

# PERBANDINGAN IMPLEMENTASI METODE *SMOTE* PADA ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DALAM ANALISIS SENTIMEN OPINI MASYARAKAT TENTANG MIXUE

Teresia Ardika Dewi <sup>1\*</sup>, Evangs Mailoa <sup>2</sup>.

<sup>1,2</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Email: 672019066@student.uksw.edu <sup>1\*</sup>, evangs.mailoa@uksw.edu <sup>2</sup>

## Histori Artikel:

Dikirim 22 Mei 2023; Diterima dalam bentuk revisi 10 Juni 2023; Diterima 1 Juli 2023; Diterbitkan 10 September 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

## Abstrak

Pada bulan Desember tahun 2022, perkembangan dari franchise gerai minuman es krim dan teh asal Cina bernama Mixue menjadi perbincangan masyarakat Indonesia, khususnya di media sosial Twitter, sehingga memunculkan berbagai opini dari masyarakat terkait gerai Mixue yang berkembang begitu pesat. Sehingga dari hal tersebut, akan dilakukan analisis sentimen dengan melakukan klasifikasi menggunakan implementasi algoritma Support Vector Machine (SVM). Dari hasil penelitian yang telah dilakukan, algoritma Support Vector Machine (SVM) dengan berbasis SMOTE menghasilkan peningkatan nilai accuracy menjadi 73.67% dan precision menjadi 75.40%, dan untuk hasil Support Vector Machine (SVM) tanpa menggunakan SMOTE, nilai accuracy sebesar 69.40% dan nilai precision sebesar 68.12%. Namun sebaliknya terjadi penurunan pada nilai recall menjadi 70.83% dan nilai F1-Score menjadi 72.79%. Sehingga dari hasil evaluasi dapat disimpulkan bahwa SMOTE berpengaruh dalam peningkatan nilai accuracy dan precision, namun terjadi penurunan pada nilai recall dan F1-Score.

**Kata Kunci:** Analisis Sentimen; Support Vector Machine (SVM); Twitter; SMOTE.

## Abstract

In December 2022, the development of a franchise for an ice cream and tea outlet from China named Mixue became the talk of the Indonesian people, especially on social media Twitter, giving rise to various opinions from the public regarding the Mixue outlet which is growing so rapidly. So from that, sentiment analysis will be carried out by classifying using the implementation of the Support Vector Machine (SVM) algorithm. From the results of research that has been done, the SMOTE-based Support Vector Machine (SVM) algorithm results in an increase in the accuracy value to 73.67% and precision to 75.40%, and for the results of Support Vector Machine (SVM) without using SMOTE, the accuracy value is 69.40% and the precision value is 68.12%. But on the contrary there was a decrease in the recall value to 70.83% and the F1-Score value to 72.79%. So from the evaluation results it can be concluded that SMOTE has an effect on increasing the accuracy and precision values, but there is a decrease in the recall value and F1-Score.

**Keyword:** Sentiment Analysis; Support Vector Machine; Twitter; SMOTE.

## 1. Pendahuluan

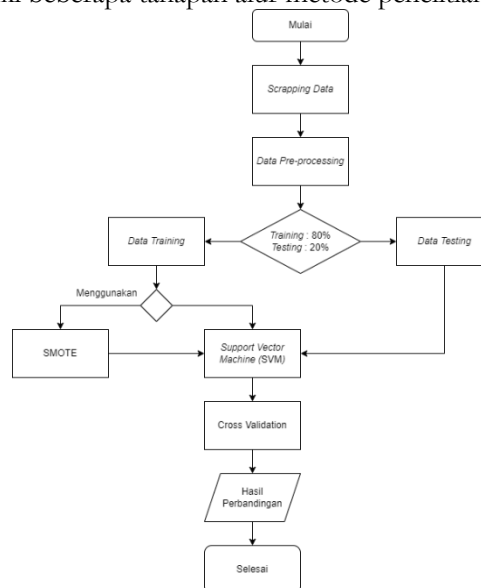
Internet telah mengalami revolusi komunikasi yang begitu pesat dibanding sebelumnya, dengan adanya internet, dapat memberikan kemudahan dan efisiensi transmisi data, dan juga menjadi saluran interaksi dari berbagai posisi geografis di seluruh dunia [1]. Salah satunya adalah penggunaan media sosial Twitter, di mana saat ini Twitter menjadi media sosial utama yang digunakan untuk berkomunikasi dan berbagi informasi kepada sesama pengguna Twitter [2]. Sehingga melalui Twitter, para pengguna dapat menyampaikan berbagai opini mereka, salah satunya adalah opini terkait sebuah gerai asal Cina yang menjual es krim dan teh yaitu Mixue, di mana pada penghujung tahun 2022 yaitu tanggal 26 Desember 2022, media sosial Twitter diramaikan dengan trendingnya “Mixue”, diambil dari website Trends24, Mixue trending pada peringkat 48 yang dicuitkan sebanyak kurang lebih 34.000 kali [3].

