

**NAMED ENTITY RECOGNITION MENGGUNAKAN  
PEMBOBOTAN TERM FREQUENCY – INVERSE  
DOCUMENT FREQUENCY DAN SUPPORT VECTOR  
MACHINES**

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika



Oleh:

Septri Putra Wiarka  
NIM : 09021381621129

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2021**

LEMBAR PENGESAHAN TUGAS AKHIR

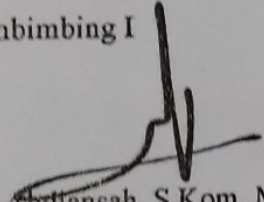
NAMED ENTITY RECOGNITION MENGGUNAKAN  
PEMBOBOTAN TERM FREQUENCY – INVERSE  
DOCUMENT FREQUENCY DAN SUPPORT VECTOR  
MACHINES

Oleh:

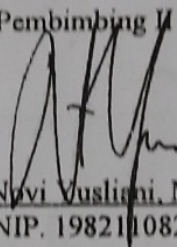
Septri Putra Wiarka  
NIM : 09021381621129

Palembang, Juni 2021

Pembimbing I

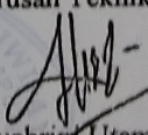
  
Dr. Abdiansah, S.Kom, M.Cs  
NIP. 198410012009121005

Pembimbing II

  
Novi Musliani, M.T.  
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika

  
Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

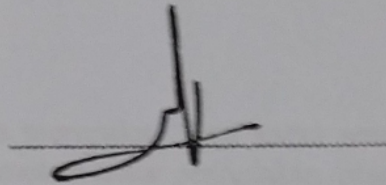
## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jum'at tanggal 11 Juni 2021 telah dilaksanakan ujian sidang tugas akhir oleh jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Septri Putra Wiarka  
NIM : 0902138162129  
Judul : *Named Entity Recognition Menggunakan Pembobotan Term Frequency – Inverse Document Frequency dan Support Vector Machines*

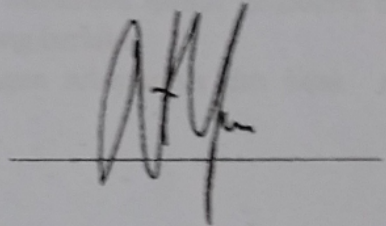
### 1. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.  
NIP. 1984100120091210005



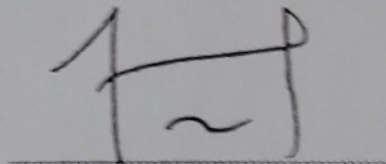
### 2. Pembimbing II

Novi Yusliani, M.T.  
NIP. 198211082012122001



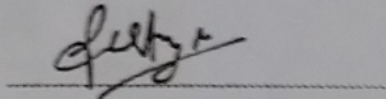
### 3. Penguji I

Muhammad Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002

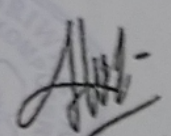


### 4. Penguji II

Desty Rodiah, M.T.  
NIP. 198912212020122011



Mengetahui  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Septri Putra Wiarka  
NIM : 09021381621129  
Program Studi : Teknik Informatika  
Judul Skripsi : *Named Entity Recognition Menggunakan Pembobotan Term Frequency – Inverse Document Frequency dan Support Vector Machines*  
Hasil Pengecekan Software : 11%  
*iThenticate/Turnitin*

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan /plagiat dalam Laporan Proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Juni 2021



Septri Putra Wiarka  
NIM. 09021381621129

## MOTTO DAN PERSEMBAHAN

### MOTTO :

**"DO WITH YOUR OWN RISK,  
BUT DON'T FORGET WHO YOU ARE"**

**Kupersembahkan karya tulis ini untuk :**

- **Orang tuaku dan saudaraku tercinta**
- **Keluarga Besarku**
- **Teman seperjuanganku**
- **Jurusan Teknik Informatika**
- **Fakultas Ilmu Komputer**
- **Universitas Sriwijaya**

