

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP DOMPET ELEKTRONIK DENGAN METODE LEXICON BASED DAN K – NEAREST NEIGHBOR

¹Siti Saidah, ²Joanna Mayary

^{1,2} Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma

^{1,2} Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

¹sitisaidah@staff.gunadarma.ac.id, ²joannamay@studentsite.gunadarma.ac.id

Abstrak

Analisis sentimen telah menempatkan peran sebagai alat bisnis intelijen. Kemampuan Analisis sentimen antara lain mengekstrak opini publik tentang topik tertentu, produk atau jasa yang di dalamnya terkandung teks-teks yang tidak terstruktur. Analisis sentimen merupakan proses untuk memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi yang akurat, penerapannya pada transaksi elektronik mempengaruhi kebutuhan gaya hidup konsumen dan aktivitas berbelanja. Dompot elektronik adalah suatu metode pembayaran yang terbaru di Indonesia. Dompot elektronik diistilahkan sebagai e-wallet, dompot digital, digital wallet atau electronic wallet. Penelitian ini mengadopsi analisis sentimen dengan metode Lexicon Based dan K-Nearest Neighbor. Tujuan penelitian ini adalah melakukan visualisasi sentimen terhadap dompot digital berdasarkan opini pengguna twitter dengan pengklasifikasian menggunakan Lexicon - based dan K-Nearest Neighbor. Tahapan penelitian terdiri dari analisis sumber data, preprocessing data, klasifikasi sentimen dengan metode Lexicon Based untuk menentukan sentimen positif dan negatif, klasifikasi dengan metode K-Nearest Neighbor untuk perhitungan akurasi dengan confusion matrix, visualisasi wordcloud dan histogram serta tampilan menggunakan packages shiny RStudio. Metode Lexicon Based menggunakan data OVO sebanyak 357 tweets, data GOPAY sebanyak 337 tweets, dan data LinkAja sebanyak 255 tweets. Hasil akhir perhitungan metode K-Nearest Neighbor dengan confusion matrix untuk tweet OVO diperoleh nilai akurasi positif sebesar 86,91% dan nilai akurasi negatif 13,09%, tweet Gopay diperoleh nilai akurasi sebesar 94,05% dan nilai akurasi negatif 5,95%, serta tweet LinkAja diperoleh nilai akurasi sebesar 76,31% dan nilai akurasi negative 23,69%.

Kata Kunci : Analisis, confusion matrix, K-Nearest Neighbor, Lexicon Based, sentimen.

Abstract

Sentiment analysis has placed its role as a business intelligence tool. Capability Analysis sentiments include extracting public opinion about a particular topic, product or service which contains unstructured texts. Sentiment analysis is a process for understanding, extracting, and processing textual data automatically to get accurate information, its application to electronic transactions affects the lifestyle needs of consumers and shopping activities. Electronic wallet is the newest payment method in Indonesia. Electronic wallet is termed as e-wallet, digital wallet, digital wallet or electronic wallet. This study adopts sentiment analysis using the Lexicon Based and K-Nearest Neighbor methods. The purpose of this study is to visualize sentiments on digital wallets based on Twitter user opinion by classifying using Lexicon-based and K-Nearest Neighbor Stages of research consists of data source analysis, preprocessing data, sentiment classification with Lexicon Based methods to determine positive and negative sentiments, classification with the K-Nearest Neighbor method for calculating accuracy with a confusion matrix, wordcloud visualization and histogram and display using RStudio shiny packages. The Lexicon Based method uses OVO data of 357 tweets, GOPAY data of 337

tweets, and LinkAja data of 255 tweets. The final result of K-Nearest Neighbor method calculation with confusion matrix for OVO tweets obtained a positive accuracy value of 86.91% and a negative accuracy value of 13.09%, Gopay tweets obtained a positive accuracy value of 94.05% and a negative accuracy value of 5.95%, and LinkAja tweets obtained a positive accuracy value of 76.31% and a negative accuracy value of 23.69%.

Keywords : *Analysis, confusion matrix, K-Nearest Neighbor, Lexicon Based, sentiment.*

PENDAHULUAN

Dompot elektronik merupakan metode pembayaran yang menjadi topik yang sedang tren di Indonesia. Penilaian masyarakat terhadap kenyamanan cara pembayaran melalui dompet elektronik ini dapat diperoleh melalui media sosial, salah satunya melalui *twitter*. *Twitter* merupakan utilitas yang memungkinkan pengguna mengirim SMS di seluruh dunia dengan mengirim *tweet* yang terdiri dari 140 karakter menjadikan pengguna dapat mengekspresikan banyak pendapat, informasi, dan kegiatan sehari-hari.

Data yang dirilis oleh *Twitter* Indonesia [1] Negara Indonesia merupakan salah satu dari 5 negara terbesar dalam penggunaan *Twitter* secara aktif, yakni 77% dari pengguna *Twitter* Indonesia merupakan pengguna aktif. Berbeda dengan sosial media lainnya, *Twitter* memiliki keterbukaan terhadap data yang dimilikinya melalui API (*Application Programming Interface*). Ketersediaan *Twitter* API yang memungkinkan pengembangan pihak ketiga untuk membuat program yang menggabungkan layanan *Twitter*. API merupakan antarmuka perangkat lunak ke perangkat lunak, dengan API maka aplikasi berbicara satu sama lain tanpa sepengetahuan

atau intervensi pengguna. API menyerupai *Software as a Service* (SaaS), karena kehadiran perangkat lunak ini memudahkan, sehingga tidak perlu memulai dari awal untuk menulis program [2].

Tweets yang dikumpulkan dan dianalisis disebut sebagai analisis sentimen. Analisis sentimen dipandang sebagai alat bisnis intelijen yang diinginkan, karena kemampuannya dapat mengekstrak opini publik tentang topik tertentu, produk atau jasa yang tertanam dalam teks-teks yang tidak terstruktur [3].

