

***NAMED ENTITY RECOGNITION (NER) PADA BAHASA LAMPUNG
BERBASIS MULTI CLASS CLASSIFICATION***

(skripsi)

Oleh

MUHAMMAD AZRIEL BINTANG SAPUTRA

1817051024



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMPUNG

BANDAR LAMPUNG

2023

***NAMED ENTITY RECOGNITION (NER) PADA BAHASA LAMPUNG
BERBASIS MULTI CLASS CLASSIFICATION***

Oleh

MUHAMMAD AZRIEL BINTANG SAPUTRA

Skripsi

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2023

ABSTRACT

NAMED ENTITY RECOGNITION IN LAMPUNG LANGUAGE BASED ON MULTI-CLASS CLASSIFICATION

By

Muhammad Azriel Bintang Saputra

The use of Lampung language in the rajabasa area is 50% of the 6812 total respondents, while 50% use Indonesian as their daily language. Language is a means of communication used to convey ideas, thoughts and information. Information can be accessed on various social media and online news portals. News portals are a form of unstructured data. Unstructured data can cause problems when trying to sort it and find information, so it requires tools and a long time. We can use Natural Language Processing (NLP) and Named Entity Recognition (NER) to analyze and shorten the time to get information from unstructured data. The way NER works is to recognize an entity and then give a label or name to a named entity (NE) in a text. SpaCy library is a library that performs NLP and NER with better performance than other python libraries. This research is done labeling and grouping named entities into several categories based on Multi Class Classification with the aim of producing an information system that can present the results of information extraction of Lampung language text data using NER based on Multi Class Classification. The method in this research uses spaCy with CNN and LSTM algorithms for machine learning implementation and extreme programming for information system implementation. An information system is obtained that can perform the NER process from the input of Lampung language text and produce extracted information and the system can classify text based on the specified entity labeling. In further research, a lot of data is needed so that there is no unbalanced data in the modeling process.

Keywords: Lampung Language, NER, SpaCy, LSTM, CNN

ABSTRACT

NAMED ENTITY RECOGNITION (NER) PADA BAHASA LAMPUNG BERBASIS MULTI CLASS CLASSIFICATION

Oleh

Muhammad Azriel Bintang Saputra

Penggunaan bahasa lampung didaerah Rajabasa sebanyak 50% dari 6812 jumlah responden, sedangkan 50% nya menggunakan bahasa Indonesia sebagai bahasa sehari-hari. Bahasa merupakan sarana komunikasi yang digunakan untuk menyampaikan ide, pikiran dan juga informasi. Informasi dapat diakses diberbagai sosial media dan portal berita online. Portal berita merupakan bentuk dari data yang tidak terstruktur. Data tidak terstruktur dapat menimbulkan masalah saat mencoba mengurutkannya dan mencari sebuah informasi, sehingga memerlukan alat dan waktu yang lama. Kita dapat menggunakan Natural Language Processing (NLP) dan Named Entity Recognition (NER) untuk menganalisis dan mempersingkat waktu dalam mendapatkan informasi dari data tidak terstruktur. Cara kerja NER adalah melakukan pengenalan terhadap suatu entitas lalu memberikan label atau nama pada sebuah entitas bernama atau Named Entity (NE) dalam sebuah teks. Library SpaCy merupakan library yang data melakuakn NLP dan NER dengan performa yang lebih baik daripada library phyton lainnya. Penelitian ini dilakukan pelabelan dan pengelompokkan entitas bernama menjadi beberapa kategori berbasis *Multi Class Classification* dengan tujuan menghasilkan sistem informasi yang dapat menyajikan hasil ekstraksi informasi data teks bahasa lampung menggunakan NER berbasis *Multi Class Classification*. Metode pada penelitian ini menggunakan spaCy dengan algoritma CNN dan LSTM untuk implementasi machine learning dan extreme programming untuk implementasi sistem informasi. Didapatkan sistem informasi yang dapat melakukan proses NER dari input teks bahasa lampung dan menghasilkan informasi hasil ekstraksi serta sistem dapat mengklasifikasikan teks berdasarkan pelabelan entitas yang ditetapkan. Pada penelitian selanjutnya diperlukan data yang banyak agar tidak terjadi unbalanced data pada proses pembuatan model.

