SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 11 No 1 – Agustus 2025

Perbandingan Pelabelan Otomatis Textblob Dan Vader Lexicon Terhadap Sentimen Ulasan *Mobile Banking* Livin' By Mandiri

Sulton Bagus Permadi[1], Mujib Ridwan[2], Ahmad Yusuf[3]

Program Studi Sistem Informasi, Fakultas Sains & Teknologi, UIN Sunan Ampel Surabaya

H76217068@uinsby.ac.id[1], mujibrw@uinsa.ac.id[2], ahmadyusuf@uinsa.ac.id [3]

Kata Kunci

Analisis Sentimen, Ulasan Google Play Store, TextBlob, Vader Lexicon, Livin' by Mandiri.

Abstrak

Salah satu produk mobile banking Livin' by Mandiri mengalami kelonjakan jumlah pengguna aktif selama periode 3 tahun terakhir. Ulasan yang diberikan oleh para nasabah atau pengguna Livin' by Mandiri bisa ditinjau dengan memanfaatkan fitur Google Play Store. Tujuan dilakukannya penelitian ini adalah demi mengetahui sentimen terhadap Livin' by Mandiri dengan membandingkan pengukuran unsupervised learning yakni TextBlob dan Vader Lexicon. Data yang diperoleh dari proses scrapping google play sebanyak 7.952 ulasan dari tanggal 6 Oktober 2023 sampai 5 Maret 2024. Hasil yang diperoleh setelah dibandingkan dengan rating, akurasi pelabelan TextBlob sebesar 44%, sedangkan pelabelan Vader Lexicon sedikit lebih tinggi sebesar 53%. Terjadinya perbedaan pelabelan sebesar 30%, sedangkan persamaan pelabelan sebesar 70%. Frekuensi kata "tidak" cukup tinggi, tidak hanya ada pada ulasan yang berlabel negatif saja, melainkan berlabel positif dan netral juga.

Keywords

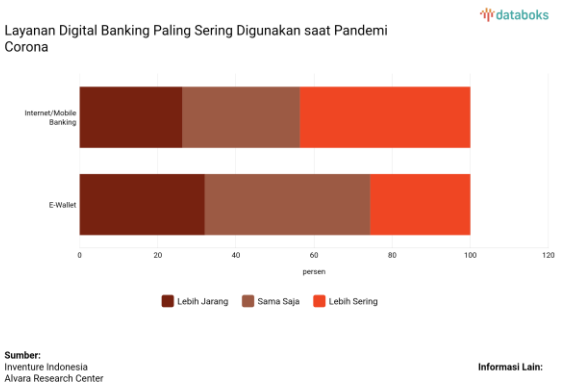
Sentiment Analysis, Google Play Store Reviews, TextBlob, Vader Lexicon, Livin' by Mandiri

Abstract

One of our mobile banking products, Livin' by Mandiri, has experienced a surge in the number of active users over the past 3 years. Reviews given by customers or users of Livin' by Mandiri can be reviewed by utilizing the Google Play Store feature. The purpose of this research is to determine the sentiment towards Livin' by Mandiri by comparing unsupervised learning measures, namely TextBlob and Vader Lexicon. The data obtained from the Google Play scrapping process is 7,952 reviews from October 6, 2023 to March 5, 2024. The results obtained after being compared with the rating, TextBlob labeling accuracy is 44%, while Vader Lexicon labeling is slightly higher at 53%. The occurrence of labeling differences was 30%, while the similarity of labeling was 70%. The frequency of the word "no" is quite high, not only in negatively labeled reviews, but also in positive and neutral ones.

1. Pendahuluan

Perkembangan teknologi informasi pada era sekarang ini semakin mendorong berbagai instansi pemerintahan maupun kalangan swasta dalam memberikan pelayanan agar menjadi lebih mudah, efisien, dan juga responsif kepada masyarakat. Salah satunya instansi publik yang dimaksud adalah Bank, yang menyediakan layanan transaksional keuangan yang paling banyak dimanfaatkan oleh masyarakat juga berupaya menciptakan inovasi pelayanannya ke dalam sebuah produk digital.

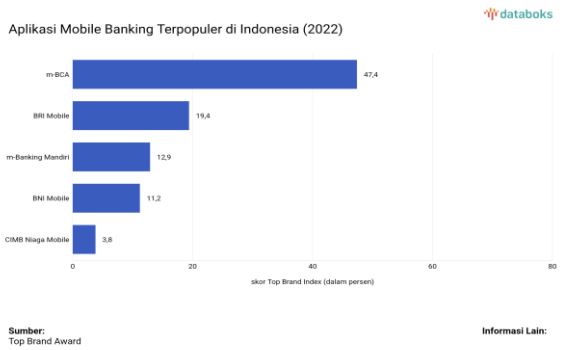


Gambar 1. Layanan transaksi digital yang sering dipakai saat pandemi (Inventure Indonesia, Alvara Research Center, Databoks, 2021)

DOI : 10.29080/systemic.v11i1.2173
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

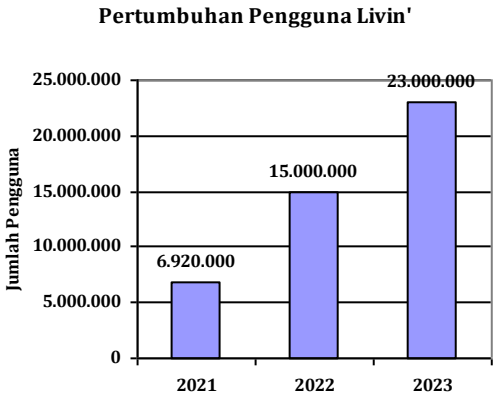
Berdasarkan Gambar 1, lembaga Inventure Indonesia bersama Alvara Research Center melakukan survei dikala pandemi Covid-19 masih merambah pada periode Agustus-September 2020. Dari pilihan layanan digital atau mobile banking dengan *e-wallet*, didapati hasil bahwa minat konsumen terhadap pemanfaatan layanan *mobile banking* sangat tinggi. Pada survei tersebut memaparkan bahwa 43,6% responden mengaku lebih sering menggunakan layanan *moblie banking*, sedangkan 25,6% responden yang lebih sering menggunakan layanan *e-wallet*.

