



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Penggunaan juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

BAB II

TINJAUAN LITERATUR

Bab ini membahas landasan teori yang digunakan sebagai dasar dalam pelaksanaan penelitian. Teori-teori yang disajikan meliputi konsep Natural Language Processing (NLP), teknik preprocessing teks, metode ekstraksi fitur menggunakan TF-IDF, SMOTE, serta algoritma klasifikasi XGBoost dan random forest. Selain itu, pada bab ini juga dibahas penelitian terdahulu yang relevan sebagai referensi dan pembandingan dalam mendukung penelitian yang dilakukan. Dengan adanya tinjauan pustaka ini, diharapkan penelitian memiliki dasar teoritis yang kuat serta arah yang jelas dalam proses analisis dan pembahasan.

2.1 Aplikasi JobStreet

JobStreet adalah aplikasi pencarian kerja yang cukup familiar di Indonesia yang telah beroperasi lebih dari 20 tahun. Aplikasi ini menyediakan layanan pencarian lowongan pekerjaan, pengiriman lamaran, pengelolaan profil, hingga rekomendasi pekerjaan berbasis kecerdasan buatan. JobStreet beroperasi di beberapa negara Asia, termasuk Indonesia, Malaysia, Singapura, Filipina, dan Thailand, serta tergabung dalam ekosistem SEEK, grup perusahaan lapangan kerja online terbesar di kawasan Asia Pasifik.

Berdasarkan data resmi dari laman Jobstreet, platform ini telah terhubung dengan lebih dari 42 juta kandidat dan 1,35 juta perusahaan di Asia. Sementara itu, berbagai sumber pemberitaan menyebutkan bahwa JobStreet memiliki lebih dari 40 juta pengguna secara global. Di Indonesia secara khusus, Jobstreet by SEEK memiliki lebih dari 20 juta kandidat terdaftar dan lebih dari setengah juta



perusahaan yang bergabung, mulai dari korporasi besar hingga pelaku UMKM. Platform ini juga telah mengiklankan lebih dari 1 juta lowongan pekerjaan dalam satu tahun terakhir melalui Gerakan #NextMillionJobs.

Tingkat efektivitas JobStreet dalam membantu pencari kerja juga tercermin dari data yang menunjukkan bahwa 1 dari 3 pencari kerja di Indonesia berhasil mendapatkan pekerjaan melalui platform ini. Selain itu, berdasarkan survei independen, Jobstreet by SEEK menguasai pangsa pasar tiga kali lipat dibandingkan pesaing terdekatnya di Indonesia. Data ini menunjukkan bahwa JobStreet bukan hanya aplikasi dengan jumlah ulasan terbanyak di Google Play Store, tetapi juga merupakan platform yang sangat vital dalam ekosistem ketenagakerjaan digital di Indonesia. Hal ini memperkuat urgensi dan relevansi penelitian analisis sentimen terhadap ulasan penggunanya.

2.2 Analisis Sentimen

suatu tahapan yang dilakukan untuk mendeteksi, mengolah, serta mengelompokkan opini atau pendapat yang terdapat dalam data berbentuk teks ke dalam kategori sentimen tertentu, seperti positif, negatif, maupun netral. Pada penelitian ini, analisis sentimen dimanfaatkan untuk mengetahui kecenderungan penilaian pengguna terhadap aplikasi JobStreet berdasarkan ulasan yang diberikan melalui Google Play Store[4][5].

Ulasan tersebut umumnya memuat berbagai bentuk tanggapan pengguna, mulai dari pengalaman yang memuaskan, kritik terhadap fitur aplikasi, keluhan mengenai layanan, hingga saran untuk pengembangan sistem. Melalui analisis sentimen, data teks yang awalnya tidak terstruktur dapat diolah menjadi informasi yang lebih sistematis dan mudah dipahami sehingga dapat digunakan sebagai bahan

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

evaluasi terhadap kualitas aplikasi.[23]

2.3 Text Mining

Text mining merupakan proses pengolahan data berbasis teks yang bertujuan untuk menemukan informasi, pola, maupun pengetahuan tertentu dari sekumpulan dokumen. Pada penelitian ini, text mining dimanfaatkan untuk mengubah ulasan pengguna aplikasi JobStreet yang masih berupa data teks menjadi bentuk data relasional sehingga dapat diproses menggunakan algoritma machine learning.

