

TUGAS AKHIR

**PENERAPAN N-GRAM TERHADAP ANALISIS SENTIMEN
PADA FITUR BELANJA TIKTOKSHOP MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR***



RISKA DWI AYU LESTARI

1910651020

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

2022

TUGAS AKHIR

PENERAPAN N-GRAM TERHADAP ANALISIS SENTIMEN PADA FITUR BELANJA TIKTOKSHOP MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Disusun Untuk Melengkapi Dan Memenuhi Syarat Kelulusan Guna Meraih Gelar
Sarjana Komputer Teknik Informatika Universitas Muhammadiyah Jember



RISKA DWI AYU LESTARI

1910651020

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

2022

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

**PENERAPAN N-GRAM TERHADAP ANALISIS SENTIMEN PADA
FITUR BELANJA TIKTOKSHOP MENGGUNAKAN ALGORITMA
*K-NEAREST NEIGHBOR***

Oleh :

RISKA DWI AYU LESTARI

1910651020

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Pembimbing I

Pembimbing II


Dr. Bagus Setya Kintyarna S.T., M.Kom
NIDN. 0729017904


Moh. Dasuki M.Kom
NIDN. 0722109103

HALAMAN PENGESAHAN

PENERAPAN N-GRAM TERHADAP ANALISIS SENTIMEN PADA FITUR BELANJA TIKTOKSHOP MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Oleh :

RISKA DWI AYU LESTARI
1910651020

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 13 November sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Penguji I



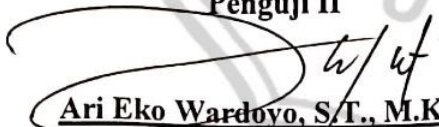
Wiwik Suharso, S.Kom., M.Kom
NIDN. 0006097601

Pembimbing I




Dr. Bagus Setya Rintvarna, S.T., M.Kom
NIDN. 0729017904

Penguji II



Ari Eko Wardoyo, S.T., M.Kom
NIDN. 0014027501

Pembimbing II



Moh. Dasuki, M.Kom
NIDN. 0722109103

Mengesahkan,

Dekan Fakultas Teknik



Dr. H. Nanang Saiful Rizal, S.T., M.T., IPM
NIDN. 0705047806

Mengetahui,

Ketua Program Studi Teknik Informatika



Ari Eko Wardoyo, S.T., M.Kom
NIDN. 0014027501

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Riska Dwi Ayu Lestari
NIM : 1910651020
Program Studi : Teknik Informatika
Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “PENERAPAN N-GRAM TERHADAP ANALISIS SENTIMEN PADA FITUR BELANJA TIKTOKSHOP MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*” adalah ASLI dan BELUM PERNAH dibuat orang lain, kecuali yang diacu dalam Daftar Pustaka pada Tugas Akhir ini.

Demikian Surat Pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak sesuai, penulis bersedia mendapat sanksi dari akademik.

Jember, 12 Desember 2022



Riska Dwi Ayu Lestari
Nim. 1910651020

KATA PENGANTAR

Alhamdulillah rabbil 'alamin segala puji dan syukur kami panjatkan atas kehadiran Allah dengan segala limpahan rahmat, nikmat dan karunia-Nya sehingga penulis bisa bertahan menahan segala gangguan dan diberikan kemudahan dalam menyelesaikan tugas akhir ini dengan judul “Penerapan N-Gram Terhadap Analisis Sentimen Pada Fitur Belanja Tiktokshop Menggunakan Algoritma *K-Nearest Neighbor*”

Laporan tugas akhir ini disusun sebagai salah satu syarat mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Muhammadiyah Jember. Adapun dalam penyelesaian tugas akhir ini, banyak pihak yang ikut serta dalam memberi dukungan, saran serta bimbingan. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Allah SWT atas segala nikmat dan karunia-Nya yang telah memberikan kemudahan dan kelancaran dalam meraih gelar Sarjana Komputer.
2. Bapak Dr. Nanang Saiful Rizal, S.T., M.T., IPM selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Ari Eko Wardoyo, S.T., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
4. Bapak Dr. Bagus Setya Rintyarna S.T, M.Kom. selaku dosen pembimbing I dan Bapak Moh. Dasuki M.Kom. selaku dosen pembimbing II yang telah meluangkan waktu untuk memberi arahan dan membimbing penulis dalam menyelesaikan tugas akhir ini.
5. Bapak Wiwik Suharso, S.Kom, M.Kom. selaku dosen penguji I dan Bapak Ari Eko Wardoyo, S.T., M.Kom. selaku dosen penguji II yang telah memberikan saran, kritik dan masukan dalam penyelesaian tugas akhir ini.
6. Kepada Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmunya kepada penulis selama menempuh perkuliahan di Universitas Muhammadiyah Jember.

