



**KLASIFIKASI SENTIMEN TWEET TENTANG SICEPAT DENGAN  
STRATEGI PENYEIMBANGAN DATA**

**LAPORAN TUGAS AKHIR**

**RIFQI AL GHIFARY**  
**41520010023**

**UNIVERSITAS**  
**MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS MERCU BUANA**  
**JAKARTA**  
**2025**



**KLASIFIKASI SENTIMEN TWEET TENTANG SICEPAT  
DENGAN STRATEGI PENYEIMBANGAN DATA**

**LAPORAN TUGAS AKHIR**

**RIFQI AL GHIFARY**  
**41520010023**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana**

**UNIVERSITAS**  
**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS MERCU BUANA**  
**JAKARTA**  
**2025**

## HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Rifqi Al Ghifary

NIM : 41520010023

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Klasifikasi Sentimen Tweet tentang SiCepat dengan Strategi Penyeimbangan Data

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.



Jakarta, 19 Juli 2025



Rifqi Al Ghifary

UNIVERSITAS  
MERCU BUANA

## HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : RIFQI AL GHIFARY  
NIM : 41520010023  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Klasifikasi Sentimen Tweet tentang SiCepat dengan Strategi Penyeimbangan Data

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Fauzi Nur Iman, S.Kom., M.Kom (  )  
NIDN : 0318088903  
Ketua Penguji : Afiyati, Dr., S.Si, MT (  )  
NIDN : 0316106908  
Penguji 1 : Dhanny Permatasari Putri, S.Kom, MT (  )  
NIDN : 0328087903  
Penguji 2 : Dwiki Jatikusumo, S.Kom, M.Kom (  )  
NIDN : 0301128903

UNIVERSITAS  
MERCU BUANA

Jakarta, 19 Juli 2025

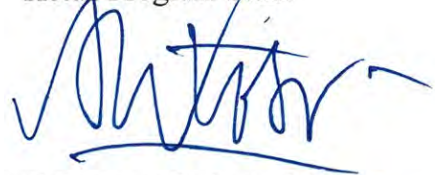
Mengetahui,

Dekan



Dr. Bambang Jokonowo, S.Si., MTI  
NIDN : 0320037002

Ketua Program Studi



Dr. Hadi Santoso, S.Kom., M.Kom  
NIDN : 0225067701

## KATA PENGANTAR

Puji syukur kehadiran Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan tugas akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Kedua Orang Tua saya yang selalu mensupport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubuana.
2. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
3. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
4. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubuana.
5. Ibu Dosen Pembimbing, S.Kom., M.Kom. selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan tugas akhir ini terjadwal dengan baik.
6. Semua teman kuliah yang selalu berbagi informasi dan memberikan dukungan dalam bentuk yang berbeda-beda.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 24 Juli 2025



Rifqi Al Ghifary

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS  
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : RIFQI AL GHIFARY  
NIM : 41520010023  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Klasifikasi Sentimen Tweet tentang SiCepat dengan Strategi Penyeimbangan Data

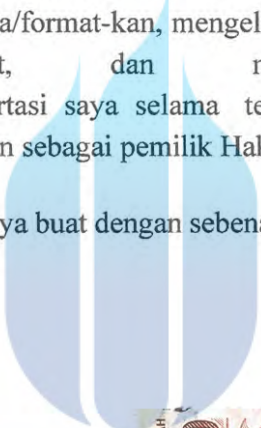
Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Juli 2025

Yang menyatakan,

  
  
  
Rifqi Al Ghifary



## ABSTRAK

Nama : Rifqi Al Ghifary  
NIM : 41520010023  
Program Studi : Teknik Informatika  
Judul Laporan Skripsi : Klasifikasi Sentimen Tweet tentang SiCepat dengan Strategi Penyeimbangan Data  
Dosen Pembimbing : Fauzi Nur Iman, S.Kom., M.Kom

Penelitian ini bertujuan membangun sistem klasifikasi sentimen otomatis terhadap tweet berbahasa Indonesia yang membahas layanan SiCepat. Sebanyak 15.000 data dikumpulkan melalui Tweet-Harvest dan diproses melalui tahap preprocessing, stemming, dan pelabelan otomatis berbasis lexicon berbobot. Masalah utama adalah ketidakseimbangan kelas, di mana tweet positif jauh lebih banyak daripada negatif. Untuk mengatasinya, digunakan strategi penyeimbangan seperti SMOTE, undersampling, dan kombinasi keduanya. Tiga algoritma pembelajaran mesin yang diuji adalah Support Vector Machine (SVM), Naive Bayes (NB), dan Logistic Regression (LR). Evaluasi dilakukan menggunakan akurasi, precision, recall, F1-score, dan confusion matrix. Hasil menunjukkan SVM dengan SMOTE memiliki kinerja terbaik (akurasi 90,5% dan F1-score 0,907), disusul Logistic Regression kombinasi (akurasi 89,2%). Naive Bayes cenderung bias terhadap kelas mayoritas. Pendekatan kombinasi data dengan SVM terbukti paling efektif dan direkomendasikan untuk implementasi analisis sentimen di industri logistik.

