

**SISTEM IDENTIFIKASI BERITA PALSU YANG  
TERDAPAT DI WEB DENGAN MENGGUNAKAN  
METODE MACHINE LEARNING**

**TUGAS AKHIR**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**



**OLEH :**

**MUHAMMAD TAMLIKHO**

**09011281621031**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA  
2020**

**HALAMAN PENGESAHAN**

**SISTEM IDENTIFIKASI BERITA PALSU YANG TERDAPAT DI WEB  
DENGAN MENGGUNAKAN METODE *MACHINE LEARNING***

**TUGAS AKHIR**

**Sebagai salah satu syarat untuk menyelesaikan  
Program Studi Sistem Komputer  
Jenjang S1**

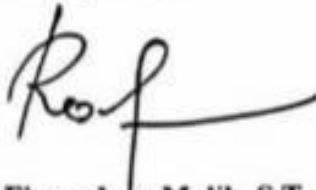
Oleh

**Muhammad Tamlikho  
09011281621031**

**Palembang, 22 Juli 2020**

**Mengetahui,**

**Pembimbing I Tugas Akhir**



**Reza Firsandaya Malik, S.T., M.T., Ph.D.  
NIP. 197604252010121001**

**Pembimbing II Tugas Akhir**



**Firdaus, S.T., M.Kom.  
NIP. 197801212008121003**

**Ketua Jurusan Sistem Komputer**



**Dr. Ir. Sukemi, M.T.  
NIP. 196612032006041001**

## HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:

Hari : Rabu

Tanggal : 1 Juli 2020

### Tim Penguji:

1. Ketua : Kemahyanto Exaudi Siahaan. S.Kom. MT.

2. Sekretaris I : Reza Firsandaya Malik, S.T., M.T., Ph.D.

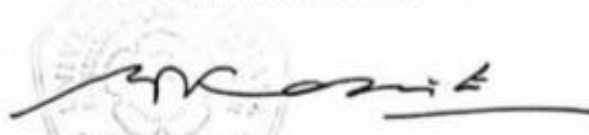
3. Sekretaris II : Firdaus, S.T., M.Kom.

4. Anggota I : Aditya Putra Perdana. S.Kom. M.T.

5. Anggota II : Rahmat Fadli Isnanto, M.T.

Mengetahui,

Ketua Jurusan Sistem Komputer



Dr. Ir. H. Sukemi, M.T.  
NIP. 196612032006041001

## HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Muhammad Tamlikho

NIM : 09011281621031

Judul : Sistem Identifikasi Berita Palsu yang Terdapat di Web dengan Menggunakan Metode *Machine Learning*

**Hasil pengecekan *Software iThenticate/Turnitin* : 4%**

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas Sriwijaya. Demikian, pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.



Palembang, 2 Juli 2020



Muhammad Tamlikho

## KATA PENGANTAR



Alhamdulillah rabbil'alamin. Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Proposal Tugas Akhir ini dengan judul “Sistem Identifikasi Berita Palsu yang Terdapat di Web dengan Menggunakan Metode *Machine Learning*”.

Dalam laporan ini penulis menjelaskan mengenai pengklasifikasian berita palsu dan berita asli dengan menggunakan *Random Forest Classifier* (RFC), *Support Vector Machine* (SVM), *Logistic Regression Classifier* (LRC), *Deep Neural Network* (DNN) dan *K-Nearest Neighbor* (KNN) dengan disertai data-data yang diperoleh penulis saat melakukan pengujian. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak.

Pada penyusunan laporan ini, penulis banyak mendapatkan ide dan saran serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah SWT dan terimakasih kepada yang terhormat:

1. Orang tua saya tercinta yang telah membesarkan saya dengan penuh kasih sayang dan selalu mengajarkan saya dalam berbuat hal yang baik. Terimakasih untuk segala doa, motivasi dan dukungannya baik moril, materiil maupun spiritual selama ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rossi Passarella, S.T., M.Eng., selaku Penasihat Akademik Jurusan Sistem Komputer.

5. Bapak Dr. Reza Firsandaya Malik, Ph.D., selaku Pembimbing I Tugas Akhir yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Tugas Akhir ini.
6. Bapak Firdaus, S.T., M.Kom., selaku Pembimbing II Tugas Akhir yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Tugas Akhir ini.
7. Teman-teman saya yang telah membantu saya dan telah memberi motivasi untuk penulis dalam menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa Laporan ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi di kemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu pembelajaran.

