

**EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN
TRADISIONAL INDONESIA DENGAN NAMED ENTITY
RECOGNITION**

Skripsi



oleh:

**ZANETAXINA PUDIHAND
71180279**

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2023

**EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN
TRADISIONAL INDONESIA DENGAN NAMED ENTITY
RECOGNITION**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh
ZANETAXINA PUDIANG

71180279

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2023

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
SKRIPSI/TESIS/DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Zanetaxina Pudihsang
NIM : 71180279
Program studi : Informatika
Fakultas : Teknologi Informasi
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

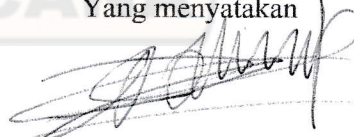
**“EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL
INDONESIA DENGAN NAMED ENTITY RECOGNITION”**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 10 Agustus 2023

Yang menyatakan



(Zanetaxina Pudihsang)

NIM.71180279

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL INDONESIA DENGAN NAMED ENTITY RECOGNITION

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 26 Juli 2023



ZANETAXINA PUDIHING

71180279

DUTA WACANA

HALAMAN PERSETUJUAN

Judul Skripsi : EKSTRAKSI INFORMASI MAKANAN DAN
MINUMAN TRADISIONAL INDONESIA
DENGAN NAMED ENTITY RECOGNITION

Nama Mahasiswa : ZANETAXINA PUDIHANG

N I M : 71180279

Matakuliah : Skripsi (Tugas Akhir)

Kode : TI0366

Semester : Genap

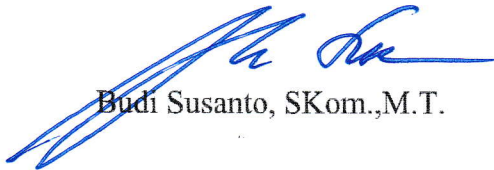
Tahun Akademik : 2022/2023

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 17 Juli 2023

Dosen Pembimbing I

Dosen Pembimbing II


Lucia Dwi Krisnawati, Dr. Phil.


Budi Susanto, SKom.,M.T.

HALAMAN PENGESAHAN

EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL INDONESIA DENGAN NAMED ENTITY RECOGNITION

Oleh: ZANETAXINA PUDIHING / 71180279

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 23 Juni 2023

Yogyakarta, 17 Juli 2023
Mengesahkan,


Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Budi Susanto, SKom., M.T.
3. Matahari Bhakti Nendya, S.Kom., M.T.
4. Sri Suwarno, Dr. Ir. M.Eng.

Dekan


(Restyandito, S.Kom., MSIS., Ph.D.)

Ketua Program Studi


(Gloria Virginia, Ph.D.)

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS
SECARA ONLINE
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA**

Saya yang bertanda tangan di bawah ini:

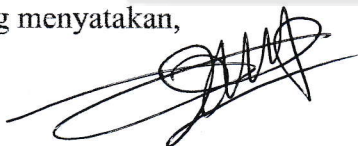
NIM : 71180279
Nama : Zanetaxina Pudihang
Prodi / Fakultas : Teknologi Informasi / Informatika
Judul Tugas Akhir : Ekstraksi Informasi Makanan dan Minuman
Tradisional Indonesia dengan Named Entity
Recognition

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty-free Right*) serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Noneklusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk *database*, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 26 Juli 2023

Yang menyatakan,



(71180279 – Zanetaxina Pudihang)



Karya sederhana ini dipersembahkan
kepada Tuhan, Keluarga Tercinta,
dan Kedua Orang Tua



It's not luck. It's God Favor

Anonim

And of his fulness have all we received, and grace for grace.

(John 1 : 16)

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Skripsi dengan judul EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL DENGAN NAMED ENTITY RECOGNITION ini telah selesai disusun.

Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Skripsi ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan yang maha kasih,
2. Orang tua, Revolius Pudihang dan Joyce Watung yang selama ini telah sabar membimbing dan mendoakan penulis tanpa kenal Lelah.
3. Bapak Restyandito, S.Kom.,MSIS.,Ph.D. selaku Dekan FTI, yang sudah mengajari dalam setiap pembelajaran terkait desain antarmuka.
4. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku Kaprodi Informatika, yang sudah memberi banyak pelajaran dan nasehat.
5. Ibu Dr. Phil. Lucia Dwi Krisnawati selaku Dosen Pembimbing 1, yang telah membimbing saya selama penyelesaian skripsi, dan mengajari saya dalam setiap pembelajaran terkait NLP
6. Bapak Budi Susanto, S.Kom., M.T., selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan membimbing penulis sejak semester 1 sampai semester terakhir ini.
7. Nettavania Pudihang selaku Kaka, dan Revazariel Pudihang selaki adik yang telah mendukung dan memberikan semangat bagi penulis
8. Riani dan Maria yang tak henti-hentinya mendengarkan keluh kesah, serta memberikan masukan dan dukungan pada penulis dalam penyelesaian skripsi ini.
9. Teman-teman Elyta, Orin, dan Gevin yang selalu memberikan dukungan dan semangat dalam penyelesaian skripsi ini.

10. Praise dan Petra yang selalu memberikan semangat serta dukungan dan motivasi dalam penyelesaian skripsi ini.
11. Teman-teman JAAVEEL, yakni Art, Taju, Verlung, Lady, Ebe, yang juga dengan setia memberikan motivasi dan penguatan bagi penulis dalam penyelesaian skripsi ini.
12. Iyan, Joel, Rindho, Hardus, Elza, Stevani, yang telah mendukung dan memberikan semangat serta selalu mengajak untuk mengerjakan skripsi bersama.
13. Teman-teman partymers di Lab FTI yang juga selalu memberikan dukungan dan semangat dalam setiap pengerjaan skripsi ini.
14. Light Cell Group yang tak henti-hentinya medoakan penulis dalam setiap proses yang dilalui oleh penulis.
15. Teman-teman grup ceria yang selalu mendengar keluh kesah penulisan skripsi ini
16. Teman-teman yang selalu memberikan motivasi dan dukungan bagi penulis dalam penulisan skripsi ini.
17. Diri sendiri yang telah berjuang ditengah setiap masalah dan keterpurukan yang ada.

Laporan proposal/skripsi ini tentunya tidak lepas dari segala kekurangan dan kelemahan, untuk itu segala kritikan dan saran yang bersifat membangun guna kesempurnaan skripsi ini sangat diharapkan. Semoga proposal/skripsi ini dapat bermanfaat bagi pembaca semua dan lebih khusus lagi bagi pengembangan ilmu komputer dan teknologi informasi.

Yogyakarta, -----

Penulis

DAFTAR ISI

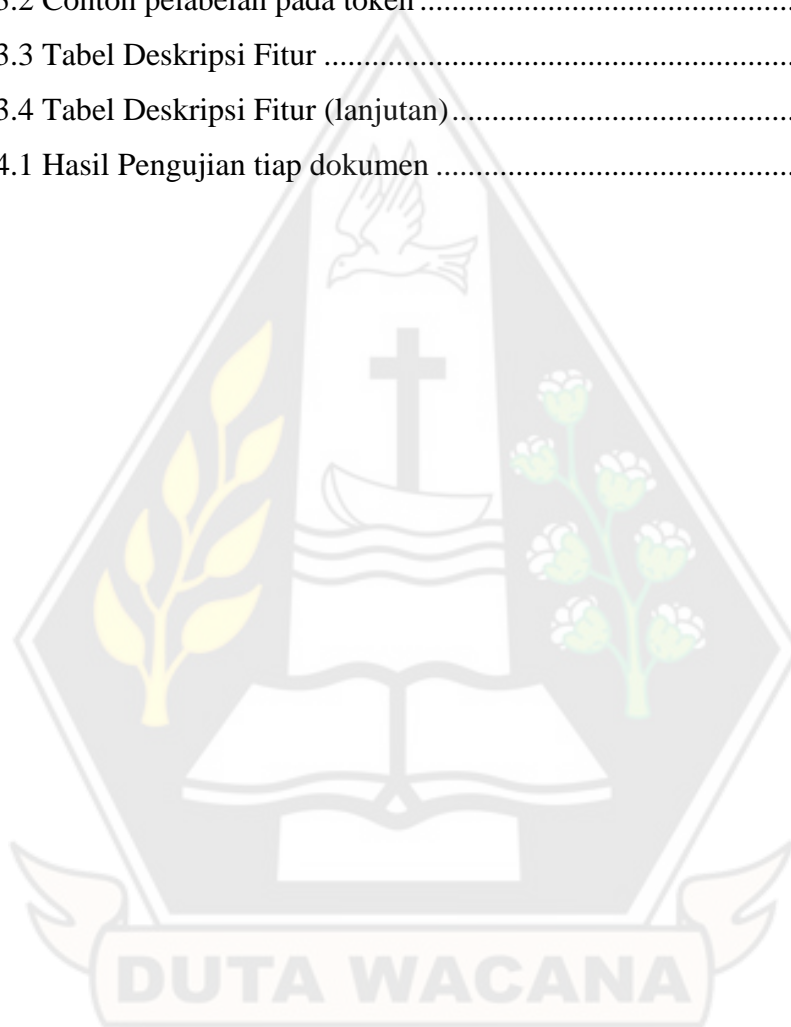
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE.....	vi
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA	vi
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
1.6. Sistematika Penulisan.....	4
BAB II.....	5
TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.1 Landasan Teori	7
2.1.1 Pemrosesan Bahasa Natural	7
2.1.2 Ekstraksi Informasi	7
2.1.3 Prapemroses Data.....	7
2.1.4 <i>Named Entity Recognition</i> (NER).....	8

