

Analisis Sentimen Twitter Terhadap Pembayaran ShopeePayLater Pada Aplikasi Belanja *Online* (Shopee) Menggunakan Metode *Lexicon Based* dan *Naïve Bayes Classifier*

Indira Mahayani, Dewi Agushinta R. dan Muhammad Edy Supriyadi

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

E-mail : indiramahayani03@gmail.com, (dewiar, edysupriyadi)@staff.gunadarma.ac.id

Abstrak

Shopee adalah sebuah aplikasi yang bergerak di bidang jual beli secara *online* dan dapat diakses secara mudah dengan menggunakan *smartphone*. Pada aplikasi Shopee terdapat fitur PayLater yaitu belanja terlebih dahulu dan membayar di awal bulan. Fitur PayLater pada aplikasi Shopee dinamakan ShopeePayLater. Setiap pengguna baru yang ingin menggunakan fitur ShopeePayLater, memiliki kemungkinan ingin mengetahui reaksi dari pengguna sebelumnya pada twitter berupa *tweets* yang telah menggunakan fitur ShopeePayLater. Sebuah analisis sentimen diperlukan untuk mengetahui opini pengguna terhadap fitur pembayaran dengan menggunakan ShopeePayLater pada aplikasi Shopee, cenderung masuk ke positif atau negatif serta sebagai bahan pertimbangan untuk menggunakan fitur ShopeePayLater bagi calon pengguna. Tahapan proses analisis sentimen ini adalah pengambilan data (*crawling*), text pre-processing, klasifikasi sentimen yang terdiri dari metode *lexicon Based* dan *Naïve Bayes Classifier* serta penyusunan hasil analisis sentimen. Perhitungan pada penelitian ini menggunakan perhitungan tingkat keakurasian pengujian *confusion matrix* lalu hasil yang didapat berupa sentimen aktual sebesar 82,52 % dan kesalahan sistem atau *error rate* sebesar 17,48 %. Presisi positif sebesar 89,54 %, presisi negatif sebesar 47,06 %, dan recall sebesar 89,54 %. Hasil analisis sentimen pada penelitian ini menunjukkan bahwa pembayaran menggunakan ShopeePayLater memiliki kecenderungan sentimen positif pada saat data *tweets* diambil.

Kata kunci : Analisis Sentimen, *Confusion Matrix*, *Lexicon Based*, *Naïve Bayes Classifier*, ShopeePayLater

Pendahuluan

Perkembangan teknologi informasi semakin maju di era digital saat ini bukan hanya sebagai media informasi dan komunikasi, namun internet mampu menambah pendapatan seseorang dalam perekonomian dengan memenuhi keinginan masyarakat dalam berbelanja secara praktis yaitu dengan adanya belanja *online*. Bagi sebagian orang, belanja *online* merupakan suatu kebutuhan yang harus dipenuhi untuk ke-

hidupannya sebagai konsumen. Shopee merupakan salah satu aplikasi yang bergerak di bidang jual beli secara *online* dan dapat diakses secara mudah dengan menggunakan *smartphone*. Aplikasi shopee memiliki fitur PayLater yaitu pengguna dapat berbelanja terlebih dahulu dengan limit yang telah ditentukan. Fitur PayLater pada Shopee dinamakan ShopeePayLater. Melalui fitur ShopeePayLater, Shopee menawarkan keuntungan dalam bentuk pinjaman dana instan dengan bunga

yang sangat minim kepada para pengguna aktif di aplikasi Shopee. Untuk bisa menggunakan layanan ini pengguna akan diminta memberikan data pribadi, foto diri dan foto KTP.

Setiap pengguna baru yang ingin menggunakan fitur ShopeePayLater, memiliki kemungkinan ingin mengetahui reaksi dari pengguna sebelumnya pada twitter berupa *tweets* yang telah menggunakan fitur ShopeePayLater. Sebuah analisis sentimen diperlukan untuk mengetahui opini pengguna mengenai seberapa besar fitur ShopeePayLater ini dapat memberikan kemudahan kepada pengguna. Untuk itulah diperlukan suatu sistem yang dapat menganalisis opini *tweets* pengguna ShopeePayLater pada Twitter dengan menerapkan analisis sentimen. Analisis sentimen adalah studi yang bertujuan untuk menganalisis opini, sentimen dan emosi yang terdapat pada dokumen atau data. Tugas dasar dari analisis sentimen adalah untuk mengelompokkan sifat dari teks yang ada di dalam kalimat maupun pendapat, yang pada umumnya terbagi menjadi kelas yaitu negatif dan positif [1]. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen [2].

Dalam melakukan proses klasifikasi analisis sentimen dapat dilakukan menggunakan berbagai metode, yaitu dengan metode *Lexicon Based* dan *Naïve Bayes Classifier*. *Lexicon Based* adalah metode untuk menentukan sentimen atau polaritas opini melalui beberapa fungsi kata opini dalam dokumen atau kalimat yang menghasilkan sentimen bernilai positif atau negatif [3]. *Naïve Bayes Classifier* adalah metode yang berpotensi baik untuk melakukan klasifikasi data dari pada metode klasifikasi lainnya dalam hal akurasi dan komputasi [4].

Berdasarkan uraian ini, maka akan dilakukan penelitian mengenai analisis sentimen *tweets* pembayaran menggunakan PayLater pada aplikasi Shopee. Alasan pemilihan objek penelitian ini karena aplikasi Shopee memiliki fitur PayLater sendiri yaitu ShopeePayLater sehingga memiliki banyak ulasan *tweets* mengenai fitur ini yang diambil dari Twitter berupa data *tweets* yang mengandung kata "ShopeePayLater". Pada penelitian ini digu-

nakan metode *Lexicon Based* dengan tujuan untuk menentukan sentimen awal dari suatu kalimat berdasarkan jumlah proporsi kata positif atau negatif yang menjadi penyusun suatu kalimat, *Naïve Bayes Classifier* ditujukan untuk mengolah dan mengklasifikasikan *tweets* sehingga akan diketahui hasil klasifikasi dari data *tweets* dalam bentuk proporsi sentimen positif dan negatif.