Palembang, Juli 2020



**Muhammad Tamlikho**  
**NIM. 09011281621031**

## ***Fake News Identification System on the Web Using Machine Learning Methods***

**Muhammad Tamlikho (09011281621031)**

*Computer Engineering Department, Computer Science Faculty,  
Sriwijaya University*

Email : tamlikhooo@gmail.com

### ***Abstract***

*With the development of the internet which continues to grow more rapidly, the internet has turned into a source of information that can be easily accessed. The development of information technology such as social media and news contained on the web helped to push the spread of fake news very easy and quick. In this study the classification was carried out using five classifiers, namely Random Forest Classifier (RFC), Support Vector Machine (SVM), Logistic Regression Classifier (LRC), Deep Neural Network (DNN) and K-Nearest Neighbor (KNN). The classification was carried out in thirty-seven experiments (models / scenarios) for two datasets. From all experiments, each classifier has the best model / scenario which is judged by the good performance value generated. Among the models / scenarios of the experiments carried out, there is each one of the best models / scenarios that produce the highest accuracy value for each dataset used. the model / scenario is also the best model / scenario for all experiments of all classifiers for each dataset used.*

**Keywords :** *Fake News Detection, Machine Learning, Random Forest Classifier, Support Vector Machine, Logistic Regression, Deep Neural Network, K-Nearest Neighbor.*

# **Sistem Identifikasi Berita Palsu yang Terdapat di Web dengan Menggunakan Metode *Machine Learning***

**Muhammad Tamlikho (09011281621031)**

*Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya*

Email : tamlikhooo@gmail.com

## **Abstrak**

Dengan perkembangan internet yang semakin hari terus berkembang semakin pesat, internet berubah menjadi sumber informasi yang dapat dengan mudah diakses. Perkembangan teknologi informasi seperti media sosial dan berita yang terdapat di web turut serta mendorong penyebaran berita palsu dengan sangat mudah dan sangat cepat. Pada penelitian ini klasifikasi dilakukan dengan menggunakan lima buah *classifier*, yaitu *Random Forest Classifier* (RFC), *Support Vector Machine* (SVM), *Logistic Regression Classifier* (LRC), *Deep Neural Network* (DNN) dan *K-Nearest Neighbor* (KNN). Pengklasifikasian dilakukan dalam tiga puluh tujuh kali percobaan (model/skenario) untuk dua buah dataset. Dari semua percobaan, masing-masing *classifier* memiliki sebuah model/skenario terbaik yang dinilai dari bagusnya nilai performa yang dihasilkan. Diantara model/skenario percobaan yang dilakukan, terdapat masing-masing satu model/skenario terbaik yang menghasilkan nilai akurasi tertinggi untuk masing-masing dataset yang digunakan. model/skenario tersebut sekaligus menjadi model/skenario terbaik untuk semua percobaan dari semua *classifier* untuk masing-masing dataset yang digunakan.

**Kata Kunci :** *Fake News Detection, Machine Learning, Random Forest Classifier, Support Vector Machine, Logistic Regression, Deep Neural Network, K-Nearest Neighbor.*



## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vii</b>
<b>Abstrak</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>ix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Tujuan dan Manfaat.....	2
1.2.1. Tujuan .....	2
1.2.2. Manfaat .....	3
1.3. Perumusan dan Batasan Masalah .....	3
1.3.1. Perumusan Masalah .....	3
1.3.2. Batasan Masalah.....	3
1.4. Metodologi Penelitian .....	4
1.4.1. Metode Studi Pustaka dan Literatur .....	4
1.4.2. Metode Konsultasi .....	4
1.4.3. Metode Pembuatan Model .....	5
1.4.4. Metode Pengujian dan Validasi .....	5

1.4.5.	Metode Hasil dan Analisa .....	5
1.4.6.	Metode Penarikan Kesimpulan dan Saran .....	5
1.5.	Sistematika Penulisan.....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>		<b>7</b>
2.1.	Misinformasi yang Terdapat di Web.....	7
2.1.1.	Berita Palsu .....	8
2.1.2.	Rumor.....	8
2.1.3.	Lain-Lain.....	9
2.2.	Model Deteksi Berita Palsu .....	9
2.3.	Pra-Pemrosesan Data.....	11
2.3.1.	Case Folding.....	11
2.3.2.	Punctuation Removal .....	12
2.3.3.	Number Removal .....	12
2.3.4.	Removal Word < N Character.....	13
2.3.5.	Stop Word Removal.....	14
2.3.6.	Stemming .....	15
2.3.7.	Lemmatization.....	16
2.4.	Ekstraksi Dan Seleksi Fitur .....	17
2.4.1.	Term-Frequency .....	17
2.4.2.	Invers Document Frequency .....	18
2.4.3.	Term Frequency-Invers Document Frequency (TF-IDF) .....	19
2.5.	Machine Learning.....	19
2.5.1.	Random Forest Classifier .....	20
2.5.2.	Support Vector Machine .....	22
2.5.3.	Logistic Regression.....	24