Penelitian ini membatasi analisis opini pengguna *twitter* terhadap dompet digital yang terdapat pada media sosial yaitu *twitter*, melakukan analisis *tweets* yang mengandung kalimat ‘promo dan OVO’ untuk analisis dompet elektronik OVO, melakukan analisis *tweets* yang mengandung kalimat ‘GOPAY dan promo’ untuk analisis dompet elektronik GOPAY dan melakukan analisis *tweets* yang mengandung kalimat ‘TCASH atau LinkAja’ untuk analisis dompet elektronik LinkAja. Data diolah dari masing-masing *tweets* sebanyak 500 *tweets* berbahasa Indonesia.

Penelitian terkait masalah di atas, fokus penelitian dilakukan terhadap implementasi algoritma K-Nearest Neighbor dalam peng-

klasifikasi follower twitter yang menggunakan Bahasa Indonesia memperoleh nilai akurasi terbesar yang didapat pada proses klasifikasi untuk empat kali pengujian adalah 68% dengan jumlah data uji yang digunakan akurasi dilakukan bertahap mulai dari mulai dari jumlah 25 data uji, dilanjutkan dengan penambahan data sebanyak 50 hingga 100 data uji dan data latih yang tersedia adalah 1054 data latih [4]. Analisis sentimen juga dapat digunakan untuk menilai kesukaan atau ketidaksukaan publik terhadap suatu barang atau jasa. Sentimen tersebut merupakan informasi tekstual yang bersifat subjektif dan memiliki polaritas positif dan negatif. Nilai polaritas ini dapat digunakan sebagai parameter dalam menentukan suatu keputusan [5]. Analisis sentimen dengan tingkat kepuasan pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia pada *Twitter* dengan menggunakan metode *Support Vector Machine* dan *Lexicon Based Features* menghasilkan nilai *accuracy* sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%. Nilai *recall* yang sangat besar tersebut dipengaruhi oleh banyaknya data uji yang sebenarnya positif dan terdeteksi sebagai positif oleh sistem [6].

Berdasarkan uraian di atas, maka dilakukan penelitian mengenai analisis sentimen pengguna twitter terhadap dompet digital dengan pengklasifikasiannya menggunakan metode *Lexicon – based* dan metode *K-Nearest Neighbor*. Metode *Lexicon - based* yang ditujukan untuk mengolah dan mengkla-

sifikasi opini, sehingga akan diketahui klasifikasi dari opini tersebut (positif atau negatif). Metode *K-Nearest Neighbor* digunakan untuk proses klasifikasi dalam sebuah penelitian karena memiliki kesederhanaan dimana prosesnya berdasarkan pada pendekatan pembobotan yang sederhana dan kemudahan dalam implementasi, adaptasi dan proses learning serta memiliki nilai akurasi yang tinggi [7]. *Twitter* sumber yang kaya untuk analisis sentimen dan penambahan kepercayaan, mengembangkan penggolong fungsional secara otomatis dan memperkenalkan dua metode dikenal sebagai algoritma klasifikasi sentimen (*SCA*) berdasarkan k-tetangga terdekat (*KNN*) dan yang lain didasarkan pada mesin vektor dukungan (*SVM*) [8]. Penelitian ini diawali dengan langkah menganalisis teks tanpa penggunaan *emoticon*, menerapkan kombinasi metode dari *Lexicon Based* dan metode *K-Nearest Neighbor* Analisis didukung oleh perangkat lunak *R-Programming* versi 3.5.1. Data diklasifikasikan oleh *Lexicon – based* dalam dua bentuk, yaitu bentuk positif dan negatif, sedangkan metode *K-Nearest Neighbor* digunakan untuk perhitungan akurasi dengan confusion matrix. Hasil dari analisis sentimen akan divisualisasikan dalam bentuk histogram dan wordcloud.

Tujuan dari penelitian ini adalah membuat website visualisasi sentimen terhadap dompet digital berdasarkan opini pengguna *twitter* dengan pengklasifikasi menggunakan *Lexicon Based* untuk mendapatkan

kelompok positif dan negatif dan metode *K-Nearest Neighbor* dibutuhkan untuk perhitungan akurasi dengan *confusion matrix*, serta visualisasi hasil analisis dalam bentuk *histogram* dan *wordcloud*.

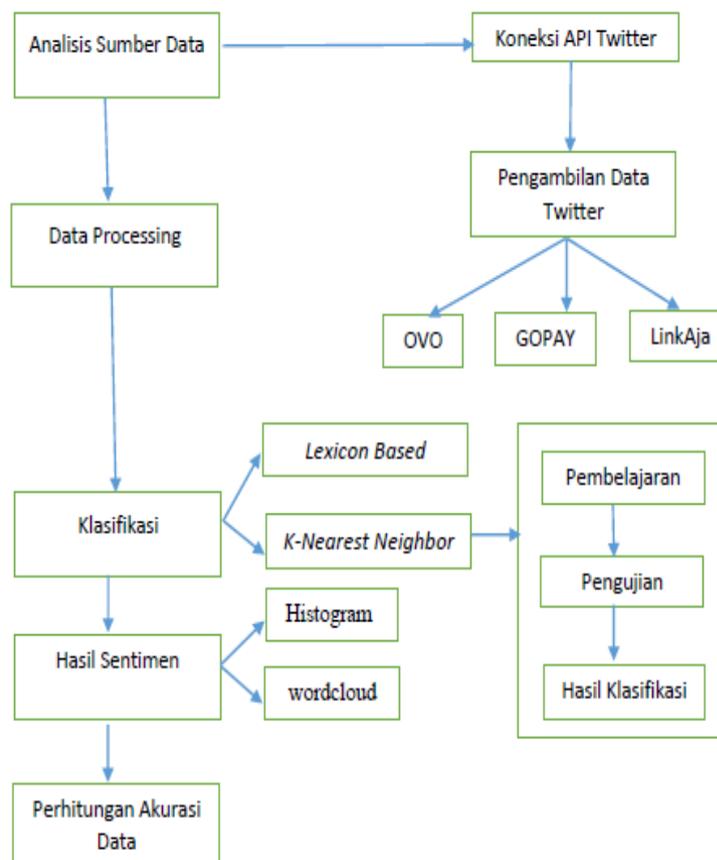
METODE PENELITIAN

Tahapan Penelitian

Gambar 1 di bawah menjelaskan tahapan penelitian analisis sentimen pengguna twitter terhadap dompet elektronik yang diamati, secara garis besar pengambilan data analisis sentimen terhadap *tweets* yang bermuatan dompet elektronik media sosial

Twitter OVO, Gopay dan LinkAja dengan menggunakan API *Twitter* sebagai data uji yang akan dianalisis.

