

Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet

Implementation of Naive Bayes Algorithm For Sentiment Analysis Of The Use Of Jobstreet Application

Bobby Kurniadi Widodo¹, Nur Hafifah Matondang², Desta Sandya Prasvita³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta
E-mail: ¹bobbyk@upnvj.ac.id, ²nurhafifahmatondang@upnvj.ac.id, ³desta.sandya@upnvj.ac.id

Abstrak

Aplikasi Jobstreet merupakan sebuah aplikasi lowongan pekerjaan yang sudah *download* oleh lebih dari 10 juta masyarakat yang menyediakan beberapa jenis pekerjaan seperti akuntansi, sumber daya manusia, pemasaran, komunikasi, pelayanan, dan lainnya. Dengan banyaknya masyarakat yang *download* aplikasi ini maka masyarakat pasti memberikan ulasan-ulasan mereka terhadap aplikasi ini. Di masa pandemi seperti ini, banyak orang yang mencari pekerjaan menggunakan aplikasi android dimana informasinya lebih cepat dan mudah untuk mencari lowongan pekerjaan, oleh karena itu aplikasi Jobstreet membantu masyarakat dalam mencari lowongan pekerjaan di perusahaan yang mereka inginkan. Ulasan komentar opini masyarakat ini bisa dijadikan peluang untuk menggali keterangan tentang evaluasi dan penilaian atas pelayanan aplikasi jobstreet yang telah berjalan menggunakan analisis sentimen. Tujuan dari penelitian ini adalah melakukan klasifikasi sentimen terhadap ulasan pada aplikasi Jobstreet dengan metode *Naive Bayes*. Dalam penelitian ini opini akan dibagi kedalam dua golongan sebagai positif dan negatif, kemudian diklasifikasikan dengan menggunakan algoritma *Naive Bayes*. Hasil pengujian yang didapat menggunakan data uji memiliki nilai akurasi sebesar 0,96; nilai precision sebesar 0,98; nilai recall sebesar 0,94.

Kata kunci: Analisis Sentimen, Jobstreet, *Naive Bayes*, Klasifikasi

Abstract

The Jobstreet application is a job vacancy application that has been downloaded by more than 10 million people which provides several types of jobs such as accounting, human resources, marketing, communication, services, and others. With so many people downloading this application, people will definitely give their reviews of this application. In times of a pandemic like this, many people are looking for work using android applications where the information is faster and easier to find job vacancies, therefore the Jobstreet application helps people find job vacancies in the companies they want. This review of public opinion comments can be used as an opportunity to dig up information about the evaluation and assessment of jobstreet application services that have been running using sentiment analysis. The purpose of this study is to classify the sentiment of reviews on the Jobstreet application using the Naive Bayes method. In this study, opinions will be divided into two groups as positive and negative, then classified using the Naive Bayes algorithm. The test results obtained using test data have an accuracy value of 0.96; precision value is 0.98; recall value of 0.94.

Keywords: Sentiment Analysis, Jobstreet, *Naive Bayes*, Classification

1. PENDAHULUAN

Perkembangan penggunaan *smartphone* pada indonesia waktu ini dirasakan sangat cepat. Penduduk indonesia yang berjumlah 250 juta ialah pasar yang sangat besar bagi pertumbuhan

pesat pengguna *smartphone*. Penggunaan aplikasi seluler internet untuk mendapatkan informasi ini masih tidak biasa hari ini. Terbukti dengan semakin kuatnya persaingan di sektor *mobile*, Android merupakan teknologi yang saat ini sedang menggebrak pasar. Ini ada hubungannya dengan sistem operasi dan aplikasi yang didukungnya. Sangat penting bagi banyak pemangku kepentingan, terutama pencari kerja, untuk menyadari pentingnya memposting lowongan pekerjaan.