Keywords: Bahasa Lampung, NER, SpaCy, LSTM, CNN

Judul Skripsi : **NAMED ENTITY RECOGNITION (NER) PADA BAHASA LAMPUNG BERBASIS MULTI CLASS CLASSIFICATION**

Nama Mahasiswa : **Muhammad Aerial Bintang Saputra**

Nomor Induk Mahasiswa : **1817051024**

Program Studi : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. **Komisi Pembimbing**

Dr. Ir. Kurnia Muludi, M.S.Sc.
NIP. 19640616 198902 1 001

2. **Ketua Jurusan Ilmu Komputer**

Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

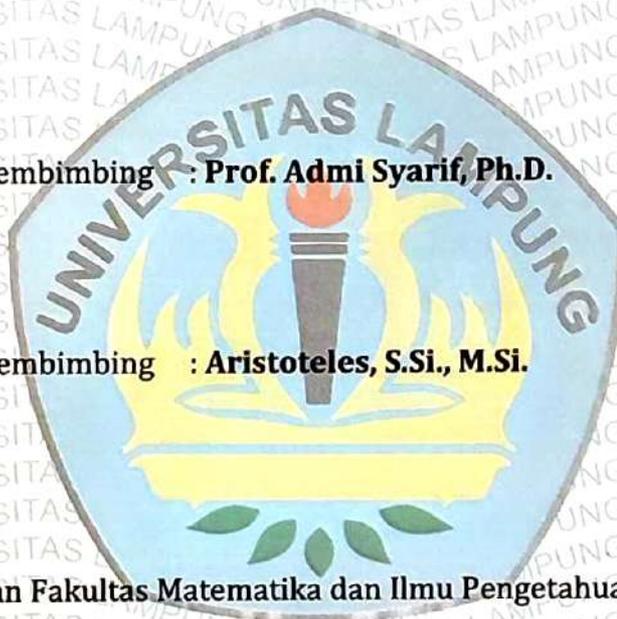
MENGESAHKAN

1. Tim Penguji

Ketua : Dr. Ir. Kurnia Muludi, M.S.Sc.

**Penguji
Bukan Pembimbing : Prof. Admi Syarif, Ph.D.**

**Penguji
Bukan Pembimbing : Aristoteles, S.Si., M.Si.**



Kurnia Muludi

Admi Syarif

2. Plt. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi : 21 Maret 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Azriel Bintang Saputra

NPM : 1817051024

Menyatakan bahwa skripsi saya yang berjudul "***Named Entity Recognition (NER) Pada Bahasa Lampung Berbasis Multi Class Classification***" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 28 Maret 2023



Muhammad Azriel Bintang Saputra

NPM. 1817051024

RIWAYAT HIDUP



Penulis dilahirkan di Brebes, Jawa tengah, pada tanggal 11 Februari 2001, Sebagai anak kedua dari tiga bersaudara.

Penulis menyelesaikan pendidikan formal SDN Jatibarang Lor 02 dan selesai pada tahun 2012, kemudian pendidikan menengah pertama SMPN 1 Cikarang Utara pada tahun

2015, lalu melanjutkan pendidikan menengah atas di SMAN

1 Cikarang Utara yang diselesaikan pada tahun 2018.

Pada tahun 2018 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMPTN. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain

1. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu komputer pada periode 2018/2019.
2. Menjadi anggota Bidang Kaderisasi Himpunan Mahasiswa Jurusan Ilmu komputer periode 2018/2019.
3. Menjadi Wakil Ketua Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2019/2020.
4. Menjadi Asisten Dosen di Jurusan Ilmu Komputer tahun 2020 hingga 2021.

5. Melaksanakan Kerja Praktek sebagai User Interface Designer pada bulan Juli periode 2020/2021 di PT Poseidon Aquakultura Sejahtera.
6. Melaksanakan Kuliah Kerja Nyata Kebangsaan dan Bersama di Desa Bakung, Kecamatan Maro Sebo, Kabupaten Muaro Jambi, Jambi pada tahun 2020/2021 dengan program pelatihan dan *micro-teaching* teknologi komputer terhadap remaja dan anak-anak Desa Bakung.
7. Menjadi IT Support Specialist di Center for Career and Entrepreneurship (CCED) Pada September 2021 sampai Maret 2022.