2.5.4.	Deep Neural Network .....	28
2.5.5.	K-Nearest Neighbor .....	31
2.6.	Performanca Measurements .....	33
<b>BAB III METODOLOGI .....</b>		<b>35</b>
3.1.	Gambaran Umum .....	35
3.2.	Akuisisi Data .....	36
3.3.	Persiapan Data .....	37
3.4.	Pra-Pemrosesan Data.....	37
3.5.	Ekstraksi Dan Seleksi Fitur .....	40
3.6.	Klasifikasi.....	40
3.6.1.	Random Forest Classifier.....	40
3.6.2.	Support Vector Machine Classifier.....	42
3.6.3.	Logistic Regression Classifier.....	43
3.6.4.	Klasifikasi Deep Neural Network .....	43
3.6.5.	Klasifikasi K-Nearest Neighbor.....	45
3.7.	Evaluasi Model.....	45
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>49</b>
4.1.	Hasil Akuisisi Data.....	49
4.2.	Hasil Persiapan Data.....	49
4.3.	Hasil Pra-Pemrosesan Data .....	50
4.4.	Hasil Ekstraksi Dan Seleksi Fitur.....	51
4.5.	Hasil Klasifikasi .....	53
4.5.1.	Hasil Klasifikasi RFC .....	53
4.5.2.	Hasil Klasifikasi SVM .....	56
4.5.3.	Hasil Klasifikasi LRC .....	60
4.5.4.	Hasil Klasifikasi DNN .....	64

4.5.5.	Hasil Klasifikasi KNN .....	68
4.6.	Komparasi Hasil .....	72
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>76</b>
5.1.	Kesimpulan.....	76
5.2.	Saran.....	78
<b>DAFTAR PUSTAKA .....</b>		<b>79</b>

## DAFTAR GAMBAR

<b>Gambar 2.1.</b> Jenis-Jenis Misinformasi yang terdapat di Web [7].....	7
<b>Gambar 2.2.</b> Proses Untuk Mendeteksi Berita Palsu [2] .....	10
<b>Gambar 2.3.</b> Contoh <i>Case Folding</i> .....	11
<b>Gambar 2.4.</b> Contoh <i>Punctuation Removal</i> .....	12
<b>Gambar 2.5.</b> Contoh <i>Number Removal</i> .....	13
<b>Gambar 2.6.</b> Contoh <i>Remove Word Contain &lt; N Character</i> .....	13
<b>Gambar 2.7</b> Contoh <i>Stop Word Removal</i> .....	14
<b>Gambar 2.8.</b> Ilustrasi Klasifikasi <i>Random Forest</i> [26].....	21
<b>Gambar 2.9.</b> Ilustrasi Klasifikasi <i>Support Vector Machine</i> [34] .....	23
<b>Gambar 2.10.</b> Contoh Kasus 1[36] .....	25
<b>Gambar 2.11.</b> Contoh Kasus 2 [36] .....	26
<b>Gambar 2.12.</b> Penerapan Regresi Linear pada Logistic Regression [36] .....	27
<b>Gambar 2.13.</b> Ilustrasi perkembangan Deep Neural Network [39] .....	29
<b>Gambar 2.14.</b> Ilustrasi K-Nearest Neighbor [44] .....	31
<b>Gambar 3.1.</b> Flow Chart Penelitian .....	35
<b>Gambar 3.2.</b> Arsitektur Untuk Membuat Sistem Identifikasi Berita Palsu .....	38
<b>Gambar 3.3.</b> Alur Pra-Pemrosesan Data.....	39
<b>Gambar 4.1.</b> Data Hasil dari Tahapan Pra-Pemrosesan Data .....	51
<b>Gambar 4.2.</b> Hasil TF-IDF dari Data Training.....	52
<b>Gambar 4.3.</b> Visualisasi Karakteristik Berita dari TF-IDF .....	52
<b>Gambar 4.4.</b> Detail Hasil Klasifikasi RFC Untuk Dataset Pertama .....	53
<b>Gambar 4.5.</b> Detail Hasil Klasifikasi RFC Untuk Dataset Kedua.....	54
<b>Gambar 4.6.</b> <i>Performance Measurements</i> RFC Untuk Dataset Pertama.....	55

