



Sentiment Analysis pada Twitter untuk Perbandingan Produk Secara Real-Time dengan Menggunakan Pendekatan Lexicon-Based

Sentiment Analysis on Twitter for Real-Time Product Comparison Using Lexicon Based Approach

Benedictus Steven Hanantyo, Rinta Kridalukmana, Dania Eridani

*Program Studi Sistem Komputer, Fakultas Teknik, Universitas Diponegoro
Jl. Prof. Soedarto, SH, Kampus Undip Tembalang, Semarang, Indonesia 50275*

How to cite: B.S. Hanantyo, R. Kridalukmana, D. Eridani, "Sentiment Analysis pada Twitter untuk Perbandingan Produk Secara Real-Time dengan Menggunakan Pendekatan Lexicon-Based", *Jurnal Teknologi Komputer*, vol. 1, no. 2, pp. 41-48, November 2022, doi: 10.14710/jtk.v1i2.36317 [Online].

Abstract – *In the digital era, it is easier for us to get information. We can get information from websites, electronic news portals, social media, and other electronic media. Social media is one of the places to express opinions such as expressing opinions on a product on the Twitter application. For this reason, companies can obtain information from opinions expressed by users of their products, from these opinions sentiment analysis can be carried out which can determine the opinion is positive, negative, or neutral. The research was carried out based on the OSEMN method which consisted of Obtain, Scrub, Explore, Model, and Interpret stages. In this study, the researcher hopes to get a high enough accuracy value so that it can help the product to develop properly according to the problem at hand. The final result of this research is to get an accuracy value of 0.68.*

Keywords – *Social media, Twitter API, Sentiment Analysis*

Abstrak – *Pada masa yang sudah serba digital seperti sekarang ini maka semakin mudah pula kita untuk mendapatkan informasi. Informasi yang kita dapatkan dapat berasal dari website, portal berita elektronik, sosial media, dan berbagai media elektronik lainnya. Media sosial merupakan sarana informasi dan juga sebagai sarana untuk menyampaikan pendapat yang paling sering digunakan contohnya ialah seperti memberikan opini terhadap suatu produk dari perusahaan tertentu pada aplikasi Twitter, maka dari itu perusahaan juga dapat mendapatkan informasi dari opini-opini pengguna produknya yang disampaikan lewat media sosial tersebut dengan cara melakukan scrapping data melalui API yang telah disediakan oleh media sosial yang ingin digunakan untuk mendapatkan informasi. Dari data tersebut dapat dilakukannya sentiment analysis terhadap opini tersebut untuk menentukan apakah opini tersebut bernilai positive,*

negative, atau netral. Penelitian dilaksanakan berdasarkan metode OSEMN yang terdiri atas tahap Obtain, Scrub, Explore, Model, dan iNterpret. Pada penelitian ini peneliti berharap untuk mendapatkan nilai akurasi yang cukup tinggi sehingga dapat membantu produk dapat berkembang dengan tepat sesuai dengan masalah yang dihadapi. Hasil akhir dari penelitian ini ialah mendapatkan nilai akurasi sebesar 0.68.

Kata kunci – *Media sosial, Twitter API, Sentiment Analysis*

I. PENDAHULUAN

Pada perkembangan zaman yang sudah maju ini, media sosial merupakan suatu hal yang paling melekat pada masyarakat khususnya pada kaum muda yang mana media sosial, merupakan tempat mereka untuk untuk berkomunikasi dengan kerabat, juga sebagai tempat bagi mereka untuk mengekspresikan pendapat. Salah satu platform yang digunakan oleh kaum muda untuk mengekspresikan pendapat ialah media sosial *Twitter*. Pendapat yang dikemukakan pada *Twitter* sendiri sebagian besar merupakan pendapat yang berasal dari pengalaman yang pernah dirasakan oleh penggunaanya, seperti memberikan ulasan terhadap suatu produk yang pernah dipakai.

Dikarenakan media sosial sebagai tempat untuk mengekspresikan pendapat terdapat banyak sekali kritik dan saran yang disalurkan oleh pengguna suatu produk pada media sosial miliknya. Maka dari itu baiknya untuk suatu perusahaan memantau *trend* terhadap produk yang telah disalurkan ke pasaran untuk mengetahui apakah *trend* produk tersebut positif atau negatif ataupun netral. Dengan adanya pendapat yang disalurkan oleh pengguna suatu produk diharapkan dapat membuat sebuah perusahaan dapat menjadi lebih baik untuk kedepannya dikarenakan semua pendapat baik itu positif ataupun negatif dapat membantu pertumbuhan sebuah perusahaan itu sendiri seperti bila mendapatkan banyak pendapat

*) Penulis Korespondensi (B. S. Hanantyo)
Email: benedictussteven@students.undip.ac.id



negatif berupa kritik itu dapat membantu sebuah perusahaan untuk memperbaiki produk tersebut.