Penelitian ini dilakukan dengan mempelajari penelitian-penelitian terdahulu yang memiliki keterkaitan tentang analisis sentimen, metode *Lexicon Based* dan *Naïve Bayes Classifier*, yaitu yang pertama adalah penelitian yang memiliki hasil bahwa menggunakan Algoritma *Naïve Bayes Classifier* dapat diterapkan atau digunakan untuk pengklasifikasian tweet review film. Dari hasil analisis sentimen yang didapatkan menunjukkan hasil sentimen negatif lebih banyak jumlahnya dari pada sentimen positif dan sentimen netral [5]. Penelitian yang kedua yaitu memiliki hasil bahwa hasil analisis sentimen terhadap jasa transportasi *online* dengan *Naïve Bayes*, didapatkan hasil sentimen positif sebesar 88.60 % dan sentimen negatif sebesar 11.40 % dengan akurasi sebesar 86.80 % [6]. Penelitian yang ketiga yaitu dengan hasil yang didapat Berdasarkan pengujian yang telah dilakukan, diperoleh hasil akurasi terbaik dari analisis sentimen tentang opini film pada dokumen Twitter berbahasa Indonesia menggunakan *Naïve Bayes* dengan perbaikan kata tidak baku adalah sebesar 98.33% dan untuk precision, recall, dan f-measure adalah 96.77%, 100%, dan 98.36% [7].

Analisis Sentimen dan Klasifikasi

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstraksi dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [8]. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang.

Klasifikasi adalah proses pencarian

sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Model itu sendiri diperoleh berdasarkan analisis dari data yang sudah diketahui label kelasnya. Dalam pengklasifikasian terdapat dua tahap di dalamnya, yaitu tahap pengamatan dan tahap pengujian. Tahap pengamatan merupakan tahap ketika algoritma membangun model klasifikasi dari data latih yang sudah diketahui label kelasnya. Sedangkan tahap pengujian merupakan langkah untuk menerapkan model tersebut pada data uji sehingga kelas yang sesungguhnya dari data uji dapat diketahui [9].

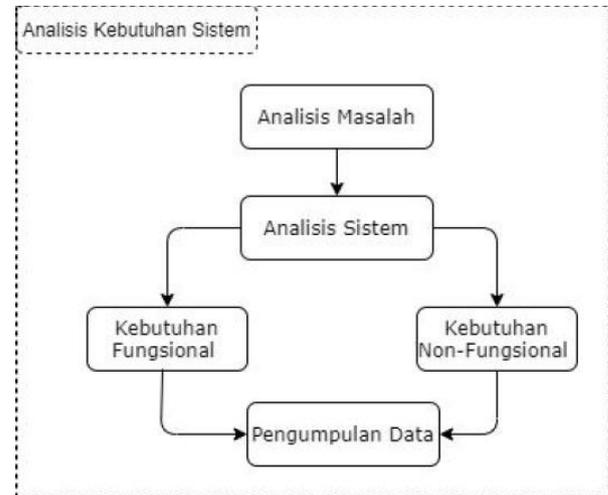
Metode Penelitian

Dalam penelitian ini menggunakan metode *System Development Life Cycle* (SDLC) atau siklus hidup pengembangan sistem untuk membangun website, dalam menganalisis menggunakan metode *Lexicon Based* dan *Naïve Bayes Classifier* untuk klasifikasi data *tweets*. Adapun tahapan-tahapan yang dilakukan dalam metode SDLC yaitu :

1. Tahap Perencanaan, yaitu tahap pengumpulan data *tweets* dan bahan yang dibutuhkan untuk analisis sentimen melalui jurnal, buku dan internet, serta pemilihan bahasa pemrograman.
2. Tahap Analisis, Penulis melakukan analisis sentimen dari data *tweets* pengguna ShopeePayLater yaitu analisis kebutuhan sistem, analisis pre-processing, dan analisis algoritma.
3. Tahap Perancangan, Penulis merancang website dengan membuat UML dan membuat rancangan antarmuka.
4. Tahap Implementasi dan uji coba, Penulis menggunakan bahasa pemrograman RStudio dengan paket Shiny. Pada tahap uji coba yaitu tahap yang penting untuk mengetahui masalah yang timbul pada saat pengoperasian program, apakah sudah sesuai dengan spesifikasi yang telah ditentukan sebelumnya serta memastikan program dapat berjalan seperti yang diharapkan.

Hasil dan Pembahasan

Pada tahap analisis kebutuhan sistem, tahap-tahap yang akan berjalan pada sistem analisis adalah analisis masalah, analisis sistem, yang terdiri dari analisis kebutuhan fungsional serta non-fungsional. Berbagai tahapan diilustrasikan pada Gambar 1.



Gambar 1: Analisis Kebutuhan Sistem

Analisis Masalah

Pendapat atau opini masyarakat terhadap pengguna ShopeePayLater pada platform Twitter dapat memberikan data yang jika dikumpulkan dan dilakukan analisa akan menghasilkan pola-pola dan informasi. Informasi diambil dari proses analisis sentimen terhadap data opini masyarakat publik dapat berupa nilai-nilai sentimen, jumlah opini untuk masing-masing klasifikasi sentimen, jumlah responden, dan lain sebagainya yang dapat bermanfaat untuk mengetahui respon pengguna terhadap fitur pembayaran dengan menggunakan Pay Later pada aplikasi Shopee, cenderung masuk ke positif atau negatif, sehingga hasil analisis sentimen terhadap opini masyarakat pada platform media sosial Twitter dapat digunakan untuk memberikan informasi serta bahan pertimbangan bagi calon pengguna fitur ShopeePayLater dan pihak Shopee dapat mengetahui respon konsumen yang dijadikan sebagai bahan untuk perbaikan dan peningkatan kualitas fitur PayLater.

Analisis Sistem

Analisis sistem merupakan penguraian dari sistem utama ke dalam sub-sub sistem dengan

tujuan untuk mengidentifikasi permasalahan yang ada dan kebutuhan-kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem baru yang lebih baik. Analisis sistem bertujuan untuk menganalisis bagaimana sistem bekerja, mulai dari data masukkan sistem, proses kerja sistem, hingga hasil keluaran dari sistem. Pada tahapan ini dilakukan analisis kebutuhan fungsional dan kebutuhan non-fungsional dari aplikasi analisis sentimen.

Analisis Kebutuhan Fungsional dan Non-Fungsional

Kebutuhan fungsional berupa kebutuhan proses-proses yang dapat dilakukan oleh aplikasi analisis sentimen. Tabel 1 menjelaskan kebutuhan fungsional aplikasi analisis sentimen. Kebutuhan non-fungsional berupa jenis kebutuhan perangkat keras dan perangkat lunak yang dibutuhkan untuk membangun atau menjalankan aplikasi analisis sentimen yang sedang dirancang. Tabel 2 menjelaskan kebutuhan non-fungsional aplikasi analisis sentiment.

