

PERBANDINGAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA JOBSTREET

Andica Parlindungan Zai^{1*}, Sushanty Saleh²

^{1,2}Sistem Informasi, Fakultas Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia

Email: ^{1*}andika.zai.2111050017@mail.darmajaya.ac.id, ²sushantysaleh@darmajaya.ac.id

(*: coresponding author)

Abstrak-Meningkatnya penggunaan platform pencarian kerja digital seperti *Jobstreet* menghasilkan banyak ulasan pengguna yang dapat dimanfaatkan untuk memahami tingkat kepuasan dan pengalaman mereka terhadap layanan tersebut. Namun, jumlah ulasan yang besar membuat analisis manual menjadi tidak efektif sehingga diperlukan pendekatan melalui analisis sentimen. Penelitian ini bertujuan untuk menganalisis 4835 data ulasan pengguna jobstreet yang dimana data diperoleh dari *Google Play Store* dengan pendekatan *text mining*, eksperimen ini berupaya membandingkan dua metode klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam mengelompokkan ulasan pengguna jobstreet kedalam sentimen positif dan negatif. Selain itu, penelitian ini juga menerapkan teknik *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk menangani ketidakseimbangan kelas pada dataset ulasan, yang sering berdampak pada penurunan kemampuan model dalam mengenali sentimen dengan jumlah data yang lebih sedikit. Hasil evaluasi menunjukkan bahwa algoritma SVM memberikan performa lebih stabil dengan akurasi meningkat dari 89% menjadi 91%, sedangkan algoritma *Naïve Bayes* meningkat dari 88% menjadi 90%. Penerapan teknik SMOTE efektif dalam mengoptimalkan performa model dalam mendeteksi sentimen negatif, yang ditunjukkan oleh kenaikan recall pada kedua model, yaitu algoritma *Naïve Bayes* dari 85% menjadi 95% dan SVM dari 87% menjadi 93%. Secara keseluruhan, metode SVM menjadi algoritma yang lebih unggul untuk analisis sentimen ulasan pengguna *Jobstreet* karena memberikan hasil yang lebih konsisten pada berbagai metrik evaluasi.

Kata Kunci: Analisis Sentimen, Jobstreet, Naïve Bayes, SVM, SMOTE

Abstract-The increasing use of digital job search platforms such as JobStreet has generated many user reviews that can be used to understand their level of satisfaction and experience with these services. However, the large number of reviews makes manual analysis ineffective, requiring an approach through sentiment analysis. This study aims to analyze 4,835 user review data from the Google Play Store using a text mining approach. This experiment attempts to compare two classification methods, namely Naïve Bayes and Support Vector Machine (SVM), in grouping JobStreet user reviews into positive and negative sentiments. In addition, this study also applies the Synthetic Minority Over-Sampling Technique (SMOTE) to address class imbalance in the review dataset, which often impacts the model's ability to recognize sentiment with fewer data points. The evaluation results show that SVM provides more stable performance with accuracy increasing from 89% to 91%, while Naïve Bayes increased from 88% to 90%. The application of SMOTE is effective in optimizing model performance in detecting negative sentiment, as evidenced by the increase in Naïve Bayes recall from 85% to 95% and SVM from 87% to 93%. Overall, the SVM method is the superior algorithm for analyzing user sentiment in JobStreet reviews because it provides more consistent results.

Keywords: Sentiment Analysis, Jobstreet, Naïve Bayes, SVM, SMOTE

1. PENDAHULUAN

Perkembangan teknologi digital yang berlangsung cepat telah mengubah cara manusia mendapatkan informasi, terutama dalam proses rekrutmen dan pencarian pekerjaan. Sebelumnya, pelamar kerja harus mencari lowongan melalui surat kabar atau datang langsung ke perusahaan, kini proses tersebut menjadi jauh lebih praktis berkat hadirnya berbagai aplikasi digital[1]. Hanya menggunakan *smartphone*, pelamar dapat mencari lowongan, melamar pekerjaan, hingga memantau proses seleksi secara mudah dan cepat. Salah satu aplikasi yang sering dipakai di Indonesia ialah Jobstreet. Platform ini didirikan di Malaysia pada tahun 1997 oleh *Mark Chang Mun Kee* dan berkembang menjadi salah satu situs rekrutmen digital terbesar di Asia Tenggara berkat meningkatnya penggunaan internet dan kebutuhan solusi perekutan digital, layanan ini kemudian meluas di bermacam negara termasuk Indonesia[2]. Sampai saat ini, Jobstreet telah banyak digunakan pengguna untuk mencari informasi pekerjaan, membangun profil profesional, dan berkomunikasi langsung dengan perusahaan. Popularitas Jobstreet ini terlihat dari banyaknya pengguna melalui platform di *Google Play Store*.