Mobile banking sendiri merupakan sebuah layanan informasi berbasis aplikasi yang disediakan dan dijamin keamanannya oleh bank tertentu guna mempermudah para nasabah dalam melakukan berbagai macam transaksi keuangan yang dapat diakses dengan menggunakan *smartphone*. Salah satu bank yang termasyhur di kalangan masyarakat yang juga turut meluncurkan produk layanan *mobile banking* adalah PT. Bank Mandiri (Persero) Tbk. dengan aplikasinya yang bernama Livin' by Mandiri. Livin' by Mandiri ialah aplikasi *mobile banking* yang dirancang dengan konsep *financial superapp*, di mana para penggunanya dapat secara praktis dalam mengakses fitur layanan Bank Mandiri melalui perangkat ponsel pintar mereka seperti membuat rekening baru, melihat daftar catatan riwayat mutasi transaksi, melakukan pembayaran tagihan, transfer uang, setor dan tarik tunai, dan transaksi perbankan lainnya [1].



Gambar 2. Aplikasi moblie banking paling populer di Indonesia (Top Brand Award, Databoks, 2022)

Menurut survei Databoks (2022), sejak pertama kali diluncurkan pada Oktober 2021 lalu Livin' by Mandiri tercatat memiliki 6,92 juta pengguna aktif dengan frekuensi transaksi mencapai 1,04 miliar dengan nominal sebesar Rp. 1.455 triliun. Tertera dalam Gambar 3, Livin' terus mengalami kenaikan jumlah pengguna aktif selama periode 3 tahun terakhir. Dari hasil wawancara CNBC Indonesia dengan Dirut Teknologi Informasi Bank Mandiri, jumlah pengguna aktif naik hingga mencapai 23 juta nasabah dengan akumulasi transaksi pada Livin' mencapai Rp. 3.271 triliun di tahun 2023 silam, sehingga mengindikasikan pertumbuhan yang baik.



Gambar 3. Pengguna Livin' by Mandiri pada periode 2021-2023 (CNBC & Databoks)

Dengan banyaknya pengguna yang beragam, tentunya tak jarang beragam timbal balik yang dilontarkan mengenai kualitas layanan perbankan Livin'. Dari situ pihak penyedia layanan tentunya bisa melihat dan menganalisis kualitas layanannya dari ulasan yang diberikan oleh para nasabah bisa ditinjau dari kolom komentar atau postingan teks pada media sosial dan platform tertentu yang menyediakan fitur ulasan suatu produk. Salah satu media digital yang menyediakan *review* dan rating yakni Google Play Store, yang dimana terdapat berbagai ulasan atau *feedback* pengguna yang terkait persepsi mengenai layanan produk dari suatu perusahaan [2]. Maka dari itu, untuk mengetahui persepsi para nasabah mengenai gambaran kualitas layanan pada Livin' perlu dilakukan analisis mengenai sentimen.

Pada penelitian ini akan dilakukan perbadingan metode pseudolabelling TextBlob dengan Vader Lexicon terhadap sentimen ulasan para nasabah aplikasi Livin' by Mandiri. Data ulasan yang digunakan berasal dari ulasan-ulasan yang terpublikasi pada *google play store*. Ulasan-ulasan nasabah mengenai Livin' yang sudah mengalami perkembangan pada tiap tahunnya cukup menarik untuk dijadikan sebagai objek penelitian. Pemilihan TextBlob dengan Vader Lexicon dikarenakan sering diulas dan dipakai dalam pelabelan otomatis pada penelitian analisis sentimen. Hasil pelabelan yang diperoleh dari komparasi keduanya diharapkan dapat dijadikan referensi dalam memilih dari metode pelabelan dari kedua pesudolabelling tersebut.

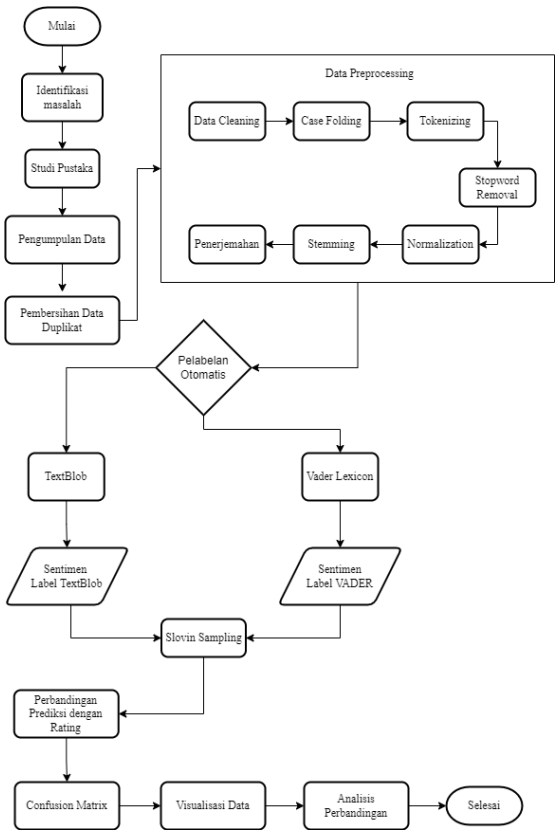
Dari uraian latar belakang yang telah diapaprkan di atas, maka peneliti menentukan rumusan masalah untuk penelitian ini sebagai berikut:

1. Bagaimana keakuratan dari pelabelan sentimen otomatis TextBlob dan Vader Lexicon terhadap sentimen ulasan Livin' by Mandiri?
2. Bagaimana perbandingan dari hasil pelabelan otomatis TextBlob dengan Vader Lexicon terhadap sentimen ulasan Livin' by Mandiri?

