

**ANALISIS SENTIMEN PELANGGAN PERUMDA GIRI TIRTA
GRESIK PADA IMBALANCE DATA DENGAN
METODE SMOTE DAN NAIVE BAYES**

SKRIPSI



Disusun Oleh :

Hanif Abdillah

210602052

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH GRESIK**

2025

KATA PENGANTAR

Dengan memanjatkan puji dan syukur ke hadirat Tuhan Yang Maha Esa, penulis mengucapkan terima kasih atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “ANALISIS SENTIMEN PELANGGAN PERUMDA GIRI TIRTA GRESIK PADA *IMBALANCE DATA* DENGAN METODE *SMOTE* DAN *NAIVE BAYES*”, sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Gresik. Skripsi ini bertujuan untuk mengidentifikasi persepsi pelanggan terhadap layanan Perumda Giri Tirta Gresik berdasarkan ulasan pada *Google Maps*, serta mengelompokkan sentimen tersebut ke dalam kategori positif, negatif, atau netral dengan memanfaatkan algoritma Naive Bayes sebagai metode klasifikasi.

Penulis menyadari bahwa penyusunan skripsi ini tidak akan dapat terselesaikan tanpa adanya dukungan, bantuan, bimbingan, serta nasihat dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya dan setulus-tulusnya kepada:

1. Bapak Prof. Dr. Khoirul Anwar, M.Pd, selaku Rektor Universitas Muhammadiyah Gresik, atas kepemimpinan dan dukungannya terhadap lingkungan akademik.
2. Bapak Harunur Rosyid, S.T., M.Kom., Ph.D, selaku Dekan Fakultas Teknik Universitas Muhammadiyah Gresik, atas arahan, dukungan, dan motivasi yang telah diberikan kepada penulis selama menjalani proses pembelajaran hingga penyusunan skripsi ini.
3. Ibu Henny Dwi Bhakti, S.Si., M.Si, selaku Kepala Program Studi Teknik Informatika dan juga Dosen Wali, atas arahan, bimbingan, dan dukungan yang diberikan kepada penulis.
4. Ibu Umi Chotijah, S.Kom., M.Kom dan Bapak Harunur Rosyid, S.T., M.Kom., Ph.D., selaku pembimbing yang telah memberikan bimbingan, masukan, dan arahan yang sangat berarti selama penyusunan proposal skripsi ini.

5. Seluruh dosen Program Studi Teknik Informatika Universitas Muhammadiyah Gresik, yang telah memberikan ilmu dan wawasan yang sangat berharga selama masa perkuliahan.
6. Keluarga tercinta, khususnya kepada kedua orang tua penulis, Bapak Abdul Hadi dan Ibu Nur Rokhah, serta kakak-kakak penulis, Kak Fathul Farikh dan Kak Dwi Rinayatus Sa'adah, yang senantiasa memberikan doa, kasih sayang, serta dukungan moral maupun material, dan telah menjadi sumber inspirasi serta motivasi yang tiada henti selama proses penyusunan proposal ini.
7. Rekan-rekan seperjuangan di Teknik Informatika Angkatan 2021 Universitas Muhammadiyah Gresik, yang telah memberikan bantuan, semangat, dan kerja sama selama proses penyusunan proposal skripsi ini.
8. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan dukungan secara langsung maupun tidak langsung dalam penyusunan skripsi ini. Semoga segala bantuan dan kebaikan yang telah diberikan mendapat balasan yang setimpal dari Tuhan Yang Maha Esa.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis terbuka terhadap segala kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat dan menjadi kontribusi positif dalam pengembangan ilmu pengetahuan.

Gresik, 16 April 2025

Hanif Abdillah
210602052

DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN	ii
KATA PENGANTAR.....	iii
LEMBAR PERNYATAAN KEASLIAN SKRIPSI.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
ABSTRAK.....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	5
BAB II LANDASAN TEORI.....	6
2.1 Analisis Sentimen	6
2.2 <i>Text Mining</i>	7
2.3 Pelanggan	8
2.4 <i>Google Maps</i>	8
2.5 <i>Text Preprocessing</i>	9
2.6 Pembobotan Kata <i>TF-IDF</i>	10
2.7 <i>Imbalance Data</i>	11
2.8 <i>SMOTE</i>	12
2.9 <i>Naive Bayes Classifier</i>	13
2.10 Evaluasi.....	15
2.11 Penelitian Terkait.....	17
BAB III ANALISIS DAN PERANCANGAN SISTEM.....	25
3.1 Analisis Sistem	25