# NAMED ENTITY RECOGNITION USING WEIGHTING TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY AND SUPPORT VECTOR MACHINES

By:

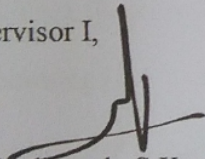
Septri Putra Wiarka  
09021381621129

## ABSTRACT

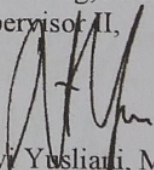
The sources of information are currently diverse, making it easier for people to get information. However, the main information in the media is not structured so that it must be carefully read so as not to get wrong information. For that we need a tool to extract information. Named Entity Recognition (NER) or the recognition of named entities is a technique in extracting information by recognizing entities that have been determined in a sentence. In this study, NER is used in Indonesian language news texts using weighting term frequency - inverse document frequency (TF-IDF) and support vector machines (SVM) for named entities such as names of people (PER), names of organizations (ORG), names of locations (LOC), adverb of time (TIME) and other entities (OTH). The recognition of named entities is done by using the TF-IDF weight feature and the weight of the Part of Speech Tagging (POS-Tag) for each word. The test was carried out on the Indonesian language news text with a total of 3538 words and the results of the performance scores for accuracy on each entity named OTH, ORG, TIME, PER and LOC each scored 60.55%, 59.85%, 53.12%, 33.01% and 4.35%.

Keywords: Named Entity Recognition (NER), Information Extraction, Term Frequency – Inverse Document Frequency (TF-IDF), Part of Speech Tagging (POS-Tag), Support Vector Machines (SVM)

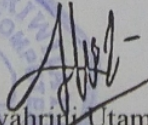
Supervisor I,

  
Dr. Abdiansah, S.Kom, M.Cs  
NIP. 198410012009121005

Palembang, Juni 2021  
Supervisor II,

  
Novi Yusliani, M.T.  
NIP. 198211082012122001

Approved,  
Head of Informatics Engineering Department

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

NAMED ENTITY RECOGNITION MENGGUNAKAN PEMBOBOTAN TERM  
FREQUENCY – INVERSE DOCUMENT FREQUENCY DAN SUPPORT  
VECTOR MACHINES

Oleh:

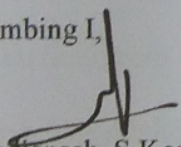
Septri Putra Wiarka  
09021381621129

ABSTRAK

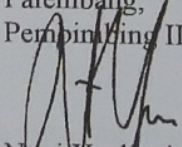
Sumber informasi saat ini sudah beragam sehingga memudahkan masyarakat dalam mendapatkan informasi. Namun informasi utama yang ada pada media tidak terstruktur sehingga harus cermat dalam membaca agar tidak mendapatkan informasi yang salah. Untuk itu diperlukan sebuah alat untuk mengekstrak informasi. *Named Entity Recognition* (NER) atau pengenalan entitas bernama merupakan salah satu teknik dalam ekstraksi informasi dengan cara mengenali entitas yang telah ditentukan pada sebuah kalimat. Pada penelitian ini NER digunakan pada teks berita berbahasa indonesia menggunakan pembobotan *term frequency – inverse document frequency* (TF-IDF) dan *support vector machines* (SVM) untuk entitas bernama seperti nama orang (PER), nama organisasi (ORG), nama lokasi (LOC), keterangan waktu (TIME) serta entitas lainnya (OTH). Pengenalan entitas bernama dilakukan dengan menggunakan fitur bobot TF-IDF dan bobot *Part of Speech Tagging* (POS-Tag) pada setiap kata. Pengujian dilakukan pada teks berita berbahasa indonesia dengan jumlah 3538 kata dan didapatkan hasil nilai kinerja untuk *Accuracy* pada tiap entitas bernama OTH, ORG, TIME, PER dan LOC masing-masing mendapat nilai 60.55%, 59.85%, 53.12%, 33.01% dan 4.35%.