Data uji akan dibersihkan terlebih dahulu dari entitas-entitas yang dapat mengganggu proses analisis, setelah proses pembersihan dilakukan dilanjutkan dengan proses pengklasifikasi menggunakan metode *Lexicon Based*, yaitu dengan cara mencocokkan data uji dengan data sampel, setelah itu masing-masing kata dijumlahkan berdasarkan sentimen positif yang bermuatan dengan nilai plus satu (+1) dan sentimen negatif yang bermuatan dengan nilai minus satu(-1).

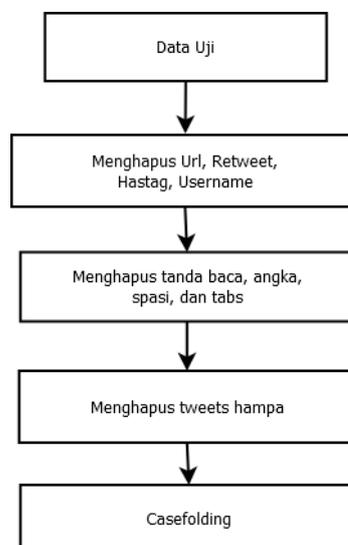


Gambar 1. Tahapan Penelitian Analisis Sentimen Pengguna *Twitter* terhadap Dompet Elektronik

Hasil akhir berupa skor dari masing-masing data uji, skor lebih besar atau sama dengan satu bernilai positif, lebih kecil atau sama dengan minus satu bernilai negatif. Data yang telah diuji akan dilanjutkan ke proses pengklasifikasian menggunakan metode *K-Nearest Neighbor*. Tahapan data uji akan dicocokkan dengan data sampel yang telah dibuat melakukan pembuatan model dan perhitungan berdasarkan metode *K-Nearest Neighbor* dengan mengumpukan pada sentimen yang dikenal dari training data ke dalam model data, lakukan proses perhitungan nilai akurasi dengan menggunakan model *confusion matrix*, di akhir proses dibuat visualisasi berupa *histogram* dan *wordcloud*. Tahapan Analisis Sumber Data merupakan tahap awal pelaksanaan dibagi menjadi beberapa tahap, dimulai dari pengambilan data uji dari *Twitter* diambil dengan menggunakan koneksi terhadap *Application Programming Interface (API) Twitter*, membuat koneksi *API Twitter* dengan

menggunakan package *Twitter*, lalu dilanjutkan ke proses pengambilan data dari *twitter* melalui proses koneksi terhadap *API Twitter* disesuaikan dengan pengaturan library yang tersedia.

Tahapan Data Preprocessing yang dimaksud adalah data uji yang dilakukan terhadap *tweets*, dengan cara membersihkan berbagai komponen yang tidak berguna dan mengganggu proses analisis, dapat dijelaskan pada Gambar 2. Penjelasan alur data *Preprocessing* pada Gambar 2 dimulai dari proses pembersihan dengan bantuan perangkat lunak R dan mengaktifkan fungsi *gsub()*, sehingga dapat digunakan untuk menghapus komponen yang tidak berguna seperti menghapus alamat situs atau *url*, menghapus *Retweet*, menghapus *username*, menghapus angka, mengubah kata negasi, menghapus *Tweets* hampa, menghapus *Tweets* berulang dan melakukan *Case Folding*, yaitu penyeragaman seluruh karakter dengan huruf kecil.

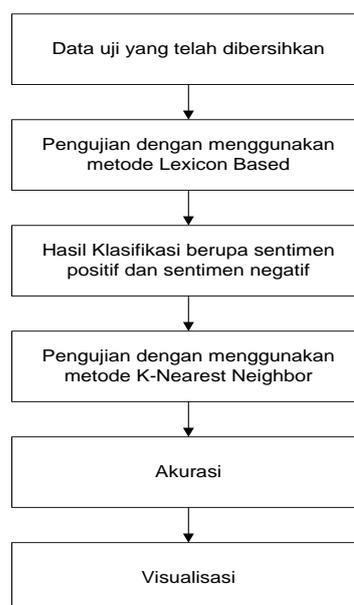


Gambar 2. Alur Data *Preprocessing*

Tahap klasifikasi sentimen adalah memproses data uji menggunakan dua metode, yaitu dengan *Lexicon Based* dan *K-Nearest Neighbor*. Alur proses pengujian ditunjukkan Gambar 3 yang menjelaskan bahwa data uji yang telah dibersihkan akan diklasifikasikan dengan dua metode yaitu metode *Lexicon Based* dan metode *K-Nearest Neighbor*. Data uji klasifikasi dengan metode *Lexicon Based* memberikan hasil klasifikasi berupa sentimen positif dan sentimen negatif, tahapan berikutnya adalah data diklasifikasi menggunakan metode *K-Nearest Neighbor* dan menghasilkan akurasi. Hasil dari klasifikasi data akan ditampilkan dalam bentuk *wordcloud* dan histogram.

