

**SISTEM PENDETEKSI INFORMASI *HOAX* MENGGUNAKAN
ALGORITMA APRIORI DAN ALGORITMA *RANDOM*
FOREST PADA APLIKASI *TWITTER***

Tesis

**Sebagai persyaratan untuk memenuhi
Mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi**



**Mailia Putri Utami
30000319410012**

**SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2021**

HALAMAN PERSETUJUAN

Tesis dengan Judul:

**SISTEM PENDETEKSI INFORMASI *HOAX* MENGGUNAKAN
ALGORITMA APRIORI DAN ALGORITMA *RANDOM*
FOREST PADA APLIKASI *TWITTER***

Disusun Oleh:

Mailia Putri Utami

30000319410012

Telah dilakukan pembimbingan tesis dan dinyatakan layak untuk mengikuti ujian tesis pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Pembimbing I,



Dr. Oky Dwi Nurhayati, ST, MT
NIP. 197910022009122001

Semarang, 18 Desember 2020
Pembimbing II,



Dr. Budi Warsito, S.Si., M.Si.
NIP.197508241999031003

HALAMAN PENGESAHAN

TESIS

SISTEM PENDETEKSI INFORMASI *HOAX* MENGGUNAKAN ALGORITMA
APRIORI DAN ALGORITMA *RANDOM FOREST*
PADA APLIKASI *TWITTER*

Oleh:
Mailia Putri Utami
30000319410012

Telah ditujikan dan dinyatakan lulus ujian tesis pada tanggal 20 Januari 2021 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 20 Januari 2021
Mengetahui,

Penguji I

Jatmiko Endro Suseno, M.Si., Ph.D
NIP. 197211211998021001

Penguji II

Dr. Eng. Adi Wibowo, S.Si., M.Kom.
NIP. 198203092006041002

Pembimbing I

Dr. Oky Dwi Nurhayati, ST, MT
NIP. 197910022009122001

Pembimbing II

Dr. Budi Warsito, S.Si., M.Si.
NIP. 197508241999031003

Mengetahui:
Dekan Sekolah Pascasarjana
Universitas Diponegoro



Dr. R.B. Sulanto, S.H., M.Hum
NIP. 196701011991031005

Ketua Program Studi
Magister Sistem Informasi

Dr. Budi Warsito, S.Si., M.Si.
NIP. 197508241999031003

PERNYATAAN PERSETUJUAN

PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Mailia Putri Utami
NIM : 30000319410012
Program Studi : Magister Sistem Informasi
Program : Sekolah Pascasarjana
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

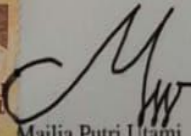
SISTEM PENDETEKSI INFORMASI *HOAX* MENGGUNAKAN ALGORITMA APRIORI DAN ALGORITMA *RANDOM FOREST* PADA APLIKASI *TWITTER*

berserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengolah dalam bentuk pengkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang
Pada tanggal : 20 Januari 2021
Yang menyatakan

Tanda tangan dengan materai 6000
(Sebelum Ujian)




Mailia Putri Utami
Nim. 30000319410012

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 20 Januari 2021




Mailia Putri Utami

KATA PENGANTAR

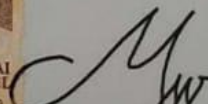
Puji syukur saya panjatkan kepada Allah SWT, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Karya Akhir ini. Penulisan Karya Akhir ini dilakukan dalam rangka memenuhi salah satu syarat dalam mencapai gelar Magister Sistem Informasi di Universitas Diponegoro. Saya menyadari bahwa tanpa bantuan dan bimbingan dari beberapa pihak, sangat banyak kendala yang saya temukan dalam menyelesaikan penelitian ini. Untuk itu, saya mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Dr. Budi Warsito, S.Si., M.Si, selaku Kepala Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro dan selaku Pembimbing II yang telah memberikan ilmu, petunjuk, dorongan serta semangat dalam pembuatan Tesis ini.
2. Ibu Dr. Oky Dwi Nurhayati, ST, MT, selaku Dosen Pembimbing I yang telah memberikan ilmu, petunjuk, dorongan serta semangat dalam pembuatan Tesis ini.
3. Bapak dan Ibu Dosen Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.
4. Seluruh staf karyawan Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.
5. Kedua Orang Tua dan adik, yang telah memberikan perhatian, semangat dan doanya.
6. Keluarga besar Program Studi Magister Sistem Informasi 2019 Sekolah Pascasarjana Universitas Diponegoro.