Kata Kunci : *Named Entity Recognition* (NER), Ekstraksi Informasi, *Term Frequency – Inverse Document Frequency* (TF-IDF), *Part of Speech Tagging* (POS-Tag), *Support Vector Machines* (SVM)

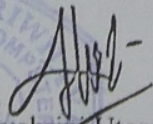
Pembimbing I,

  
Dr. Abdiansah, S.Kom, M.Cs  
NIP. 198410012009121005

Palembang, Juni 2021  
Pembimbing II,

  
Novi Yustiani, M.T.  
NIP. 19821082012122001

Mengetahui,  
Ketua Jurusan Teknik Informatika,

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## KATA PENGANTAR

Puji syukur kepada Allah atas berkah dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir dengan baik. Tugas Akhir ini disusun untuk memenuhi salah satu persyaratan guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada :

1. Orang Tuaku, M. Yusuf dan Rohmah, Saudaraku Yusto Wisnu Kawirian dan Arwin Cholik, S.T. dan seluruh Keluarga Besarku yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika dan Ibu Mastura Diana Marieska, S.T., M.T. selaku Sekretaris Jurusan Teknik Informatika.
3. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Pembimbing I dan Ibu Novi Yusliani, M.T. selaku pembimbing II yang telah membimbing, memotivasi serta mengarahkan Penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.



4. Bapak Osvari Arsalan, M.T. selaku Pembimbing Akademik yang telah membimbing, mengarahkan dan memberika motivasi pada Penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak M. Fachrurrozi, M.T. selaku dosen Penguji I dan Ibu Desty Rodiah, M.T. selaku dosen Penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh Dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Bapak Ricy dan seluruh staff akademik dan tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Zain, Syahid, Ade, Rifal, Bagus, Tara, Bayu serta teman-teman seperjuangan jurusan Teknik Informatika dan Organisasi yang telah berbagi keluh kesah, semangat, motivasi, duka cita dan canda tawa selama masa perkuliahan.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan dikarenakan keterbatasan pengetahuan dan pengalaman Penulis, karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Palembang, Juni 2021

Septri Putra Wiarka

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR .....	iii
HALAMAN PERNYATAAN .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT.....	vi
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xv
BAB I PENDAHULUAN .....	I-1
1.1 Pendahuluan .....	I-1
1.2 Latar Belakang Masalah .....	I-1
1.3 Rumusan Masalah .....	I-3
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR .....	II-1
2.1 Pendahuluan .....	II-1
2.2 Landasan Teori .....	II-1
2.2.1 <i>Named Entity Recognition</i> .....	II-1
2.2.2 Prapengolahan .....	II-2

2.2.3	Pembobotan TF-IDF .....	II-5
2.2.4	<i>Support Vector Machines</i> .....	II-6
2.2.5	<i>Confusion Matrix</i> .....	II-13
2.2.6	<i>Rational Unified Process</i> .....	II-15
2.3	Penelitian Lain Yang Relevan .....	II-16
2.4	Kesimpulan.....	II-17
BAB III	METODE PENELITIAN.....	III-1
3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data .....	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Metode Pengumpulan Data.....	III-1
3.3	Tahapan Penelitian .....	III-2
3.3.1	Menetapkan Kerangka Kerja Penelitian .....	III-2
3.3.2	Menetapkan Kriteria Pengujian .....	III-5
3.3.3	Menetapkan Format Data Pengujian.....	III-6
3.3.4	Menentukan Alat yang Digunakan dalam Penelitian .....	III-6
3.3.5	Melakukan Pengujian Penelitian .....	III-7
3.3.6	Melakukan Analisis Pengujian dan Membuat Kesimpulan.....	III-7
3.4	Metode Pengembangan Perangkat Lunak .....	III-8
3.4.1	Fase Insepsi.....	III-8
3.4.2	Fase Elaborasi .....	III-9
3.4.3	Fase Kontruksi .....	III-9
3.4.4	Fase Transisi .....	III-10
3.5	Kesimpulan.....	III-10
BAB IV	PENGEMBANGAN PERANGKAT LUNAK .....	IV-1
4.1	Pendahuluan .....	IV-1
4.2	<i>Ratioanl Unified Process (RUP)</i> .....	IV-1
4.2.1	Fase Insepsi.....	IV-1
4.2.2	Fase Elaborasi .....	IV-22