Agar penelitian mengenai rumusan masalah bisa lebih terfokus, maka cakupan penelitian ditentukan dengan Google Play Store menjadi sumber data. Data ulasan yang dipakai penelitian adalah ulasan-ulasan berbahasa indonesia dari periode tanggal 6 Oktober 2023 sampai dengan 5 Maret 2024. Data akan diproses menggunakan pengukuran *unsupervised learning* dan tidak menangani ulasan yang mengandung sarkasme. Kemudian hasil dari pelabelan dikelompokkan menjadi sentimen positif, negatif, dan netral.

2. Metode Penelitian

Jenis penelitian yang dilakukan adalah penelitian yang bersifat kuantitatif deskriptif, dimana penerapannya dilakukan berdasarkan sudut pandang mengenai perhitungan yang didapatkan dari hasil penelitian kemudian dipaparkan secara tekstual secara rinci. Data ulasan yang dipergunakan pada penelitian kali ini bersumber dari ulasan-ulasan pengguna Livin' yang telah terposting pada *google play store*. Data diambil dengan *web scrapping* menggunakan program bahasa python. Data yang diperoleh merupakan data sekunder atau data yang diperoleh tanpa melalui proses wawancara dengan narasumber terkait. Penelitian ini mengambil data ulasan terbaru mengenai Livin' by Mandiri. Langkah-langkah penelitian yang dilakukan diilustrasikan ke dalam bentuk flowchart di Gambar 4.



Gambar 4. Alur Penelitian

2.1 Pengumpulan Data

Setelah berbagai referensi penelitian sudah didapatkan, selanjutnya adalah melakukan pengumpulan data. Ulasan mengenai aplikasi Livin' by Mandiri yang tersedia di play sotre akan menjadi sumber data. Pengambilan ulasan play store menggunakan *package google play scrapper* milik JoMingYu di GitHub yang dijalankan dengan menggunakan Google Colab.

2.2 Pembersihan data duplikat

Proses ini perlu dilakukan karena pada saat pengambilan data pada google play store kemungkinan terdapat banyak sekali ulasan yang diambil dengan kondisi waktu ataupun isi yang serupa atau isi ulasan yang *template* seperti oke, ok, mantap, keren, bagus, puas, dan masih banyak lagi. Untuk itu ulasan-ulasan yang terindikasi duplikat harus dihilangkan agar tidak terjadi redundan atau pengulangan proses data yang sama sehingga dapat mempersingkat waktu dan menghasilkan data-data yang unik.

2.3 Data Preprocessing

Pada proses ini data-data sekiranya kurang lengkap, tidak konsisten, maupun kesalahan penulisan akan diolah menjadi sebuah format yang terstruktur dan mudah dipahami. Proses dari *data preprocessing* pada penelitian ini akan melalui beberapa tahapan sebagai berikut.

1. Data Cleaning
   Karakter-karakter yang tidak penting dalam sebuah data seperti tanda baca, link, emoji, link, simbol akan dihilangkan.
2. Case Folding
   Merupakan proses merubah semua data diselaraskan ke dalam format *lowercase* atau berhuruf kecil.
3. Tokenizing
   Proses dimana memecah atau membagi teks ke dalam beberapa kata dengan spasi yang sebagai patokan pemisah.
4. Stopword Removal
   Proses dimana menghilangkan kata-kata penghubung atau sambung, seperti di, ke, dari, dan. Pada tahapan ini juga menggunakan library NLTK terutama *corpus stopword*-nya dengan parameter bahasa Indonesia.
5. Normalization
   Mengubah kembali kata-kata yang masih tidak baku menjadi kata-kata yang sesuai dengan aturan KBBI. Pada proses ini memanfaatkan *corpus* kata baku dari repositori jbkalit di GitHub.
6. Stemming
   Merupakan proses menghilangkan imbuhan dari kata yang telah terbagi menjadi kata dasar, seperti "me-kan", "-nya", "ter-" kemudian digabungkan kembali menjadi satu kalimat.

Stemming dilakukan juga dengan bantuan dari *corpus* kata dasar yang disediakan dalam library Sastrawi, kemudian data yang masih dalam kondisi terpisah selepas dari *tokenizing*, akan digabungkan menjadi satu kalimat kembali untuk mempermudah proses selanjutnya.

7. Penerjemahan Data

Data yang telah dilolah dari data preprocessing akan diterjemahkan ke dalam bahasa inggris dikarenakan library pelabelan otomatis dari python hanya bisa memproses data dalam bahasa inggris saja. Dalam penerjemahan data, penulis menggunakan library google translate dengan versi 3.1.0.

2.4 Pelabelan Data

Seluruh data ulasan yang telah diolah melalui bebagai tahapan *data preprocessing* kemudian dilabeli secara otomatis dengan library python TextBlob dan Vader Lexicon. Label sentimen dengan TextBlob berdasarkan pada nilai polaritas, sedangkan Vader Lexicon berdasarkan pada nilai compoundnya berasarkan dengan kriteria pada Tabel 2. Untuk menggunakan Vader Lexicon cukup dengan menggunakan fitur vader sentimen yang telah disediakan oleh library NLTK.

Nilai polaritas yang dipakai sebagai acuan pengelompokan dalam TextBlob diantara berkisar -1 hingga 1, yang lebih detailnya lagi jika dengan nilai polaritas -1 hingga 0 dinyatakan sebagai sentimen negatif, rentang 0 hingga 1 dinyatakan sebagai sentimen positif, dan terakhir 0 sebagai kriteria sentimen bersifat netral [3]. Uuntuk menentukan indeks tingkantan nilai polaritas dengan memakai fungsi polarity_scores(), yang nantinya akan menghasilkan nilai metrik *compound* <= -0.05 untuk sentimen negatif, nilai metrik *compound* > -0.05 hingga < 0.05 untuk sentimen netral, dan sentimen positif dengan nilai metrik *compound* >= 0.05 untuk kalimat tertentu [4].

