

Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE

Christian Cahyaningtyas¹, Yessica Nataliani², Indrastanti Ratna Widiarsari³
^{1,2,3} Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga
Email : christiancahyaningtyas@gmail.com

Received: 14-07-2021

Riwayat artikel:
Revised: 24-08-2021

Accepted: 12-10-2021

Abstract

Text mining research that performs the categorization procedure on text documents is known as sentiment analysis. Sentiment analysis is the process of extracting a person's ideas, emotions, and evaluations expressed about a certain issue using natural language approaches. Researchers used the Decision Tree approach to do sentiment analysis on the Shopee application rating. The goal of this research is to find out how accurate this Shopee app is and what users think about it. The accuracy value of 99.91 percent, AUC (Area Under Curve) 0.999, recall 99.88 percent, and precision value 99.98 percent were obtained using the Decision Tree method using SMOTE (Synthetic Minority Oversampling Technique). The accuracy of the Decision Tree method without SMOTE is 99.89 percent, the AUC (Area Under Curve) is 0.950, the recall is 99.88 percent, and the precision is 99.98 percent. Based on the findings of the current study, it can be concluded that SMOTE has an effect on value accuracy and AUC (Area Under Curve), but has no effect on recall and precision values, and that the results are the same whether SMOTE is used or not. The accuracy value achieved differs by 0.02 percent, whereas the AUC differs by 0.049

Keywords: Sentiment Analysis, Rating, Decision Tree

Abstrak

Analisis sentimen adalah cabang penelitian *text mining* yang melakukan proses dalam klasifikasi pada dokumen teks. Analisis sentimen merupakan mengekstraksi pendapat, emosi dan evaluasi seseorang yang tertulis mengenai suatu topik tertentu dengan memanfaatkan teknik pemrosesan bahasa alami. Peneliti melakukan penelitian tentang analisis sentimen pada *rating* aplikasi Shopee dengan menggunakan metode *Decision Tree*. Tujuan penelitian ini untuk mengetahui tingkat keakurasian dan mengetahui pendapat pengguna mengenai aplikasi Shopee ini. Hasil penelitian dengan menggunakan algoritma *Decision Tree* dengan SMOTE (*Synthetic Minority Oversampling Technique*) nilai *accuracy*-nya menghasilkan 99,91 persen, AUC (*Area Under Curve*) 0,999, *recall* 99,88 persen dan nilai *precision* 99,98 persen. Hasil menggunakan algoritma *Decision Tree* tanpa SMOTE nilai *accuracy*-nya menghasilkan 99,89 persen, AUC (*Area Under Curve*) 0,950, *recall* 99,88 persen dan nilai *precision* 99,98 persen. Dari hasil evaluasi yang ada dapat ditarik kesimpulan SMOTE dapat berpengaruh terhadap nilai *accuracy* dan AUC (*Area Under Curve*), serta untuk nilai *recall* dan *precision* tidak berpengaruh atau hasilnya tetap

sama walau menggunakan SMOTE atau tanpa SMOTE. Selisih nilai *accuracy* yang didapat adalah 0,02 persen dan untuk AUC-nya sebesar 0.049.

Kata Kunci : Analisis Sentimen, *Rating*, *Decision Tree*

Pendahuluan

Shopee merupakan salah satu *marketplace* yang populer di Indonesia. Saat ini Shopee berada di peringkat nomor dua sebagai *e-commerce* dengan pengunjung situs bulanan pada kuartal pertama tahun 2021 di Indonesia. Berdasarkan data yang diambil dari databoks, Shopee mempunyai jumlah pengunjung sebanyak 127,4 juta pengunjung di kuartal pertama tahun 2021 [1]. Shopee merupakan aplikasi jual beli elektronik yang dapat diunduh di situs *Google Play Store*. *Google Play* adalah situs yang menyediakan layanan konten yang dimiliki oleh *Google* yang telah menyediakan beberapa toko produk *online*, contohnya seperti film, musik, buku, *game* dengan berbagai macam kategori [2]. *Google Play* memiliki berbagai macam fitur, salah satunya adalah *rating* dan ulasan. *Rating* dan fitur ini digunakan untuk para pengguna menilai sebuah aplikasi. *Review* adalah teks yang mempunyai isi pendapat atau komentar seseorang terhadap suatu karya orang lain. Dalam mendapatkan informasi suatu produk atau jasa yang baik dapat dengan menggunakan ulasan dari pengguna lama, hal ini akan menjadi lebih efektif dan efisien. Banyak pengguna internet yang bergantung dari rekomendasi atau pendapat dari pengguna sebelumnya untuk menggunakan suatu produk atau jasa [3].

