

Bidang Keahlian : Data Mining

**IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES DALAM
KLASIFIKASI JUDUL BERITA CLICKBAIT**

SKRIPSI



Oleh :

Dicky Satria Mahendra

NPM. 20081010132

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**


LEMBAR PENGESAHAN SKRIPSI

Judul : IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES
DALAM KLASIFIKASI JUDUL BERITA CLICKBAIT
Oleh : Dicky Satria Mahendra
NPM : 20081010132

Telah Diseminarkan Dalam Ujian Skripsi Pada :
Hari Rabu, Tanggal 22 Mei 2024


Mengetahui


Dosen Pembimbing

1. 
Dr. Basuki Rahmat, S.Si., MT.
NIP. 19690723 2021211 002

2. 
Retno Mumpuni, S.Kom., M.Sc.
NPT. 172198 70 716054

Dosen Penguji

1. 
Chrystia Kji Putra, S.Kom, M.T.
NIP. 19861008 202121 1 001

2. 
Retno Mumpuni, S.Kom., M.Sc.
NPT. 172198 70 716054

Menyetujui


Dekan
Fakultas Ilmu Komputer
Prof. Dr. Ir. Novirina Hendrasarie, M.T.
NIP. 19681126 199403 2 001


Koordinator Program Studi
Informatika
Fetty Tri Anggraeny, S.Kom. M.Kom.
NIP. 19820211 202121 2 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Dicky Satria Mahendra

NPM : 20081010132

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

**“IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES DALAM KLASIFIKASI
JUDUL BERITA CLICKBAIT”**

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.

Surabaya, 04 Juni 2024

Hormat saya,



Dicky Satria Mahendra

NPM. 20081010132

IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES DALAM KLASIFIKASI JUDUL BERITA CLICKBAIT

Nama Mahasiswa : Dicky Satria Mahendra
NPM : 20081010132
Program Studi : Informatika
Dosen Pembimbing : Dr. Basuki Rahmat, S.Si., MT.
Retno Mumpuni, S.Kom., M.Sc.

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan judul berita menjadi clickbait dan non-clickbait menggunakan metode *Multinomial Naive Bayes*. Data yang digunakan berasal dari dataset *CLICK-ID: A Novel Dataset for Indonesian Clickbait Headlines*. Proses penelitian melibatkan tahapan pengumpulan data, preprocessing, ekstraksi fitur, pelatihan model, evaluasi model, dan analisis hasil. Hasil pengujian menunjukkan bahwa algoritma *Multinomial Naive Bayes* mampu menghasilkan tingkat akurasi yang konsisten sekitar 78%.

Melalui optimasi menggunakan *Grid Search*, tidak menghasilkan peningkatan akurasi. Meskipun tidak mengalami peningkatan akurasi, terdapat perbaikan dalam nilai *recall* untuk kelas non-clickbait dari 76% menjadi 80% parameter terbaik yang ditemukan adalah *alpha* sebesar 0.15. Dengan demikian, Algoritma *Multinomial Naive Bayes* dapat digunakan secara efektif dalam mengatasi masalah klasifikasi judul berita clickbait, dengan potensi untuk memberikan kontribusi dalam upaya pencegahan clickbait di masa mendatang.

Kata kunci: *Multinomial Naive Bayes, Optimasi Grid Search, Klasifikasi, Berita Clickbait*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT, yang telah memberikan kesabaran dan kekuatan, rahmat serta hidayah-nya karena atas izin dan keridhoannya, penulis mampu berfikir dan menyelesaikan skripsi dengan judul :

“IMPLEMENTASI METODE MULTINOMIAL NAIVE BAYES DALAM KLASIFIKASI JUDUL BERITA CLICKBAIT”

Banyak dukungan maupun bantuan yang penulis terima selama proses pelaksanaan penelitian skripsi ini, membuat penulis mudah dalam mengerjakan penelitian ini. Diharapkan dengan adanya penyusunan skripsi ini, dapat menambah ilmu serta wawasan baru bagi yang membaca.

Penulis menyadari masih terdapat banyak kekurangan pada laporan penelitian skripsi ini, oleh karena itu, penulis menerima segala bentuk kritik, saran, dan masukan dari semua pihak yang bertujuan membangun penelitian ini menjadi lebih baik dan sempurna.

