

**ANALISIS SENTIMEN TWITTER MENGGUNAKAN
METODE N-GRAM, K-MEANS DAN MULTINOMIAL NAIVE
BAYES (STUDI KASUS: BPJS)**

SKRIPSI



Oleh:

QUINCY MAYFERTA PRADANI PUTRI ARIANTO

19081010127

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2023

LEMBAR PENGESAHAN

Judul : ANALISIS SENTIMEN TWITTER MENGGUNAKAN
METODE N-GRAM, K-MEANS DAN
MULTINOMIAL NAIVE BAYES (STUDI KASUS:
BPJS)

Oleh : QUINCY MAYFERTA PRADANI PUTRI ARIANTO

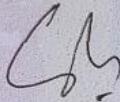
NPM : 19081010127

Telah Diseminarkan Dalam Ujian Skripsi Pada:
Hari Rabu Tanggal 24 Mei 2023

Mengetahui

Dosen Pembimbing

1.



Eva Yulia Puspaningrum, S.Kom, M.Kom

NIP. 19890705 2021212 002

2.

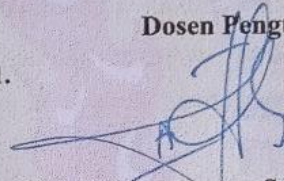


Hendra Maulana, S.Kom, M.Kom

NPT. 201198 31 223248

Dosen Penguji

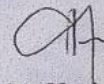
1.



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

2.



Afina Lina Nurlaili, S.Kom, M.Kom

NIP. 1993121 3202203 2010

Menyetujui

Dekan

Fakultas Ilmu Komputer

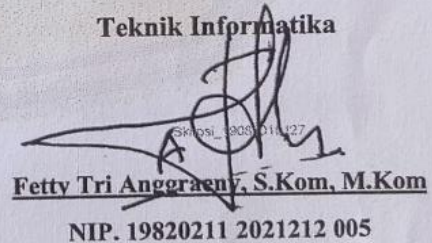


Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

Koordinator Program Studi

Teknik Informatika



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Informatika UPN “Veteran” Jawa Timur, yang bertanda tangan di bawah ini :

Nama : Quincy Mayferta Pradani Putri Arianto

NPM : 19081010127

Menyatakan bahwa judul skripsi yang saya ajukan dan kerjakan, yang berjudul:

“ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE N-GRAM, K-MEANS DAN MULTINOMIAL NAIVE BAYES (STUDI KASUS: BPJS)”

Bukan merupakan plagiat dari skripsi/tugas akhir/penelitian orang lain dan juga bukan merupakan produk dan atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Surabaya, 29 Mei 2023

Hormat Saya,



Quincy Mayferta Pradani Putri Arianto

NPM. 19081010127

ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE N-GRAM, K-MEANS DAN MULTINOMIAL NAIVE BAYES (STUDI KASUS: BPJS)

Nama Mahasiswa : Quincy Mayferta Pradani Putri Arianto

NPM 19081010127

Program Studi : Teknik Informatika

Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom, M.Kom

Hendra Maulana, S.Kom, M.Kom

ABSTRAK

Badan Penyelenggara Jaminan Sosial (BPJS) adalah badan hukum yang dibentuk oleh pemerintah negara Indonesia untuk memberikan jaminan sosial kepada masyarakatnya. Dengan meningkatnya jumlah peserta BPJS, opini terkait badan hukum tersebut pun semakin bervariasi. Opini-opini dari pengguna BPJS dapat menjadi sebuah masukan yang membangun bagi pemerintah untuk meningkatkan kualitas pelayanan sesuai dengan yang dibutuhkan oleh peserta. Media sosial di masa kini memiliki peran penting dalam penyebaran informasi maupun sebagai media pengutaraan perasaan maupun opini. Banyaknya data informasi yang tersebar di media sosial menarik perhatian bagi beberapa pihak dalam mengumpulkan data penting untuk berbagai tujuan seperti peningkatan kualitas layanan. Opini terkait diskusi atau produk dalam kanal media sosial saat ini menjadi dinamis dan konstruktif yang menyebabkan makna kalimat menjadi lebih kompleks untuk dipahami. Untuk memudahkan dalam memahami makna dari suatu opini dalam sosial media, dilakukan ekstraksi informasi berupa analisis sentimen.

