Tabel 5: Hasil Perbaikan Kata Tidak Baku

Contoh Data			
Data Uji		Tahapan Kata Tidak Baku	
<i>Input</i>	bpjs bisa membantu yg kesusahan	<i>Output</i>	bpjs bisa membantu yang kesusahan

4. Mengubah Kata Negasi

Mengubah kata-kata negasi diperlukan karena kata-kata negasi memiliki pengaruh besar pada nilai sentimen pada suatu kalimat. Jika kata negasi ditemukan, kata itu akan digabungkan dengan kata di sebelahnya dengan menggunakan tanda (_). Contoh transformasi kata negasi (lihat Gambar 8).



Gambar 8: Perubahan Tranformasi Kata Negasi

Hasil proses dari tranformasi kata bernegasi dapat dilihat pada tabel 6.

Tabel 6: Hasil Transformasi Kata Negasi

Contoh Data			
Data Uji		Tahapan Kata Bernegasi	
<i>Input</i>	bpjs tidak bisa membantu masyarakat	<i>Output</i>	bpjs tidak_bisa membantu masyarakat

5. Stopword Removal

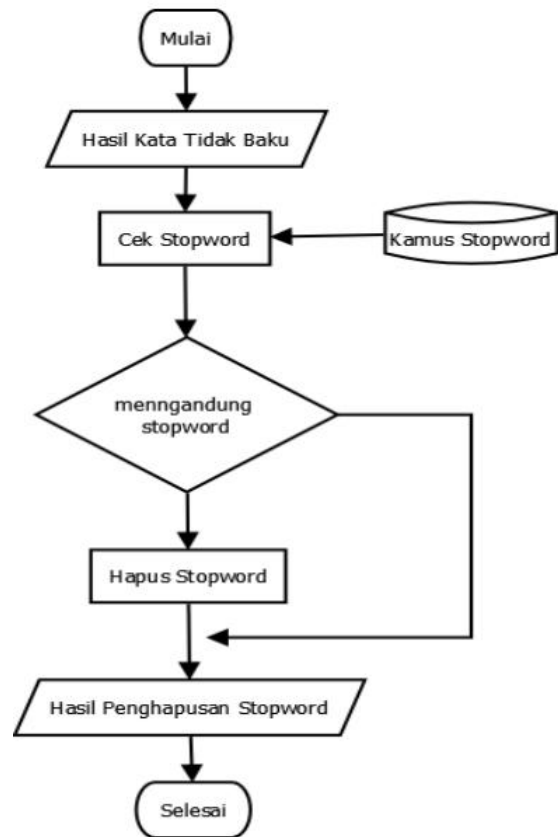
Penghapusan stopwords adalah proses untuk menghilangkan konjungsi, preposisi dan kata ganti. Proses ini dilakukan dengan mencocokkan kata-kata dalam dataset dengan kamus stopwords, jika ditemukan, stopwords akan dihapus. Berikut adalah contoh kamus stopwords dapat dilihat pada tabel 7.

Tabel 7: Kamus Stopword

No	Stopword
1	Ada
2	Adalah
3	Adanya
4	Adapun
5	Agak
6	Agaknya
7	Agar
8	Akan
9	Akankah
10	Akhir
11	Akhiri

Kamus stopwords merupakan kamus yang memiliki kata konjungsi, preposisi dan kata

ganti. Proses ini dilakukan dengan mencocokkan setiap kata dalam dataset dengan kamus stopwords, jika ditemukan, stopwords akan dihapus (lihat Gambar 9).



Gambar 9: Diagram Alur Stopword Removal

Hasil proses dari penghapusan stopwords dapat dilihat pada tabel 8.