Aplikasi Jobstreet adalah aplikasi lowongan pekerjaan dengan lebih dari 10 juta unduhan yang menawarkan berbagai posisi di berbagai bidang seperti akuntansi, sumber daya manusia, pemasaran, komunikasi, layanan, dan banyak lagi. Di masa pandemi seperti ini, banyak orang yang mencari pekerjaan menggunakan aplikasi android dimana informasinya lebih cepat dan mudah untuk mencari lowongan pekerjaan, oleh karena itu aplikasi Jobstreet membantu masyarakat dalam mencari lowongan pekerjaan di perusahaan yang mereka inginkan. Perusahaan berperan sebagai perantara dalam pemilihan lowongan kerja dan pertukaran informasi antara pelamar kerja dengan perusahaan di negara tetangga. Visi JobStreet adalah menggabungkan bisnis dan bakat serta meningkatkan standar hidup melalui karir yang lebih baik[1].

Dengan begitu banyak orang yang memasang aplikasi ini, orang pasti akan meninggalkan ulasan tentangnya, tentang aplikasi mereka sendiri, tentang keadaan bisnis saat ini, dan sebagainya. Pemeriksaan komentar opini publik ini dapat dimanfaatkan untuk menggali informasi mengenai evaluasi dan penilaian layanan aplikasi jobstreet yang telah menerapkan analisis sentimen.

Analisis sentimen atau disebut juga *opinion mining* adalah proses pengolahan data teks secara otomatis, juga dikenal sebagai analisis sentimen atau mining opinion analysis, untuk memahami informasi tentang opini atau tanggapan[2]. Analisis sentimen menandai opini individu secara positif atau negatif[3].

Text mining merupakan salah satu teknik yang dapat digunakan untuk klasifikasi, di antaranya text mining merupakan salah satu varian dari *data mining*, yang bertujuan untuk menemukan pola-pola menarik dari sejumlah besar data teks[4]. Melalui text mining, informasi yang akan diekstraksi jelas dan jelas di dalam teks[5].

Berdasarkan pada penelitian Sepri[6], sentimen analisis masyarakat sehingga didapatkan tingkat akurasi menggunakan algoritma *Naïve Bayes*. Penelitian ini membagi opini yang didapatkan kedalam dua jenis, yaitu positif dan negatif. Dimana hasil yang didapatkan hasil akurasi sebesar 0,87 atau (87 %). Berdasarkan pada penelitian lainnya yang dilakukan oleh Daryfayi[7], mengklasifikasi teks ulasan untuk mendapat sentimen yang diberikan dari ulasan tersebut dan menganalisis performa sistem klasifikasi sentimen dengan metode MNB yang hanya menggunakan *library* Sastrawi dan NLTK. Hasil yang didapatkan dari penelitian yang telah dilakukan dengan metode *Naïve Bayes* akurasi tertinggi 78,9% untuk kasus klasifikasi ulasan berdasarkan teks

Berdasarkan hal tersebut, penelitian ini menganalisis masalah yang terkait dengan beberapa ulasan aplikasi Jobstreet di *Google Play Store* dan menentukan akurasi analisis sentimen yang dihasilkan oleh algoritma *Naïve Bayes*. Dalam algoritma ini diimplementasikan *Teorema Bayes* yang menilai bahwa setiap variabel yang bersifat berdikari atau independen untuk mempertimbangkan nilai variabel kelas[8].

Tujuan penelitian ini adalah untuk menganalisa melakukan pembangunan model sentimen terhadap ulasan pada aplikasi Jobstreet dengan metode *Naïve Bayes* dan performa *Naïve Bayes* saat mengklasifikasikan data sentimen pengguna di Jobstreet dengan meningkatkan *text preprocessing* menggunakan kamus tambahan.

2.METODE PENELITIAN

2.1 Pengumpulan Data

Bagian pertama dari penelitian ini adalah pengumpulan data, yang dilakukan dengan menggunakan 1000 data scraping. Prosedur scraping dilakukan melalui paket *Google Play*

Scraper, yang ditawarkan oleh bahasa pemrograman Python. Operasi scraping dilakukan secara otomatis untuk mendapatkan data review untuk aplikasi Jobstreet, yang meliputi komentar Google Play Store. Hasil akhir dari prosedur ini adalah kumpulan data berformat CSV.