Klasifikasi dengan *Lexicon Based* akan membandingkan masing-masing data uji terhadap data kata sentimen Bahasa Indonesia

dan akan menghasilkan sebuah nilai sentimen. Data kata sentimen berbahasa Indonesia diperoleh berdasarkan kata yang mengandung sentimen yang diambil dalam Kamus Besar Bahasa Indonesia (KBBI) dan beberapa kata tidak baku yang sering digunakan. Data kata sentimen diperoleh dengan cara mengunduh dari github.com dan mengetik secara manual. Data yang sudah diperoleh, kemudian digabung antara data kata sentimen unduh dari github.com dan cara manual. Proses pemasukan berdasarkan kata sifat (*adjective*) dipisahkan menurut pengelompokan lexicon negatif, menyaring kata bermakna negatif, sebagai contoh alasan, aneh, beban, merugikan, hilang, dan sebagainya. Lexicon positif, menyaring kata bermakna positif, sebagai contoh semangat, gratis, baik, patuh, unggul dan sebagainya.



Gambar 3. Alur Proses Klasifikasi Data

Klasifikasi dengan metode *K-Nearest Neighbor* memiliki 3 tahap yang terkandung dalam algoritma *K-Nearest Neighbor* yaitu tahap pembelajaran, tahap pengujian dan hasil sentimen. Tahap Pembelajaran mengolah data term matrix yaitu menghitung jumlah kosakata (term) pada masing-masing dokumen (*tweets*) dalam dataset, proses berikutnya adalah data training yaitu data yang digunakan sebagai data latih untuk sistem belajar [9]. Data training adalah data hasil dari pengklasifikasi metode *Lexicon Based*.

Data training diambil 70% secara acak dari dataset setelah pengklasifikasi metode *Lexicon Based*, untuk data training OVO terdapat 357 *tweets*, data training Gopay terdapat 337 *tweets* dan data training LinkAja terdapat 255 *tweets*, proses akhir dilakukan terhadap data test atau data yang digunakan sebagai data uji sistem. Data test merupakan hasil dari pengklasifikasi metode *Lexicon Based*. Data diambil dari sisa dataset yang bukan merupakan menjadi data training, untuk data test OVO terdapat 107 *tweets*, data test Gopay terdapat 101 *tweets* dan data test LinkAja terdapat 76 *tweets*. Tahap pengujian ini menjelaskan bagaimana sistem menggunakan algoritma *K-Nearest Neighbor* terhadap suatu *tweets*. *K-Nearest Neighbor* akan memproses data yang dihasilkan dari pengklasifikasi metode *Lexicon Based*. Klasifikasi metode *Lexicon Based* yang telah selesai dilakukan, dilanjutkan dengan membuat matriks istilah dokumen term-matrix (DTM), model *K-Nearest Neighbor*

membutuhkan tiga set data yaitu train, test dan *classifier*. Ketiga set tersebut harus memiliki jumlah baris yang sama, membuat data train yang diambil sebanyak 70% dari dataset.

Data test akan menyimpan semua sisa data atau 30% dari dataset, membuat *classifier*, menggunakan data dalam model *K-Nearest Neighbor*, proses *confusion matrix*, yaitu matriks yang memberi tahu dokumen yang diprediksi model dengan benar dan dokumen apa yang tidak diprediksi dengan benar, langkah terakhir menghitung akurasi model *confusion matrix*. Perhitungan dilanjutkan ke proses untuk mengetahui tingkat kemiripan antar dua dokumen menggunakan persamaan *similarity* dengan mencari hasil dari jarak terdekat. Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data latih, dan nilai k harus ganjil dan lebih dari satu. Dekat atau jauhnya jarak data latih yang paling dekat dengan objek yang akan diklasifikasi dapat dihitung dengan menggunakan metode *cosine similarity* [4]. *Cosine similarity* berfungsi untuk menguji ukuran yang dapat digunakan sebagai interpretasi kedekatan jarak berdasarkan kemiripan dokumen. Persamaan 1 digunakan untuk menghitung kemiripan kasus antar 2 objek [10].

$$\text{Similarity (q,d)} = \frac{\sum_{j=1}^t (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij})^2 \cdot \sum_{j=1}^t (d_{ij})^2}} \quad (1)$$

Keterangan :

q_{ij} = bobot istilah j pada dokumen i

d_{ij} = bobot istilah j pada dokumen i

Berikut diberikan contoh dokumen untuk menghitung similaritas

Term	go pay	promo	jan	sepunya	langsung	cash back
1	1	1	1	1	1	1
X	1	1	0	0	0	0

Hasil kemiripan dihitung dengan menggunakan persamaan (1) menghasilkan nilai 0,25. Jarak yang sudah diketahui dari nilai tertinggi sampai terendah, akan diambil sebanyak k data tertinggi, dari k data tersebut akan dilihat nilai sentimen mana yang paling banyak muncul sehingga kelas sentimen yang paling banyak muncul kelas/nilai sentimen sebagai data yang sedang dihitung [11].

Tujuan dari algoritma ini adalah mengklasifikasikan objek berdasarkan atribut dan training sample. *Classifier* tidak menggunakan apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah k objek atau titik *training* yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k objek Algoritma *K-Nearest Neighbor* (KNN) menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru [12].

Hasil klasifikasi dari keseluruhan proses diperoleh sebagai berikut; data test

OVO dengan klasifikasi menggunakan *K-Nearest Neighbor* terdapat 2 *tweets* dengan 16 sentimen negatif dan 91 review dengan sentimen positif, data test GOPAY dengan klasifikasi menggunakan *K-Nearest Neighbor* terdapat 2 *tweets* dengan 11 sentimen negatif dan 90 review dengan sentimen positif, data test LinkAja dengan klasifikasi menggunakan *K-Nearest Neighbor* terdapat 2 *tweets* dengan 25 sentimen negatif dan 51 review dengan sentimen positif. Hasil akhir analisis sentimen dari pengolahan data proses pengklasifikasian analisis sentimen berdasar-kan metode *Lexicon Based* dan *K-Nearest Neighbor* ke dalam bentuk website dilengkapi dengan tampilan hasil sentimen dalam bentuk *wordcloud* dan *histogram*.