Tabel 8: Hasil Stopword Removal

Contoh Data			
Data Uji		Tahapan Stopword Removal	
<i>Input</i>	Pemerintah akan menaikan bpjs	<i>Output</i>	Pemerintah menaikan bpjs

6. Penghapusan Spasi Berlebih

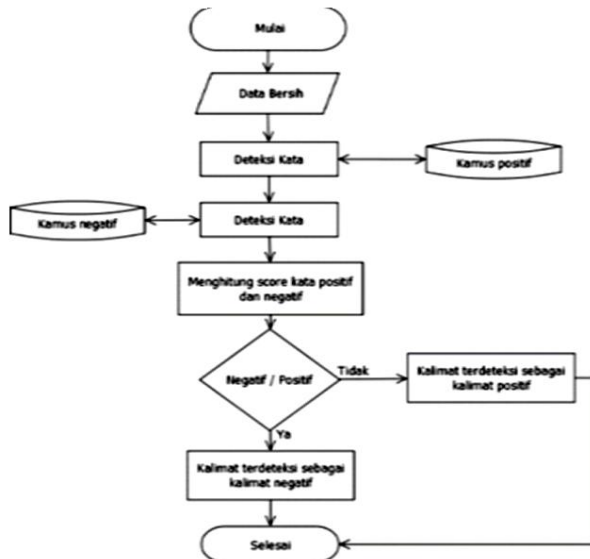
Pada tahap ini data tweet yang memiliki spasi lebih dari satu akan dihapus sepasinya dan dijadikan satu spasi. Berikut adalah contoh penghapusan spasi berlebih dapat dilihat pada tabel 9.

Tabel 9: Hasil Penghapusan Spasi Berlebih

Contoh Data			
Data Uji		Tahapan Penghapusan spasi berlebih	
<i>Input</i>	Pemerintah menaikkan bpjs	<i>Output</i>	Pemerintah menaikkan bpjs

Klasifikasi Sentimen Lexicon Based

Klasifikasi Lexicon Based akan mengklasifikasikan tweets berdasarkan kata positif dan kata negatif pada hasil data tweets yang telah dibersihkan pada tahap preprocessing dan mendefinisikan sebagai sebuah tweets yang cenderung memiliki sentimen positif atau negative masyarakat pada BPJS. Tweet yang memiliki kecenderungan positif, negatif dan netral tersebut diharapkan dapat membantu pihak BPJS dalam melihat pendapat pengguna twitter di Indonesia mengenai pelayanan pada BPJS, dan diklasifikasikan tweet kedalam sentimen positif, negatif, dan netral menggunakan metode lexicon base dan naïve bayes classifier, lalu di visualisasikan dalam bentuk diagram dan menghitung akurasi dari metode naïve bayes classif tweet yang telah melalui proses klasifikasi menggunakan Metode Lexicon Based akan terkandung dalam hasil tabel analisis yang berisi tweet, skor positif, skor negatif, dan skor netral, kelas sentimen (lihat Gambar 10).



Gambar 10: Diagram Alur *Lexicon Based*

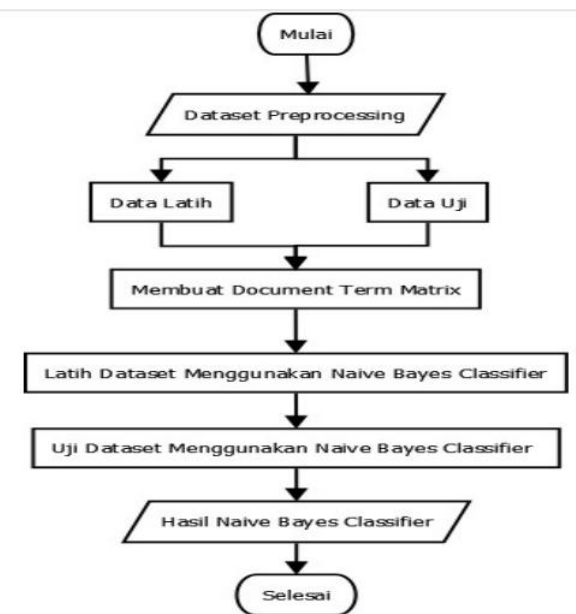
Tweet yang telah melalui proses klasifikasi menggunakan Metode *Lexicon Based* akan terkandung dalam hasil tabel analisis yang berisi tweet, skor positif, skor negatif, dan skor netral, kelas sentimen dapat dilihat seperti yang terdapat pada tabel 10.

