



**PENERAPAN *ADAPTIVE SYNTHETIC MULTINOMIAL NAÏVE*
BAYES UNTUK MENGATASI *IMBALANCED CLASS DATA* PADA
KLASIFIKASI SENTIMEN KENAIKAN BBM**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

ISMATULLAH

B2A019002

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG**

2022

PERSETUJUAN PEMBIMBING

Skripsi dengan Judul “Penerapan *Adaptive Synthetic Multinomial Naive Bayes* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Sentimen Kenaikan BBM” yang disusun oleh:

Nama : Ismatullah
NIM : B2A019002
Program Studi : S1 Statistika
Judul Skripsi : Penerapan *Adaptive Synthetic Multinomial Naive Bayes* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Sentimen Kenaikan BBM

telah disetujui oleh pembimbing pada tanggal: 21 Maret 2023.

Pembimbing Utama



Fatkhuokhman Fauzi, M.Stat
NIK 28.6.1026.468

Pembimbing Pendamping



Indah Manfaati Nur, S.Si, M.Si
NIK 28.6.1026.221

Mengetahui

Ketua Program Studi



Indah Manfaati Nur, S.Si, M.Si
NIK 28.6.1026.221

PENGESAHAN KELULUSAN

Skripsi dengan Judul “Penerapan *Adaptive Synthetic Multinomial Naïve Bayes* untuk Mengatasi *Imbalanced Class Data* pada Klasifikasi Sentimen Kenaikan BBM” yang disusun oleh:

Nama : Ismatullah
NIM : B2A019002
Program Studi : S1 Statistika

telah diujikan dalam Sidang Panitia Ujian Skripsi Program Sarjana, Universitas Muhammadiyah Semarang pada tanggal : 21 Maret 2023.

Panitia Ujian

Ketua Tim Penguji



M. Al Haris, S.Si., M.Si
NIK. 28.6.1026.442

Anggota Tim Penguji I



Prizka Rismawati Arum, M.Stat
NIK. 28.6.1026.441

Anggota Tim Penguji II



Fatkhurokhan Fauzi, M.Stat
NIK. 28.6.1026.468

Anggota Tim Penguji III



Indah Manfaati Nur, S.Si., M.Si
NIK. 28.6.1026.221

Mengetahui

Ketua Program Studi



Indah Manfaati Nur, S.Si., M.Si
NIK. 28.6.1026.221

PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa :

1. Karya tulis saya, skripsi ini, adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, magister, dan/atau doktor), baik di Universitas Muhammadiyah Semarang maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing dan masukan Tim Penguji.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi ini.

Semarang, 24 Maret 2023

Yang membuat pernyataan,



Ismatullah
NIM. B2A019002

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Segala puji dan syukur penulis ucapkan kehadiran Allah SWT dan mengharapkan ridho yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan *Adaptive Synthetic Multinomial Naïve Bayes* untuk mengatasi *Imbalanced Class Data* pada Klasifikasi Sentimen Kenaikan BBM”. Skripsi ini disusun sebagai salah satu persyaratan meraih gelar Sarjana Statistika pada Program Studi Statistika Universitas Muhammadiyah Semarang. Shalawat dan salam disampaikan kepada junjungan alam Nabi Muhammad SAW, mudah-mudahan kita semua mendapatkan safaatNya di yaumul akhir nanti, Amin.

Penelitian ini diangkat sebagai upaya untuk merealisasikan pembelajaran yang berpusat pada siswa (*Student Centered*), keterlibatan siswa dalam proses mempelajari prinsip dan konsep statistika, menumbuh kembangkan keterampilan proses dan memicu keaktifan siswa dalam proses pembelajaran.

Penulis menyadari sepenuhnya, bahwa dalam penyelesaian skripsi ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Untuk itu pada kesempatan ini penulis ingin menyampaikan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Ibu Dr. Eny Winaryati, M.Pd, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang.
2. Ibu Indah Manfaati Nur, S.Si., M.Si., selaku Ketua Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang serta selaku dosen pembimbing II yang telah meluangkan waktu, pikiran, dukungan dan bimbingannya kepada penulis dalam penyusunan skripsi ini.
3. Bapak Fatkhurohman Fauzi, M.Stat, selaku dosen pembimbing I yang telah meluangkan waktu, pikiran, dukungan dan bimbingannya kepada penulis dalam penyusunan skripsi ini.
4. Bapak M. Al Haris, S.Si., M.Si., selaku dosen penguji I atas bimbingan, koreksi, dan arahan yang diberikan.

5. Ibu Prizka Rismawati Arum, M.Stat, selaku dosen penguji II atas bimbingan, koreksi, dan arahan yang diberikan.
6. Orang tua penulis yang tercinta dan tersayang. Merekalah orang-orang hebat yang selalu mengalirkan do'a serta memberikan semangat dan motivasi yang tinggi.
7. Mahasiswa dengan NIM B2A019024 yang telah memberikan motivasi dan dukungan kepada penulis dalam penyusunan skripsi ini.
8. Seluruh pihak yang turut membantu dan mendukung penulis selama penyusunan skripsi ini, yang tak mungkin penulis sebutkan satu per satu.