MOTTO

1. “Jangan menjelaskan dirimu kepada siapa pun, karena yang menyukaimu tidak butuh itu. Dan yang membencimu tidak percaya itu”

(Ali bin Abi Thalib)

2. “Orang yang hebat adalah orang yang memiliki kemampuan menyembunyikan kesusahan, sehingga orang lain mengira bahwa ia selalu senang”

(Imam Syafi’i)

3. “Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya”

(QS Al-Baqarah: 286)

4. “Jangan bersedih atas apa yang telah berlalu, kecuali kalau itu bisa membuatmu bekerja lebih keras untuk apa yang akan datang”

(Umar bin Khatab)

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'ala' atas segala

Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini.

Shalawat serta salam selalu tercurahkan Kepada Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Yang senantiasa memberikan yang terbaik, dan melantunkan do'a yang selalu

Menyertaiku. Kuucapkan pula terimakasih sebesar-besarnya karena telah

mendidik dan membesarkanku dengan cara yang dipenuhi kasih sayang,

dukungan, dan pengorbanan yang belum bias terbalaskan.

Seluruh Keluarga Besar Ilmu Komputer 2018

Yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat bernaung melahap semua ilmu untuk menjadi bekal hidup

SANWACANA

Puji syukur kehadiran Allah SWT atas berkah, rahmat dan hidayat-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad Sholallahu Alaihi Wasallam penulis dapat menyelesaikan skripsi yang berjudul “*Named Entity Recognition (NER) Pada Bahasa Lampung Berbasis Multi Class Classification*” dengan baik dan lancar. Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun skripsi ini, antara lain.

1. Kedua orang tua serta kakak dan adikku tercinta yang memberi dukungan, do’a, semangat, motivasi, dan kasih sayang yang luar biasa tak terhingga. Semua yang telah kalian berikan tidak akan pernah mampu untukku balas. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan akhirat.
2. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. sebagai pembimbing utama yang telah memberikan arahan, ide, motivasi, kritik serta saran kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik
3. Bapak Prof. Admi Syarif, Ph.D sebagai pembahas 1 yang telah memberikan masukan yang bermanfaat dalam perbaikan skripsi ini.
4. Bapak Aristoteles, S.Si., M.Si. sebagai pembahas 2 penulis yang telah memberikan masukan yang bermanfaat dalam perbaikan skripsi ini.
5. Bapak Dr. Eng. Heri Satria, S.Si M.Si. selaku Plt. Dekan FMIPA Universitas Lampung.

6. Bapak Didik Kurniawan, S.Si., M.T. selaku ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
7. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung sekaligus pembimbing akademik penulis yang mendukung peningkatan akademik penulis selama perkuliahan.
8. Ibu Ade Nora Maela, Bang Zainuddin dan Mas Nofal yang telah membantu segala urusan administrasi penulis di Jurusan Ilmu Komputer.
9. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman dalam hidup untuk menjadi lebih baik.
10. Rini Anggraini S.Si yang senantiasa memberikan semangat kepada penulis dan menjadi tempat berkeluh kesah.
11. Sahabat-sahabat Paris Kost yang selalu menemani serta menjadi tempat bertukar pikiran dan berbagi ide gagasan.
12. Vindo Rizkiyanto sebagai sahabat perjuangan dalam mengerjakan skripsi dari awal penetapan tema hingga tulisan ini selesai.
13. Keluarga Ilmu Komputer 2018 yang tidak bisa penulis sebut satu persatu. Keluarga kedua penulis, rekan kelompok, rekan diskusi, rekan bercanda, dan telah memberi arti dan warna serta pengalaman tak ternilai semasa duduk di bangku kuliah.
14. Seluruh kakak tingkat dan adik tingkat Ilmu Komputer yang tidak bisa disebutkan satu persatu yang telah menjadi warna selama masa perkuliahan penulis.

15. Teman-teman Himakom yang sudah mengajarkan banyak hal dalam berorganisasi, memberikan banyak pengalaman.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, semoga skripsi ini membawa manfaat dan keberkahan bagi semua civitas Ilmu Komputer Universitas Lampung aamiin ya rabbal aalamiin.

