

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
POSTINGAN HOAX TERHADAP COVID-19 DI TWITTER**

SKRIPSI



disusun oleh

Dina Amalia

17.11.1174

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
POSTINGAN HOAX TERHADAP COVID-19 DI TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Dina Amalia

17.11.1174

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

IMPLEMENTASI NAIVE BAYES CLASSIFIER UNTUK MENDETEKSI POSTINGAN HOAX TERHADAP COVID-19 DI TWITTER

yang dipersiapkan dan disusun oleh

Dina Amalia

17.11.1174

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Mei 2020

Dosen Pembimbing,

Sumarni Adi, S.Kom,M.Cs
NIK. 190302256

PENGESAHAN

SKRIPSI

IMPLEMENTASI NAIVE BAYES CLASSIFIER UNTUK MENDETEKSI POSTINGAN HOAX TERHADAP COVID-19 DI TWITTER

yang dipersiapkan dan disusun oleh

Dina Amalia

17.11.1174

telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Februari 2021

Susunan Dewan Penguji

Nama Penguji

Robert Marco, M.T
NIK. 190302228

Ainul Yaqin, M.Kom.
NIK. 190302255

Sumarni Adi, S.Kom, M.Cs.
NIK. 190302256

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

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 tanggung jawab saya pribadi.

Yogyakarta, 23 Februari 2021



Dina Amalia

NIM. 17.11.1174

MOTTO

“Balas dendam terbaik adalah dengan memperbaiki dirimu.” – Ali Bin Abi Thalib

“Jika kau tak suka sesuatu, ubahlah. Jika tak bisa, maka ubahlah cara pandangmu tentangnya.” – Maya Angelou

“Nikmat paling indah adalah takut pada Tuhan.” – SYP

“Time keep ticking, a well time spent is a must.” – DW

“Your time is yours. My time is mine. But everytime is God’s” – DI

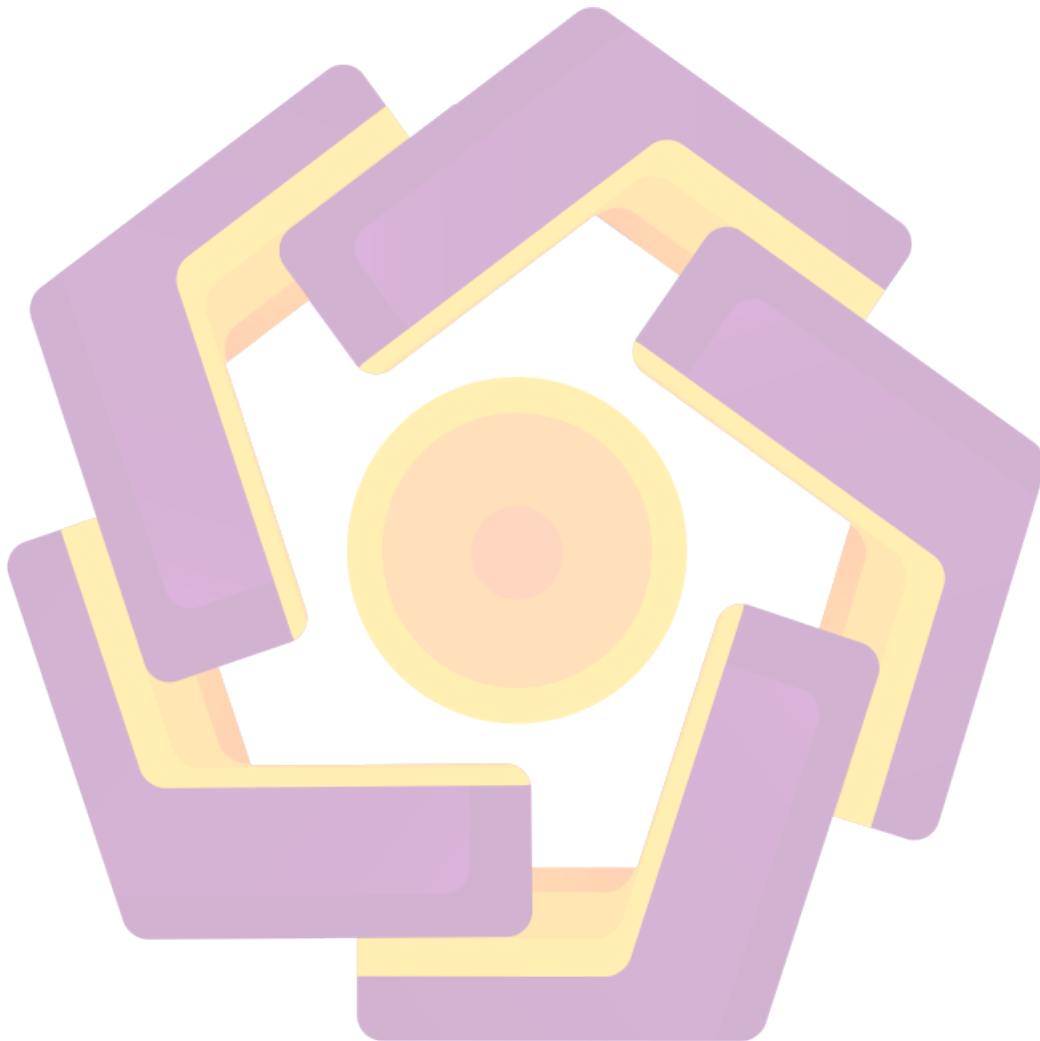
“Setiap baik dan buruk yang terjadi pasti ada maksudnya. Tapi selalu kembali pada Tuhan adalah jawabannya.” – DI