7. Kepada orang tua, adik, serta keluarga dan kerabat yang selalu mendoakan, memberikan kasih sayang, memberikan semangat, *support*, nasihat sehingga penulis dapat menyelesaikan tugas akhir ini
8. Kepada Tiara, Ardiyah, Citra dan Retha teman seperjuangan dalam menempuh studi mulai dari awal kuliah sampai akhir, juga teman-teman dari seluruh mahasiswa Prodi Teknik Informatika khususnya angkatan 2019 yang telah berproses bersama selama di masa perkuliahan.
9. Semua pihak yang tidak dapat disebutkan satu per satu yang telah turut mendoakan dan mendukung penulis selama proses penyelesaian tugas akhir.
10. *Last but not least. I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for no days off, I wanna thank me for never quitting, I wanna thank me for always being a giver and tryna give more than I receive.*

Akhirnya, dengan segala kerendahan hati, penulis menyadari bahwa karya ilmiah skripsi ini masih jauh dari kata sempurna. Karena itu, penulis berharap kepada pembaca untuk memberikan kritik dan saran yang membangun demi kesempurnaan penulisan. Penulis berharap semoga karya yang sederhana ini dapat bermanfaat bagi semua pihak. Aamiin ya Rabbal ‘Alamin

MOTTO

“Only you can change your life, nobody else can do it for you.”

Orang lain gak akan bisa paham *struggle* dan masa sulitnya kita, yang mereka ingin tahu hanya bagian *success stories*. Berjuanglah untuk diri sendiri walaupun gak ada yang tepuk tangan, kelak diri kita di masa depan akan sangat bangga dengan apa yang kita perjuangkan hari ini.

(Fardi Yandi)

“Ketika dalam kesulitan orang-orang meninggalkanmu, bisa jadi Allah sendirilah yang akan mengurusmu”

(Imam Syafi’i)

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
LEMBAR PERNYATAAN KEASLIAN	iv
KATA PENGANTAR.....	vii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Manfaat.....	4
1.5 Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	6
2.1 Analisis Sentimen	6
2.2 TikTok Shop	6
2.3 <i>Text Mining</i>	8
2.4 Twitter	10
2.5 <i>N-Grams Sequences</i>	11
2.6 <i>Term Weighting TF-IDF</i>	12
2.7 <i>Nearest Neighbor Classifier</i>	14
2.7.1 Algoritma K-NN	14
2.7.2 Karakteristik <i>K-Nearest Neighbor Classifier</i>	16
2.8 <i>Imbalanced Dataset</i>	17
2.9 <i>K-Fold Cross Validation</i>	17
2.10 <i>Confusion Matrix</i>	19
2.11 <i>Python</i>	20
2.12 Penelitian Terdahulu	21
BAB III METODOLOGI PENELITIAN	23

3.1 Tahapan Penelitian	23
3.2 Uraian Metodologi Penelitian	24
3.2.1 <i>Crawling Data</i>	24
3.2.2 Pelabelan Data	24
3.2.3 <i>Text Preprocessing</i>	25
3.2.4 <i>N-Gram Sequences</i>	30
3.2.5 <i>Term Weighting TF-IDF</i>	31
3.2.6 <i>Random Under Sampling (RUS)</i>	35
3.2.7 Partisi Data	35
3.2.8 Pemodelan Data dengan <i>K-Fold Cross Validation</i>	35
3.2.9 Implementasi <i>K-Nearest Neighbor</i>	36
3.2.10 Evaluasi <i>Confusion Matrix</i>	45
BAB IV IMPLEMENTASI DAN HASIL	47
4.1 Persiapan Data	47
4.1.1 <i>Crawling Dataset</i>	47
4.1.2 Pelabelan Data	48
4.1.3 <i>Preprocessing Data</i>	50
4.1.4 Pembobotan TF-IDF	52
4.1.5 <i>Undersampling Data</i>	54
4.1.6 Partisi Data	54
4.2 <i>K-Fold Cross Validation</i>	55
4.3 Implementasi <i>K-Nearest Neighbor</i> dan N-Gram.....	56
4.3.1 Pemodelan K-NN.....	56
4.3.2 Pengujian Validasi Data	58
4.4 Implementasi K-NN dan N-Gram dengan <i>Undersampling</i>	61
4.4.1 Pemodelan K-NN Terhadap Hasil <i>Undersampling</i>	61
4.4.2 Pengujian Validasi <i>Undersampling</i>	63
4.5 Hasil	66
4.5.1 Hasil Analisis K-NN	66
4.5.2 Hasil Klasifikasi Sistem	68
4.5.3 Hasil Analisis Visual N-Gram.....	69

BAB V KESIMPULAN DAN SARAN.....	71
5.1 Kesimpulan.....	71
5.2 Saran.....	71
DAFTAR PUSTAKA.....	72
LAMPIRAN.....	76



