

Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data

Riska Aryanti¹, Titik Misriati^{2,*}, Asriyani Sagiyanto³

¹Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

²Sistem Informasi Akuntansi Kampus Kabupaten Karawang, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

³Hubungan Masyarakat, Fakultas Komunikasi dan Bahasa, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ¹riska.rts@bsi.ac.id, ^{2,*}titik.tmi@bsi.ac.id, ³asriyani.ayy@bsi.ac.id

Email Penulis Korespondensi: titik.tmi@bsi.ac.id

Submitted: 10/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak—Aplikasi Primaku adalah aplikasi yang dapat dimanfaatkan sebagai alat bantu dalam memantau pertumbuhan anak balita, aplikasi ini dapat digunakan untuk mengumpulkan data pertumbuhan anak balita, selain itu aplikasi ini juga dapat memberikan informasi dan visualisasi yang jelas tentang pertumbuhan anak balita, termasuk status gizi dan perkembangan pertumbuhan sesuai dengan standar yang telah ditetapkan dengan demikian aplikasi primaku dapat membantu orang tua atau petugas kesehatan dalam memantau pertumbuhan anak balita secara rutin dan mendeteksi dini potensi risiko stunting, stunting merupakan kondisi gangguan pertumbuhan yang terjadi pada anak balita akibat kekurangan gizi yang ditandai dengan tinggi badan anak yang lebih pendek dari standar usianya. Stunting dapat memiliki dampak jangka panjang terhadap kualitas hidup anak, seperti gangguan perkembangan fisik, kognitif dan sosial, serta meningkatkan risiko penyakit kronis di masa dewasa. Aplikasi primaku sudah banyak digunakan sebanyak lebih dari 500.000 pengguna telah mendownload aplikasi ini dan 44.700 ulasan yang diberikan oleh pengguna terhadap aplikasi ini, namun, membaca semua ulasan mungkin memerlukan waktu, namun jika ulasan yang dibaca sedikit, maka hasil review akan bias. Oleh karena itu, analisis sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan ulasan pengguna menjadi ulasan positif dan negatif. Oleh karenanya, penelitian deteksi pertumbuhan balita untuk mengetahui respon masyarakat terhadap aplikasi Primaku dapat manfaat yang besar dalam upaya mencegah stunting pada anak balita di Indonesia. Dalam penelitian ini, algoritma random forest dengan teknik SMOTE digunakan untuk melakukan analisis sentimen ulasan aplikasi Primaku. Algoritma random forest adalah algoritma pembelajaran mesin yang berbasis pohon keputusan. Teknik SMOTE digunakan untuk mengatasi masalah ketidakseimbangan data dan mampu mengurangi overfitting sekaligus meningkatkan kinerja algoritma Random Forest. Data yang digunakan dalam penelitian ini adalah data ulasan aplikasi primaku yang diperoleh dari hasil *scrapping* dari *Google play store*. Data tersebut berisi komentar pengguna aplikasi yaitu positif, dan negatif. Hasil dari analisis sentimen ini menampilkan pemahaman mendalam tentang persepsi pengguna terhadap aplikasi primaku. Analisis sentimen ini dapat menjadi landasan untuk perbaikan dan pengembangan lebih lanjut pada aplikasi primaku, dengan fokus pada aspek-aspek yang mempengaruhi kepuasan pengguna dan hasil penelitian menunjukkan bahwa algoritma random forest dengan teknik SMOTE dapat menghasilkan akurasi yang cukup baik dalam analisis sentimen aplikasi primaku akurasi yang diperoleh pada penelitian ini adalah 88%.

Kata Kunci: Text Mining; Analisis Sentimen; Stunting; *Google Play Store*; Primaku

Abstract— The Primaku application is an application that can be used as a tool to monitor the growth of children under five, this application can be used to collect data on the growth of children under five, apart from that this application can also provide clear information and visualization about the growth of children under five, including nutritional status and growth development In accordance with the standards that have been set, the Primaku application can help parents or health workers in routinely monitoring the growth of children under five and early detecting the potential risk of stunting. Stunting is a growth disorder that occurs in children under five due to malnutrition which is characterized by the child's height, which is shorter than the age standard. Stunting can have a long-term impact on a child's quality of life, such as disrupting physical, cognitive and social development, as well as increasing the risk of chronic disease in adulthood. The primaku application has been widely used, more than 500,000 users have downloaded this application and 44,700 reviews have been given by users to this application, however, reading all the reviews may take time, but if there are few reviews read, then the review results will be biased. Therefore, sentiment analysis aims to overcome this problem by automatically grouping user reviews into positive and negative reviews. Therefore, research on toddler growth detection to determine the public's response to the Primaku application can be of great benefit in efforts to prevent stunting in children under five in Indonesia. In this research, the random forest algorithm with the SMOTE technique was used to carry out sentiment analysis of Primaku application reviews. The random forest algorithm is a machine learning algorithm based on decision trees. The SMOTE technique is used to overcome data imbalance problems and is able to reduce overfitting while increasing the performance of the Random Forest algorithm. The data used in this research is Primaku application review data obtained from scrapping results from the Google Play Store. This data contains comments from application users, namely positive and negative. The results of this sentiment analysis show a deep understanding of user perceptions of the Primaku application. This sentiment analysis can be a basis for further improvement and development of the Primaku application, with a focus on aspects that influence user satisfaction and the research results show that the random forest algorithm with the SMOTE technique can produce quite good accuracy in sentiment analysis of the Primaku application. obtained in this study was 88%.