PERSEMBAHAN

Terdapat banyak pihak yang berperan dalam pembuatan skripsi ini hingga akhirnya dapat terselesaikan. Penulis mengucapkan terima kasih kepada :

1. Allah SWT dengan segala kuasa-Nya yang telah membimbing penulis dalam berikhtiar serta memberikan kemampuan untuk dapat menyelesaikan skripsi ini,
2. Kedua orang tua penulis yang tidak pernah berhenti berdoa dan memberikan dukungan kepada penulis dalam bentuk apapun,
3. Kakak dan adik penulis yang selalu mendoakan dan memberikan dukungan,
4. Ibu Sumarni Adi, S.Kom., M.Cs., yang telah memberikan bimbingan dan arahan kepada penulis,
5. Seluruh guru dan dosen yang telah berbagi ilmu dan pengetahuan dengan tulus dan ikhlas sehingga dapat dijadikan bekal dalam pengerjaan skripsi ini baik secara langsung maupun tidak,
6. Shidiq Pangestu yang bersedia meluangkan waktu dan menyalurkan energi positif untuk tetap semangat dalam menyelesaikan skripsi ini,
7. Teman-teman student staff DAAK dan DPK yang senantiasa membantu penulis
8. Teman-teman 17 Informatika 04, Keluarga AMCC Universitas AMIKOM Yogyakarta 2018/2019, keluarga Senat Mahasiswa 2018/2019, serta penghuni Kost Wisma Ukhti. penulis mengucapkan terima kasih atas segala dukungan dan doa yang diberikan,

9. Semua pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat penulis sebutkan satu persatu,
10. Diri sendiri.



KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah SWT yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Implementasi Naive Bayes Classifier Untuk Mendeteksi Postingan Hoax Terhadap Covid-19 di Twitter.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program Strata 1 dan untuk memperoleh gelar Sarjana Komputer.