2.2 Proses Pelabelan Sentimen pada Ulasan

Proses pelabelan adalah mengidentifikasi ulasan yang telah dikumpulkan secara manual setiap ulasan hanya memiliki satu nilai, yang mungkin positif atau negatif. Untuk menjaga substansi informasi review, pelabelan dilakukan sebelum tahap preprocessing. Sebagai konsekuensi dari langkah ini, semua ulasan hanya dibagi menjadi dua kategori ulasan dengan nilai positif dan ulasan dengan nilai negatif. Jika ditemukan data yang bukan milik keduanya (netral), maka akan ditambahkan ke kelas positif[9].

Dikarenakan pelabelan dilakukan dengan lebih dari 2 bantuan manusia, diyakini bahwa kesalahan akan terjadi, maka metode yang dikenal sebagai *leiss kappa value* digunakan untuk memeriksa reliabilitas antar anotator. Rumus untuk menghitung nilai dari persamaan *kappa value* adalah sebagai berikut :

$$kappa\ value = \frac{p_{\alpha} - p_{\epsilon}}{1 - p_{\epsilon}} \quad (1)$$

Keterangan:

p_{α} : Persentase jumlah pengukuran antar *annotator*

p_{ϵ} : Persentase jumlah perubahan antar *annotator*

Hasil perhitungan kappa akan dijadikan sebagai hasil kesepakatan pada setiap rater yang akan digunakan sebagai tolak ukur kesepakatan label data dalam penelitian yang dilakukan oleh Nichols[10]. Berikut merupakan tabel interpretasi kappa:

Tabel 1 Tabel Skala Nilai Kappa

<i>Indeks Kappa</i>	<i>kesepakatan</i>
1.00 – 0.81	Kesepakatan yang hampir sempurna
0.80 – 0.61	Perjanjian substansial
0.60 – 0.41	Perjanjian moderat
0.40 – 0.21	Perjanjian yang adil
0.20 – 0.00	Sedikit kesepakatan
<0	Perjanjian yang buruk atau tidak ada kesepakatan

2.3 Text Preprocessing

Tahap terpenting dalam penambahan data adalah *pre-processing*, yaitu proses mengubah data mentah menjadi data yang kompatibel dengan prosedur penambahan yang akan digunakan[11]. *Pre-processing* membuat indeks kumpulan dokumen selama proses klasifikasi dokumen. Indeks adalah kumpulan istilah yang menunjukkan konten atau topik dokumen[12]. Ekstraksi istilah biasanya terdiri dari beberapa operasi besar, termasuk:

a. Case Folding

Case folding ialah proses konversi seluruh karakter/case dalam sebuah dokumen ke format baku. Ini membantu untuk membakukan karakter karakter pada teks pada format terpadu. Ubah ke huruf kecil kata-kata yang ditulis menggunakan huruf besar serta kata-kata yang ditulis menggunakan huruf kecil diklaim tidak ada bedanya, sehingga semua huruf besar biasanya diubah menjadi huruf kecil sebelum klasifikasi dilakukan.

Tabel 2 Data Hasil Proses Case Folding

Text Input	Text Output
Aku kasih bintang 5. Aplikasi yang bermanfaat banget, ngebantu aku cari kerjaan yang sesuai sama passion ditambah lagi info dari tiap2 perusahaannya detail banget. 👍	aku kasih bintang 5. aplikasi yang bermanfaat banget, ngebantu aku cari kerjaan yang sesuai sama passion ditambah lagi info dari tiap2 perusahaannya detail banget. 👍

b. Cleaning

Cleaning adalah langkah menghilangkan kata-kata yang tidak perlu dari teks untuk mengoptimalkan dan mengurangi munculnya noise pada fase klasifikasi, seperti emoticon, tanda baca (!/?.), dan lainnya.

