
Analisis Sentimen Berbasis Aspek Aplikasi Brimo berdasarkan Ulasan Pengguna di Google Playstore.

Yosia Azarya¹, Indra Budi²

yosia.azarya@ui.ac.id¹, indra@cs.ui.ac.id²

^{1,2} Universitas Indonesia

Informasi Artikel

Diterima : 4 Jan 2025
Direvisi : 29 Jan 2025
Disetujui : 7 Feb 2025

Kata Kunci

mobile banking app, model PACMAD, analisis sentimen berbasis aspek, ulasan pengguna, pemodelan topik

Abstrak

BRI memiliki fokus untuk meningkatkan layanan keuangan yang dapat menjadi andalan bagi nasabahnya melalui superapps mobile banking yang dikenal dengan BRImo. Adapun ulasan pengguna yang terdapat di halaman layanan penyedia aplikasi Google Playstore dapat diklasifikasikan baik untuk aspek *usability* PACMAD maupun sentimennya dengan menggunakan beberapa model klasifikasi yaitu Random Forest, Decision Tree, dan Extreme Gradient Boosting dengan memanfaatkan TF-IDF sebagai metode untuk melakukan *feature extraction* dan metode Random Oversampling (ROS) dan Synthetic Minority Oversampling Technique (SMOTE) sebagai *treatment* tambahan untuk menanggulangi *imbalanced class* pada data ulasan pengguna aplikasi BRImo. Selain itu dilakukan pemodelan topik menggunakan metode LDA untuk menemukan kata-kata kunci serta topik pembahasan utama pada masing-masing aspek *usability* PACMAD beserta sentimennya sehingga didapatkan topik yang jelas yang dapat digunakan sebagai fokus pengembangan bagi pengembang aplikasi BRImo. Hasil penelitian menunjukkan bahwa penggunaan model klasifikasi XGBoost dengan menggunakan metode sampling SMOTE menunjukkan performa terbaik dalam melakukan klasifikasi aspek *usability* PACMAD dan sentimennya dengan perhitungan *F1-score* memiliki nilai tertinggi dibandingkan dengan penerapan model *machine learning* lainnya. Selain itu didapatkan juga topik-topik utama dari masing-masing aspek *usability* PACMAD beserta sentimennya.

Keywords

mobile banking app, PACMAD usability model, Aspect Based Sentiment Analysis, user review, topic modeling

Abstract

BRI is focused through their mobile banking superapp known as BRImo. However, based on discussions with the BRImo application research team, The user reviews available on the Google Play Store application service page could be classified based on the PACMAD usability aspects and sentiment using several classification models, including Random Forest, Decision Tree, and Extreme Gradient Boosting, with TF-IDF employed as the feature extraction method. Additionally, the Random Oversampling and Synthetic Minority Oversampling Technique (SMOTE) methods were applied as supplementary treatments to address the issue of imbalanced classes in BRImo application user review data. Topic modeling was also conducted using the LDA method to identify keywords and the main discussion topics for each PACMAD usability aspect and its sentiment, resulting in clear topics that can serve as a focus for the development of the BRImo application. The research findings indicate that the XGBoost classification model, combined with the SMOTE sampling method, demonstrated the best performance in classifying PACMAD usability aspects and sentiments, achieving F1-scores of 86.55% and 89.59%, respectively. Furthermore, the key topics for each PACMAD usability aspect and their associated sentiments were identified.

A. Pendahuluan

PT Bank Rakyat Indonesia (BRI) merupakan salah satu bank yang dimiliki oleh pemerintah terbesar di Indonesia. BRI memiliki lebih dari 150 juta nasabah [1]. Untuk menjadi perusahaan yang memiliki kapitalisasi pasar tertinggi, maka BRI perlu memiliki performa yang baik yang ditunjukkan melalui laba dan perkembangan dari perusahaan.

Dalam perkembangan teknologi digital di layanan finansial, transformasi teknologi menjadi kunci untuk meningkatkan *market share* melalui kemudahan dalam menyediakan layanan secara efisien dan menjaga kenyamanan nasabah yang ada. BRI terus memperkuat aspek pelayanan dan operasional perbankan agar selalu mengutamakan kenyamanan nasabah dan memperluas pelayanan ke seluruh Indonesia [2]. Oleh karena itu BRI memiliki fokus untuk meningkatkan layanan keuangan yang dapat menjadi andalan bagi nasabahnya, sehingga diharapkan adanya pertumbuhan CASA yang mencerminkan kesetiaan nasabah BRI melalui penguatan layanan keuangan kepada nasabah umum melalui *superapps mobile banking* yang dikenal dengan BRImo [1].

Kemudahan pengguna dalam menggunakan aplikasi merupakan salah satu aspek penting yang perlu dievaluasi untuk memperoleh informasi mengenai cara mengembangkan aplikasi agar lebih baik. People At the Center of Mobile Application Development (PACMAD) menjadi *framework usability* yang relevan untuk aplikasi yang bersifat *mobile* [3]. *Framework* ini merupakan pengembangan dari *framework usability* yang dikenalkan oleh Bevan sebelumnya [4]. Evaluasi yang berfokus pada aspek usability tentu melibatkan pengguna sebagai referensi utama. Salah satu bentuk masukan dari pengguna dapat diperoleh melalui ulasan. Ulasan pengguna adalah kumpulan teks yang berisi informasi tentang pengalaman mereka saat menggunakan aplikasi, baik pengalaman positif maupun negatif [5]. Saat ini, salah satu sumber data ulasan pengguna adalah ulasan pengguna pada platform penyedia aplikasi Google Playstore.