Analisis sentimen adalah cabang penelitian *text mining* yang melakukan proses dalam klasifikasi pada dokumen teks. Analisis sentimen merupakan mengekstraksi pendapat, emosi dan evaluasi seseorang yang tertulis mengenai suatu topik tertentu dengan memanfaatkan teknik pemrosesan bahasa alami [4]. Menurut Nurjanah analisis sentimen adalah menentukan pendapat tentang masalah, apakah permasalahan tersebut mempunyai kecenderungan ke arah positif ataupun negatif serta dapat digunakan sebagai tolak ukur dalam meningkatkan pelayanan ataupun kualitas produk [5].

Penelitian dengan menggunakan metode *Decision Tree* pernah dilakukan oleh Fajriansyah dan Siswanto [6]. Penelitian ini mengenai analisis sentimen partai politik yang dilakukan untuk menganalisis pendapat positif atau pendapat negatif dari para pengguna Twitter. Hasil dari penelitian ini dapat ditarik kesimpulan bahwa aplikasi analisis sentimen cocok dan bagus untuk menambah referensi kepada masyarakat mengenai partai politik. Serta untuk nilai *accuracy* dengan menggunakan pohon keputusan telah yang dibangun oleh algoritma C4.5 menghasilkan rata-rata error 36 persen berdasarkan data latih dan data uji yang telah digunakan.

Penelitian mengenai pemilihan metode dan algoritma analisis sentimen pernah dilakukan oleh Singgalen dengan menggunakan pendekatan *literature*

review [7]. Hasil dari penelitian tersebut terdapat kesenjangan berdasarkan studi kasus dan sumber data yang menunjukkan bahwa kuantitas kajian analisis sentimen perlu ditingkatkan lagi pada topik yang berhubungan dengan penyelenggaraan pendidikan, dinamika birokrasi, layanan dan fasilitas kesehatan serta aktivitas lembaga swadaya masyarakat. Pada aspek metode atau algoritma juga terdapat kesenjangan yang perlu ditingkatkan kuantitas riset yaitu terhadap metode *BM25*, *Decision Tree*, *K-Means* dan *Neural Network* yang sesuai dengan studi kasus dan sumber data yang ada.

Perbandingan metode *Decision Tree*, KNN dan *Naïve Bayes* sudah pernah dilakukan oleh Puspita dan Widodo pada tahun 2020 [8]. Perbandingan metode tersebut mengenai analisis sentimen pada pengguna layanan BPJS. Data diambil untuk dilakukannya penelitian ini merupakan data dari Twitter. Hasil penelitian dengan menggunakan KNN tingkat *accuracy*-nya mencapai 95,58 persen. Lalu dengan memanfaatkan metode *Decision Tree* dihasilkan nilai *accuracy* sebesar 96,13 persen. Serta yang terakhir dengan menggunakan *Naïve Bayes* untuk tingkat *accuracy* mencapai 89,14 persen. Dari hasil *accuracy* menggunakan ketiga metode tersebut dapat ditarik kesimpulan bahwa metode *Decision Tree* merupakan metode yang terbaik dengan *accuracy* tertinggi dari pada kedua metode yang lain.

Berdasarkan penjelasan dan latar belakang yang telah dijelaskan sebelumnya, maka dilakukannya sebuah penelitian mengenai analisis sentimen. Peneliti akan melakukan penelitian tentang analisis sentimen pada *rating* aplikasi Shopee dengan memanfaatkan metode *Decision Tree*. Data *rating* yang akan diambil merupakan data *rating* skor penilaian pada bintang satu dan bintang lima. Peneliti menggunakan metode *Decision Tree* dikarenakan sudah ada penelitian mengenai kurangnya sebuah penelitian analisis sentimen menggunakan metode *Decision Tree* serta sudah dilakukannya juga sebuah penelitian bahwa metode *Decision Tree* mempunyai tingkat *accuracy* yang cukup tinggi dibandingkan metode yang lain. Tujuan dilakukannya penelitian ini adalah untuk mengetahui tingkat *accuracy* dan mengetahui pendapat pengguna mengenai aplikasi Shopee ini.