Surabaya, Mei 2024

Dicky Satria Mahendra

20081010132

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur kepada Allah SWT, penelitian dan penyusunan laporan ini dapat diselesaikan. Penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan bantuan selama proses penelitian ini. Khususnya, penulis berterima kasih kepada semua pihak yang telah berkontribusi secara langsung maupun tidak langsung dalam menyelesaikan skripsi ini. Dukungan dan bantuan tersebut sangat berarti bagi penulis. Pada Kesempatan ini penulis menyampaikan rasa terimakasih yang sebesar besarnya kepada :

1. Allah SWT yang telah memberikan kekuatan rahmat, serta hidayah kepada penulis, hingga menyelesaikan skripsi ini sampai selesai.
2. Orang tua yang telah memberikan dukungan secara materi dan non-materi sehingga penulis dapat menyelesaikan perkuliahan dari awal hingga akhir dengan baik.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
5. Bapak Firza Prima Aditiawan, S.Kom., MTI. selaku dosen wali saya yang telah membantu saya dalam perwalian yang saya alami sejak semester satu hingga semester delapan ini.
6. Bapak Dr. Basuki Rahmat, S.Si., MT. selaku dosen pembimbing pertama saya yang sangat membantu dan memberikan arahan sehingga saya dapat menyelesaikan tugas akhir skripsi ini dengan baik.
7. Ibu Retno Mumpuni, S.Kom., M.Sc. selaku dosen pembimbing kedua saya yang telah membimbing saya dalam melakukan penulisan tugas akhir ini dengan maksimal.

8. Seluruh Dosen Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu yang bermanfaat selama perkuliahan.
9. Semua pihak yang penulis tidak sebutkan satu persatu.

Akhir kata, semoga laporan skripsi ini dapat memberikan manfaat dan menambah wawasan pengetahuan bagi berbagai pihak yang membaca laporan skripsi ini. Semoga Allah Subhanahu Wa Ta'ala memberikan balasan yang berlipatganda atas kebaikan yang telah diberikan

Surabaya, Mei 2024

Dicky Satria Mahendra

DAFTAR ISI

| | |
|----------------------------------|-----|
| ABSTRAK | iii |
| KATA PENGANTAR | iv |
| UCAPAN TERIMA KASIH..... | v |
| DAFTAR ISI..... | vii |
| DAFTAR GAMBAR | x |
| DAFTAR TABEL..... | xi |
| DAFTAR KODE PROGRAM..... | xii |
| BAB I..... | 1 |
| PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Tujuan..... | 4 |
| 1.4 Manfaat..... | 5 |
| 1.5 Batasan Masalah..... | 5 |
| BAB II..... | 6 |
| TINJAUAN PUSTAKA | 6 |
| 2.1 Penelitian Terdahulu..... | 6 |
| 2.2 Berita | 7 |
| 2.3 <i>Text Processing</i> | 8 |
| 2.3.1 Case Folding..... | 8 |
| 2.3.2 Cleansing | 9 |
| 2.3.3 Tokenizing..... | 9 |
| 2.3.4 <i>Stopword</i> | 9 |
| 2.3.5 Stemming..... | 10 |

| | | |
|------------------|---|----|
| 2.4 | Klasifikasi Teks | 11 |
| 2.5 | <i>Naive Bayes Classifier</i> | 11 |
| 2.5.1 | Multinomial Naive Bayes..... | 12 |
| 2.6 | <i>Term Frequency Inverse Document Frequency (TF-IDF)</i> | 12 |
| 2.7 | <i>Confusion Matrix</i> | 13 |
| 2.8 | <i>Machine Learning</i> | 13 |
| 2.9 | Data Mining..... | 14 |
| 2.10 | Optimasi <i>Grid Search</i> | 14 |
| BAB III | | 15 |
| METODOLOGI | | 15 |
| 3.1 | Alur Penelitian..... | 15 |
| 3.2 | Pengumpulan Data | 15 |
| 3.3 | Text Processing | 16 |
| 3.4 | Pembobotan Kata | 18 |
| 3.4.1 | Menghitung <i>Term Frequency</i> | 19 |
| 3.4.2 | Menghitung <i>Inverst Document Frequency</i> | 21 |
| 3.4.3 | Menghitung TF-IDF | 22 |
| 3.5 | Optimasi <i>Grid Search</i> | 23 |
| 3.6 | <i>Multinomial Naive Bayes</i> | 27 |
| 3.6.1. | Input data uji..... | 27 |
| 3.6.2. | Hitung Probabilitas <i>Prior</i> | 28 |
| 3.6.3. | Hitung Probabilitas <i>Likelihood</i> | 29 |
| 3.6.4. | Hitung Probabilitas Akhir | 31 |
| 3.6.5. | Hasil Probabilitas Akhir | 32 |
| 3.7 | Evaluasi Model..... | 32 |
| 3.8 | Analisis Hasil | 33 |