Maka dari itu peneliti melakukan penelitian pada beberapa produk dari beberapa perusahaan untuk mengetahui *trend* apa yang dimiliki perusahaan tersebut apakah itu positif, negatif, ataupun netral. Sehingga dilakukannya *sentiment analysis* ini dengan menggunakan pendekatan *lexicon based* yang mana nantinya memiliki hasil akhir yaitu menentukan apakah suatu opini tersebut positif, negatif, ataupun netral, dengan adanya sistem ini diharapkan dapat membantu baik itu perseorangan maupun sebuah perusahaan untuk dapat mengetahui *trend* yang dimiliki oleh suatu produk yang ingin diketahui. Pada penelitian ini menghasilkan hasil akhir yaitu berupa nilai akurasi yang mana pada penelitian ini memiliki nilai akurasi sebesar 0.68, yang mana nilai akurasi didapatkan dari perbandingan hasil *survey* terhadap *tweet dataset* yang dimiliki dan juga hasil *tweet* yang dihasilkan menggunakan *library*.

II. TINJAUAN LITERATUR

Pada penelitian ini digunakan beberapa referensi yaitu berupa penelitian terdahulu yang mana penelitian tersebut sejenis dengan penelitian ini yaitu bertemakan tentang *sentiment analysis* dan dapat dijadikan sebagai pedoman dalam melakukan penelitian ini. Penelitian yang pertama ialah memiliki hasil bahwa tidak semua tahapan *text pre-processing* perlu dilakukan, pada tahapan menghapus atau memperbaiki *slang word* itu sangat penting dikarenakan dapat menaikkan keefektivitasan dari hasil akhir, dan juga menghapus *stopword*, angka, dan URL hanya berdampak kecil sehingga bisa saja diabaikan [1].

Penelitian yang kedua menggunakan metode *Naïve Bayes*, *J48 Classification*, *Breadth First Search*, dan *One Rule Method* semua algoritma yang dipakai memiliki nilai *precision* dan *F-measure* diatas 80% [2]. Penelitian yang ketiga menggunakan metode *Naïve Bayes*, *Rocchio Classifier*, *Perceptron Classifier* dan dapat disimpulkan bahwa berdasarkan dengan *F-score* maka *Rocchio Classifier* yang paling baik dari semuanya, *Naïve Bayes* memiliki nilai yg hampir sama dengan *Rocchio Classifier* tetapi memiliki nilai *recall* yang lebih kecil [3].

Pada penelitian yang keempat metode yang digunakan ialah *K-Nearest Neighbors* dan juga *Support Vector Machine*. *K-Nearest Neighbors* menghasilkan nilai *accuracy* sebesar 84.32% dan *Support Vector Machine* menghasilkan nilai *accuracy* sebesar 77,97%. [4] Pada penelitian yang kelima Metode yang digunakan ialah *Naïve Bayes*, *K-Nearest Neighbors*, dan *Support Vector Machine*. *Naïve Bayes* memiliki nilai *accuracy* yang paling baik yaitu dengan nilai 80,90%, *K-Nearest Neighbors* menghasilkan nilai *accuracy* sebesar 75,58%, dan *Support Vector Machine* menghasilkan nilai *accuracy* sebesar 63,99% [5].

Dari keseluruhan penelitian yang terlebih dahulu tidak terdapatnya penelitian yang menggunakan metode *lexicon based* yang mana dikarenakan hal tersebut peneliti ingin melakukan penelitian menggunakan

metode *lexicon based* untuk mengetahui keefektivitasan metode *lexicon based* bila digunakan untuk melakukan *sentiment analysis* dan juga dikarenakan belum adanya penelitian yang membahas tentang sistem *sentiment analysis* secara *real-time*.

III. METODE PENELITIAN

Pada penelitian tugas akhir ini menerapkan model OSEMN yang mana model OSEMN merupakan langkah dalam *data science* yang terdiri atas urutan langkah kronologis: *Obtain*, *Scrub*, *Explore*, *Model*, dan *iNterpret*. Untuk penjabaran pada setiap Langkah adalah sebagai berikut

A. Tahap *Obtain*

Pada tahap ini peneliti melakukan pengambilan data yang diperlukan, pada penelitian ini data berupa *tweet* yang ada pada *Twitter* API yang diambil menggunakan *library tweepy* yang mana *tweet* yang diambil memiliki kriteria seperti: kata kunci (*samsung*, *iphone*, *xiaomi*), berbahasa indonesia, dan juga dibuat pada batas tanggal tertentu. Data yang diambil dari *Twitter* API ialah berupa: *full text* yang mana berisi semua isi dari *tweet* tersebut, nama *username* dari penulis *tweet*, dan yang terakhir ialah tanggal dibuatnya *tweet* tersebut. Semua data tersebut akan di simpan pada format *.csv*.

