

**KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL TWITTER
DENGAN METODE NAÏVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Mifta Aji Zulfiqar

17.11.1404

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
2022**

**KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL TWITTER
DENGAN METODE NAÏVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Mifta Aji Zulfiqar
17.11.1404

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Mifta Aji Zulffkar

17.11.1404

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 November 2021

Dosen Pembimbing,

Atnul Yaqin, M.Kom.

NIK. 190302255

PENGESAHAN

SKRIPSI

KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Mifta Aji Zulfiqar

17.11.1404

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Desember 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom

NIK. 190302150

Mardhiva Havaty, S.T., M.Kom

NIK. 190302108

Alnul Yaqln, M.Kom

NIK. 190302255

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Desember 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

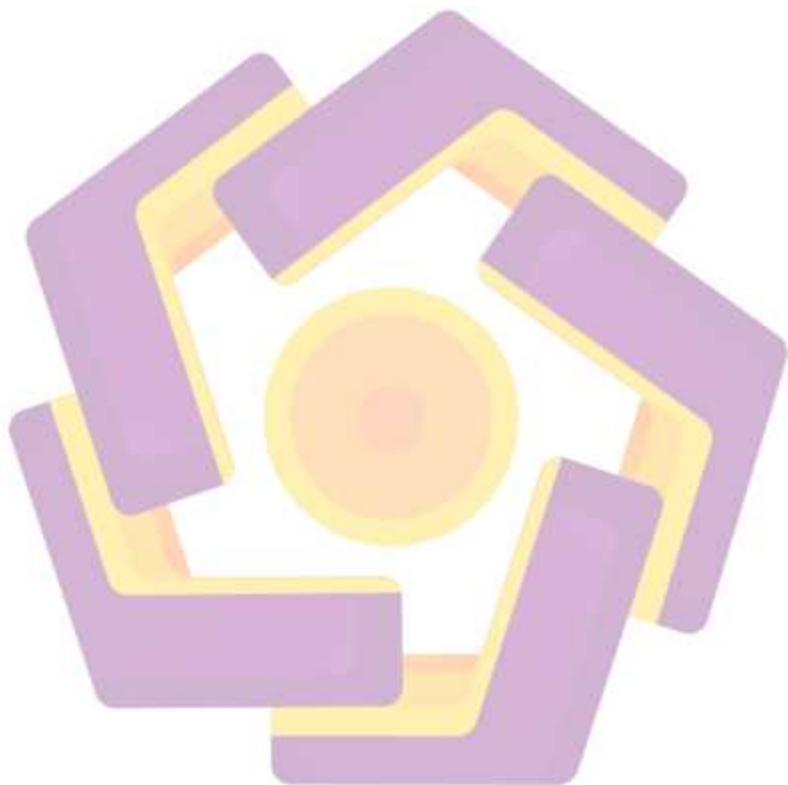
Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Desember 2021

Mifta Aji Zulfikar
17.11.1404

MOTTO

*"Teringat kuteringat, pada janjimu kuterikat, hanya sekejap ku berdiri,
kulakukan sepenuh hati." – Letto*



PERSEMBAHAN

Saya memersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam proses pembuatan skripsi. Berkat rahmat Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya hingga detik ini.

1. Kedua orang tua yang dengan ikhlas tanpa mengharapkan imbalan apapun dalam memberikan dukungan, do'a dan restunya tanpa henti untuk penulis.
2. Adik-adik dari penulis yang selalu menjadi sumber semangat dan motivasi bagi penulis untuk selalu semangat.
3. Keluarga besar H. Ahmat Thohir yang selalu memberi dukungan serta do'a untuk kelancaran pendidikan bagi penulis.
4. Bapak Ainul Yaqin selaku dosen pembimbing yang dengan kesabarannya membimbing dan mengajari banyak hal untuk penulis.
5. *My Precious*, Shafira Apriliani Iswanto yang selalu memberikan semangat, dorongan serta menemani penulis dari awal hingga saat ini.
6. Teman-teman Trivastore yang menjadi bagian dari perjalanan dalam menyusun skripsi.
7. Teman-teman dekat penulis, Khairin Pratama, Bayu Wicahyono, Putri Handayani, Riana Anggita Saputri, Ruth Mufidah Devi, Natasha Berliana Aurora yang selalu mendukung penulis.
8. Semua dosen dan karyawan Universitas Amikom Yogyakarta.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shalawat serta salam juga tidak lupa penulis panjatkan kepada junjungan nabi kita Nabi Muhammad SAW yang telah memberikan teladan mulia dan menuntun umatnya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul “Klasifikasi Berita Hoax pada Media Sosial Twitter dengan Metode Naïve Bayes Classifier” ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas Amikom Yogyakarta.