Keywords: Text Mining; Sentiment Analysis; Stunting; Google Play Store; Primaku

1. PENDAHULUAN

Dalam era digital seperti saat ini, teknologi dapat dimanfaatkan sebagai alat bantu dalam memantau pertumbuhan anak balita. Salah satu aplikasi yang dapat digunakan adalah primaku, sebuah aplikasi yang dikembangkan oleh Kementerian Kesehatan Indonesia untuk memantau pertumbuhan anak balita secara mandiri oleh orang tua atau wali anak. Aplikasi ini dilengkapi dengan fitur yang memudahkan pengguna dalam memasukkan data pertumbuhan anak, serta memberikan informasi dan saran gizi yang tepat sesuai dengan usia dan kondisi anak. Aplikasi Primaku adalah sebuah aplikasi mobile yang dirancang untuk membantu orang tua dalam memantau pertumbuhan dan perkembangan anak balita secara teratur dan akurat. Aplikasi ini dikembangkan oleh Kementerian Kesehatan Republik Indonesia dan dapat diunduh secara gratis melalui *Google play store* atau *App Store*. Aplikasi Primaku memungkinkan orang tua untuk mencatat berbagai informasi terkait pertumbuhan anak, seperti berat badan, tinggi badan, dan lingkaran kepala. Selain itu, aplikasi ini juga dilengkapi dengan grafik pertumbuhan dan rekomendasi pemberian makanan yang sesuai dengan usia dan kebutuhan anak balita[1]. Kelebihan aplikasi Primaku adalah dapat membantu orang tua[1] untuk memantau pertumbuhan anak secara berkala dan teratur, sehingga dapat lebih cepat mendeteksi potensi masalah pertumbuhan seperti stunting dan segera mengambil tindakan preventif. Selain itu, aplikasi Primaku juga memberikan edukasi tentang nutrisi dan kesehatan anak yang dapat membantu orang tua dalam merawat anak dengan lebih baik sehingga dapat mencegah stunting melalui deteksi dini menggunakan aplikasi Primaku, kualitas hidup anak balita dapat dijaga dan ditingkatkan. Aplikasi primaku sudah diunduh lebih dari 500.000 pengguna dengan rating 4.8 dan 44.700 ulasan pengguna pada aplikasi ini yang berisi ulasan positif dan negatif seperti keluhan, kritik, dan saran terhadap aplikasi ini yang bisa dijadikan data untuk menganalisis tren dan topik tertentu. Upaya untuk menganalisis data ini disebut analisis sentimen atau penambangan opini [2].

Analisis sentimen mencakup bidang penambangan teks, pemrograman bahasa alami, dan kecerdasan buatan dan dilakukan untuk secara otomatis mengekstraksi, memahami, dan memproses data teks untuk memperoleh informasi berguna dan pengetahuan baru[3], [4]. Proses analisis sentimen menentukan bagaimana review masyarakat terhadap suatu ulasan terhadap aplikasi tersebut cenderung dengan membagi sentimen menjadi dua kelas atau lebih[5].

Penelitian tentang deteksi pertumbuhan balita untuk mengetahui respon masyarakat terhadap aplikasi Primaku sangat penting[6] untuk meningkatkan kesadaran masyarakat tentang pentingnya deteksi dini dan pencegahan stunting pada anak balita. Menurut studi WHO pada tahun 2019, kurangnya kesadaran dan pengetahuan masyarakat tentang pentingnya pencegahan stunting menjadi salah satu faktor utama terjadinya stunting pada anak balita. Wulan Alinda Wahyuni dalam penelitiannya melakukan analisis sentimen ulasan aplikasi investasi online pada google playstore menggunakan algoritma random forest hasil implementasi algoritma random forest pada aplikasi Stockbit pada penelitian ini menghasilkan nilai akurasi sebesar 62,50%, untuk aplikasi seed pada penelitian ini menghasilkan nilai akurasi sebesar 63,39%. dan untuk aplikasi HSB pada penelitian ini menghasilkan nilai akurasi sebesar 96,25%[7]. Penelitian lainnya juga telah dilakukan oleh Fanka, Dian dan Buce dalam penelitiannya menggunakan data ulasan dari aplikasi dana pada google playstore sebanyak 1354 ulasan, Metode yang digunakan Random Forest dan TF-IDF, tahapan preprocessing yang dilakukan Case folding, Tokenizing, Stopword removal dan Stemming, hasil tingkat akurasi mencapai 84%[8]. Penelitian Muhammad, Azhari dan Ultach melakukan analisis sentimen ulasan aplikasi MOLA, data yang digunakan merupakan ulasan aplikasi MOLA sebanyak 520 data yang terdiri dari 312 ulasan positif dan 208 ulasan negatif tahapan preprocessing pada penelitian ini adalah cleaning, Case folding, Tokenizing, Normalization, Stopword removal dan Stemming sedangkan algoritma yang digunakan menggunakan algoritma SVM[9].