B. Tahap *Scrub* dan *Explore*

Pada tahap *scrub* dan *explore* ini peneliti melakukan beberapa langkah dengan urutan *Scrub-1*, *Explore-1*, *Scrub-2*, *Explore-2*, *Scrub-3*. Pada tahapan ini *scrub* dan *explore* sangat kebergantungan dikarenakan dengan dilakukannya eksplorasi pada *dataset* yang dimiliki maka hasil pembersihan dan pemfilteran data menjadi lebih baik. Pada tahap *scrub-1* ini peneliti melakukan *text preprocessing* berupa *case folding* yang mana terdiri dari:

1. *Lower case* yang mana yaitu membuat semua kata menjadi huruf kecil sehingga memudahkan peneliti untuk memproses *dataset* yang dimiliki
2. Menghapus *tab*, *enter*, dan juga *back slice* yang terdapat pada kalimat *dataset* yang dimiliki
3. Menghapus semua *non ASCII* yang terdapat pada *dataset* seperti *emoticon*, *chinese word*, dan lain sebagainya
4. Menghapus *mention*, *link*, dan juga *hashtag*
5. Menghapus URL yang tidak lengkap
6. Menghapus tanda baca

Setelah menerapkan tahapan *scrub-1* maka dilakukannya tahapan *explore-1* dimana pada tahap *explore-1* ini peneliti melakukan *FreqDist* pada *dataset* sehingga *dataset* dapat digunakan untuk melihat kata apa saja yang paling banyak keluar pada *dataset*, dan juga melakukan distribusi kata yang mana digunakan agar dapat mengetahui konteks kata yang terdapat pada *dataset*. Pada tahap ini didapatkan bahwa terdapat kata seperti *tokopedia*, *shopee*, dan juga *grabfood* yang mana tidak relevan dikarenakan kata tersebut berisi data yang berulang-ulang yang mana berisi tentang kuis yang



diadakan oleh pihak *shopee*, *tokopedia*, dan juga *grabfood*. Setelah dilakukannya tahap *explore-1* dilakukannya tahap *scrub-2*, pada tahap *scrub-2* ini dilakukannya penghapusan terhadap data yang terdapat kata *shopee*, *tokopedia*, *grabfood* dengan alasan yang telah dijelaskan pada langkah *explore-1*.

Setelah dilakukannya tahap *scrub-2* dilakukannya tahap *explore-2* yang mana pada tahap *explore-2* ini dilakukannya hal yang sama seperti pada langkah *explore-1*, dan didapatkan hasil beberapa *stopword* yang dapat dimasukkan pada *list stopwords*, yaitu 7fbd640f, pin, 087889858171, gkl0838m yang mana kata tersebut tidak memiliki dampak perubahan makna pada kalimat bila kata tersebut dihapuskan.

Setelah dilakukannya tahap *explore-2* dilakukannya tahap *scrub-3* yang mana pada tahap *scrub-3* ini peneliti melakukan *tokenizing data* dan *stopword removal* yang mana *dataset stopwords* didapatkan dari beberapa sumber yaitu *library nltk*, *list stopwords* yang didapatkan dari hasil langkah *explore-2*, dan juga *stopwords.txt* yang diunduh dari internet. Setelah dilakukannya *tokenizing* dan juga *stopword removal* dilakukannya penghapusan baris yang kosong dikarenakan yang diakibatkan oleh *stopword removal* dan juga dilakukan penghapusan duplikat data.

C. Tahap Model

Pada tahap *model* ini peneliti melakukan pengubahan bahasa pada *dataset* yang ada dari bahasa indonesia menjadi bahasa inggris dikarenakan *library vaderSentiment* tidak dapat digunakan bila data menggunakan bahasa indonesia, untuk *library* yang digunakan untuk menerjemahkan data ialah *library googletrans*. Setelah *dataset* diterjemahkan maka *dataset* disimpan menjadi *file csv* dengan nama *clean_data* dan berikut merupakan kode yang digunakan untuk menerjemahkan *dataset* menjadi berbahasa inggris

Setelah dilakukan penerjemahan bahasa maka dapat dilakukannya *sentiment analysis* pada *dataset* yang dipunya yang mana akan menghasilkan *compound score* yang mana dari *compound score* tersebut akan dirubah menjadi kata positif, negatif, maupun netral berdasarkan kategori yang telah dijelaskan pada poin 2.6 VADER. Berikut merupakan kode yang digunakan untuk melakukan *sentiment analysis* menggunakan *vaderSentiment*.

D. Tahap Interpret

Pada tahap ini peneliti melakukan pengecekan hasil *sentiment analysis* dengan cara melakukan *sentiment analysis* secara manual pada *dataset* yang telah dimiliki yang mana nantinya akan dilakukan pengecekan akurasi terhadap model yang telah dibuat dan juga dilakukannya *confusion matrix*. Dilakukannya pemecahan *dataset* yang dimiliki menjadi tiga *dataset* yang mengandung kata kunci *samsung*, *iphone*, dan *xiaomi* dan setelah itu dibandingkan produk manakah

yang lebih baik dengan kriteria nilai positif dibagi dengan jumlah data yang dimiliki.