Tabel 10: Hasil Klasifikasi *Lexicon Based*

Tweet	Positif	Negatif	Hasil	Sentimen
Keren banget semoga bpjs sukses melayani orang membutuhkan	3	0	3	Positif
Kesal sama bpjs	0	1	-1	Negatif
Bayar kacamata pakai bpjs kali mayaan	0	0	0	Netral

Klasifikasi Naïve Bayes Classifier

Pada klasifikasi menggunakan metode Naïve Bayes, data akan diuji untuk menentukan apakah data termasuk kedalam sentimen positif, negatif, atau netral.



Gambar 11: Diagram Alur *Naïve Bayes Classifier*

Pada metode *Naïve Bayes Classifier* ini terdiri dari 4 proses, yaitu proses menentukan data latih dan data uji, proses pembuatan data termmatrix, pelatihan, dan proses pengujian.

1. Data Latih dan Data Uji

Tahap ini adalah proses pelabelan untuk menentukan data yang masuk kategori data latih dan data uji. Setelah dilakukan tahap *Preprocessing* data yang didapatkan berjumlah 1.261. Perbandingan antara data latih dan data uji adalah 90:10 persen dari jumlah

data. field yang digunakan adalah field sentimen karena pokok perhitungan dari penelitian adalah sentimen positif dan negatif, selain itu untuk melatih model sebaik mungkin maka rasio partisi untuk data latih harus lebih besar dari rasio partisi data uji.

(a) Data Latih

Data latih adalah data yang digunakan sebagai data latih untuk sistem belajar. Data latih diambil 90% dari data 1 – 1.135 data dari dataset.

(b) Data Uji

Data uji adalah data yang digunakan sebagai data uji sistem. Data diambil dari sisa dataset sebesar 10% dari data 1.135 – 1.257.

2. Data Term Matrix

Tahap ini adalah tahap pembobotan kata menggunakan sintaks *Document Term Matrix* (DTM) pada Rstudio. Untuk masuk ke tahap DTM, diperlukan pendefinisian fungsi menggunakan sintaks corpus. Corpus berfungsi untuk membaca data dari suatu dataset. Berikut adalah salah satu sintaks untuk membuat *Document Term Matrix*.

```
nbc_dtm <- DocumentTermMatrix(
nbc_data_corpus)inspect(nbc_dtm[1:5,
1:10])
```

3. Tahap Pelatihan

Tahap ini dilakukan proses pembelajaran menggunakan metode *Naïve Bayes Classifier*. Data dimasukan menggunakan data latih.

Simulasi Tahap Pelatihan Data Latih

Pada tahap ini adalah simulasi perhitungan data latih menggunakan metode Naive Bayes Classifier. Tahap awal yang dilakukan yaitu memilih beberapa himpunan data latih yang sudah selesai tahap preprocessing, dapat dilihat pada Tabel 11.

Tabel 11: Himpunan Data Latih

Komentar	Sentimen
Membantu	Positif
bpjs dapat meringankan masyarakat	Positif
jelek	Negatif
Pelayanan bpjs kurang_baik	Negatif
Bpjs	Netral
Badan Penyelenggara Jaminan Sosial	Netral

Selanjutnya dilakukan perhitungan prior menggunakan rumus :

$$P(V_j) = \frac{|docs_j|}{|contoh|}$$

Positif/negatif/netral =

$$\frac{jumlah\ data\ latih\ positif/negatif/netral}{jumlah\ semua\ data\ latih}$$

Positif : $2/6 = 0,3$

Negatif : $2/6 = 0,3$

Netral : $2/6 = 0,3$

Setelah menentukan keenam data latih, lalu membuat matriks untuk keenam data latih tersebut seperti Tabel 12. Apabila mengandung kata yang didefinisikan maka diberikan tanda yaitu dengan mengisi angka 1, apabila tidak maka dikosongkan

Tabel 12: Matriks 1

Membantu	bpjs	dapat	Meringankan	Masyarakat	Summary
1	-	-	-	-	Positif
-	1	1	1	1	Positif
-	-	-	-	-	Negatif
-	-	-	-	-	Negatif
-	-	-	-	-	Netral
-	-	-	-	-	Netral

Pada tabel DTM 1 ini merupakan tabel *document term matrix* dengan kata positif jika ada kata muncul 1 kata di kata positif jika kata itu positif maka akan dikasih angka satu. artinya kata tersebut sudah muncul 1 kali.