HASIL DAN PEMBAHASAN

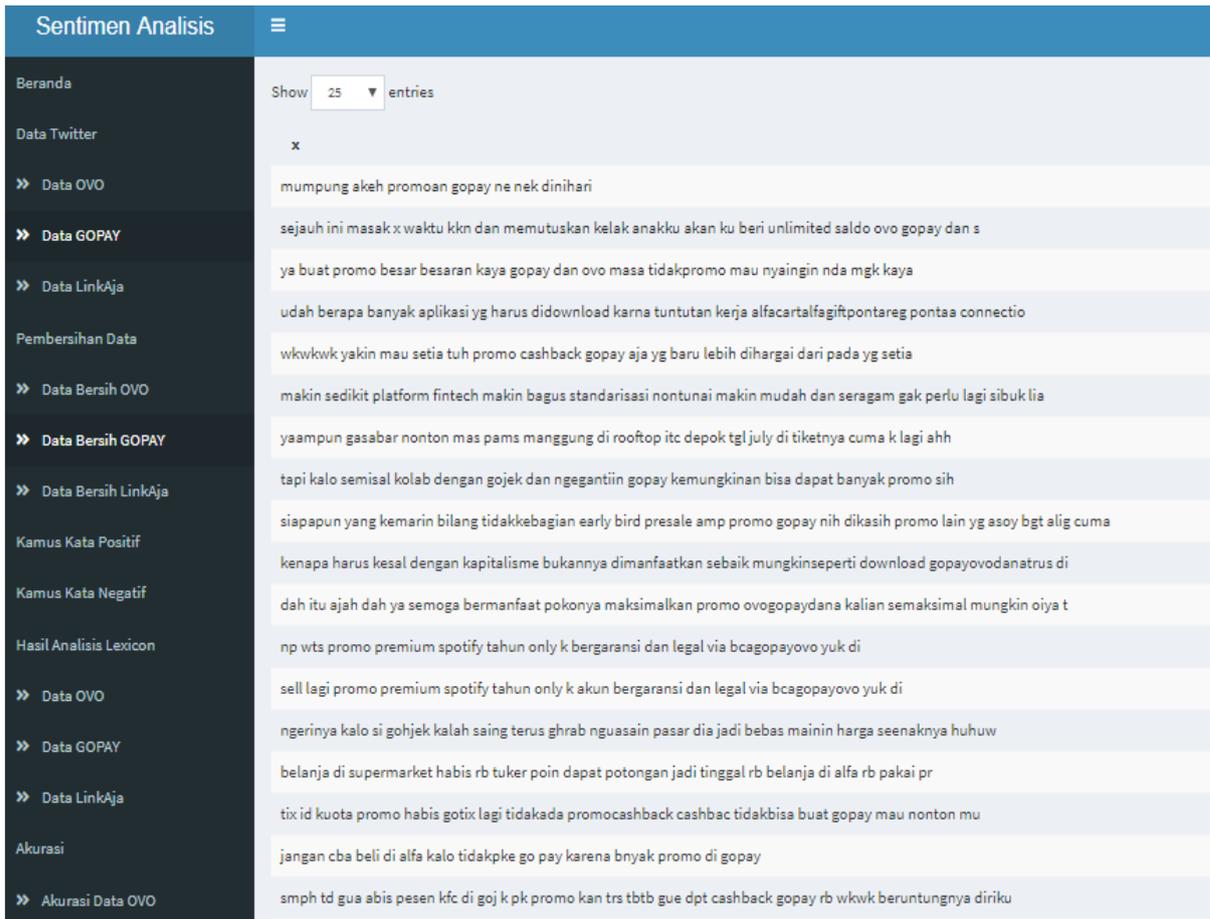
Website sebagai hasil akhir dari semua tahapan yang telah dilakukan ditunjukkan pada halaman beranda dengan menu yang terdiri dari beranda, data twiter, kamus kata positif, kamus kata negatif, hasil lexicon, akurasi, tampilan *wordcloud* dan tampilan *histogram*, serta terdapat sub menu yaitu data OVO, data GOPAY, data LinkAja, data bersih OVO, data bersih GOPAY, data bersih LinkAja, akurasi data OVO, akurasi data GOPAY, akurasi data LinkAja, *wordcloud* data OVO, *wordcloud* data GOPAY, *wordcloud* data LinkAja. Halaman data twitter OVO, GOPAY dan LinkAja memiliki tampilan yang sama, yang membedakan

adalah isi datanya, ssebagai salah satu contoh akan diilustrasikan dataset yang digunakan pada proses analisis sentimen dari GOPAY. Tampilan dataset berbentuk tabel hasil dari pemanggilan data twitter GOPAY. Tampilan halaman data twitter GOPAY dapat dilihat pada Gambar 4. Dataset pada Gambar 4 di atas memiliki beberapa komponen yang harus dibersihkan untuk membuat dataset yang lebih efisien. Hasil pembersihan dataset yang berawal 500 *tweets* menjadi 381 *tweets*. Tampilan halaman pembersihan data twitter GOPAY dapat dilihat pada Gambar 5. Tampilan halaman pembersihan pada gambar

5 di atas menjelaskan data twitter GOPAY sudah menggunakan huruf kecil, tidak mengandung angka, tidak ada alamat situs atau *url*, tidak ada *tweet* yang sama, tidak ada *username* dan tidak ada *tweet* kosong. Banyaknya kata yang dimasukkan ke dalam kamus positif dan kamus negatif bergantung pada komentar yang ada pada *tweet* dari pengguna, sehingga jumlah kata yang ada pada kamus positif dan kamus negatif untuk produk baru seperti OVO dan LinkAja memiliki kecenderungan lebih sedikit daripada produk Gopay.