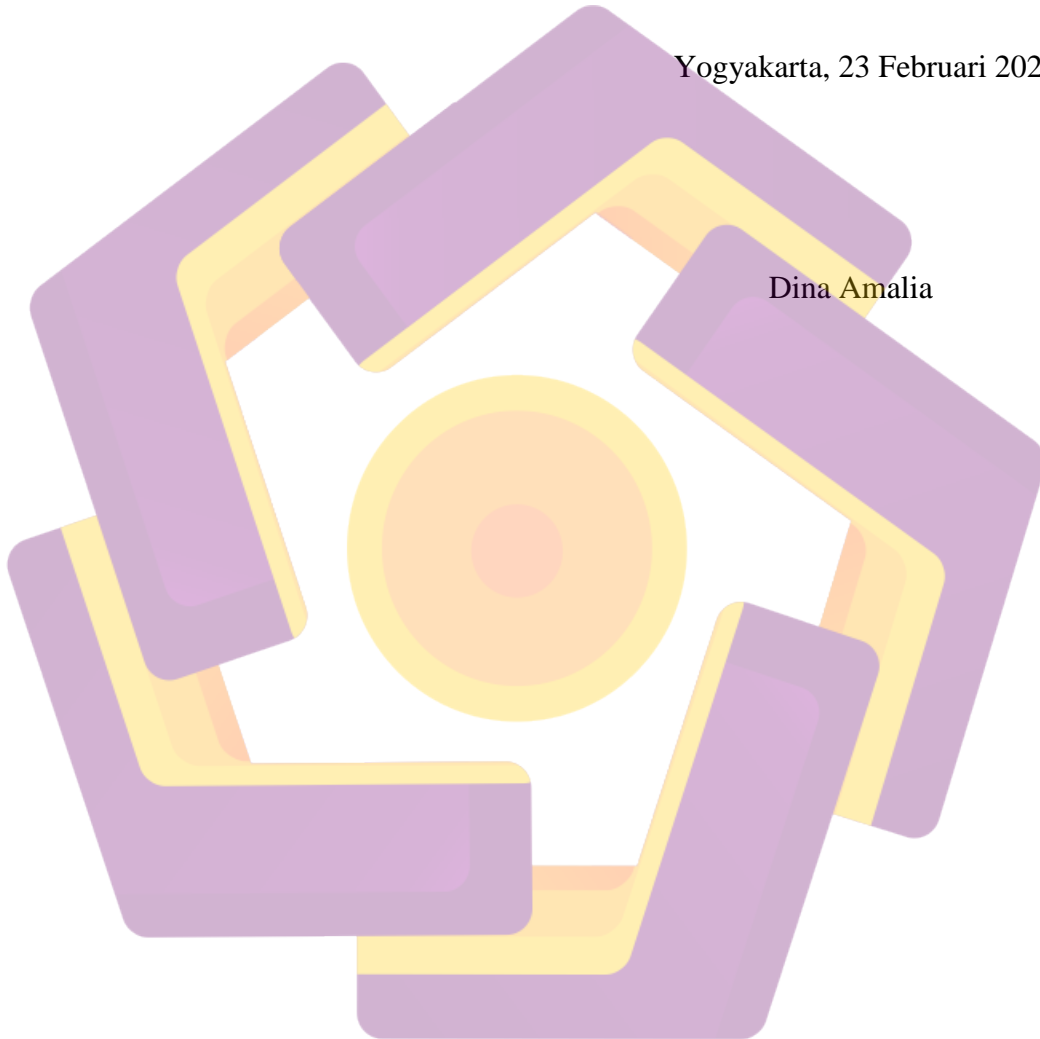
Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Sumarni Adi, S.Kom, M.Cs. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji (Robert Marco, M.T. dan Ainul Yaqin, M.Kom.) dan segenap Dosen dan Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamanya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat.
6. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Tuhan Yang Maha Esa memberikan balasan yang lebih kepada semua yang telah ikut membantu saya dan menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun akan diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 23 Februari 2021

Dina Amalia




DAFTAR ISI

PERNYATAAN	iv
PERSEMBAHAN.....	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvii
INTISARI	xix
ABSTRACT	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Maksud dan Tujuan Penelitian.....	4
1.5 Metodologi Penelitian	5
1.5.1 Metode Pengumpulan Data	5
1.5.1.1 Studi Pustaka.....	5
1.5.1.2 Crawling Data	5
1.5.2 Metode Analisis.....	6
1.5.3 Metode Perancangan Sistem	6

1.5.4	Metode Implementasi	6
1.5.5	Metode Pengujian.....	9
1.6	Sistematika Penulisan.....	9
BAB II LANDASAN TEORI.....		13
2.1	Kajian Pustaka.....	13
2.2	Dasar Teori.....	20
2.2.1	Media Sosial.....	20
2.2.2	Twitter.....	21
2.2.3	COVID-19.....	22
2.2.4	Hoax.....	23
2.2.5	Data Mining.....	31
2.2.6	Text Mining.....	32
2.2.6.1	Text Preprocessing.....	32
2.2.7	Klasifikasi.....	34
2.2.8	Naïve Bayes Classifier.....	34
2.3	Metode Pengembangan Sistem.....	37
2.4	Analisis.....	41
2.4.1	Analisis Kebutuhan.....	41
2.5	Perancangan.....	42
2.5.1	Data Flow Diagram (DFD).....	42
2.5.2	Flowchart.....	46