DAFTAR GAMBAR

Gambar 2.1 Cara Penggunaan Fitur Belanja Tiktokshop	7
Gambar 2.2 Diagram Alur <i>Text Preprocessing</i>	9
Gambar 2.3 N-Grams Pengolahan Kata.....	11
Gambar 2.4 Ilustrasi K-NN (Guifa Teng et al., 2006).....	15
Gambar 2.5 Diagram Alur K-NN.....	16
Gambar 2.6 <i>K-Fold Cross Validation</i> (Wikipedia, 2017).....	18
Gambar 3.1 Alur Tahapan Penelitian	23
Gambar 3.2 Contoh Hasil <i>Crawling Data Tweet</i>	24
Gambar 3.3 Diagram Alir TF-IDF	32
Gambar 3.4 Partisi Data.....	35
Gambar 4.1 <i>Source Code</i> Proses <i>Crawling</i>	47
Gambar 4.2 Potongan Data Asli Hasil <i>Crawling</i>	48
Gambar 4.3 Hasil Pemetaan Klasifikasi Sentimen	48
Gambar 4.4 Hasil Pelabelan Sentimen	49
Gambar 4.5 Partisi Data Awal (Sebelum <i>Undersampling</i>).....	54
Gambar 4.6 Partisi Data <i>Undersampling</i>	55
Gambar 4.7 Pengelompokan <i>10-Fold Cross Validation</i>	56
Gambar 4.8 Hasil <i>Confusion Matrix</i> Validasi <i>Unigram</i>	59
Gambar 4.9 Hasil <i>Confusion Matrix</i> Validasi <i>Bigram</i>	60
Gambar 4.10 Hasil <i>Confusion Matrix</i> Validasi <i>Trigram</i>	60
Gambar 4.11 Hasil <i>Confusion Matrix</i> Validasi <i>Unigram Undersampling</i>	63
Gambar 4.12 Hasil <i>Confusion Matrix</i> Validasi <i>Bigram Undersampling</i>	64
Gambar 4.13 Hasil <i>Confusion Matrix</i> Validasi <i>Trigram Undersampling</i>	65
Gambar 4.14 Grafik Analisis Kesimpulan Hasil Uji Validasi.....	67
Gambar 4.15 Visualisasi <i>Unigram</i> Positif (Kiri), Negatif (Kanan).....	69
Gambar 4.16 Visualisasi <i>Bigram</i> Positif (Kiri), Negatif (Kanan)	69
Gambar 4.17 Visualisasi <i>Trigram</i> Positif (Kiri), Negatif (Kanan)	70

DAFTAR TABEL

Tabel 2.1 Contoh Potongan <i>Word N-Grams</i>	12
Tabel 2.2 Dua Kelas <i>Confusion Matrix</i>	19
Tabel 3.1 Contoh Hasil Pelabelan Data	25
Tabel 3.2 Penerapan Hasil <i>Cleansing</i>	26
Tabel 3.3 Penerapan Hasil <i>Case Folding</i>	27
Tabel 3.4 Penerapan Hasil <i>Stopword Removal</i>	28
Tabel 3.5 Penerapan Hasil <i>Stemming</i>	28
Tabel 3.6 Penerapan Hasil <i>Tokenization</i>	29
Tabel 3.7 Implementasi <i>Unigram</i>	30
Tabel 3.8 Implementasi <i>Bigram</i>	31
Tabel 3.9 Implementasi <i>Trigram</i>	31
Tabel 3.10 Perhitungan TF-IDF <i>Unigram</i>	33
Tabel 3.11 Perhitungan TF-IDF <i>Bigram</i>	33
Tabel 3.12 Perhitungan TF-IDF <i>Trigram</i>	34
Tabel 3.13 Perkalian Skalar dan Panjang Vektor <i>Unigram</i>	38
Tabel 3.14 Hasil <i>Cosine Similarity Unigram</i>	39
Tabel 3.15 Perkalian Skalar dan Panjang Vektor <i>Bigram</i>	40
Tabel 3.16 Hasil <i>Cosine Similarity Bigram</i>	41
Tabel 3.17 Perkalian Skalar dan Panjang Vektor <i>Trigram</i>	41
Tabel 3.18 Hasil <i>Cosine Similarity Trigram</i>	43
Tabel 3.19 Perangkingan <i>Cosine Similarity</i>	44
Tabel 3.20 Contoh Tabel <i>Confusion Matrix</i>	46
Tabel 4.1 Hasil <i>Preprocessing</i> Sistem	50
Tabel 4.2 Hasil Pembobotan TF-IDF	52
Tabel 4.3 Komposisi Data Awal dan Setelah <i>Undersampling</i>	54
Tabel 4.4 Hasil Uji K-Fold K-NN <i>Unigram</i>	57
Tabel 4.5 Hasil Uji K-Fold K-NN <i>Bigram</i>	57
Tabel 4.6 Hasil Uji K-Fold K-NN <i>Trigram</i>	58
Tabel 4.7 Hasil Uji K-Fold K-NN <i>Unigram Undersampling</i>	61
Tabel 4.8 Hasil Uji K-Fold K-NN <i>Bigram Undersampling</i>	62

Tabel 4.9 Hasil Uji K-Fold K-NN *Trigram Undersampling* 62
Tabel 4.10 Rekapitulasi Hasil Pemodelan dan Uji Validasi 66
Tabel 4.11 Hasil Klasifikasi pada Pengujian Validasi..... 68