Table 2. Kriteria Pelabelan TextBlob dan Vader Lexicon

Label	Polaritas TextBlob	Compound Vader Lexicon
Positif	>= 1	>= 0.05
Netral	0	0.0
Negatif	<= -1	<= -0.05

2.5 Perbandingan Prediksi Label dengan Rating

Dalam penelitian Analisis Sentimen Ulasan Aplikasi Transportasi Online oleh [5], menggunakan skema *average labeling*, dimana berdasarkan pada rating ulasan aplikasi Gojek dan Grab dengan rating 1 – 4 dilabelkan sebagai negatif dan rating 5 sebagai positif. Lalu Analisa Sentimen Pada Ulasan Google Untuk Hotel Gran Mahakam Jakarta oleh [6], yang melabeli ulasan rating 1-3 sebagai label negatif dan rating 4-5 sebagai label positif. Pada penelitian kali ini, ulasan Livin' dinyatakan data pelabelan aktual dimana ulasan yang memiliki rating 1-2 sebagai ulasan negatif, rating 3 sebagai label netral, dan 4-5 sebagai ulasan positif. Data pelabelan prediktif masing-masing dari TextBlob dan Vader Lexicon akan dibandingkan dengan data pelabelan aktual tersebut dan dipilih menggunakan teknik *slovin sampling* ulasan periode waktu paling terbaru. Berdasarkan data aktual tersebut, akan dianalisis keakuratan hasil dari data prediksinya. Berdasarkan dari penelitian yang telah dilakukan oleh Dodi Sukma R.A dkk [7], teknik pengambilan sampel dengan *Slovin* dirumuskan sebagai berikut:

$$n = \frac{N}{1+N(e^2)} \quad (1)$$

Keterangan:

- n: Jumlah sampel yang yang akan diambil atau digunakan
- N: Jumlah populasi data
- e: Batas maksimal toleransi kesalahan yang ditolerir dalam sampel

2.6 Visualisasi Hasil Pelabelan

Hasil dari pelabelan otomatis sebelunya, lalu disajikan ke dalam tampilan diagram batang, lingkaran, dan juga *wordcloud*. Visualsiasi ke dalam grafik batang akan menjadi bahan perbandingan banyaknya jumlah antara data yang berlabel positif, negatif, dan netral. Sedangkan *wordcloud* memaparkan kata-kata yang sering banyak digunakan atau digunakan sebagai ulasan pada label positif, negatif, dan netral masing-masing. Kemudian untuk visualisasi prosentase dari perbandingan hasil kedua pelabelan akan disajikan ke dalam sebuah grafik lingkaran.

2.7 Analisis Perbandingan Pelabelan

Secara keseluruhan hasil yang diperoleh dari pelabelan TextBlob dan Vader Lexicon terhadap ulasan Livin' by Mandiri akan dianalisis menggunakan confusion matrix sehingga diperoleh nilai akurasi, *recall*, presisi, rasio error, dan *f-measure*. Kemudian hasil dari kedua pelabelan tersebut akan dibandingkan dan dianalisis mengenai faktor penyebab yang mempengaruhi persamaan dan perbedaan hasil pelabelan sentimen. Berikut persamaan dari model Confusion Matrix [8]:

a. Nilai akurasi, dirumuskan dengan persamaan:

$$Accuracy = \frac{TP+TN+TNet}{(TP+FP+TN+FN+TNet+FNet)} \quad (2)$$

b. *Recall*, persamaannya dirumuskan sebagai berikut pada label positif:

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

c. Presisi, berikut perumusan persamanya pada label positif:

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

d. Rasio Eror, untuk perhitungan rasio error dapat dirumuskan seperti berikut:

$$Rasio\ Eror = \frac{FP+FN+FNet}{(TP+FP+TN+FN+TNet+FNet)} \quad (5)$$

e. *F-1 Score*, atau *F-Measure*, dirumuskan sebagai berikut [9].

$$F-1\ Score = 2 \times \frac{precission \times recall}{precission+recall} \quad (6)$$

Keterangan:

1. TP (True Positive): Prediksi label positif pada data yang benar benilai positif.
2. FP (False Positive): Prediksi label positif pada data yang tidak bernilai positif.
3. TN (True Negative): Prediksi label negatif pada data yang benar bernilai negatif.
4. FN (False Negative) : Prediksi label negatif pada data yang tidak bernilai negatif.
5. TNet (True Netral) : Prediksi label netral pada data yang benar bernilai netral.
6. FNet (False Netral) : Prediksi label netral pada data yang tidak bernilai netral.

3. Hasil Dan Pembahasan

3.1 Pengumpulan Data

Proses pengambilan data dari play store Livin'telah dilakukan dengan memanfaatkan package google play scrapper milik JoMingYu menggunakan Google Colab. Data yang diambil dengan parameter penyortiran paling terbaru berhasil memperoleh data-data ulasan dari periode tanggal 6 Oktober 2023 hingga 5 Maret 2024. Total data yang diperoleh sebanyak 7.952 ulasan dengan rincian sebagai berikut:

Tabel 3. Jumlah Ulasan Terkumpul

Bulan	Jumlah Ulasan
Oktober	88
November	133
Desember	798
Januari	361
Februari	4.846
Maret	1.726
Total	7.952

Tabel 4. Contoh Data Ulasan

Hasil Pengambilan Data Ulasan
Dengan adanya living lebih mudah bertransaksi semoga ke depan bisa lebih baik lagi,tanks bank mandiri ♡♡♡♡♡
Padahal sudah diperbarui kenapa makin lama makin merugikan gini yaa😥😵😭 sama aja bohong cuma bisa cek saldo😕 mau transaksi gak bisa Ini sih bukan mobile banking namanya kalo cuma buat cek saldo doang😪😔😥 Mau transfer harus tarik tunai dulu cari atm dulu
Selama ini tidak ada kendala, membantu proses transaksi saat malas pergi ke ATM.