Kajian Pustaka

Analisis sentimen telah banyak dilakukan implementasi dengan berbagai objek, metode, data atau parameter. Analisis sentimen dapat dikatakan sebagai teknik untuk mengenali opini seseorang berupa pendapat positif maupun pendapat negatif. Analisis sentimen atau *opinion mining* merupakan cabang ilmu dari *data mining* yang biasanya digunakan untuk menganalisis data tekstual berupa sebuah opini yang mengandung polaritas sehingga nantinya akan menghasilkan suatu informasi yang memiliki nilai baik itu positif, negatif maupun netral [9].

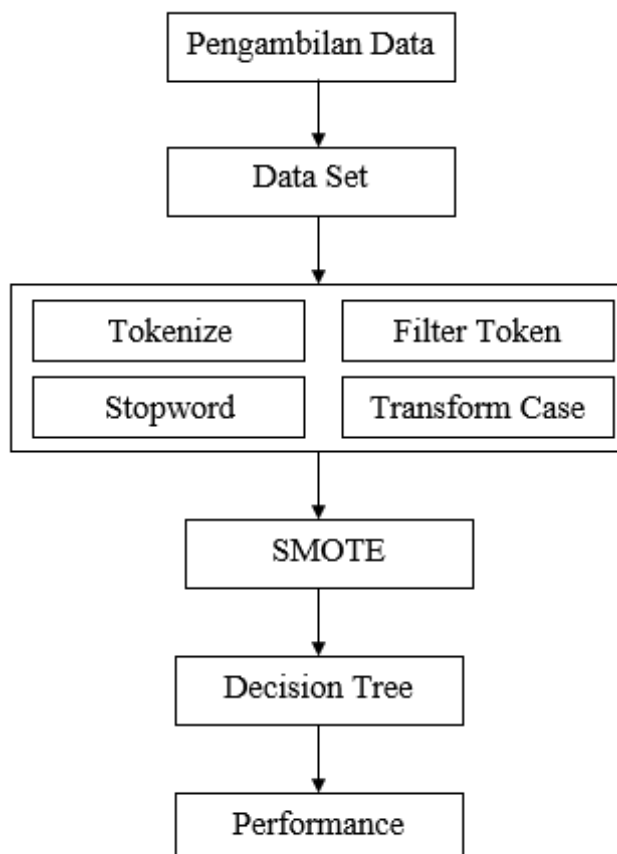
Rating merupakan bentuk kepuasan dari pengguna yang telah menggunakan suatu produk atau jasa [10]. *Rating* menjadi tolak ukur pengguna baru dalam mempertimbangkan produk atau jasa tersebut untuk digunakan. Pada umumnya

rating ditentukan pada jumlah bintang yang telah ditentukan yaitu mulai dari satu sampai lima bintang. Suatu produk bagus atau tidak dapat dinilai dari *rating*-nya yang sudah dinilai dari pengguna lama, sehingga dapat disimpulkan bahwa kualitas produk yang bagus mempunyai jumlah bintang yang makin banyak [11].

Decision Tree adalah metode yang populer dan kuat untuk melakukan prediksi dan klasifikasi. *Decision Tree* dapat merubah fakta menjadikan sebuah pohon keputusan yang dapat mempresentasikan sebuah aturan. Peraturan yang ada bisa dipahami dengan mudah. Proses yang dilakukan pada metode *Decision Tree* adalah dengan merubah suatu data kedalam bentuk pohon keputusan, merubah bentuk pohon kedalam bentuk *role* dan disederhanakannya *role* [12]. *Internal node*, *root node*, dan *terminal node* merupakan bagian dari pohon di dalam metode *Decision Tree*. Untuk sementara sebuah variabel atau fitur merupakan *root node* dan *internal node*, dan label kelas merupakan *terminal node*. Didalam sebuah klasifikasi, *root node* dan *internal node* akan ditelusuri oleh data *query* sampai ke *terminal node*. Untuk label kelas pada *data query* ini berdasarkan label yang telah ada di *internal node* [13].