Bandar Lampung, 28 Maret 2023



Muhammad Azriel Bintang Saputra
NPM. 1817051024

DAFTAR ISI

	Halaman
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA	5
2.1. Penelitian Terdahulu.....	5
2.2. Uraian Tinjauan Pustaka	9
2.2.1. Bahasa Daerah	9
2.2.2. Bahasa Lampung.....	9
2.2.3. Portal Berita <i>Online</i>	9
2.2.4. Sistem Informasi	10
2.2.5. <i>Framework</i> Flask Python.....	10
2.2.6. <i>Extreme Programming</i>	11
2.2.7. <i>Black Box Testing</i>	13
2.2.8. HTML (<i>Hypertext Markup Language</i>).....	13
2.2.9. Python	14
2.2.10. UML (Unified Modeling Language)	14
2.2.11. Wireframe	16
2.2.12. Natural Language Processing	17
2.2.13. Text Mining	17
2.2.14. Information Extraction.....	18
2.2.15. Named Entity Recognition.....	18
2.2.16. SpaCy.....	20
2.2.17. <i>Multi Class Classification</i>	21

2.2.18.	<i>Convolutional Neural Network</i>	22
2.2.19.	<i>Long Short Term Memory</i>	24
2.2.20.	Hyperparameter	26
III. METODOLOGI PENELITIAN		27
3.1.	Waktu dan Tempat	27
3.1.1.	Waktu.....	27
3.1.2.	Tempat	27
3.2.	Data dan alat	29
3.2.1.	Data.....	29
3.2.2.	Alat.....	31
3.3.	Alur Kerja Penelitian.....	32
3.3.1.	Studi Literatur	34
3.3.2.	Pembuatan Model	34
3.3.3.	Perencanaan sistem.....	37
3.3.4.	Perancangan sistem.....	37
3.3.5.	Pengkodean sistem.....	46
3.3.6.	Pengujian sistem	46
IV. HASIL DAN PEMBAHASAN		47
4.1.	Hasil.....	47
4.1.1.	Hasil Pengumpulan Data	48
4.1.2.	Hasil Pelabelan Data	48
4.1.3.	Hasil Pelatihan Data.....	50
4.1.4.	Hasil Evaluasi Kinerja Model.....	50
4.2.	Pembahasan	52
4.2.1.	Hasil Implementasi <i>Machine Learning</i>	52
4.2.2.	Hasil Implementasi Metode <i>Extreme Programming</i>	58
4.2.3.	Hasil Pengkodean Sistem.....	64
4.2.4.	Hasil Pengujian Sistem	67
V. KESIMPULAN DAN SARAN		72
5.1.	Kesimpulan.....	72
5.2.	Saran.....	73
DAFTAR PUSTAKA		74

DAFTAR TABEL

	Halaman
Tabel 1. Penelitian Terdahulu	5
Tabel 2. Contoh Multi Class Classification	21
Tabel 3. Kegiatan dan Jadwal Penelitian	28
Tabel 4. Kelompok Entitas.....	29
Tabel 5. Skenario Pelatihan Data	36
Tabel 6. Deskripsi Usecase Diagram	38
Tabel 7. Jumlah Data.....	48
Tabel 8. Jumlah Penyebaran Label	48
Tabel 9. Hasil Pelatihan Data.....	50
Tabel 10. Hasil Evaluasi Kinerja Model.....	50
Tabel 11. Hasil Evaluasi Setiap Label	51
Tabel 12. Dokumentasi Pengembangan Sistem Dengan Metode Extreme Programming.....	59
Tabel 13. Hasil Blackbox Testing.....	68