3.2	Hasil analisis Sistem	25
3.3	Representasi Model	47
3.4	Perancangan system.....	50
3.4.1	Diagram Konteks	50
3.4.2	Diagram Jenjang	51
3.4.3	<i>Data Flow Diagram</i>	52
3.5	Perancangan basis data.....	54
3.5.1	Tabel <i>user</i>	54
3.5.2	Tabel <i>pdam</i>	54
3.5.3	Tabel <i>preprocessing</i>	55
3.5.4	Tabel Split Data.....	55
3.5.5	Tabel klasifikasi	55
3.6	Perancangan Antarmuka	57
3.7	Perancangan pengujian system.....	62
3.8	Spesifikasi Pengembangan system	63
3.9	Skenario Pengujian Dan Evaluasi Model.....	64
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM		65
4.1	Implementasi Sistem.....	65
4.1.1	Pelabelan Awal	65
4.1.2	Proses Preprocessing	65
4.1.3	Split Data.....	68
4.1.4	Ekstraksi Fitur Dengan TF-IDF.....	69
4.1.5	Proses SMOTE	71
4.1.6	Klasifikasi <i>Naive Bayes</i>	73
4.2	Pengujian Sistem	75
4.2.1	Halaman Login	76
4.2.2	Halaman Beranda.....	77
4.2.3	Halaman Dataset	77
4.2.4	Halaman Preprocessing	78
4.2.5	Halaman Split Data	78
4.2.6	Halaman <i>TF-IDF</i>	79
4.2.7	Halaman <i>SMOTE</i>	79

4.2.8	Halaman Evaluasi	80
4.2.9	Halaman Visualisasi.....	81
4.3	Analisa Hasil Pengujian Sistem.....	84
4.3.1	Evaluasi Pengujian Sistem	84
4.3.2	Evaluasi Pengujian Metode <i>Naive Bayes</i>	85
BAB V PENUTUP.....		89
5.1	Kesimpulan.....	89
5.2	Saran	89
DAFTAR PUSTAKA		90
LAMPIRAN		94



DAFTAR GAMBAR

Gambar 3. 1 Metodologi Penelitian.....	26
Gambar 3. 2 Alur Scraping Dataset.....	28
Gambar 3. 3 Distribusi Sentimen Terhadap PDAM Giri Tirta Gresik.....	30
Gambar 3. 4 Alur Preprocessing	31
Gambar 3. 5 Alur Proses Casefolding.....	32
Gambar 3. 6 Alur Proses Cleaning	33
Gambar 3. 7 Alur Proses Tokenizing.....	35
Gambar 3. 8 Alur Proses Normalisasi.....	37
Gambar 3. 9 Alur Proses Stopword	39
Gambar 3. 10 Alur Proses Stemming	40
Gambar 3. 11 Flowchart Algoritma Naive Bayes.....	46
Gambar 3. 12 Diagram Konteks Sistem.....	50
Gambar 3. 13 Diagram Jenjang Sistem	51
Gambar 3. 14 Data Flow Diagram Level 1 Sistem.....	52
Gambar 3. 15 ERD (Entity Relationship Diagram).....	56
Gambar 3. 16 Perancangan Antarmuka Halaman Login	57
Gambar 3. 17 Prancangan Antarmuka Halaman Beranda.....	58
Gambar 3. 18 Perancangan Antarmuka Halaman Dataset	58
Gambar 3. 19 Perancangan Antarmuka Halaman Preprocessing	59
Gambar 3. 20 Perancangan Antarmuka Halaman Split Data	59
Gambar 3. 21 Perancangan Antarmuka Halaman TF-IDF.....	60
Gambar 3. 22 Perancangan Antarmuka Halaman Evaluasi	61
Gambar 3. 23 Peancangan Antarmuka Halaman Visualisasi	61
Gambar 4. 1 Halaman Login.....	76
Gambar 4. 2 Halaman Beranda	77
Gambar 4. 3 Halaman Dataset.....	77
Gambar 4. 4 Halaman Preprocessing.....	78
Gambar 4. 5 Halaman Split Data.....	78
Gambar 4. 6 Halaman TF-IDF	79
Gambar 4. 7 Halaman SMOTE.....	79
Gambar 4. 8 Halaman Evaluasi Perbandingan Probabilitas.....	80

Gambar 4. 9 Halaman Evaluasi Hasil Klasifikasi Naive Bayes.....	80
Gambar 4. 10 Halaman Evaluasi Performa klasifikasi	81
Gambar 4. 11 Halaman Visualisasi.....	82
Gambar 4. 12 Grafik Distribusi Awal Pelabelan	82
Gambar 4. 13 Grafik Perbandingan Sentimen Tanpa SMOTE	83
Gambar 4. 14 Grafik Perbandingan Sentimen Dengan SMOTE.....	83



DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	15
Tabel 2. 2 Penelitian Sejenis	17
Tabel 3. 1 Contoh hasil scraping menggunakan Instant Data Scraper	29
Tabel 3. 2 Contoh Hasil Proses Casefolding	32
Tabel 3. 3 Contoh Hasil Proses Cleaning.....	34
Tabel 3. 4 Contoh Hasil Tokenizing	35
Tabel 3. 5 Contoh Hasil Normalisasi	37
Tabel 3. 6 Contoh Hasil Stopword	39
Tabel 3. 7 Contoh Hasil Stemming.....	41
Tabel 3. 8 Contoh Ulasan Pelanggan Dari Data Uji.....	42
Tabel 3. 9 Nilai Bobot Masing-Masing Kata	42
Tabel 3. 10 Hasil <i>Oversampling SMOTE</i>	45
Tabel 3. 11 Hasil Perhitungan Probabilitas Kata Dalam Setiap Kategori.....	48
Tabel 3. 12 Hasil Data Uji.....	50
Tabel 3. 13 Tabel user	54
Tabel 3. 14 Tabel pdam.....	54
Tabel 3. 15 Tabel Preprocessing	55
Tabel 3. 16 Tabel Split Data.....	55
Tabel 3. 17 Tabel Klasifikasi.....	56
Tabel 3. 18 Perancangan Pengujian Sistem	62
Tabel 3. 19 Tabel Pembagian Dataset.....	64
Tabel 4. 1 Evaluasi Pengujian Sistem.....	84
Tabel 4. 2 Confusion Matrix Pada Data Uji.....	86
Tabel 4. 3 Hasil Evaluasi Model Sebelum dan Sesudah SMOTE.....	87

ABSTRAK

Dalam pengelolaan layanan publik seperti penyediaan air bersih, memahami persepsi dan kepuasan pelanggan menjadi faktor penting untuk peningkatan kualitas layanan. Perumda Giri Tirta Gresik merupakan perusahaan daerah yang menyediakan layanan air bersih kepada masyarakat, dan ulasan pelanggan di Google Maps dapat menjadi sumber informasi untuk menganalisis sentimen masyarakat. Penelitian ini melakukan analisis sentimen terhadap 571 ulasan menggunakan algoritma Naive Bayes. Data diperoleh melalui scraping dari Google Maps dan diproses melalui tahapan preprocessing, seperti casefolding, cleansing, tokenizing, normalisasi, stopword, dan stemming. Setelah itu, dilakukan pembobotan TF-IDF dan penyeimbangan data dengan teknik SMOTE untuk mengatasi ketidakseimbangan kelas. Sentimen diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Evaluasi dilakukan dengan membandingkan performa model sebelum dan sesudah SMOTE menggunakan akurasi, presisi, recall, dan F1-score. Hasil klasifikasi divisualisasikan dalam bentuk grafik untuk memudahkan interpretasi opini pelanggan. Berdasarkan hasil pengujian, metode Naive Bayes mampu mengelompokkan sentimen dengan cukup baik. Penerapan SMOTE juga terbukti meningkatkan kemampuan model dalam mempelajari pola sentimen dari kelas minoritas. Penelitian ini menekankan pentingnya preprocessing yang tepat serta penggunaan teknik penyeimbangan data dalam pengembangan sistem analisis sentimen yang andal dan informatif.

Kata Kunci : Analisi Sentimen, Google Maps, SMOTE, Klasifikasi

ABSTRACT

In the management of public services such as the provision of clean water, understanding customer perception and satisfaction is an important factor for improving service quality. Perumda Giri Tirta Gresik is a regional company that provides clean water services to the community, and customer reviews on Google Maps can be a source of information to analyze community sentiment. This study conducted a sentiment analysis of 571 reviews using the Naive Bayes algorithm. Data is obtained through scraping from Google Maps and processed through preprocessing stages, such as casefolding, cleansing, tokenizing, normalization, stopwording, and stemming. After that, TF-IDF weighting and data balancing were carried out with the SMOTE technique to overcome class imbalances. Sentiment is classified into three categories, which are positive, negative, and neutral. Evaluation was carried out by comparing the performance of the model before and after SMOTE using accuracy, precision, recall, and F1-score. The results of the classification are visualized in the form of a graph to facilitate the interpretation of customer opinions. Based on the test results, the Naive Bayes method is able to group sentiment quite well. The application of SMOTE has also been shown to improve the model's ability to study sentiment patterns from minority classes. This study emphasizes the importance of proper preprocessing as well as the use of data balancing techniques in the development of a reliable and informative sentiment analysis system.

Keywords : Sentiment Analysis, Google Maps, SMOTE, Classification