Analisis Sumber Data

Data *tweets* diambil menggunakan koneksi RStudio terhadap Application Programming Interface (API) Twitter dengan kata kunci "ShopeePayLater". Jumlah data yang diambil kurang lebih sebanyak 373 data *tweets* dengan rentang waktu dari tanggal 10 Mei 2020 hingga 03 Juni 2020. Data yang digunakan adalah *tweets* dari para pengguna Shopee yang proses pembayarannya menggunakan ShopeePayLater. Berikut ini adalah salah satu contoh tweet pengguna ShopeePayLater yang memiliki makna sentimen positif "@nanaslagi gampang, banyak voucher, enak bgt pake ShopeePayLater ehe #PromoShopeeTHREkstra".

Tabel 1: Analisis Kebutuhan Fungsional

No	Parameter	Kebutuhan Fungsional
1.	Pengambilan Data (<i>Text Crawling</i>)	<ol style="list-style-type: none"> 1. Sistem harus terhubung ke API <i>Search Twitter</i> untuk dapat menelusuri dan mendapatkan <i>dataset</i> teks sumber (<i>tweets</i>). 2. Sistem harus melakukan filtrasi dan eliminasi data yang tidak diperlukan, sehingga tidak terjadi redundansi dan mengurangi <i>noise</i> pada data yang dianalisis. 3. Sistem harus menampilkan <i>dataset</i> teks sumber (<i>tweets</i>) yang telah didapatkan dalam bentuk file berekstensi <i>.csv</i>.
2.	Tahap <i>Pre-processing</i> (<i>Text Pre-processing</i>)	<ol style="list-style-type: none"> 1. Sistem harus menghapus elemen-elemen yang tidak diperlukan seperti tanda baca, <i>hashtag</i>, <i>mention</i>, <i>link</i> tautan dan lain-lain pada teks sumber (<i>tweets</i>). 2. Sistem harus merubah kapitalisasi dokumen menjadi huruf kecil. 3. Sistem harus melakukan perubahan <i>slang words</i> menjadi kata baku. 4. Sistem harus menghapus kata-kata yang tidak diperlukan (<i>stop words</i>) 5. Sistem harus melakukan <i>stemming</i> terhadap kata berimbuhan menjadi bentuk kata dasar.
3.	Klasifikasi	<ol style="list-style-type: none"> 1. Sistem harus melakukan klasifikasi sentimen menggunakan metode <i>lexicon Based</i>. 2. Sistem harus melakukan klasifikasi sentimen menggunakan metode <i>Naïve Bayes classification</i>.
4.	Keluaran data (Hasil Klasifikasi dan Visualisasi data)	<ol style="list-style-type: none"> 1. Sistem harus menampilkan hasil analisis dengan visualisasi data berupa <i>word cloud</i>, <i>diagram pie</i>, diagram batang dan <i>histogram</i>. 2. Menampilkan akurasi sistem dengan uji coba <i>Confusion Matrix</i>.

Tabel 2: Analisis Kebutuhan Non-Fungsional

1.	Perangkat Keras	1. Asus TP301U dengan prosesor Intel® Core™ i5-6200U 2.3GHz 3M Cache. 2. Memori RAM 4GB. 3. Storage 1TB. 4. Monitor 13,3”.
2.	Perangkat Lunak	1. Sistem Operasi Windows 10 Home Pro 64-bit. 2. R-Programming 3.6.2. 3. IDE RStudio 1.2.1333.

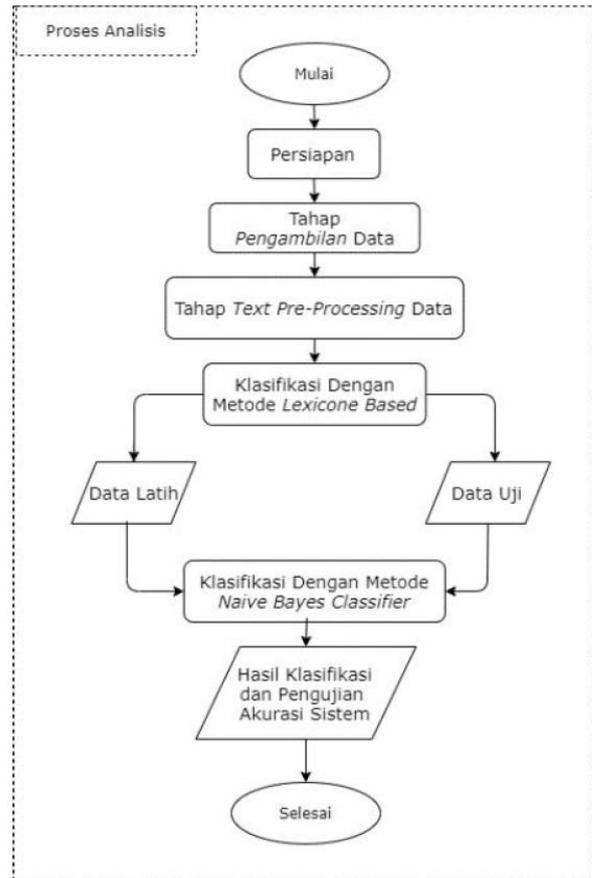
Sedangkan berikut adalah salah satu contoh tweet pengguna “ShopeePayLater” yang memiliki makna sentimen negatif “@rlthingy Emang ngeselin sih shopee sekarang btw. ShopeePayLater ga dapet voucher gratis ongkir jg. Jd males bgt.”. Data masukan untuk analisis sentimen pada penelitian ini merupakan kumpulan *tweets* yang sudah dilakukan proses text-preprocessing menjadi sebanyak 344 data *tweets* selanjutnya dikategorikan ke dalam data uji dan data latih. Data *tweets* terdiri dari 241 data latih dan 103 data uji.

Perancangan Proses Analisis Sentimen

Kegiatan analisis sentimen terhadap data *tweets* terhadap pengguna ShopeePayLater terdiri dari proses-proses dan sub-proses yang saling berhubungan. Dimulai dari proses pengambilan data, pemrosesan data, klasifikasi sentimen, hingga pengujian akurasi dan visualisasi hasil sentimen. Berikut langkah-langkah dalam melakukan perancangan proses analisis sentimen, dijabarkan pada Gambar 2.