Google Play Store ialah platform berbasis android untuk menyediakan beragam aplikasi seperti toko online, permainan, hingga pencarian pekerjaan dan layanan digital lainnya. Pada setiap halaman aplikasi di *Google Play Store* memiliki bagian komentar yang memungkinkan pengguna sering menuangkan pendapat mereka terhadap aplikasi yang sudah digunakan atau diunduh, baik dalam bentuk rating maupun ulasan[3]. Ulasan pengguna menjadi sumber data penting untuk menilai sejauh mana tingkat kepuasan terhadap layanan Jobstreet. Namun, ulasan tersebut jumlahnya yang sangat banyak dan beragam sehingga sulit dianalisis secara manual. Oleh karena itu, diperlukan analisis sentimen agar opini setiap pengguna dapat diolah dan dimanfaatkan secara lebih optimal.

Analisis sentimen merupakan proses pengelompokan teks dalam suatu kalimat atau dokumen untuk mengetahui pendapat yang disampaikan di dalamnya. Melalui proses ini, teks diklasifikasikan ke dalam sentimen bersifat positif atau negatif sehingga dapat menggambarkan pandangan atau respons pengguna terhadap suatu objek, layanan, maupun isu tertentu[4].

Dalam penelitian ini, menentukan metode klasifikasi yang sesuai sangat diperlukan agar memperoleh output yang akurat. Dua metode yang banyak memanfaatkan dalam studi analisis sentimen adalah *Naïve Bayes* dan *Support Vector Machine*[5]. *Naïve Bayes* ialah salah satu metode *machine learning* yang memakai probabilitas dalam upaya pengklasifikasian terhadap sebuah dokumen. Pendekatan tersebut memiliki keunggulan berupa model yang sederhana, tetapi mampu menghasilkan ketepatan akurasi yang tinggi[6]. Selain itu, *Support Vector Machine* adalah algoritma kategori yang dirancang untuk memisahkan dua kelas dengan membangun sebuah *hyperplane* dengan disertai margin pada ruang fitur. Metode ini sangat sesuai digunakan dalam penelitian ini karena proses klasifikasi hanya melibatkan 2 sentimen yaitu, positif dan negatif, sehingga kemampuan SVM dalam memaksimalkan pemisahan antar kelas dapat dimanfaatkan secara optimal untuk menghasilkan akurasi yang baik[7].

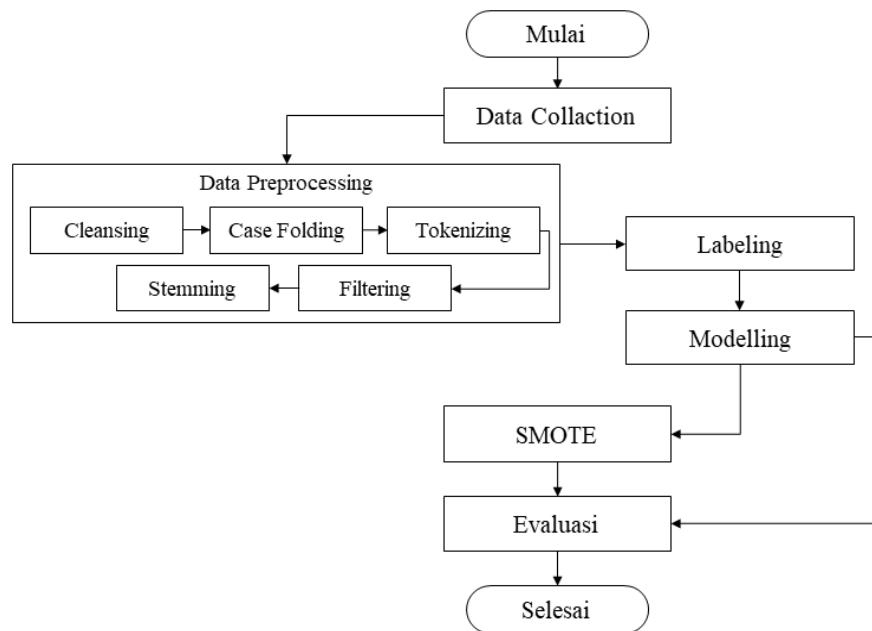
Beberapa penelitian sebelumnya telah dilakukan dalam bidang analisis sentimen pada ulasan aplikasi. Penelitian oleh [8] menganalisis sentimen terkait ulasan aplikasi *Threads* memakai metode *Naïve Bayes* dimana data di dapat dari aplikasi *Google Play Store* dengan memanfaatkan cara *web scrapping*, dan diperoleh akurasi yang baik sebesar 73%. Penelitian lainnya mengenai analisis sentimen ulasan aplikasi PLN mobile dari *Google Play Store* dengan memakai algoritma *Naïve bayes* dan *K-Nearest Neighbor*, hasil menunjukkan *Naïve Bayes* lebih tinggi daripada KNN dengan akurasi 77,69%, *recall* 53,14%, *precision* 59,84% dan *f1-score* 54,09%[9]. Berikutnya penelitian tentang komparasi metode analisis sentimen menggunakan *Naïve Bayes* dan *Support Vector Machine* pada aplikasi BRImo, pada metode SVM skor akurasi 97,69% sementara *Naïve Bayes* dengan skor akurasi 96,53%[10].