Tabel 3 Data Hasil Proses Data Cleaning

Text input	Text output
aku kasih bintang 5. aplikasi yang bermanfaat banget, ngebantu aku cari kerjaan yang sesuai sama passion ditambah lagi info dari tiap2 perusahaannya detail banget. 👍	aku kasih bintang aplikasi yang bermanfaat banget ngebantu aku cari kerjaan yang sesuai sama passion ditambah lagi info dari tiap perusahaannya detail banget

c. Normalisasi Kata

. Normalisasi data adalah mengubah ulasan / komentar yang mungkin ada yang typo atau singkatan – singkatan, komentar yang diberikan seseorang tidak semuanya menggunakan bahasa baku akan diubah sesuai kamus normalisasi. Untuk normalisasi ini menggunakan bantuan kamus yang telah dibuat oleh peneliti dan beberapa orang lainnya.

Tabel 4 Data Hasil Proses Normalisasi Data

Text input	Text output
aku kasih bintang aplikasi yang bermanfaat banget ngebantu aku cari kerjaan yang sesuai sama passion ditambah lagi info dari tiap perusahaannya detail banget	kasih bintang aplikasi yang bermanfaat sangat bantu cari kerjaan yang sesuai sama passion ditambah lagi informasi dari tiap perusahaannya detail sangat

d. Stopword Removal

Stopword Removal merupakan langkah memilih hanya kata-kata penting dan menghapus kata-kata yang tidak secara signifikan mempengaruhi isi dokumen, seperti kata-kata, “dan, adalah, di, yang”.

Tabel 5 Data Hasil Proses Stopword Removal

Text input	Text output
kasih bintang aplikasi yang bermanfaat sangat bantu cari kerjaan yang sesuai sama passion ditambah lagi informasi dari tiap perusahaannya detail sangat	kasih bintang aplikasi bermanfaat bantu cari kerjaan sesuai passion ditambah informasi perusahaannya detail

e. Stemming

Stemming merupakan Proses sangat penting dalam memperkecil indeks yang terdapat dalam kalimat utuh, sehingga karakter kata yang memiliki makna yang sama akan dihitung sebagai kelompok kata yang sama contohnya kata memakan, makanan, dimakan, dan termakan akan dikembalikan ke dasar kata utamanya yaitu “makan”.

Tabel 6 Data Hasil Proses Stemming

Text Input	Text Output
kasih bintang aplikasi bermanfaat bantu cari kerjaan sesuai passion ditambah informasi perusahaannya detail	kasih bintang aplikasi manfaat bantu cari kerja sesuai passion tambah informasi usaha detail

f. *Tokenizing*

Tokenizing adalah proses membagi korpus teks menjadi kalimat yang berfungsi sebagai token tingkat pertama dalam korpus, juga dikenal sebagai segmentasi kalimat karena mengatur teks menjadi frasa yang koheren. Strategi dasar melibatkan pencarian pembatas antara kalimat, seperti titik (.) atau karakter baris baru (/n), dan kadang-kadang bahkan titik koma (;).

Tabel 7 Data Hasil Proses Tokenizing

Text Input	Text Output
kasih bintang aplikasi manfaat bantu cari kerja sesuai passion tambah informasi usaha detail	['kasih', 'bintang', 'aplikasi', 'manfaat', 'bantu', 'cari', 'kerja', 'sesuai', 'passion', 'tambah', 'informasi', 'usaha', 'detail']

2.4 *TF-IDF*

Untuk mengubah data menjadi angka menggunakan metode pembobotan TF-IDF, data yang telah diproses sebelumnya harus dalam format numerik[13]. Dengan menetapkan bobot pada setiap kata, pendekatan ini menentukan seberapa dekat frasa tersebut terkait dengan dokumen. *Term-weighting* dilakukan untuk memberikan bobot kata pada tiap dokumen atau ulasan supaya mempunyai nilai yang mampu membantu proses klasifikasi menggunakan perhitungan yang terdapat pada rumus.

$$TF = \begin{cases} 1 + \log_{10}(tf_{t,d}), & \text{if } tf_{t,d} > 0 \\ 0, & \text{if } tf_{t,d} = 0 \end{cases} \quad (1)$$

