



**KLASIFIKASI TEKS PADA BERITA *HOAX* DI INDONESIA
MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON***

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
RABIATUL ZANNAH
NIM. 1911017220005**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2026**



**KLASIFIKASI TEKS PADA BERITA *HOAX* DI INDONESIA
MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON***

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
RABIATUL ZANNAH
NIM. 1911017220005**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2026**

SKRIPSI

KLASIFIKASI TEKS PADA BERITA HOAX DI INDONESIA MENGGUNAKAN METODE MULTI-LAYER PERCEPTRON

Oleh:
Rabiatul Zannah
191101722005

Telah dipertahankan pada hari Rabu, tanggal 1-07-2026 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



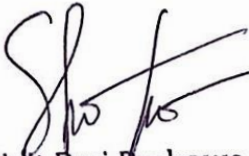
Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

Penguji I



Selvi Annisa, S.Si., M.Si.
NIP. 199212262022032016

Pembimbing II



Sigit Dwi Prabowo, S.Mat., M.Stat.
NIP. 199605282024061003

Penguji II



Hj. Maisarah, S.Pd.I., M.Pd.
NIP. 19890713201801213001

Banjarbaru, 23 Juli 2026

Mengetahui,

Koordinator Program Studi Statistika
FMIPA ULM



Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

PERNYATAAN

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

Banjarbaru, 1 Juli 2026



Rabiatul Zannah
NIM. 1911017220005

PRODI STATISTIKA

ABSTRAK

Klasifikasi Teks pada Berita *Hoax* di Indonesia Menggunakan Metode *Multi-layer Perceptron* (Oleh: Rabiatul Zannah; Dewi Anggraini dan Sigit Dwi Prabowo, 2026; 39 halaman)

Pesatnya perkembangan ilmu pengetahuan dan teknologi saat ini sangat berpengaruh pada penyebaran informasi yakni dengan memanfaatkan penggunaan internet. Namun dalam penyebarannya, pengguna internet sendiri bisa saja abai dalam pengecekan kebenaran dari informasi yang disebarkan tersebut hingga menimbulkan kerugian apalagi jika informasi tersebut skala nasional maka lebih banyak lagi pihak yg dirugikan akibat kesalahan pada informasi yang disebarkan. Penelitian ini bertujuan untuk mengklasifikasikan data berupa teks pada berita *hoax* di Indonesia pada tahun 2015-2020. Data yang digunakan merupakan data sekunder dari *website* Kaggle yang berjumlah 4,231 data. Metode yang digunakan ialah *multi-layer perceptron* dengan klasifikasi dua kelas yaitu *hoax* dan tidak *hoax*. Hasil yang diperoleh antara lain *wordcloud* dan diagram batang yang menampilkan kata "orang" sebagai kata yang paling sering kemunculannya pada berita *hoax* di Indonesia pada tahun 2015-2020. Model *multi-layer perceptron* pada klasifikasi teks berita *hoax* di Indonesia dibangun menggunakan arsitektur 1 *hidden layer* dengan 50 *neuron*, fungsi aktivasi ReLU pada *hidden layer* dan Sigmoid pada *output layer*, serta *optimizer* RMSProp dengan *learning rate* 0.001 telah berhasil dilatih untuk mengklasifikasikan berita *hoax* dan tidak *hoax*. Adapun hasil evaluasi performa model diantaranya akurasi sebesar 80.91% terlihat cukup baik, namun tidak cukup untuk menyimpulkan performa model terutama pada kasus data yang tidak seimbang (*imbalanced*). Oleh karena itu, penting untuk melihat metrik lain seperti *Specificity*, Presisi, *Recall* (Sensitivitas), dan *F1-Score* secara bersamaan. Kombinasi Presisi 86.04% dan *Recall* 91.87% yang tinggi ini tercermin pada *F1-Score* sebesar 88.86% yang mengindikasikan model cukup andal dalam mengklasifikasikan label *hoax* secara seimbang antara ketepatan dan kelengkapan deteksi. Namun, *Specificity* yang hanya 28.04% menunjukkan bahwa model kesulitan mengenali data sebagai label tidak *hoax*.