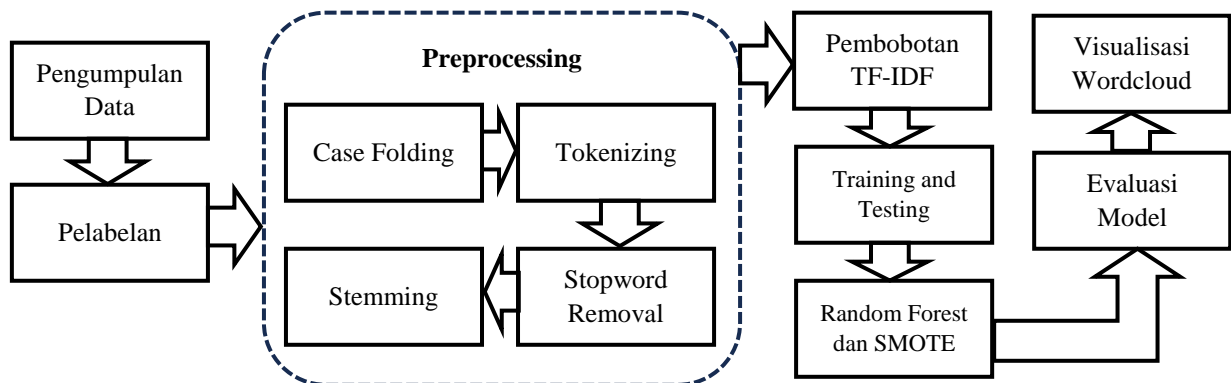
Penelitian Tinaliah dan Triana melakukan analisis sentimen ulasan primaku menggunakan algoritma SVM, data yang digunakan adalah hasil scrapping google playstore hanya 1000 ulasan, namun dalam penelitiannya data yang digunakan tidak seimbang yakni 787 ulasan positif dan 213 ulasan negatif[8]. Hal ini dapat menyebabkan model klasifikasi tidak akurat dan cenderung memprediksi kelas mayoritas[10]. Untuk mengatasi masalah ketidakseimbangan data, Penggunaan teknik oversampling dapat menyebabkan overfitting, metode SMOTE (Synthetic Minority Oversampling Technique) diusulkan untuk menangani overfitting[11]. Selain itu penggunaan teknik SMOTE juga dapat meningkatkan kinerja algoritma yang dibangun[12], [13][14], [15].

Berdasarkan hal tersebut, dalam penelitian ini akan dilakukan analisis sentimen publik untuk mengukur persepsi dan respon masyarakat terhadap aplikasi Primaku menggunakan algoritma Random Forest dan teknik SMOTE. Teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data dan mengurangi overfitting data yang disebabkan cenderung memprediksi kelas mayoritas selain itu penggunaan teknik SMOTE dalam penelitian ini juga dapat meningkatkan kinerja model Random Forest, sehingga dengan mengetahui respon masyarakat terhadap aplikasi Primaku, penelitian ini dapat membantu dalam meningkatkan kualitas dan efektivitas aplikasi, serta menentukan strategi pemasaran yang tepat agar aplikasi dapat diterima dan digunakan secara luas oleh masyarakat. Selain itu, hasil penelitian ini dapat membantu dalam mengidentifikasi faktor-faktor yang mempengaruhi penerimaan dan penggunaan aplikasi Primaku, sehingga dapat dikembangkan solusi untuk mengatasi hambatan dan meningkatkan penggunaan aplikasi secara efektif. Hal ini dapat membantu

meningkatkan efektivitas deteksi dini dan pencegahan stunting pada anak balita, serta memberikan kontribusi positif dalam upaya mencegah stunting di Indonesia.

2. METODOLOGI PENELITIAN

Pada penelitian ini, dataset yang digunakan berupa data ulasan pengguna Aplikasi Primaku yang berasal dari Google Play. Pengambilan data menggunakan *Python* yang diperoleh dari *scraping data google play store* aplikasi Primaku. Tahap berikutnya yaitu dilakukan pelabelan berdasarkan skor, *preprocessing* (*case folding*, *tokenizing*, *stop word removal* dan *stemming*), pembobotan TF-IDF, klasifikasi Random Forest dengan SMOTE, dan evaluasi model, seperti pada gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pengumpulan data ulasan aplikasi Primaku di halaman Google Play Store dengan *web scrapping* di *Python* menghasilkan 2252 ulasan.

2.2 Pelabelan

Pelabelan ulasan aplikasi primaku berdasarkan rating atau skor untuk menentukan sentimen yang diberikan oleh pengguna. Rating atau skor ulasan yang lebih tinggi menunjukkan sentimen positif, sedangkan rating atau skor ulasan yang lebih rendah menunjukkan sentimen negatif.

2.3 Preprocessing

Preprocessing merupakan langkah penting untuk menyiapkan data sebelum digunakan dalam analisis lebih lanjut. Tujuan dari *preprocessing* adalah untuk membersihkan data dari *noise* dan kesalahan, serta mengubah data menjadi bentuk yang terstruktur dan siap untuk diproses. *Preprocessing* dilakukan dengan *case folding*, *tokenizing*,

1. Case Folding

Case folding adalah proses mengubah semua huruf dalam suatu teks menjadi huruf kecil untuk meningkatkan efisiensi dalam pengolahan teks. *Case folding* dilakukan untuk menyeragamkan huruf dalam suatu teks, sehingga dapat mempermudah proses analisis teks.

2. Tokenizing

Tokenizing merupakan proses untuk memecah teks menjadi unit-unit yang lebih kecil, seperti memecah kalimat menjadi kata.

3. Stopword Removal

Stopword Removal adalah proses untuk menghilangkan kata-kata yang tidak penting dari ulasan aplikasi Primaku seperti kata penghubung, kata depan, dan kata bantu.