Perumusan IDF

$$IDF = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Perkalian dari perhitungan TF dan IDF dapat menghasilkan bobot kata yang disebut TF-IDF, yaitu dengan perhitungan berikut:

$$W_{t,d} = TF \times IDF \quad (3)$$

2.5 *Naïve Bayes Classifier*

Salah satu algoritma yang dipergunakan dalam perhitungan kombinasi nilai dan frekuensi data pada dataset sederhana dikenal sebagai algoritma *Naïve Bayes*[8]. Perhitungan bersyarat serta asumsi independensi serau berlaku pada aplikasi dunia nyata, sehingga ditandai sebagai *Naïve*, namun algoritma condong belajar terampil dalam berbagai persoalan klasifikasi terkontrol. *Naïve bayes* dapat diformulasikan menjadi persamaan.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipótesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipótesis H berdasar kondisi X (posteriori probabilitas)

- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Algoritma *Multinomial Naïve Bayes* digunakan dalam penelitian ini karena memiliki akurasi yang tinggi dan mudah untuk dihitung[14]. Dimana dapat di formulasikan menjadi persamaan.

$$C_{MAP} = \underset{c \in V}{arg \max} P(p) \prod_{t=i}^{|V|} P(W_i|p) \tag{5}$$

Keterangan :

$P(W_i|p)$: Peluang kemunculan W_i pada kelas p

$P(p)$: Peluang kemunculan dokumen yang berada pada kelas p

1.6 SMOTE (Synthetic Minority Over-Sampling Technique)

Oversampling dengan SMOTE (Synthetic Minority Over-Sampling Technique) digunakan untuk mengatasi ketidakseimbangan data. Teknik SMOTE membangkitkan data palsu untuk menambah jumlah data kelas minor (kelas negatif) hingga sama dengan jumlah data kelas mayor (kelas positif)[15].

3.HASIL DAN PEMBAHASAN

3.1 Pelabelan Data

Pelabelan teks atau ulasan memerlukan bantuan dari spesialis multibahasa yang secara otomatis dapat mengembangkan struktur klasifikasi teks. Ini dapat digunakan untuk membandingkan analisis sistem dengan analisis pakar. Namun dalam penelitian ini, prosedur kategorisasi manual dilakukan oleh tiga responden yang semuanya berlatar belakang informatika. Prosedur kategorisasi membagi kumpulan data hingga 1000 ulasan menjadi dua kategori positif dan negatif. Kelas yang paling dominan akan dikategorikan ke dalam kelas yang paling dominan. Jika ditemukan hasil yang imbalanced, maka akan dilakukan teknik sampling SMOTE sehingga data akan balanced dengan harapan nilai akan lebih baik.

Tabel 8 Contoh Pelabelan Data Ulasan

Ulasan	Anatator 1	Anatator 2	Anatator 3	Hasil Akhir
Kenapa tidak bisa upload CV ? Sudah saya coba di beberapa device tetap gak bisa upload.	0	1	1	1
Bintang 4 dulu kalo misal memang efektif dalam membantu dan terpercaya sehingga bisa bekerja melalui aplikasi Job Street ini saya akan kasih tambah bintang menjadi bintang 5 🙌	1	0	0	0

Sangat penting untuk menentukan apakah pelabelan konsisten setelah setiap individu memberikan label untuk setiap bagian dari data yang ada. Akibatnya, nilai kappa diperlukan untuk menilai konsistensi label; nilai kappa dapat dihitung dengan menggunakan persamaan (9) sebagai berikut :

$$kappa \ value = \frac{p_{\alpha} - p_{\epsilon}}{1 - p_{\epsilon}}$$

$$kappa \ value = \frac{p_{\alpha} - p_{\epsilon}}{1 - p_{\epsilon}}$$

$$\begin{aligned} \text{kappa value} &= \frac{0,989333333 - 0,724003556}{1 - 0,724003556} \\ &= \frac{0,26533}{0,275996} = 0,961352159 \end{aligned}$$

Nichols mengemukakan skala untuk menginterpretasi nilai kappa[10]. Skala nilai kappa dapat dilihat pada Tabel 3.1 yang terdapat pada bab 3. Berdasarkan interpretasi dari Tabel 3.2 tersebut, maka nilai kappa sebesar 0,961352159 pada data ulasan aplikasi Jobstreet berada pada kategori Kesepakatan yang hampir sempurna yang berarti konsistensi dari label sangat sempurna..