Tabel 13: Matriks 2

jelek	pelayanan	bpjs	Kurang_baik	Summary
-	-	-	-	Positif
-	-	-	-	Positif
1	-	-	-	Negatif
-	1	1	1	Negatif
-	-	-	-	Netral
-	-	-	-	Netral

Pada tabel DTM 2 ini merupakan tabel dokumen matriks term matrix dengan kata negatif jika ada kata muncul 1 kata di kata negatif jika kata itu negatif maka akan di kasih angka satu. artinya kata tersebut sudah muncul 1 kali.

Tabel 14: Matriks 3

jelek	pelayanan	bpjs	Kurang_baik	Summary
-	-	-	-	Positif
-	-	-	-	Positif
1	-	-	-	Negatif
-	1	1	1	Negatif
-	-	-	-	Netral
-	-	-	-	Netral

Pada tabel 14, DTM 3 ini merupakan tabel dokumen term matrix dengan kata netral jika ada kata muncul 1 kata di kata netral jika kata itu netral maka akan di kasih angka satu. Artinya kata tersebut sudah muncul 1 kali. Untuk jumlah kemunculan kata kategori positif sebanyak 5 kata, untuk

kategori negatif sebanyak 4 kata, dan kategori netral sebanyak 5 kata. Untuk jumlah kata ada 15. Berikut adalah rumus perhitungan pembobotan :

$$P(x_i|V_j) = \frac{\frac{|nk+1|}{n+|kosakata|}}{\frac{\text{jumlah kemunculan kata}+1}{\text{jml kemunculan kata tiap kategori}+\text{jml semua kata(term)}}$$

Tabel 15: Hasil Perhitungan Tahap Pelatihan *Naïv Bayes Classifier*

Positif	Negatif	Netral
P(membantu 1+1/5+14= 0,105	P(membantu 0+1/4+14= 0,056	P(membantu 0+1/5+14= 0,053
P(bpjs 1+1/5+14= 0,105	P(bpjs 1+1/4+14= 0,111	P(bpjs 1+1/5+14= 0,105
P(dapat 1+1/5+14= 0,105	P(dapat 0+1/4+14= 0,056	P(dapat 0+1/5+14= 0,053
P(meringankan 1+1/5+14= 0,105	P(meringankan 0+1/4+14= 0,056	P(meringankan 0+1/5+14= 0,053
P(masyarakat 1+1/5+14= 0,105	P(masyarakat 0+1/4+14= 0,056	P(masyarakat 0+1/5+14= 0,053
P(jelek 0+1/5+14= 0,053	P(jelek 1+1/4+14= 0,111	P(jelek 0+1/5+14= 0,053
P(pelayanan 0+1/5+14= 0,053	P(pelayanan 1+1/4+14= 0,111	P(pelayanan 0+1/5+14= 0,053
P(bpjs 0+1/5+14= 0,053	P(bpjs 1+1/4+14= 0,111	P(bpjs 0+1/5+14= 0,053
P(kurang_baik 0+1/5+14= 0,053	P(kurang_baik 1+1/4+14= 0,111	P(kurang_baik 0+1/5+14= 0,053
P(bpjs 0+1/5+14= 0,053	P(bpjs 0+1/4+14= 0,056	P(bpjs 1+1/5+14= 0,105
P(badan 0+1/5+14= 0,053	P(badan 0+1/4+14= 0,056	P(badan 1+1/5+14= 0,105
P(penyelenggara 0+1/5+14= 0,053	P(penyelenggara 0+1/4+14= 0,056	P(penyelenggara 1+1/5+14= 0,105
P(jaminan 0+1/5+14= 0,053	P(jaminan 0+1/4+14= 0,056	P(jaminan 1+1/5+14= 0,105
P(kesehatan 0+1/5+14= 0,053	P(kesehatan 0+1/4+14= 0,056	P(kesehatan 1+1/5+14= 0,105