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Ainul Yaqin, M. Kom selaku dosen pembimbing yang telah banyak memberikan arahan bagi penulis agar penulis dapat menyelesaikan skripsi ini dengan baik.
3. Kedua orang tua, saudara, dan keluarga besar yang telah memberikan dukungan, doa, dan kasih sayangnya kepada penulis.
4. Seluruh dosen dan staff Universitas Amikom Yogyakarta yang telah membantu dan membimbing selama proses perkuliahan.
5. Teman-teman kelas 16 S1IF 02 dan teman-teman dekat.
6. Teman-teman kelas 16 S1IF 02 dan teman-teman dekat. Seluruh teman-teman dan keluarga besar Universitas Amikom Yogyakarta dan semua pihak yang tidak dapat saya sebutkan satu persatu, terima kasih banyak atas segala bantuannya dalam menyelesaikan skripsi ini.

Penulis menyadari skripsi ini masih ada kekurangan, maka dari itu kritik dan saran yang membangun serta teguran dari semua pihak, penulis menerima dengan lapang dada untuk kesempurnaan karya selanjutnya. Akhirnya kepada Allah SWT semoga skripsi yang sederhana ini dapat memberikan manfaat. Khususnya bagi penulis dan pembaca yang yang budiman pada umumnya.

Yogyakarta, 20 Desember 2021


Mifta Aji Zulfikar
17.11.1404



DAFTAR ISI

JUDUL	II
PERSETUJUAN.....	III
PENGESAHAN	IV
PERNYATAAN	V
MOTTO.....	VI
PERSEMBAHAN	VII
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	X
DAFTAR TABEL	XIII
DAFTAR GAMBAR.....	XIV
INTISARI	XVI
ABSTRACT	XVII
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	3
1.5 MANFAAT PENELITIAN	3
1.6 METODE PENELITIAN.....	4
1.7 SISTEMATIKA PENULISAN.....	5
BAB II LANDASAN TEORI.....	7
2.1 KAJIAN PUSTAKA	7
2.2 DASAR TEORI	12
2.2.1 Pengertian Hoax	12
2.2.2 Pengertian Twitter	13

2.2.3	Text Mining.....	13
2.2.4	Text Preprocessing.....	14
2.2.5	Naive Bayes.....	15
2.2.6	Term Frequency-Inverse Document Frequency (TF-IDF).....	16
2.2.7	Confusion Matrix.....	17
2.2.8	Precision.....	18
2.2.9	Recall.....	19
2.2.10	F1-Score.....	19
2.2.11	K-Fold Cross Validation.....	20
2.2.12	Python.....	20
BAB III ANALISIS DAN PERANCANGAN SISTEM.....		21
3.1	IDENTIFIKASI MASALAH.....	21
3.2	ANALISIS MASALAH.....	21
3.3	HASIL ANALISIS.....	22
3.4	ANALISIS KEBUTUHAN.....	22
3.4.1	Analisis Kebutuhan Fungsional.....	22
3.4.2	Analisis Kebutuhan Non-Fungsional.....	23
3.5	DESKRIPSI SISTEM.....	23
3.6	KLASIFIKASI ALGORITMA NAÏVE BAYES.....	25
3.6.1	Pengumpulan Dataset.....	25
3.6.2	Cleaning Data.....	26
3.6.3	Case Folding.....	28
3.6.4	Tokenization.....	29
3.6.5	Stemming.....	31
3.6.6	TF-IDF.....	33
3.6.7	Naïve Bayes.....	38
3.7	PERANCANGAN SISTEM.....	45
3.7.1	Diagram Pengembangan Sistem.....	45
3.7.2	Diagram Alur Preprocessing.....	46
3.7.3	Proses Klasifikasi Naïve Bayes.....	47
BAB IV HASIL DAN PEMBAHASAN.....		49