Secara umum, tahapan text mining yang diterapkan dalam penelitian ini mencakup proses pengumpulan data, preprocessing teks, ekstraksi fitur, penanganan ketidakseimbangan data, proses klasifikasi, hingga evaluasi performa model yang digunakan. Melalui tahapan tersebut, data ulasan dapat dianalisis secara lebih sistematis untuk menghasilkan informasi yang relevan terhadap sentimen pengguna aplikasi JobStreet[5][6].

2.4 Pelabelan Sentimen

Pelabelan sentimen merupakan proses pemberian kategori sentimen pada setiap data ulasan yang dianalisis. Kategori sentimen tersebut umumnya dibedakan menjadi tiga kelas, yaitu positif, negatif, dan netral. Pada ulasan di Google Play Store, tahapan pelabelan dapat dijalankan berlandaskan nilai rating yang diberikan pengguna maupun melalui penentuan label secara manual. Dalam penelitian ini, rating digunakan sebagai acuan awal dalam proses pelabelan, di mana rating 1-2 dikategorikan sebagai sentimen negatif, rating 3 sebagai sentimen netral, dan rating 4-5 sebagai sentimen positif. Mesk demikian, proses uji secara manual tetap dilakukan untuk memastikan bahwa label sentimen yang diberikan telah sesuai dengan isi dan makna dari ulasan pengguna aplikasi JobStreet[2][5].





2.5 Preprocessing Teks

Preprocessing teks merupakan tahap awal dalam pengolahan data berbasis teks yang dilakukan untuk membersihkan serta mempersiapkan data sebelum masuk ke proses ekstraksi fitur dan klasifikasi. Tahapan ini memiliki peran penting karena ulasan pengguna umumnya masih mengandung berbagai elemen yang tidak relevan, seperti penggunaan huruf kapital, tanda baca, angka, emoji, URL, kata tidak baku, hingga kata-kata universal yang kurang memiliki nilai. Pada penelitian ini, proses preprocessing yang diterapkan meliputi case folding, cleaning, tokenizing, normalisasi kata, stopwords removal, dan stemming[5]. Melalui tahapan tersebut, data teks diharapkan menjadi lebih terstruktur dan lebih optimal untuk diproses dalam analisis sentimen.

Adapun tahapan preprocessing yang digunakan dalam penelitian ini:

1. Cleaning adalah proses membersihkan teks dari karakter yang tidak diperlukan.
2. Case folding adalah proses mengubah seluruh huruf menjadi huruf kecil.
3. Normalisasi adalah proses mengubah kata tidak baku atau singkatan menjadi kata baku.
4. Stopword removal adalah proses menghapus kata-kata umum yang kurang memiliki makna penting dalam proses klasifikasi.
5. Stemming adalah proses mengubah kata berimbuhan menjadi bentuk kata dasar.

2.6 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk mengukur tingkat pentingnya suatu kata dalam sebuah dokumen. Metode ini memberikan nilai bobot tinggi pada kata yang



sering muncul dalam satu dokumen, tetapi memiliki frekuensi rendah pada dokumen lainnya. Dalam library Scikit-learn, proses tersebut dapat dilakukan menggunakan *TfidfVectorizer*, yaitu fitur yang berfungsi mengubah kumpulan dokumen teks menjadi matriks berbasis bobot TF-IDF[7][5].