Kata Kunci: Berita *Hoax*, Klasifikasi Teks, *Multi-layer Perceptron*

ABSTRACT

Text Classification of Hoax News in Indonesia Using the Multi-layer Perceptron Method (By: Rabiatul Zannah; Dewi Anggraini and Sigit Dwi Prabowo, 2026; 39 pages)

The rapid advancement of science and technology today has a significant impact on the dissemination of information through the use of the internet. However, in the process of dissemination, internet users themselves may neglect to verify the accuracy of the information being shared, leading to harm, especially if the information is on a national scale, in which case even more parties are harmed as a result of errors in the disseminated information. This study aims to classify text data from hoax news in Indonesia from 2015 to 2020. The data used consists of 4,231 secondary data points from the Kaggle website. The method employed is a multi-layer perceptron with a two-class classification: hoax and non-hoax. The results include a word cloud and a bar chart showing that the word "Indonesia" appeared most frequently in hoax news in Indonesia from 2015 to 2020. The multi-layer perceptron model for classifying hoax news articles in Indonesia built using an architecture with one hidden layer containing 50 neurons, a ReLU activation function in the hidden layer and a Sigmoid activation function in the output layer, and an RMSProp optimizer with a learning rate of 0.001 was successfully trained to classify hoaxes and non-hoaxes. The results of the model performance evaluation including an accuracy of 80.91% appear quite good, but are not sufficient to draw conclusions about the model's performance, especially in cases of imbalanced data. Therefore, it is important to consider other metrics such as Specificity, Precision, Recall (Sensitivity), and F1-Score simultaneously. This combination of high Precision (86.04%) and Recall (91.87%) is reflected in an F1-Score of 88.86%, indicating that the model is quite reliable in classifying hoax labels while maintaining a balance between accuracy and completeness of detection. However, a Specificity of only 28.04% shows that the model struggles to identify data as non-hoax.

Keywords: Hoax News, Text Classification, Multi-layer Perceptron

PRAKATA

Dengan syukur kehadiran Allah *Subhanahu wa ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan Skripsi yang berjudul **“Klasifikasi Teks pada Berita Hoax di Indonesia Menggunakan Metode *Multi-layer Perceptron*”**. Tujuan penulisan Skripsi ini ialah untuk memenuhi salah satu persyaratan dalam menyelesaikan program sarjana di Program Studi Statistika FMIPA ULM.

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bantuan dan dukungan dalam menyelesaikan penelitian ini, diantaranya:

1. Kedua orang Tua dan kedua kakak penulis yang selalu mendoakan serta mendukung penulis sepenuhnya, baik secara moral maupun material.
2. Ibu Dewi Anggraini S.Si., M.App.Sci., Ph.D selaku Koordinator Program Studi Statistika FMIPA ULM dan dosen pembimbing akademik serta dosen pembimbing Skripsi yang telah memberikan arahan dan bimbingan selama perkuliahan hingga penyusunan Skripsi ini.
3. Bapak Sigit Dwi Prabowo, S.Mat., M.Stat. selaku dosen pembimbing yang telah bersedia meluangkan waktu untuk memberikan bimbingan, dan bantuan dalam melakukan penelitian dan penyusunan Skripsi ini.
4. Ibu Selvi Annisa, S.Si., M.Si. dan Ibu Hj. Maisarah, S.Pd.I., M.Pd. selaku dosen penguji yang telah memberikan arahan dan masukan dalam rangka perbaikan Skripsi ini.
5. Seluruh jajaran dosen dan staff Program Studi Statistika FMIPA ULM yang telah memberikan bantuan dan dukungan dalam penyusunan Skripsi ini.
6. Teman-teman dan semua pihak yang telah membantu dan memotivasi baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa Skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan adanya kritik dan saran terkait penulisan Skripsi ini dari semua pihak sebagai evaluasi untuk penulisan berikutnya. Akhir kata, penulis berharap agar Skripsi ini dapat bermanfaat bagi semua pihak.