Berdasarkan hasil penelitian-penelitian tersebut dapat dijadikan acuan untuk menulis jurnal penelitian saat ini, umumnya menunjukkan bahwa performa kedua metode dapat berbeda tergantung pada karakteristik dataset yang digunakan, seperti jumlah data, pesebaran kelas, serta kualitas fitur. Meskipun begitu, hanya sedikit studi yang secara khusus membandingkan kedua algoritma dalam lingkup analisis sentimen ulasan pengguna jobstreet. Selain itu, pada studi sebelumnya belum menerapkan optimasi *Synthetic Minority Over-sampling Technique* (SMOTE) pada masing-masing metode klasifikasi yang digunakan sehingga kinerja setiap algoritma belum mencapai hasil yang maksimal.

Penelitian ini bertujuan untuk membandingkan performa metode klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam melakukan analisis sentimen terhadap ulasan pengguna Jobstreet. Selain itu, penelitian ini juga menerapkan teknik SMOTE sebagai upaya mengatasi ketidakseimbangan kelas pada dataset sehingga kualitas model yang dilatih semakin lebih baik.

2. METODE PENELITIAN

Penelitian ini dilaksanakan melalui sejumlah tahapan, sehingga memiliki alur yang tersusun dari awal hingga akhir. Tahapan penelitian ini bisa ditinjau di Gambar 1.



Gambar 1. Alur Penelitian

Pada Gambar 1, menunjukkan alur penelitian ini dimulai pada tahap pengumpulan data (*Data Collection*) yakni kumpulan ulasan pengguna aplikasi Jobstreet yang diperoleh melalui *Google Play Store*. Setelah diperoleh, tahapan berikutnya melakukan data preprocessing agar data siap digunakan pada tahapan pemodelan dengan dua algoritma, seperti *Naïve Bayes* dan *Support Vector Machine*. Penggunaan algoritma tersebut bertujuan untuk membandingkan kinerja setiap model untuk memperoleh performa terbaik. Setelah didapat performa awal selanjutnya akan dilakukan optimasi menggunakan *SMOTE* untuk meningkatkan akurasi dari setiap model klasifikasi. Pada tahap akhir, peneliti melakukan evaluasi terhadap model yang telah dibangun. Uraian lebih detail akan dijelaskan dibawah ini.

2.1 Data Collection

Dalam penelitian ini, data diperoleh dengan cara memanfaatkan bahasa pemrograman Python memakai library *google-play-scraper* untuk melakukan proses *web scraping* dari platform *Google Play Store*[11]. Library ini digunakan untuk mengekstrasi ulasan pengguna dari aplikasi Jobstreet dengan mengakses aplikasi berdasarkan *package name* “*com.jobstreet.jobstreet*”, yang merupakan identitas unik aplikasi di *Google Play Store*. Setelah data terkumpul akan disimpan dalam file CSV. Gambaran datanya dapat diperhatikan dalam Tabel 1.

Tabel 1. Scraping Data

Tanggal	Ulasan	Rating
11/11/2025	mantul	5
11/11/2025	Sangat berguna buat saya dalam mencari pekerjaan	4
11/11/2025	apk jelek, kebanyakan notif gajelas	1
11/11/2025	Bertahun tahun tepi belum pernah berhasil dapat pekerjaan di jobstreet, semoga kali ini berhasil	4
11/11/2025	Sangat informatif	5

2.2 Data Preprocessing

Data Preprocessing adalah langkah pertama pada proses klasifikasi teks yang berfungsi mempersiapkan data sebelum memasuki tahapan berikutnya. Pada proses ini, data diolah dan diubah dalam bentuk yang lebih terstruktur sehingga meningkatkan hasil analisis yang baik [12]. Tahapan yang dilakukan *preprocessing* adalah sebagai berikut.

a. *Cleansing*

Cleansing merupakan dokumen teks yang dibersihkan dari elemen-elemen yang tidak relevan atau dinilai sebagai gangguan. Pada Tabel 2, hasil *cleansing* ini meliputi pengapusan tanda baca seperti koma (,) dan emoji.

Tabel 2. Cleansing

Sebelum	Sesudah
Tolong bantu saya verifikasi dong, saya udah nyoba berulang kali tetep gagal di bagian selfie, padahal udah bener tiap syaratnya 😊	Tolong bantu saya verifikasi dong saya udah nyoba berulang kali tetep gagal di bagian selfie padahal udah bener tiap syaratnya

b. *Case Folding*

Case Folding ialah proses standarisasi penulisan huruf pada dokumen teks agar mempermudah proses pengolahan teks. Pada langkah ini, seluruh karakter alfabet dikonversi ke bentuk huruf kecil. Hasil *case folding* bisa ditinjau dalam Tabel 3.

Tabel 3. Case Folding

Sebelum	Sesudah
Tolong bantu saya verifikasi dong saya udah nyoba berulang kali tetep gagal di bagian selfie padahal udah bener tiap syaratnya	tolong bantu saya verifikasi dong saya udah nyoba berulang kali tetep gagal di bagian selfie padahal udah bener tiap syaratnya

c. *Tokenizing*

Tokenizing merupakan proses memisahkan sebuah kalimat atau paragraf menjadi beberapa individu yang disebut nama token [13]. Hasil *tokenizing* bisa ditinjau dalam Tabel 4.