3.2 Penghapusan Data Duplikat

Proses selanjutnya adalah menghapus ulasan-ulasan yang diperoleh dari pengambilan data sebelunya yang terindikasi duplikat. Proses ini dilakulan supaya nantinya tidak memproses data yang sama berulang-ulang dan mempercepat proses selanjutnya dengan menggunakan script python. Jumlah data yang tersisa setelah dilakukannya pembersihan data duplikat dicantumkan pada Tabel 5 berikut:

Tabel 5. Jumlah Ulasan Setelah Pembersihan Data Duplikat

Bulan	Jumlah Ulasan
Oktober	64
November	93
Desember	527
Januari	245
Februari	2.063
Maret	511
Total	3.503

3.3 Data Preprocessing

Data ulasan akan diolah menjadi format kalimat atau data yang sama agar menjadi data yang lebih mudah dipahami dan memperoleh hasil yang maksimal. Dalam menangani data preprocessing, penelitian ini memanfaatkan Google Colab sebagai sarana proses pengolahan dengan spesifikasi RAM sebesar 12,67 GB dan Disk sebesar 107,72 GB.

Terjadi perubahan jumlah total ulasan setelah melalui proses *data cleaning*, dikarenakan ada ulasan yang hanya menyantumkan *emoticon* seperti acungan jempol dan hati/love, maka akan terhapus total data menjadi 3.460 ulasan. Lalu pada tahap stopword removal terjadi perubahan

jumlah total data ulasan lagi, dikarenakan beberapa ulasan hanya mengandung unsur kata yang termasuk ke dalam corpus stopword bahasa Indonesia yang akan dihapuskan, sehingga menyisakan 3.434 ulasan.

Tabel 6. Hasil akhir *data prerpocessing*

Hasil *Data Preprocessing*
living transactions are easy, hopefully bank mandiri will continue to tank
i just made a loss, it's just a lie, check the transaction balance or not, mobile banking, if you check the balance, just transfer, withdraw cash, look for an atm
obstacles in helping process transactions, lazy to go to atm

3.4 Pelabelan Data

Dari data yang telah diterjemahkan, selanjutnya bisa dilabeli oleh library TextBlob dan Vader Lexicon yang dari NLTK. Ulasan dilabeli sentimen positif, negatif, dan netral. TextBlob menilai senitmen berdasarkan tingkatan nilai polaritasnya, sendangkan Vader Lexicon tergantung pada besaran nilai *compound*-nya. Pada Tabel 7 ditampilkan hasil dari ulasan yang dilabeli secara otomatis oleh TextBlob dengan Vader Lexicon. Untuk TexBlob akan disimbolkan TB dan Vader Lexicon disimbolkan dengan VL.

Table 7. Hasil Pelabelan TextBlob dan Vader Lexicon

Ulasan	Label TB	Label VL
living transactions are easy, hopefully bank mandiri will continue to tank	Positif	Positif
i just made a loss, it's just a lie, check the transaction balance or not, mobile banking, if you check the balance, just transfer, withdraw cash, look for an atm	Netral	Negatif
obstacles in helping process transactions, lazy to go to atm	Negatif	Negatif

3.5 Perbandingan Prediksi Label dengan Rating

Hasil pelabelan dari TextBlob dan Vader Lexicon dibandingkan dengan memanfaatkan nilai rating yang diberikan pengguna pada ulasannya sebagai data aktualnya. Untuk ulasan dengan nilai rating dari 1 sampai 2 sebagai label negatif, yang bernilai 3 sebagai label netral, sedangkan rating nilai 4 sampai 5 sebagai label positif. Pengambilan beberapa sampel ulasan dilakukan dengan menggunakan teknik *simple random samling* yakni *slovin sampling*.

Dari total data yang berjumlah 3.434 ulasan berlabel, data yang telah dilabeli secara otomatis oleh TextBlob adalah 1.648 ulasan bersentimen positif, 819 ulasan bersentimen negatif, dan 967 ulasan bersentimen netral. Sedangkan data yang dilabeli otomatis oleh Vader Lexicon berjumlah 1.775 ulasan bersentimen positif, 967 ulasan bersentimen negatif, dan 693 ulasan bersentimen netral. Dari banyaknya data ulasan tersebut, diambil sampel dari masing-masing populasi data yang telah dilabeli oleh TextBlob maupun Vader Lexicon dengan menerapkan *margin error* sebesar 0.5% pada rumus 1.

$$n = \frac{3.434}{1 + 3.434\ (0.05^2)} = 358{,}268127$$

Jumlah sampel yang mewakili populasi data ulasan untuk dibandingkan adalah sebanyak 358 ulasan dan dipilih berdasarkan periode waktu yang paling terbaru yakni 1 Maret sampai 5 Maret 2024. Pada Tabel 8 ditampilkan beberapa contoh dari hasil dari perbandingan pelabelan TextBlob dengan nilai rating ulasan atau data aktual, sedangkan untuk pelabelan Vader Lexicon ada pada Tabel 9.