Banjarbaru, 1 Juli 2026

Rabiatul Zannah

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kajian Penelitian Terdahulu	4
2.2 Analisis Teks.....	5
2.3 Ekstraksi Fitur (<i>Feature Extraction</i>).....	7
2.4 Klasifikasi Teks (<i>Text Clasification</i>).....	7
2.5 <i>Multi-layer Perceptron</i>	8
2.6 Fungsi Aktivasi (<i>Activation Function</i>).....	11
2.6.1 ReLU	11
2.6.2 Sigmoid.....	12
2.7 <i>Loss Function</i>	12
2.8 <i>Optimizer Function</i>	12
2.9 Evaluasi Model Klasifikasi	13
2.10 Berita <i>Hoax</i>	14
BAB III METODE PENELITIAN	15
3.1 Jenis dan Sumber data	15
3.2 Variabel Penelitian	15
3.3 Prosedur Penelitian	15
BAB IV HASIL DAN PEMBAHASAN	18
4.1 <i>Preprocessing</i> data	18
4.1.1 Data.....	18
4.1.2 Pembersihan Teks	20
4.1.3 Koreksi Ejaan.....	21
4.1.4 <i>Stemming</i> dan <i>Lemmatization</i>	22
4.1.5 Penghapusan <i>Stopwords</i> dan Tokenisasi.....	23
4.2 Visualisasi Data	24
4.3 Ekstraksi Fitur	25
4.4 Pemodelan <i>Multi-layer Perceptron</i>	27

4.4.1 Analisis Deskriptif.....	28
4.4.2 Pembagian Data	28
4.4.3 <i>Multi-layer Perceptron</i> Model	29
4.5 Evaluasi Model Klasifikasi	36
BAB V PENUTUP	38
5.1 Kesimpulan	38
5.2 Saran.....	38
DAFTAR PUSTAKA	40
LAMPIRAN	43
RIWAYAT HIDUP	56

PRODI STATISTIKA

DAFTAR GAMBAR

Gambar 2.1 Jaringan <i>Multilayer Perceptron</i>	8
Gambar 3.1 Alur Prosedur Penelitian.....	17
Gambar 4.1 <i>Wordcloud</i> dengan 40 Kata Teratas yang Muncul dari Kolom Narasi pada Berita <i>Hoax</i> di Indonesia.....	24
Gambar 4.2 Diagram Batang Berdasarkan Nilai Frekuensinya dengan 10 kata Teratas dari Data Berita <i>Hoax</i> di Indonesia	25
Gambar 4.3 Grafik <i>Loss Training</i> dan <i>Loss Validasi</i>	35
Gambar 4.4 Grafik Akurasi <i>Training</i> dan Akurasi Validasi.....	35

PRODI STATISTIKA

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu	4
Tabel 2.2	<i>Confusion Matrix</i> dari Klasifikasi Dua Kelas	13
Tabel 4.1	Data pada Berita <i>Hoax</i> di Indonesia	18
Tabel 4.2	Hasil Pembersihan Teks Kolom Narasi pada Berita <i>Hoax</i> di Indonesia.....	20
Tabel 4.3	Hasil Koreksi Ejaan Kolom Narasi pada Berita <i>Hoax</i> di Indonesia	21
Tabel 4.4	Hasil <i>Stemming</i> dan <i>Lemmatization</i> Kolom Narasi pada Berita <i>Hoax</i> di Indonesia.....	22
Tabel 4.5	Hasil Penghapusan <i>Stopwords</i> dan Tokenisasi Kolom Narasi pada Berita <i>Hoax</i> di Indonesia	23
Tabel 4.6	Nilai Frekuensi Kemunculan Kata, TF, IDF, dan TF-IDF Kolom Narasi pada Berita <i>Hoax</i> di Indonesia.....	26
Tabel 4.7	Distribusi Kelas pada Data Berita <i>Hoax</i> di Indonesia	28
Tabel 4.8	Pembagian Data <i>Training</i> dan Data <i>Testing</i>	28
Tabel 4.9	Data <i>Training</i>	28
Tabel 4.10	Data <i>Testing</i>	29
Tabel 4.11	Nilai Bobot dan Bias Antara <i>Input Layer</i> (i) dengan <i>Hidden Layer</i> (j).....	30
Tabel 4.12	Nilai Bobot dan Bias Antara <i>Hidden Layer</i> (j) dengan <i>Output Layer</i> (k)	30
Tabel 4.13	<i>Confusion Matrix</i> Model <i>Multi-layer Perceptron</i>	36
Tabel 4.14	Hasil Kriteria Evaluasi Model Klasifikasi.....	36