Akhir kata, saya berharap semoga Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Karya Akhir ini bermanfaat untuk pengembangan ilmu pengetahuan.

Semarang, 20 Januari 2021




Mailia Putri Utami

DAFTAR ISI

	Halaman
Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Pernyataan Persetujuan	iii
Pernyataan	Error! Bookmark not defined.
Kata Pengantar	Error! Bookmark not defined.
Daftar Isi.....	vii
Daftar Gambar.....	ix
Daftar Tabel	x
Daftar Lampiran	xi
Daftar Arti Lambang dan Singkatan	xii
Abstrak	xiii
Abstract	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Tujuan	4
1.3. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Dasar Teori	7
2.2.1. Berita Palsu (<i>Hoax</i>)	7
2.2.2. Media Sosial	9
2.2.3. <i>Twitter</i>	10
2.2.4. <i>Data Mining</i>	11
2.2.5. Algoritma Apriori.....	12
2.2.6. <i>Machine Learning</i>	15
2.2.7. Algoritma <i>Random Forest</i>	15
2.2.8. <i>Text Preprocessing</i>	18
2.2.9. Indeks Evaluasi Kerja.....	19
2.2.10. Metode <i>Undersampling</i>	21
2.2.11. <i>Mean Absolute Error</i>	21
BAB III METODOLOGI PENELITIAN.....	22
3.1. Bahan dan Alat Penelitian	22

3.2. Prosedur Penelitian	23
3.2.1. Proses <i>Learning</i>	25
3.2.2. Proses <i>Prediction</i>	29
3.3. Kerangka Sistem Informasi	30
3.3.1. <i>Input</i>	31
3.3.2. Proses	31
3.3.3. <i>Output</i>	32
BAB IV HASIL DAN PEMBAHASAN	33
4.1. Hasil Penelitian.....	33
4.1.1. Hasil Terhadap Proses Pengambilan Data.....	33
4.1.2. Hasil Terhadap Analisis Proses <i>Learning</i>	33
4.1.3. Hasil Terhadap Analisis Proses <i>Prediction</i>	35
4.1.4. Hasil Pengamatan Percobaan Klasifikasi <i>Hoax</i> dan <i>Non-Hoax</i>	37
4.1.5. Hasil Evaluasi Sistem.....	39
4.2. Pembahasan	40
4.2.1. Pembangunan Sistem	41
4.2.1.1. Teknik Pengambilan Data.....	41
4.2.1.2. <i>Text Preprocessing</i>	42
4.2.1.3. Proses <i>Undersampling</i>	45
4.2.1.4. Proses <i>Training</i>	46
4.2.1.5. Proses <i>Testing</i>	48
4.2.1.6. Proses Klasifikasi	48
4.2.1.7. Proses <i>Prediction</i>	49
4.2.1.8. <i>Confusion Matrix</i>	50
4.2.2. Program Sistem	51
BAB V KESIMPULAN DAN SARAN.....	55
5.1. Kesimpulan.....	55
5.2. Saran	56
DAFTAR PUSTAKA	57
LAMPIRAN.....	60