Beberapa penelitian yang terkait dengan pemanfaatan ulasan pengguna aplikasi BRImo untuk mendapatkan sentimen dari aplikasi dengan melakukan pemanfaatan *machine learning* dengan model Naïve Bayes [6]. Penelitian terdahulu juga telah menggunakan data dari Google Playstore sebagai sumber data untuk melakukan pembelajaran mesin [7]. Kedua penelitian terdahulu mengenai BRImo berfokus pada pembuatan model untuk melakukan klasifikasi sentimen secara umum terhadap aplikasi, namun belum dapat menunjukkan sentimen untuk aspek *usability* yang bermanfaat bagi pengembang aplikasi untuk melakukan evaluasi terhadap aplikasinya. Model *machine learning* lain seperti Decision Tree, ataupun XGBoost disertai dengan *sampling* dengan Random Oversampling (ROS) ataupun Synthetic Minority Oversampling Technique (SMOTE) dapat dimanfaatkan untuk mendapatkan model yang lebih baik dalam melakukan prediksi terhadap sentimen yang berbasis terhadap aspek *usability* PACMAD. Berdasarkan wawancara dengan *product owner* dan tim riset BRImo, terdapat beberapa hal yang menyebabkan masalah yang terjadi. Terdapat akar masalah berupa ulasan pengguna yang belum dapat dimanfaatkan dan diolah secara optimal. Melalui wawancara juga didapatkan bahwa *feedback* yang diberikan masyarakat melalui helpdesk, Google Play, App Store, maupun media sosial terkait dengan bagaimana pengalaman masyarakat dalam menggunakan aplikasi BRImo akan selalu dipantau sentimennya guna

monitoring berkala. Sekarang, semua ulasan yang diberikan pengguna di platform yang tersedia dipantau secara umum sentimennya, namun belum ada pembagian secara lebih detil terkait dengan sentiment yang terbentuk sehingga hasil dari riset masih belum matang untuk menggunakan sentimen dari ulasan yang ada.

Hal ini mendorong peneliti untuk meneliti lebih lanjut terkait dengan pemanfaatan *feedback* yang diberikan oleh pengguna BRImo karena ulasan yang diberikan oleh pengguna dapat menggambarkan permasalahan yang dihadapi, serta membantu pengembang untuk mengetahui kebutuhan dari pengguna khususnya terkait dengan *usability* dari BRImo dengan bantuan *framework* PACMAD. Analisis sentimen terkait *usability* BRImo berdasarkan ulasan pengguna dapat dilakukan untuk melihat respon masyarakat terhadap *update* yang dilakukan setiap minggunya. Sentimen pengguna terkait dengan *update* yang dilakukan dapat membantu pengembang untuk merespon dengan cepat bila terdapat isu yang ditandai dengan sentimen negatif terkait dengan *usability* dari BRImo sehingga dapat melengkapi proses pengembangan BRImo untuk memenuhi kebutuhan pengguna.

Akar masalah yang menjadi fokus penelitian adalah pemanfaatan *feedback* dari pengguna BRImo terutama pada sentimen terkait *usability* berdasarkan PACMAD dari ulasan yang diberikan oleh pengguna melalui permodelan dengan model *machine learning*. Selain itu, peneliti mampu untuk memberikan rekomendasi untuk pengembangan BRImo berdasarkan sentimen serta topik-topik pembahasan dari ulasan pengguna BRImo. Sehingga pertanyaan dari penelitian ini adalah:

1. Apa model terbaik untuk melakukan prediksi aspek *usability* PACMAD dan sentimen aplikasi BRImo berdasarkan ulasan pengguna?
2. Topik apakah yang banyak dibahas pada ulasan pengguna aplikasi BRImo pada setiap aspek *usability* PACMAD serta sentimennya?

Mengacu pada pertanyaan-pertanyaan dari penelitian yang dilakukan, maka tujuan dari penelitian ini adalah:

1. Menentukan model *machine learning* terbaik untuk melakukan analisis sentimen berbasis aspek *usability* PACMAD terhadap ulasan pengguna aplikasi BRImo.
2. Menentukan topik yang banyak dibahas pada ulasan pengguna aplikasi BRImo berdasarkan aspek *usability* PACMAD serta sentimennya.

Penelitian ini diharapkan memberikan manfaat bagi pengembang aplikasi BRImo berupa rekomendasi pengembangan berdasarkan analisis sentimen berbasis aspek yang dilakukan pada ulasan pengguna. Selain itu penelitian ini memiliki manfaat teoretis terkait penggunaan model PACMAD sebagai aspek *usability* untuk melakukan analisis sentimen pada aplikasi di Google Play.

B. Tinjauan Literatur

Model *Usability* PACMAD

Menurut ISO 9241-11, *usability* didefinisikan sebagai kemampuan sistem, produk, ataupun layanan untuk mencapai suatu tujuan dengan mempertimbangkan efektifitas, efisiensi, serta kepuasan dari pengguna dalam suatu konteks tertentu [4]. *Usability* juga disebutkan sebagai atribut kualitas yang menilai dan mengevaluasi seberapa mudah tampilan pengguna untuk digunakan [8]. Baik penelitian Bevan maupun penelitian Nielsen & Budiu mendefinisikan model *usability* secara umum dalam konteks penggunaan suatu sistem, terutama untuk aplikasi *desktop*. Di sisi lain, penelitian Harrison et al. pada tahun 2013 berargumen bahwa model tersebut memiliki limitasi atau batasan untuk aplikasi berbasis *mobile* pada perangkat seluler. Beberapa limitasi tersebut merupakan perbedaan-perbedaan antara aplikasi *desktop* dan *mobile*, di antaranya yaitu konteks *mobile*, konektivitas, ukuran layar kecil, resolusi layar yang berbeda, kemampuan dan daya pemrosesan yang terbatas, dan metode entri data yang berbeda [9].

Harrison et al. mengatakan bahwa untuk aplikasi *mobile* di dalam perangkat seluler butuh model *usability* yang lebih spesifik. Oleh karena itu, dikembangkan model yang diberi nama People at the Center of Mobile Application Development (PACMAD). Model *usability* PACMAD memodifikasi model *usability* yang dikembangkan oleh Bevan dan Nielsen & Budiu. Penelitian Bevan mengidentifikasi lima dimensi *usability*, yaitu *efficiency*, *satisfaction*, *learnability*, *memorability*, dan *errors*, sedangkan penelitian Nielsen & Budiu hanya menganggap *effectiveness*, *efficiency*, dan *satisfaction* sebagai tiga dimensi dari *usability*. Di sisi lain, terdapat tiga faktor yang perlu dipertimbangkan dalam mengevaluasi *usability*, yaitu pengguna, tujuan, dan konteks penggunaan [4]. Oleh karena itu, Harrison et al. merancang model PACMAD dengan menggabungkan kedua model *usability* penelitian sebelumnya dan menambahkan satu dimensi baru, yaitu *cognitive load* karena dimensi ini secara langsung memengaruhi dan dapat dipengaruhi oleh *usability* dari aplikasi *mobile*. Berikut adalah aspek-aspek *usability* menurut model PACMAD [3]:

1. *Effectiveness*; adalah aspek yang menjelaskan bagaimana pengguna mampu menyelesaikan tugas dalam konteks tertentu.
2. *Efficiency*; adalah aspek yang menggambarkan kemampuan pengguna dalam menyelesaikan tugas yang ada dengan cepat dan akurat.
3. *Satisfaction*; adalah aspek yang berkaitan dengan tingkat kenyamanan dan kesenangan yang dirasakan serta yang diberikan kepada pengguna oleh suatu perangkat lunak dalam penggunaannya.
4. *Learnability*; adalah kemudahan pengguna dalam memperoleh kemahiran pada aplikasi.
5. *Memorability*; adalah kemudahan pengguna untuk mengingat cara kerja aplikasi.
6. *Errors*; adalah aspek yang berkaitan pada kegagalan atau eror yang terjadi pada pengguna dalam menggunakan aplikasi.
7. *Cognitive load*; menjelaskan mengenai aspek kognitif yang bekerja dalam pengguna menggunakan aplikasi.

Analisis Sentimen Berbasis Aspek

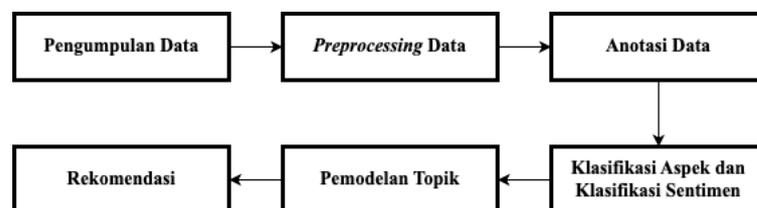
Analisis sentimen yang dilakukan untuk mengungkap opini, penilaian, sikap, dan emosi dari publik pada suatu objek tertentu seperti produk, layanan, organisasi, individu, isu, kejadian, topik, dan sebagainya. Hasil dari opini publik yang terbentuk adalah dalam bentuk sentimen positif dan negatif [10]. Analisis sentimen sudah digunakan dalam banyak bidang ilmu, mulai dari pengembangan produk, bidang ilmu kesehatan, layanan finansial, hingga politik. Analisis sentimen dalam dilakukan dengan metode *supervised learning* maupun *unsupervised learning* [10].

Analisis sentimen dalam dilakukan dalam tiga tingkatan besar, yaitu pada tingkat dokumen, pada tingkat kalimat, dan pada tingkat aspek. Pada tingkat dokumen dan kalimat, analisis sentimen dapat melihat sentimen positif ataupun negatif yang terbentuk pada suatu dokumen ataupun kalimat. Namun untuk mendapatkan detail informasi dari entitas ataupun fitur yang dilihat sentimennya diperlukan adanya sentimen analisis pada tingkat aspek. Salah satu keuntungan melakukan sentimen analisis berbasis aspek adalah peneliti mampu menyesuaikan aspek yang akan digunakan untuk penelitian [11]. Analisis sentimen berbasis aspek dapat dilakukan dengan melakukan *text mining* melalui klasifikasi sentimen maupun aspek dari setiap data yang berupa ulasan.

Text mining merupakan bagian di dalam bidang *data mining* di mana data yang diolah merupakan data berbentuk teks [12]. *Text mining* adalah sebuah proses untuk melakukan penelusuran dan penelitian lebih dalam terkait dengan data teks yang bersifat tidak terstruktur. *Text mining* dapat dilakukan bantuan perangkat lunak berupa program yang dapat menemukan konsep, pola, kata kunci serta hal-hal lain yang tersembunyi yang ada di dalam data teks yang mentah. Secara umum, proses-proses *text mining* menurut kerangka kerja Hu dan Liu dimulai dari data collection, dilanjutkan dengan *text processing*, kemudian *text representation*, lalu tahapan terakhir adalah *knowledge discovery*.

C. Metode Penelitian

Penelitian ini dilakukan dengan 6 tahapan yang ditunjukkan pada Gambar 1 yaitu melakukan pengumpulan data, melakukan *preprocessing* data, melakukan pelabelan data melalui anotasi data, melakukan klasifikasi aspek *usability* PACMAD dan klasifikasi sentimen, melakukan pemodelan topik, serta memberikan rekomendasi berdasarkan hasil klasifikasi aspek dan sentimen beserta hasil pemodelan topik.



Gambar 1. Diagram Alur Tahapan Penelitian

Pengumpulan Data

Data yang digunakan adalah ulasan pengguna aplikasi BRImo pada halaman penyedia layanan Google Play. Data ini diambil dengan menggunakan bantuan

library google-play-scrapper pada bahasa pemrograman Python. Pada prinsipnya data dikumpulkan dengan teknis *web scraping* pada halaman web yang menyediakan informasi mengenai ulasan pengguna [13][14]. Informasi yang diperlukan untuk mendapatkan ulasan ini adalah id dari aplikasi BRImo yang dapat ditemukan melalui *query parameter* id pada uniform resource identifier di situs Google Play. Ulasan yang dikumpulkan dibatasi sampai waktu rilis BRImo. Informasi utama yang digunakan sebagai masukan untuk menjalankan fungsi tersebut adalah id dari aplikasi dan *country_id* dari aplikasi.

Adapun data tersebut merupakan ulasan pengguna sejak bulan April 2021 hingga bulan Agustus 2024 seperti dapat dilihat pada Gambar 2. Data ulasan pengguna BRImo yang dikumpul berisikan informasi mengenai reviewID yaitu id dari ulasan, nama dari user yang memberikan ulasan, gambar dari user yang memberikan ulasan, content dari ulasan yang diberikan, nilai yang diberikan untuk aplikasi BRImo, jumlah jempol yang diberikan untuk ulasan tersebut, versi aplikasi ketika ulasan dibuat, waktu dibuatnya ulasan, balasan dari admin aplikasi BRImo terhadap ulasan, waktu dibuatnya balasan terhadap ulasan, serta versi aplikasi. Data hasil scraping kemudian disimpan dalam file dengan format csv.