4.2.3 Fase Konstruksi.....	IV-34
4.2.4 Fase Transisi .....	IV-42
4.3 Kesimpulan.....	IV-52
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan .....	V-1
5.2 Data Hasil Percobaan/Penelitian .....	V-1
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Hasil Pengujian .....	V-4
5.3 Analisis Hasil Penelitian .....	V-5
5.4 Kesimpulan.....	V-9
BAB VI KESIMPULAN DAN SARAN .....	VI-1
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2
DAFTAR PUSTAKA .....	VII-1

## DAFTAR TABEL

	Halaman
II-1	Contoh Segmentasi pada Kalimat ..... II-2
II-2	Contoh Normalisasi pada Kalimat ..... II-3
II-3	Contoh Tokenisasi pada kalimat ..... II-3
II-4	Tabel <i>Tagset</i> Bahasa Indonesia yang akan digunakan ..... II-4
II-5	<i>Confusion Matrix</i> Kelas Positif dan Kelas Negatif ..... II-13
II-6	<i>Confusion Matrix</i> n kelas ..... II-14
III-1	Tabel Bobot <i>Part of Speech</i> ..... III-4
III-2	Rancangan Tabel Hasil Uji NER ..... III-6
III-3	Rancangan Tabel Hasil Analisis NER ..... III-8
IV-1	Tabel Kebutuhan Fungsional ..... IV-3
IV-2	Tabel Kebutuhan Non-Fungsional ..... IV-4
IV-3	Contoh Data Latih ..... IV-5
IV-4	Contoh Data Uji ..... IV-6
IV-5	Segmentasi ..... IV-6
IV-6	Normalisasi..... IV-7
IV-7	Tokenisasi..... IV-8
IV-8	POS-Tag..... IV-8
IV-9	Pembobotan TF-IDF ..... IV-10
IV-10	Contoh Data..... IV-11
IV-11	Klasifikasi <i>Multiclass SVM One Versus One</i> ..... IV-14
IV-12	Definisi Aktor..... IV-17
IV-13	Definisi <i>Use Case</i> ..... IV-17
IV-14	Skenario Melakukan Prapengolahan..... IV-18
IV-15	Skenario Melakukan Pelatihan..... IV-19
IV-16	Skenario Melakukan Klasifikasi ..... IV-20
IV-17	Skenario Melakukan Hitung Kinerja..... IV-22
IV-18	Implementasi Kelas ..... IV-38
IV-19	Rencana Pengujian <i>Use Case</i> Melakukan Prapengolahan ..... IV-44
IV-20	Rencana Pengujian <i>Use Case</i> Melakukan Tahap Pelatihan..... IV-44

IV-21	Rencana Pengujian <i>Use Case</i> Melakukan Tahap Klasifikasi.....	IV-45
IV-22	Rencana Pengujian <i>Use Case</i> Melakukan Hitung Kinerja.....	IV-46
IV-23	Pengujian <i>Use Case</i> Melakukan Prapengolahan.....	IV-48
IV-24	Pengujian <i>Use Case</i> Melakukan Tahap Pelatihan.....	IV-48
IV-25	Pengujian <i>Use Case</i> Melakukan Tahap Klasifikasi .....	IV-50
IV-26	Pengujian <i>Use Case</i> Melakukan Hitung Kinerja.....	IV-51
V-1	Tabel <i>Confusion Matrix</i> Hasil Pengujian .....	V-4
V-2	Nilai Kinerja.....	V-5