DAFTAR LAMPIRAN

Lampiran 1. Data Berita <i>Hoax</i> Di Indonesia pada Tahun 2015-2020	43
Lampiran 2. Data pada Kolom ke-5 Narasi pada Berita <i>Hoax</i> Di Indonesia pada Tahun 2015-2020 Setelah <i>Preprocessing</i> Data.....	45
Lampiran 3. Hasil Perhitungan TF-IDF Kolom ke-5 Narasi pada Berita <i>Hoax</i> Di Indonesia pada Tahun 2015-2020	48
Lampiran 4. Data <i>Training</i> dan Data <i>Testing</i> Kolom ke-5 Narasi pada Berita <i>Hoax</i> Di Indonesia pada Tahun 2015-2020.....	50
Lampiran 5. <i>Syntax</i> pada Aplikasi R	51

PRODI STATISTIKA

DAFTAR LAMBANG

w_{dt}	bobot dokumen ke-d terhadap kata ke-t
TF_{dt}	banyaknya kata ke-t yang dicari dari sebuah dokumen ke-d
IDF_{ft}	<i>Invers Document Frequency</i> dari frekuensi ke-f dari kata ke-t
n	total dokumen
df	banyaknya dokumen ke-d yang mengandung kata yang dicari
x_i	nilai <i>neuron input layer</i> ke-i
v_{ij}	bobot antara <i>input layer</i> ke-i dan <i>hidden layer</i> ke-j
v_{0j}	bias pada <i>neuron hidden layer</i> ke-j
z_{in_j}	sinyal dari <i>neuron input layer</i> ke-i menuju <i>neuron hidden layer</i> ke-j
z_j	nilai <i>neuron hidden layer</i> ke-j
w_{jk}	bobot antara <i>hidden layer</i> ke-j dan <i>output layer</i> ke-k
w_{0k}	bias <i>neuron output layer</i> ke-k
y_{in_k}	sinyal dari <i>neuron hidden layer</i> ke-j menuju <i>neuron output layer</i> ke-k
δ_k	nilai kesalahan <i>neuron output layer</i> ke-k
t_k	nilai target (nilai sebenarnya) pengamatan ke-k
y_k	nilai <i>neuron output layer</i> pengamatan ke-k (nilai prediksi)
$f'(y_{in_k})$	fungsi aktivasi dari sinyal dari <i>neuron hidden layer</i> ke-j menuju <i>neuron output layer</i> ke-k
α	<i>learning rate</i>
$g_{t_{jk}}$	nilai <i>gradient descent epoch</i> ke-t pada bobot antara <i>hidden layer</i> ke-j dan <i>output layer</i> ke-k
δ_{in_j}	nilai kesalahan pada sinyal dari <i>neuron output layer</i> ke-k menuju <i>neuron hidden layer</i> ke-j
δ_j	nilai kesalahan <i>neuron output layer</i> ke-j
$g_{t_{ij}}$	nilai <i>gradient descent epoch</i> ke-t pada bobot antara <i>input layer</i> ke-i dan <i>hidden layer</i> ke-j
w_{jk}	bobot antara <i>hidden layer</i> ke-j dan <i>output layer</i> ke-k
Δw_{jk}	suku perubahan bobot w_{jk}
v_{ij}	bobot antara <i>input layer</i> ke-i dan <i>hidden layer</i> ke-j
Δv_{ij}	suku perubahan bobot v_{ij}
y	label kelas nilai sebenarnya (0 atau 1)
p	probabilitas prediksi untuk kelas positif (1)
$1 - p$	probabilitas prediksi untuk kelas negatif (0)
w_t	bobot <i>epoch</i> ke-t
w_{avgt}	bobot rata-rata <i>epoch</i> ke-t
g_t	nilai <i>gradient descent epoch</i> ke-t
ε	bilangan kecil untuk mencegah pembagian nol