Table 8. Perbandingan Prediksi TextBlob dengan Rating

Ulasan	TB	DA	B/S
Kenapa Harga Kuota internet mahal" dan tidak ada Promo sama sekali Tolong di tinjau kembali..tq	1	1	B
Nyaman dan nda ribet pokoknya beda jauh dgn brimo	0	1	S
Aneh.. Jadi ga bsa login.. Setelah update.parah nih ga profesional	-1	-1	B

Table 9. Perbandingan Prediksi Vader Lexicon dengan Rating

Ulasan	TB	DA	B/S
Kenapa Harga Kuota internet mahal" dan tidak ada Promo sama sekali Tolong di tinjau kembali..tq	1	1	B
Nyaman dan nda ribet pokoknya beda jauh dgn brimo	1	1	B
Aneh.. Jadi ga bsa login.. Setelah update.parah nih ga profesional	-1	-1	B

Keterangan:

- TB: Label TextBlob
- VL: Label Vader Lexicon
- DA : Data Aktual/Rating
- 1: Label Positif

- 0: Label Netral
- -1: Label Negatif
- B: Prediksi label benar
- S: Prediksi label salah

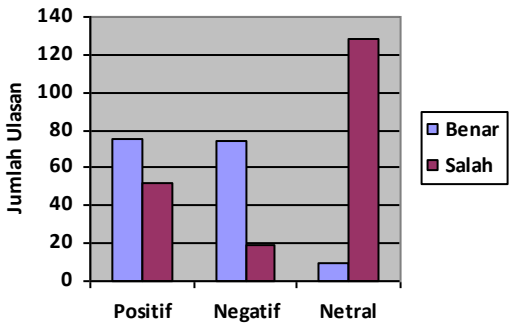
Hasil dari perbandingan hasil pelabelan TextBlob berdasarkan rating disusun ke dalam tabel Confusion Matrix seperti dirincikan pada Tabel 10. Pelabelan TextBlob hanya bisa menghasilkan akurasi sekitar 44% dan rasio errornya jika dibulatkan mendapatkan nilai 56%.

Table 10. Confusion Matrix Sampel Pelabelan TextBlob

Sampel Ulasan		Label Prediksi TextBlob		
		Positif	Negatif	Netral
Label Aktual (Rating)	Positif	75	11	59
	Negatif	45	74	69
	Netral	7	8	10

$$Accuracy = \frac{75 + 74 + 10}{75 + 45 + 7 + 74 + 11 + 8 + 10 + 59 + 69} = 0{,}444134$$

$$Rasio\ Error = \frac{52 + 19 + 128}{75 + 45 + 7 + 74 + 11 + 8 + 10 + 59 + 69} = 0{,}555866$$



Gambar 4. Grafik Akurasi Prediksi TextBlob

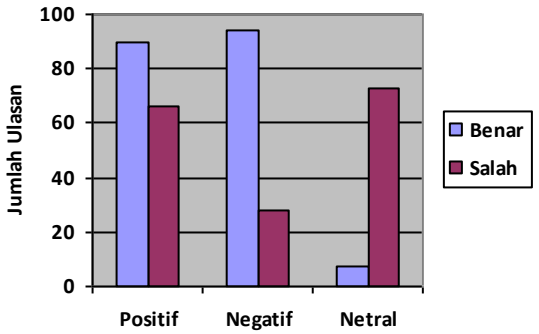
Begitu pula dengan ricinan Confusion Matrix hasil pelabelan Vader Lexicon. Bisa terlihat pada Tabel 11 yang didasarkan pada rating, akurasi yang diperoleh dari pelabelan Vader Lexicon sedikit lebih tinggi daripada TextBlob yakni sebesar 53% dan dengan rasio error sebesar 47%.

Table 11. Confusion Matrix Sampel Pelabelan Vader Lexicon

Sampel Ulasan		Label Prediksi Vader Lexicon		
		Positif	Negatif	Netral
Label Aktual (Rating)	Positif	90	17	39
	Negatif	56	94	34
	Netral	10	9	7

$$Accuracy = \frac{90 + 94 + 7}{90 + 56 + 10 + 94 + 17 + 9 + 7 + 39 + 34} = 0{,}53352$$

$$Rasio\ Error = \frac{66 + 28 + 73}{90 + 56 + 10 + 94 + 17 + 9 + 7 + 39 + 34} = 0{,}46648$$



Gambar 5. Grafik Akurasi Prediksi Vader Lexicon

Dari Gambar 4 dan 5 dapat diperoleh bahwa tingkat kesalahan prediksi pelabelan netral baik TextBlob dan Vader Lexicon bisa dikatakan sangatlah tinggi. Kemungkinan kesalahan prediksi pelabelan netral yang menjadi penyebab terbesar dari kurangnya nilai akurasi dari kedua pelabelan otomatis tersebut.

Table 12. Metrik Evaluasi Hasil Pelabelan TextBlob

Label	Metrik Evaluasi TextBlob		
	Precission	Recall	F1-Score
Positif	0,590551	0,072464	0,429799
Negatif	0,795699	0,291339	0,426513
Netral	0,072464	0,123457	0,091324

Table 13. Metrik Evaluasi Hasil Pelabelan Vader Lexicon

Label	Metrik Evaluasi Vader Lexicon		
	Precission	Recall	F1-Score
Positif	0,576923	0,471204	0,518732
Negatif	0,770492	0,403433	0,529577
Netral	0,0875	0,069307	0,077348

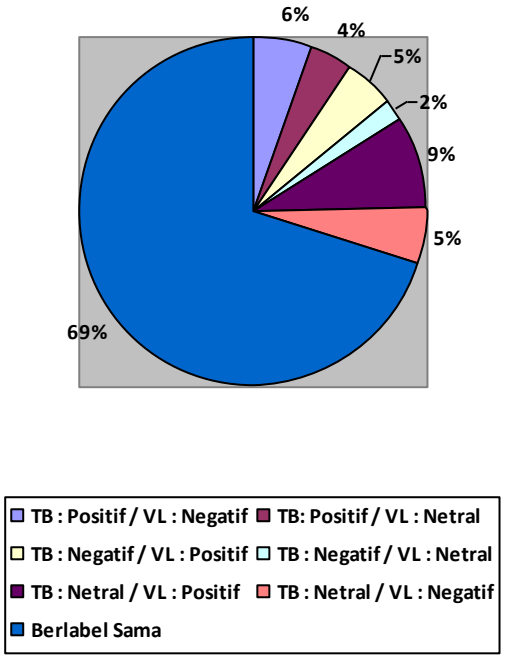
Berdasarkan evaluasi metrik kedua pelabelan pada Tabel 12 dan Tabel 13, diperoleh hasil presisi dari pelabelan sentimen netral pada ulasan cenderung sangat rendah dibanding pelabelan sentimen lainnya. Begitu pula pada hasil pemetaan tabel-tabel confusion matrix sebelumnya, peneliti beranggapan jika kurangnya nilai akurasi disebabkan karena didapati ada ulasan yang diberikan pengguna tidak sesuai dengan nilai rating, membuat prediksi label yang paling parah adalah pelabelan netral diantara label lainnya. Ditambah lagi dengan adanya ulasan yang berisikan kata-kata yang tidak jelas makna atau

maksudnya sehingga diprediksi TextBlob dan Vader Lexicon sebagai label netral.