Banyaknya cuitan tentang Mixue ini, karena Mixue dapat membuka cabang dari gerainya dengan begitu pesat, bahkan ketika terdapat ruko kosong yang terbengkalai, dalam waktu singkat telah berubah menjadi gerai Mixue, sehingga dari hal tersebut masyarakat Indonesia dapat dengan mudah menemukan cabang dari Mixue yang memunculkan sebuah julukan baru yaitu “Malaikat pencabut ruko kosong” [4]. Bukan hanya karena gerai Mixue yang sudah menjamur di berbagai daerah saja, tapi Mixue juga digemari oleh masyarakat Indonesia karena harganya murah dan memiliki rasa yang enak, membuat Mixue sangat disukai oleh berbagai kalangan mulai dari anak-anak, remaja, dewasa, hingga orang tua [5]. Tidak heran membuat banyak pengguna Twitter yang muncul dan menyampaikan berbagai tanggapan mereka terhadap fenomena gerai Mixue ini baik berupa tanggapan positif, negatif, atau pun netral, di mana dari tanggapan-tanggapan tersebut akan dikelompokkan dan mengetahui bagaimana emosi dari tanggapan-tanggapan yang disampaikan oleh pengguna Twitter tersebut [1].

Dari permasalahan tersebut perlunya sebuah pengolahan data menggunakan metode teknik *text mining*, dan pada penelitian ini akan menggunakan metode analisis sentimen dengan algoritma *Support Vector Machine (SVM)* berbasis *SMOTE (Synthetic Minority Over-sampling Technique)*, yang bertujuan untuk memahami, dan mengolah data yang dihubungkan dengan proses analisis sentimen pada data *tweet* oleh pengguna media sosial Twitter di Indonesia [6].

## 2. Metode Penelitian

Pada penelitian memiliki beberapa tahapan alur metode penelitian sebagai berikut :



Gambar 1. Metode Penelitian

## 2.1 Scrapping Data

Proses *scrapping data* atau pengumpulan data diambil menggunakan bantuan *tools Orange Data Mining*, untuk menarik data berupa *tweet* pada Twitter. Penarikan data dilakukan menggunakan fitur *Scrapping Twitter* yang terdapat di *Orange Data Mining* yang sudah terintegrasi dengan *API Twitter*. Data yang diperoleh menggunakan kata kunci “Mixue” saat puncak *trending* topik dari Mixue yaitu pada tanggal 26 Desember 2022. Data yang dihasilkan dalam bentuk format .csv, yang kemudian dikonversi menjadi bentuk tabel menggunakan fitur *Data Table*, sehingga dihasilkan data yang lebih rapi. Data *tweet* tersebutlah yang akan digunakan sebagai dataset, namun masih dengan kondisi dataset kotor.

## 2.2 Data Pre-Processing

Dataset yang sebelumnya telah diperoleh melalui tahap *Scrapping Data* sebelumnya, selanjutnya akan dilakukan *Data Pre-Processing* untuk mengubah data yang tidak terstruktur menjadi data terstruktur, sehingga menghasilkan dataset bersih yang siap untuk dianalisis [7]. Berikut adalah beberapa tahapan dari *Data Pre-Processing*, di antaranya adalah sebagai berikut, yaitu:

- 1) *Data Cleaning* : *Data cleaning* yang dilakukan pertama kali adalah menghapus *url*, RT, *username*, *hashtags*, ‘@’ dan simbol-simbol lainnya dari *tweet* yang telah ditarik, sehingga data yang didapatkan sudah bersih dari simbol-simbol dan *url*. Kemudian pada tahap kedua, dilakukan *data cleaning* manual menggunakan *microsoft excel*, dilakukan untuk menyaring *tweet* yang kurang sesuai, serta jika terdapat *tweet* berulang yang tidak terdeteksi pada proses *data cleaning* pertama. Setelah dari tahap *data cleaning* kedua, didapatkan dataset *final* yang bersih.
- 2) *Tokenization* : Dari dataset yang telah bersih, selanjutnya akan ditokenisasi untuk segmentasi kata per katanya dari kalimat teks pada *tweet*.
- 3) *Stopwords* : Pada tahap ini, *stopwords* digunakan untuk menghapus atau menghilangkan kata-kata yang tidak diperlukan, namun dengan tidak mengubah makna dari kata tersebut.
- 4) *Transform Case* : Proses *transform case* adalah untuk merubah huruf kapital yang ada pada kata menjadi huruf kecil secara menyeluruh.