Pada tahap pengambilan data merupakan proses utama dalam mendapatkan data *tweets* berbentuk ekstensi .csv (comma-separated values) dari Twitter. selanjutnya pada analisis sentimen adalah *text pre-processing* yang berfungsi untuk mempersiapkan data *tweets*. Tujuan utama dari tahap ini adalah menghilangkan kata dan simbol yang tidak diperlukan, menyeragamkan bentuk kata dan mengurangi panjang kata dalam satuan data tweet sehingga menghasilkan data yang singkat, padat dan jelas untuk dilakukan analisis sentimen pada tahap berikutnya. Tahap ini meliputi beberapa sub-proses, yaitu case folding, filtering, mengganti kata negasi, *stopword removal*, *stemming* kata dasar, dan diakhiri dengan penyimpanan data

yang merupakan output dari proses *text pre-processing*.



Gambar 2: Perancangan Proses Analisis Sentimen

Tahap Pre-Processing

Tahap ini meliputi beberapa sub-proses, yaitu case folding, filtering, mengganti kata negasi, *stopword removal*, *stemming* kata dasar, dan diakhiri dengan penyimpanan data yang merupakan output dari proses *text pre-processing*. Berikut adalah penggalan perintah untuk memanggil file yang akan dijadikan master data (dataset) untuk tahapan *text pre-processing*.

```

shopeepaylater<-read.csv ('data_shopeepaylaterbaru.csv'),
shopeepaylater<-as.character(shopeepaylater$text)
    
```

1. Case Folding Case folding adalah tahap mengubah huruf pada data *tweets* yang awalnya merupakan huruf kapital menjadi huruf kecil. Tabel 3 merupakan contoh data *tweets* sebelum dan sesudah melalui tahap case folding.

Tabel 3: Case Folding

Contoh Data	
Data Latih/ Uji (Input)	Tahap Case Folding (Output)
@sanjunhoe Hi Kak, maaf ya. Saat ini akun Shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan ShopeePayLater dari salah satu pengguna. Kamu ngga perlu khawatir karena hal ini ngga akan berpengaruh ke saldo ShopeePay. Pastiin pengguna udah lunasin tagihannya ya :) ^DA	@sanjunhoe hi kak, maaf ya. saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna. kamu ngga perlu khawatir karena hal ini ngga akan berpengaruh ke saldo shopeepay. pastiin pengguna udah lunasin tagihannya ya :) ^da

Tabel 4: Filtering

Contoh Data	
Tahap Case Folding (Input)	Tahap Filtering (Output)
@sanjunhoe hi kak, maaf ya. saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna. kamu ngga perlu khawatir karena hal ini ngga akan berpengaruh ke saldo shopeepay. pastiin pengguna udah lunasin tagihannya ya :) ^da	hi kak maaf ya saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna kamu ngga perlu khawatir karena hal ini ngga akan berpengaruh ke saldo shopeepay pastiin pengguna udah lunasin tagihannya ya da

2. Filtering Filtering merupakan tahapan untuk menghapus elemen-elemen dari teks sumber yang tidak diperlukan dan melakukan standarisasi untuk setiap teks sumber. Tabel 4 adalah hasil tahap filtering.

3. Mengganti kata negasi Mengganti kata negasi merupakan penggantian kata negasi dilakukan agar kata yang memiliki makna negasi seperti kata “tdk”, “tak”, “g”, “gk”, “ga”, “gak”, “ngga”, “engga”, “enggak”, dan “tanpa” menjadi kata “tidak”. Tabel 5 adalah hasil tahap mengganti kata negasi.

4. Stopword Removal Penghapusan stop-words adalah untuk mengurangi jumlah kata yang terdapat dalam satu teks sumber. Tabel 6 berikut adalah hasil tahap stopword removal

Tabel 5: Mengganti kata negasi

Contoh Data	
Tahap Filtering (Input)	Tahap Mengganti Kata Negasi (Output)
hi kak maaf ya saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna kamu ngga perlu khawatir karena hal ini ngga akan berpengaruh ke saldo shopeepay pastiin pengguna udah lunasin tagihannya ya da	hi kak maaf ya saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna kamu tidak perlu khawatir karena hal ini tidak akan berpengaruh ke saldo shopeepay pastiin pengguna udah lunasin tagihannya ya da

Tabel 6: *Stopword Removal*

Contoh Data	
Tahap Mengganti Kata Negasi (Input)	Tahap <i>Stopword Removal</i> (Output)
hi kak maaf ya saat ini akun shopee kamu dibatasi sementara karena masih menunggu pembayaran tagihan shopeepaylater dari salah satu pengguna kamu tidak perlu khawatir karena hal ini tidak akan berpengaruh ke saldo shopeepay pastiin pengguna udah lunasin tagihannya ya da	maaf akun shopee dibatasi menunggu pembayaran tagihan shopeepaylater salah pengguna tidak khawatir tidak berpengaruh saldo shopeepay pastiin pengguna lunasin tagihan

Tabel 7: *Stemming* Kata Dasar

Contoh Data	
Tahap <i>Stopword Removal</i> (Input)	Tahap <i>Stemming</i> (Output)
maaf akun shopee dibatasi menunggu pembayaran tagihan shopeepaylater salah pengguna tidak khawatir tidak berpengaruh saldo shopeepay pastiin pengguna lunasin tagihan	maaf akun shopee batas tunggu bayar tagih shopeepaylater salah guna tidak khawatir tidak pengaruh saldo shopeepay pasti guna lunas tagih

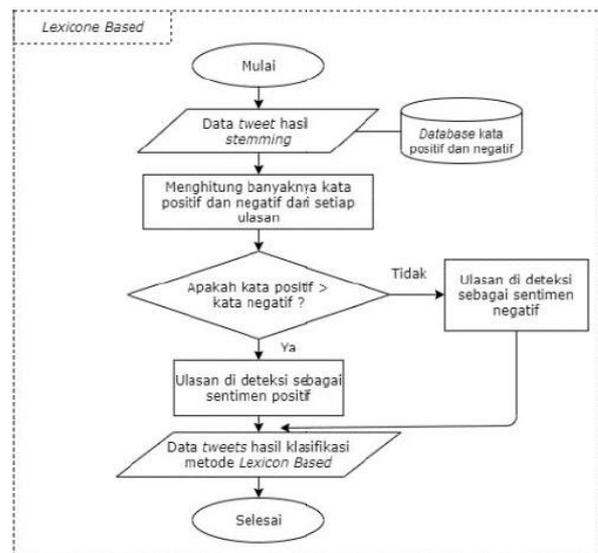
5. *Stemming* Kata Dasar Pada tahap *stemming* kosakata berimbuhan yang terdapat dalam kalimat teks sumber akan diubah menjadi kata dasar dengan melakukan pemotongan terhadap imbuhan dari kata berimbuhan. Tabel 7 adalah hasil tahap *stemming*.