Penulis menyadari akan segala keterbatasan dan kekurangan dari isi maupun tulisan skripsi ini. Oleh karena itu, kritik dan saran yang bersifat membangun dari semua pihak akan diterima dengan senang hati. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi bagi pengembangan pembelajaran statistika.

Wassalamu'alaikum Wr. Wb.

Semarang, 23 Maret 2023



Ismatullah

DAFTAR ISI

Halaman	
LEMBAR JUDUL.....	i
LAMAN PERSETUJUAN.....	ii
LAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.4.1 Manfaat Teoritis	4
1.4.2 Manfaat Praktis.....	4
1.5 Batasan Penelitian	5
BAB II TINJAUAN PUSTAKA	
2.1 <i>Text Mining</i>	6
2.2 Analisis Sentimen.....	6
2.3 Algoritma <i>Multinomial Naïve Bayes</i>	6
2.4 <i>Text Preprocessing</i>	8
2.4.1 <i>Case Folding</i>	8
2.4.2 <i>Word Normalization</i>	8
2.4.3 <i>Cleansing</i>	9
2.4.4 <i>Filtering/Stopword Removal</i>	9

2.4.5 <i>Stemming</i>	9
2.4.6 <i>Tokenization</i>	9
2.5 <i>Feature Extraction</i> dengan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	10
2.5.1 TF-IDF	10
2.6 <i>Feature Selection</i>	11
2.7 ADASYN	11
BAB III METODE PENELITIAN	
3.1 Sumber Data	15
3.2 Variabel Penelitian dan Struktur Data	16
3.3 Tahapan Penelitian	16
3.4 <i>Flowchart Text Preprocessing</i>	23
3.5 <i>Flowchart ADASYN</i>	24
3.6 <i>Flowchart Multinomial Naïve Bayes</i>	25
3.7 Alur Penelitian.....	26
BAB IV HASIL DAN PEMBAHASAN	
4.1 <i>Text Preprocessing</i>	27
4.2 Gambaran Analisis Sentimen Pada Setiap Provinsi Yang Ada Di Indonesia Mengenai Kenaikan BBM	29
4.2.1 Peta Sebaran Sentimen	29
4.2.2 <i>Wordcloud</i> Sentimen Kenaikan BBM.....	33
4.3 Hasil Pengujian <i>Multinomial Naïve Bayes</i> Setelah Diterapkan ADASYN.....	36
4.5 Prediksi Sentimen.....	42
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	44
5.2 Saran	45
DAFTAR PUSTAKA.....	46
LAMPIRAN	49

DAFTAR TABEL

Tabel	Halaman
3.1 Sampel <i>Scraping</i> Data dari <i>Twitter</i>	15
3.2 Struktur Data	16
3.3 Rumus <i>Multiclass Confusion Matrix</i>	21
4.1 <i>Case Folding</i>	27
4.2 <i>Word Normalization</i>	27
4.3 <i>Cleansing</i>	28
4.4 <i>Filtering/Stopwords Removal</i>	28
4.5 <i>Stemming</i>	28
4.6 <i>Tokenization</i>	29
4.7 Kata dengan Frekuensi Terbanyak (Persentase).....	34
4.8 Sampel Data <i>Training</i> untuk Klasifikasi	38
4.9 Sampel Data <i>Testing</i> untuk Klasifikasi	38
4.10 Tokenisasi dan Frekuensi Kemunculan Kata Tanpa <i>Preprocessing</i>	38
4.11 Probabilitas Kemunculan Kata dengan Polaritas p	39
4.12 Probabilitas Munculnya Dokumen d dengan Polaritas p	40
4.13 Probabilitas Data <i>Testing</i> Masuk ke Polaritas p	40
4.14 Tingkat Akurasi Setiap Provinsi.....	41

DAFTAR GAMBAR

Gambar	Halaman
3.1 <i>Flowchart Text Preprocessing</i>	23
3.2 <i>Flowchart ADASYN</i>	24
3.3 <i>Flowchart Multinomial Naïve Bayes</i>	25
3.4 Alur Penelitian.....	26
4.1 Distribusi Sentimen Positif.....	30
4.2 Distribusi Sentimen Netral	31
4.3 Distribusi Sentimen Negatif	32
4.4 Sampel <i>Wordcloud</i> Sentimen Kenaikan BBM.....	34
4.5 Sampel Jumlah Sentimen Sebelum dan Setelah ADASYN	37
4.6 <i>Confusion Matrix</i> Provinsi Aceh	40
4.7 <i>Form Input</i> Sentimen.....	42
4.8 Hasil <i>Output</i> Prediksi Sentimen	42

DAFTAR LAMPIRAN

Lampiran	Halaman
1. <i>Wordcloud</i> Sentimen	49
2 Syntax <i>Scraping</i> Data Berdasarkan Latitude dan Longitude.....	55
3 Syntax Skripsi.....	56
4 Syntax Visualisasi <i>Choropleth Map</i>	68
5 Syntax Visualisasi <i>Bar Plot</i>	69
6 Visualisasi Jumlah Sentimen Sebelum ADASYN	71
7. <i>Confusion Matrix</i> Setiap Provinsi	73
8. Contoh Perhitungan <i>Multinomial Naïve Bayes</i> Secara Manual.....	82
9. <i>Oversampling</i> dengan ADASYN	83
10 Data IP-TIK tahun 2020	84