Simulasi Perhitungan Data Uji

Pada tahap simulasi perhitungan data uji ini merupakan proses inti dari penelitian, yakni proses pengujian data uji berdasarkan hasil dari proses pelatihan menggunakan *Naive Bayes Classifier*, data latih yang menjadi contoh adalah kalimat "Pelayanan BPJS Kurang_Baik". Berdasarkan hasil pada proses pelatihan maka dilakukan perhitungan, sebagai berikut :

Sentimen positif : Pelayanan * BPJS * Kurang_Baik = 0,053 * 0,105 * 0,053 = 0,000295

Sentimen negatif : Pelayanan * BPJS * Kurang_Baik = 0,111 * 0,111 * 0,111 = 0,0014

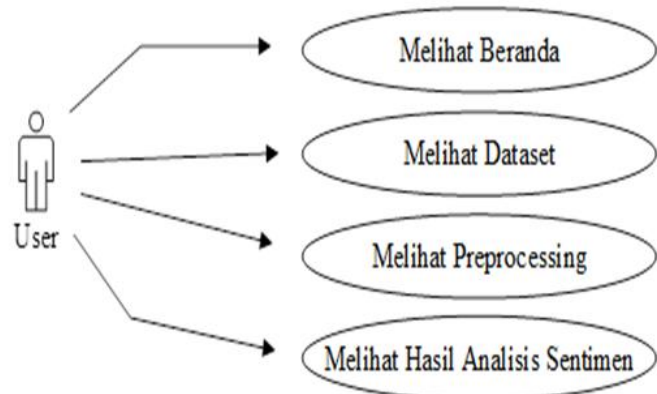
Sentimen netral : Pelayanan * BPJS * Kurang_Baik = 0,053 * 0,105 * 0,053 = 0,000295

Berdasarkan hasil perhitungan diatas, maka kalimat "Pelayanan BPJS Kurang_Baik" dikategorikan sebagai sentimen negatif karena memiliki nilai yang paling besar.

Rancangan

Rancangan UML

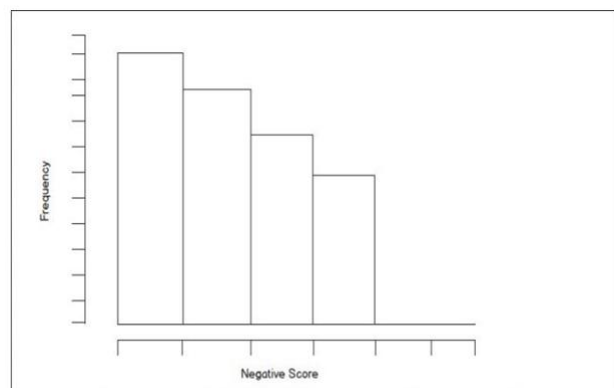
Tahap ini mengenai pembuatan *use case diagram*, menggambarkan secara umum mengenai sistem yang akan berjalan pada web yang dibuat dengan *Shiny package*. Alur website dalam *use case diagram*.