Tabel 4. Tokenizing

Sebelum	Sesudah
tolong bantu saya verifikasi dong saya udah nyoba berulang kali tetep gagal di bagian selfie padahal udah bener tiap syaratnya	['tolong', 'bantu', 'saya', 'verifikasi', 'dong', 'saya', 'udah', 'nyoba', 'berulang', 'kali', 'tetep', 'gagal', 'di', 'bagian', 'selfie', 'padahal', 'udah', 'bener', 'tiap', 'syaratnya']

d. *Filtering*

Filtering adalah tahap yang dilakukan untuk menyingkirkan istilah yang tidak dibutuhkan setelah proses tokenisasi. Istilah yang kemunculanya sangat sering namun tidak memberikan makna penting disebut *stopword* [14]. Hasil *filtering* bisa ditinjau dalam Tabel 5.

Tabel 5. Filtering

Sebelum	Sesudah
['tolong', 'bantu', 'saya', 'verifikasi', 'dong', 'saya', 'udah', 'nyoba', 'berulang', 'kali', 'tetep', 'gagal', 'di', 'bagian', 'berulang', 'kali', 'tetep', 'gagal', 'selfie', 'udah', 'padahal', 'udah', 'bener', 'tiap', 'syaratnya']	['tolong', 'bantu', 'verifikasi', 'udah', 'nyoba', 'berulang', 'kali', 'tetep', 'gagal', 'selfie', 'udah', 'bener', 'syaratnya']

e. *Stemming*

Stemming merupakan langkah pemrosesan teks yang bertujuan mengembalikan ke bentuk asalnya dengan cara menghapus imbuhan yang ada dalam kalimat [15]. Hasil *stemming* bisa ditinjau dalam Tabel 6.

Tabel 6. Stemming

Sebelum	Sesudah
['tolong', 'bantu', 'verifikasi', 'udah', 'nyoba', 'berulang', 'kali', 'tetep', 'gagal', 'selfie', 'udah', 'bener', 'syaratnya']	tolong bantu verifikasi udah nyoba ulang kali tetep gagal selfie udah bener syarat

2.3 Labeling dan Ekstraksi Fitur

Pada langkah ini, data yang terkumpul dan sudah dibersihkan akan di beri label berdasarkan skor ulasan yang bernilai 1 sampai 5 kemudian dikelompokkan ke dalam dua kategori sentimen yaitu negatif untuk skor 1 sampai 3 kemudian kategori positif apabila skor berupa pada nilai 4 sampai 5.

2.4 Modeling

Pada bagian ini, dibangun metode klasifikasi menggunakan data yang telah diproses pada tahap sebelumnya, dengan menerapkan algoritma *Naïve Bayes* dan *Support Vector Machine* sebagai pendekatan pemodelan dalam penelitian ini. *Naïve Bayes* merupakan teknik klasifikasi yang berlandaskan *Teorema Bayes*. Pendekatan ini memanfaatkan konsep probabilitas dan statistik untuk melakukan pengelompokan data dengan cara menghitung peluang suatu kelas berdasarkan data sebelumnya. Keuntungan dari metode ini adalah data pelatihan relatif sedikit untuk menghitung parameter yang dibutuhkan dalam proses klasifikasi [16].

$$P(H | X) = \frac{P(X | H).P(H)}{P(X)} \quad (1)$$

X adalah sekumpulan data yang belum memiliki label, Sementara H adalah perkiraan label untuk data X yang dianggap paling mungkin, $P(H|X)$ menunjukkan peluang bahwa dugaan H benar berdasarkan kondisi X (*posterior probability*), $P(H)$ merupakan peluang awal dari dugaan H sebelum mempertimbangkan data (*prior probability*), $P(X|H)$ artinya seberapa besar kemungkinan X muncul jika dugaan H benar, $P(X)$ merupakan peluang keseluruhan dari munculnya data X.

Support Vector Machine (SVM) merupakan sebuah teknik dalam *machine learning* untuk mengklasifikasikan data dan tergolong ke dalam kategori *supervised learning*, sehingga membutuhkan data yang telah diberi label. SVM bekerja dengan memisahkan data berdasarkan pencarian *hyperplane* yang terbaik yang memiliki margin paling lebar. *Hyperplane* berfungsi sebagai garis atau bidang pemisah antara dua kelas, sedangkan *margin* adalah jarak antara *support vector*, yaitu titik-titik data yang posisinya berada paling dekat dengan batas pemisah [17]. Bentuk rumus dasar SVM di tunjukan pada persamaan 2.

$$w.x - b = 0 \quad (2)$$

Simbol w merupakan *vector* parameter yang menentukan arah *hyperplane*, x adalah *vector* atau fitur yang sedang diprediksi, b nilai bias yang menggeser posisi *hyperplane*.

2.5 Optimasi SMOTE

SMOTE (*Synthentic Minority Oversampling Tecnique*) adalah salah satu teknik pengolahan data untuk menyeimbangkan data antar kelas dengan cara mengambil sampel dari kelas minoritas dan menambahkan data sintetis baru untuk meningkatkan jumlah pada kelas tersebut [18]. Hal ini dapat membantu model dalam melakukan klasifikasi dengan lebih optimal.