2.1.5	<i>POS Tagging</i>	9
2.1.6	Ekstraksi Fitur	9
2.1.7	<i>Naïve Bayes Classifier (NBC)</i>	9
2.1.8	<i>Confusion Matrix</i>	10
2.1.9	<i>Evaluation Metric</i>	11
2.1.10	<i>Overfitting</i>	12
2.1.11	<i>Cross Validation</i>	12
2.1.12	<i>Conditional Random Field</i>	13
2.1.13	Makanan dan Minuman Tradisional	13
BAB III		15
METODOLOGI PENELITIAN		15
3.1	Analisis Kebutuhan Sistem	15
3.1.1	Kebutuhan Fungsional	15
3.1.2	Kebutuhan Non Fungsional.....	16
3.2	Perancangan Penelitian.....	16
3.2.1	Pengumpulan Data	16
3.2.2	Prapemroses Data.....	16
3.2.3	Pelabelan	17
3.2.4	Ekstraksi Fitur	18
3.2.5	Implementasi Naïve Bayes.....	19
3.2.6	<i>Named Entity Recognition</i>	19
3.2.7	Evaluasi	19
3.3	Diagram Alir.....	20
3.4	Perancangan Pengujian Sistem.....	20
BAB IV		22
IMPLEMENTASI DAN PEMBAHASAN.....		22
4.1	Implementasi Awal.....	22

4.1.1	Pengumpulan Data	22
4.1.2	Prapemrosesan.....	23
4.1.3	Mempersiapkan data latih dan data uji	24
•	Pelabelan	24
4.2	Implementasi Sistem	26
4.2.1	Ekstraksi fitur	26
4.2.2	Klasifikasi dengan Naïve Bayes.....	29
4.2.3	<i>Named Entity</i>	30
4.2.4	Antarmuka Sistem.....	31
4.3	Pengujian dan Analisis	32
BAB V	41
KESIMPULAN DAN SARAN	41
5.1	Kesimpulan.....	41
5.2	Saran.....	41
DAFTAR PUSTAKA	43
LAMPIRAN A	46
KODE SUMBER PROGRAM	46
LAMPIRAN B	56
KARTU KONSULTASI DOSEN 1	56
LAMPIRAN C	57
KARTU KONSULTASI DOSEN 2	57

DAFTAR TABEL

Tabel 2.1 Contoh <i>Named Entity Recognition</i>	8
Tabel 2.2 Tabel <i>Confusion Matrix</i>	10
Tabel 3.1 Contoh penerapan <i>Preprocessing</i>	17
Tabel 3.2 Contoh pelabelan pada token	17
Tabel 3.3 Tabel Deskripsi Fitur	18
Tabel 3.4 Tabel Deskripsi Fitur (lanjutan).....	19
Tabel 4.1 Hasil Pengujian tiap dokumen	36