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
a660b170-3e1e-4237-a143-5a8619ea9fa5	cris yanty	lh.googleusercontent.com/a/ACg8oc...	Mantaps 🙌	5	0	2.67.1	2024-08-29 07:30:44	NaN	NaN	2.67.1
20c928ba-bd57-481c-91c5-e236ce701304	Armanda Kholili	lh.googleusercontent.com/a-/ALV-U...	udah normal lagi aplikasinya	4	1	2.67.1	2024-08-29 08:37:30	NaN	NaN	2.67.1
c5ff60d0-cf05-4991-9bab-1b3464929b5d	Erte Bahlul	lh.googleusercontent.com/a-/ALV-U...	kenapa gak bisa-bisa daftar? error terus	3	0	2.67.1	2024-08-29 08:36:37	NaN	NaN	2.67.1
1519398a-42d0-4608-b230-87a9003e8af0	Salman Alfariis	lh.googleusercontent.com/a-/ALV-U...	so helpfull and has a low rate of error	5	0	2.67.0	2024-08-29 06:12:27	NaN	NaN	2.67.0
84413323-97f2-4e2b-a973-aade4cb624d	Jhovan Prasetyo	lh.googleusercontent.com/a/ACg8oc...	Jossss	5	0	2.67.1	2024-08-29 05:54:45	NaN	NaN	2.67.1

Gambar 2. Contoh Ulasan Pengguna BRImo

Preprocessing Data

Ulasan pengguna yang dikumpulkan merupakan dataset mentah (*raw dataset*) yang perlu diolah terlebih dahulu. Pengolahan perlu dilakukan untuk menghilangkan pengotor pada data. Pengotor pada data antara lain adalah *stopword*, angka, tanda baca, serta imbuhan. Maka dari itu dilakukan tahapan-tahapan *preprocessing* yaitu: *short review removal*, *casefolding*, *stopword removal*, *punctuation removal*, *number removal*, *stemming*, dan *lemmatization*. *Short review removal* dilakukan dengan cara memotong ulasan pengguna yang memiliki jumlah kata kurang dari lima kata. Hal tersebut dilakukan untuk menghilangkan ulasan yang tidak memiliki konteks yang spesifik baik secara sentimen maupun aspek *usability* PACMAD [15]. *Stopword removal*, *punctuation removal*, dan *number removal* dilakukan untuk menghilangkan pengotor berupa tanda baca, angka, serta *stopword*. *Stopword* perlu dihilangkan karena memiliki pengaruh yang minim terhadap sentimen, selain itu *stopword*, angka, dan tanda baca perlu dihilangkan untuk mengurangi data yang perlu diolah pada tahapan klasifikasi sentimen [16][17][18]. Imbuhan pada kata juga dihilangkan dengan cara mengubah kata menjadi bentuk dasarnya melalui proses *stemming* [17]. Tahapan-tahapan tersebut akan menghasilkan dataset yang siap untuk menuju proses selanjutnya, atau disebut

dengan *preprocessed dataset*. Maka contoh hasil data ulasan yang telah melalui tahapan-tahapan tersebut ditunjukkan oleh Tabel 1.

Tabel 1. Contoh Hasil *Preprocessing* Data

Ulasan sebelum <i>preprocessing</i>	Ulasan setelah <i>preprocessing</i>
<ul style="list-style-type: none"> • Jossss • Jelek. Lamban dalam menangani keluhan nasabah. Tidak profesional • GREAT • BAGUS • Easy to use and very secure for transection. Thanks 	<ul style="list-style-type: none"> • jelek lamban menangani keluhan nasabah profesional • easy use secure transection thanks

Anotasi Data

Dataset hasil *scraping* dan *preprocessing* selanjutnya dilabel secara manual baik untuk aspek *usability* PACMAD maupun sentimennya. Data yang dianotasi berjumlah 9541 ulasan yang dilakukan oleh tiga anotator selama rentang waktu 14 hari. Kriteria anotator yang digunakan oleh peneliti adalah minimal memiliki ijazah Sarjana, menggunakan aplikasi *mobile banking*, dan fasih dalam bahasa inggris. Dalam menentukan aspek *usability* PACMAD dari ulasan pengguna, anotator memiliki patokan terhadap definisi aspek-aspek PACMAD pada penelitian Harrison tahun 2013. Selain itu dilakukan pengarahan oleh peneliti untuk menjelaskan aspek *usability* PACMAD dan penentuan sentimen untuk setiap ulasan. Dalam menentukan sentimen, anotator menentukan berdasarkan kata-kata dan tendensi dari ulasan tersebut. Hasil anotasi dikumpulkan menjadi satu file dengan format xlsx dan diolah untuk mendapatkan label final. Dari tiga label yang diberikan oleh anotator baik untuk aspek *usability* PACMAD maupun sentimennya akan dipilih berdasarkan majority voting.

Jika hasil anotasi tidak dapat menentukan label final akibat terjadinya deadlock karena semua anotator memberikan label berbeda untuk ulasan tertentu, maka label final akan ditentukan oleh peneliti. Selain itu, jika anotator memiliki kebingungan dalam menentukan label, maka anotator akan berkonsultasi dengan peneliti untuk menentukan label. Hal ini akan berdampak pada bias yang mungkin terjadi akibat hasil label akan mengarah pada pengertian dari peneliti.

Klasifikasi Aspek dan Klasifikasi Sentimen

Dataset yang telah melalui tahapan *preprocessing* dan telah memiliki label selanjutnya dapat digunakan untuk melakukan klasifikasi aspek *usability* PACMAD dan sentimen. Adapun terdapat tujuh kelas pada aspek *usability* PACMAD dan dua kelas pada sentimen. Dikarenakan terdapat adanya imbalanced class pada hasil anotasi data aspek *usability* PACMAD dan sentimennya, maka diterapkan skenario dengan adanya penerapan metode *oversampling* dengan menggunakan teknik ROS dan SMOTE. Model klasifikasi yang digunakan adalah RF, DT, dan XGBoost. *Feature extraction* dilakukan dengan TF-IDF untuk semua scenario. Seluruh skenario eksperimen untuk melakukan klasifikasi aspek *usability* PACMAD dan sentimennya dapat dilihat pada Tabel 2. Dari setiap skenario eksperimen kemudian dibandingkan

hasilnya dengan menggunakan *confusion matrix* dan perhitungan *F1-score* untuk menentukan model klasifikasi terbaik yang dapat digunakan.