2.6	Pengujian Sistem	50
2.6.1	Black Box Testing.....	50
2.6.2	K Fold Cross Validation.....	51
BAB III ANALISIS DAN PERANCANGAN		53
3.1	Analisis Masalah	53
3.1.1	Identifikasi Masalah	53
3.1.2	Hasil Analisis	54
3.2	Analisis Sistem.....	56
3.2.1	Analisis Kebutuhan	56
3.2.2.1	Analisis Kebutuhan Fungsional	56
3.2.2.2	Analisis Kebutuhan Non Fungsional.....	56
3.3	Analisis Algoritma	57
3.3.1	Analisis Kebutuhan Data.....	57
3.3.1.1	Pengumpulan Data	57
3.3.2	Pelabelan Data dan Perhitungan Manual.....	59
3.3.3	Preprocessing	72
3.3.3.1	Case Folding.....	73
3.3.3.2	Tokenization.....	74
3.3.3.3	Stopwords Removal	76
3.3.3.4	Stemming.....	77
3.3.4	Klasifikasi.....	78

3.3.4.1	Proses Training.....	80
3.3.4.2	Proses Klasifikasi	84
3.4	Perancangan Sistem.....	86
3.4.1	DFD.....	86
3.4.2	Perancangan Antarmuka.....	94
BAB IV IMPLEMENTASI DAN PEMBAHASAN		99
4.1	Implementasi	99
4.1.1	Pengumpulan Dataset Training	99
4.1.2	Implementasi Antar Muka.....	100
4.1.2.1	Implementasi Halaman Utama.....	100
4.1.2.2	Implementasi Halaman Classification.....	103
4.1.2.3	Implementasi Halaman Evaluation	105
4.2	Pembahasan Source Code	106
4.2.1	Source Code Pengambilan Data.....	106
4.2.2	Source Code Training Data.....	107
4.2.3	Source Code Preprocessing Data	108
4.2.4	Source Code Klasifikasi.....	109
4.3	Pengujian Sistem	110
4.3.1	Black-box Testing	110
4.4	Hasil dan Analisis Pengujian Model	112
4.4.1.1	10 Fold Cross Validation.....	112