X	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource
1	@olvrin Mumpung akeh promo an gopay ne nek dinihari <U+0001F605>	FALSE	0	olvrin	2019-07-02 17:47:34	FALSE	1.146112e+18	1.146113e+18	7.322176e+17	Twitter for Android
2	#hikmahKkn sejauh ini masak 4c waktu KKN dan memutuskan, kelak anakku akan ku beri unlimited saldo ovo, gopay dan s... https://t.co/M4ON2MvvDG	FALSE	0		2019-07-02 16:56:42	TRUE		1.146100e+18		Twitter for Android
3	@detikcom @detikfinance ya buat promo besar besaran kaya gopay dan ovo. masa tanpa promo mau nyaingi? nda mgk kaya... https://t.co/2mM1T0Jyf	FALSE	0	detikcom	2019-07-02 15:28:57	TRUE	1.146077e+18	1.146078e+18	6.918318e+07	Twitter for iPhone
4	Udah berapa banyak aplikasi yg harus didownload karna tuntutan kerja, alfacart, alifagift, ponta, reg ponta, a connectio... https://t.co/9TJ1kP7c	FALSE	0	tlieko	2019-07-02 15:09:24	TRUE	1.146073e+18	1.146073e+18	2.363468e+08	Twitter for iPhone
5	@deviannisaaa_ @ayangayy Wkwkwk yakin mau setia? Tuh promo cashback gopay aja yg baru lebih dihargai dari pada yg setia <U+0001F917>	FALSE	0	deviannisaaa_	2019-07-02 14:54:14	FALSE	1.146063e+18	1.146070e+18	1.136257e+18	Twitter for iPhone

Gambar 4. Halaman Data Twitter GOPAY



Gambar 5. Tampilan Halaman Pembersihan Data Twitter GOPAY

Halaman kamus positif dan kamus negatif menampilkan gabungan daftar kamus positif yang sudah diunduh pada *website* github.com dan cara manual.

Hasil penggabungan kamus positif dapat ditunjukkan pada Gambar 6 dan hasil penggabungan kamus negatif ada pada Gambar 7. Gambar 6 dan Gambar 7 memperlihatkan bahwa halaman kamus positif pada file *ui.R*, berfungsi untuk

menampilkan kamus positif yang terdapat pada variabel kamus positif. Variabel kamus positif terdapat pada server.R yaitu sebagai variabel penampung untuk memanggil kamus positif, sedangkan halaman kamus negatif berfungsi untuk menampilkan kamus negatif yang terdapat pada variabel kamus negatif. Variabel kamus negatif terdapat pada server.R yaitu sebagai variabel penampung untuk memanggil kamus negatif.

Sentimen Analisis	
Beranda	Show 25 entries
Data Twitter	a.
» Data OVO	acungan jempol
» Data GOPAY	adaptif
» Data LinkAja	adil
Pembersihan Data	afinitas
» Data Bersih OVO	afirmasi
» Data Bersih GOPAY	agilely
» Data Bersih LinkAja	agung
Kamus Kata Positif	ahli
Kamus Kata Negatif	ahlinya
Hasil Analisis Lexicon	ajaib
» Data OVO	aklamasi
» Data GOPAY	akomodatif
» Data LinkAja	akurat
Akurasi	alam mimpi
» Akurasi Data OVO	alhamdulillah
» Akurasi Data GOPAY	allahu akbar
	altruistis
	aman
	amanah

Gambar 6. Halaman Kamus Positif

Sentimen Analisis	
Beranda	Show 25 entries
Data Twitter	abnormal
» Data OVO	absurd
» Data GOPAY	acak
» Data LinkAja	acak-acakan
Pembersihan Data	acuh
» Data Bersih OVO	acuh tak acuh
» Data Bersih GOPAY	adiktif
» Data Bersih LinkAja	adil
Kamus Kata Positif	agresi
Kamus Kata Negatif	agresif
Hasil Analisis Lexicon	agresor
» Data OVO	aib
» Data GOPAY	air terjun
» Data LinkAja	akurat
Akurasi	alarm
» Akurasi Data OVO	alasan
» Akurasi Data GOPAY	alat permainan
	alergi
	alergik
	amat ketakutan

Gambar 7. Halaman Kamus Negatif

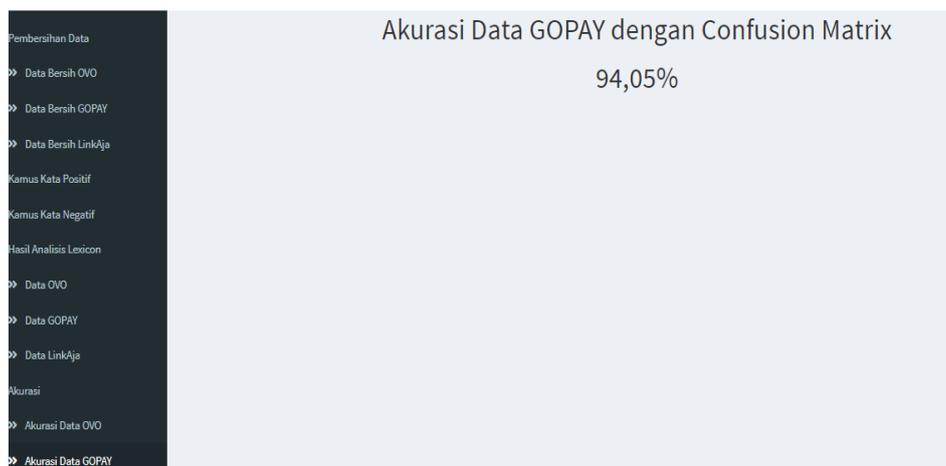
Halaman Hasil Klasifikasi Data GOPAY dengan *Lexicon Based*

Halaman klasifikasi data GOPAY, dataset yang telah dibersihkan memiliki 381 tweets yang terdiri dari 303 *tweets* positif, 36

tweets negatif dan 42 *tweets* netral. Tampilan halaman hasil klasifikasi data GOPAY dengan *Lexicon Based* dapat dilihat pada Gambar 8.