2.7 SMOTE

SMOTE atau *Synthetic Minority Over-sampling Technique* merupakan metode oversampling yang digunakan untuk mengatasi permasalahan ketidakseimbangan data pada proses klasifikasi. Berbeda dengan metode oversampling konvensional yang hanya menduplikasi data pada kelas minoritas, SMOTE menghasilkan data sintetis baru berdasarkan karakteristik sampel dari kelas minoritas tersebut. Pendekatan ini membuat distribusi data menjadi lebih seimbang sehingga model dapat mempelajari pola pada kelas minoritas dengan lebih optimal[8][9][10].

2.8 XGBoost

XGBoost atau *Extreme Gradient Boosting* merupakan algoritma *machine learning* yang dikembangkan berdasarkan metode *gradient boosting* dengan proses pembentukan model secara bertahap. Setiap model yang dibangun bertujuan untuk memperbaiki kesalahan prediksi dari model sebelumnya sehingga kinerja prediksi dapat meningkat secara optimal. XGBoost dikenal memiliki efisiensi komputasi yang baik, mendukung fitur regularisasi untuk mengurangi risiko *overfitting*, serta mampu menangani data dalam jumlah besar dan berdimensi tinggi[4][11][12].

2.9 Random Forest

Random Forest merupakan algoritma *machine learning* berbasis ensemble learning yang bekerja dengan membangun banyak decision tree untuk

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



menghasilkan prediksi akhir. Setiap pohon dibentuk dari sampel data dan fitur yang dipilih secara acak, kemudian hasil seluruh pohon digabungkan melalui voting untuk menentukan kelas akhir. Algoritma ini banyak digunakan dalam klasifikasi karena mampu menangani data berdimensi tinggi, stabil, dan dapat mengurangi risiko overfitting dibandingkan decision tree tunggal[2][5]. Dalam penelitian ini, Random Forest digunakan untuk mengklasifikasikan sentimen ulasan aplikasi JobStreet.

2.10 Confusion Matrix

Confusion matrix merupakan tabel evaluasi yang digunakan untuk menampilkan hasil prediksi dari suatu model klasifikasi dengan membandingkan label aktual dan label hasil prediksi model kesalahan prediksi[13].

2.11 Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat kinerja algoritma dalam proses klasifikasi sentimen. Pada penelitian ini, metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk menunjukkan persentase jumlah prediksi yang benar dibandingkan dengan seluruh data yang diuji. Precision menggambarkan tingkat ketepatan model dalam memprediksi suatu kelas tertentu. Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang termasuk ke dalam kelas tertentu. Sementara itu, F1-score merupakan nilai rata-rata harmonik antara precision dan recall, sehingga metrik ini dinilai efektif untuk digunakan pada data dengan distribusi kelas yang tidak seimbang[14][15].

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

2.12 Penelitian Terdahulu

Berikut Tabel 2.1 menyajikan Penelitian sebelumnya dijadikan sebagai acuan dalam menentukan metode dan pendekatan yang digunakan pada analisis sentimen.

Tabel 2. 1 Penelitian Terdahulu

No	Artikel	Judul	Hasil Penelitian
(1)	(2)	(3)	(4)
1	[2]	Perbandingan Algoritma Machine Learning untuk Sentimen Ulasan Aplikasi JobStreet	Penelitian ini menganalisis sentimen ulasan aplikasi JobStreet di Google Play Store menggunakan algoritma Naïve Bayes, SVM, Logistic Regression, dan Random Forest. Penelitian tersebut relevan karena menggunakan objek yang sama, yaitu ulasan aplikasi JobStreet. Hasil penelitian menunjukkan bahwa Naïve Bayes memperoleh tingkat akurasi tertinggi dibandingkan algoritma lainnya.
2	[16]	Sentiment Analysis of Jobstreet Application Reviews on Google Play Store Using Support Vector Machine Algorithm with Adaptive Synthetic	Penelitian lain membahas analisis sentimen ulasan aplikasi JobStreet menggunakan algoritma SVM dengan teknik oversampling ADASYN. Hasil penelitian menunjukkan bahwa penerapan ADASYN mampu meningkatkan akurasi SVM dari 89,08% menjadi 89,95%. Penelitian ini relevan karena sama-sama menggunakan objek JobStreet serta membahas penanganan data tidak seimbang.
3	[17]	Analisis Sentimen Negatif pada Aplikasi JobStreet Menggunakan Hadoop Distributed File System (HDFS)	Penelitian ini berfokus pada analisis sentimen negatif pada aplikasi JobStreet. Metode yang digunakan berkaitan dengan pengolahan data ulasan dan klasifikasi sentimen. Penelitian ini relevan karena sama-sama menggunakan objek aplikasi JobStreet, tetapi berbeda pada fokus dan metode penelitian.





Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
4. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
5. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

(1)	(2)	(3)	(4)
4	[4]	Analisis Sentimen Ulasan Pengguna Aplikasi Threads di Google Play Menggunakan Algoritma XGBoost dengan Penguatan SMOTE	Penelitian ini menggunakan algoritma XGBoost dengan teknik SMOTE pada ulasan aplikasi Threads di Google Play. Data direpresentasikan menggunakan TF-IDF dan dianalisis menggunakan XGBoost sebelum dan sesudah penerapan SMOTE. Hasil penelitian menunjukkan bahwa SMOTE mampu meningkatkan performa model, terutama pada nilai recall kelas negatif. Penelitian ini relevan karena sama-sama menggunakan metode XGBoost dan SMOTE.
5	[3]	Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers	Penelitian ini membandingkan algoritma Logistic Regression, SVM, XGBoost, Naïve Bayes, dan Random Forest pada data ulasan Google Reviews Puskesmas di Bali. Penelitian menggunakan metode TF-IDF dan SMOTE untuk mengatasi ketidakseimbangan data. Penelitian ini relevan karena sama-sama menerapkan TF-IDF, SMOTE, XGBoost, dan Random Forest.
6	[5]	Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest	Penelitian ini menggunakan algoritma Random Forest untuk mengklasifikasikan ulasan aplikasi di Google Play Store ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses klasifikasi dilakukan berdasarkan isi ulasan pengguna untuk mengetahui kecenderungan opini terhadap aplikasi yang dianalisis. Penelitian ini relevan karena sama-sama menggunakan data ulasan dari Google Play Store serta menerapkan algoritma Random Forest dalam analisis sentimen.



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
4. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
5. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

(1)	(2)	(3)	(4)
7	[18]	Analisis Sentimen Aplikasi M-Pajak Menggunakan Machine Learning Berbasis Algoritma XGBoost	Penelitian ini menganalisis sentimen ulasan pengguna aplikasi M-Pajak di Google Play Store menggunakan algoritma XGBoost. Proses analisis dilakukan untuk mengidentifikasi kecenderungan opini pengguna terhadap aplikasi berdasarkan ulasan yang diberikan. Penelitian ini relevan karena sama-sama menerapkan algoritma XGBoost dalam analisis sentimen ulasan aplikasi.
8	[19]	Analisis Sentimen Aplikasi Halo BCA di Google Play Store Menggunakan Metode Naive Bayes, Support Vector Machine dan Random Forest	Penelitian ini menganalisis ulasan aplikasi Halo BCA di Google Play Store menggunakan algoritma Naïve Bayes, SVM, dan Random Forest. Hasil penelitian menunjukkan bahwa Random Forest memiliki performa terbaik dibandingkan algoritma lainnya dalam klasifikasi sentimen. Penelitian ini relevan karena sama-sama menggunakan data ulasan dari Google Play Store serta menerapkan algoritma Random Forest.
9	[20]	Analisis Sentimen Traveloka Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine dan Random Forest	Penelitian ini menganalisis sentimen ulasan aplikasi Traveloka di Google Play Store menggunakan algoritma SVM dan Random Forest. Analisis dilakukan untuk mengetahui kecenderungan opini pengguna terhadap aplikasi berdasarkan ulasan yang diberikan. Penelitian ini relevan karena sama-sama membahas analisis sentimen ulasan aplikasi serta menggunakan Random Forest sebagai salah satu algoritma klasifikasi.