## 2.3 Splitting Data

Dataset yang telah melalui proses *data cleaning*, selanjutnya akan dilakukan pembagian data ke dalam data training dan data testing, dengan perbandingan antara data training dan data testing adalah 8:2.

## 2.4 SMOTE (Synthetic Minority Oversampling Technique)

Metode SMOTE (*Synthetic Minority Oversampling Technique*) adalah salah satu metode yang begitu populer untuk digunakan dalam mengatasi sebuah ketidakseimbangan kelas. Dengan menggunakan metode SMOTE, sampel yang baru akan disintesis dari kelas minoritas yang kemudian akan menyeimbangkan dataset, dengan cara membangun sebuah *instance* baru dari kelas minoritas dengan tujuan meningkatkan *performance* dari metode klasifikasi [7], [8].

## 2.5 Support Vector Machine (SVM)

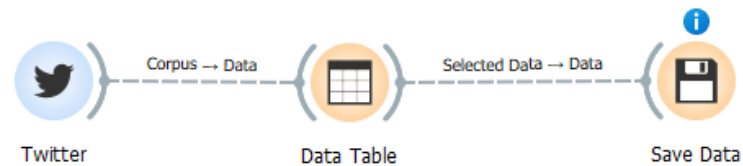
Algoritma *Support Vector Machine* (SVM) merupakan salah satu metode dari Supervised Learning, di mana metode *Support Vector Machine* (SVM) digunakan untuk dapat mengenali pola dan juga mengklasifikasikan data. Algoritma *Support Vector Machine* (SVM) merupakan algoritma pembelajaran dalam melakukan kategorisasi teks, sehingga pada penelitian ini, algoritma *Support Vector Machine* (SVM) diimplementasikan, yang kemudian akan dianalisis seberapa akurat yang dihasilkan dari algoritma *Support Vector Machine* (SVM) terhadap sentimen dari *tweet* twitter mengenai Mixue [9][10].

## 3. Hasil dan Pembahasan

Pada penelitian terkait analisis sentimen untuk mengetahui bagaimana opini masyarakat mengenai Mixue, melalui implementasi algoritma *Support Vector Machine* (SVM) dengan berbasis

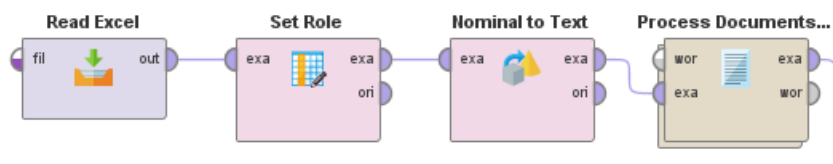
SMOTE dan tanpa berbasis SMOTE digunakan aplikasi bernama Orange untuk melakukan pengambilan data tweet yang ada pada Twitter dengan kata kunci “Mixue”, dan untuk pengolahan data lebih lanjut digunakan aplikasi RapidMiner dengan tools-tools yang memadai yang ada di dalam RapidMiner yaitu Text Processing.

Dalam melakukan pengambilan data atau scrapping data menggunakan Orange, digunakan tools Text Mining. Perlu diperhatikan untuk dapat melakukan scrapping data, diperlukan Twitter API Key. Kemudian setelah proses scrapping data telah selesai, maka hasil scrapping data disimpan dalam bentuk file excel.



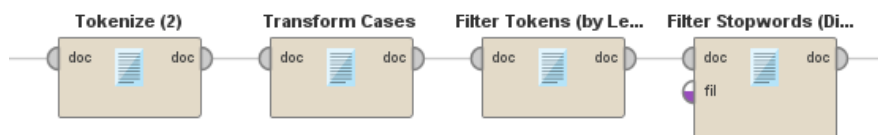
Gambar 2. Proses *Scrapping Data* menggunakan Orange

Dari hasil *scrapping data* yang telah didapatkan, diambil sebanyak 1000 *tweet* untuk digunakan sebagai dataset, yang selanjutnya akan dilakukan proses *Data Pre-Processing* menggunakan RapidMiner dan juga pelabelan data *tweet* secara manual. Langkah yang dilakukan pertama kali adalah membaca *file excel* yang berisikan hasil dari *scrapping data* sebelumnya dengan *import data* dengan *operator read excel*, kemudian mengatur *attribute name* pada *field* dan *target role* pada data. Agar dapat mempermudah dalam pengolahan data, maka perlu dilakukan konversi data yang masih berupa nominal menjadi teks.