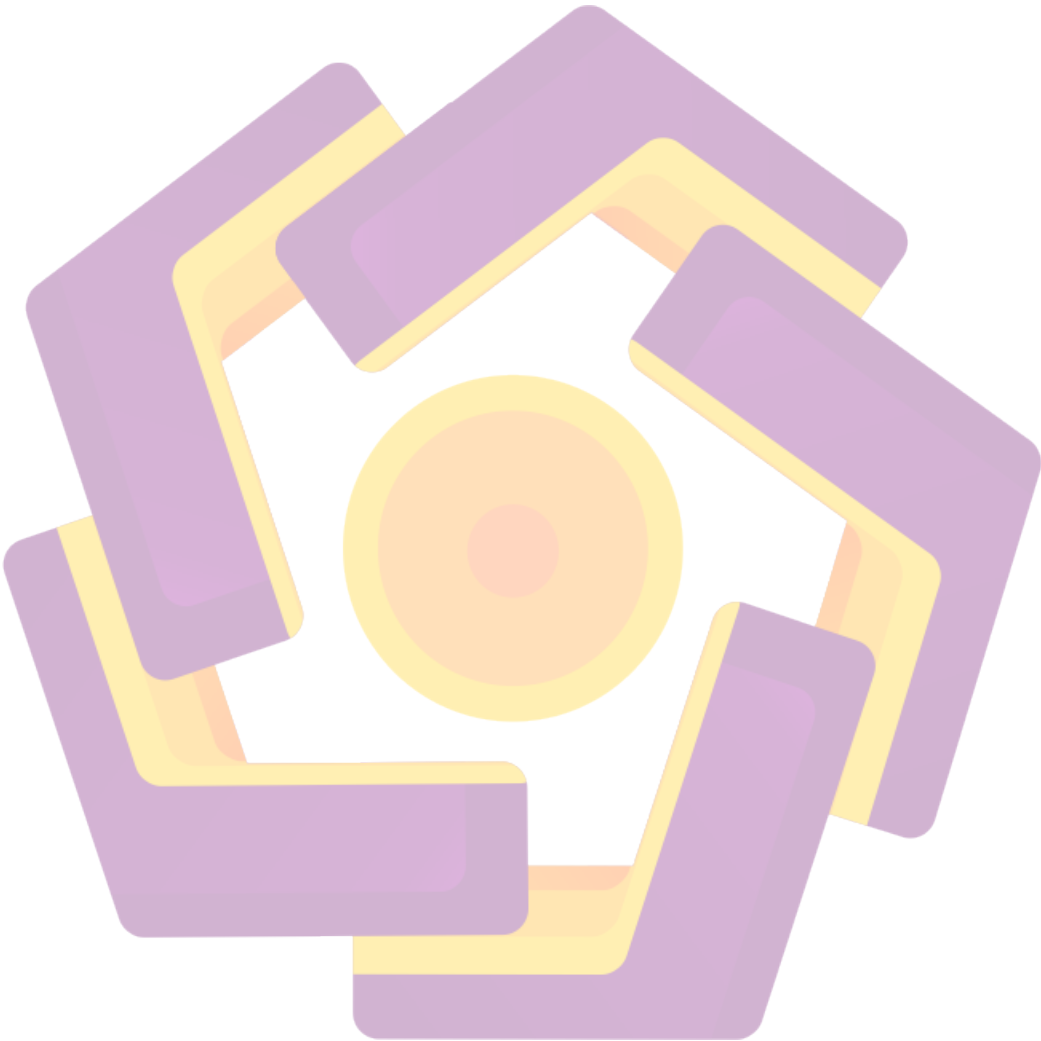


4.4.1.2	Confusion Matrix	113
4.4.2.1	Hasil Pengujian Fold 1	114
4.4.2.2	Hasil Pengujian Fold 2	115
4.4.2.3	Hasil Pengujian Fold 3	115
4.4.2.4	Hasil Pengujian Fold 4	116
4.4.2.5	Hasil Pengujian Fold 5	117
4.4.2.6	Hasil Pengujian Fold 6	118
4.4.2.7	Hasil Pengujian Fold 7	119
4.4.2.8	Hasil Pengujian Fold 8	119
4.4.2.9	Hasil Pengujian Fold 9	120
4.4.2.10	Hasil Pengujian Fold 10	121
BAB V	Penutup	123
5.1	Kesimpulan.....	123
5.2	Saran.....	123
DAFTAR PUSTAKA	125

DAFTAR TABEL

Tabel 2. 1 Tabel Referensi Penelitian	15
Tabel 2. 2 Fitur	30
Tabel 2. 3 Simbol-simbol Data Flow Diagram (DFD)	44
Tabel 2. 4 Tabel Flowchart	47
Tabel 3. 1 Contoh Postingan.....	57
Tabel 3. 2 Fitur yang Mengidentifikasi Hoax	59
Tabel 3. 3 Penentu Postingan	60
Tabel 3. 4 Perhitungan Probabilitas kemunculan setiap nilai untuk masing-masing atribut	62
Tabel 3. 5 Contoh Pelabelan Postingan	71
Tabel 3. 6 Contoh Case Folding Postingan	73
Tabel 3. 7 Contoh tokenization Postingan	74
Tabel 3. 8 Stopwords Removal Postingan	76
Tabel 3. 9 Contoh Stemming Postingan.....	77
Tabel 3. 10 Jumlah Dokumen	80
Tabel 3. 11 Token Frekuensi.....	81
Tabel 4. 1 Black box Testing	111
Tabel 4. 2 Hasil Pengujian Confusion Matrix Fold 1	114
Tabel 4. 3 Hasil Pengujian Confusion Matrix Fold 2	115
Tabel 4. 4 Hasil Pengujian Confusion Matrix Fold 3	116
Tabel 4. 5 Hasil Pengujian Confusion Matrix Fold 4	117
Tabel 4. 6 Hasil Pengujian Confusion Matrix Fold 5	117
Tabel 4. 7 Hasil Pengujian Confusion Matrix Fold 6	118
Tabel 4. 8 Hasil Pengujian Confusion Matrix Fold 7	119
Tabel 4. 9 Hasil Pengujian Confusion Matrix Fold 8	120
Tabel 4. 10 Hasil Pengujian Confusion Matrix Fold 9	120
Tabel 4. 11 Hasil Pengujian Confusion Matrix Fold 10	121

Tabel 4. 12 Hasil Pengujian K Fold Cross Validation Seluruh Fold 122



DAFTAR GAMBAR

Gambar 2. 1 Tahapan SDLC	37
Gambar 2. 2 Skenario K Fold Cross Validation	53
Gambar 3. 1 Alur Sistem.....	56
Gambar 3. 2 Flowchart Preprocessing	74
Gambar 3. 3 Flowchart Naïve Bayes Classifier	81
Gambar 3. 4 Diagram Konteks.....	88
Gambar 3. 5 DFD Level 1.....	89
Gambar 3. 6 DFD Level 2 Proses Training.....	90
Gambar 3. 7 DFD Level 2 Proses Klasifikasi.....	91
Gambar 3. 8 DFD Level 2 Proses Evaluasi.....	92
Gambar 3. 9 DFD Level 3 Proses Training.....	93
Gambar 3. 10 DFD Level 3 Proses Klasifikasi.....	95
Gambar 3. 11 DFD Level 3 Proses Evaluasi.....	96
Gambar 3. 12 UI Halaman Utama.....	97
Gambar 3. 13 UI Navigation Bar	98
Gambar 3. 14 UI Menu Training.....	99
Gambar 3. 15 UI Menu Classification	99
Gambar 3. 16 UI Menu Evaluation.....	100
Gambar 4. 1 Data Tweet Hoax.....	101