4.1	DESKRIPSI IMPLEMENTASI	49
4.2	IMPLEMENTASI PENGUMPULAN DATA TRAINING	49
4.3	IMPLEMENTASI PELABELAN DATA	50
4.4	IMPLEMENTASI PERANCANGAN FUNGSI DAN SISTEM	51
4.4.1	Pengaksesan Data Training	51
4.4.2	Implementasi Data Cleaning dan Case Folding	52
4.4.3	Implementasi Tokenizing	52
4.4.4	Implementasi Stemming	53
4.4.5	Implementasi Pembobotan TF-IDF	53
4.4.6	Implementasi Naïve Bayes Classifier	54
4.4.7	Implementasi Pengukuran Akurasi Naïve Bayes	54
4.5	EVALUASI PROGRAM	55
BAB V PENUTUP		64
5.1	KESIMPULAN	64
5.2	SARAN	64
DAFTAR PUSTAKA		65

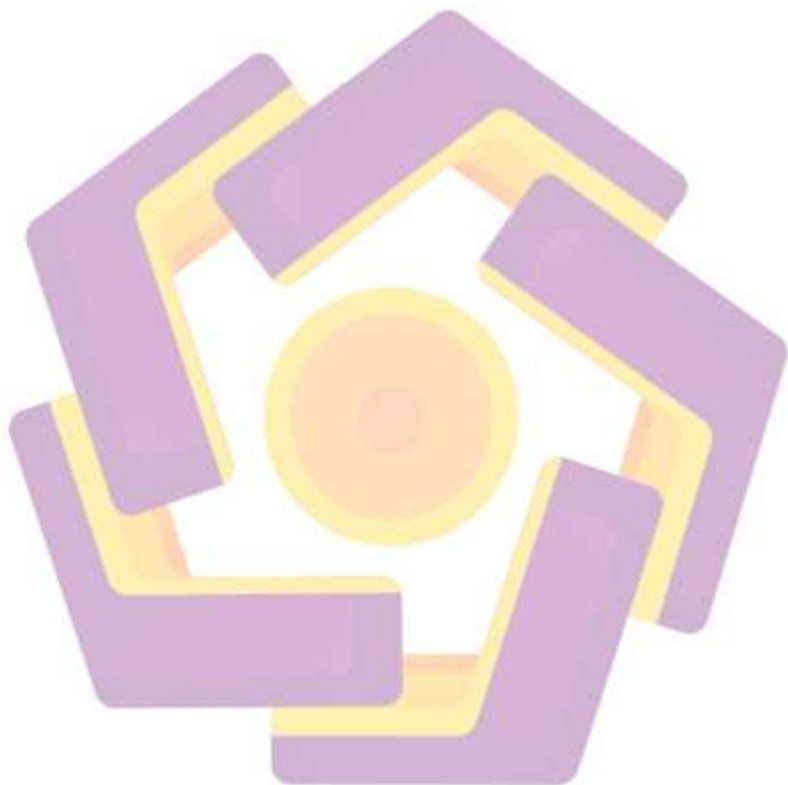
DAFTAR TABEL

TABLE 2. 1 TABEL PENELITIAN SEBELUMNYA.....	9
TABLE 2. 2 CONFUSION MATRIX	18
TABLE 3. 1 CONTOH DATA LATIH	26
TABLE 3. 2 CLEANING DATA.....	27
TABLE 3. 3 CASE FOLDING	28
TABLE 3. 4 TOKENIZATION.....	29
TABLE 3. 5 STEMMING.....	31
TABLE 3. 6 FEATURE LIST.....	32
TABLE 3. 7 TF-IDF FEATURE TRUE	34
TABLE 3. 8 TF-IDF FEATURE FALSE.....	35
TABLE 3. 9 DATA UJI NAÏVE BAYES.....	39
TABLE 3. 10 PENGECEKAN KATA PADA FEATURE MODEL	41
TABLE 3. 11 PENENTUAN PRIOR KELAS.....	41
TABLE 3. 12 HASIL DAN PRESENTASE PREDIKSI.....	43
TABLE 4. 1 CONFUSION MATRIX MODEL NAÏVE BAYES.....	55
TABLE 4. 2 HASIL 10 FOLD CROSS VALIDATION	56
TABLE 4. 3 CONFUSION MATRIX PADA K-FOLD CROSS VALIDATION.....	58