DAFTAR GAMBAR

Gambar 3.1 Diagram alir penelitian.....	20
Gambar 4.1 Contoh dataset dari Korpus Artikel objek budaya	22
Gambar 4.2 Contoh dataset yang telah diubah formatnya	22
Gambar 4.3 <i>Prapemroses</i> dengan Penghapusan Tanda Baca dan tokenisasi	23
Gambar 4.4 Hasil Penghilangan Tanda Baca.....	23
Gambar 4.5 Hasil Tokenisasi	24
Gambar 4.6 Pelabelan oleh peneliti pada data latih Ayam Betutu.....	25
Gambar 4.7 Source Code Ekstraksi fitur bigram	26
Gambar 4.8 Source Code fitur <code>next_word</code> dan <code>prev_word</code>	27
Gambar 4.9 Source Code fitur POS Tag.....	27
Gambar 4.10 Menyimpan fitur ke dictionary.....	28
Gambar 4.11 Data Latih.....	28
Gambar 4.12 Pelatihan <i>Naive Bayes</i>	29
Gambar 4.13 Melakukan prediksi label pada data uji.....	30
Gambar 4.14 Filtering pada data hasil prediksi entitas	30
Gambar 4.15 Hasil <i>Named Entity</i> dari dokumen Bubur Manado	31
Gambar 4.16 Halaman Awal Tampilan Antarmuka Sistem.....	32
Gambar 4.17 Pengujian pada data latih.....	32
Gambar 4.18 Confusion Matrix	33
Gambar 4.19 Hasil Evaluasi data uji Bubur Manado.....	36
Gambar 4.20 Confusion Matrix hasil evaluasi.....	36
Gambar 4.21 Hasil Evaluasi pada data uji	38
Gambar 4.22 Hasil Validasi pada data latih.....	39
Gambar 4.23 Hasil Cross Validation	39

INTISARI

EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL DENGAN NAMED ENTITY RECOGNITION

Oleh

ZANETAXINA PUDIHING

71180279

Salah satu bagian pembelajaran dari *Information Extraction* yang paling sering digunakan saat ini adalah *Named Entity Recognition* (NER). NER berguna dalam mempermudah serta menghemat waktu dalam pencarian informasi penting dari suatu data yang besar. Penggunaan NER

Selain itu, penerapan NER membutuhkan dataset yang banyak untuk mendapatkan hasil keakuratan yang tinggi. Banyaknya dataset yang ada tentunya akan menjadi semakin sulit untuk dievaluasi dan diteliti keakuratannya. Salah satu metode yang paling sering digunakan adalah pengklasifikasian statistika dengan *Naïve Bayes*. Proses ekstraksi NER dimulai dengan prapemroses pada data, kemudian melalui proses ekstraksi fitur untuk setiap data yang ada, Hasil ekstraksi fitur berupa data latih kemudian dilatih dengan menggunakan model *Naïve Bayes*. Klasifikasi kemudian dilakukan pada data uji dengan menggunakan data training yang ada. Entitas hasil klasifikasi kemudian dievaluasi kinerjanya dengan menggunakan metrik evaluasi seperti precision, recall, dan F1-Score.

Adapun hasil dari penerapan *Naïve Bayes* dalam pencarian *Named Entity Recognition* menunjukkan hasil *precision* sebesar 75%, *recall* sebesar 78%, akurasi sebesar 87%, dan *F1-Score* adalah sebesar 76%.

Kata-kata kunci : *Naïve Bayes*, *Named Entity Recognition*, Ekstraksi Informasi

ABSTRACT

EKSTRAKSI INFORMASI MAKANAN DAN MINUMAN TRADISIONAL DENGAN NAMED ENTITY RECOGNITION

By

ZANETAXINA PUDIHAND

71180279

One part of learning from Information Extraction most frequently used nowadays is Named Entity Recognition (NER). The use of Named Entity Recognition in simplifying and saving time for searching for important information from larges data is very beneficial.

In addition, the Named Entity Recognition application requires large datasets to obtain highly accurate results. A large existing dataset will become hard to evaluate and examine for accuracy. One of the most frequently used methods is statistical classification with Naïve Bayes. The Named Entity Recognition extraction process begins with pre-processing data and then goes through a feature extraction process for each existing data. The results of feature extraction in the form of training data are then used for training using the Naïve Bayes model. Then a classification process is applied using training data and test data. The results of the named entity classifications are evaluated with evaluation metrics such as precision, recovery, and F1-Score.