Tabel 2. Skenario Eksperimen Klasifikasi Aspek dan Sentimen

<i>Eksperimen</i>	<i>Feature Extraction</i>	<i>Sampling</i>	<i>Model</i>
1			RF
2		Original	DT
3			XGBoost
4			RF
5	TF - IDF	ROS	DT
6			XGBoost
7			RF
8		SMOTE	DT
9			XGBoost

Pemodelan Topik

Tahapan *topic modeling* ini dilakukan dengan menggunakan *preprocessed data* untuk kemudian didapatkan topik-topik utama yang ada pada setiap ulasan yang terkumpul pada aspek *usability* PACMAD beserta sentimennya. Kelompok-kelompok ulasan didapatkan berdasarkan label aspek *usability* PACMAD dan label sentimen hasil anotasi. Tahapan ini akan menghasilkan kata-kata kunci untuk topik-topik ulasan yang paling dibahas pada setiap aspek *usability* PACMAD beserta sentimennya.

Rekomendasi

Pada tahapan ini, peneliti akan memberikan rekomendasi terhadap penggunaan model *machine learning* yang baik digunakan untuk melakukan klasifikasi terhadap aspek *usability* PACMAD serta sentimennya. Rekomendasi akan diberikan berdasarkan hasil evaluasi dari masing-masing skenario yang diterapkan. Selain itu juga akan diberikan rekomendasi terkait dengan hal-hal yang menjadi perhatian khusus untuk pengembang BRImo berdasarkan topik-topik utama yang dihasilkan pada tahapan pemodelan topik.

D. Hasil dan Pembahasan

Hasil Klasifikasi Aspek

Eksperimen klasifikasi aspek *usability* PACMAD untuk ulasan pengguna dilakukan dalam 9 skenario berdasarkan tiga metode *sampling* dan tiga model klasifikasi dan kemudian dibandingkan dengan menggunakan *F1-score* yang ditunjukkan pada Tabel 3. Pada hasil tersebut dapat terlihat bahwa dengan berbagai macam metode *sampling*, model XGBoost menunjukkan performa terbaik dalam melakukan klasifikasi aspek *usability* PACMAD.

Hasil terbaik ditunjukkan oleh model klasifikasi XGBoost dengan bantuan metode *sampling* SMOTE. Selain itu dari hasil klasifikasi, penambahan *treatment sampling* dengan metode ROS memiliki performa yang tidak lebih baik dibandingkan klasifikasi tanpa adanya *treatment* dengan metode ROS. Namun pada skenario tersebut model dapat melakukan prediksi dengan baik untuk kelas *Effectiveness* dibandingkan dengan model dari skenario lain. Adapun hasil prediksi

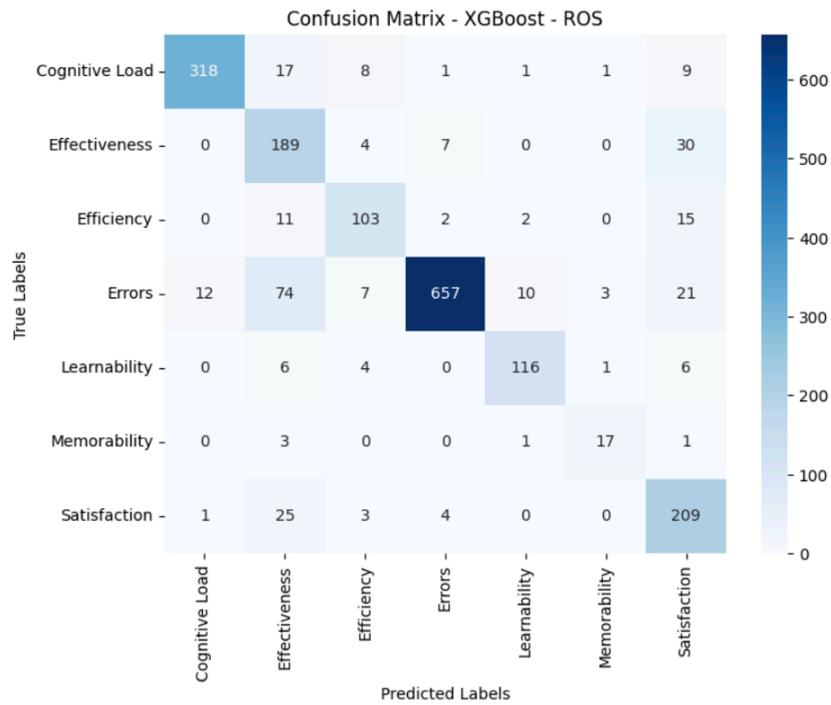
beserta nilai sebenarnya dari hasil klasifikasi dapat dilihat pada Gambar 3 dan Gambar 4.