<b>Gambar 4.7.</b> <i>Performance Measurements</i> RFC Untuk Dataset Kedua .....	56
<b>Gambar 4.8.</b> Detail Hasil Klasifikasi SVM Untuk Dataset Pertama .....	57
<b>Gambar 4.9.</b> Detail Hasil Klasifikasi SVM Untuk Dataset Kedua.....	58
<b>Gambar 4.10.</b> <i>Performance Measurements</i> SVM Untuk Dataset Pertama.....	59
<b>Gambar 4.11.</b> <i>Performance Measurements</i> SVM Untuk Dataset Kedua .....	60
<b>Gambar 4.12.</b> Detail Hasil Klasifikasi LRC Untuk Dataset Pertama .....	61
<b>Gambar 4.13.</b> Detail Hasil Klasifikasi LRC Untuk Dataset Kedua.....	62
<b>Gambar 4.14.</b> <i>Performance Measurements</i> LRC Untuk Dataset Pertama .....	63
<b>Gambar 4.15.</b> <i>Performance Measurements</i> LRC Untuk Dataset Kedua .....	64
<b>Gambar 4.16.</b> Detail Hasil Klasifikasi DNN Untuk Dataset Pertama .....	65
<b>Gambar 4.17.</b> Detail Hasil Klasifikasi DNN Untuk Dataset Kedua .....	66
<b>Gambar 4.18.</b> <i>Performance Measurements</i> DNN Untuk Dataset Pertama .....	67
<b>Gambar 4.19.</b> <i>Performance Measurements</i> DNN Untuk Dataset Kedua .....	68
<b>Gambar 4.20.</b> Detail Hasil Klasifikasi KNN Untuk Dataset Pertama .....	69
<b>Gambar 4.21.</b> Detail Hasil Klasifikasi KNN Untuk Dataset Kedua.....	70
<b>Gambar 4.22.</b> <i>Performance Measurements</i> KNN Untuk Dataset Pertama .....	71
<b>Gambar 4.23.</b> <i>Performance Measurements</i> KNN Untuk Dataset Kedua .....	72
<b>Gambar 4.24.</b> <i>Performance Measurements</i> Untuk Dataset Pertama .....	73
<b>Gambar 4.25.</b> <i>Performance Measurements</i> Untuk Dataset Kedua.....	74

## DAFTAR TABEL

<b>Tabel 2.1.</b> Contoh <i>Stemming</i> .....	15
<b>Tabel 2.2.</b> Contoh <i>Lemmatization</i> .....	16
<b>Tabel 3.1.</b> Skenario Percobaan RFC.....	41
<b>Tabel 3.2.</b> Skenario Percobaan SVM .....	42
<b>Tabel 3.3.</b> Skenario Percobaan LRC .....	43
<b>Tabel 3.4.</b> Skenario Percobaan DNN .....	44
<b>Tabel 3.5.</b> Skenario Percobaan KNN .....	45
<b>Tabel 3.6.</b> Tabel <i>Confussion Matrix</i> [49] .....	46

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Internet sudah menjadi bagian dari hidup pada saat ini, tidak mudah untuk dapat lepas dari internet. Dengan perkembangan internet yang semakin hari terus berkembang semakin pesat, internet berubah menjadi sumber informasi yang dapat dengan mudah diakses. Berita menjadi lebih murah, lebih cepat dan lebih mudah diakses. Perkembangan ini selain membawa dampak positif tentu saja juga membawa dampak negatif di lingkungan masyarakat. Perkembangan teknologi informasi seperti media sosial dan berita yang terdapat di web turut serta mendorong penyebaran berita palsu dengan sangat mudah dan sangat cepat [1].

Dampak buruk dari penyebaran berita palsu apabila tidak dicegah sedini mungkin diantaranya adalah dapat mengakibatkan perpecahan antar kelompok, dapat memberikan reputasi buruk akan seseorang/sesuatu kemudian menguntungkan pihak tertentu, berita palsu juga dapat menjadikan masyarakat menjadi panik akan berita yang diberikan oleh pembuat berita palsu tersebut.