Metode Penelitian

Didalam penelitian kali ini dalam menghasilkan tingkat *accuracy* yang baik dalam analisis sentimen pada *rating* aplikasi Shopee di situs *Google Play* maka dilakukan dengan menggunakan metode *Decision Tree* berbasis SMOTE. Berikut alur penelitian yang ada pada gambar 1.



Gambar 1. Tahapan Penelitian

1) Pengumpulan Data

Proses pengumpulan data didapat dari pengambilan data *rating* pada aplikasi Shopee di situs *Google Play*. Pengambilan data dilakukan dengan teknik *scraping* menggunakan ekstensi dari *Google Chrome*. Terdapat beberapa macam ekstensi yang ada pada *Google Chrome* yang dapat digunakan dalam mengambil data pada situs *online*. Salah satu ekstensi yang dapat dimanfaatkan adalah data *scraper*. Data *scraper* dapat digunakan untuk mengekstraksi data pada suatu situs secara gratis. Data yang diekstraksi dapat di-*import* menjadi *spreadsheet Microsoft Excel* maupun CSV. Data *review* yang telah diambil akan dijadikan sebagai dataset.

2) Preprocessing Data

Tahap selanjutnya adalah tahap *preprocessing data*. *Preprocessing data* dilakukan dengan beberapa tahap. Tahap *preprocessing* adalah dengan dilakukannya penyiapan data yang didapat supaya data dapat diolah pada saat dilakukannya pemodelan. Tahapan *preprocessing* melingkupi kegiatan memperbaiki data serta menghapus data yang tidak diperlukan supaya dapat diolah ke tahap berikutnya. Tahapan - tahapan *preprocessing* data adalah sebagai berikut:

- *Tokenize*

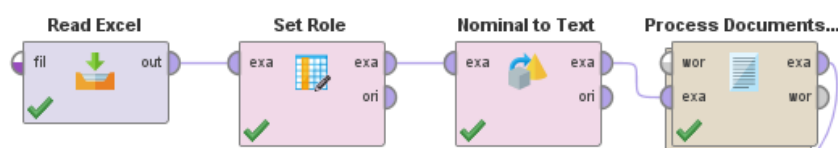
Hasil dan Pembahasan

Pada penelitian ini *rating* atau ulasan pengguna aplikasi Shopee data yang digunakan diambil dari situs *Google Play*. Pengambilan data dilakukan dengan cara menggunakan ekstensi dari *Google Chrome* yaitu data *scraper*. Data *rating* atau ulasan yang diambil sebanyak 2.705 data dengan skor penilaian bintang satu dan bintang lima. Data yang telah diambil dimasukkan ke dalam bentuk *excel* untuk memudahkan dalam proses pengolahan data dan *tools* yang digunakan dalam olah data adalah *Rapidminer*.



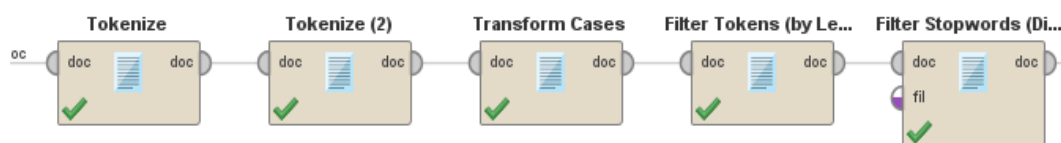
Gambar 2. Proses Olah Data Menggunakan *Tools RapidMiner*.

Langkah pertama adalah dengan meng-*import* dataset yang telah disiapkan pada operator *read excel*. Hubungkan operator *read excel* dengan operator *set role* kemudian *setting set role* di bagian *attribute* dan label terlebih dahulu. Langkah selanjutnya melakukan konversi dari bentuk nominal ke dalam bentuk teks.



Gambar 3. *Process Documents from Data*

Setelah terkonversi menjadi binomial, tahap selanjutnya adalah tahap proses *preprocessing* dengan menggunakan operator *process documents from data*. Setelah data terkumpul mulailah ke tahap pengolahan data. Data yang sudah ada ini tidak dapat langsung digunakan dalam pengolahan untuk analisis sentimen. Sehingga dilanjutkan ke tahap data *preparation*.