DAFTAR GAMBAR

	Halaman
Gambar 1. Fase Pada Extreme Programming (Pressman, 2005)	12
Gambar 2. Notasi Usecase Diagram (Ardian et al, 2018).....	15
Gambar 3. Notasi Pada Activity Diagram (Ardian et al, 2018).....	16
Gambar 4. Contoh Named Entity Recognition (Brown , 2020).....	19
Gambar 5. Arsitektur spaCy (SpaCy, 2022).	21
Gambar 6. Arsitektur CNN (Yahia dan Abdulazeez, 2021).	22
Gambar 7. Arsitektur Model CNN-LSTM (Hermanto et al, 2021).	23
Gambar 8. Arsitektur Model LSTM-CNN (Hermanto et al, 2021).	24
Gambar 9. Ilustrasi LSTM (Rizki et all, 2020).....	25
Gambar 10. Diagram Alur Penelitian.....	33
Gambar 11. Usecase Diagram Sistem.....	38
Gambar 12. Activity Diagram Melihat Halaman Beranda.....	40
Gambar 13. Activity Diagram Memasukan Input Teks Ekstraksi	41
Gambar 14. Activity Diagram Melihat Hasil Output Ekstraksi.....	41
Gambar 15. Activity Diagram Melihat Halaman Tentang.....	42
Gambar 16. Rancangan Tampilan Antarmuka Sistem Beranda.....	43
Gambar 17. Rancangan Tampilan Antarmuka Sistem Input Teks Ekstraksi.....	44
Gambar 18. Rancangan Tampilan Antarmuka Sistem Hasil Teks Ekstraksi.....	45
Gambar 19. Rancangan Tampilan Antarmuka Sistem Tentang.....	45
Gambar 20. Proses Pelabelan Data	53

Gambar 21. Hasil Tahap 1 Desain Wireframe Halaman Beranda	61
Gambar 22. Hasil Tahap 2 Perbaikan Desain Wireframe	62
Gambar 23. Desain user interface sebelum perbaikan	63
Gambar 24. Desain user interface sesudah perbaikan.....	63
Gambar 25. Hasil Tahap 3 Penerapan NER pada Sistem	64
Gambar 26. Halaman Beranda	65
Gambar 27. Hasil Input Teks Ekstraksi	65
Gambar 28. Halaman Hasil Ekstraksi	66
Gambar 29. Halaman Tentang	66

I. PENDAHULUAN

1.1. Latar Belakang

Bahasa merupakan sarana komunikasi yang dipergunakan oleh manusia untuk menyampaikan ide, pikiran dan juga informasi. Bahasa memungkinkan manusia untuk menyampaikan suatu informasi baik secara lisan ataupun tulisan (Devianty, 2017). Menurut studi yang dilakukan Agustina pada tahun 2015, menyebutkan bahwa pemakaian bahasa lampung di daerah Rajabasa tidak digunakan oleh seluruh masyarakat daerah tersebut. Hasil penelitian dilakukan dengan mengambil responden sebanyak 6812 masyarakat Rajabasa. Terdapat 50% masyarakat Rajabasa yang menggunakan bahasa lampung dan 50% sisanya menggunakan bahasa indonesia sebagai bahasa sehari-hari. Penggunaan bahasa lampung tidak intensif digunakan pada masyarakat yang berasal dari keluarga perkawinan antar etnis, padahal bahasa daerah khususnya di daerah lampung yaitu bahasa lampung adalah suatu ciri khas dan budaya yang seharusnya harus terus digunakan dan dipahami oleh seluruh kalangan.

Pada zaman modern seperti ini, informasi dapat dengan mudah diakses di berbagai macam sosial media dan portal berita online. Berdasarkan topik pencarian pada Google Trends, pada tahun 2022 topik peristiwa nasional nomor 1 yang paling sering dicari oleh orang indoneisa adalah minyak goreng, dimana topik tersebut dapat ditemui pada pada artikel-artikel berita pertanian. Berita pertanian sering dicari untuk mendapatkan informasi mengenai perkembangan naik turunnya harga produk pertanian. Namun, informasi - informasi yang berada di sosial media ataupun portal berita online masih merupakan bentuk dari data tidak terstruktur. *Unstructured* data merupakan sekelompok data teks dengan format tidak pasti yang belum memiliki bentuk secara khusus. Contoh

unstructured data ialah foto, gambar, video, suara, dan berbagai macam teks. Pada teks laporan, email, postingan sosial media, dan portal berita *online* masih merupakan bentuk dari data tidak terstruktur. *Unstructured* data bisa menimbulkan masalah saat kita mencoba mengurutkannya dan mencari sebuah informasi pada data tersebut, sehingga memerlukan alat dan waktu yang lebih untuk mendapatkan informasi dari data yang tidak terstruktur (Maryanto 2017). Untuk menganalisis serta mempersingkat waktu dalam mendapatkan informasi dari suatu data yang tidak terstruktur dapat digunakan *Natural Language Processing* (NLP) dan *Named Entity Recognition* (NER).