| | |
|--|----|
| BAB IV | 34 |
| HASIL DAN PEMBAHASAN..... | 34 |
| 4.1 Pengumpulan Data | 34 |
| 4.2 Preprocessing Data | 35 |
| 4.2.1. Case Folding..... | 36 |
| 4.1.2. Cleansing | 37 |
| 4.1.3. Tokenizing..... | 38 |
| 4.1.4. Stopword | 39 |
| 4.1.5. Stemming..... | 40 |
| 4.3 Pembobotan Kata | 41 |
| 4.4 Pembagian Data..... | 42 |
| 4.5 Model Grid Search | 43 |
| 4.6 <i>Multinomial Naive Bayes</i> | 45 |
| 4.7 Evaluasi Model..... | 47 |
| 4.8 Analisis Hasil | 60 |
| BAB V..... | 62 |
| KESIMPULAN DAN SARAN..... | 62 |
| 5.1 Kesimpulan..... | 62 |
| 5.2 Saran | 63 |
| DAFTAR PUSTAKA | 64 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 3. 1. Alur Penelitian..... | 15 |
| Gambar 3. 2. Flowchart Ekstraksi Fitur | 19 |
| Gambar 3. 3. Flowchart Optimasi Grid Search..... | 23 |
| Gambar 3. 4. Flowchart Pelatihan Model | 27 |
| Gambar 4. 1. Hasil Tampilan DataFrame | 35 |
| Gambar 4. 2. Case Folding..... | 36 |
| Gambar 4. 3. Hasil Cleansing Data..... | 38 |
| Gambar 4. 4. Hasil Tokenizing Data..... | 39 |
| Gambar 4. 5. Hasil Stopword Data | 40 |
| Gambar 4. 6. Hasil Stemming | 41 |
| Gambar 4. 7. Hasil TF-IDF | 42 |
| Gambar 4. 8. Confussion Matrix Skenario 1 | 49 |
| Gambar 4. 9. Classification Report Skenario 1 | 50 |
| Gambar 4. 10. Confussion Matrix Skenario 2 | 51 |
| Gambar 4. 11. Classification Report Skenario 2..... | 51 |
| Gambar 4. 12. Confussion Matrix Skenario 3 | 53 |
| Gambar 4. 13. Classification Report Skenario 3..... | 53 |
| Gambar 4. 14. Confussion Matrix Skenario 4 | 55 |
| Gambar 4. 15. Classification Report Skenario 4..... | 55 |
| Gambar 4. 16. Confussion Matrix Skenario 5 | 57 |
| Gambar 4. 17. Classification Report Skenario 5..... | 57 |
| Gambar 4. 18. Confussion Matrix Skenario 6 | 59 |
| Gambar 4. 19. Classification Report Skenario 6..... | 59 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2. 1 Case Folding | 8 |
| Tabel 2. 2. Cleansing..... | 9 |
| Tabel 2. 3. Tokenizing | 9 |
| Tabel 2. 4. Stopword | 10 |
| Tabel 2. 5. Stemming | 11 |
| Tabel 3. 1. Contoh Dataset..... | 16 |
| Tabel 3. 2. Hasil Proses Case Folding | 17 |
| Tabel 3. 3. Hasil Proses Cleansing..... | 17 |
| Tabel 3. 4. Hasil Proses Tokenizing..... | 17 |
| Tabel 3. 5. Hasil Proses Stopword | 18 |
| Tabel 3. 6. Hasil Proses Stemming | 18 |
| Tabel 3. 7. Contoh Dokumen | 19 |
| Tabel 3. 8. Hasil Perhitungan TF | 20 |
| Tabel 3. 9. Hasil Perhitungan IDF | 21 |
| Tabel 3. 10. Hasil Perhitungan TF-IDF | 22 |
| Tabel 3. 11. Hasil Perhitungan Fold | 26 |
| Tabel 3. 12. Contoh Data | 27 |
| Tabel 3. 13. Hasil Proses Tokenizing..... | 28 |
| Tabel 3. 14. Hasil proses probabilitas Likelihood..... | 30 |
| Tabel 3. 15. Confussion Matrix..... | 33 |
| Tabel 4. 1. Perbandingan Hasil Skenario Pengujian..... | 61 |

DAFTAR KODE PROGRAM

| | |
|---|----|
| Kode Program 4. 1Menampilakan Dataset | 35 |
| Kode Program 4. 2. Case Folding | 36 |
| Kode Program 4. 3. Cleansing | 37 |
| Kode Program 4. 4. Tokenizing | 38 |
| Kode Program 4. 5. Stopword..... | 39 |
| Kode Program 4. 6. Stemming..... | 40 |
| Kode Program 4. 7. Proses Pembobotan Kata | 42 |
| Kode Program 4. 8. Pembagian Data | 43 |
| Kode Program 4. 9. Install dan Import Library..... | 43 |
| Kode Program 4. 10. Hyperparameter Tunning..... | 44 |
| Kode Program 4. 11. Mendefinisikan Range Alpha | 44 |
| Kode Program 4. 12. Mendapatkan Nilai Parameter Terbaik | 44 |
| Kode Program 4. 13. MNB | 46 |
| Kode Program 4. 14. Memanggil Kelas MNB..... | 47 |
| Kode Program 4. 15. Evaluasi Model | 48 |