Pada penelitian ini, penerapan rating 3 ulasan sebagai label data aktual untuk netral. Ulasan dengan nilai rating 3 diantaranya ada yang tidak mengandung ulasan yang bersentimen netral dan berdasarkan analisis peneliti, rating 3 cenderung masih digolongkan oleh para pengguna sebagai tingkat penilaian yang rendah untuk kekurangan yang dimiliki Livin' seperti penilaian ulasan yang buruk dengan pemberian rating 1 dan 2. Data aktual netral yang dipakai dengan acuan rating 3 masih sangat sedikit kemungkinannya bersentimen netral.

3.6 Analisis Perbandingan Hasil Pelabelan

Pada penelitian ini diperoleh total 1.028 ulasan dengan hasil pelabelan berbeda atau sebesar 30% perbedaan pelabelan dari TextBlob dengan Vader Lexicon. Kasus perbedaan pelabelan ulasan yang sering terjadi adalah TextBlob memprediksi sebagai label netral namun pada Vader Lexicon diprediksi sebagai label positif yang mencakup sebesar 9% atau sebanyak 301 ulasan. Disusul dengan ulasan yang diprediksi label positif oleh TextBlob namun berbeda dengan perdiksi Vader Lexicon yaitu dilabeli sebagai negatif sebesar 6% atau seumlah 192 ulasan. Kemudian lebih sedikit ke bawah lagi sebesar 5% atau dengan total 177 ulasan untuk kasus ulasan yang mengalami perbedaan label netral oleh TextBlob sedangkan negatif dari Vader Lexicon.



Gambar 6. Grafik Perbandingan Hasil Pelabelan

Kasus terjadinya perbedaan pelabelan antara hasil label negatif dari TextBlob dan label netral dari Vader Lexicon menjadi paling sedikit terjadi yakni sebesar 2% atau sebanyak 66 ulasan. Total sebanyak 2.406 ulasan memiliki kesamaan hasil label dari kedua sisi pelabelan otomatis, baik itu yang berlabel positif, negatif, ataupun netral. Persamaan hasil prediksi pelabelan oleh kedua pelabelan yang berlabel positif terdapat sebanyak 1.319 ulasan, untuk data yang berlabel negatif sebanyak 598 ulasan, sedangkan data yang berlabel netral diperoleh sebanyak 489 ulasan.

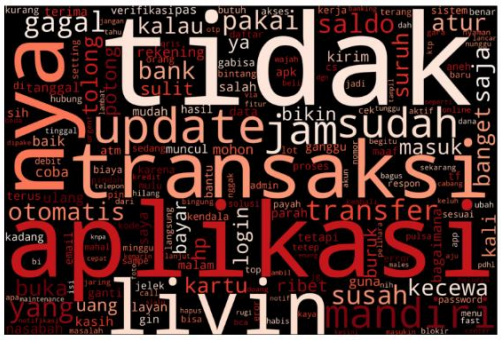
Data ulasan yang telah berlabel sentimen oleh TextBlob juga Vader Lexicon kemudian disajikan ke dalam bentuk *wordcloud*. Kata-kata yang disusun ke dalam wordcloud adalah data ulasan yang telah diproses *stemming* sebelumnya. Berikut *wordcloud* ulasan yang berlabel positif dari TextBlob pada Gambar 7, sedangkan Vader Lexicon pada Gambar 8.

Gambar 7. Wordcloud Positif TextBlob

Gambar 8. Wordcloud Positif Vader Lexicon

Kata-kata yang paling sering muncul atau digunakan dalam ulasan positif dari pelabelan TextBlob adalah mulai dari "tidak", "mudah", "aplikasi", dan "livin. Tak sedikit berbeda dengan Vader Lexicon, hanya runtutannya yang bermula dari kata "aplikasi", "mudah", "livin", dan terakhir adalah "tidak". Kata "mudah", "aplikasi", dan "livin" mengindikasikan bahwa aplikasi Livin' tergolong masih mudah digunakan dalam kegiatan transaksi perbankan oleh banyak pengguna.

Selanjutnya *wordcloud* ulasan-ulasan yang diprediksi sebagai label negatif. Wordcloud pelabelan negatif TextBlob terdapat pada Gambar 9 dan Vader Lexicon ada pada Gambar 10 di bawah berikut.



Gambar 9. Wordcloud Negatif TextBlob

Gambar 10. Wordcloud Negatif Vader Lexicon

Dari *wordcloud* TextBlob dapat diperoleh yang bernilai negatif diantaranya "tidak", "nya", "transaksi", "aplikasi", dan "livin". Sedikit berbeda dari *wordcloud* label negatif dari Vader Lexicon yang juga ada tambahan kata yang tersorot seperti "update" dan "jam". Kata imbuhan "–nya" pada wordcloud negatif TextBlob dan Vader Lexicon termasuk yang sering muncul. Kata imbuhan terlewatkan dari proses *stemming* kemungkinan dikarenakan kebiasaan pengguna dalam penulisan atau pengetikan imbuhan "-nya" selalu dipisah atau diberi jarak spasi.