2.6 Evaluasi

Pada bagian penutup, proses evaluasi dilakukan untuk menilai keakuratan model yang dihasilkan dengan memanfaatkan *confusion matrix*. Penilaian performa algoritma *Naïve Bayes* dan *Support Vector Machine* didasarkan pada empat metrik utama yaitu, *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Dari keempat metrik tersebut dihitung menurut nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), *True Positive* (TP) [19].

Tabel 7. Label Confusion Matrix

Sentimen	Prediksi Negatif	Prediksi Positif
Negatif	<i>True Negative</i> (TN)	<i>False Positive</i> (FP)
Positif	<i>False Negative</i> (FN)	<i>True Positive</i> (TP)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

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

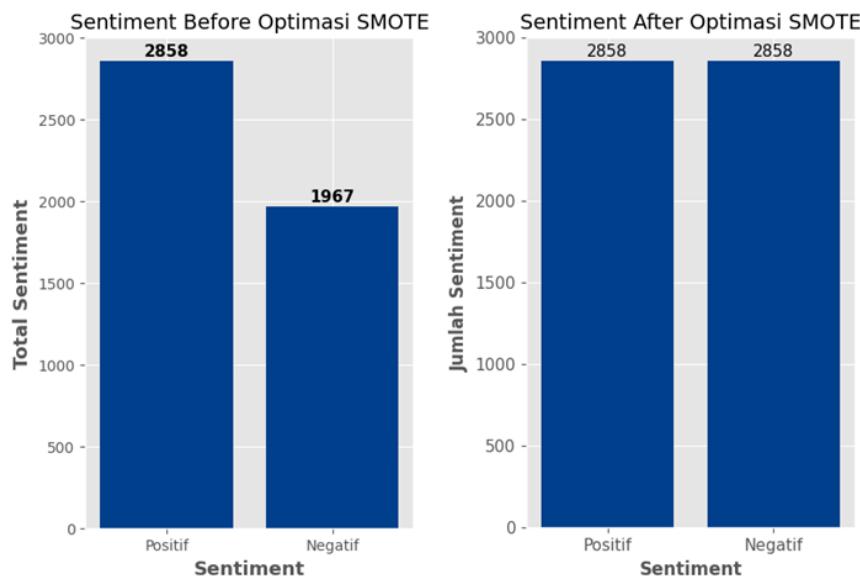
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Pada penelitian ini, data yang dianalisis berjumlah 4825 ulasan user aplikasi Jobstreet yang didapatkan dari platform *Google Play Store*. Pada gambar 2, sebelum optimasi terlihat bahwa jumlah dengan sentimen positif mencapai 2858 dan sentimen negatif hanya berjumlah 1967 data. Kondisi ini menggambarkan adanya ketidakseimbangan data, di mana kelas positif mendominasi dataset yang menyebabkan model klasifikasi menjadi cenderung lebih berpihak pada kelas terbesar sehingga menurunkan kemampuan sistem dalam mendeteksi kelas negatif. Setelah penerapan SMOTE, jumlah sampel pada kedua sentimen menjadi seimbang, yaitu masing-masing 2858 data. Hal ini menunjukkan bahwa SMOTE berhasil menambah sampel sintesis pada kelas minoritas sehingga distribusi data menjadi proporsional. Keseimbangan data dalam proses pelatihan model *Naïve Bayes* dan SVM menjadi lebih optimal karena model tidak kembali cenderung pada salah satu kelas. Hasil SMOTE bisa ditinjau pada Gambar 2.



Gambar 2. Hasil SMOTE

3.2 Evaluasi Perbandingan Model

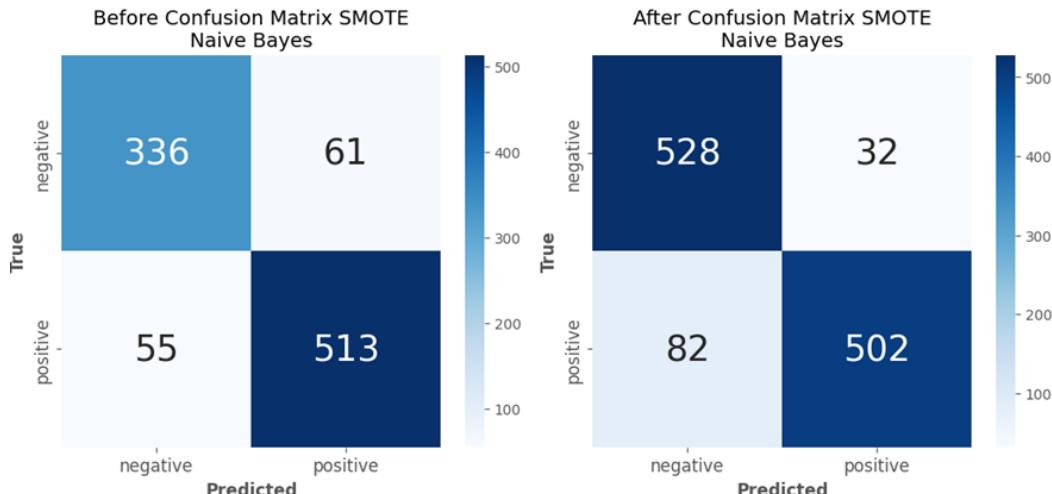
Pada sesi ini, evaluasi dilaksanakan untuk membandingkan kinerja algoritma berdasarkan beberapa metrik, ialah *accuracy*, *precision*, *recall* dan *F1-Score*, pada dua kelas sentimen dalam melakukan klasifikasi sentimen. Pengujian dilakukan menggunakan data uji sebesar 20% dari keseluruhan data, sementara 80% digunakan sebagai data latih.