(1)	(2)	(3)	(4)
10	[21]	Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models	Penelitian ini menganalisis sentimen ulasan Tokopedia menggunakan TF-IDF, SMOTE, dan beberapa model machine learning, termasuk Random Forest. Penelitian ini relevan karena menggunakan TF-IDF dan SMOTE dalam proses klasifikasi sentimen..

Berdasarkan Tabel 2.1 Penelitian Terdahulu, dapat diketahui bahwa penelitian mengenai analisis sentimen pada ulasan aplikasi telah banyak dilakukan dengan menggunakan berbagai objek dan metode yang berbeda. Sejumlah penelitian memanfaatkan ulasan pengguna di Google Play Store sebagai sumber data, seperti pada aplikasi JobStreet, Threads, Halo BCA, Traveloka, dan M-Pajak. Kondisi tersebut menunjukkan bahwa ulasan pengguna pada platform digital memiliki potensi yang besar untuk dimanfaatkan dalam mengidentifikasi opini, persepsi, dan pengalaman pengguna terhadap suatu aplikasi.

Penelitian yang memiliki keterkaitan paling kuat dengan penelitian ini adalah penelitian yang dilakukan oleh Marzuqi dkk. serta Shantika dan Abidin karena sama-sama menggunakan ulasan aplikasi JobStreet di Google Play Store sebagai objek penelitian. Marzuqi dkk. melakukan perbandingan beberapa algoritma *machine learning*, termasuk *Random Forest*, sedangkan Shantika dan Abidin menerapkan algoritma SVM dengan teknik oversampling ADASYN. Meskipun demikian, kedua penelitian tersebut belum mengombinasikan teknik SMOTE dengan algoritma XGBoost dan *Random Forest* pada analisis sentimen ulasan aplikasi JobStreet.



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Di sisi lain, penelitian yang dilakukan oleh Setiyono dkk serta Budaya dan Suniantara menunjukkan bahwa teknik SMOTE efektif digunakan untuk mengatasi permasalahan ketidakseimbangan data pada analisis sentimen. Hasil penelitian mereka juga membuktikan bahwa algoritma XGBoost dan *Random Forest* mampu memberikan performa yang baik dalam proses klasifikasi sentimen berbasis ulasan pengguna. Namun, objek penelitian yang digunakan berbeda dengan penelitian ini, yaitu aplikasi Threads dan Google Reviews Puskesmas. Oleh karena itu, penelitian ini menawarkan kebaruan melalui penerapan kombinasi SMOTE dengan algoritma XGBoost dan *Random Forest* pada ulasan aplikasi JobStreet di Google Play Store.

Berdasarkan uraian tersebut, penelitian ini berupaya melanjutkan sekaligus mengembangkan penelitian terdahulu dengan mengintegrasikan beberapa pendekatan yang relevan. Penelitian menggunakan data ulasan aplikasi JobStreet, melakukan pelabelan sentimen positif dan negatif, menerapkan tahapan *preprocessing* teks, memanfaatkan TF-IDF sebagai metode ekstraksi fitur, serta menggunakan SMOTE untuk mengatasi ketidakseimbangan data. Selanjutnya, proses klasifikasi sentimen dilakukan dengan algoritma XGBoost dan *Random Forest*. Melalui pendekatan tersebut, penelitian ini diharapkan dapat mengidentifikasi algoritma yang memiliki performa terbaik berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

Berdasarkan hasil kajian penelitian terdahulu, dapat disimpulkan bahwa analisis sentimen terhadap ulasan aplikasi JobStreet telah banyak dilakukan, namun masih terdapat peluang pengembangan terutama dalam aspek penanganan ketidakseimbangan data dan pemilihan algoritma klasifikasi. Penelitian sebelumnya pada objek JobStreet umumnya menggunakan algoritma seperti *Naïve*



Bayes, *SVM*, *Logistic Regression*, *Random Forest*, serta teknik oversampling ADASYN. Sementara itu, penerapan kombinasi SMOTE dengan algoritma XGBoost dan *Random Forest* pada analisis sentimen ulasan aplikasi JobStreet masih jarang ditemukan dalam penelitian yang telah ada. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan menerapkan teknik SMOTE guna mengatasi ketidakseimbangan data serta membandingkan performa algoritma XGBoost dan *Random Forest* dalam proses klasifikasi sentimen ulasan aplikasi JobStreet.