Gambar 4. Proses *Pre-Processing*

Didalam proses *preprocessing document* terdapat beberapa operator yang digunakan yaitu :

- *Tokenize*

Proses *tokenize* disini yaitu mengumpulkan semua kata di dalam setiap dokumen dan menghilangkan karakter-karakter yang ada seperti simbol, angka, tanda baca, atau apapun yang bukan berupa huruf.

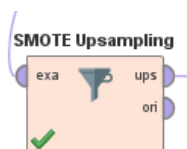
- *Transform Case*

Pada proses ini operator *transform case* digunakan untuk merubah huruf kapital menjadi huruf kecil yang ada pada teks. Ini dilakukan supaya ketika melakukan proses kedalam bentuk klasifikasi ada kesamaan huruf serta terhindar dari kesalahan didalam proses *token*.

- *Filter Token (by length)*

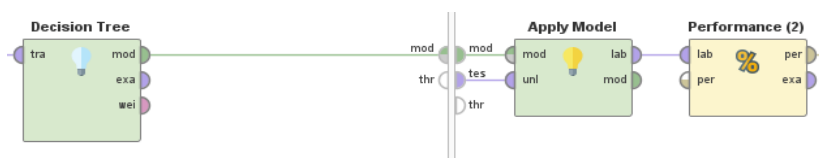
Pada operator *filter token* ini digunakan dalam menghilangkan atau menghapus sejumlah kata yang telah ditentukan. Di dalam penelitian ini peneliti menggunakan panjang karakter minimum empat karakter serta maximum panjang karakternya adalah 25 karakter. Dengan arti lain panjang kalimat kurang dari empat dan lebih dari 25 akan dihapus.

Setelah proses *preprocessing* atas dataset yang ada. Maka selanjutnya adalah melakukan peningkatan kinerja pada metode *Decision Tree* dengan menggunakan SMOTE (*Synthetic Minority Oversampling Technique*).



Gambar 5. Operator *SMOTE Upsampling*

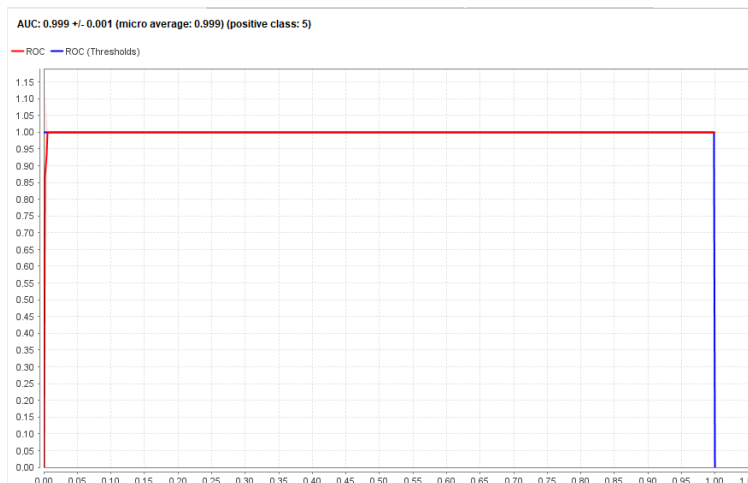
Setelah dilakukan peningkatan kinerja maka selanjutnya dilakukannya proses validasi algoritma. Proses validasi algoritma dengan memanfaatkan operator *cross validation* dengan $k=10$ fold dan *sampling type automatic*. Yang didalamnya terdapat beberapa operasi yang dapat dilihat pada gambar 6.



