

TUGAS AKHIR

**PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI
SENTIMEN TEKS BAHASA INDONESIA DENGAN
*MACHINE LEARNING***



INDAH PURWANTI

20110007

**PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2024**

TUGAS AKHIR

**PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI
SENTIMEN TEKS BAHASA INDONESIA DENGAN
*MACHINE LEARNING***

***IMPLEMENTATION OF NORMALIZATION TEXT FOR
SENTIMENT ANALYSIS OF INDONESIAN TEXTS USING
MACHINE LEARNING***

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains Data



INDAH PURWANTI

20110007

PROGRAM STUDI SAINS DATA

FAKULTAS INFORMATIKA

INSTITUT TEKNOLOGI TELKOM PURWOKERTO

2024

HALAMAN PERSETUJUAN PEMBIMBING

**PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI
SENTIMEN TEKS BAHASA INDONESIA DENGAN
*MACHINE LEARNING***

***IMPLEMENTATION OF NORMALIZATION TEXT FOR
SENTIMENT ANALYSIS OF INDONESIAN TEXTS USING
MACHINE LEARNING***

Dipersiapkan dan Disusun Oleh

INDAH PURWANTI

20110007

Fakultas Informatika

Institut Teknologi Telkom Purwokerto

Pada Tanggal: 15 Mei 2024

Pembimbing utama,



(Siti Khomsah, S.Kom., M.Cs)
NIDN: 0517108101

Pembimbing Pendamping,



(Yohani Setiya Rafika Nur, S.Kom., M.Kom)
NIDN: 0627099501

HALAMAN PENGESAHAN

**PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI
SENTIMEN TEKS BAHASA INDONESIA DENGAN
*MACHINE LEARNING***

***IMPLEMENTATION OF NORMALIZATION TEXT FOR
SENTIMENT ANALYSIS OF INDONESIAN TEXTS USING
MACHINE LEARNING***

Disusun Oleh
INDAH PURWANTI
20110007

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir pada
Hari, Tanggal 11 Juni 2024

Penguji I,



Sudianto, S.Pd., M.Kom
NIDN: 0605049301

Penguji II,



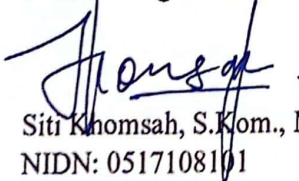
Nia Annisa Ferani Tanjung, S.Si., M.Sc
NIDN: 0630049203

Penguji III,



Mahazam Afrad, S.Kom., M.Kom
NIDN: 0624039305

Pembimbing utama,



Siti Khomsah, S.Kom., M.Cs
NIDN: 0517108101

Pembimbing Pendamping,



Yohani Setiya Rafika Nur, S.Kom., M.Kom
NIDN: 0627099501

Dekan



Auliya Buchanudin S.Si., M.Kom.
NIK. 19800008

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : Indah Purwanti
NIM : 20110007
Program Studi : S1 Sains Data

Menyatakan bahwa Tugas Akhir dengan judul berikut:

**PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI SENTIMEN
TEKS BAHASA INDONESIA DENGAN *MACHINE LEARNING***

Dosen Pembimbing Utama : Siti Khomsah, S.Kom M.Cs

Dosen Pembimbing Utama : Yohani Setiya Rafika Nur, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Purwokerto, 15 Mei 2024

Yang Menyatakan,



(Indah Purwanti)

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “PENERAPAN NORMALISASI TEKS UNTUK KLASIFIKASI SENTIMEN TEKS BAHASA INDONESIA DENGAN *MACHINE LEARNING*”. Shalawat dan Salam senantiasa tercurah kepada Baginda Rasulullah Muhammad SAW yang telah membawa manusia dari alam jahiliyah kepada alam yang terang menerang yang penuh ilmu pengetahuan seperti saat sekarang ini.