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Volume, Kebenaran, dan Kecepatan Berita Palsu.....	8
Gambar 3.1 Alur Umum Prosedur Penelitian.	24
Gambar 3.2 Kerangka Sistem Informasi.	30
Gambar 4.1 Halaman Hasil Proses Percobaan Pengamatan Terhadap Apriori Salah.	38
Gambar 4.2 Halaman Hasil Proses Percobaan Pengamatan Terhadap Apriori Benar.	39
Gambar 4.3 Proses <i>Case Folding</i>	44
Gambar 4.4 Proses <i>Tokenizing</i>	44
Gambar 4.5 Proses <i>Filtering (Stopword Removal)</i>	45
Gambar 4.6 Proses <i>Stemming</i>	45
Gambar 4.7 Tampilan Halaman <i>Login User</i>	51
Gambar 4.8 Tampilan Halaman <i>Session Berhasil Login</i>	52
Gambar 4.9 Tampilan Halaman <i>Session Gagal Login</i>	52
Gambar 4.10 Tampilan Halaman Utama Program.....	53
Gambar 4.11 Tampilan Halaman Pemilihan Sumber Berita.....	53
Gambar 4.12 Tampilan Halaman <i>Session Save Excel</i>	53
Gambar 4.13 Tampilan Halaman <i>Excel Hasil Program</i>	54
Gambar 4.14 Tampilan Halaman <i>Excel Hasil Program Nilai Dari Confusion Matrix</i>	54

DAFTAR TABEL

	Halaman
Tabel 2.1 Matriks Klasifikasi Model Dua Kelas.....	20
Tabel 4.1 <i>Word Occurrence</i> dari Proses Algoritma Apriori Salah	34
Tabel 4.2 <i>Word Occurrence</i> dari Proses Algoritma Apriori Benar	35
Tabel 4.3 Asumsi Nilai dari <i>Mean Absolute Error</i> Berdasarkan Pengamatan Pada Proses Prediksi.....	36
Tabel 4.4 Hasil Prediksi dan Dilakukan Pengamatan Bersarkan Bentuk dari <i>Confusion Matrix</i>	39

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Hasil Deteksi Secara Manual.....	60
Lampiran 2. Sumber Data Proses Learning	65
Lampiran 3. Beberapa Data Proses Learning.....	66
Lampiran 4. Beberapa Berita Yang Diolah Dari Aplikasi Twitter	69

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti
n	Jumlah pengamatan
p	Variabel penjelas
m	Variabel penjelas secara <i>random</i>
$L(\theta)$	Sampel <i>bootstrap</i>
(L)	Set pembelajaran
$TL(\theta)$	<i>Tree</i> prediktor
θ	Vektor acak
k	Indeks <i>tree</i>
y'	Hasil prediksi
<i>Support</i> (X)	Ukuran frekuensi <i>itemset</i> dalam <i>database</i>
N	Total ukuran frekuensi <i>itemset</i> dalam <i>database</i>
k	Indeks <i>itemset</i>
Tp	Jumlah kelompok (kelas) positif yang diklasifikasi sebagai positif
Fp	Jumlah kelompok (kelas) negatif yang diklasifikasi sebagai positif
Tn	Jumlah kelompok (kelas) positif yang diklasifikasi sebagai negative
Fn	Jumlah kelompok (kelas) negatif yang diklasifikasi sebagai negative
X_{ip}	Nilai hasil peramalan atau perkiraan
X_{im}	Nilai terukur dari indeks ke- i
n	Jumlah keseluruhan data yang digunakan
<i>MAE</i>	<i>Mean absolute error</i>
Σ	Operator penjumlahan

SINGKATAN

Singkatan	Kepanjangan Singkatan
Corpus	Penggunaan bahasa dalam bentuk tertulis atau lisan
AI	<i>Artificial Intellegent</i>
KDD	<i>Knowledge Discovery in Database</i>
OOB	<i>Out-Of-Bag</i>
URL	<i>Uniform Resource Locator</i>
PC	<i>Personal Computer</i>
API	<i>Application Programming Integration</i>
GUI	<i>Graphic User Interface</i>
MAE	<i>Mean absolute error</i>
RAM	<i>Random Access Memory</i>