Beberapa penelitian terkait analisis sentimen telah dilakukan dengan berbagai metode untuk meningkatkan akurasi dalam ekstraksi informasi. Pendekatan dengan metode gabungan terkait penelitian analisis sentimen juga telah dilakukan dengan menggabungkan dua metode seperti k-means dengan naive bayes, dan n-gram dengan naive bayes. Penggabungan metode tersebut menghasilkan akurasi yang cukup baik namun kurang optimal. Metode k-means dalam penggabungan metode k-means dan naive bayes masih mengalami beberapa kesalahan dalam clustering sentimen sehingga masih memerlukan suatu asistensi manual dalam mengelompokkan sentimen.

Maka dari itu, penelitian ini bertujuan untuk melakukan analisis sentimen dengan gabungan metode N-Gram, K-Means, dan Multinomial Naive Bayes untuk

meningkatkan akurasi dalam memahami nilai sentimen dalam suatu data teks dari sosial media. Metode n-gram dapat membantu dalam *clustering* data dengan metode k-means dengan mereduksi kata-kata yang tidak penting, dan mengoptimalkan representasi kata-kata yang akan di kelompokkan. Dengan begitu, diasumsikan jika nilai akurasi pengelompokan dengan k-means meningkat maka akurasi multinomial naive bayes juga akan turut meningkat dalam mengklasifikasikan data sentimen. Data yang digunakan dalam penelitian ini berjumlah sebesar 500 data dengan penggunaan data untuk data training sebesar 75% dan data uji sebesar 25%. Berdasarkan penilaian evaluasi, analisis sentimen dengan penggabungan n-gram, k-means, dan multinomial naïve bayes menghasilkan akurasi, presisi, *recall*, dan *F1-score* yang cukup baik dengan nilai akurasi sebesar 80,80%, presisi sebesar 79,38%, recall sebesar 80,80%, dan F1-Score sebesar 80,08%.

Kata kunci : Analisis Sentimen, N-gram, K-Means, Multinomial Naive Bayes, Twitter, BPJS

KATA PENGANTAR

Dengan mengucapkan puji syukur ke hadirat Allah SWT Yang Maha Pengasih dan Maha Penyayang karena atas segala rahmat, ridho, dan karunia-Nya penulis dapat menyelesaikan tugas akhir dengan judul “Analisis Sentiment Twitter Menggunakan Metode N-Gram, K-Means, dan Multinomial Naïve Bayes”.

Dengan rasa hormat, penulis mengucapkan terima kasih atas segala bentuk bantuan maupun dukungan yang telah diberikan kepada penulis dari awal hingga akhir penelitian tugas akhir ini.

Penulis menyadari bahwa masih terdapat kekurangan dalam laporan tugas akhir ini, mengingat keterbatasan kemampuan ataupun pengetahuan penulis. Oleh karena itu, penulis mengharapkan adanya masukan, kritik, dan saran yang bersifat membangun dalam penyempurnaan laporan ini.

Surabaya, 29 Mei 2023

Penulis

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Allah SWT atas kemudahan maupun kelancaran yang diberikan sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.