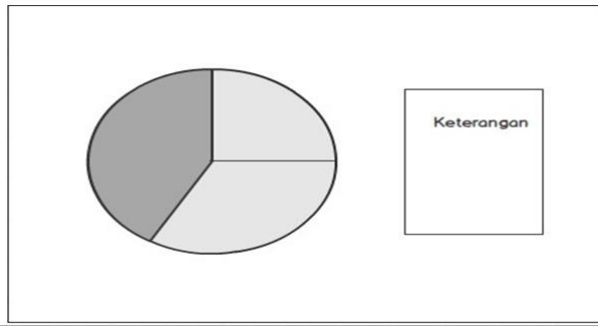
Gambar 12: Use Case Alur Website

Rancangan Visualisasi

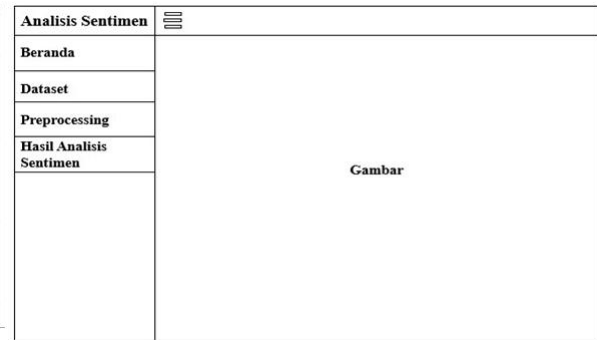
Pada tahap ini rancangan yang akan dibuat terdiri dari visualisasi Histogram, Diagram PieChart, Wordcloud, dan Diagram Batang. Rancangan visualisasi histogram dapat dilihat pada Gambar 13. Rancangan visualisasi diagram Pie Chart disajikan pada Gambar 14. Rancangan visualisasi wordcloud dapat dilihat pada Gambar 15. Rancangan visualisasi diagram batang disajikan pada Gambar 16.



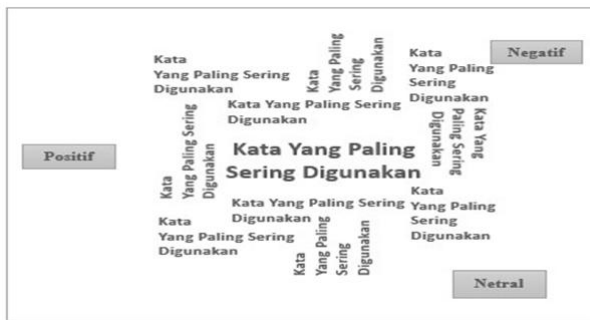
Gambar 13: Rancangan Visualisasi Histogram



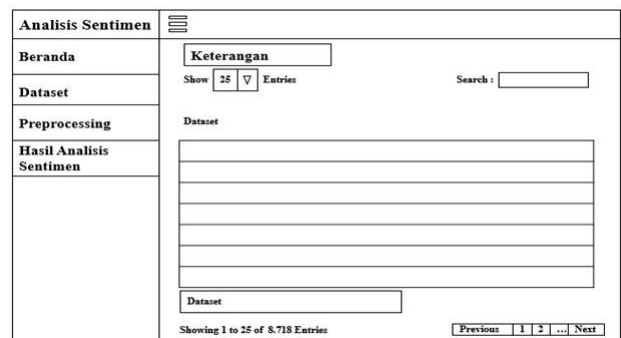
Gambar 14: Rancangan Visualisasi Pie Chart



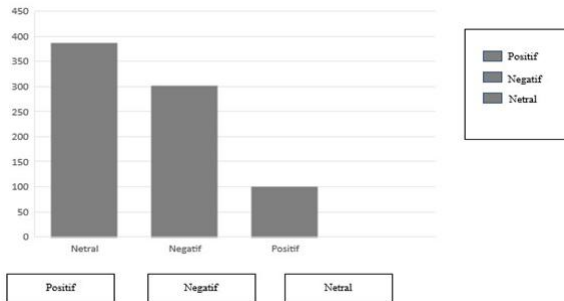
Gambar 17: Rancangan Tampilan Beranda



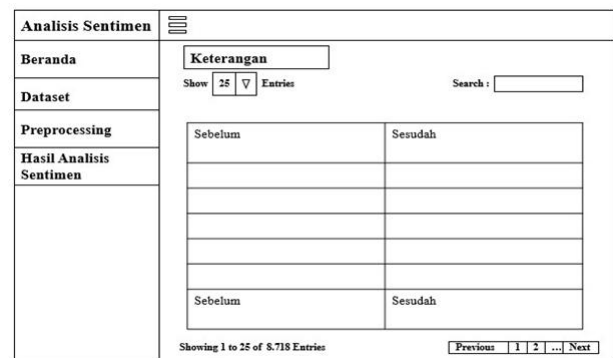
Gambar 15: Rancangan Visualisasi WordCloud



Gambar 18: Rancangan Tampilan Dataset



Gambar 16: Rancangan Visualisasi Diagram Batang



Gambar 19: Rancangan Tampilan Preprocessing

Rancangan Tampilan

Pada tahapan ini merancang tampilan website dari Shiny package Rstudio. Rancangan tampilan website terdiri dari rancangan sidebar menu, halaman Beranda, halaman Dataset, halaman hasil pembersihan data preprocessing, halaman hasil analisis sentimen. rancangan tampilan beranda dapat dilihat pada Gambar 17. Rancangan tampilan dataset disajikan pada Gambar 18. Rancangan tampilan preprocessing disajikan pada Gambar 19. Rancangan tampilan tabel hasil *Lexicon Based* dapat dilihat pada Gambar 20. Rancangan tampilan tabel hasil *Naïve Bayes Classifier* tersaji pada Gambar 21.