Dalam penyelesaian studi dan penulisan tugas akhir ini, penulis banyak memperoleh bantuan baik dari pengajaran, bimbingan, dan arahan dari berbagai pihak baik secara langsung maupun tidak langsung. Untuk itu penulis menyampaikan penghargaan dan terima kasih yang tak terhingga kepada :

1. Ibu Dr. Tenia Wahyuningrum S.Kom., M.Kom, selaku Rektor Institut Teknologi Telkom Purwokerto.
2. Bapak Auliya Burhanuddin S.Si., M.Kom, selaku Dekan Fakultas Informatika Institut Teknologi Telkom Purwokerto.
3. Ibu Siti Khomsah S.Kom., M.Cs. selaku Pembimbing Utama.
4. Ibu Yohani Setiya Rafika Nur, S.Kom., M.Kom. selaku Pembimbing Pendamping
5. Seluruh Dosen Prodi S1 Sains Data Institut Teknologi Telkom Purwokerto yang telah memberikan bekal ilmu dan pengetahuan sehingga dapat digunakan dalam penyusunan skripsi serta untuk bekal hidup di kehidupan mendatang.
6. Kedua orang tua penulis, Hadi Nur Waras dan Wati serta kakak dan adik yang telah memberikan bantuan dan dukungan baik secara moril maupun material.
7. Teman – teman Sains data Angkatan 20 yang telah mendukung dan berjuang bersama saya selama ini yang namanya tidak bisa saya sebutkan satu per satu.
8. Seluruh pihak yang telah membantu penulis dalam menyelesaikan tugas akhir ini. Penulis menyadari sepenuhnya bahwa hasil penelitian ini masih jauh dari sempurna, oleh karena itu penulis sangat mengharapkan kesediaan pembaca untuk memberikan kritik dan saran.

Akhirnya penulis menyampaikan permohonan maaf yang sebesar-besarnya kepada pembaca seandainya terdapat kesalahan – kesalahan di dalam skripsi ini dan penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan.

Purwokerto, 15 Mei 2024



Indah Purwanti

DAFTAR ISI

TUGAS AKHIR.....	ii
HALAMAN PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	ii
DAFTAR LAMPIRAN	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Pertanyaan Penelitian.....	4
1.4 Batasan Masalah	4
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Kajian Pustaka	6
2.2 Landasan Teori.....	14

2.2.1	Kata Slang.....	14
2.2.2	<i>Natural Language Processing</i>	15
2.2.3	Analisis Sentimen	15
2.2.4	<i>Exploratory Data Analysis (EDA)</i>	15
2.2.5	<i>IndoBERT</i>	16
2.2.6	<i>Text Preprocessing</i>	16
2.2.7	<i>Feature Extraction</i>	17
2.2.8	<i>Naïve Bayes Classifier</i>	18
2.2.9	<i>Decision Tree</i>	19
2.2.10	<i>Confusion Matrik</i>	20
BAB III		23
METODELOGI PENELITIAN		23
3.1	Subjek dan Objek Penelitian.....	23
3.2	Alat dan Bahan.....	23
3.2.1	Perangkat Keras	23
3.2.2	Perangkat Lunak	23
3.2.3	Bahan	23
3.3	Alur Penelitian	23
3.3.1	Studi Literatur	24
3.3.2	Pengumpulan Data	24
3.3.3	<i>Preprocessing</i> Awal	26
3.3.7	<i>Exploratory Data Analysis</i>	26
3.3.6	<i>Text Preprocessing</i>	27
3.3.7	Pelabelan <i>IndoBERT</i>	32
3.3.8	<i>Vektorisasi</i>	33

3.3.9	Model Klasifikasi.....	36
3.3.10	Evaluasi dan Analisis	40
BAB IV	42
HASIL DAN PEMBAHASAN	42
4.1.	Pelabelan Data	42
4.2.	Hasil <i>Exploratory Data Analysis</i>	42
4.3.	Pembuatan Kopus <i>Slang</i>	46
4.4.	<i>Text Preprocessing</i>	48
4.3.1.	Skenario Pertama	48
4.3.2.	Skenario Kedua dengan Normalisasi	50
4.5.	Hasil Vektorisasi	51
4.5.	Pemodelan.....	53
4.6.	Perbandingan Evaluasi Model	57
4.6.1.	<i>Naïve Bayes Classifier</i>	57
4.6.2.	<i>Decision Tree Classifier</i>	58
BAB IV	60
KESIMPULAN & SARAN	60
5.1.	Kesimpulan	60
1.2.	Saran	61
Daftar Pustaka	62
LAMPIRAN	66