IV. HASIL DAN PEMBAHASAN

A. Hasil Tahap Obtain

Pada tahap *obtain* ini peneliti melakukan pengambilan data pada *Twitter API* dengan menggunakan *library* bernama *tweepy*, tetapi sebelum peneliti dapat mengambil data pada *Twitter API* peneliti diwajibkan untuk mendaftar terlebih dahulu pada *Twitter developer* agar bisa mendapatkan *API key*, *access token*, dan *bearer token* yang mana digunakan untuk menyambungkan *library tweepy* dengan *Twitter API* untuk tata cara agar mendapatkan *Twitter API* dapat dilihat pada sumber berikut [6]. Pada penelitian ini digunakan *Twitter API v1.1*.

Pada *library tweepy* digunakan parameter seperti *auth* yang digunakan sebagai autentikasi, *wait_on_rate_limit* yang digunakan bila mana peneliti sedang mengambil data pada *Twitter* dan batas pengambilan *tweet* terjadi maka program tidak berhenti dan tetap berjalan hingga waktu menunggu selesai, *cursor* yang digunakan untuk mengambil data sesuai dengan jumlah yang kita butuhkan, *search_tweets* yang digunakan untuk mengambil *tweet* pada *Twitter* yang mana memiliki parameter kata kunci, bahasa, mode *tweet*, hasil, dan batas waktu, untuk hasil dari parameter *search_tweets* itu sendiri berupa *full_text*, *username*, dan *timestamp*.

Data yang diambil merupakan *username*, *timestamp*, dan juga *tweet* itu sendiri yang mana kita mengambil menjadi tiga bagian yang dibedakan berdasarkan kata kunci yang berbeda yaitu *samsung*, *iphone*, dan juga *xiaomi* dan juga ketiga bagian tersebut memiliki beberapa persamaan, yaitu memiliki batas tanggal 22 September 2022 pukul 00.00 waktu server, *tweet* yang diambil bukan merupakan *retweet*, dan juga data *tweet* yang diambil berjumlah maksimal 500 *tweet*. Lalu setelah didapatkannya tiga data yang berbeda dilakukan penggabungan data terhadap ketiga data yang dimiliki dan diberi nama variabel data *raw* lalu data disimpan menjadi *file csv*. Tabel 1 merupakan beberapa contoh *dataset* yang didapatkan dari hasil tahap *obtain*.

Tabel 1. Contoh *dataset* yang dihasilkan pada tahap *obtain*

<i>user name</i>	<i>timestamp</i>	<i>tweet</i>
User 1	2022-09-21 23:59:41+00:00	Wahai langit tolong kirimka duit buat beli nih hp ada fitur kesukaan gue Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/S2YZS8QDqy



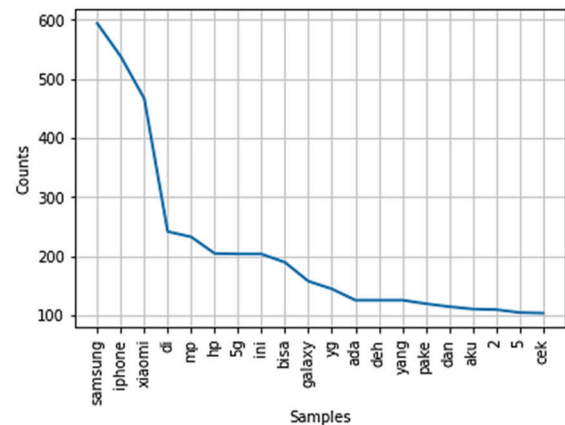
<i>user name</i>	<i>timestamp</i>	<i>tweet</i>
User 2	2022-09-21 23:58:49+00:00	Mau videocallan sama doi udh ga malu lagi soalnya muka ku tertutupi Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/OGUP8nt5c4
User 3	2022-09-21 23:57:53+00:00	Mau fotoin kucing aku yg unyunya seribu bahasa pun skrg udh gampang ada Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/EJEPmgOe36
User 4	2022-09-21 23:57:41+00:00	@convomfs Samsung J1 wkwkwk
User 5	2022-09-21 23:55:03+00:00	[Samsung] TV Samsung 43" 4K 2022 R\$ 2399,99 #desconto #oferta #promocao https://t.co/9RDGz9KsPQ

B. Hasil Tahap Scrub dan Explore

Pada penelitian ini peneliti memiliki *dataset* awal yang berisi sebanyak 1500 baris data tentang *tweet* yang diambil dari *Twitter* menggunakan *library tweepy*. Setelah dilakukannya tahap *explore-1*, yaitu melakukan perhitungan *FreqDist*, yang digunakan untuk menampilkan kata-kata yang paling banyak keluar, didapatkan bahwa kata *samsung*, *iphone*, dan juga *xiaomi* menjadi 3 kata terbanyak yang terdapat pada *dataset* karena ketiga kata tersebut merupakan kata utama pada *dataset* ini, dapat dilihat dari Gambar 1 dan Gambar 2.