## DAFTAR GAMBAR

	Halaman
II-1	Contoh dua kelas yang dipisahkan oleh <i>hyperplane</i> ..... II-7
II-2	Pemetaan data ke ruang vektor berdimensi tinggi..... II-10
II-3	Contoh <i>Multiclass</i> SVM menggunakan <i>one versus all</i> ..... II-12
II-4	Contoh <i>Multiclass</i> SVM menggunakan <i>one versus one</i> ..... II-12
II-5	Arsitektur <i>Rational Unified Process</i> ..... II-15
III-1	Diagram Tahapan Penelitian ..... III-2
III-2	Diagram Tahapan Proses Perangkat Lunak..... III-3
IV-1	Diagram <i>Use Case</i> ..... IV-16
IV-2	Rancangan Data Latih ..... IV-23
IV-3	Rancangan Data Uji ..... IV-23
IV-4	Rancangan Antarmuka Tahap Pelatihan ..... IV-24
IV-5	Rancangan Antarmuka Tahap Klasifikasi ..... IV-25
IV-6	Diagram Aktivitas Prapengolahan Data ..... IV-27
IV-7	Diagram Aktivitas Tahap Pelatihan ..... IV-28
IV-8	Diagram Aktivitas Tahap Klasifikasi ..... IV-29
IV-9	Diagram Aktivitas Hitung Kinerja ..... IV-30
IV-10	Diagram <i>Sequence</i> Melakukan Prapengolahan ..... IV-31
IV-11	Diagram <i>Sequence</i> Melakukan Tahap Pelatihan ..... IV-32
IV-12	Diagram <i>Sequence</i> Melakukan Tahap Klasifikasi ..... IV-33
IV-13	Diagram <i>Sequence</i> Melakukan Hitung Kinerja..... IV-34
IV-14	Kelas Diagram..... IV-36
IV-15	Kelas Analisis Melakukan Tahap Prapengolahan ..... IV-37
IV-16	Kelas Analisis Melakukan Tahap Pelatihan ..... IV-37
IV-17	Kelas Analisis Melakukan Tahap Klasifikasi ..... IV-37
IV-18	Kelas Analisis Melakukan Tahap Hitung Kinerja..... IV-38
IV-19	Tampilan Halaman Antarmuka Tahap Pelatihan ..... IV-41
IV-20	Tampilan Halaman Antarmuka Tahap Klasifikasi..... IV-42
V-1	Hasil Tahap Pelatihan..... V-2
V-2	Hasil Tahap Klasifikasi ..... V-3
V-3	Perbandingan Nilai <i>F-Measure</i> Pada Entitas Bernama..... V-6

V-4	Perbandingan Nilai <i>Precision</i> Pada Entitas Bernama .....	V-6
V-5	Perbandingan Nilai <i>Recall</i> Pada Entitas Bernama .....	V-7
V-6	Perbandingan Nilai <i>Accuracy</i> Pada Entitas Bernama .....	V-8



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan ini memaparkan secara umum terkait latar belakang penelitian *Named Entity Recognition* menggunakan *Support Vector Machines* dan *TF-IDF*, rumusan masalah, tujuan penelitian, manfaat penelitian, serta batasan masalah pada penelitian ini.

### **1.2 Latar Belakang Masalah**

Berdasarkan survei Katadata Insight Center (KIC), masyarakat Indonesia cenderung memanfaatkan media sosial sebagai sumber informasi. Selain media sosial, televisi dan berita dalam jaringan (daring) menjadi opsi selanjutnya untuk mendapatkan informasi<sup>1</sup>. Menurut Ilyas dan Khodra (2015), dengan banyaknya informasi menimbulkan masalah kesulitan pencarian informasi penting karena tertutup oleh informasi yang tidak dibutuhkan (*information overload*). Untuk mendapatkan informasi utama, diperlukan sistem yang mampu mengekstraksi informasi. Salah satu kegiatan dari Ekstraksi Informasi adalah *Named Entity Recognition* (NER).

NER bertujuan untuk memudahkan mencari informasi dengan cara pemberian nama entitas pada setiap kata dalam sebuah teks (Setiyoaji dkk., 2017). NER juga sering digunakan sebagai prapengolahan untuk menyelesaikan

---

<sup>1</sup> <https://katadata.co.id/ekarina/brand/5fcfc332efab0/transformasi-bisnis-media-di-era-digital-terus-bergulir>

permasalahan yang lebih kompleks pada bidang pemrosesan bahasa lainnya seperti *question-answering system*, *machine translation*, *automatic text summarization*, dan lain sebagainya (Goyal dkk., 2018).