Gambar 6. Kumpulan Operator didalam *Cross Validation*

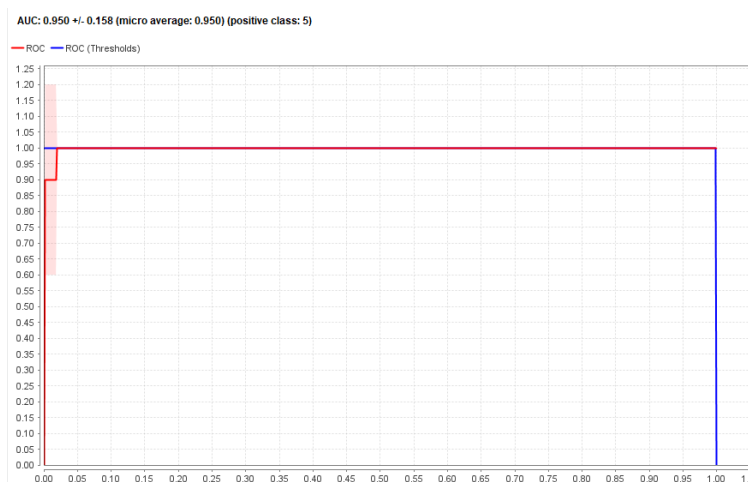
Pada kumpulan operator di *cross validation* ini yang pertama gunakan operator metode *Decision Tree* selanjutnya *apply model* atau memodelkan metode *Decision Tree*. Dan yang terakhir gunakan *performance (binomial classification)*. Pada *performance* pilih untuk memunculkan tingkat *accuracy*, *AUC*, *recall* dan *precision*.

Dalam penelitian ini untuk dilakukannya analisis sentimen menggunakan algoritma *Decision Tree* dan untuk evaluasinya menggunakan nilai dari *accuracy*, *AUC (Area Under Curve)*, *recall* dan *precision*. Berdasarkan hasil penelitian yang sudah dilakukan menggunakan algoritma *Decision Tree* dengan SMOTE. Hasil nilai *AUC* dapat dilihat pada gambar 7.



Gambar 7. Bentuk AUC untuk *Decision Tree* dengan SMOTE

Pada gambar 7 merupakan bentuk AUC untuk algoritma *Decision Tree* menggunakan SMOTE. Dapat dilihat pada gambar nilai AUC dengan SMOTE mendapatkan nilai sebesar 0,999. Sedangkan nilai AUC tanpa menggunakan SMOTE dapat dilihat pada gambar 8.



Gambar 8. Bentuk AUC untuk *Decision Tree* tanpa SMOTE

Pada gambar 8 merupakan bentuk AUC tanpa SMOTE. Dapat dilihat pada gambar bahwa tanpa menggunakan SMOTE hasil dari AUC adalah sebesar 0,950. Dari tahapan yang sudah dilakukan maka dihasilkan nilai dari *accuracy*, AUC, *recall* dan *precision* dengan menggunakan SMOTE dan tanpa SMOTE yang dapat dilihat pada tabel 1.

Tabel 1. Tabel Perbandingan Hasil *Cross Validation* dengan SMOTE dan Tanpa Menggunakan SMOTE

	<i>Decision Tree</i> + SMOTE	<i>Decision Tree</i> tanpa SMOTE	Selisih
<i>Accuracy</i>	99,91%	99,89 %	0,02 %

<i>AUC</i>	0,999	0,950	0,049
<i>Recall</i>	99,88%	99,88%	0
<i>Precision</i>	99,94%	99,94%	0

Dapat dilihat pada tabel 1 merupakan hasil *cross validation* menggunakan SMOTE dan tanpa menggunakan SMOTE. Dapat dilihat pada tabel bahwa hasil menggunakan SMOTE mendapatkan nilai *accuracy* yaitu 99,91 persen, *AUC* 0,999, *recall* 99,88 persen dan *precision* 99,94 persen. Sedangkan hasil tanpa menggunakan SMOTE nilai *accuracy* sebesar 99,89 persen, *AUC* 0,950, *recall* 99,88 persen dan *precision* 99,94 persen. Selisih hasil yang didapat menggunakan SMOTE dan tanpa SMOTE untuk *accuracy* memiliki selisih sebesar 0,02 persen dan selisih untuk *AUC* sebesar 0,049.