Gambar 1. Cloudword yang terdapat pada *dataset* pada tahap *explore-1*

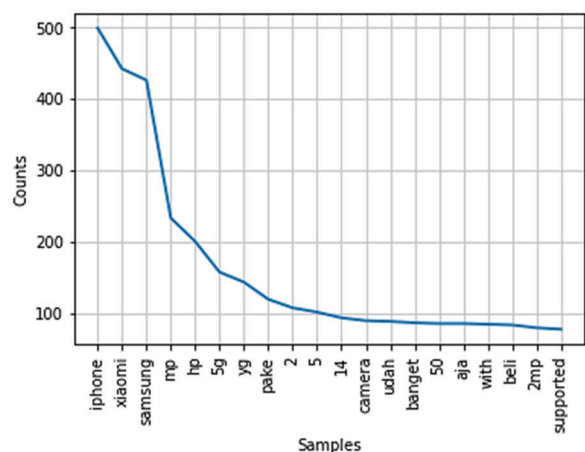


Gambar 2. Grafik frekuensi kata yang terdapat pada tahap *explore-1*

Lalu dilakukannya pemangkasan data pada tahap *scrub-2* yang mana akhirnya menghasilkan perubahan hasil dari *explore-2* yaitu terjadi perubahan jumlah *dataset* yang dimiliki dari 1500 baris menjadi 1247 baris dan juga terjadinya perubahan frekuensi pada ketiga kata utama sehingga menjadi seperti Gambar 3 dan Gambar 4.



Gambar 3. Cloudword yang terdapat pada tahap *explore-2*



Gambar 4. Grafik frekuensi kata yang terdapat pada tahap *explore-2*

Setelah dilakukannya *explore-2* maka dilakukannya *scrub-3* yang mana dilakukannya *tokenizing* dan juga *stopword removal*. Contoh hasil pada *dataset* setelah dilakukannya tahapan *scrub* dan juga *explore* dapat dilihat pada Tabel 2.



Tabel 2. Hasil tahapan *scrub* dan *explore*

<i>user name</i>	<i>tweet</i>	<i>clean_tweet</i>
<i>User 1</i>	Wahai langit tolong kirimka duit buat beli nih hp ada fitur kesukaan gue Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/S2YZS8QDqy	langit tolong kirimka duit beli nih hp fitur kesukaan gue quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g
<i>User 2</i>	Mau videocallan sama doi udh ga malu lagi soalnya muka ku tertutupi Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/OGUP8nt5c4	videocallan doi udh ga malu muka ku tertutupi quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g
<i>User 3</i>	Mau fotoin kucing aku yg unyunya seribu bahasa pun skrg udh gampang ada Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/EJEpMgOe36	fotoin kucing yg unyunya seribu bahasa skrg udh gampang quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g
<i>User 4</i>	@convomfs Samsung J1 wkwwk	samsung j1 wkwwk
<i>User 5</i>	[Samsung] TV Samsung 43" 4K 2022 R\$ 2399,99 #desconto #oferta #promocao https://t.co/9RDGz9KsPQ	samsung tv samsung 43 4k 2022 r 239999

C. Hasil Sentiment Analysis secara Manual

Pada saat melakukannya *sentiment analysis* secara manual melibatkan lima responden termasuk dengan peneliti, hasil survei tersebut akan dijadikan satu pada kolom *sentiment* akhir yang mana berisi nilai terbanyak yang dihasilkan dari kelima responden. Responden mengisi survei dengan angka 2, 1, ataupun 0 yang memiliki arti 2 merupakan *tweet* bersifat positif, 1 merupakan *tweet* bersifat netral, dan 0 merupakan *tweet* bersifat negatif. Dari survei tersebut menghasilkan data netral sebanyak 681 data, data positif sebanyak 428 data,

dan data negatif sebanyak 138 data. Beberapa contoh hasil dari dilakukannya *sentiment analysis* secara manual dapat dilihat pada Tabel 3.