Klasifikasi *Lexicon Based*

Metode *Lexicon Based* merupakan salah satu klasifikasi yang digunakan dalam penelitian ini. Data tweet yang telah melalui tahap *text pre-processing* akan dibaca per kata dan dicocokkan

dengan kamus kata bersentimen, dengan hasil apabila suatu teks sumber memiliki jumlah kata bersentimen positif lebih banyak dari jumlah kata bersentimen negatif, maka teks sumber tersebut diklasifikasikan sebagai teks sumber bersentimen positif, begitu pula sebaliknya.

Gambar 3 adalah gambaran tahap *Lexicon Based* secara otomatis. Klasifikasi dengan metode *lexicon* adalah klasifikasi tweet berdasarkan kata positif dan kata negatif pada hasil data *tweets* yang telah dibersihkan pada tahap *text pre-processing*.

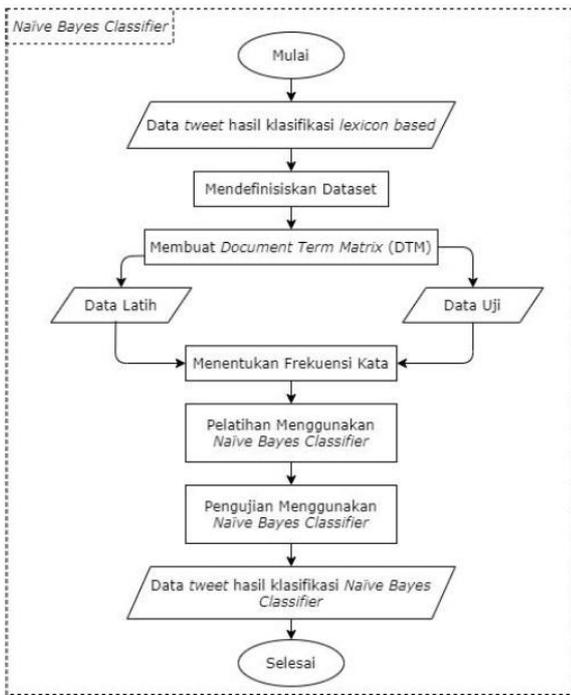


Gambar 3: Klasifikasi Dengan Metode *Lexicon Based*

Klasifikasi *Naïve Bayes Classifier*

Metode *Naïve Bayes* merupakan metode klasifikasi kedua yang digunakan pada kegiatan analisis sentimen dalam penelitian ini. Klasifikasi dengan metode *Naïve Bayes* adalah klasifikasi dokumen yang didasari pada pencarian nilai probabilitas tertinggi pada suatu dokumen yang diklasifikasi berdasarkan pola-pola yang ditemukan pada dokumen latih sebagai *knowledge base*. Metode ini terdiri dari 2 (dua) proses, yaitu proses pelatihan dan proses pengujian.

Berikut akan diterangkan tahapan proses klasifikasi menggunakan metode *Naïve Bayes Classifier* pada Gambar 4.



Gambar 4: Klasifikasi Dengan Metode *Naïve Bayes Classifier*

Berikut akan diterangkan tahapan proses klasifikasi menggunakan metode *Naïve Bayes Classifier*:

1. Mendefinisikan Dataset: pendefinisian dataset yang digunakan pada penelitian ini adalah hasil data tweet opini masyarakat terhadap pengguna ShopeePayLater yang telah melalui seluruh tahapan *pre-processing* dan telah dilakukan klasifikasi dengan metode *Lexicon Based*.
2. Membuat Document Term Matrix (DTM): tahap ini merupakan tahap pembobotan kata menggunakan perintah Document Term Matrix pada RStudio. Sebelum masuk ke tahap DTM, dilakukan pendefinisian fungsi menggunakan perintah *corpus*. *Corpus* memiliki fungsi untuk membaca data dari suatu dataset. Berikut ini adalah penggalan perintah untuk mem buat Document Term Matrix.
3. Pelabelan Data Latih dan Data Uji: ada tahap pelabelan data akan dilakukan proses pelabelan untuk menentukan rentang data yang masuk kategori

data latih dan data uji. Proporsi pembagian data latih dan data uji adalah 70:30 persen dari jumlah data. Field yang digunakan adalah field sentimen karena pada penelitian ini yang menjadi pokok perhitungan adalah sentimen positif atau negatif. Pada pelabelan data latih proporsi data latih adalah 70 persen dari jumlah data yaitu dari data ke 1 sampai data ke 241. Pada pelabelan data uji proporsi data uji pada penelitian ini sebesar 30 persen dari jumlah dataset yaitu dari data ke 242 sampai data ke 344. Data uji merupakan data yang digunakan sebagai data pengujian sistem. Berikut penggalan perintah untuk melakukan proses pelabelan data uji serta mengetahui jumlah sentimen positif dan negatif serta probabilitasnya

4. Menentukan Frekuensi Kata: Klasifikasi menggunakan metode *Lexicon Based* pada 344 data *tweets* menghasilkan sentimen positif sebanyak 290 *tweets* dan sentimen negatif sebanyak 54 *tweets*. Menentukan frekuensi kata merupakan tahap selanjutnya pada klasifikasi *Naïve Bayes Classifier* setelah melakukan pelabelan data. Pada tahap ini, dilakukan pencarian frekuensi kata-kata yang sering muncul pada data latih dengan batasan minimal kemunculan adalah 5 kali. Proses penentuan frekuensi kata dapat dilakukan dengan menuliskan perintah berikut pada RStudio.
5. Pelatihan Set Data Latih Menggunakan *Naïve Bayes Classifier*: Tahap pelatihan set data latih menggunakan *Naïve Bayes Classifier* termasuk dalam tahapan selanjutnya yaitu proses pembelajaran. Data masukan yang digunakan adalah data latih. Proses pembelajaran set data latih menggunakan algoritma *Naïve Bayes Classifier* terdiri dari 2 buah perhitungan, yaitu perhitungan untuk mencari probabilitas kemunculan label sentimen dan perhitungan untuk mencari probabilitas kemunculan tiap term untuk tiap klasifikasi. Proses latih dapat dituliskan dengan persamaan (1) berikut:

$$P(Label_1) = \frac{jumlah\ Label_1}{Panjang\ Dala\ Latih} \quad (1)$$

$$P(Term_n | Label_i) = \frac{fkwensi\ kemunculan\ Term_n\ pada\ teks\ dengan\ Label_i}{jumlah\ Label_1} \quad (2)$$

Pada persamaan (1) P didefinisikan sebagai probabilitas dan $Label_1$ adalah label-label yang digunakan pada penelitian ini dengan i adalah “positif” atau “negatif”. Adapun Termn adalah term ke-n yang terdapat pada set data latih. Untuk eksekusi program pada RStudio, dibutuhkan package “e1071”. Package “e1071” digunakan untuk implementasi *Naïve Bayes Classifier* pada bahasa pemrograman R.