Penelitian terkait berita palsu pernah dilakukan oleh F. A. Ozbay & B. Alatas pada tahun 2019 lalu dengan menggunakan metode *Decision Stump*, *Logistic Model Tree* dan J48 dengan mengklasifikasikan antara berita palsu dan berita asli, penelitian tersebut mendapatkan persentase hasil berturut-turut sebesar 56,4% untuk *Decision Stump*, 60,7% untuk *Logistic Model Tree* dan 55,8% untuk J48 [2].



Pada penelitian lain, terkait klasifikasi *Logistic Regression* untuk penilaian kerentanan longsor yang dilakukan oleh Tsangaratos, P., & Ilia, I. (2016), penelitian tersebut menunjukkan persentase hasil yang cukup baik, dengan tingkat persentase sebesar 84,4% [3]. Pada penelitian yang dilakukan oleh Dogru, N., & Subasi, A. (2018), mengenai klasifikasi *Random Forest Classifier* dan *Support Vector Machine* untuk mendeteksi kecelakaan lalu lintas, penelitian tersebut menunjukkan akurasi yang cukup baik, dengan persentase sebesar 91,56% untuk *Random Forest Classifier* dan 88,71% untuk *Support Vector Machine* [4]. Pada penelitian yang dilakukan oleh K. K. Gulhare, S. P. Shukla, & L. K. Sharma untuk mendeteksi penyakit *Alzheimer* menggunakan klasifikasi *Deep Neural Network*, penelitian tersebut mendapatkan akurasi yang cukup baik dengan nilai persentase sebesar 96,6% [5].

Berdasarkan pembahasan beberapa referensi sebelumnya, maka penulis menerapkan metode *Machine Learning* untuk melakukan identifikasi berita palsu yang terdapat di web untuk memperoleh hasil identifikasi yang lebih akurat terhadap berita palsu. Dengan adanya penelitian ini diharapkan dapat membantu mengidentifikasi berita palsu yang terdapat di web sedini mungkin untuk mengurangi dampak negatif penggunaan internet.

## **1.2. Tujuan dan Manfaat**

### **1.2.1. Tujuan**

Tujuan dari penulisan tugas akhir ini yaitu:

1. Membuat sistem identifikasi dalam pembuktian berita palsu yang terdapat di web dengan menggunakan metode *Machine Learning*.
2. Menentukan *Hyperparameter* terbaik untuk masing-masing algoritma *Machine Learning* dan masing-masing dataset.
3. Menentukan algoritma *Machine Learning* terbaik untuk mengidentifikasi berita palsu.

### 1.2.2. Manfaat

Manfaat dari penulisan Tugas Akhir ini, yaitu :

1. Masyarakat mampu membedakan mana yang merupakan berita palsu dan mana yang bukan merupakan berita palsu yang terdapat di web secara otomatis dengan bantuan sistem identifikasi yang telah dibuat.
2. Mencegah terjadinya fitnah yang bisa menyebabkan perpecahan antar kelompok ataupun merugikan pihak-pihak tertentu.
3. Mencegah semaksimal dan sedini mungkin penyebaran berita palsu yang terdapat di web.

## 1.3. Perumusan dan Batasan Masalah

### 1.3.1. Perumusan Masalah

Bagaimana cara membuat sistem dengan menggunakan metode *Machine Learning* untuk mengidentifikasi dan melakukan klasifikasi berita palsu yang terdapat di web untuk mendapatkan hasil yang akurat.

### 1.3.2. Batasan Masalah

Berikut batasan masalah pada Tugas Akhir ini, yaitu :

1. Berita palsu yang digunakan hanya berasal dari web.
2. Kedua dataset, yaitu *Isot Fake News Dataset* yang bersumber dari *University of Victoria* dan *fake\_or\_real\_news* dataset yang didapat melalui website *kagge.com* yang bersumber dari *Biology of Aging Group, Institute of Biochemistry of the Rumanian Academy*. Data yang dikumpulkan telah dibersihkan dan diproses, namun, tanda baca dan kesalahan yang ada dalam berita palsu disimpan dalam teks.

3. Berita yang terdapat di web yang diidentifikasi hanya berbahasa Inggris.
4. Output yang dihasilkan dari penelitian ini hanya berupa nilai akurasi dari metode yang digunakan untuk mengidentifikasi apakah merupakan berita palsu atau bukan berita palsu.
5. Algoritma *Machine Learning* yang digunakan sebagai pembanding adalah *Random Forest Classifier*, *Support Vector Machine*, *Logistic Regression Classifier*, *Deep Neural Network* dan *K-Nearest Neighbor*.