Dari kata "upadate" yang sering muncul pada wordcloud negatif, menujukkan bahwa mayoritas pengguna mengeluhkan adanya saran pembaharuan versi terbaru terhadap aplikasi Livin' yang namun merasa tidak ada perubahan signifikan, loading pemrosesan waktu membuka aplikasi atau sendang dalam suatu transaksi yang memakan waktu, maupun mengalami kegagalan pembaharuan dikarena spesifikasi device yang kurang mumpuni. Kemudian kata "jam" dimana banyak yang mengeluhkan ketika malam terutama dini hari para pengguna aplikasi Livin' tidak bisa mengakses layanan dikarenakan pada periode tersebut sedang dilakukan *maintenance*.

Terakhir adalah wordcloud hasil pelabelan perdiktif sentimen netral pada Gambar 11 untuk TextBlob dan Gambar 12 untuk Vader Lexicon.

Pada wordcloud label netral TextBlob muncul kembali kata-kata yang ada dalam wordcloud positif dan negatif sebelumnya, begitu pula dengan wordcloud Vader Lexicon. Seperti kata "tidak" dan "aplikasi" misalnya. Kata yang "bantu" dan "update" menjadi pembeda TextBlob dari Vader Lexicon yang muncul juga kata "buka", "jam", dan juga kata imbuhan "-nya" lagi. Berdasarkan kata "bantu", "buka", dan "jam" yang sering muncul, menyatakan bahwa para pengguna dalam memberikan ulasan netral ini, cenderung meminta bantuan solusi atau jawaban dari ketidaktahuan penyebab Livin' tidak bisa diakses pada waktu malam atau dini hari, lama proses membuka aplikasi atau layanan Livin', hingga kendala lainnya dalam penggunaan aplikasi Livin' untuk para petugas atau *customer service* yang bertugas melayani/membalas ulasan-ulasan yang telah dilontarkan.

Gambar 11. Wordcloud Netral TextBlob

Gambar 12. Wordcloud Netral Vader Lexicon

4. Kesimpulan

Berdasarkan dari rangkaian proses penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut bahwa:

1. Berdasarkan data ulasan Livin' by Mandiri pada periode 6 Oktober 2023 sampai dengan 5 Maret 2024, hasil akurasi dari pelabelan TextBlob setelah dilakukannya pembandingan data label prediksi berdasarkan dengan nilai rating masing-masing ulasan diperoleh sebesar 44%, sedangkan akurasi dari pelabelan Vader Lexicon sedikit lebih tinggi dengan nilai sebesar 53%. Meski belum masuk dalam kriteria yang dikatakan akurat, pada penelitian ini diperoleh hasil akurasi Vader Lexicon lebih tinggi dibanding dengan TextBlob.
2. Perbandingan hasil prediktif pelabelan TextBlob dan Vader Lexicon dalam penelitian ini terdapat perbedaan jumlah ulasan yang diprediksi. Jumlah ulasan yang diprediksi sebagai label positif dan negatif oleh Vader Lexicon lebih tinggi dari TextBlob, namun

TextBlob mempredikisi ulasan sebagai label netral lebih banyak dari Vader Lexicon. Kasus perbedaan pelabelan terjadi sebesar 31% atau mencakup 1.028 ulasan, sedangkan kasus persamaan pelabelan terjadi sebesar 69 % mencakup 2.406 ulasan.

Daftar Pustaka

[1] T. Setiawan, "Sosialisasi Pemanfaatan Aplikasi Digital Platform Livin' byMandiri pada MasyarakatPenerima Bantuan Sosial," *AbdiMU Pengabdi. Kpd. Masy.*, vol. 3, no. 1, pp. 1–5, 2023, [Online]. Available: https://jurnal.masoemuniversity.ac.id/index.php/abdimu/article/view/722/494.

[2] N. Nurfadila, M. Ariyanti, and N. Trianasari, "Analisis Kualitas Layanan Mobile Banking New Livin' By Mandiri Menggunakan Sentiment Analysis," *JIBR J. Indones. Bus. Res.*, vol. 1, no. 1, pp. 77–82, 2023, [Online]. Available: http://doi.org/10.25124/logic.v1i1.6486.

[3] T. S. Az-Zahra, "Analisis sentimen terhadap belajar daring menggunakan optimasi naive bayes classifier dengan adaboost," 2021, [Online]. Available: digilib.uinsby.ac.id.

[4] F. Fazrin, N. O. Pratiwi, and R. Andreswari, "Perbandingan Algoritma K-Nearest Neighbor dan Logistic Regression pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan Vader dan Textblob," *J. e-Proceeding Eng.*, vol. 10, no. 2, pp. 1596–1604, 2023.

[5] Y. Anugrah *et al.*, "Analisis Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Multinomial Naïve Bayes dan Query Expansion Ranking," *J. Elektron. Ilmu Komput. Udayana*, vol. 11, no. 1, pp. 121–128, 2022.

[6] M. R. Prasetyo and A. Fahrurozi, "ANALISA SENTIMEN PADA ULASAN GOOGLE UNTUK HOTEL GRAN MAHAKAM JAKARTA MENGGUNAKAN HOTEL GRAN MAHAKAM JAKARTA MENGGUNAKAN," *J. Ilm. Inform. Komput.*, vol. 28, pp. 203–217, 2023.

[7] Dodi Sukma R.A, R. Hardianto, and Heleni Filtri, "Analisa Tingkat Kepuasan Mahasiswa Terhadap Perkuliahan Daring Pada Era Pandemi COVID-19," *Zo. J. Sist. Inf.*, vol. 3, no. 2, pp. 130–142, 2021, doi: 10.31849/zn.v3i2.8353.

[8] M. Karim, *Analisis Sentimen Pada Twitter Menggunakan Support Vector Machine Dengan Modifikasi Lexicon Inset Dan Sentis-Trength _ Id (Studi Kasus : Vaksinasi Covid-19)*. 2022.

[9] M. Hanafi, "Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM)," 2023, [Online]. Available: digilib.uinsa.ac.id.