Penelitian NER pada Bahasa Indonesia sudah pernah dilakukan dengan beberapa pendekatan seperti menggunakan pendekatan berbasis aturan yang dilakukan Budi dkk., (2005) dan pembelajaran mesin oleh Suryana dan Ipnuwati (2016). Sistem NER yang dilakukan dengan pendekatan berbasis aturan mendapatkan hasil yang baik namun terbatas hanya pada domain tertentu. Sedangkan pada sistem modern sekarang banyak dilakukan dengan pendekatan berbasis pembelajaran mesin untuk mengatasi kelemahan sistem sebelumnya (Goyal dkk., 2018).

Metode *Support Vector Machines* (SVM) merupakan salah satu algoritma berbasis pembelajaran mesin yang sering digunakan pada berbagai permasalahan seperti klasifikasi teks, analisis sentimen dan sebagainya. SVM dikenal dapat mengatasi permasalahan *overfitting* ketika fitur yang digunakan sudah banyak (Ekbal & Bandyopadhyay, 2009).

Permasalahan metode SVM pada NER dalam bahasa Indonesia adalah kesalahan dalam memberikan label entitas dikarenakan fitur kata dan dataset yang digunakan saat pelatihan (Saniyah, 2019). Hal ini dapat diatasi dengan pembobotan kata seperti *Term Frequency – Inverse Document Frequency* (TF-IDF) dan dengan POS Tag (Goyal dkk., 2018).

TF-IDF merupakan teknik yang digunakan untuk memberikan bobot pada sebuah kata apakah termasuk kata yang penting dalam sebuah dokumen atau teks

(Putong, 2020). Sedangkan POS Tag digunakan untuk mengetahui apakah kata merupakan kelas kata seperti kata benda, kata kerja, kata sifat atau kata keterangan (Dinakaramani dkk., 2014).

Pada penelitian NER untuk bahasa Inggris oleh Sravani (2018) menggunakan *word embedding* sebagai fitur utama, yaitu Word2vec, Glove dan TF-IDF. Metode klasifikasi yang digunakan pada penelitian tersebut antara lain *Support Vector Machine (SVM)*, *Random Forest*, *Logistic Regression*, *Gaussian Naive Bayes*, *K-Nearest Neighbor (KNN)* dan *Adaboost*. Penelitian tersebut menunjukkan TF-IDF adalah model *word embedding* dengan akurasi terbaik pada kasus tersebut yaitu 84,33% yang diperoleh dengan menggunakan metode TF-IDF dan algoritma SVM.

Berdasarkan uraian di atas pada penelitian ini akan menerapkan metode pembobotan TF-IDF dan algoritma SVM untuk *Named Entity Recognition* pada teks bahasa Indonesia.

### **1.3 Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah bagaimana kinerja metode pembobotan TF-IDF dan algoritma *Support Vector Machines* untuk *named entity recognition* pada bahasa Indonesia. Beberapa pertanyaan yang akan dijawab pada usulan penelitian ini, yaitu:

1. Bagaimana menerapkan metode pembobotan TF-IDF dan algoritma *Support Vector Machines* untuk *Named Entity Recognition*?

2. Bagaimana mengembangkan perangkat lunak *Named Entity Recognition* untuk melakukan ekstraksi informasi pada teks Bahasa Indonesia?
3. Bagaimana tingkat akurasi metode pembobotan TF-IDF dan algoritma *Support Vector Machines* untuk *Named Entity Recognition* pada teks bahasa Indonesia?