Tabel 3. Hasil *sentiment analysis* secara manual

<i>tweet</i>	S 1	S 2	S 3	S 4	S 5	<i>Sentiment Akhir</i>
Wahai langit tolong kirimka duit buat beli nih hp ada fitur kesukaan gue Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/S2YZS8QDqy	2	2	2	1	2	2
Mau videocallan sama doi udh ga malu lagi soalnya muka ku tertutupi Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/OGUP8nt5c4	2	2	2	2	2	2
Mau fotoin kucing aku yg unyunya seribu bahasa pun skrg udh gampang ada Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/EJEpMgOe36	2	2	2	2	2	2
@convomfs Samsung J1 wkwwk	1	1	1	1	1	1
[Samsung] TV Samsung 43" 4K 2022 R\$ 2399,99 #desconto #oferta #promocao https://t.co/9RDGz9KsPQ	1	1	1	1	1	1



D. Hasil *Sentiment Analysis* Menggunakan *Library vaderSentiment*

Sebelum melakukan *sentiment analysis* menggunakan *library vaderSentiment* perlu dilakukannya penerjemahan data menjadi berbahasa Inggris menggunakan *library googletrans* yang mana nantinya hasil dari penerjemahan *dataset* disimpan pada kolom `clean_tweet_eng`. Hasil penerjemahan data menggunakan *library googletrans* dapat dilihat pada Tabel 4.

Tabel 4. Hasil penerjemahan data

<i>username</i>	<i>clean_tweet</i>	<i>clean_tweet_eng</i>
<i>User 1</i>	langit tolong kirimka duit beli nih hp fitur kesukaan gue quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g	<i>sky please send money buy this my favorite feature phone quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>
<i>User 2</i>	videocallan doi udh ga malu muka ku tertutupi quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g	<i>video call doi already not ashamed of my face covered quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>
<i>User 3</i>	fotoin kucing yg unyunya seribu bahasa skrg udh gampang quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g	<i>take pictures of the cute cat in a thousand languages now it's easy quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>
<i>User 4</i>	samsung j1 wkwkwk	<i>samsung j1 wkwkwk</i>
<i>User 5</i>	samsung tv samsung 43 4k 2022 r 239999	<i>samsung tv samsung 43 4k 2022 r 239999</i>

Setelah dilakukannya penerjemahan dilakukannya *sentiment analysis* menggunakan *library vaderSentiment* yang nantinya akan menghasilkan *compound score* yang mana dari nilai *compound score* tersebut akan dirubah menjadi kata positif, negatif, dan netral. Hasil dari dilakukannya *sentiment analysis* menggunakan *library vaderSentiment* ialah 600 data bersifat netral, 511 data bersifat positif, dan 136 data bersifat negatif. Contoh

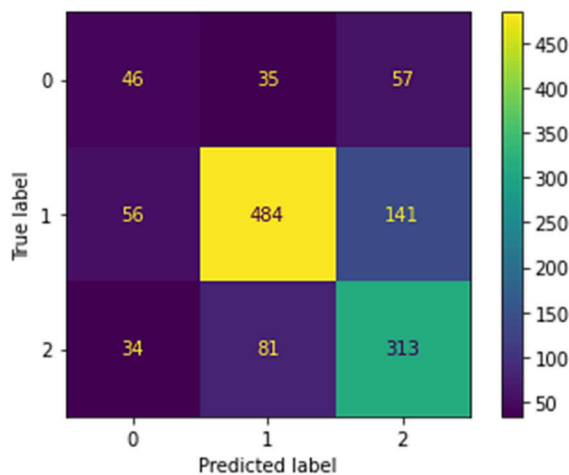
hasil dari *sentiment analysis* menggunakan *vaderSentiment* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *sentiment analysis* menggunakan *library vaderSentiment*

<i>username</i>	<i>clean_tweet_eng</i>	<i>Compound_Score</i>	<i>Sentiment</i>
<i>User 1</i>	<i>sky please send money buy this my favorite feature phone quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>	0.765	<i>Positive</i>
<i>User 2</i>	<i>video call doi already not ashamed of my face covered quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>	0.5932	<i>Positive</i>
<i>User 3</i>	<i>take pictures of the cute cat in a thousand languages now it's easy quad camera 50 mp supported with ois 5 mp 2mp 2 mp samsung 5g</i>	0.802	<i>Positive</i>
<i>User 4_</i>	<i>samsung j1 wkwkwk</i>	0	<i>Neutral</i>
<i>User 5</i>	<i>samsung tv samsung 43 4k 2022 r 239999</i>	0	<i>Neutral</i>

E. Analisa dan Pembahasan

Setelah dilakukannya *sentiment analysis* secara manual dan juga menggunakan *library* maka peneliti membuat *confusion matix* yang ditampilkan pada Gambar 5 dan dihasilkan data sebagai berikut.