DAFTAR GAMBAR

GAMBAR 3. 1 FLOWCHART SISTEM	25
GAMBAR 3. 2 MODEL 4-FOLD CROSS VALIDATION	44
GAMBAR 3. 3 DIAGRAM PENGEMBANGAN SISTEM.....	45
GAMBAR 3. 4 DIAGRAM ALUR PREPROCESSING.....	46
GAMBAR 3. 5 PROSES KLASIFIKASI NAÏVE BAYES.....	47
GAMBAR 4. 1 LINK DATASET.....	50
GAMBAR 4. 2 SALINAN ISI LINK DATASET.....	50
GAMBAR 4. 3 IMPLEMENTASI PELABELAN DATASET	51
GAMBAR 4. 4 PENGAKSESAN DATA TRAINING	52
GAMBAR 4. 5 IMPLEMENTASI DATA CLEANING DAN CASE FOLDING	52
GAMBAR 4. 6 PROSES TOKENIZING	53
GAMBAR 4. 7 IMPLEMENTASI STEMMING	53
GAMBAR 4. 8 IMPLEMENTASI PEMBOBOTAN TF-IDF.....	54
GAMBAR 4. 9 IMPLEMENTASI NAÏVE BAYES CLASSIFIER.....	54
GAMBAR 4. 10 IMPLEMENTASI PENGUKURAN AKURASI NAÏVE BAYES.....	55
GAMBAR 4. 11 EVALUASI K-FOLD CROSS VALIDATION	58
GAMBAR 4. 12 IMPLEMENTASI TRAIN TEST SPLIT	59
GAMBAR 4. 13 FOLD – 1 TABEL CONFUSION MATRIX	59
GAMBAR 4. 14 FOLD – 2 TABEL CONFUSION MATRIX	59
GAMBAR 4. 15 FOLD – 3 TABEL CONFUSION MATRIX	60
GAMBAR 4. 16 FOLD – 4 TABEL CONFUSION MATRIX	60
GAMBAR 4. 17 FOLD – 5 TABEL CONFUSION MATRIX	60
GAMBAR 4. 18 FOLD – 6 TABEL CONFUSION MATRIX	61

GAMBAR 4. 19 FOLD – 7 TABEL CONFUSION MATRIX	61
GAMBAR 4. 20 FOLD – 8 TABEL CONFUSION MATRIX	61
GAMBAR 4. 21 FOLD – 9 TABEL CONFUSION MATRIX	62
GAMBAR 4. 22 FOLD – 10 TABEL CONFUSION MATRIX	62



INTISARI

Berita hoax yang tersebar di media sosial menjadi salah satu penyebab perpecahan atau pertengkaran yang terjadi di dunia maya. Salah satu media yang menjadi tempat penyebaran berita hoax adalah Twitter. Dengan segala kemudahannya membuat berita hoax dengan mudah tersebar dan dikonsumsi oleh masyarakat tanpa adanya penyaringan.

Klasifikasi teks dengan menggunakan metode Naïve Bayes dapat melakukan prediksi terhadap tweet yang mengarah pada berita hoax dengan melakukan klasifikasi dari data yang telah dikumpulkan. Implementasi model Naive Bayes ini diharapkan mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi.

Penelitian dengan menggunakan 1000 dataset yang dibagi menjadi 80% data training dan 20% data testing ini menghasilkan nilai accuracy sebesar 78,9%, 77% untuk nilai precision dan 99,5% untuk nilai recall.

Kata Kunci: Berita, Hoax, Twitter, Naïve Bayes



ABSTRACT

Hoax news spread on social media is one of the causes of divisions or fights that occur in cyberspace. One of the media that is a place for spreading hoax news is Twitter. With all its convenience, hoax news is easily spread and consumed by the public without any filtering.

Text classification using the Naïve Bayes method can predict tweets that lead to hoax news by classifying the data that has been collected. The implementation of the Naïve Bayes model is expected to be able to produce predictions with a high level of accuracy.

The research using 1000 datasets which are divided into 80% training data and 20% testing data produces an accuracy value of 78.9%, 77% for precision values and 99.5% for recall values.

Keyword: *News, Hoax, Twitter, Naïve Bayes*