2.13 Rangkuman

Secara umum, penelitian terdahulu, Berdasarkan penelitian terdahulu yang telah dikaji, dapat diketahui bahwa analisis sentimen pada ulasan aplikasi telah banyak dilakukan dengan memanfaatkan berbagai algoritma *machine learning*, seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Logistic Regression*, *Random Forest*, dan XGBoost. Selain itu, beberapa penelitian juga telah menerapkan teknik oversampling, seperti SMOTE dan ADASYN, untuk mengatasi permasalahan ketidakseimbangan data. Penelitian yang menggunakan ulasan aplikasi JobStreet sebagai objek kajian juga telah dilakukan, namun sebagian besar masih berfokus pada penggunaan algoritma tertentu dan belum mengombinasikan teknik SMOTE dengan algoritma XGBoost dan *Random Forest* dalam satu penelitian.

Penelitian yang dilakukan oleh Marzuqi dkk dan Shantika & Abidin menggunakan objek yang sama, yaitu ulasan aplikasi JobStreet pada Google Play Store, tetapi metode yang diterapkan berbeda dengan penelitian ini. Di sisi lain, penelitian Setiyono dkk serta Budaya dan Suniantara menunjukkan bahwa penerapan SMOTE mampu meningkatkan kinerja model klasifikasi pada data yang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



tidak seimbang. Meskipun demikian, perbedaan objek penelitian menyebabkan hasil yang diperoleh belum dapat menggambarkan kondisi sentimen pengguna aplikasi JobStreet secara spesifik.

Berdasarkan kajian tersebut, masih terdapat peluang untuk mengembangkan penelitian dengan menerapkan teknik SMOTE pada data ulasan aplikasi JobStreet yang berpotensi memiliki distribusi kelas sentimen yang tidak seimbang. Selain itu, penggunaan algoritma XGBoost dan *Random Forest* pada objek penelitian yang sama memungkinkan dilakukan perbandingan performa untuk mengetahui algoritma yang lebih efektif dalam mengklasifikasikan sentimen pengguna.

Secara teoritis, penelitian ini didasarkan pada konsep analisis sentimen yang merupakan bagian dari *text mining* untuk mengidentifikasi opini pengguna berdasarkan data berbentuk teks. Sebelum dilakukan analisis, data ulasan terlebih dahulu melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Tahapan tersebut bertujuan untuk membersihkan serta menstandarkan data sehingga lebih siap digunakan pada proses pengolahan berikutnya.

Data yang telah melalui proses *preprocessing* kemudian dikonversi ke dalam bentuk numerik menggunakan metode TF-IDF. Selanjutnya, teknik SMOTE diterapkan untuk menyeimbangkan distribusi data sentimen sebelum dilakukan proses klasifikasi. Tahap klasifikasi menggunakan dua algoritma *machine learning*, yaitu XGBoost dan *Random Forest*. Hasil klasifikasi kemudian dievaluasi menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan model yang memiliki performa terbaik.

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Berdasarkan penelitian terdahulu dan landasan teori yang digunakan, kebaruan penelitian ini terletak pada penerapan teknik SMOTE yang dikombinasikan dengan algoritma XGBoost dan *Random Forest* untuk analisis sentimen ulasan aplikasi JobStreet pada Google Play Store. Penelitian ini diharapkan dapat memberikan gambaran mengenai sentimen pengguna terhadap aplikasi JobStreet sekaligus mengidentifikasi algoritma yang memiliki performa paling optimal dalam proses klasifikasi sentimen.



Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