Gambar 5. Tabel confusion matrix

Dari Gambar 5 dapat diterjemahkan sebagai berikut, yaitu data dibagi menjadi tiga hasil akhir yaitu label angka 1 merupakan netral, label bernilai 2 merupakan positif, dan label dengan angka 0 merupakan negatif sehingga menghasilkan confusion matrix sebanyak 6 jenis, yaitu:

1. True Neutral yaitu untuk actual value dan predicted value bernilai netral yaitu sebesar 484 baris.
2. False Neutral yaitu predicted value bernilai netral dan actual value selain nilai netral yaitu berjumlah 116 baris.
3. True Positive yaitu untuk actual value dan predicted value bernilai positif yaitu sebesar 313 baris.
4. False Positive yaitu predicted value bernilai positif dan actual value selain nilai positif yaitu berjumlah 198 baris.
5. True Negative yaitu untuk actual value dan predicted value bernilai negatif yaitu sebesar 46 baris.
6. False Negative yaitu predicted value bernilai negatif dan actual value selain nilai negatif yaitu berjumlah 90 baris.

Dari hasil confusion matrix diatas kita dapat melihat metrik klasifikasinya, pada Gambar 6.

	precision	recall	f1-score	support
Negative	0.34	0.33	0.34	138
Neutral	0.81	0.71	0.76	681
Positive	0.61	0.73	0.67	428
accuracy			0.68	1247
macro avg	0.59	0.59	0.59	1247
weighted avg	0.69	0.68	0.68	1247

Gambar 6. Gambar nilai matriks klasifikasi

Dari Gambar 6 dapat dilihat bahwa nilai dari metrik klasifikasi bahwa nilai recall, precision dan juga f1-score untuk bagian negatif sangat kecil yaitu dibawah 0.4 dan untuk nilai yang paling tinggi yaitu bagian netral. Dapat dilihat juga bahwa dari hasil pemodelan ini hanya mendapatkan nilai accuracy sebesar 0.68. Untuk nilai accuracy yang relatif kecil ini dapat terjadi dikarenakan beberapa faktor, seperti:

1. Kurang optimal dalam melakukan text preprocessing
2. Terjadinya human error dalam melakukan sentiment analysis secara manual dikarenakan banyaknya dataset yang dimiliki
3. Banyaknya tweet yang bukan berbahasa indonesia walaupun sudah dilakukannya pembatasan dalam pengambilan data pada Twitter yaitu dengan syarat tweet bahasa indonesia
4. Banyaknya slangword dan juga salah penulisan yang ada pada tweet dataset dikarenakan pengguna Twitter tidak menggunakan bahasa baku pada saat melakukan tweet
5. Terdapat sejumlah tweet yang bersifat satire sehingga sulit untuk dibedakan

Setelah dilakukannya penghitungan confusion matrix pada keseluruhan dataset yang dimiliki, dilakukannya pemecahan dataset menjadi tiga bagian dengan kriteria memiliki kata kunci samsung, iphone, dan xiaomi. Untuk hasil pemecahan dapat dilihat pada Tabel 6 untuk data yang mengandung kata kunci samsung, Tabel 7 untuk data yang mengandung kata kunci iphone dan Tabel 8 untuk data yang mengandung kata kunci xiaomi.

Tabel 6. Tabel dataset samsung

user name	tweet	Sentiment
User 1	Wahai langit tolong kirimka duit buat beli nih hp ada fitur kesukaan gue Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/S2YZS8QDqy	Positive
User 2	Mau videocallan sama doi udh ga malu lagi soalnya muka ku tertutupi Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/OGUP8nt5c4	Positive
User 3	Mau fotoin kucing aku yg unyunya seribu bahasa pun skrg udh gampang ada Quad Camera 50 MP supported with OIS 5 MP + 2MP + 2 MP #JelasPastiAwesome Samsung #GalaxyA23 5G https://t.co/EJEpMgOe36	Positive
User 4	@convomfs Samsung J1 wkwkwk	Neutral
User 5	[Samsung] TV Samsung 43" 4K 2022 R\$ 2399,99 #desconto #oferta #promocao https://t.co/9RDGz9KsPQ	Neutral



Tabel 7. Tabel *dataset iphone*

<i>user name</i>	<i>tweet</i>	<i>Senti-ment</i>
User 1	Hp SuperCopy iPhone 5s 16GB Case Besi Rp 1.150.000 More Info Pin :7FBD640F	Neutral
User 2	@theganatanatda itu keknyaa user android kalo bikin story fontnya sering kek gitu dehh 🤔 aku juga pas make andro ngerepost story org lain pasti usernamenya fontnya ga bisa ke-baca kek gitu di iphone baik2 ajaa	Positive
User 3	@freaky_hamna Thanks ❤️ Insha Allah Btw iPhone kahan gya?	Positive
User 4	iPhone 14 lebih kurang sama ja dgn 13. Selain warna dgn crash detection, dgn Plus size, yang lain copy paste ja. Kalau nak upgrade 14, baik ambik Pro version. https://t.co/AhhJYUd8bs	Negative
User 5	hope y'all choke on that money you used 🤔 pastok keluar la meli iphone baru nak? oops	Negative