Tabel 3. Hasil F1-Score Klasifikasi Aspek

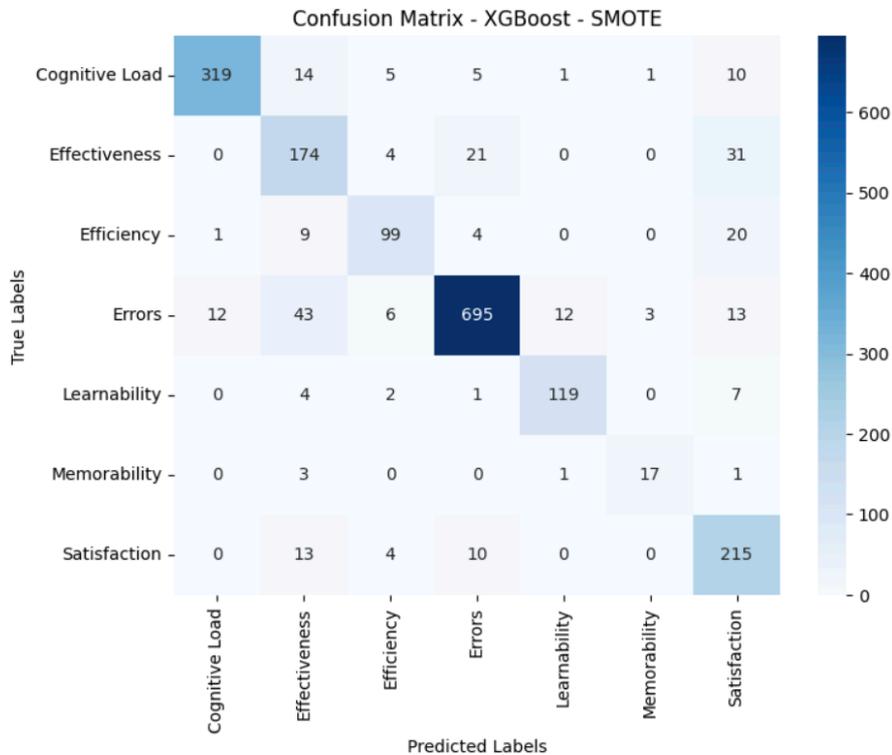
Model Klasifikasi	Metode <i>Sampling</i>		
	Original	ROS	SMOTE
RF	73.64%	80.44%	79.55%
DT	80.14%	77.12%	79.69%
XGBoost	86.06%	85.44%	86.55%

Berdasarkan pada Gambar 4 *confusion matrix* prediksi aspek *usability* PACMAD dengan menggunakan metode klasifikasi XGBoost dengan teknik *sampling* SMOTE, kesalahan prediksi banyak terjadi pada prediksi aspek *Effectiveness*, sebanyak 43 ulasan pada data testing salah diprediksi sebagai aspek *Effectiveness* sedangkan secara actual merupakan aspek *Errors*. Hal tersebut juga terjadi pada prediksi aspek *usability* PACMAD dengan menggunakan metode klasifikasi XGBoost dengan teknik *sampling* ROS seperti terlihat pada Gambar 3, dimana 74 ulasan pada data testing salah diprediksi sebagai aspek *Effectiveness* sedangkan secara actual merupakan aspek *Errors*.

Sebaliknya pada prediksi aspek *usability* PACMAD dengan menggunakan metode klasifikasi XGBoost pada ulasan tanpa *treatment sampling*, kesalahan prediksi justru terjadi pada aspek *Errors*, dimana 51 ulasan pada data testing salah diprediksi sebagai aspek *Errors* sedangkan secara actual merupakan aspek *Effectiveness*. Hal ini kemungkinan besar terjadi karena aspek *Errors* dan *Effectiveness* memiliki irisan secara definisi, aspek *Errors* merupakan aspek yang berkaitan dengan ulasan pengguna terkait dengan gangguan secara sistem yang mengakibatkan tidak dapat digunakannya aplikasi, sedangkan aspek *Effectiveness* merupakan aspek yang berkaitan dengan kemampuan suatu aplikasi untuk dapat menyelesaikan suatu tugas tanpa adanya masalah yang mengakibatkan tidak dapat diselesaikannya suatu tugas. Hal yang dapat menyebabkan terjadinya langkah yang tidak efektif merupakan suatu eror, sehingga aspek *Errors* dan *Effectiveness* memiliki irisan.



Gambar 3. Confusion Matrix Model XGBoost ROS



Gambar 4. Confusion Matrix Model XGBoost SMOTE

Hasil Klasifikasi Sentimen

Eksperimen klasifikasi sentimen untuk ulasan pengguna dilakukan dalam sembilan skenario berdasarkan tiga metode sampling dan tiga model klasifikasi dan kemudian dibandingkan dengan menggunakan *F1-score* yang ditunjukkan pada Tabel 4. Pada hasil tersebut dapat terlihat bahwa dengan berbagai macam metode sampling, model XGBoost menunjukkan performa terbaik dalam melakukan klasifikasi sentimen.

Hasil terbaik ditunjukkan oleh model klasifikasi XGBoost dengan bantuan metode *sampling* SMOTE. Selain itu dari hasil klasifikasi, penambahan *treatment sampling* dengan metode ROS memiliki performa yang lebih baik dibandingkan klasifikasi tanpa adanya *treatment* dengan metode ROS. Berdasarkan hasil tersebut dapat terlihat bahwa penambahan *treatment sampling* memiliki dampak positif dalam melakukan klasifikasi sentimen.

Tabel 4. Hasil F1-Score Klasifikasi Sentimen

Model Klasifikasi	Metode Sampling		
	Original	ROS	SMOTE
RF	88.88%	88.88%	89.01%
DT	83.27%	83.27%	82.26%
XGBoost	88.43%	88.43%	89.59%

Hasil klasifikasi sentimen menunjukkan bahwa seringkali terjadi kesalahan berupa *False Negative*. Pada Gambar 5.12 dapat terlihat bahwa ulasan yang salah diprediksi sebagai negatif memiliki kata-kata seperti “repot” “bosen” “keluar” “error”, selain itu kata “login” “fitur” “saldo” muncul pada ulasan yang salah diprediksi. Hal ini dapat menjadi indikasi bahwa kata-kata “login” “fitur” “saldo” seringkali muncul pada ulasan yang memiliki sentimen negatif secara aktual.

review	label	label_pred
pakai brimo hidup lebih sat set gak repot atm dulu	POSITIVE	NEGATIVE
stable ca n login device	POSITIVE	NEGATIVE
like app stable never get error long use	POSITIVE	NEGATIVE
bri mo turut cukup mungkin tinggal tampil aplikasi aja baru engga bosen	POSITIVE	NEGATIVE
mantapppp aplikasi nya saing dng bank swasta	POSITIVE	NEGATIVE
bagus kalau mohon tambah fitur sm saldo masuk terima kasih	POSITIVE	NEGATIVE
terima kasih solusi sekarang aplikasi tdk keluar sendiri	POSITIVE	NEGATIVE

Gambar 5. Contoh Hasil Prediksi Sentimen *False Negative*

Hasil Pemodelan Topik

Berdasarkan hasil *topic modeling* dengan LDA untuk ulasan pengguna dengan aspek *usability Cognitive Load* didapatkan topik utama ulasan tersebut. Secara umum pada sentimen negatif berbicara tentang sulitnya mendaftar di aplikasi BRImo, pada bagian verifikasi foto ktp. Pada sentimen positif berbicara tentang

banyaknya fitur yang dapat digunakan untuk bertransaksi serta kemudahan dalam akses fitur-fitur yang ada. Berdasarkan hasil tersebut, maka diperlukan adanya perbaikan pada verifikasi foto ktp pada saat pengguna baru BRImo melakukan pendaftaran. Jika hal tersebut tidak diperbaiki, maka BRImo akan memiliki potential loss customer akibat tidak bisanya pengguna melakukan registrasi. Selain itu kemudahan pengguna untuk melakukan transaksi juga perlu dipertahankan.