The results of applying Naïve Bayes in the search for Named Entity Recognition show a precision value of 75%, recall of 78%, accuracy of 87% and an F1-Score of 76%.

Keywords : Naïve Bayes, Named Entity Recognition, Extraction Information

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Indonesia merupakan negara yang kaya akan budaya dan tentunya hal ini menjadi suatu kebanggaan tersendiri bagi setiap warga negaranya. Setiap Objek budayanya beragam dan tinggal di daerahnya sebagai ciri khas dari daerah tersebut. Sama halnya dengan makanan dan minuman tradisional Indonesia yang bermacam-macam dan sangat unik, sesuai dengan tanah kelahirannya. Hal ini tentunya menarik banyak perhatian masyarakat, apalagi dalam bidang kuliner dan penikmatnya. Ini dibuktikan dengan banyaknya ulasan kuliner tradisional Indonesia lewat artikel-artikel, dan tidak sedikit pula makanan dan minuman tradisional Indonesia yang sering diberitakan lewat berbagai media sosial maupun portal berita lainnya. Berdasarkan IDN Research Institute (2020), sebanyak 90% penggunaan internet oleh milenial adalah untuk menjelajahi internet. Hal ini disebabkan dengan adanya rasa keingintahuan dari milenial untuk mempelajari berbagai hal baru dan trend yang berkembang, dan salah satu hal yang paling sering dijelajahi diantaranya adalah kuliner.

Hal tersebut tentu saja sangat membantu masyarakat dalam mempromosikan makanan dan minuman tradisional daerah mereka, dan juga sangat membantu wisatawan dalam menjelajahi keberagaman makanan dan minuman tradisional di Indonesia. Namun dengan berkembangnya teknologi, semakin banyak artikel dan dokumen ulasan yang ditulis dan tersebar di internet, sehingga terkadang menjadi sulit untuk mencari informasi penting yang diperlukan terkait kuliner tradisional Indonesia ini. Selain itu, jika dilakukan pencarian informasi secara manual maka hal tersebut pasti akan memakan waktu yang lama. Dalam mengatasi masalah ini, diperlukan adanya ekstraksi informasi dari suatu dokumen agar informasi penting yang diperlukan dapat ditemukan dengan mudah.

Salah satu bagian pembelajaran dari *Information Extraction* yang paling sering digunakan saat ini adalah *Named Entity Recognition* (NER). *Named Entity*

Recognition memiliki peran yang penting dalam mencari suatu informasi penting terkait dokumen, dan tentunya hal ini sangat berguna dalam mengekstraksi informasi pada makanan dan minuman tradisional Indonesia untuk penelitian ini. Dalam penerapannya, *Named Entity Recognition* melakukan pengkategorian data sesuai dengan kebutuhan dan mengklasifikasikan data yang ada dalam berbagai kategori tersebut (Elsebai, 2019). Dengan demikian pencarian informasi menjadi lebih mudah dan efisien.

Berdasarkan hal ini, penelitian ini dilakukan untuk mengekstrak informasi dari dokumen makanan dan minuman tradisional Indonesia menggunakan *Named Entity Recognition*. Selain itu, penerapan NER membutuhkan data latih yang banyak untuk mendapatkan hasil keakuratan yang tinggi. Hal ini menyebabkan dibutuhkan dataset yang banyak untuk penelitian ini. Banyaknya dataset yang ada tentunya akan menjadi semakin sulit untuk dievaluasi dan diteliti keakuratannya. Salah satu metode yang paling sering digunakan dalam melakukan pengklasifikasian statistika adalah metode *Naïve Bayes*. Pengklasifikasian dengan menggunakan metode ini menguntungkan karena berdasarkan probabilitas dan setiap variabelnya diasumsikan tidak memiliki keterkaitan dengan variabel lain (Indriani, 2014). Selain itu, penggunaan metode ini terhadap dataset yang besar dapat menghasilkan tingkat kecepatan dan akurasi yang tinggi. Oleh karena itu, penelitian ini menerapkan NER dengan menggunakan metode *Naïve Bayes* untuk pengklasifikasian dan evaluasi hasil dari pengkategorian informasi tersebut. Penerapan metode *Naïve Bayes* diharapkan dapat memberikan hasil klasifikasi yang baik pada data yang ada. Penelitian ini juga merupakan bagian dari Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) dengan dana hibah RISTEK-BRIN berjudul “Portal Objek Budaya Berbasis Semantic Web”.