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah penelitian di atas maka tujuan penelitian ini, yaitu :

1. Menerapkan Metode pembobotan *Term Frequency – Inverse Document Frequency* dan *Support Vector Machines* untuk *Named Entity Recognition*.
2. Membuat perangkat lunak *Named Entity Recognition* pada Bahasa Indonesia menggunakan metode pembobotan *Term Frequency -Inverse Document Frequency* dan *Support Vector Machines*.
3. Mengetahui tingkat akurasi *Named Entity Recognition* pada Bahasa Indonesia menggunakan metode pembobotan *Term Frequency –Inverse Document Frequency* dan *Support Vector Machines*.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang di harapkan dari penelitian yang diusulkan ini adalah sebagai berikut :

1. Dapat digunakan sebagai referensi untuk penelitian lain di bidang Pemrosesan Bahasa Alami (PBA) khususnya *Named Entity Recognition*.

2. Dapat digunakan sebagai prapengolahan pada penelitian PBA lainnya.
3. Dapat digunakan untuk mengetahui entitas kata pada teks.

### **1.6 Batasan Masalah**

Untuk membatasi penelitian agar tidak menyimpang dari pembahasan, maka diberikan sejumlah batasan masalah penelitian, meliputi :

1. Data yang digunakan berbentuk teks berita berbahasa Indonesia.
2. Tipe data yang digunakan sebagai masukan sistem adalah tipe data txt.
3. Entitas yang digunakan adalah nama orang (PER), nama organisasi (ORG), nama lokasi (LOC), keterangan waktu (TIME) dan entitas lainnya (OTH).

### **1.7 Sistematika Penulisan**

Sistematika penulisan penelitian ini berdasarkan standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain:

## **BAB I. PENDAHULUAN**

Dalam bab ini dijelaskan tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah serta sistematika penulisan.

## **BAB II. KAJIAN LITERATUR**

Dalam bab membahas landasan teori yang dipakai pada penelitian ini, meliputi penelitian terkait, definisi *Named Entity Recognition*, Prapengolahan, *Part of Speech Tagging* (POS-Tag), TF-IDF, Algoritma *Support Vector Machines*, *Confusion Matrix* dan *Rational Unified Process*.

### **BAB III. METODOLOGI PENELITIAN**

Dalam bab ini dijelaskan tahap-tahap yang akan dilakukan pada penelitian ini. Setiap tahapan penelitian digambarkan secara detail berdasarkan suatu kerangka kerja.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan dibahas mengenai tahapan-tahapan dalam mengembangkan perangkat lunak pada penelitian ini.

### **BAB V. HASIL DAN ANALISIS PENELITIAN**

Dalam bab ini membahas mengenai hasil penelitian yang telah dilakukan berdasarkan usulan penelitian. Analisis hasil penelitian ini dijadikan sebagai acuan dalam memberikan kesimpulan pada penelitian ini.

### **BAB VI. KESIMPULAN DAN SARAN**

Bab ini mengandung kesimpulan dan saran hasil penelitian ini. Kesimpulan dan saran yang diberikan mengacu pada bab-bab sebelumnya.

#### **1.8 Kesimpulan**

Dalam penelitian ini memfokuskan untuk menerapkan metode pembobotan kata terhadap *Named Entity Recognition* untuk mendapatkan informasi pada teks berita berbahasa Indonesia menggunakan algoritma *Support Vector Machines*. Metode pembobotan yang digunakan yaitu *Term Frequency - Inverse Document Frequency* (TF-IDF).