4. Stemming

Stemming adalah proses untuk mengurangi kata-kata ke bentuk dasarnya. *Stemming* dilakukan untuk mempermudah proses analisis teks, seperti untuk analisis sentimen, klasifikasi teks, dan pencarian informasi.

2.4 Pembobotan TF-IDF

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode pembobotan kata yang digunakan untuk mengukur relevansi suatu kata dalam suatu dokumen dimana kata-kata yang sering muncul dalam suatu dokumen akan diberi bobot yang lebih tinggi daripada kata-kata yang jarang muncul. Selain itu, kata-kata yang hanya muncul dalam beberapa dokumen akan diberi bobot yang lebih tinggi daripada kata-kata yang muncul dalam banyak dokumen. Pembobotan TF-IDF terdiri dari dua komponen, yaitu frekuensi kemunculan suatu kata dalam suatu dokumen dan kebalikan dari frekuensi kemunculan suatu kata dalam keseluruhan dokumen[16], [17].

Pembobotan TF-IDF dapat dihitung dengan menggunakan rumus berikut:

$$TF - IDF = TF * IDF \tag{1}$$

Term Frequency (TF) dihitung dengan menggunakan rumus berikut:

$$TF = \frac{\text{Jumlah kemunculan kata}}{\text{Jumlah total kata dalam dokumen}} \tag{2}$$

Inverse Document Frequency (IDF) dihitung dengan menggunakan rumus berikut:

$$IDF = \text{Log} \frac{\text{Jumlah dokumen}}{\text{Jumlah dokumen yang mengandung kata}} \tag{3}$$

Dengan menggunakan pembobotan TF-IDF, kata-kata yang sering muncul dalam suatu dokumen akan diberi bobot yang lebih tinggi daripada kata-kata yang jarang muncul. Selain itu, kata-kata yang hanya muncul dalam beberapa dokumen akan diberi bobot yang lebih tinggi daripada kata-kata yang muncul dalam banyak dokumen.

2.5 Split Data

Pembagian data dalam penelitian ini terdiri dari 80% data latih dan 20% data uji.

2.6 Random Forest dan SMOTE

Random forest adalah metode klasifikasi yang membandingkan sekelompok pohon keputusan dengan data pelatihan hingga fitur acak independen dengan berbagai karakteristik berbeda. Random Forest adalah algoritma yang dapat mengklasifikasikan data dalam jumlah besar secara akurat [18], [19]. Hasil akhir diperoleh dengan menentukan simpul akar dan diakhiri dengan beberapa simpul daun. SMOTE (Synthetic Minority Over-sampling Technique) adalah sebuah teknik *oversampling* yang digunakan dalam pemrosesan data untuk menangani ketidakseimbangan kelas dalam masalah klasifikasi [20], [21].

2.7 Evaluasi Model

Evaluasi dilakukan berdasarkan akurasi, presisi (*precision*), recall dan F1 score. Akurasi mengukur sejauh mana model dapat mengklasifikasikan data dengan benar. Ini dihitung dengan membagi jumlah prediksi yang benar dengan jumlah total prediksi.

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Prediksi}} \tag{4}$$

Presisi mengukur sejauh mana kelas positif yang diprediksi oleh model adalah benar. Ini dinyatakan sebagai rasio antara true positive (TP) dan total prediksi positif (TP + false positive (FP)).

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{5}$$

Recall mengukur sejauh mana model dapat menemukan semua instans positif. Ini dinyatakan sebagai rasio antara true positive (TP) dan total instance positif (TP + false negative (FN)).

$$\text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

F1-score adalah rata-rata harmonik dari presisi dan recall. Ini memberikan keseimbangan antara kedua metrik ini dan berguna ketika distribusi kelas tidak seimbang.

$$F1 - \text{Score} = 2x \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \tag{7}$$