#### **1.4. Metodologi Penelitian**

Metodologi yang digunakan pada Tugas Akhir ini akan melewati beberapa tahapan sebagai berikut:

##### **1.4.1. Metode Studi Pustaka dan Literatur**

Pada metode ini, dilakukan pencarian dan pengumpulan referensi yang berupa literatur yang terdapat pada buku dan internet mengenai identifikasi berita palsu dengan menggunakan beberapa metode **Machine Learning**.

##### **1.4.2. Metode Konsultasi**

Pada metode kedua, dilakukan konsultasi secara langsung dan atau tidak langsung kepada semua pihak narasumber yang memiliki pengetahuan serta wawasan yang baik dalam mengatasi permasalahan yang ditemui pada penulisan Tugas Akhir ini tentang sistem identifikasi berita palsu menggunakan metode *Machine Learning*.

#### **1.4.3. Metode Pembuatan Model**

Pada metode ketiga, membuat sistem identifikasi dengan menggunakan simulasi program.

#### **1.4.4. Metode Pengujian dan Validasi**

Pada metode keempat, dilakukan pengujian terhadap sistem identifikasi dengan menggunakan simulasi program yang telah dibuat, untuk melihat batasan-batasan simulasi tersebut dapat menghasilkan nilai akurasi yang baik atau tidak.

#### **1.4.5. Metode Hasil dan Analisa**

Hasil yang didapat dari pengujian penulisan Tugas Akhir ini akan dianalisa mengenai kelebihan maupun kekurangannya, agar dapat digunakan sebagai referensi yang baik untuk penelitian selanjutnya.

#### **1.4.6. Metode Penarikan Kesimpulan dan Saran**

Berdasarkan hasil dan analisis penelitian yang didapat dalam penulisan Tugas Akhir maka akan didapatkan kesimpulan dan saran untuk penelitian selanjutnya.

### **1.5. Sistematika Penulisan**

Agar mempermudah penyusunan Tugas Akhir ini, maka dibuat sistematika penulisan sebagai berikut :

## **BAB I – PENDAHULUAN**

Bab ini akan membahas tentang Latar Belakang Masalah, Tujuan dan Manfaat, Perumusan dan Batasan Masalah, Metode Penelitian, dan Sistematika Penulisan dari penelitian yang dilakukan.

## **BAB II – TINJAUAN PUSTAKA**

Bab selanjutnya merupakan penjelasan Landasan Teori yang dibutuhkan untuk sebagai pondasi untuk menyelesaikan masalah dalam penelitian yang dilakukan.

## **BAB III – METODOLOGI**

Bab ini akan menjelaskan rincian tentang teknik, metode, dan alur proses yang dilakukan dalam penelitian.

## **BAB IV – HASIL DAN PEMBAHASAN**

Bab ini merupakan hasil yang didapat dari pengujian metode yang digunakan meliputi kelebihan dan kekurangan dari penelitian yang telah dilakukan.

## **BAB V – KESIMPULAN DAN SARAN**

Pada bab terakhir berisi kesimpulan yang didapat berdasarkan hasil penelitian dan juga saran untuk penelitian selanjutnya.

## DAFTAR PUSTAKA

- [1] L. Y. Pan and J. S. Chiou, “How Much Can You Trust Online Information? Cues for Perceived Trustworthiness of Consumer-generated Online Information,” *J. Interact. Mark.*, vol. 25, no. 2, pp. 67–74, May 2011, doi: 10.1016/j.intmar.2011.01.002.
- [2] F. A. Ozbay and B. Alatas, “Fake news detection within online social media using supervised artificial intelligence algorithms,” *Phys. A Stat. Mech. its Appl.*, Feb. 2019, doi: 10.1016/j.physa.2019.123174.
- [3] P. Tsangaratos and I. Ilia, “Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size,” *Catena*, vol. 145, pp. 164–179, 2016, doi: 10.1016/j.catena.2016.06.004.
- [4] N. Dogru and A. Subasi, “Traffic accident detection using random forest classifier,” *2018 15th Learn. Technol. Conf. LT 2018*, pp. 40–45, 2018, doi: 10.1109/LT.2018.8368509.
- [5] K. K. Gulhare, S. P. Shukla, and L. K. Sharma, “Deep Neural Network Classification method to Alzheimer’s Disease Detection,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 7, no. 6, pp. 1–4, 2017, doi: 10.23956/ijarcsse/v7i6/0259.
- [6] E. C. Tandoc, Z. W. Lim, and R. Ling, “Defining ‘Fake News’: A typology of scholarly definitions,” *Digital Journalism*, vol. 6, no. 2. Routledge, pp. 137–153, 07-Feb-2018, doi: 10.1080/21670811.2017.1360143.
- [7] A. Bondielli and F. Marcelloni, “A survey on fake news and rumour detection techniques,” *Inf. Sci. (Ny)*, vol. 497, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05