Tabel 8. Hasil Modelling

Matriks	NB	SVM	NB SMOTE	SVM SMOTE
Accuracy	88%	89%	90%	91%
Sentimen Negatif	Precision	86%	87%	88%
	Recall	85%	87%	93%
	F1-Score	85%	87%	91%
Sentimen Positif	Precision	89%	91%	93%
	Recall	90%	90%	88%
	F1 Score	90%	91%	91%

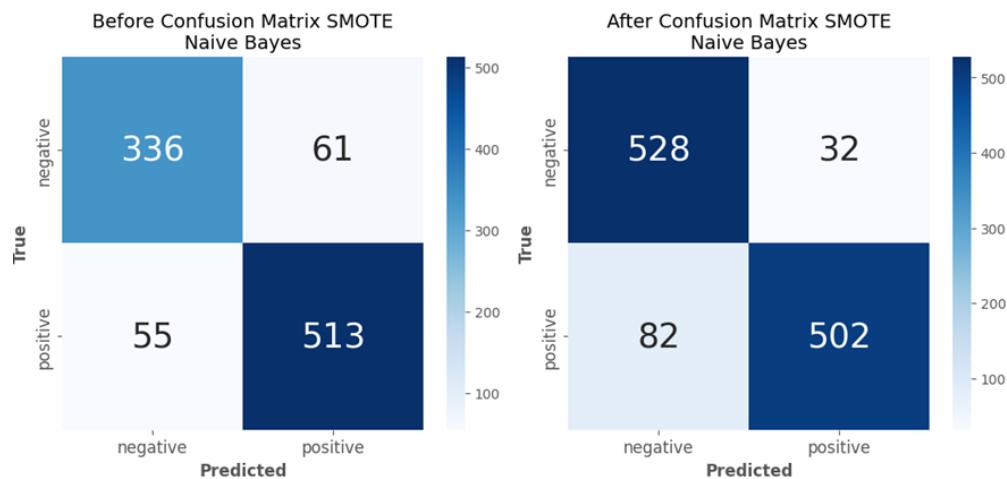
Berdasarkan hasil evaluasi pada tabel 8, SVM memiliki akurasi yang sedikit lebih tinggi dibandingkan *Naïve Bayes*, baik sebelum dan sesudah penerapan SMOTE, yaitu 89% menjadi 91% sedangkan *Naïve Bayes* meningkat dari 88% menjadi 90%. Pada sentimen negatif, kedua algoritma mengalami peningkatan signifikan setelah menggunakan SMOTE, terutama pada nilai *recall* dengan nilai sebesar 85% menjadi 94% pada *Naïve Bayes*, sementara SVM meningkat dari 87% menjadi 93%. Peningkatan *recall* ini menunjukkan kedua algoritma menjadi lebih sensitif dalam mengenali sentimen negatif yang sebelumnya merupakan kelas minoritas. Namun, pada sentimen positif kedua algoritma mengalami sedikit penurunan *recall* setelah SMOTE, yang merupakan dampak dari distribusi data yang telah seimbang, sehingga model tidak kembali cenderung mengarah kelas tertinggi. Walaupun demikian, kedua algoritma ini juga mempertahankan nilai *precision* yang tinggi pada kelas positif. Pada hasil keseluruhan, SVM tampil lebih stabil dan konsisten pada seluruh metrik, sedangkan *Naïve Bayes* menunjukkan peningkatan signifikan dalam mendekripsi sentimen negatif setelah data diseimbangkan.

Gambaran mendalam mengenai kinerja setiap model dalam melakukan klasifikasi, peneliti menganalisis output yang dihasilkan oleh *confusion matrix*. Tujuan penggunaan *Confusion Matrix* adalah untuk menilai kemampuan algoritma dalam membedakan sentimen positif atau negatif secara rinci, serta untuk mengidentifikasi kesalahan yang masih terjadi pada setiap model. Dalam hasil ini, peneliti dapat menentukan perbandingan algoritma *Naïve Bayes* dan *Support Vector Machine* pada ulasan pengguna Jobstreet.