Tabel 8. Tabel *dataset xiaomi*

<i>user name</i>	<i>tweet</i>	<i>Senti-ment</i>
User 1	Xiaomi Kucik Charger https://t.co/RcvPUVznhd	Neutral
User 2	@Askrlfess Kalo hp nya xiaomi ada apk bawaannya	Neutral
User 3	@XiaomiIndonesia min, service center xiaomi di daerah cilacap purwokerto dimana yah min? mau service HP redmi note 10, overheating/panas, dan suka re-start dari menu home ke menu logo terus2an tolong infoin estimasi habis nya skitar berapa dan paling pahitnya abis berapa?	Posi-tive
User 4	Amazon (TR): Mi Robot VacuumMop P Disposable Mop Pad Yedek Parça https://t.co/ma3AxU9wNH 211,58 TL (%31 indirim) #ApolloDeals #işbirliği https://t.co/M9H8V0YPAh	Posi-tive
User 5	Xiaomi Mijia Smart Yuba opens pre-sale https://t.co/RmpwfkzuGe #Xiaomi #Xiaomimijia #smartyuba	Posi-tive

Pada Tabel 6 terdapat data sebanyak 329 baris data yang mengandung kata kunci *samsung*, dari 329 baris data menghasilkan data *sentiment analysis* yaitu untuk data netral sebanyak 108 data, data positif sebanyak 208 data, dan data negatif sebanyak 13 data sehingga data pada Tabel 6 memiliki nilai presentase positif sebesar 63,2%. Pada Tabel 7 terdapat data sebanyak 454 baris data yang mengandung kata kunci *iphone*, dari 454 baris data menghasilkan data *sentiment analysis* yaitu untuk data netral sebanyak 266 data, data positif sebanyak 127 data, dan data negatif sebanyak 61 data sehingga data pada Tabel 7 memiliki nilai presentase positif sebesar 27,97%. Pada Tabel 8 terdapat data sebanyak 464 baris data yang mengandung kata kunci *xiaomi*, dari 464 baris data menghasilkan data *sentiment analysis* yaitu untuk data netral sebanyak 226 data, data positif sebanyak 176 data, dan data negatif sebanyak 62 data sehingga data pada Tabel 8 memiliki nilai presentase positif sebesar 37,93%. Dari hasil Tabel 6 hingga Tabel 8 dapat disimpulkan bahwa produk *Samsung* merupakan yang memiliki opini paling baik dikarenakan memiliki nilai presentase positif sebesar 63,2%.

V. PENUTUP

Berdasarkan hasil penelitian yang berjudul “*Sentiment Analysis* pada *Twitter* untuk Perbandingan Produk Secara *Real-Time* dengan Menggunakan Pendekatan *Lexicon Based*” ini, peneliti memperoleh beberapa kesimpulan yaitu: Nilai *accuracy* yang diperoleh dari penelitian ini sebesar 68%, klasifikasi untuk kategori negatif memiliki nilai terkecil dibandingkan dengan kategori yang lainnya dapat dilihat dari nilai *f1-score*nya hanya sebesar 34% saja, produk *Samsung* merupakan produk yang memiliki opini yang paling baik dikarenakan memiliki nilai presentase positif sebesar 63,2%, pada penelitian ini kurang optimalnya dalam melakukan *text preprocessing*, dan banyaknya *slangword*, salah pengejaan kata, dan juga kalimat satire mempersulit untuk mendapatkan hasil yang optimal.

Saran-saran yang dapat penulis berikan berdasarkan penelitian ini yaitu: sebaiknya melakukan *text processing* dengan lebih baik lagi dengan cara menambahkan lebih banyak langkah pada saat melakukan tahap *text processing*, sebaiknya menggunakan metode lain selain *lexicon based* agar dapat memiliki pembandingan dari penelitian ini, dan yang terakhir sebaiknya untuk pendekatan *lexicon based* diterapkan pada *dataset* yang memiliki bahasa baku seperti portal berita sehingga dapat menghasilkan hasil akhir yang maksimal.

DAFTAR PUSTAKA

- [1] Z. Jianqiang and G. Xiaolin, “Comparison research on text pre-processing methods on twitter sentiment analysis,” *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.
- [2] S. H. Biradar, J. V. Gorabal, and G. Gupta, “Machine learning tool for exploring sentiment



- analysis on twitter data,” *Mater Today Proc*, vol. 56, pp. 1927–1934, 2022, doi: 10.1016/j.matpr.2021.11.199.
- [3] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, “Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning.”
- [4] M. R. Huq, A. Ali, and A. Rahman, “Sentiment Analysis on Twitter Data using KNN and SVM,” 2017. [Online]. Available: www.ijacsa.thesai.org
- [5] M. Wongkar and A. Angdresey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter.”
- [6] “Cara Mendapatkan API Twitter dan Menggunakannya.” <https://klikjon.com/cara-mendapatkan-api-twitter/> (accessed Sep. 29, 2022).



©2022. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).