Tidak lupa, dalam penulisan tugas akhir ini penulis mendapatkan bimbingan, dukungan, dan bantuan dari beberapa pihak. Dalam kesempatan ini, penulis ingin mengucapkan terimakasih kepada yang terhormat:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur
2. Ibu Dr. Novirina Hendrasarie, S.T, M.T. Selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur
4. Ibu Eva Yulia Puspaningrum, S.Kom, M.Kom selaku dosen pembimbing I dan Bapak Hendra Maulana, S.Kom, M.Kom selaku dosen pembimbing II yang telah meluangkan banyak waktu dan memberikan arahan selama pengerjaan tugas akhir ini dari awal hingga akhir
5. Kedua orang tua penulis, Irwan Dwi Arianto, S.Sos, M.I.Kom dan Metta Vitria Maya Saputri Permadi, S.Sos atas ilmu, masukan, dukungan, doa, dan motivasi yang diberikan pada penulis.
6. Kedua saudara penulis, Chakra Satria Pradana Putra Arianto dan Farhan Dequika Pradana Putra Arianto atas segala semangat, motivasi, dan dukungan selama pengerjaan skripsi penulis
7. Teman-teman Teknik Informatika angkatan 2019 yang bersedia membantu dan menjadi bagian dari kisah hidup selama perkuliahan penulis
8. Semua pihak yang tidak dapat disebutkan satu persatu atas segala doa, motivasi, masukan, dan semangat yang diberikan selama ini

Semoga segala kebaikan dan doa-doa yang diberikan pihak-pihak tersebut kepada penulis selama ini dibalas oleh Allah SWT dengan setimpal ataupun dengan hal yang lebih baik lagi.

Surabaya, 19 Mei 2023

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	i
SURAT PERNYATAAN ANTI PLAGIAT	ii
ABSTRAK	iii
KATA PENGANTAR.....	v
UCAPAN TERIMA KASIH.....	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan.....	4
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Pendukung	6
2.2. Twitter	7
2.3. Web Scraping	7
2.4. Text Mining.....	8
2.5. Preprocessing.....	8
2.5.1 <i>Case Folding</i>	8
2.5.2 Normalisasi	9
2.5.3 <i>Tokenizing</i>	9

2.5.4	<i>Stopword Removal</i>	10
2.5.5	<i>Stemming</i>	10
2.6.	Ekstraksi Fitur	11
2.6.1	N-Gram	11
2.6.2	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	12
2.7.	K-Means	13
2.8.	Multinomial Naïve Bayes.....	14
2.9.	Analisis Sentimen.....	15
2.10.	<i>Confusion Matrix</i>	15
2.10.1	<i>Accuracy</i>	16
2.10.2	<i>Precision</i>	16
2.10.3	<i>Recall</i>	17
2.10.4	<i>F1-Score</i>	17
BAB III	METODOLOGI	18
3.1.	Rancangan Tahap Penelitian	18
3.2.	Scraping Data	19
3.3.	Preprocessing.....	21
3.3.1	<i>Cleaning</i>	23
3.3.2	<i>Slang Replacement</i>	24
3.3.3	<i>Tokenization</i>	26
3.3.4	<i>Stopword Removal</i>	27
3.3.5	<i>Stemming</i>	29
3.4.	Ekstraksi Fitur	30
3.5.	<i>Data Splitting</i>	31
3.6.	<i>Clustering</i>	31
3.7.	Klasifikasi Multinomial Naïve Bayes	33

3.8. Evaluasi Model.....	35
BAB IV HASIL DAN PEMBAHASAN.....	36
4.1. Pengambilan Data.....	36
4.1.1 Proses Pengambilan Data	36
4.1.2 Proses Penyimpanan Data	38
4.2. <i>Filtering</i> dan <i>Labelling</i>	39
4.3. Preprocessing Data	40
4.4. Ekstraksi Fitur dan Data <i>Splitting</i>	51
4.5. <i>Clustering</i>	52
4.6. Klasifikasi Data	55
4.7. Evaluasi Model.....	56
4.8. Skema Uji Coba.....	60
BAB V PENUTUP	64
5.1. Kesimpulan.....	64
5.2. Saran	64
DAFTAR PUSTAKA	65
LAMPIRAN	69
<i>Filtered Data</i> yang terlabeli	69
Hasil Preprocessing	137