**Gambar 3.** Hasil SMOTE *Naïve Bayes*

Menurut ilustrasi gambar 3, *Confusion matrix* sebelum SMOTE menunjukkan bahwa model *Naïve Bayes* berhasil mengenali kelas positif cukup baik, terlihat dari jumlah *True Positive* (513) yang tinggi. Namun, model masih melakukan kesalahan yang cukup besar pada kelas negatif, ditunjukkan oleh nilai *False Positive* (61) dan *False Negative* (55). Setelah SMOTE di terapkan, nilai TN meningkat signifikan menjadi 528 dan FP menurun menjadi 32, yang menunjukkan peningkatan kemampuan dalam mengenali kelas negatif, meskipun FN meningkat dari 55 menjadi 82, nilai TP tetap tinggi yaitu 502 sehingga model masih cukup baik dalam mengidentifikasi ulasan positif.

Berdasarkan ilustrasi pada gambar 4, model SVM sebelum SMOTE diperoleh TP sebesar 512, TN sebesar 346, FP sebanyak 51, dan FN sebesar 56. Hasil ini memperlihatkan bahwa SVM sudah cenderung lebih seimbang dibanding *Naïve Bayes* dalam mengenali kedua kelas. Setelah SMOTE, performa SVM meningkat terutama pada kelas negatif, ditunjukkan oleh kenaikan TN menjadi 522 dan FP menurun menjadi 38, sementara TP tetap tinggi sebesar 516. Secara keseluruhan, model *SVM* menunjukkan konsistensi yang lebih baik, karena peningkatannya lebih merata pada kedua kelas dan tidak menimbulkan kenaikan FN yang sebesar pada *Naïve Bayes*. Hasil ini menunjukkan bahwa SVM dapat dikatakan lebih stabil dibandingkan model *Naïve Bayes*.

**Gambar 4.** Hasil SMOTE SVM

3.3 Visualisasi Data

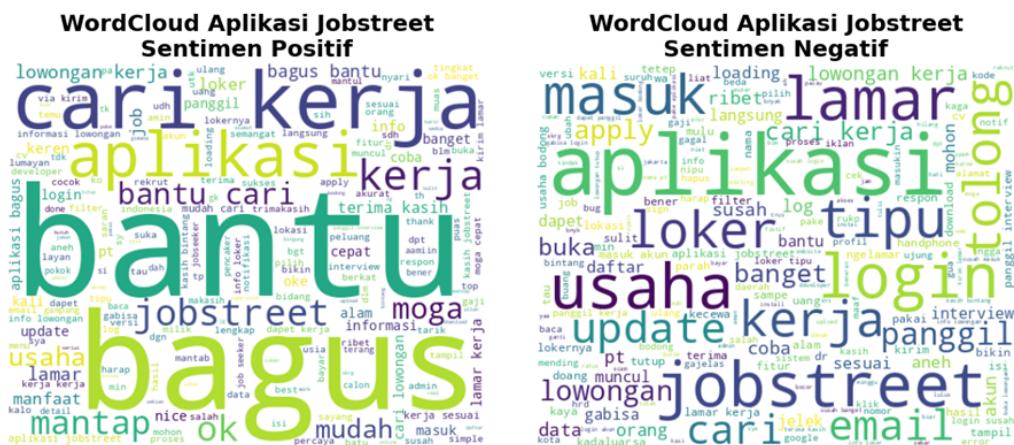
Dalam visualisasi data, diterapkan analisis terhadap frekuensi kemunculan kata untuk melihat pola umum dalam ulasan pengguna pengguna jobstreet. Gambar berikut menampilkan sepuluh kata dengan kemunculan tertinggi setelah proses pembersihan data dilakukan.

**Gambar 5.** Frekuensi Kata

Berdasarkan gambar 5, grafik frekuensi kata diatas menunjukkan kata yang paling sering muncul pada ulasan pengguna Jobstreet. Kata “kerja” menjadi yang paling dominan dengan 1223 kemunculan, disusul “aplikasi” dan “bagus”. Kemunculan kata-kata tersebut menunjukkan bahwa ulasan pengguna banyak berfokus pada penggunaan aplikasi dan proses pencarian kerja. *Wordcloud* adalah gambar teks yang menampilkan kumpulan kata berdasarkan tingkat kemunculan [20]. Kata yang berukuran besar adalah kata yang seringkali muncul, ini akan memudahkan pembaca untuk mengidentifikasi beberapa tema yang muncul dalam dataset. Hasil *wordcloud* kedua sentimen bisa diperlihatkan pada gambar 6.

Pada *wordcloud* sentimen positif menunjukkan bahwa pengguna menilai aplikasi jobstreet sangat membantu dalam mencari kerja. terlihat dari dominasi kata “bantu”, “bagus”, “kerja”, “cari”, serta dungungan kata “mudah” dan “mantap” yang mencerminkan kemudahan dan kepuasan penggunaan aplikasi. Sementara itu pada *wordcloud* sentimen negatif menunjukkan bahwa banyak pengguna mengeluhkan aplikasi yang sulit “login”, membutuhkan *update* untuk memperbaiki *bug*, sering meminta “tolong” karena kendala penggunaan, serta muncul kekhawatiran terkait

validitas lowongan yang dianggap tipu atau sudah kedaluwarsa. *Wordcloud* ini mampu menyajikan representasi visual terkait beberapa tema yang dibahas oleh pengguna Jobstreet.