Sentimen Analisis					
Beranda					
Show 25 entries					
	Text	Positive	Negative	Score	Summary
» Data OVO	mumpung akeh promoan gopay ne nek dinihari	1	0	1	Positive
» Data GOPAY	sejauh ini masak x waktu kkn dan memutuskan kelak anakku akan ku beri unlimited saldo ovo gopay dan s	4	6	-2	Negative
» Data LinkAja	ya buat promo besar besaran kaya gopay dan ovo masa tidakpromo mau nyaingin nda mgk kaya	7	6	1	Positive
Pembersihan Data	udah berapa banyak aplikasi yg harus didownload karna tuntutan kerja alfacartalfajitpontareg pontaa connectio	3	2	1	Positive
» Data Bersih OVO	wkwkwk yakin mau setia tuh promo cashback gopay aja yg baru lebih dihargai dari pada yg setia	14	9	5	Positive
» Data Bersih GOPAY	makin sedikit platform fintech makin bagus standarisasi nontunai makin mudah dan seragam gak perlu lagi sibuk lia	4	10	-6	Negative
» Data Bersih LinkAja	siapapun yang kemarin bilang tidakkebagian early bird presale amp promo gopay nih dikasih promo lain yg asoy bgt alig cuma	5	2	3	Positive
Kamus Kata Positif	kenapa harus kesal dengan kapitalisme bukannya dimanfaatkan sebaik mungkinseperti download gopayovodanaturus di	4	5	-1	Negative
Kamus Kata Negatif	ngerinya kalo si gojkek kalah saing terus ghrab nguasain pasar dia jadi bebas mainin harga seenaknya huhuw	5	4	1	Positive
Hasil Analisis Lexicon	belanja di supermarket habis rb tuker poin dapat potongan jadi tinggal rb belanja di alfa rb pakai pr	6	5	1	Positive
» Data OVO	tix id kuota promo habis gotix lagi tidakada promocashback cashbac tidakbisa buat gopay mau nonton mu	2	1	1	Positive
» Data GOPAY	jangan cba beli di alfa kalo tidakpke go pay karena bnyak promo di gopay	4	3	1	Positive
» Data LinkAja	smph td gua abis pesen kfc di goj k pk promo kan trs tbtb gue dpt cashback gopay rb wkwk beruntungnya diriku	6	4	2	Positive

Gambar 8. Tampilan Halaman Hasil Klasifikasi Data GOPAY dengan *Lexicon Based*



Gambar 9. Tampilan Halaman Akurasi Data GOPAY

Gambar 8 menjelaskan tampilan hasil dari klasifikasi *lexicon based* data GOPAY terdapat pada variabel *hasillexiconGopay*. Variabel hasil *lexicon Gopay* terdapat pada server.R yaitu sebagai variabel penampung untuk memanggil variabel hasil *lexicon Gopay*.

Tampilan Halaman Akurasi Data GOPAY

Halaman akurasi data GOPAY merupakan halaman yang digunakan untuk menampilkan hasil akurasi menggunakan

confusion matrix dari sentimen analisis algoritma K-Nearest Neighbor. Hasil akurasi yang tercantum adalah hasil akurasi proses sentimen analisis pada data twitter GOPAY yang telah diproses sebelumnya. Tampilan halaman akurasi ditampilkan secara *text* dengan nilai akurasi yang didapatkan dari hasil perhitungan sistem. Halaman akurasi data GOPAY dapat dilihat pada Gambar 9 yang diperoleh dari perhitungan dengan model *Confusion Matrix* untuk nilai akurasi dengan kecenderungan positif.

Gambar 11 menjelaskan bahwa histogram sisi kiri menunjukkan warna orange adalah jumlah data negatif dan warna hijau adalah jumlah data positif digabung dalam satu grafik. Histogram sisi kanan menunjukkan bahwa data GOPAY dengan sentimen positif lebih banyak daripada data GOPAY dengan sentimen negative dipisah menjadi 2 grafik.

Pengujian Confusion Matrix Data GOPAY

Pengujian confusion matriks data GOPAY dilakukan secara manual dengan melihat tabel confusion matrix oleh sistem pada tabel. Pengujian akurasi dilakukan untuk mengetahui tingkat akurasi yang dilakukan oleh sistem dengan menggunakan *K-Nearest Neighbor*. Pengujian dilakukan dengan menggunakan *confusion matrix* yaitu sebuah matriks dari prediksi yang akan dibandingkan dengan kelas yang asli dari data twitter [8]. Hasil pengujian akurasi klasifikasi dapat dilihat Tabel 1 yang menunjukkan data pengujian akurasi, proses perhitungan difokuskan pada nilai *true-*

positive, *true-negative*, *false-positive* dan *false negative*. Berdasarkan pengujian dan perhitungan nilai akurasi data OVO dari sistem analisis sentimen dengan menggunakan *k-nearest neighbor* sebanyak 107 *tweets* didapatkan hasil akurasi sebesar 86,91% dengan menggunakan persamaan 2.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (2)$$

$$Akurasi = \frac{89 + 4}{89 + 13 + 1 + 4}$$

$$Akurasi = 0,8691$$

Data pengujian akurasi yang digunakan untuk Gopay sebanyak 101 *tweets*. Berdasarkan pengujian akurasi, didapatkan hasil akurasi data Gopay dari sistem analisis sentimen dengan menggunakan *K-Nearest Neighbor* sebesar 94,05%.

$$Akurasi = \frac{93 + 2}{93 + 6 + 0 + 2}$$

$$Akurasi = 0,9405$$

Tabel 1. Confusion Matrix Data GOPAY

Predicted	Actual Negative	Positive	Row Total
Negative	2 1.000 0.250	0 0.000 0.000	2 0.020
Positive	6 0.061 0.750	93 0.939 1.000	99 0.980
Column Type	8 0.079	93 0.921	101

Data pengujian akurasi yang digunakan untuk LinkAja sebanyak 76 *tweets*. Berdasarkan pengujian akurasi, didapatkan hasil akurasi data LinkAja dari sistem analisis sentiment dengan menggunakan k-nearest neighbor sebesar 76,31%.

$$Akurasi = \frac{49 + 9}{49 + 16 + 2 + 9}$$

$$Akurasi = 0,7631$$

Hasil proses yang memiliki nilai akurasi paling tinggi dari ketiga twitter yang diamati, yaitu *twitter* Gopay.