DAFTAR TABEL

Tabel 2. 1 Implementasi <i>Case Folding</i>	9
Tabel 2. 2. Implementasi Normalisasi.....	9
Tabel 2. 3. Implementasi <i>Tokenizing</i>	10
Tabel 2. 4. Implementasi <i>Stopword Removal</i>	10
Tabel 2. 5. Implementasi <i>Stemming</i>	11
Tabel 2. 6. Implementasi Jenis N-Gram.....	12
Tabel 2. 7. <i>Confusion Matrix Multiclass</i>	16
Tabel 3. 1. Contoh <i>Slang Replacement</i>	25
Tabel 3. 2. Contoh Dataset Kata Gaul.....	26
Tabel 4. 1. Contoh Data Tweet Tidak Relevan	39
Tabel 4. 2. Contoh <i>Labelling</i> Sentimen	40
Tabel 4. 3. Implementasi <i>Cleaning</i>	43
Tabel 4. 4. Implementasi <i>Slang Replacement</i>	45
Tabel 4. 5. Implementasi Tokenisasi.....	46
Tabel 4. 6. Implementasi Stopword Removal.....	48
Tabel 4. 7. Implementasi <i>Stemming</i>	49

DAFTAR GAMBAR

Gambar 3. 1. Alur Rancangan Tahapan Penelitian	18
Gambar 3. 2. Laman Profil Pengguna AnyPicker.....	20
Gambar 3. 3. Tampilan Pencarian Twitter BPJS	20
Gambar 3. 4. Tampilan Jendela Pengambilan Data	21
Gambar 3. 5. Alur Proses Preprocessing.....	22
Gambar 3. 6. Alur Proses <i>Cleaning</i>	23
Gambar 3. 7. Alur Proses Pergantian Kata Gaul.....	24
Gambar 3. 8. Alur Proses <i>Tokenization</i>	27
Gambar 3. 9. Alur Proses <i>Stopword Removal</i>	28
Gambar 3. 10. Alur Proses <i>Stemming</i>	29
Gambar 3. 11. File Hasil Preprocessing.....	30
Gambar 3. 12. Alur Proses Ekstraksi Fitur	31
Gambar 3. 13. Alur Proses Clustering	32
Gambar 3. 14. Alur Proses Multinomial Naive Bayes.....	34
Gambar 4. 1. Laman Pencarian BPJS di Twitter	36
Gambar 4. 2. Proses Pemilihan Elemen Data yang Diambil.....	37
Gambar 4. 3. Proses Pengambilan Data AnyPicker	37
Gambar 4. 4. Download Hasil Pengambilan Data dalam Bentuk CSV	38
Gambar 4. 5. Hasil Data Dalam Bentuk File Excel	38
Gambar 4. 6. Hasil Penerapan Potongan Kode Membaca Data.....	42
Gambar 4. 7. Hasil Penerapan Potongan Kode <i>Cleaning</i>	43
Gambar 4. 8. Hasil Penerapan Potongan Kode <i>Slang Replacement</i>	44
Gambar 4. 9. Hasil Penerapan Potongan Kode Tokenisasi.....	46
Gambar 4. 10. Hasil Penerapan Potongan Kode <i>Stopword Removal</i>	47
Gambar 4. 11. Hasil Penerapan Potongan Kode <i>Stemming</i>	50
Gambar 4. 12. Hasil Setiap Tahap Preprocessing	50
Gambar 4. 13. Dataframe "hasilp"	51
Gambar 4. 14. Hasil Pembobotan TF-IDF.....	52
Gambar 4. 15. y_train Hasil Clustering K-Means.....	54

Gambar 4. 16. Visualisasi Clustering K-Means	55
Gambar 4. 17. y_pred Prediksi Multinomial Naive Bayes	56
Gambar 4. 18. Hasil Penerapan Potongan Kode Evaluasi Model.....	57
Gambar 4. 19. Hasil Penerapan Potongan Kode <i>Confusion Matrix</i>	58
Gambar 4. 20. Hasil Penerapan Potongan Kode Report Evaluasi	59
Gambar 4. 21 y_pred Skema Uji Coba	60
Gambar 4. 22. Visualisasi K-Means Uji Coba	61
Gambar 4. 23. y_pred Uji Coba	61
Gambar 4. 24. Evaluasi Uji Coba.....	62
Gambar 4. 25. Confusion Matrix Skema Uji Coba.....	62
Gambar 4. 26. Report Skema Uji Coba.....	63