3.2 Pembagian Data

Setelah semua kata sudah selanjutnya adalah melakukan pembagian data label menjadi data latih dan data uji, data yang akan dibagi kedalam data latih dan data uji dengan perbandingan 80:20. Pembagian data uji dan data latih dapat dilihat pada Tabel 4.12 sebagai berikut :

Tabel 9 Pembagian Data

	Label Positif	Label Negatif	Total Jumlah
Data Latih	665	135	800
Data Uji	167	33	200
Total	832	168	1000

3.3 Pengujian Naïve Bayes

Setelah dilakukan pemodelan dengan metode Multinomial Naïve Bayes, maka selanjutnya melakukan proses evaluasi dengan menggunakan confusion matrix. Tujuan dilakukannya evaluasi untuk mengukur kinerja klasifikasi yang dilakukan. Berikut adalah tabel confusion matrix hasil klasifikasi menggunakan model:

Tabel 10 Confusion Matrix Hasil Evaluasi Naïve Bayes

		True Values	
		Positive	Negative
Prediction	Positive	165	25
	Negative	2	8

Hasil penghitungan nilai akurasi (*accuracy*), *precision*, *recall*, *specificity* dideskripsikan sebagai berikut.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{165 + 8}{165 + 8 + 25 + 2} = \frac{173}{200} = 0,865 = 0,86$$

$$\text{precision} = \frac{TP}{TP + FP} = \frac{165}{165 + 25} = \frac{165}{190} = 0,868 = 0,87$$

$$\text{recall} = \frac{TP}{TP + FN} = \frac{165}{165 + 2} = \frac{165}{167} = 0,988 = 0,99$$

$$\text{specificity} = \frac{TN}{TN + FP} = \frac{8}{8 + 25} = \frac{8}{33} = 0,242 = 0,24$$

Akurasi yang dihasilkan adalah 0,86, seperti yang dapat dilihat. Dari 200 data tes, 165 diklasifikasikan dengan benar sebagai positif, dan 8 diklasifikasikan dengan benar sebagai negatif. Sementara itu, dua titik data positif salah diklasifikasikan sebagai positif, dan 25 titik data negatif salah diklasifikasikan sebagai negatif.

3.4 Pengujian Naïve Bayes Dengan Oversampling SMOTE

Penerapan metode SMOTE pada data latih dideskripsikan pada Tabel 11.

Tabel 11 Penerapan Metode SMOTE

	Positif	Negatif	Total
Data Latih	665	665	1330

Dari Tabel 4.18, kedua kelas telah seimbang, dimana masing-masing kelas pada data latih berjumlah 665 data. Metode SMOTE menambah jumlah kelas minor (negatif) sebanyak 530 data dari sebelumnya hanya berjumlah 135 data, sehingga jumlah kelas negatif sama dengan jumlah kelas positif pada data latih.

Setelah dilakukan penyeimbangan data, maka model dibuat kembali dan dievaluasi untuk mengetahui apakah kinerja model yang dibuat lebih baik dari sebelumnya. Hasil evaluasi model dengan *oversampling* SMOTE dideskripsikan pada Tabel 12.

Tabel 12 Confusion Matrix Hasil Evaluasi Dengan SMOTE

		True Values	
		Positive	Negative
Prediction	Positive	155	12
	Negative	9	24

Untuk mengukur nilai akurasi, *Precision*, dan *recall* dari model klasifikasi yang telah dilakukan *oversampling* SMOTE. Penghitungan masing-masing nilai tersebut dideskripsikan pada persamaan (5), (6), (7), dan (8). Hasil penghitungan nilai akurasi (*accuracy*), *precision*, *recall*, *specificity* dideskripsikan sebagai berikut.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{155 + 24}{155 + 24 + 9 + 12} = \frac{179}{200} = 0,895 = 0,90$$