Pada aspek *Effectiveness*, secara umum ulasan negatif dari aspek ini berbicara tentang tidak berhasilnya transaksi akibat saldo yang tidak mencukupi. Untuk sentimen positif secara umum menunjukkan aktifitas pembayaran dengan metode qris. Kegagalan pengguna dalam melakukan pembayaran melalui qris perlu menjadi perhatian bagi pengembang BRImo, karena kegagalan untuk melakukan kegiatan ini akan berdampak pada ketidaknyamanan dan sulit dipercayanya BRImo oleh nasabah dan penggunanya untuk melakukan transaksi. Di lain sisi, pembayaran kartu kredit dirasa cukup efektif untuk digunakan. Terdapat juga ulasan positif untuk efektifitas penggunaan layanan qris dalam melakukan pembayaran, hal ini perlu diteliti lebih lanjut terkait dengan versi dan potensi tidak berjalan baiknya layanan qris, sehingga setiap adanya update pada BRImo, layanan qris dapat tetap digunakan dengan baik.

Topik utama pada ulasan dengan label negatif pada aspek *efficiency* adalah kendala koneksi untuk melakukan transfer melalui BI FAST. Untuk sentimen positif pada aspek ini adalah fitur top rekening untuk mempercepat transfer ke rekening yang sering. Sedangkan Topik utama pada aspek *error* negatif adalah kegagalan saat melakukan pendaftaran dan verifikasi, selain itu juga banyak error yang terjadi pada versi android tertentu. Sedangkan pada aspek error negatif disampaikan bahwa login yang sebelumnya *error* sudah diperbaiki. Berdasarkan topik- topik ulasan tersebut, maka pengembang BRImo perlu aware terhadap versi android mendatang sehingga tidak berdampak secara negatif atau menghasilkan *error* pada aplikasi BRImo. Setiap update pada versi android perlu diperhatikan, mengingat isu ini merupakan topik yang paling sering muncul pada ulasan pengguna di aspek *error*.

Berdasarkan hasil *topic modeling* dengan LDA untuk ulasan pengguna dengan aspek *usability learnability*, topik utama pada aspek ini adalah kurang mudahnya dalam melakukan penarikan laporan mutasi dan mengganti *password*. Sedangkan sisi positif dari ulasan di aspek ini adalah kemudahannya untuk menggunakan aplikasi bagi pengguna baru karena UI yang simpel dan menampilkan fitur-fitur yang berguna. Dapat disimpulkan bahwa diperlukan adanya perbaikan pada fitur cetak mutasi agar lebih memudahkan pengguna, terutama pada bagian cetak mutasi secara 1 bulan penuh. Selain itu kemudahan pengguna dapat menggunakan aplikasi ini secara umum perlu dipertahankan. Terlebih karena aspek ini merupakan aspek yang berkaitan dengan kemudahan pengguna dalam menggunakan fitur ataupun aplikasi ini secara pertama kali. Sehingga topik-topik utama pembahasan, terlebih pada sentimen negatif perlu menjadi perhatian khusus sehingga tidak terjadinya churn akibat kebingungan pengguna dalam menggunakan aplikasi BRImo ini.

Topik utama pada aspek *memorability* adalah mengenai lupa username dan password hingga terblokir dari aplikasi. Pada sentimen positif tidak dapat ditentukan topik utama akibat kata-kata yang bersifat umum. Aspek *memorability* berfokus pada kemudahan pengguna untuk menggunakan aplikasi setelah waktu yang cukup lama. Kebingungan pengguna dapat disebabkan akibat adanya

perubahan interface pada update BRImo. Salah satu hal yang dapat dilakukan oleh pengembang adalah dengan memberikan bantuan pada pengguna yang sudah lama tidak menggunakan aplikasi BRImo ini.

Topik utama pada aspek *satisfaction* adalah aplikasi BRImo yang membantu nasabah dan meningkatkan kepuasan nasabah. Secara umum ulasan-ulasan yang ada pada aspek *satisfaction* ini berisikan ulasan yang menggambarkan kepuasan pengguna aplikasi BRImo secara umum, namun masih terdapat ulasan yang bersifat negatif secara umum yang berkaitan dengan diperlukan adanya fitur yang lebih baik serta harapan bahwa aplikasi BRImo dapat berfungsi dengan lebih baik lagi.

E. Simpulan

Penelitian ini bertujuan untuk menentukan model terbaik dalam melakukan analisis sentimen berbasis aspek *usability* PACMAD. Terdapat tiga model klasifikasi yang digunakan beserta tiga perlakuan sampling yang diterapkan dalam melakukan klasifikasi aspek *usability* PACMAD dan klasifikasi sentimen. Model klasifikasi yang digunakan adalah Random Forest (RF), Decision Tree (DT), dan Extreme Gradient Boosting (XGBoost) yang diterapkan pada data ulasan tanpa *treatment sampling*, dengan Random Oversampling (ROS), dan Synthetic Minority Over-sampling Technique (SMOTE). Berdasarkan hasil penelitian, maka dapat ditentukan bahwa model klasifikasi XGBoost yang didukung oleh metode *sampling* SMOTE merupakan model terbaik dalam melakukan klasifikasi aspek *usability* PACMAD dengan nilai F1-score sebesar 86,55% dan model terbaik dalam melakukan klasifikasi sentimen ulasan pengguna dengan nilai F1-score sebesar 89,59%. Hasil dari penelitian ini menunjukkan bahwa pada skenario terbaik pun masih terdapat kesempatan terjadinya error pada hasil klasifikasi sentimen ataupun aspek *usability* PACMAD. Adanya kemungkinan terjadinya kesalahan klasifikasi dapat terjadi pada aspek *Effectiveness* dan *Satisfaction*. Selain itu juga model klasifikasi sentimen cenderung untuk melakukan klasifikasi pada sentimen negatif ketimbang positif. Adanya kesalahan pada klasifikasi sentimen tersebut dapat ditolerir dikarenakan sentimen negatif akan menjadi perhatian lebih, sehingga adanya hasil klasifikasi bersifat *false negative* akan dimasukkan dalam penelitian lebih lanjut. Dapat disimpulkan bahwa hasil klasifikasi dapat digunakan untuk membantu pengembang dalam mengerucutkan kategori pengembangan berdasarkan aspek *usability* PACMAD dan sentimennya.