1. Pengujian Menggunakan *Naïve Bayes Classifier*: Pada tahap pengujian ini akan dilakukan proses inti dari penelitian ini, yakni proses pengujian data uji berdasarkan hasil dari proses pelatihan menggunakan metode *Naïve Bayes Classifier*.

```
> conf.mat <- confusionMatrix(shopeepaylater_test,
> conf.mat
Confusion Matrix and Statistics

          Reference
Prediction Negatif Positif
Negatif      8      9
Positif      9     77

              Accuracy : 0.8252
              95% CI   : (0.7379, 0.893)
              No Information Rate : 0.835
              P-Value [Acc > NIR] : 0.6638

              Kappa : 0.3659

              Mcnemar's Test P-value : 1.0000

              Sensitivity : 0.47059
              Specificity : 0.89535
              Pos Pred Value : 0.47059
              Neg Pred Value : 0.89535
              Prevalence : 0.16505
              Detection Rate : 0.07767
              Detection Prevalence : 0.16505
              Balanced Accuracy : 0.68297

              'Positive' Class : Negatif
```

Gambar 5: Hasil Akurasi Sistem Menggunakan *Naïve Bayes Classifier*

2. Akurasi Sistem: Hasil akurasi dari sistem pengujian menggunakan fungsi Confusion Matriks adalah sebesar 82.52%. Gambar 5 adalah akurasi dari sistem pengujian menggunakan fungsi ConfusionMatrix.

Tampilan Website

Setelah melakukan tahap analisis dan perancangan, tahap berikutnya adalah merepresen-

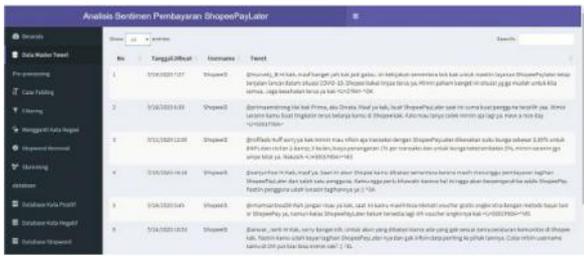
tasikan hasil dari tahap-tahap yang sudah dilakukan sebelumnya. Tahap ini menjelaskan hasil dari setiap tahapan proses analisis sentimen dan hasil akurasi sentimen yang divisualisasikan ke dalam tampilan halaman berbasis web server dengan menggunakan paket Shiny dari perangkat lunak RStudio.

1. Tampilan Halaman Beranda Halaman Beranda merupakan halaman awal yang akan muncul saat program Shiny dijalankan. Pada halaman ini ada memunculkan pesan selamat datang dan informasi mengenai penelitian apa yang dibuat, seperti Gambar 6.



Gambar 6: Tampilan Halaman Beranda

2. Tampilan Halaman Data Master Tweet Halaman data master tweet menampilkan halaman data master tweet menampilkan master data dari penelitian ini yang telah disimpan dalam bentuk file berekstensi .csv. Field data berisi hasil dari pengambilan data (*crawling*). Field yang digunakan yaitu no, tanggal dibuat, username, dan tweet yang akan dijadikan data untuk dilakukan proses *pre-processing*. Gambar 7 adalah tampilan halaman data master tweet.



Gambar 7: Tampilan Halaman Data Master Tweet

3. Tampilan Halaman Case folding Halaman case folding menampilkan data asli sebelum dan sesudah dilakukan proses case folding pada data master tweet. Proses case folding menggunakan fungsi tolower yang membuat string berbentuk kapital berubah menjadi string huruf kecil atau tidak kapital. Data berisi field dari Data Master Tweet dengan data dari hasil Case Folding yang telah disimpan dalam bentuk file berekstensi .csv. Field yang ditampilkan yaitu no, Data Master Tweet dan Case Folding. Gambar 8 adalah tampilan halaman case folding.



Gambar 8: Tampilan Halaman Case folding

4. Tampilan Halaman Filtering Pada halaman filtering akan menampilkan data filtering yaitu elemen-elemen dari teks sumber yang dihapus tidak diperlukan, menghilangkan tweet ganda dan melakukan standarisasi untuk setiap teks sumber yang telah disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu no, dan hasil dari tahap filtering ini akan dijadikan data untuk ke tahap proses *pre-processing* berikutnya. Gambar 9 adalah tampilan halaman filtering.



Gambar 9: Tampilan Halaman Filtering

5. Tampilan Halaman Mengganti Kata Negasi Pada halaman ini akan menampilkan data yang telah dilakukan proses filtering, kemudian akan dilakukan proses mengganti kata negasi yang terdapat pada data yang telah disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu no, hasil dari tahap filtering dan hasil dari tahap mengganti kata negasi. Selanjutnya hasil dari tahap mengganti kata negasi akan dijadikan data untuk ke tahap proses *pre-processing* berikutnya. Gambar 10 adalah tampilan halaman mengganti kata negasi.



Gambar 10: Tampilan Halaman Mengganti Kata Negasi

6. Tampilan Halaman *Stopword removal* Halaman *stopword removal* akan menampilkan hasil data tweet yang sudah dihilangkan kata negasinya dan daftar tweet yang mengandung kata yang terdapat pada daftar kata tidak perlu, maka kata tersebut akan dihapus kemudian file tersebut disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu no, hasil dari tahap mengganti kata negasi dan hasil dari tahap *stopword removal*. Selanjutnya hasil dari tahap *stopword removal* akan dijadikan data untuk ke tahap *pre-processing* berikutnya. Gambar 11 adalah tampilan halaman *stopword removal*.



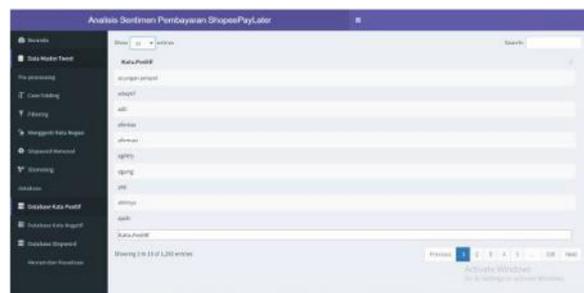
Gambar 11: Tampilan Halaman *Stopword removal*

7. Tampilan Halaman *Stemming* Pada halaman *stemming* akan menampilkan data yang telah dilakukan proses *stopword removal*, kemudian akan dilakukan proses mengubah suatu kata menjadi bentuk kata dasarnya menggunakan algoritma *stemming* Nazief dan Andriani yang telah disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu no, hasil dari tahap *stopword removal* dan hasil dari tahap *stemming*. Gambar 12 adalah tampilan halaman *stemming*.