Gambar 4. 2 Data Tweet Not Hoax.....	102
Gambar 4. 3 Implementasi Halaman Utama.....	103
Gambar 4. 4 Implementasi Navigation Bar.....	104
Gambar 4. 5 Implementasi Halaman Training.....	105
Gambar 4. 6 Halaman Training Result	105
Gambar 4. 7 Implementasi Halaman Classification.....	106
Gambar 4. 8 Halaman Classification Result	107
Gambar 4. 9 Implementasi Halaman Evaluation	108
Gambar 4. 10 Halaman Evaluation Result.....	108
Gambar 4. 11 Source Code Scrapping Data.....	109
Gambar 4. 12 Source Code Training Data	110
Gambar 4. 13 Source Code Preprocessing Data	111
Gambar 4. 14 Source Code Klasifikasi	112
Gambar 4. 15 Grafik batang presentase akurasi.....	115

INTISARI

Media sosial adalah sarana untuk berkomunikasi dan bertukar informasi antar sesama pengguna media sosial, media sosial Twitter adalah salah satunya. Tetapi informasi yang disebarluaskan tidak sepenuhnya benar, ada beberapa berita yang tidak sesuai dengan kebenaran atau sering disebut hoax. Ada banyak kasus penyebaran hoax yang menimbulkan kekhawatiran dan sering membahayakan individu atau kelompok. sehingga dalam penelitian ini, penulis membangun sistem untuk mengidentifikasi berita hoax pada media sosial Twitter.

Penelitian ini menggunakan metode Naïve Bayes Classifier untuk mengklasifikasikan postingan-postingan dengan topik covid-19 di Twtter. Postingan tersebut akan digolongkan ke dalam dua kelas yaitu kelas hoax untuk postingan yang tergolong tidak benar dan kelas non hoax untuk postingan yang tergolong benar.

Dataset yang digunakan untuk proses training sebanyak 500 data, dengan komposisi 250 postingan hoax dan 250 postingan non hoax. Penelitian ini diawali dengan proses preprocessing, dilanjutkan proses training dan klasifikasi. Berdasarkan hasil evaluasi yang menggunakan 10 fold cross validation, didapatkan hasil akurasi rata-rata sebesar 76 %, rata-rata error 24 %, rata-rata precision 39% dan rata-rata recall 74%. Dari 10 fold, akurasi tertinggi terdapat pada fold 8 dengan hasil akurasi 96%, sedangkan akurasi terendah yaitu pada fold 10 sebesar 50%. Dari hasil tersebut, Naïve Bayes Classifier dapat diterapkan untuk mengklasifikasikan postingan hoax di Twitter.

Kata kunci: Klasifikasi, Naïve Bayes Classifier, Hoax, Twitter, 10-Fold Cross Validation.

ABSTRACT

Social media is a means to communicate and exchange information among fellow social media users, Twitter social media is one of them. But the information disseminated is not entirely true, there are some news that do not correspond to the truth or are often called hoaxes. There are many cases of the spread of hoaxes that raise concerns and often harm individuals or groups. so that in this study, the author built a system to identify hoax news on social media Twitter.

This study used Naïve Bayes Classifier method to classify posts with covid-19 topic on Twtter. The post will be classified into two classes, namely hoax class for posts that are classified as incorrect and non-hoax class for posts that are classified as true.

Dataset used for the training process as much as 500 data, with the composition of 250 hoax posts and 250 non-hoax posts. This research begins with the preprocessing process, followed by the training and classification process. Based on evaluation results using 10 fold cross validation, the average accuracy result was 76.%, with an average error of 24%,. Of the 10 folds, the highest accuracy is found in fold 8 with an accuracy of 96%, while the lowest accuracy is at fold 10 of 50%. From these results, Naïve Bayes Classifier can be applied to classify hoax posts on Twitter.

Keywords: *Classification, Naïve Bayes Classifier, Hoax, Twitter, 10-Fold Cross Validation.*