KESIMPULAN DAN SARAN

Penelitian ini telah berhasil membuat website untuk melakukan analisis sentimen terhadap dompet elektronik dari *tweet* dengan kata OVO, Promo, Gopay, LinkAja dengan metode *Lexicon Based* dan *K-Nearest Neighbor*. Hasil dari penelitian ini mencakup tahapan penelitian yang terdiri dari analisis sumber data, *preprocessing* data, klasifikasi sentimen dengan metode *Lexicon Based* untuk menentukan sentimen positif dan negatif, klasifikasi dengan metode *K-Nearest Neighbor* untuk perhitungan akurasi dengan confusion matrix, visualisasi wordcloud dan histogram serta tampilan menggunakan packages shiny RStudio. Pengujian menggunakan confusion matrix memperoleh hasil akurasi dari masing-masing data menggunakan metode *K-Nearest*

Neighbor untuk *tweet* OVO cenderung positif ditunjukkan oleh nilai akurasi data 86,91% dan kecenderungan negatif *tweet* OVO memiliki nilai akurasi data 13,09%, *tweet* Gopay cenderung positif dengan akurasi data 94,05% dan kecenderungan negatif *tweet* Gopay memiliki nilai akurasi data 5,95%, serta *tweet* LinkAja cenderung positif dengan akurasi data 76,31% dan kecenderungan negatif *tweet* LinkAja memiliki nilai akurasi data 23,69%.

Kecenderungan negatif *tweet* yang diperoleh sangat bergantung pada pemilihan tahapan *pre-processing*. Analisis sentimen ini masih memerlukan pengembangan selanjutnya, seperti pengambilan data dari *Twitter* dalam jumlah yang besar dan dibuatkan skala berdasarkan periode waktu yang telah ditentukan secara interaktif. Penggunaan metode yang berbeda sebagai bahan perbandingan, untuk mengetahui hasil akurasi yang lebih baik. Aplikasi dapat dikembangkan dengan penambahan fitur *real-time* pada visualisasi data, sehingga dapat diakses kapan saja. Pemilihan tingkat ketepatan dalam melakukan prediksi, membutuhkan konsultasi dengan pakar bahasa untuk mengatasi data uji yang memiliki kalimat ambigu.

DAFTAR PUSTAKA

- [1] I. Utami, dan M. Marzuki, "Analisis sistem informasi banjir berbasis media

- twitter”, *Jurnal Fisika Unand*, vol. 9, no. 1, pp. 67-72, 2020.
- [2] S.K. Ravindran, and V. Garg, “*Mastering social media mining with R*”, Packt Publishing Ltd. UK., 2015, [ebook].
- [3] S. Jusoh, and HM. Alfawareh, “Applying fuzzy sets for opinion mining”, *International Conference on Computer Applications Technology (ICCAT)*, pp. 1–5, 2013, <http://doi.org/10.1109/ICCAT.2013.6521965>.
- [4] M. Rivki, dan AM. Bachtiar, “Implementasi algoritma K-Nearest Neighbor dalam pengklasifikasian follower twitter yang menggunakan Bahasa Indonesia”, *Jurnal Sistem Informasi*, vol. 13, no. 1, pp. 31-27, 2017.
- [5] I. Indriati, dan A. Ridok, “Analisis sentimen untuk meninjau aplikasi seluler menggunakan Metode Neighbor Weighted K-Nearest Neighbor (Nwknn)”, *Jurnal Teknik Lingkungan dan Teknologi Berkelanjutan*, vol. 3, no. 1, pp. 23-32, 2016.
- [6] U. Rofiqoh, RS. Perdana dan MA. Fauzi, “Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode Support Vector Machine dan Lexicon Based Features”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, e-ISSN, vol. 2548, pp. 964x, 2017.
- [7] RI. Pristiyanti, MA. Fauzi, dan L. Muflikhah, “Sentimen Analisis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN., vol. 2548, pp. 964x, 2018.
- [8] MR. Huq, A. Ali, dan A. Rahman, “Analisis Sentimen Pada Data Twitter Menggunakan KNN dan SVM”, *Jurnal Internasional Ilmu dan Aplikasi Komputer Lanjutan (IJACSA)*, vol. 8, no. 6, pp. 19-25, 2017.
- [9] YF. Safri, R. Arifudin, dan MA. Muslim, “K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor”, *Sci. J. Informatika*, vol. 5, no. 1, pp. 18, 2018.
- [10] O. Nurdiana, J. Jumadi, dan D. Nursantika, “Perbandingan metode Cosine Similarity dengan metode Jaccard Similarity pada aplikasi pencarian terjemah Al-Qur’an dalam Bahasa Indonesia”, *Jurnal Online Informatika*, vol. 1, nO. 1, pp. 59-63, 2016.
- [11] A. Deviyanto, dan MDR. Wahyudi, “Penerapan analisis sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor”, *Jurnal Informatika Sunan Kalijaga (JISKA)*, vol. 3, no. 1, pp. 1-13, 2018.

- [12] SK. Lidya, OS. Sitompul, dan S. Efendi, “Sentimen analisis pada teks Bahasa Indonesia menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN)”, Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA), Universitas Sumatera Utara, ISSN: 2089-9815. 2015.