- [8] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," doi: 10.1257/jep.31.2.211.
- [9] "Ini 12 Ciri Berita Hoax - BeritaSatu.com." [Online]. Available: <https://www.beritasatu.com/digital/547545/digital/547545-ini-12-ciri-berita-hoax>. [Accessed: 23-Mar-2020].
- [10] A. Zubiaga *et al.*, "Detection and Resolution of Rumours in Social Media: A Survey," *ACM Comput. Surv.*, vol. 51, no. 2, p. 32, 2018, doi: 10.1145/3161603.
- [11] X. Hu, J. Tang, Y. Zhang, and H. Liu, "Social Spammer Detection in Microblogging."
- [12] S. Vijayarani and M. R. Janani, "Advanced Computational Intelligence," *An Int. J.*, vol. 3, no. 1, 2016, doi: 10.5121/acii.2016.3104.
- [13] T. Korenius, J. Laurikkala, K. Järvelin, M. Juhola, and K. Jarvelin, "Stemming and Lemmatization in the Clustering of Finnish Text Documents," 2004.
- [14] J. Plisson, N. Lavrac, and D. Mladenic, "A Rule based Approach to Word Lemmatization."
- [15] V. Balakrishnan and E. Lloyd-Yemoh, "Stemming and lemmatization: A comparison of retrieval performances," 2014.
- [16] S. Daniel Rudiaman and A. Anita, "Sistem Temu Balik Informasi, Clustering, K-Means," Aug. 2016.
- [17] T. Tokunaga, T. Tokunaga, and I. Makoto, "Text Categorization Based on Weighted Inverse Document Frequency," *Spec. Interes. GROUPS Inf. Process Soc. JAPAN (SIG-IPJS)*, pp. 33--39, 1994.
- [18] K. Church and W. Gale, "Inverse Document Frequency (IDF): A Measure

of Deviations from Poisson,” 1999, pp. 283–295.

- [19] S. Robertson, “Understanding inverse document frequency: On theoretical arguments for IDF,” *J. Doc.*, vol. 60, no. 5, pp. 503–520, 2004, doi: 10.1108/00220410410560582.
- [20] J. Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries.”
- [21] A. Aizawa, “An information-theoretic perspective of tf-idf measures,” *Inf. Process. Manag.*, vol. 39, no. 1, pp. 45–65, Jan. 2003, doi: 10.1016/S0306-4573(02)00021-3.
- [22] “Apa itu Machine Learning dan Cara Kerjanya - Advernesia.” [Online]. Available: <https://www.advernesia.com/blog/data-science/machine-learning-adalah/>. [Accessed: 27-Feb-2020].
- [23] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, Nov. 2003, doi: 10.1021/ci034160g.
- [24] A. Liaw and M. Wiener, “Classification and Regression by RandomForest,” 2002.
- [25] E. A. Laksana and F. Sulianta, “Analisis dan studi komparatif algoritma klasifikasi genre musik,” pp. 67–72, 2017.
- [26] “Random Forest Simple Explanation - Will Koehrsen - Medium.” [Online]. Available: <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>. [Accessed: 01-Mar-2020].
- [27] Y. S. Nugroho, “Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest,” *J. Tek. Elektro*, vol. 9, no. 1, pp. 24–29, 2017.