1.2. Perumusan Masalah

Berdasarkan Latar Belakang masalah tersebut, maka rumusan masalah yang dikaji dalam penelitian ini adalah

1. Bagaimana cara membangun dan melatih entitas makanan, bahan dan daerah dari informasi makanan dan minuman tradisional Indonesia pada model *Named Entity Recognition*?
2. Bagaimana pembuatan model *Named Entity Recognition* untuk Bahasa Indonesia menggunakan algoritma klasifikasi *Naïve Bayes*?
3. Bagaimana hasil evaluasi *accuracy*, *precision*, *recall*, dan *f1-Score* pada penerapan *Named Entity Recognition*?

1.3. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Model Pengetahuan Makanan dan Minuman Tradisional Indonesia yang digunakan didasarkan pada data korpus dokumen <https://alunalun.info/>.
2. Dataset yang diuji adalah dataset yang didasarkan pada data korpus dokumen <https://alunalun.info/>.
3. *Format file* dari data masukan yang digunakan adalah *.txt.
4. Entitas yang akan diekstraksi pada penelitian ini adalah entitas nama makanan, bahan dan daerah.
5. *Named Entity Recognition* dilakukan pada token dengan memperhatikan konteks dari token tersebut.
6. *Preprocessing* pada digunakan pada hanya berupa *tokenization*, *punctuation removal*, dan *stopword removal*.
7. Data teks yang digunakan merupakan data berbahasa Indonesia.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun dan melatih entitas nama makanan, bahan dan daerah dari informasi makanan dan minuman tradisional Indonesia pada model *Named Entity Recognition* untuk Bahasa Indonesia.
2. Membuat model *Named Entity Recognition* untuk Bahasa Indonesia menggunakan algoritma klasifikasi *Naïve Bayes*.

3. Mengevaluasi penerapan *Named Entity Recognition* dengan menggunakan parameter pengukuran *accuracy*, *precision*, *recall*, dan *f1-Score*.

1.5. Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah sebagai berikut:

1. Mempermudah dalam mencari informasi terkait makanan dan minuman Tradisional Indonesia.
2. Menambah wawasan terkait makanan dan minuman Tradisional Indonesia kepada pembaca.

1.6. Sistematika Penulisan

Bab I PENDAHULUAN, pada bab ini akan dibahas latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab II TINJAUAN PUSATAKA DAN LANDASAN TEORI berisi tinjauan pustaka dan landasan teori, pada tinjauan pustaka akan dibahas mengenai beberapa jurnal dari penelitian-penelitian sebelumnya yang berkaitan dengan penelitian ini. Selain itu, pada landasan teori berisi teori-teori terkait penelitian ini yang telah ditemukan oleh para ahli.

Bab III ANALISIS DAN PERANCANGAN, bab ini berisi pembahasan terkait analisis dan perancangan sistem yang digunakan dalam melaksanakan penelitian ini.

Bab IV HASIL DAN ANALISIS, bab ini membahas hasil dari penelitian ini, dan analisis dari peneliti terkait penelitian ini.

Bab V KESIMPULAN DAN SARAN, bab ini berisi kesimpulan dari penelitian dan saran untuk penelitian terkait pada masa yang akan datang.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Ekstraksi Informasi dengan *Named Entity Recognition*, dapat diterapkan pada data Makanan Tradisional Indonesia. Entitas yang diekstraksi dalam penelitian ini adalah entitas nama makanan, nama bahan, dan daerah di Indonesia. Implementasi NER pada data menggunakan pendekatan statistik, yaitu *Naïve Bayes*. Proses ekstraksi NER dimulai dengan prapemroses pada data, kemudian melalui proses ekstraksi fitur untuk setiap data yang ada, Hasil ekstraksi fitur kemudian dilatih dengan model *Naïve Bayes* menggunakan fitur-fitur hasil ekstraksi. Entitas hasil klasifikasi kemudian dievaluasi kinerjanya dengan menggunakan metrik evaluasi seperti precision, recall, dan F1-Score.