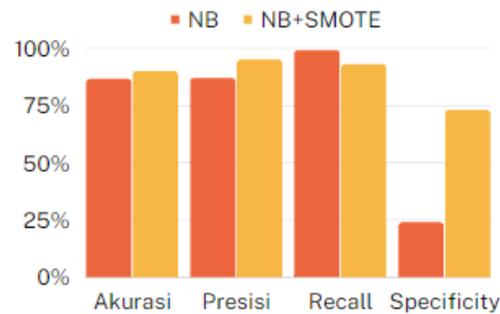
$$precision = \frac{TP}{TP + FP} = \frac{155}{155 + 9} = \frac{155}{164} = 0,945 = 0,95$$

$$recall = \frac{TP}{TP + FN} = \frac{155}{155 + 12} = \frac{155}{167} = 0,928 = 0,93$$

$$specificity = \frac{TN}{TN + FP} = \frac{24}{24 + 9} = \frac{24}{33} = 0,727 = 0,73$$

3.5 Pembahasan

Dibawah ini merupakan gambar grafik perbandingan hasil *accuracy*, *precision*, *recall* dari metode Naïve Bayes tanpa *Oversampling* SMOTE dan metode Naïve Bayes dengan *Oversampling* Smote.



Gambar 1 Grafik Perbandingan Naïve Bayes dan Naïve Bayes dengan SMOTE

Dari hasil evaluasi yang telah dilakukan, nilai akurasi (*accuracy*) mengalami peningkatan sebelumnya yaitu sebesar 0,86 menjadi 0,90. Nilai *precision* terjadi peningkatan dari 0,87 menjadi 0,95. Namun nilai *recall* menurun menjadi 0,93 dari sebelumnya sebesar 0,99 dan menaikkan nilai *specificity* sebesar 0,73 dari sebelumnya sebesar 0,24. Dan dari penelitian terdahulu terdapat kenaikan yang awalnya 0,789 menjadi 0,93 karena penambahan sebuah text preprocessing yaitu Normalisasi Data yang mendasar pada kamus yang dibuat oleh peneliti.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan temuan analisis sentimen aplikasi Jobstreet di Play Store menggunakan metode Naïve Bayes (NB), dapat diambil beberapa kesimpulan:

1. Proses pengklasifikasian sentimen positif dan negatif pada Playstore Review aplikasi Jobstreet meliputi pengumpulan dan pelabelan data dengan tiga anotasi, pembersihan data dari kebisingan, pembobotan kata dengan TF-IDF, dan memisahkan data ke dalam data pelatihan dan pengujian untuk diklasifikasikan lebih lanjut berdasarkan kelas menggunakan algoritma Naïve Bayes. Setiap objek/kata dalam ulasan diberi probabilitas berdasarkan kelas yang disediakan, yang kemudian dikalikan dengan probabilitas jumlah dokumen di kelas tersebut dibagi dengan jumlah total dokumen. Dokumen ini diklasifikasikan sebagai positif jika hasil perkalian menunjukkan bahwa peluang kelas positif pada dokumen lebih besar daripada peluang kelas negatif, dan sebaliknya.
2. Nilai akurasi, presisi, recall dan specificity yang diperoleh dari klasifikasi sentimen aplikasi Jobstreet menggunakan metode Naive Bayes (NB) dengan data pelatihan dan pengujian 80% dan 20% dan oversampling SMOTE masing-masing sebesar 0,90, 0,95, 0,93, 0,73.
3. Berdasarkan peneliti terdahulu akurasi mengalami kenaikan yang awalnya 0,789 menjadi 0,93 karena penambahan sebuah text preprocessing yaitu Normalisasi Data yang mendasar pada kamus yang dibuat oleh peneliti.
4. Berdasarkan data review, sentimen terhadap aplikasi Jobstreet mendapat respons positif 83,2% dibandingkan 16,8% respon negatif.

1.2 Saran

Saran yang dapat penulis berikan untuk perbaikan penelitian ini di masa yang akan datang, yaitu:

1. Penelitian selanjutnya diharapkan dapat menggunakan bentuk tambahan Naïve Bayes (NB) untuk membandingkan kinerja jenis Naïve Bayes (NB) yang berbeda ataupun jenis algoritma lainnya seperti SVM, KNN, Random Forest dan lainnya yang digunakan pada penelitian analisis sentimen Aplikasi Jobstreet ini, selain itu bisa juga menerapkan seleksi fitur.
2. Penelitian selanjutnya diharapkan dapat membandingkan Algoritma Naïve Bayes (NB)

dengan menggunakan data aplikasi pencari kerja lainnya seperti LinkedIn.