**Kata kunci:** Analisis Sentimen, SMOTE, Logistic Regression, SVM, Naive Bayes, Ketimpangan Kelas, SiCepat, Tweet-Harvest, Media Sosial

## ABSTRACT

Name : Rifqi Al Ghifary  
NIM : 41520010023  
Study Program : Informatics Engineering  
Thesis Title : Tweet Sentiment Classification on SiCepat  
Delivery Service Using Data Balancing Techniques  
Counsellor : Fauzi Nur Iman, S.Kom., M.Kom

*This study aims to develop an automatic sentiment classification system for Indonesian-language tweets related to SiCepat delivery services. A total of 15,000 tweets were collected using Tweet-Harvest and processed through stages of preprocessing, stemming, and automatic labeling based on a weighted lexicon. The main issue encountered was class imbalance, where positive tweets significantly outnumbered negative ones. To address this, balancing strategies such as SMOTE, undersampling, and a combination of both were applied. Three machine learning algorithms were tested: Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR). The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Results show that SVM with SMOTE achieved the best performance (90.5% accuracy and 0.907 F1-score), followed by Logistic Regression with a combined balancing approach (89.2% accuracy). Naive Bayes tended to be biased toward the majority class. Overall, the combined data balancing approach with SVM proved to be the most effective and is recommended for sentiment analysis implementation in the logistics industry.*

**Keywords:** Sentiment Analysis, Machine Learning, Class Imbalance, SMOTE, SVM, Logistic Regression, Naive Bayes, Twitter, SiCepat.



## DAFTAR ISI

|                                                                                                  |             |
|--------------------------------------------------------------------------------------------------|-------------|
| <b>HALAMAN JUDUL .....</b>                                                                       | <b>i</b>    |
| <b>HALAMAN PERNYATAAN KARYA SENDIRI .....</b>                                                    | <b>ii</b>   |
| <b>HALAMAN PENGESAHAN.....</b>                                                                   | <b>iii</b>  |
| <b>KATA PENGANTAR.....</b>                                                                       | <b>iv</b>   |
| <b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR<br/>UNTUK KEPENTINGAN AKADEMIS .....</b> | <b>v</b>    |
| <b>ABSTRAK .....</b>                                                                             | <b>vi</b>   |
| <b>ABSTRACT .....</b>                                                                            | <b>vii</b>  |
| <b>DAFTAR ISI.....</b>                                                                           | <b>viii</b> |
| <b>DAFTAR TABEL .....</b>                                                                        | <b>x</b>    |
| <b>DAFTAR GAMBAR.....</b>                                                                        | <b>xi</b>   |
| <b>DAFTAR LAMPIRAN .....</b>                                                                     | <b>xii</b>  |
| <b>BAB I PENDAHULUAN.....</b>                                                                    | <b>1</b>    |
| 1.1 Latar Belakang .....                                                                         | 1           |
| 1.2 Perumusan Masalah .....                                                                      | 2           |
| 1.3 Tujuan Penelitian .....                                                                      | 2           |
| 1.4 Manfaat Penelitian .....                                                                     | 3           |
| 1.5 Batasan Masalah .....                                                                        | 3           |
| <b>BAB II TINJAUAN PUSTAKA .....</b>                                                             | <b>5</b>    |
| 2.1 Penelitian Terdahulu .....                                                                   | 5           |
| 2.2 Teori Pendukung.....                                                                         | 17          |
| 2.2.1 Naïve Bayes.....                                                                           | 17          |
| 2.2.2 Logistic Regression .....                                                                  | 19          |
| 2.2.3 Support Vector Machine .....                                                               | 19          |
| 2.2.4 Python.....                                                                                | 20          |
| 2.2.5 Twitter .....                                                                              | 21          |
| 2.2.6 Metode Lexicon.....                                                                        | 21          |
| 2.2.7 Confusion Matrix .....                                                                     | 21          |
| <b>BAB III METODE PENELITIAN .....</b>                                                           | <b>23</b>   |
| 3.1 Jenis Penelitian.....                                                                        | 23          |