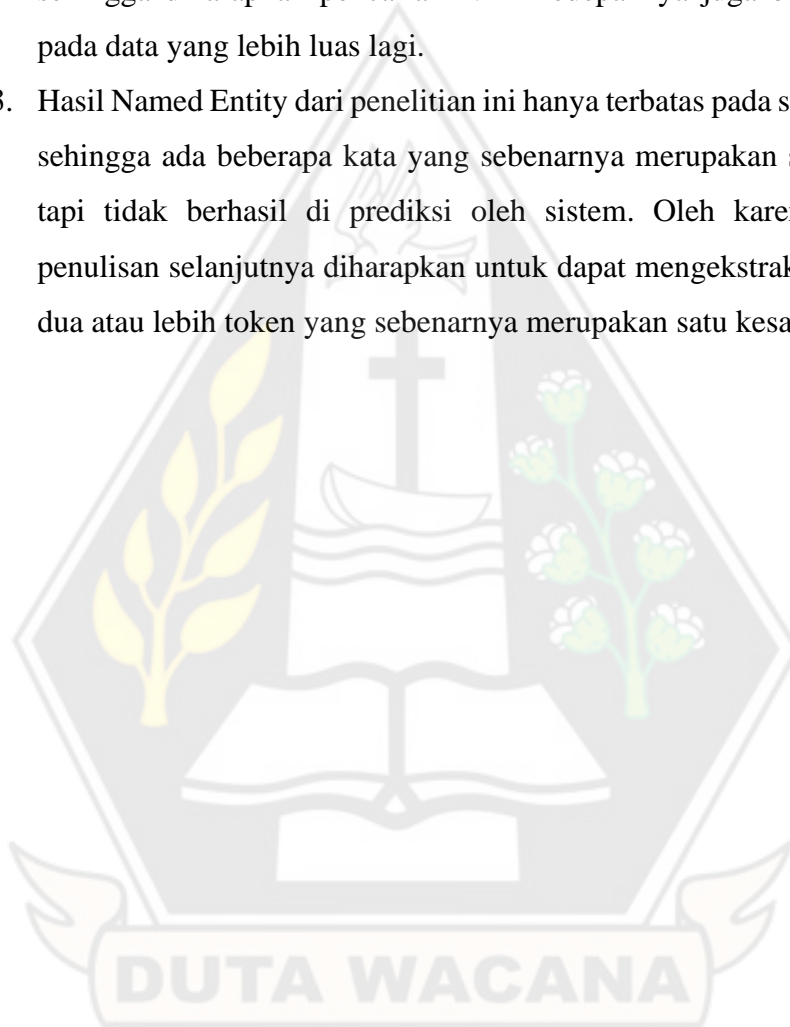
Naïve Bayes pada penelitian ini diterapkan pada 200 dokumen latih yang terdiri dari 13.197 token dengan 1262 entitas MAKANAN, 1460 entitas BAHAN 524 entitas DAERAH, serta 9951 token tanpa entitas. Data uji yang digunakan untuk evaluasi adalah sebanyak 50 dokumen, dan berhasil diekstraksikan sebanyak 449 entitas BAHAN, 223 entitas DAERAH, 294 entitas MAKANAN, dan 2568 token tanpa entitas. Adapun hasil dari penerapan *Naïve Bayes* dalam pencarian *Named Entity Recognition* menunjukkan hasil *precision* sebesar 75%, *recall* sebesar 78%, *accuracy* sebesar 87% dan F1-Score adalah sebesar 76%.

Metode *Naïve Bayes* dapat bekerja dengan mudah dalam mengatasi data dengan dimensi tinggi seperti klasifikasi teks, termasuk NER. Meskipun demikian, penerapan *Naïve Bayes* pada penelitian ini menunjukkan adanya kemungkinan *overfitting*

5.2 Saran

Adapun beberapa saran untuk pengembangan terkait penulisan serupa adalah sebagai berikut:

1. Penelitian ini hanya melakukan ekstraksi pada 3 entitas (makanan, bahan, dan daerah). Terdapat banyak entitas lainnya yang bisa membantu dalam ekstraksi informasi, terlebih untuk memperkaya informasi terkait budaya Indonesia, misalnya entitas objek budaya Indonesia.
2. Penelitian hanya terbatas pada data makanan tradisional Indonesia, sehingga diharapkan pencarian NER kedepannya juga bisa diterapkan pada data yang lebih luas lagi.
3. Hasil Named Entity dari penelitian ini hanya terbatas pada satu token saja, sehingga ada beberapa kata yang sebenarnya merupakan satu kesatuan, tapi tidak berhasil di prediksi oleh sistem. Oleh karena itu, untuk penulisan selanjutnya diharapkan untuk dapat mengekstraksi entitas dari dua atau lebih token yang sebenarnya merupakan satu kesatuan kata.



DAFTAR PUSTAKA

- Bhattacharjee, K., Karthik, S. S., Mehta, S., Kumar, A., Mehta, R., Pandya, D., . . . Verma, D. (2019). Named Entity Recognition: A Survey for Indian Languages. *2019 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, (pp. 217-220). Pune.
- Bustami. (2014). Penerapan Algoritma Naive Bayes untuk mengklasifikasi Data Nasabah Asuransi Jurnal Informatika. 8(1), 884-898.
- Chopra, A., Prashar, A., & Sain, C. (2013). Natural Language Processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1(4), 131-134.
- Cowie, J., & Lehnert, W. (1996, January). Information extraction. *Commun. ACM*, 39(1), 80-91. doi:10.1145/234173.234209
- Desikan, B. S. (2018). *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Birmingham: Packt Publishing Ltd.
- Elsebai, A. (20019). *A Rules Based System for Named Entity Recognition in Modern Standard Arabic*. Salford.
- Geetha, T., & Sendhilkumar, S. (2023). *Machine Learning: Concepts, Techniques and Application*. Oxon: Chapman & Hall/CRC Press.
- Harsana, M., Baiquni, M., Harmayani, E., & Widyaningsih, Y. A. (2018, Oktober). Potensi Makanan Tradisional Kue Kolombeng Sebagai Daya Tarik Wisata di Daerah Istimewa Yogyakarta. *Home Economics Journal (HCE)*, 1(2), 40-47.
- IDN Research Institute. (2020). *Indonesia Millennial Report 2020*. Jakarta: IDN.
- Indriai, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, (pp. G5-G10). Yogyakarta.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

- Mansouri, A., Affendy, L. S., & Mamat, A. (2008). Named Entity Recognition Approaches. *International Journal of Computer Science and Network Security*, 8(2), 339-344.
- Mariyah, S. (2017, Juni 30). Named Entity Recognition on A Collection of Research Titles. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 9(1), 61-72.
- Nurhayati, P. (2022). *Teknik Ensemble Learning untuk Peningkatan Performa Akurasi Model Prediksi (Seleksi Mahasiswa Penerima Beasiswa)*. Tangerang Selatan: Pascal Books.
- Popovski, G., Kochev, S., Seljak, B. K., & Eftimov, T. (2019). FoodIE: A Rule-based Named-entity Recognition Method for Food Information Extraction. *8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)* (pp. 915-922). Makedonia: SCITEPRESS – Science and Technology Publications.
- Rachmad, D. S. (2020, Juni). Review Named Entity Recognition Dengan Menggunakan Machine Learning. *Jurnal Sains dan Informatika*, 6(1), 28-33.
- Randy, Hasniati, & Musdar, I. A. (2018). Aplikasi Prediksi Kerusakan Smartphone menggunakan Metode Naive Bayes dan Laplace Smoothing. *Journal of Technology Research in Information System and Engineering*, 5(2), 8-16.
- Saefulloh, F. (2017). Part of Speech Tagger untuk Bahasa Indonesia Menggunakan Conditional Random Field (CRF).
- Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- T. V., G. (n.d.). *Machine Learning: Concepts, Techniques and Applications*.
- Thanaki, J. (2017). *Python Natural Language Processing*. Mumbai: Packt Publishing.
- Wulandari, D. W., Adikara, P. P., & Adinugroho, S. (2018, November). Named Entity Recognition (NER) Pada Dokumen Biologi Menggunakan Rule

Based dan Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2, 4555-4563.

Zhang, F. (2011). *Cross-validation and Regression Analysis in High-dimensional Sparse Linear Models*. California: Stanford University.