Text	Positive	Negative	score	Summary	PosPrecent	NegPresent
Text 1						
Text 2						
Text 3						
Text 4						
Text 5						

Gambar 20: Rancangan Tampilan Tabel Hasil Lexicon

yang telah dilakukan sebelumnya berdasarkan rumus :

$$Akurasi = \frac{TN + TNe + TP}{TN + FN + TNe + FNe + TP + FP} \quad (5)$$

$$Akurasi = \frac{49+36+4}{49+13+36+23+4+1}$$

$$Akurasi = 0,7063492$$

Berdasarkan perhitungan akurasi tersebut, akurasi klasifikasi sentimen menggunakan metode *Lexicon- Based* dan *Naïve Bayes Classifier* mendapatkan hasil akurasi sebesar 71 %.

Penutup

Hasil penelitian analisis sentimen pada tweet bahasa Indonesia dengan kata kunci “BPJS” menggunakan *Lexicon Based* dan *Naive Bayes Classifier* , mendapatkan hasil kesimpulan analisis sentimen :

1. Data tweet yang diolah sebanyak 1.261.
2. Menggunakan data latih sebesar 90% dari dataset yaitu sebanyak 1.135 data dan data test sebesar 10% dari dataset sebanyak 126 data.
3. Melakukan pengujian menggunakan confusion matrix dan didapatkan hasil akurasi dari analisis sentimen terhadap tweet dengan teks Bahasa Indonesia menggunakan *Lexicon Based* dan *Naive Bayes Classifier* sebesar 71 %.

Daftar Pustaka

- [1] Ika Widiastuti, “Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Jawa

Barat, Public Inspiration”, Jurnal Administrasi Publik, Universitas Krisna Dwipayana, Vol.2, No.2, Desember 2017, ISSN: 2581-2378, 2017.

- [2] Annisa Husnusyifa, “Pengaruh Penggunaan Media Sosial Twitter Terhadap Sikap Fanatisme Penggemar (Studi Pada Media Sosial Twitter @Btobindonesia Terhadap Sikap Fanatisme Penggemar)”, IDEA:Jurnal Humaniora, Vol.2 No.2, Oktober 2019, ISSN:2655-3139, (p) : 120-133, 2019.
- [3] Imam Fahrur Rozi, Sholeh Hadi Pramono dan Erfan Achmad Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi”, Jurnal EECCIS Vol. 6, No. 1, Juni 2012, (p) : 37-43, 2012.
- [4] Raflialdy Raksanagara, Yulison Herry Chrisnanto, dan Asep Id Hadiana, “Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes”, Prosiding SNST ke-7, ISBN : 978-602- 99334-5-1, (p) : 19- 24, Fakultas Teknik Universitas Wahid Hasyim, Semarang, 2016.
- [5] Mutiara Pujiasepti, “Analisis Sentimen Terhadap Aplikasi E-Wallet X dari Ulasan Pengguna Menggunakan Metode Lexicon Based, Naïve Bayes Classifier dan Analisis Tingkat Kepuasan Pengguna Menggunakan Pieces”, Skripsi, Jurusan Sistem Informasi, Universitas Gunadarma, 2019.
- [6] Lia Durrotul Mahbubah, Eri Zuliarso, “Analisis Sentimen Twitter Pada PILPRES 2019 Menggunakan Algoritma Naïve Bayes”, Prosiding SINTAK 2019, ISBN: 978-602-8557-20-7, (p) : 193- 199, 2019.
- [7] Munawar, “Analisis Perancangan Sistem Berorientasi Objek Dengan UML”, Bandung: Penerbit Informatika, 2018.