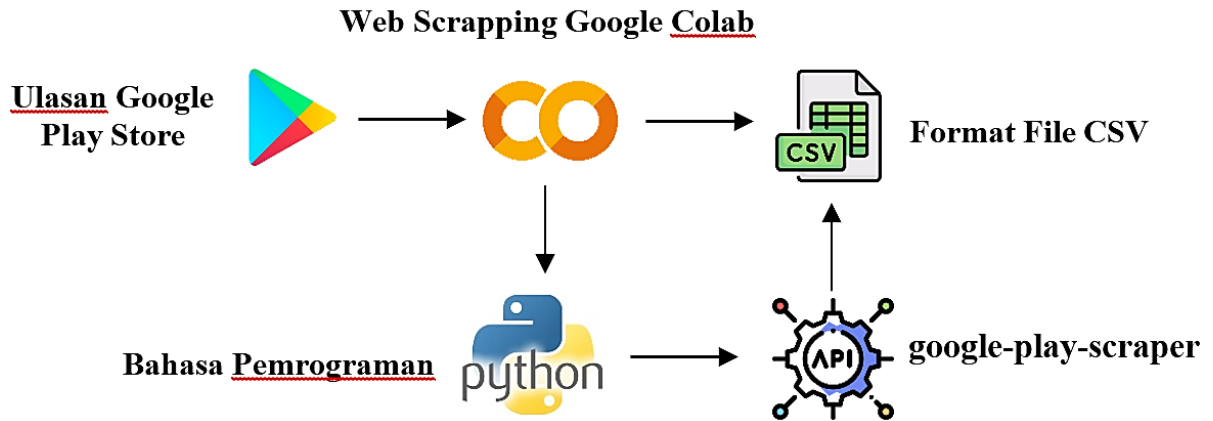
2.8 Visualisasi Wordcloud

Wordcloud adalah representasi visual dari teks di mana kata-kata yang paling sering muncul dalam teks diberikan ukuran yang lebih besar dan ditampilkan dalam tata letak yang acak atau artistik.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset dikumpulkan dari ulasan aplikasi primaku dari *google play store* dengan teknik web scrapping menggunakan google colab dengan bahasa pemrograman *python* dan menggunakan *library Google-Play-Scraper* yang tersedia pada *python* dalam pengambilan data [22], [23]. *Web scraping* adalah metode mengekstraksi informasi dari situs web [24], [25]. Metode ini digunakan untuk mengumpulkan data dari ulasan play store, dataset yang telah diambil dengan teknik web scrapping kemudian disimpan dan di ubah dengan format file comma-separated values (CSV), dapat dilihat pada gambar 2.



Gambar 2. Proses Pengumpulan Data

Dataset yang dikumpulkan dari hasil web scrapping pada gambar 1 tidak dapat langsung digunakan pada proses mining dengan algoritma machine learning. Data ulasan perlu diberi label dan pra-pemrosesan teks untuk menghilangkan data *noise*, *missing value*, atau data ulasan yang tidak memiliki informasi yang berguna, tidak memiliki makna, dan data yang sekiranya dapat mengganggu proses mining. Jumlah data ulasan yang terkumpul dari hasil *web scraping* sebanyak 2252 data, dapat dilihat pada gambar 3.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	d555b646-f24c-460d-b031-af7618562481	Enriko Siahaan	lh.googleusercontent.com/a-/ALV-U...	Aplikasi ini sangat hebat, sangat membantu unt...	5	0	3.33.1	2023-11-03 14:42:03	None	NaT	3.33.1
1	51ac76ad-4f24-48f2-9691-e70524a46ffe	Arina Susanti	lh.googleusercontent.com/a/ACg8oc...	bagus karena bisa untuk tracking BB dan TB bayi tiap bulan...	5	1	3.33.1	2023-11-01 11:54:46	None	NaT	3.33.1
2	8ce826d6-8196-4333-95d8-1abc7e207eb	Dewinta Sari	lh.googleusercontent.com/a/ACg8oc...	sangat membantu	5	0	3.33.1	2023-10-28 12:37:36	None	NaT	3.33.1
3	ef030d8e-a916-4e8a-a9e0-324d20034e08	Saka Agustiana	lh.googleusercontent.com/a/ACg8oc...	Apk untuk memantau tumbuh kembang anak 🍌🍌🍌	5	0	3.32.0	2023-10-27 18:21:41	None	NaT	3.32.0
4	3309e714-5743-49c6-a322-e3e6485d01c0	Ulfah Sandah	lh.googleusercontent.com/a-/ALV-U...	bagus	5	0	3.33.1	2023-10-25 23:27:50	None	NaT	3.33.1

Gambar 3. Ulasan Data Primaku

3.2 Pelabelan

Pada tahapan ini pelabelan data dilakukan berdasarkan skor [16], [17], [28], [29] dimana setiap skor 1, 2, 3, 4 dan 5 kemudian dibagi menjadi dua kelas yaitu positif dan negatif. Dapat dilihat pada gambar 4.

- Apabila skornya lebih besar dari 3 maka label dianggap positif.
- Sebaliknya jika skornya kurang dari 3 maka label tersebut dianggap negatif.

	userName	score	at	content	label
0	Enriko Siahaan	5	2023-11-03 14:42:03	Aplikasi ini sangat hebat, sangat membantu unt...	positif
1	Arina Susanti	5	2023-11-01 11:54:46	bagus karena bisa tracking BB dan TB bayi tiap bulan...	positif
2	Dewinta Sari	5	2023-10-28 12:37:36	sangat membantu	positif
3	Saka Agustiana	5	2023-10-27 18:21:41	Apk untuk membantu tumbuh kembang anak 🍌🍌🍌	positif
4	Ulfah Sandah	5	2023-10-25 23:27:50	bagus	positif
...
2247	Pengguna Google	5	2017-11-08 09:01:21	menarik banget aplikasinya. memudahkan penggun...	positif
2248	Pengguna Google	5	2017-11-07 14:33:42	Bermanfaat bgt, sebelumnya punya jg aplikasi P...	positif
2249	Pengguna Google	1	2017-09-23 20:21:40	tiba2 ke logout sendiri mau login lagi susah ba...	negatif
2250	Pengguna Google	5	2017-08-12 04:02:26	aplikasi yg 🍌🍌🍌🍌	positif
2251	Pengguna Google	5	2017-08-10 22:26:52	Aplikasi mudah digunakan	positif

2252 rows × 5 columns

Gambar 4. Proses Pelabelan

3.3 Preprocessing

Preprocessing adalah suatu proses dimana dataset yang di collect dari ulasan *google play store* akan dibersihkan dari unsur-unsur yang tidak diperlukan untuk selanjutnya memperoleh data berkualitas yang sesuai dengan kebutuhan.

a. *Case folding*

Pada proses *case folding* dilakukan tahapan untuk mengubah kalimat pada data menjadi huruf kecil atau lower case secara keseluruhan. Hasil *case folding* dapat dilihat pada tabel 1.