- [28] “Optimizing Hyperparameters in Random Forest Classification.” [Online]. Available: <https://towardsdatascience.com/optimizing-hyperparameters-in-random-forest-classification-ec7741f9d3f6>. [Accessed: 14-Mar-2020].
- [29] “Penjelasan Sederhana tentang Apa Itu SVM? - Samsudiney - Medium.” [Online]. Available: <https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02>. [Accessed: 03-Feb-2020].
- [30] A. B. Prasetyo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, “Hoax detection system on Indonesian news sites based on text classification using SVM and SGD,” *Proc. - 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2017*, vol. 2018-Janua, no. September 2018, pp. 45–49, 2017, doi: 10.1109/ICITACEE.2017.8257673.
- [31] “Support Vector Machines: Theory and Applications - Google Buku.” [Online]. Available: [https://books.google.co.id/books?hl=id&lr=&id=uTzMPJjVjsMC&oi=fnd&pg=PA1&dq=Support+Vector+machine&ots=GEHDds\\_Li7&sig=3fXLnW5d4c4iTr1bUrEh4IGaiFs&redir\\_esc=y#v=onepage&q=Support+Vector+machine&f=false](https://books.google.co.id/books?hl=id&lr=&id=uTzMPJjVjsMC&oi=fnd&pg=PA1&dq=Support+Vector+machine&ots=GEHDds_Li7&sig=3fXLnW5d4c4iTr1bUrEh4IGaiFs&redir_esc=y#v=onepage&q=Support+Vector+machine&f=false). [Accessed: 03-Feb-2020].
- [32] A. S. Nugroho, A. B. Witarto, and D. Handoko, “Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1,” 2003.
- [33] J. Mase, M. T. Furqon, and B. Rahayudi, “Penerapan Algoritme Support Vector Machine ( SVM ) Pada Pengklasifikasian Penyakit Kucing,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3648–3654, 2018.
- [34] “Support Vector Machine vs Logistic Regression - Georgios Drakos - Medium.” [Online]. Available: <https://medium.com/@george.drakos62/support-vector-machine-vs-logistic-regression-94cc2975433f>. [Accessed: 01-Mar-2020].

- [35] “sklearn.svm.SVC — scikit-learn 0.22.2 documentation.” [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>. [Accessed: 15-Mar-2020].
- [36] “Machine Learning: Logistic Regression - Artificial intelligence and data science.” [Online]. Available: <https://www.megabagus.id/machine-learning-logistic-regression/>. [Accessed: 01-Mar-2020].
- [37] “(Tutorial) Understanding Logistic REGRESSION in PYTHON - DataCamp.” [Online]. Available: <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>. [Accessed: 01-Mar-2020].
- [38] “sklearn.linear\_model.LogisticRegression — scikit-learn 0.22.2 documentation.” [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). [Accessed: 15-Mar-2020].
- [39] “Deep Learning made easy with Deep Cognition - Becoming Human: Artificial Intelligence Magazine.” [Online]. Available: <https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351>. [Accessed: 01-Mar-2020].
- [40] S. Nurmaini, P. R. Umi, R. M. Naufal, and A. Gani, “Cardiac arrhythmias classification using Deep Neural Networks and principle component analysis algorithm,” *Int. J. Adv. Soft Comput. its Appl.*, vol. 10, no. 2, pp. 14–32, 2018.
- [41] “Activation Functions in Neural Networks - Towards Data Science.” [Online]. Available: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. [Accessed: 16-Mar-2020].
- [42] “What are Hyperparameters ? and How to tune the Hyperparameters in a

- Deep Neural Network?” [Online]. Available: <https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a>. [Accessed: 16-Mar-2020].
- [43] “Algoritma K-Nearest Neighbor (K-NN) | INFORMATIKALOGI.” [Online]. Available: <https://informatikalogi.com/algoritma-k-nn-k-nearest-neighbor/>. [Accessed: 16-May-2020].
- [44] “Klasifikasi K-NN Menggunakan Data Tooth Growth di R.” [Online]. Available: <https://medium.com/@16611110/klasifikasi-k-nn-menggunakan-data-tooth-growth-di-r-ec18148f1cd9>. [Accessed: 10-Jun-2020].
- [45] “Pengertian dan Cara Kerja Algoritma K-Nearest Neighbors (KNN).” [Online]. Available: <https://www.advernesia.com/blog/data-science/pengertian-dan-cara-kerja-algoritma-k-nearest-neighbours-knn/>. [Accessed: 16-May-2020].
- [46] “Notes on Sensitivity, Specificity, Precision, Recall and F1 score.” [Online]. Available: <https://medium.com/analytics-vidhya/notes-on-sensitivity-specificity-precision-recall-and-f1-score-e34204d0bb9b>. [Accessed: 16-Mar-2020].
- [47] “ISOT Fake News Dataset.”
- [48] “Understanding Confusion Matrix - Towards Data Science.” [Online]. Available: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>. [Accessed: 17-Mar-2020].
- [49] “Measuring Performance: The Confusion Matrix – Glass Box.” [Online]. Available: <https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/>. [Accessed: 17-Mar-2020].