|                                        |                            |           |
|----------------------------------------|----------------------------|-----------|
| 3.2                                    | Tahapan Penelitian.....    | 23        |
| <b>BAB IV PEMBAHASAN .....</b>         |                            | <b>32</b> |
| 4.1                                    | Dataset.....               | 32        |
| 4.2                                    | Preprocessing Data.....    | 34        |
| 4.2.1                                  | Cleansing .....            | 34        |
| 4.2.2                                  | Case Folding.....          | 34        |
| 4.2.3                                  | Normalisasi.....           | 35        |
| 4.2.4                                  | Tokenization .....         | 35        |
| 4.2.5                                  | Stopword Removal .....     | 35        |
| 4.2.6                                  | Stemming.....              | 36        |
| 4.2.7                                  | Visualisasi WordCloud..... | 36        |
| 4.3                                    | Pelabelan Data .....       | 38        |
| 4.4                                    | Split Data .....           | 43        |
| 4.5                                    | Pelatihan Model .....      | 44        |
| 4.5.1                                  | Tabel Perbandingan .....   | 49        |
| 4.6                                    | Evaluasi Model .....       | 49        |
| <b>BAB V KESIMPULAN DAN SARAN.....</b> |                            | <b>55</b> |
| 5.1                                    | Kesimpulan .....           | 55        |
| 5.2                                    | Saran .....                | 56        |
| <b>DAFTAR PUSTAKA.....</b>             |                            | <b>58</b> |
| <b>LAMPIRAN.....</b>                   |                            | <b>62</b> |

## DAFTAR TABEL

|                                                  |    |
|--------------------------------------------------|----|
| Tabel 2. 1 Tinjauan Penelitian .....             | 17 |
| Tabel 4. 1 Hasil Cleaning.....                   | 34 |
| Tabel 4. 2 Hasil CaseFolding.....                | 34 |
| Tabel 4. 3 Hasil Normalisasi.....                | 35 |
| Tabel 4. 4 Hasil Tokenization .....              | 35 |
| Tabel 4. 5 Hasil Stopword Removal .....          | 36 |
| Tabel 4. 6 Hasil Stemming.....                   | 36 |
| Tabel 4. 7 Hasil setelah melakukan labeling..... | 42 |



## DAFTAR GAMBAR

|                                                          |    |
|----------------------------------------------------------|----|
| Gambar 3. 1 Tahap Penelitian.....                        | 24 |
| Gambar 4. 1 Tahap Crawling.....                          | 32 |
| Gambar 4. 2 Dataset.....                                 | 33 |
| Gambar 4. 3 Dataset.....                                 | 33 |
| Gambar 4. 4 Visualisasi WordCloud .....                  | 37 |
| Gambar 4. 5 Frekuensi Kata .....                         | 38 |
| Gambar 4. 6 Kamus kata Positive.....                     | 39 |
| Gambar 4. 7 Kamus Kata Negative .....                    | 40 |
| Gambar 4. 8 Distribusi kelas awal .....                  | 44 |
| Gambar 4. 9 Distribusi kelas setelah penyeimbangan ..... | 45 |
| Gambar 4. 10 Perbandingan antara setiap algoritma .....  | 46 |
| Gambar 4. 11 Perbandingan antara setiap algoritma .....  | 47 |
| Gambar 4. 12 perbandingan berdasarkan Accurasy.....      | 47 |
| Gambar 4. 13 Confusion Matrix SVM .....                  | 50 |
| Gambar 4. 14 Confusion Matrix Logistic Regression .....  | 51 |
| Gambar 4. 15 Confusion Matrix Naïve Bayes .....          | 52 |
| Gambar 4. 16 Perbandingan matrix evaluasi .....          | 53 |



## DAFTAR LAMPIRAN

|                                              |    |
|----------------------------------------------|----|
| Lampiran 1 Kartu Asensi .....                | 62 |
| Lampiran 2 Halaman Persetujuan .....         | 63 |
| Lampiran 3 Form Revisi Dosen Penguji.....    | 64 |
| Lampiran 4 Sertifikat BNSP .....             | 66 |
| Lampiran 5 Cek Turnitin .....                | 67 |
| Lampiran 6 Surat Pernyataan HAKI.....        | 68 |
| Lampiran 7 Curriculum Vitae .....            | 70 |
| Lampiran 8 Pernyataan Similarity Check ..... | 71 |