Tabel 1. Hasil *Case folding*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
Sangat berguna untuk memantau perkembangan anak □□	sangat berguna untuk memantau perkembangan anak □□
Primaku makin keren, ada fitur pemantau tumbuh kembang anak, konsultasi, forum. Makin membantu para ibu. ♥□	primaku makin keren ada fitur pemantau tumbuh kembang anak konsultasi forum. makin membantu para ibu ♥□
MantaP... Jadi tahu TUMBANG anak2 kita... Terimakasih PRIMAKU□□	mantap jadi tahu tumbang anak kita terimakasih primaku□□

b. *Tokenizing*

Pada proses berikutnya dilakukan tahap *tokenizing* untuk melakukan proses pengubahan kalimat pada data menjadi beberapa potong bentuk kata yang lebih kecil. Hasil *tokenizing* dapat dilihat pada tabel 2

Tabel 2. Hasil *Tokenizing*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
sangat berguna untuk memantau perkembangan anak □□	[sangat, berguna, untuk, memantau, perkembangan, anak]
primaku makin keren ada fitur pemantau tumbuh kembang anak konsultasi forum. makin membantu para ibu ♥□	[primaku, makin, keren, ada, fitur, pemantau, tumbuh, kembang, anak, konsultasi, forum, makin, membantu, para, ibu]
mantap jadi tahu tumbang anak kita terimakasih primaku□□	[mantap, jadi, tahu, tumbang, anak, kita, terimakasih, primaku]

c. *Stopword removal*

Pada proses berikutnya dilakukan tahap filter *stopword* untuk melakukan penghapusan beberapa kata umum yang tidak penting dan yang tidak ada relevansinya berdasarkan kamus *stopword*. Hasil *stopword* dapat dilihat pada tabel 3.

Tabel 3. Hasil *Stopword removal*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
[sangat, berguna, untuk, memantau, perkembangan, anak]	[berguna, memantau, perkembangan, anak]
[primaku, makin, keren, ada, fitur, pemantau, tumbuh, kembang, anak, konsultasi, forum, makin, membantu, para, ibu]	[primaku, keren, ada, fitur, pemantau, tumbuh, kembang anak, konsultasi, forum, membantu, ibu]
[mantap, jadi, tahu, tumbang, anak, kita, terimakasih, primaku]	[mantap, tumbang, anak, terimakasih, primaku]

d. *Stemming*

Pada proses berikutnya dilakukan tahap *stemming*. Proses *stemming* pada penelitian ini menggunakan library Sastrawi yang tersedia pada *Python*. Proses ini bertujuan untuk mereduksi kata menjadi bentuk dasarnya sesuai kaidah Kamus Besar Bahasa Indonesia (KBBI), misalnya kata “perkembangan”, “memantau” menjadi kata dasar “kembang”, “pantau” dan lainnya. Hasil *stemming* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Stemming*

Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
[berguna, memantau, perkembangan, anak]	guna pantau kembang anak
[primaku, keren, ada, fitur, pemantau, tumbuh, kembang anak, konsultasi, forum, membantu, ibu]	primaku keren ada fitur pantau tumbuh kembang anak konsultasi forum bantu ibu
[mantapp, tumbang, anak, terimakasih, primaku]	mantap tumbang anak terimakasih primaku

3.4 Pembobotan TF-IDF

Setelah tahapan *preprocessing*, tahapan selanjutnya adalah pembobotan TF-IDF dimana ulasan yang ada akan dibagi menjadi 80% untuk data training, dan 20% data testing. Setelah itu dilakukan pembobotan kata dengan

dua konsep yaitu TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*). TF (*Term Frequency*) mempunyai konsep menampilkan kata yang paling sering muncul dan memiliki jumlah yang sebanding dengan bobot kata tersebut, dimana semakin sering muncul maka semakin tinggi bobotnya. IDF (*Inverse Document Frequency*) memiliki konsep menampilkan bobot dalam satu dokumen, misalnya pada ulasan “bagus pantau tumbuh kembang anak” pembobotan TF-IDFnya dapat dilihat pada gambar 5.

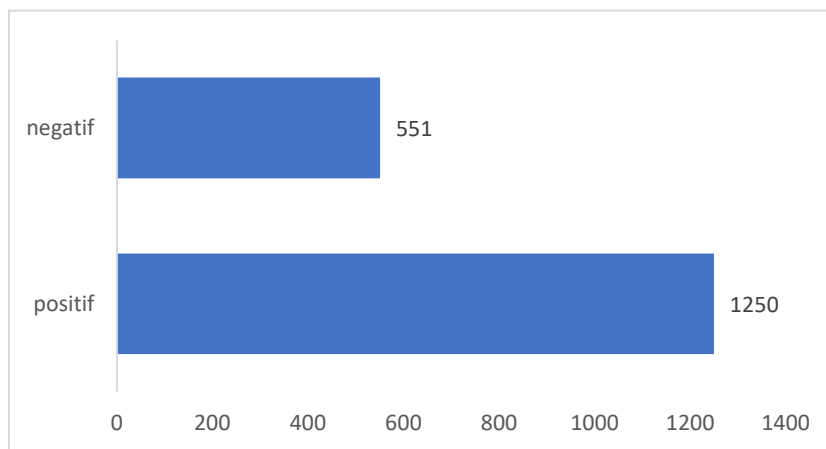
Berdasarkan gambar 5 sebelah kiri adalah pembobotan dengan *Term Frequency* dan sebelah kanan adalah pembobotan berdasarkan *Inverse Document Frequency*, hasil perhitungan TF-IDF yang dilakukan pada penelitian ini dilakukan secara manual menggunakan *python*[30], [31], dimana kolom pertama menunjukkan jumlah dokumen, kolom kedua menunjukkan letak katanya dan kolom ketiga adalah hasil *Term Frequency*.