SISTEM PENDETEKSI INFORMASI *HOAX* MENGGUNAKAN ALGORITMA APRIORI DAN ALGORITMA *RANDOM FOREST* PADA APLIKASI *TWITTER*

ABSTRAK

Penelitian ini didasari dengan adanya gangguan yang sering kali ditemukan pengguna *social media Twitter*, yaitu terkait penyebaran informasi *hoax* berbentuk teks. Salah satu upaya untuk menanggulangi permasalahan tersebut yaitu membangun sistem untuk mendeteksi berita *hoax* pada aplikasi *Twitter*. Kumpulan dari informasi berbentuk teks pada *social media*, dapat menangkap penggunaan Bahasa dalam bentuk tertulis maupun lisan (*corpus*). Berdasarkan kelebihan dari algoritma Apriori yaitu algoritma tersebut mampu menambang data teks dari banyaknya *dataset* yang digunakan, serta mampu menemukan pola hubungan atau kombinasi *itemset* dalam suatu *database* sebagai rekomendasi dari pola yang dibangkitkan. Selain itu, sistem pendeteksi juga diperlukan metode untuk mengklasifikasikan informasi berdasarkan kelasnya *hoax* atau *non-hoax*. Salah satu algoritma yang akan digunakan yaitu menggunakan algoritma *Random Forest*, yang mana algoritma tersebut mampu mengekstrak dari bentuk *decision tree* menjadi sebuah keputusan yang terlihat jelas dengan kata lain mampu mengklasifikasi dengan cepat dan mampu menghasilkan solusi keputusan yang paling optimal. Tujuan dari penelitian ini selain menerapkan dan mengintegrasikan Algoritma Apriori dan Algoritma *Random Forest* adalah untuk mempermudah peneliti menganalisis dan mengevaluasi hasil dari sistem yang dibangun dalam mendeteksi informasi *hoax* dengan tingkat akurasi yang paling optimal. Hasil penelitian menunjukkan bahwa sistem dapat mendeteksi berita *hoax* dan *non-hoax* yang datanya terintegrasi langsung dengan aplikasi *Twitter*, yang mana memiliki nilai *mean absolute error* mencapai 0 untuk masing-masing percobaan dalam mengklasifikasikan berita *hoax* dan berita *non-hoax*. Tingkat akurasi dari sistem yang telah dibangun dalam mendeteksi informasi berita pada aplikasi *Twitter* dengan nilai *minimum support* 2 yaitu 99.4%.

Kata Kunci : Algoritma Apriori, *Text Mining*, Algoritma *Random Forest*, *Machine Learning*, Pendeteksi *hoax*

HOAX INFORMATION DETECTION SYSTEM USING APRIORI ALGORITHM AND RANDOM FOREST ALGORITHM IN TWITTER APPLICATION

ABSTRACT

This research is based on the disturbance that is often found by Twitter social media users, namely related to the dissemination of hoax information in the form of text. One of the efforts to overcome this problem is to build a system to detect hoax news on the Twitter application. A collection of information in the form of text on social media, can capture the use of language in written and spoken form (corpus). Based on the advantages of the Apriori algorithm, the algorithm is able to mine text data from the number of datasets used, and is able to find relationship patterns or itemset combinations in a database as a recommendation of the generated patterns. In addition, a detection system also requires a method to classify information based on its class as hoax or non-hoax. One of the algorithms to be used is the Random Forest algorithm, in which the algorithm is able to extract from the form of a decision tree into a clearly visible decision, in other words it is able to classify quickly and to produce the most optimal decision solution. The purpose of this research, apart from implementing and integrating the Apriori Algorithm and the Random Forest Algorithm, is to make it easier for researchers to analyze and evaluate the results of the system built to detect hoax information with the most optimal level of accuracy. The results showed that the system can detect hoax and non-hoax news, whose data is directly integrated with the Twitter application, which has a mean absolute error value of 0 for each experiment in classifying hoax news and non-hoax news. The level of accuracy of the system that has been built in detecting news information on the Twitter application with a minimum value of support 2 is 99.4%.

Keyword : Apriori algorithm, Text Mining, Random Forest algorithm, Machine Learning, Hoax Detection