## DAFTAR PUSTAKA

- Budi, I., Bressan, S., Wahyudi, G., Hasibuan, Z. A., & Nazief, B. A. A. (2005). Named Entity Recognition for the Indonesian language: Combining contextual, morphological and part-of-speech features into a knowledge engineering approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3735 LNAI, 57–69. [https://doi.org/10.1007/11563983\\_7](https://doi.org/10.1007/11563983_7)
- Dinakaramani, A., Rashel, F., Luthfi, A., & Manurung, R. (2014). Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus. *Proceedings of the International Conference on Asian Language Processing 2014, IALP 2014*, 66–69. <https://doi.org/10.1109/IALP.2014.6973519>
- Ekbal, A., & Bandyopadhyay, S. (2009). Named entity recognition using support vector machine: A Language independent approach. *World Academy of Science, Engineering and Technology*, 39, 548–563.
- Emcha, A. C., Widyawan, & Adji, T. B. (2019). Quotation extraction from Indonesian online news. *2019 International Conference on Information and Communications Technology, ICOIACT 2019*, 408–412. <https://doi.org/10.1109/ICOIACT46704.2019.8938558>
- Fachri, M. (2014). Named Entity Recognition For Indonesian Text Using Hidden Markov Model. Universitas Gadjah Mada.
- Goyal, A., Gupta, V., & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques: A systematic review. *Computer Science Review*, 29, 21–43. <https://doi.org/10.1016/j.cosrev.2018.06.001>
- Kamayani, M. (2019). Perkembangan Part-of-Speech Tagger Bahasa Indonesia. *Jurnal Linguistik Komputasional (JLK)*, 2(2), 34. <https://doi.org/10.26418/jlk.v2i2.20>
- Krutchén, P. (2000). *The Rational Unified Process—An Introduction*.
- Kuspriyanto, Santoso, O. S., Widyantoro, D. H., Sastramihardja, H. S., Muludi, K., & Maimunah, S. (2010). Performance evaluation of SVM-based information extraction using  $\tau$  margin values. *International Journal on Electrical Engineering and Informatics*, 2(4), 256–265. <https://doi.org/10.15676/ijeei.2010.2.4.1>

- Muljono, Muljono & Afini, Umriya & Supriyanto, Catur & Nugroho, Raden. (2017). *The Development of Indonesian POS Tagging System for Computer-aided Independent Language Learning*. *International Journal of Emerging Technologies in Learning (IJET)*. 12. 138. 10.3991/ijet.v12i11.7383.
- Munawarah, R., Soesanto, O., & Faisal, M. R. (2016). Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, Volume 04, (February), 103–113. <https://doi.org/10.20527/klik.v3i1.39>
- Nugraha, A. S., & Purnamasari, K. K. (2019). Penerapan Metode Support Vector Machine Pada Part Of Speech Tag Bahasa Indonesia. Universitas Komputer Indonesia.
- Pisceldo, F., Adriani, M., & Manurung, R. (2009). Probabilistic Part of Speech Tagging for Bahasa Indonesia. *Proceedings of the 3rd International MALINDO Workshop, Colocated Event ACL-IJCNLP, May*.
- Purnamasari, N. M. G. D., Fauzi, M. A., Indriarti, & Dewi, L. S. (2018). Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine ( SVM ) dan Information Gain ( IG ) sebagai Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5326–5332. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3276/1280>
- Putong, M. W. (2020). *Classification Model of Contact Center Customers Emails Using Machine Learning Study of*. 5(1), 174–182.
- Saniyah (2019). *Named Entity Recognition pada Teks Berita Menggunakan Support Vector Machines*. Universitas Sriwijaya.
- Setiyoadji, A., Muflikhah, L., & Fauzi, M. A. (2017). Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1858–1864. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/673>
- Sravani, L., Reddy, A. S., & Thara, S. (2018). A Comparison Study of Word Embedding for Detecting Named Entities of Code-Mixed Data in Indian Language. *2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018*, 2375–2381. <https://doi.org/10.1109/ICACCI.2018.8554918>
- Suryana, A., & Ipnuwati, S. (2016). *Vector Machine Learning Method for Text Mining Indonesian Social Media Named Entity Recognition*. In *The 5th*



*Internatioanl Conference on Information Technology and Engineering Application (ICIBA2016). Bina Darma University.*

Syaifudin, Y., & Nurwidyantoro, A. (2016). Quotations identification from Indonesian online news using rule-based method. 187-194. 10.1109/ISITIA.2016.7828656.

Wicaksono, A. F., & Purwarianti, A. (2010). *HMM Based Part-of-Speech Tagger for Bahasa Indonesia HMM Based Part-of-Speech Tagger for Bahasa Indonesia*. In *Fourth International MALINDO Workshop, Jakarta*.