DAFTAR GAMBAR

Gambar 2. 1 Decision Tree	20
Gambar 2. 2 Confusion Matrix	21
Gambar 3. 1 Diagram Alur Penelitian.....	24
Gambar 3. 2 Visualisasi Bar Chart Frekuensi Kata Slang	26
Gambar 3. 3 Tahapan Preprocessing dengan Normalisasi.....	29
Gambar 3. 4 Alur Pembuatan Kopus Kata Slang.....	30
Gambar 3. 5 Tahap Normalisasi.....	31
Gambar 3. 6 Proses Kerja pelabelan IndoBERT	32
Gambar 3. 7 Alur Pembuatan Pohon Keputusan	38
Gambar 3. 8 Node Akar	40
Gambar 4. 1 (a) Visualisasi top 20 stopword data tokopedia & (b) Visualisasi top 20 stopword data lazada.....	43
Gambar 4. 2 (a) Wordcloud data tokopedia berdasarkan sentimen &(b) wordcloud data Lazada berdasarkan sentimen	44
Gambar 4. 3 (a) Frekuensi Panjang karakter 1 dan 2 lazada & (b) Frekuensi Panjang karakter 1 dan 2 Tokopedia.....	45
Gambar 4. 4 (a) Top 10 kata dengan Panjang karakter 1 dan 2 Lazada & (b) Top 10 kata dengan Panjang karakter 1 dan 2 Tokopedia.....	46
Gambar 4. 5 Persentase kata slang pada data Lazada	47
Gambar 4. 6 Persentase kata slang pada data Tokopedia Tokopedia	48
Gambar 4. 7 skenario 1 preprocessing review lazada.....	49
Gambar 4. 8 skenario 1 preprocessing review tokopedia	49
Gambar 4. 9 skenario ke 2 preprocessing review lazada	50
Gambar 4. 10 skenario 2 preprocessing review tokopedia	51

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	9
Tabel 3. 1 Review Tokopedia.....	25
Tabel 3. 2 Review Lazada.....	25
Tabel 3. 3 Hasil Case Folding.....	27
Tabel 3. 4 Hasil Cleaning.....	28
Tabel 3. 5 Hasil Case Tokenizing.....	28
Tabel 3. 6 Hasil Stopword Removal.....	29
Tabel 3. 7 Kamus Slang word.....	30
Tabel 3. 8 contoh hasil preprocessing.....	33
Tabel 3. 9 Perhitungan Tf.....	34
Tabel 3. 10 Perhitungan DF.....	34
Tabel 3. 11 Perhitungan IDF.....	34
Tabel 3. 12 Perhitungan TF-IDF.....	35
Tabel 3. 13 Label Data latih.....	36
Tabel 3. 14 Sampel Data Training.....	38
Tabel 3. 15 Data Decision Tree.....	39
Tabel 4. 1 Distribusi Sentimen.....	42
Tabel 4. 2 kata – kata slang yang teridentifikasi.....	47
Tabel 4. 3 Tampilan Korpus kata slangWord.....	48
Tabel 4. 4 Perbandingan 5 kata dengan TF-IDF teratas dari.....	52
Tabel 4. 5 Perbandingan kata dengan TF-IDF tertinggi dari.....	53
Tabel 4. 6 Confusion Matrix model Naïve Bayes.....	54
Tabel 4. 7 Confusion Matrix model Decision Tree.....	56
Tabel 4. 8 Evaluasi Model Naïve Bayes.....	57
Tabel 4. 9 Evaluasi Model Decision Tree.....	58

DAFTAR LAMPIRAN

Lampiran 1 Code Labelling menggunakan <i>IndoBERT</i>	66
Lampiran 2: Code Exploratory Data Analysis (EDA)	68
Lampiran 4: Code Filtering kata <i>slang</i>	72
Lampiran 5: Code Preprocessing Tanpa Normalisasi Tokopedia.....	75
Lampiran 6: Code Modelling Tanpa Normalisasi Tokopedia	77
Lampiran 7: Code Evaluasi Model Tanpa Normalisasi Tokopedia.....	79
Lampiran 8: Code Preprocessing Dengan Normalisasi Tokopedia.....	79
Lampiran 9: Code Modelling Dengan Normalisasi Tokopedia.....	82
Lampiran 10: Code Evaluasi Model Dengan Normalisasi Tokopedia	84
Lampiran 11: Code Preprocessing Tanpa Normalisasi Lazada	85
Lampiran 12: Code Modelling Tanpa Normalisasi Lazada.....	86
Lampiran 13: Code Evaluasi Model Tanpa Normalisasi Lazada	88
Lampiran 14: Code Preprocessing Dengan Normalisasi Lazada	89
Lampiran 15: Code Modelling Dengan Normalisasi Lazada.....	92
Lampiran 16: Code Evaluasi Model Dengan Normalisasi Lazada	94
Lampiran 17: Korpus <i>Slang</i>	95