3. Diharapkan para ahli dapat digunakan untuk melabeli data sehingga hasil pelabelan lebih tepat.
4. Label yang digunakan tidak hanya positif dan negatif, tetapi juga label netral dan emosional seperti gembira, sedih, marah, dan lain-lain.
5. Penelitian lebih lanjut direncanakan untuk meningkatkan jumlah data ulasan dan memastikan bahwa data yang digunakan seimbang secara proporsional untuk setiap jenis kelas.

DAFTAR PUSTAKA

- [1]J. Widiatoro, "About Us," 2020. <https://www.jobstreet.co.id/id/about-us/> (diakses Jun 04, 2022).
- [2]H. Februariyanti, M. Firmansyah, J. S. Wibowo, dan M. S. Utomo, "Analisis Sentimen Tanggapan Terhadap Aplikasi Layanan Informasi Penginapan Menggunakan Metode Klasifikasi Naive Bayes," *semanTIK*, vol. 6, no. 2, hal. 115–124, 2020, doi: 10.5281/zenodo.4399381.
- [3]H. S. Utama, D. Rosiyadi, D. Aridarma, dan B. S. Prakoso, "SENTIMEN ANALISIS KEBIJAKAN GANJIL GENAP DI TOL BEKASI MENGGUNAKAN ALGORITMA NAIVE BAYES DENGAN OPTIMALISASI INFORMATION GAIN," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, hal. 247–254, Sep 2019, doi: 10.33480/pilar.v15i2.705.
- [4]N. L. Ratniasih dan N. Gunantara, "Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat," *Teknologi Elektro*, vol. 16, no. 3, 2017, [Daring]. Tersedia pada: <http://untroubled.org/spam/>.
- [5]S. Džeroski, "Data Mining," in *Encyclopedia of Ecology, Five-Volume Set*, 2008, hal. 821–830.
- [6]D. Sepri, "Penerapan Algoritma Naïve Bayes Untuk Analisis Kepuasan Penggunaan Aplikasi Bank," *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 1, hal. 135–139, 2020.
- [7]D. P. Daryfayi Edyt dan I. Asror, "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes," *Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes*, vol. 7, no. Ulasan Pada Google Play Store, hal. 11, 2020.
- [8]A. Yasar dan M. M. Saritas, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification Gastric Cancer and Image Processing View project International Journal of Intelligent Systems and Applications in Engineering Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 7, no. 2, hal. 88–91, 2019, doi: 10.1039/b000000x.
- [9]E. B. Santoso dan A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Informatika*, vol. 9, no. 1, hal. 60–69, Sep 2019, doi: 10.30864/eksplora.v9i1.254.
- [10]T. R. Nichols, P. M. Wisner, G. Cripe, dan L. Gulabchand, "Putting the kappa statistic to use," *Quality Assurance Journal*, vol. 13, no. 3–4, hal. 57–61, Jul 2010, doi: 10.1002/qaj.481.
- [11]M. I. Fikri, T. S. Sabrila, dan Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 02, hal.

71–76, 2020, doi: 10.32664/smatika.v10i02.455.

- [12]S. Umami Masruroh dan L. Kesuma Wardhani, *PERBANDINGAN KINERJA ALGORITMA NAIVE BAYES DAN K-NN PENDEKATAN LEXICON PADA ANALISIS SENTIMEN DI MEDIA TWITTER*. 2019.
- [13]B. Herwijayanti, D. E. Ratnawati, dan L. Muflikhah, “Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity,” *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, hal. 306–312, 2018, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>.
- [14]A. Sabrani, I. G. W. Wedashwara W., dan F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)*, vol. 2, no. 1, hal. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.
- [15]E. Sutoyo dan M. A. Fadlurrahman, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network,” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 6, no. 3, hal. 379, 2020, doi: 10.26418/jp.v6i3.42896.