Selain itu, penelitian ini juga bertujuan untuk menemukan topik-topik utama yang dibahas dalam ulasan pengguna berdasarkan aspek-aspek *usability* PACMAD yang telah dikelompokkan. Dalam aspek *cognitive load*, kesulitan dalam melakukan pendaftaran serta kemudahan dalam melakukan transaksi merupakan topik utama pembahasan. Untuk aspek *effectiveness*, kegagalan dalam melakukan transaksi transfer akibat saldo yang tidak cukup menjadi hal yang paling sering dibicarakan. Dalam aspek *efficiency*, kesulitan untuk menyelesaikan layanan transfer BI FAST dikarenakan terlalu lama dalam melakukan transfer menjadi topik bahasan, beserta kemudahan untuk akses fitur-fitur yang ada di halaman utama aplikasi yang mempercepat langkah menuju fitur terkait menjadi topik utama. *Error* banyak menjadi hal yang dibahas di ulasan pengguna, Adapun *error* yang berkaitan dengan kegagalan dalam melakukan aktivasi akun BRImo beserta validasi foto KTP menjadi kendala yang paling sering dibicarakan, selain itu aplikasi yang sering keluar tiba-

tiba juga menjadi isu yang sering dibicarakan pada aspek error di ulasan pengguna BRImo. Pada aspek *memorability*, banyaknya kendala pengguna pada saat memasukkan informasi untuk login seperti *username* dan *password* membuat seringkali akses ke aplikasi terkendala. Pada bagian *learnability*, kebingungan nasabah untuk merubah *username* dan *passwords* beserta pujian bagi aplikasi terkait dengan kemudahan dalam melakukan akses ke layanan finansial bagi pengguna baru menjadi topik yang sering dibicarakan berdasarkan ulasan pengguna. Pada aspek terakhir, *satisfaction*, banyak terdapat pujian dari nasabah yang merasa terbantu dan puas dengan layanan yang ada. Secara umum hasil pemodelan topik pada aspek *effectiveness*, *error*, *cognitive load*, serta *learnability* memiliki *insight* yang dapat secara konkrit diterapkan oleh pengembang BRImo.

F. Referensi

- [1] BRI, "Memberi Makna Indonesia." Accessed: Mar. 21, 2024. [Online]. Available: <https://e-recruitment.bri.co.id/#:~:text=Di%20BRI%2C%20kami%20menamakan%20diri,BRILiaN%20BRI%20dan%20membangun%20bangsa!>
- [2] BRI, "Proposisi Investasi," 2023, Accessed: May 20, 2024. [Online]. Available: https://www.ir-bri.com/investment_securing.html
- [3] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *J Interact Sci*, vol. 1, no. 1, p. 1, 2013, doi: 10.1186/2194-0827-1-1.
- [4] N. Bevan, "Measuring usability as quality of use," *Software Quality Journal*, vol. 4, no. 2, pp. 115–130, 1995, doi: 10.1007/BF00402715.
- [5] I. Atoum and A. Ootom, "Mining Software Quality from Software Reviews: Research Trends and Open Issues," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 31, pp. 74–83, Jan. 2016, doi: 10.14445/22312803/IJCTT-V31P114.
- [6] M. Umair and E. Susanto, "Analisis Sentimen Ulasan Pengguna Pada Aplikasi BRImo BRI Menggunakan Metode Klasifikasi Algoritma Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, p. 1149, Apr. 2024, doi: 10.30865/mib.v8i2.7381.
- [7] M. Insan, U. Hayati, and O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, pp. 478–483, Mar. 2023, doi: 10.36040/jati.v7i1.6373.
- [8] Jakob Nielsen and Raluca Budiu, *Mobile usability*. New Riders, 2013.
- [9] D. Zhang and B. Adipat, "Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications," *Int J Hum Comput Interact*, vol. 18, no. 3, pp. 293–308, Jul. 2005, doi: 10.1207/s15327590ijhc1803_3.
- [10] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.
- [11] H. Samy, A. M. Helmy, and N. Ramadan, "Aspect-based Sentiment Analysis of Mobile Apps Reviews using Class Association Rules and LDA," in *Proceedings - 2021 IEEE 10th International Conference on Intelligent Computing and Information Systems, ICICIS 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 183–189. doi: 10.1109/ICICIS52592.2021.9694242.

- [12] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [13] E. Uzun, "A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages," *IEEE Access*, vol. 8, pp. 61726–61740, Dec. 2020, doi: 10.1109/ACCESS.2020.2984503.
- [14] B. Dumbache, L. Kaili Diamond, and L. aKailiDiamond, "SABLE: Tools for Web Crawling, Web Scraping, and Text Classification," 2012.
- [15] M. Ghasemaghaei, S. P. Eslami, K. Deal, and K. Hassanein, "Reviews' length and sentiment as correlates of online reviews' ratings," *Internet Research*, vol. 28, no. 3, pp. 544–563, 2018, doi: 10.1108/IntR-12-2016-0394.
- [16] C. C. Aggarwal and C. Zhai, "Mining Text Data," 2012, doi: 10.1007/978-1-4614.
- [17] Matthew Denny and Arthur Spirling, "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It," *SSRN Electronic Journal*, pp. 1–49, 2017.
- [18] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," *IEEE Access*, vol. 5, pp. 2870–2879, Dec. 2017, doi: 10.1109/ACCESS.2017.2672677.