TF (Term Frequency)		IDF (Inverse Document Frequency)	
bagus pantau tumbuh kembang anak		bagus pantau tumbuh kembang anak	
(0, 113)	1	[6.6405764	8.02687076 7.62140565 ... 7.33372358 7.33372358 7.11058003]
(0, 197)	1	(0, 2151)	0.4699692132318761
(0, 976)	1	(0, 1534)	0.5236629168984095
(0, 1534)	1	(0, 976)	0.4553078791363483
(0, 2151)	1	(0, 197)	0.3942313363159537
		(0, 113)	0.37707091739162124

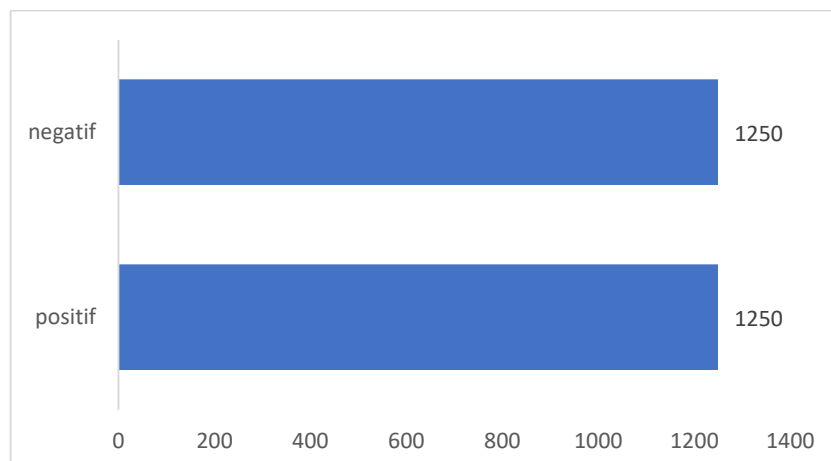
Gambar 5. Pembobotan TF-IDF

3.5 SMOTE (*Synthetic Minority Oversampling Technique*)

Synthetic Minority Oversampling Technique atau yang sering dikenal istilah SMOTE adalah teknik pengambilan sampel untuk kelas minoritas. SMOTE bekerja dengan memodifikasi kumpulan data yang tidak seimbang dengan membuat kumpulan data sintetik baru dari kelas minoritas dengan tujuan meningkatkan efisiensi metode klasifikasi[32]. Pada penelitian ini menggunakan *library imblearn*[11] yang tersedia pada *python* untuk implementasi teknik SMOTE. Adapun distribusi *class* setelah menggunakan teknik SMOTE dapat dilihat pada gambar 6 dan gambar 7.



Gambar 6. Perbandingan class sebelum menggunakan teknik SMOTE



Gambar 7. Perbandingan class setelah menggunakan teknik SMOTE

4. KESIMPULAN

Analisis sentimen terhadap aplikasi Primaku menunjukkan persepsi masyarakat terhadap aplikasi ini secara umum positif. Hal ini menunjukkan bahwa aplikasi Primaku dapat menjadi sarana yang bermanfaat bagi orang tua dalam memantau tumbuh kembang anak balita. Analisis sentimen ini dilakukan menggunakan metode Random Forest Classifier dan SMOTE untuk mengatasi ketidakseimbangan data. Hasil analisis menunjukkan bahwa metode ini dapat menghasilkan akurasi yang baik dalam menentukan sentimen dari suatu teks dengan tingkat akurasi mencapai 86%, precision 84%, recal 82%, F1-Score 83% setelah menggunakan teknik SMOTE terjadi peningkatan sebesar 2% sehingga tingkat akurasi 88%, precision 85%, recal 86%, F1-Score 85%. Sehingga dapat disimpulkan bahwa implementasi teknik SMOTE yang dikombinasikan dengan algoritma klasifikasi random forest untuk analisis sentimen aplikasi primaku dapat meningkatkan akurasi.

UCAPAN TERIMAKASIH

Ucapan terimakasih kami sampaikan kepada Universitas Bina Sarana Informatika atas pendanaan penelitian dengan skema penelitian dosen yayasan yang diberikan. Terimakasih juga kami sampaikan kepada Lembaga Penelitian dan Pengabdian Masyarakat yang telah memberikan motivasi dan kesempatan kepada kami sehingga penelitian ini dapat selesai tepat waktu.