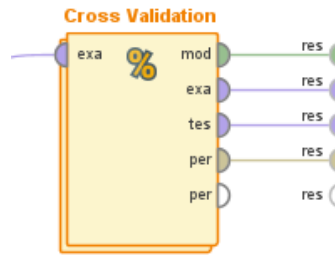
Gambar 3. Proses mempersiapkan data sebelum *Data Pre-Processing*

Pada tahap *Data Pre-Processing* yang ada di dalam *operator Process Documents from Data*, berisikan beberapa tahapan yaitu *Tokenize* untuk menghilangkan karakter-karakter atau simbol-simbol tertentu dan menghapus data yang bukan merupakan huruf, *Transform Case* untuk merubah seluruh kata yang ada di dalam data menjadi *lower case* atau huruf kecil dengan tujuan data yang akan diolah memiliki kesamaan untuk meminimalisir kesalahan, *Filter Token (by Length)* untuk menghapus data dengan ketentuan minimal terdapat empat karakter dan maksimal 25 karakter, dan *Filter Stopwords (Dictionary)* untuk menghapus kata-kata yang tidak dibutuhkan tanpa menghilangkan makna dari kata tersebut apabila berdiri sendiri.

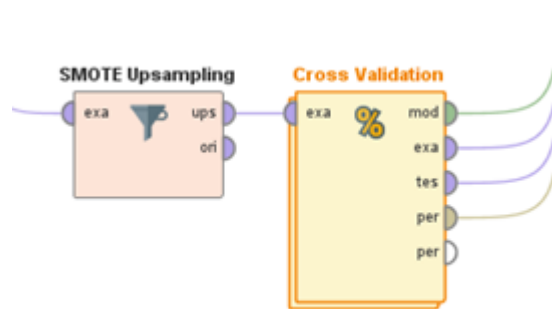


Gambar 4. Proses *Data Pre-Processing*

Terdapat dua skenario yang akan dilakukan, yaitu mengolah data dari hasil *Data Pre-Processing* menggunakan algoritma tanpa SMOTE dan dengan berbasis SMOTE (*Synthetic Minority Oversampling Technique*). Seperti yang ada di Gambar 5, menunjukkan penggunaan operator SMOTE (*Synthetic Minority Oversampling Technique*) pada setiap skenario.

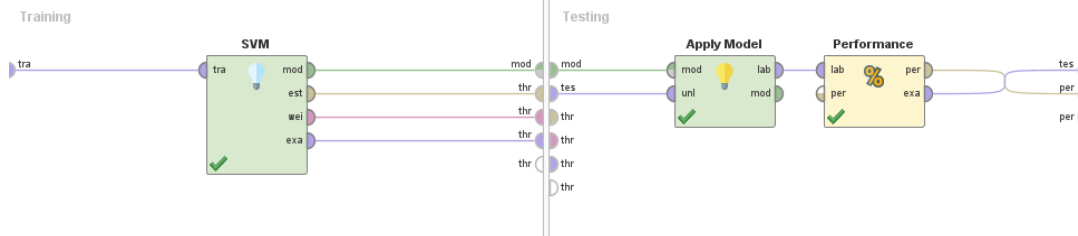


Gambar 5. Tanpa Menggunakan operator SMOTE (*Synthetic Minority Oversampling Technique*)



Gambar 6. Menggunakan operator SMOTE (*Synthetic Minority Oversampling Technique*)

Pada langkah berikutnya yang ada di dalam *operator cross validation* terdapat proses implementasi algoritma *Support Vector Machine* (SVM), dengan meliputi dua proses yaitu *training* dan *testing*. Sehingga menghasilkan nilai *performance* (*binomial classification*) berupa *accuracy*, *precision*, *recall*, dan *F1-Score*.



Gambar 7. Proses pada *operator cross validation* menggunakan algoritma *Support Vector Machine* (SVM)

Berikut adalah tabel perbandingan dari hasil pengujian dengan menggunakan algoritma *Support Vector Machine* (SVM) tanpa SMOTE dan *Support Vector Machine* (SVM) berbasis SMOTE.