Gambar 12: Tampilan Halaman *Stopword removal*

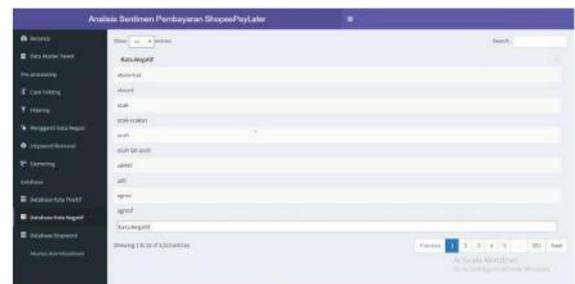
8. Tampilan Halaman Database Kata Positif Pada halaman database kata positif akan ditampilkan daftar kata yang masuk kategori kata positif. Daftar kata tersebut digunakan pada metode *Lexicon Based*. Kata Positif disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu isi dari kata positif.



Gambar 13: Tampilan Halaman Database Kata Positif

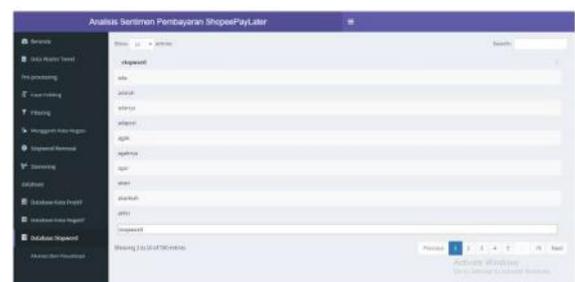
Gambar 13 adalah tampilan halaman database kata positif

9. Tampilan Halaman Database Kata Negatif Pada halaman database kata negatif akan ditampilkan daftar kata yang masuk kategori kata negatif. Daftar kata tersebut digunakan pada metode *Lexicon Based*. Kata Negatif disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu isi dari kata negatif. Gambar 4.14 adalah tampilan halaman database kata negatif.



Gambar 14: Tampilan Halaman Database Kata Negatif

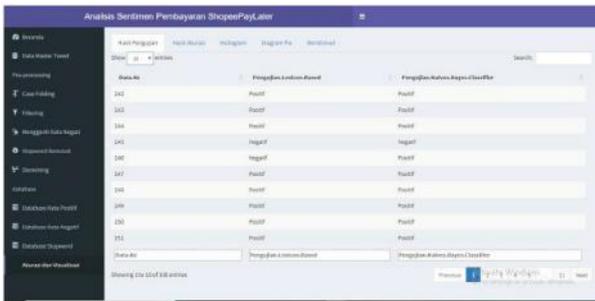
10. Tampilan Halaman Database Stopword Pada halaman ini akan ditampilkan daftar kata yang masuk kategori stopword yaitu kata tidak terpakai, sehingga apabila pada setiap data ulasan mengandung kata yang terdapat pada daftar kata tidak perlu, maka kata tersebut akan dihapus. Database stopword disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu isi dari stopword. Gambar 15 adalah tampilan halaman database stopword.



Gambar 15: Tampilan Halaman Database Stopword

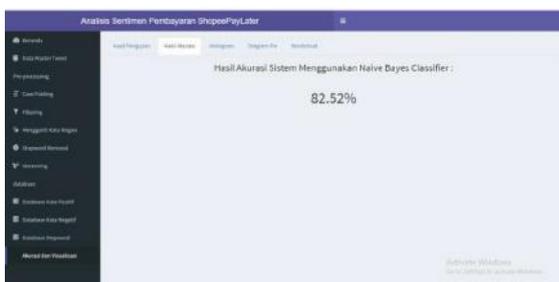
11. Tampilan Halaman Hasil Pengujian Halaman hasil pengujian akan menampilkan perbandingan hasil klasifikasi sistem

menggunakan metode *Naïve Bayes Classifier* pada data uji dengan proses klasifikasi yang dilakukan menggunakan metode *Lexicon Based*. kemudian file tersebut disimpan dalam bentuk file berekstensi .csv. Field yang digunakan yaitu dataKe untuk pengujian *Lexicon Based* dan pengujian *Naïve Bayes Classifier*. Gambar 16 adalah tampilan halaman hasil pengujian



Gambar 16: Tampilan Halaman Hasil Pengujian

12. Tampilan Halaman Hasil Akurasi Halaman hasil akurasi akan menampilkan informasi hasil akurasi sistem klasifikasi yang telah dihitung menggunakan metode *Naïve Bayes Classifier* terhadap proses klasifikasi menggunakan metode *Lexicon Based*. Gambar 17 adalah gambar informasi akurasi hasil klasifikasi sistem menggunakan metode *Naïve Bayes Classifier* yaitu sebesar 82,52%.



Gambar 17: Tampilan Halaman Hasil Akurasi

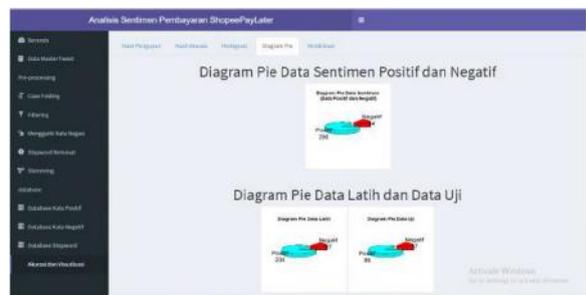
13. Tampilan Halaman Visualisasi Histogram Halaman visualisasi histogram akan menampilkan gambar visualisasi data dalam bentuk histogram dengan cara

menyimpan gambar pada folder yang sama dengan program sehingga gambar dapat terpanggil dan tampil pada sistem. Gambar 18 adalah gambar visualisasi data dalam bentuk histogram.