REFERENCES

- [1] L. Yanti and H. F., "Pemberdayaan Kader Posyandu Dalam Pemantauan Tumbuh Kembang Balita Menggunakan Kpsp Dan Aplikasi 'Primaku' Sebagai Deteksi Dini Stunting Di Desa Tanah Merah Kabupaten Kampar," *Krida Cendekia*, vol. 1, no. 9, 2022.
- [2] Z. Munawar, Iswanto, D. Widhiantoro, and N. I. Putri, "Analisis Sentimen Covid-19 Pada Media Sosial Dengan Model Neural Machine Translation Analysis of Covid-19 Sentiment on Social Media with Neural Machine Translation Model," *TEMATIK (Jurnal teknologi Informasi Komunikasi)*, vol. 9, no. 1, pp. 15–20, 2022, doi: 10.38204/tematik.v9i1.917.
- [3] R. Safitri, N. Alfira, D. Tamitiadini, W. W. A. Dewi, and N. Febriani, *Analisis Sentimen: Metode Alternatif Penelitian Big Data*. Universitas Brawijaya Press, 2021.
- [4] S. Wahyuddin *et al.*, *Data Mining*. Global Eksekutif Teknologi, 2023.
- [5] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Playstore menggunakan Algoritma NBC," *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 42–48, 2023.
- [6] T. Tinaliah and T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, 2022, doi: <https://doi.org/10.35957/jatisi.v9i4.3586>.
- [7] W. A. Wahyuni, S. Saepudin, and F. Sembiring, "Sentiment Analysis of Online Investment Applications on Google Play Store using Random Forest Algorithm Method," *Jurnal Mantik*, vol. 5, no. 4, 2022.
- [8] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] M. Diki Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 5, no. 1, 2022.
- [10] R. Aryanti *et al.*, "The use of resampling techniques to overcome imbalance of data on the classification algorithm," in *AIP Conference Proceedings*, American Institute of Physics Inc., May 2023. doi: 10.1063/5.0128424.
- [11] G. Gumelar, Q. Ain, R. Marsuciati, S. Agustanti Bambang, A. Sunyoto, and M. Syukri Mustafa, "Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance," in *SISFOTEK : Sistem Informasi dan Teknologi*, 2021, pp. 250–255.
- [12] M. I. Prasetyowati, N. U. Maulidevi, and K. Surendro, "The accuracy of Random Forest performance can be improved by conducting a feature selection with a balancing strategy," in *PeerJ Computer Science*, 2022.
- [13] E. C. Gök and M. O. Olgun, "SMOTE-NC and gradient boosting imputation based random forest classifier for predicting severity level of covid-19 patients with blood samples," *Neural Comput Appl*, vol. 33, no. 22, pp. 15693–15707, 2021.
- [14] A. N. Kasanah, M. Muladi, and U. Pujiyanto, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2. pp. 196–201, 2018. doi: <https://doi.org/10.29207/resti.v3i2.945>.

- [15] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, 2022, doi: <https://doi.org/10.30812/matrik.v21i3.1726>.
- [16] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 6, no. 4, pp. 1702–1708, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10908>
- [17] T. Asgari, A. Daneshvar, A. P. Chobar, M. Ebrahimi, and S. Abrahamyan, "Identifying key success factors for startups With sentiment analysis using text data mining," *International Journal of Engineering Business Management*, vol. 14, 2022, doi: 10.1177/18479790221131612.
- [18] N. Istiqamah, B. Surarso, and B. Warsito, "Classification of customer review using random forest classifier," *AIP Conf Proc*, vol. 2738, no. 1, 2023.
- [19] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 4, pp. 1702–1708, 2022.
- [20] Prasetya J. and Abdurakhman A., "Comparison of Smote Random Forest and Smote K-Nearest Neighbors Classification Analysis on Imbalanced Data," *Media Statistika*, vol. 15, no. 2, pp. 198–208, 2023, doi: 10.14710/medstat.15.2.198-208.
- [21] R. Alshamy, M. Ghurab, S. Othman, and F. Alshami, "Intrusion Detection Model for Imbalanced Dataset Using SMOTE and Random Forest Algorithm," in *Advances in Cyber Security. ACeS 2021. Communications in Computer and Information Science*, vol. 1487, N. Abdullah, S. Manickam, and M. Anbar, Eds., Singapore: Springer, 2021. doi: 10.1007/978-981-16-8059-5_22.
- [22] V. Christanti Mawardi and E. Darmaja, "Logistic Regression Method for Sentiment Analysis Application on Google Playstore," *International Journal of Application on Sciences, Technology and Engineering*, vol. 1, no. 1, pp. 241–247, Feb. 2023, doi: 10.24912/ijaste.v1.i1.241-247.
- [23] A. V Saurkar, K. G. Pathare, and S. A. Gode, "An overview on web scraping techniques and tools," *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 4, no. 4, pp. 363–367, 2018.
- [24] A. Ulfah and I. Najiah, "Implementasi Web Scraping Pada Situs Jurnal Sinta Menggunakan Framework Selenium Web Driver Python," *JIKA (Jurnal Informatika)*, vol. 7, no. 1, pp. 29–36, 2023.
- [25] M. Djufri, "Penerapan Teknik Web Scraping Untuk Penggalan Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak)," *Jurnal BPPK: Badan Pendidikan Dan Pelatihan Keuangan*, vol. 13, no. 2, pp. 65–75, 2020.
- [26] D. P. Santoso and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store," *Jurnal Sains dan Seni ITS*, vol. 11, no. 2, pp. 190–196, 2022.
- [27] I. Muslim Karo Karo *et al.*, "Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine," *Journal of Information System Research*, vol. 4, no. 4, pp. 1423–1430, 2023, doi: 10.47065/josh.v4i4.3943.
- [28] Z. Min, "Drugs Reviews Sentiment Analysis using Weakly Supervised Model*," in *International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2019, pp. 332–336.
- [29] HaqueTanjim Ul, N. N. Saber, and F. M. Shah, "Sentiment Analysis on Large Scale Amazon Product Reviews," in *International Conference on Innovative Research and Developme*, 2018, pp. 190–196.
- [30] S. Rabbani, D. Safitri, N. Rahmadhani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023.
- [31] R. A. Supono and M. A. Suprayogi, "Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 911–918, 2021.
- [32] A. Syukron, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," 2023. [Online]. Available: <https://doi.org/10.25047/jtit.v10i1.312>