Simpulan

Setelah dilakukannya penelitian dengan menggunakan model algoritma *Decision Tree* menggunakan SMOTE dan tanpa menggunakan SMOTE dapat disimpulkan bahwa penggunaan SMOTE mendapatkan nilai yang lebih baik dibandingkan tanpa menggunakan SMOTE. Hasil penelitian dengan menggunakan algoritma *Decision Tree* dengan SMOTE nilai *accuracy*-nya menghasilkan 99,91 persen, *AUC* 0,999, *recall* 99,88 persen dan nilai *precision* 99,98 persen. Hasil menggunakan algoritma *Decision Tree* tanpa SMOTE nilai *accuracy*-nya menghasilkan 99,89 persen, *AUC* 0,950, *recall* 99,88 persen dan nilai *precision* 99,98 persen. Dari hasil nilai *accuracy* dan *AUC* yang ada dapat ditarik kesimpulan SMOTE dapat berpengaruh terhadap nilai *accuracy* dan *AUC*, dan untuk nilai *recall* dan *precision* tidak berpengaruh atau hasilnya tetap sama walau menggunakan SMOTE atau tanpa SMOTE. Selisih nilai *accuracy* yang didapat adalah 0,02 % dan untuk *AUC*-nya sebesar 0.049.

Daftar Pustaka

- [1] D. H. Jayani, "Jumlah Pengunjung Tokopedia Kalahkan Shopee Pada Kuartal I-2021," 2021. <https://databoks.katadata.co.id/datapublish/2021/06/11/jumlah-pengunjung-tokopedia-kalahkan-shopee-pada-kuartal-i-2021#>.
- [2] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, pp. 293–298, 2020, doi: 10.24114/cess.v5i2.18186.
- [3] M. Rezki, D. N. Kholifah, M. Faisal, Priyono, and R. Suryadithia, "Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes," *J. Infortech*, vol. 2, no. 2, pp. 264–270, 2020, doi: 10.31294/infortech.v2i2.9286.
- [4] A. R. Sentiaji and A. M. Bachtiar, "Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik," *J. Ilm. Komput. dan Inform.*, pp. 1–6,

- 2014.
- [5] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [6] M. R. Fajriansyah and Siswanto, “Analisis Sentimen Pengguna Twitter Terhadap Partai Politik Pendukung Calon Gubernur Di Jakarta Menggunakan Algoritma C4 . 5 Decision Tree Learning,” *SKANIKA*, vol. 1, no. 2, pp. 697–703, 2018, [Online]. Available: <https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/278>.
- [7] Y. A. Singgalen, “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sitematic Literature Review,” *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, 2021.
- [8] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2020, doi: 10.32493/informatika.v5i4.7622.
- [9] B. Laurensz and E. Sedyono, “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021.
- [10] R. B. Trianto, A. Triyono, and D. M. P. Arum, “Klasifikasi Rating Otomatis pada Dokumen Teks Ulasan Produk Elektronik Menggunakan Metode N-gram dan Naïve Bayes,” *J. Inform. Univ. Pamulang*, vol. 5, no. 3, pp. 295–301, 2020, doi: 10.32493/informatika.v5i3.6110.
- [11] L. H. Tampubolon, M. A. Fauzi, and Y. A. Sari, “Prediksi Rating pada Ulasan Produk Kecantikan menggunakan Metode SO-CAL in an Inheritance-based,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 5765–5771, 2019.
- [12] K. G. Prasetyo and S. M. Pahlevi, “Analisis Perbandingan Algoritma Decision Tree Dengan Support Vector Machine Untuk Mendeteksi Kompetensi Mahasiswa Konsentrasi Informatika Komputer Studi Kasus : Politeknik Lp3I Jakarta , Kampus Depok,” *J. Lentera ICT*, vol. 5, no. 2, pp. 11–26, 2019.
- [13] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes , Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [14] R. Siringoringo, “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor,” *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.
- [15] L. D. Perwara, F. A. Bachtiar, and Indriati, “Penerapan Algoritma Decision Tree C4 . 5 Untuk Deteksi Fraud Pada Kartu Kredit dengan Oversampling Synthetic Minority Technique (SMOTE),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 8, pp. 2664–2669, 2020.

- [16] A. Franseda, W. Kurniawan, A. Sita, and W. Gata, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 08, no. 3, pp. 282–290, 2020, doi: 10.26418/justin.v8i3.40982.
- [17] D. A. Ningtyas, M. Wahyudi, and N. Nurajijah, "Klasifikasi Siswa SMK Berpotensi Putus Sekolah Menggunakan Algoritma Decision Tree, Support Vector Machine Dan Naive Bayes," *J. Khatulistiwa Inform.*, vol. 7, no. 2, pp. 85–90, 2019, doi: 10.31294/jki.v7i2.6839.