Tabel 1. Contoh tabel

	<i>Support Vector Machine</i> (SVM)	<i>Support Vector Machine</i> (SVM) + SMOTE	Selisih
Accuracy	69.40%	73.67%	4.27%
Precision	68.12%	75.40%	7.28%
Recall	83.11%	70.83%	-12.28%
F1-Score	74.78%	72.79%	-1.99%

Dapat dilihat pada Tabel 1 di atas, menunjukkan hasil perbandingan nilai dari *Performance Vector* pada dua skenario yang telah dilakukan, untuk skenario pertama yaitu menggunakan algoritma *Support Vector Machine* (SVM) tanpa SMOTE menghasilkan *accuracy* 69.40%, *precision* 68.12%, *recall* 83.11%, dan *F1-Score* 74.78%. Kemudian pada skenario kedua yaitu menggunakan algoritma *Support Vector Machine* (SVM) berbasis SMOTE menghasilkan *accuracy* 73.67%, *precision* 75.40%, *recall* 70.83%, dan *F1-Score* 72.79%. Sehingga dari hasil yang telah didapatkan, peningkatan nilai *accuracy* sebesar 4.27% dan *precision* sebesar 7.28% apabila menggunakan SMOTE. Namun mengalami penurunan pada nilai *recall* sebesar 12.28% dan *F1-Score* sebesar 1.99% jika menggunakan SMOTE.

#### 4. Kesimpulan

Berdasarkan hasil analisis dan pengujian yang telah dilakukan pada penelitian ini, menghasilkan kesimpulan bahwa penggunaan SMOTE dalam algoritma *Support Vector Machine* (SVM) menunjukkan peningkatan nilai *accuracy* sebesar 4.27% menjadi 73.67% dan nilai *precision* 7.28% menjadi 75.40% dibandingkan tanpa menggunakan SMOTE. Sehingga dari hal tersebut membuktikan bahwa penggunaan SMOTE dinilai sudah cukup baik. Kemudian analisis sentimen dengan mengimplementasikan algoritma *Support Vector Machine* (SVM) menghasilkan keberhasilan dalam mengklasifikasikan sentimen dari opini masyarakat di media sosial *Twitter* tentang Mixue. Dapat disimpulkan SMOTE dapat berpengaruh dalam meningkatkan nilai *accuracy* dan *precision*, namun sebaliknya menurunkan nilai *recall* dan *F1-Score*.

#### 5. Daftar Pustaka

- [1] Mohamad, A. K., Jayakrishnan, M., & Nawi, N. H. (2020). Classification of Twitter Data by Sentiment Analysis in the Malay Language. *International Journal*, 8(6). DOI: 10.30534/ijeter/2020/83862020.
- [2] C. Reviews, (2020). An efficient Decision Tree Algorithm for analyzing the Twitter Sentiment Analysis, 7(4), pp. 1010–1018.
- [3] T. Madura, (2023). Apa itu Mixue, Sempat Trending di Twitter dan Jadi Guyonan, Gerai Minuman yang Tersebar di Indonesia. <https://www.msn.com/id-id/berita/other/apa-itu-mixue-sempt-trending-di-twitter-dan-jadi-guyonan-gerai-minuman-yang-tersebar-di-indonesia/ar-AA15Jjz1> (accessed Jan. 21, 2023).
- [4] Lintang Siltya Utami, (2023). Ngakak! Mixue Jadi Trending Topic Twitter, Warganet Sebut Indonesia Dijajah Tiga Kali. <https://www.suara.com/tekno/2022/12/28/122252/ngakak-mixue-jadi-trending-topic-twitter-warganet-sebut-indonesia-dijajah-tiga-kali> (accessed Feb. 07, 2023).
- [5] C. I. Tim Redaksi, (2023). Pantasan Harganya Murah, Ternyata Ini Rahasia Bisnis Mixue, [Online]. Available: <https://www.cnbcindonesia.com/lifestyle/20230104182558-33-402849/pantasan-harganya-murah-ternyata-ini-rahasia-bisnis-mixue>
- [6] Rizkia, S., Setiawan, E. B., & Puspendari, D. (2019). Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF. *eProceedings of Engineering*, 6(2).



- [7] Cahyaningtyas, C., Nataliani, Y., & Widiyanti, I. R. (2021). Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE. *AITI*, 18(2), 173-184. DOI: 10.24246/aiti.v18i2.173-184.
- [8] Normah, N., Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174-180. DOI: 10.31294/jtk.v4i2.
- [9] Iftikar, M. A., & Sibaroni, Y. (2022). Analisis Sentimen Twitter: Penanganan Pandemi Covid-19 Menggunakan Metode Hybrid Naïve Bayes, Decision Tree, dan Support Vector Machine. *eProceedings of Engineering*, 9(3).
- [10] Mufidah, F. S., Winarno, S., Al Zami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat terhadap Layanan Shopeefood Melalui Media Sosial Twitter dengan Algoritma Naïve Bayes Classifier. *vol*, 7, 14-25. DOI: 10.33633/joins.v7i1.5883.