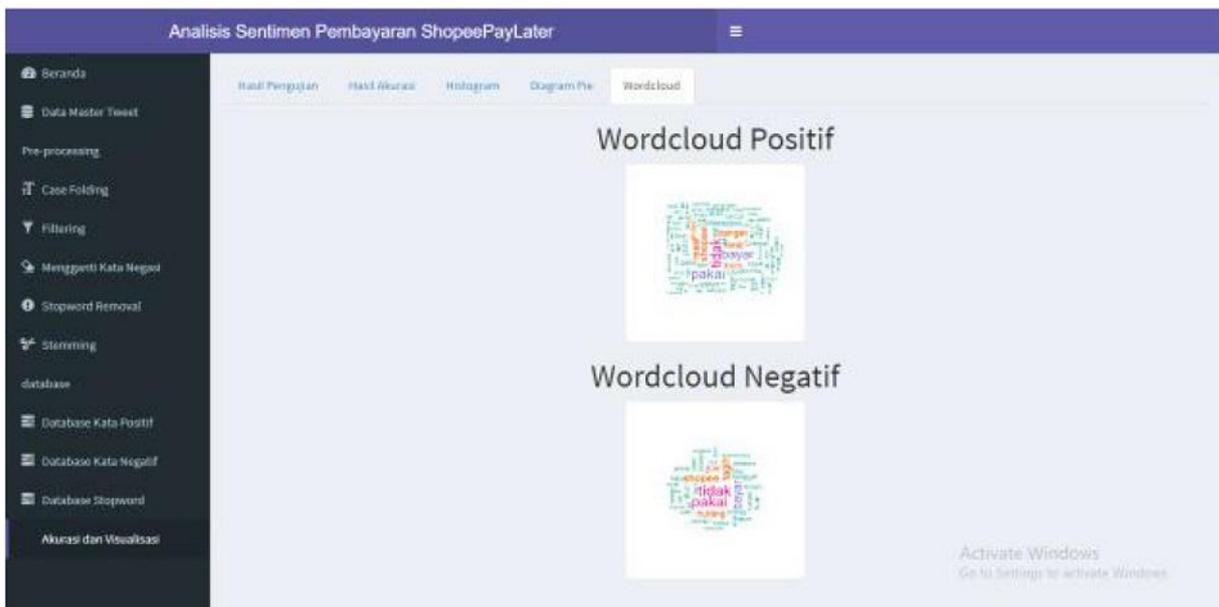
Gambar 18: Tampilan Halaman Visualisasi Histogram

14. Tampilan Halaman Visualisasi Diagram Pie Halaman visualisasi diagram pie akan menampilkan gambar visualisasi data dalam bentuk diagram pie dengan cara menyimpan gambar pada folder yang sama dengan program sehingga gambar dapat terpanggil dan tampil pada sistem. Gambar 19 adalah gambar visualisasi data dalam bentuk diagram pie.



Gambar 19: Tampilan Halaman Visualisasi Diagram Pie

15. Tampilan Halaman Visualisasi Wordcloud Halaman visualisasi wordcloud akan menampilkan gambar visualisasi data dalam bentuk wordcloud dengan cara menyimpan gambar pada folder yang sama dengan program sehingga gambar dapat terpanggil dan tampil pada sistem, lihat Gambar 20.



Gambar 20: Tampilan Halaman Visualisasi *Wordcloud*

Penutup

Berdasarkan penelitian dan hasil uji coba aplikasi yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Penelitian analisis sentimen Twitter terhadap pembayaran ShopeePayLater pada aplikasi belanja *online* (Shopee) berhasil dilakukan menggunakan bahasa R dengan IDE RStudio dengan koneksi terhadap Twitter Application Programming Interface (API).
2. Sistem yang dibuat terdiri dari 2 (dua) bagian yang berbeda yakni sistem perhitungan analisis sentimen dan website analisis sentimen. Proses analisis sentimen diawali dari pengambilan data (*crawling*) pada Twitter, selanjutnya yaitu tahap text *pre-processing* yang terdiri dari case folding, filtering, mengganti kata negasi, *stopword removal*, dan *stemming*.
3. Proses klasifikasi dilakukan menggunakan metode *Lexicon Based* untuk menentukan kalimat supaya memiliki makna sentimen positif atau negatif, dan *Naïve Bayes Classifier* untuk menghitung akurasi antara prediksi dan hasil actual dari sistem.
4. Hasil analisis sentimen divisualisasikan dalam bentuk histogram, diagram *pie* dan *wordcloud*.

5. Hasil akurasi klasifikasi data *tweets* dari jumlah data *tweets* sebanyak 344 *tweets* dengan data latih sebanyak 241 *tweets* dan data uji sebanyak 103 *tweets* yaitu sebesar 82,52 % .

6. Hasil analisis sentimen pada penelitian ini menunjukkan bahwa pembayaran menggunakan ShopeePayLater memiliki kecenderungan sentimen positif pada saat data *tweets* diambil.

Penelitian analisis sentimen yang telah dilakukan ini masih memerlukan pengembangan seperti pengambilan data untuk analisis sentimen, proses pengolahan, perhitungan hingga hasil visualisasinya bisa dikerjakan secara real-time dalam suatu sistem website agar hasil akurasi dapat langsung terlihat apabila data berganti. Selain itu dapat melakukan penanganan terhadap misspelling atau singkatan karena dapat mempengaruhi hasil akurasi.

Daftar Pustaka

- [1] Walaa Medhat, A. Hasan, & H. Korashy, "Sentiment Analysis algorithm and applications : A survey", Ains Shans Engineering Journal Cairo Campus of CBU, Egypt, 2014.
- [2] A. Go, L. Huang & R. Bhayani, "Twitter Sentiment Analysis", Final Project Report, Stanford University, California, 2009.

- [3] D. Moeljadi, T. A. Le, T. A. Miura & T. Ohkuma, "Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian *Tweets*", Division of Linguistics and Multilingual Studies Nanyang Technological University, Singapore, 2016.
- [4] M. Joshi & H. Vala, "Opinion Mining For Sentiment Data Classification", International Journal of Research in Information Technology, India, 2014.
- [5] Susanti Fransiska & Yolanda, "Analisis Sentimen Twitter Untuk Review Film Menggunakan Algoritma *Naïve Bayes Classifier* (NBC) Pada Sentimen R Programming", Jurnal Siliwangi, Tasikmalaya, 2019.
- [6] Brata Mas Pintoko & Kemas Muslim L, "Analisis Sentimen Jasa Transportasi *Online* pada Twitter Menggunakan Metode *Naïve Bayes Classifier*", e-Proceeding of Engineering, Bandung, 2018.
- [7] P. Antinasari, R. Perdana & M. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku", Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Malang 2017.
- [8] Fahrur Rozi Imam., Hadi Pramono Sholeh & Achmad Dahlan Erfan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi", Jurnal EECCIS, Malang, 2012.
- [9] J. Han and M. Kamber, "Data Mining Concepts and Techniques", Second Edition, California: Morgan Kaufman, 2006.