Gambar 6. Hasil *Wordcloud*

4. KESIMPULAN

Menurut hasil temuan dan pengujian yang sudah dijalankan dalam perbandingan algoritma *Naïve Bayes* dan SVM untuk analisis sentimen pengguna Jobstreet, diperoleh bahwa SVM menunjukkan kinerja yang lebih stabil dan konsisten dibandingkan *Naïve bayes*. Hal ini terlihat dari akurasi SVM yang lebih tinggi baik sebelum maupun sesudah penerapan SMOTE, yaitu meningkat dari 89% menjadi 90%, sementara *Naïve Bayes* meningkat dari 88% menjadi 90%. Penerapan SMOTE terbukti efektif dalam mengoptimalkan kemampuan kedua model untuk mendeteksi sentimen negatif yang sebelumnya merupakan kelas minoritas. *Naïve Bayes* mengalami peningkatan *recall* signifikan dari 85% menjadi 94%, sedangkan SVM meningkat dari 87% menjadi 93%. Namun, pada sentimen positif, kedua algoritma mengalami sedikit penurunan *recall* setelah SMOTE karena distribusi data yang telah seimbang sehingga model tidak kembali terlalu fokus pada kelas mayoritas. Meski demikian, kedua algoritma tetap mempertahankan nilai *precision* yang tinggi pada kelas positif. Secara keseluruhan, bisa dikatakan bahwa SVM algoritma yang lebih unggul untuk analisis sentimen pengguna Jobstreet, karena memberikan performa yang lebih stabil pada seluruh metrik. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan dua algoritma klasifikasi teks, yaitu *Naïve Bayes* dan *Support Vector Machine*. Oleh karena itu, studi berikutnya disarankan untuk mengevaluasi lebih banyak algoritma klasifikasi teks untuk memperoleh hasil *accuracy*, *precision*, *recall*, dan *F1-Score* yang lebih optimal.

DAFTAR PUSTAKA

- [1] H. Margahana, “Business Development in the Digital Age,” *Int. J. Adv. Sci. Comput. Appl.*, vol. 3, no. 2, pp. 2787–2792, 2024, doi: 10.47679/ijasca.v3i2.55.
- [2] K. et al 2023, “Manajemen Profil Pekerja di Website Jobstreet Sebagai Kesempatan Karier di Kalangan Gen Z,” vol. 4, no. 9, pp. 167–186, 2021.
- [3] R. D. Wahyuni and A. N. Utomo, “Penggunaan Metode Lexicon Untuk Analisis Sentimen pada Ulasan Aplikasi KAI Access di Google Play Store,” *J. Rekayasa Inf.*, vol. 11, no. 2, pp. 134–145, 2022.
- [4] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [5] P. Arsi and R. Waluyo, “Sentiment Analysis of Discourse on Moving the Indonesian Capital City Using the Support Vector Machine (SVM) Algorithm,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.202183944.
- [6] Y. Yunitasari and A. R. Putera, “Analisis Sentimen Masyarakat di Twitter Terkait Pandemi Covid-19,” *Smatika J.*, vol. 11, no. 01, pp. 22–26, 2021, doi: 10.32664/smatika.v11i01.520.
- [7] A. Setiawan and R. R. Suryono, “Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.
- [8] N. Nurzaman, N. Suarna, and W. Prihartono, “Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.
- [9] S. Syafrizal, M. Afdal, and R. Novita, “Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
- [10] A. Puji Astuti, S. Alam, and I. Jaelani, “Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRIMo,” *J. Bangkit Indones.*, vol. 11, no. 2, pp. 1–6, 2022, doi: 10.52771/bangkitindonesia.v11i2.196.

- [11] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 8, no. 2, p. 77, 2023, doi: 10.31328/jointecs.v8i2.4747.
- [12] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [13] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 544, 2021, doi: 10.32493/informatika.v5i4.6760.
- [14] K. D. Indarwati and H. Febraryanti, "Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naïve Bayes Classifier," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, 2023, doi: 10.35957/jatisi.v10i1.2643.
- [15] A. B. Muzayyanah, R. E. Pawening, and Z. Arifin, "Analisis Sentimen Pada Ulasan Aplikasi Ehadrah Di Google Playstore Menggunakan Support Vector Machine (Svm)," *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 258–266, 2024, doi: 10.36080/idealis.v7i2.3250.
- [16] L. Nursinggah, R. Ruuhwan, and T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4336.
- [17] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 534–539, 2022, doi: 10.14569/IJACSA.2022.0130665.
- [18] C. Candra, K. W. Chandra, and H. Irsyad, "Efektifitas SMOTE dalam Mengatasi Imbalanced Class Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 31–42, 2024, doi: 10.54082/jiki.132.
- [19] T. Safitri, Y. Umaidah, and I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 28–35, 2023, doi: 10.30871/jaic.v7i1.5039.
- [20] D. P. Santoso and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store," *J. Sains dan Seni ITS*, vol. 11, no. 2, 2022, doi: 10.12962/j23373520.v11i2.72534.