Named Entity Recognition (NER) merupakan bagian dari *Text Mining* dan *Natural Language Processing*, keduanya sangat penting untuk mengekstraksi sebuah informasi khususnya informasi dari sebuah teks (Zhai & Charu, 2012). *Named Entity Recognition* (NER) merupakan sebuah turunan dari ekstraksi informasi yang memiliki tujuan untuk memudahkan melakukan pencarian informasi dengan melakukan pemberian nama entitas pada setiap kata dalam suatu teks. Cara kerja dari *Named Entity Recognition* (NER) adalah melakukan pengenalan terhadap suatu entitas lalu memberikan label atau nama pada sebuah entitas bernama atau *Named Entity* (NE) dalam sebuah teks. Pelabelan atau penamaan entitas bernama pada setiap kata dapat mempercepat proses ekstraksi informasi yang dilakukan (Setioaji et al, 2017). *Named Entity* atau entitas bernama adalah kata atau ekspresi yang secara unik menggambarkan suatu elemen di antara sekumpulan elemen yang memiliki atribut serupa. Contoh entitas bernama seperti organisasi, nama orang, nama tempat dan penyakit yang ada pada bidang biomedis (Hidayatullah dan Saputro, 2021).

Pada penelitian yang dilakukan oleh Atika pada tahun 2021, ia melakukan penelitian mengenai ekstraksi informasi pada portal berita *online* dengan menggunakan *named entity recognition* menggunakan *library* pada python yang dapat melakukan tugas *natural language processing* yaitu *spaCy*. pada penelitian yang dilakukannya dengan menggunakan *library spaCy*, *library* ini dapat melakukan tugas NLP dengan baik dan model buatannya dapat

diimplementasikan menjadi website yang dapat mengekstraksi berita *online*. SpaCy adalah *open source library* yang dapat menangani permasalahan mengenai *Natural Language Processing* berbasis Python yang dapat melakukan tugas seperti *dependency parsing*, *part-of-speech tagging*, *text classification*, *named entity recognition*, *dependency parsing*, *text classification*, *sentence segmentation*, dan sebagainya (Yanti et al, 2021). SpaCy dapat mengidentifikasi secara cepat entitas-entitas yang ada pada dokumen. Entitas dapat terdiri dari berbagai jenis seperti orang, lokasi, organisasi, tanggal, angka dan label entitas yang kita buat. SpaCy juga *library* yang mudah digunakan, dan memberikan nilai performa keseluruhan terbaik jika dibandingkan dengan Stanford CoreNLP Suite, Google's SyntaxNet, dan NLTK Python *library* (Bolkyci et al, 2020).

Pada penelitian ini akan dilakukan pelabelan dan pengelompokan entitas bernama menjadi beberapa kategori berbasis *Multi Class Classification*, Serta dibuatnya sistem informasi berbasis website yang dapat mengekstraksi informasi dari data yang telah di kelompokkan tersebut menggunakan *Named Entity Recognition*. Oleh sebab itu perlu dilakukannya penelitian mengenai *Named Entity Recognition* Bahasa Lampung Berbasis *Multi Class Classification*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan maka dapat dirumuskan suatu permasalahan sebagai berikut :

1. Mengembangkan sistem yang dapat mengekstraksi informasi suatu teks bahasa Lampung menggunakan *Named Entity Recognition* berbasis *Multi Class Classification*".
2. Membuat Model *Named Entity Recognition* berbasis *Multi Class Classification* berbahasa Lampung.

1.3. Batasan Masalah

Batasan masalah pada penelitian ini yaitu :

1. Metode yang digunakan untuk pengembangan sistem adalah *extreme programming*, dan menggunakan framework Flask.
2. Model *Named Entity Recognition* dibuat dengan library spaCy dengan algoritma LSTM & CNN.
3. Penelitian ini berfokus pada data bidang pertanian berbahasa lampung dialek api dengan entitas yang diberi tanda nama, tanggal, hari, lokasi, penyakit, produk pertanian, musim, organisasi, dan harga.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan sistem informasi yang dapat menyajikan hasil ekstraksi informasi data teks Bahasa lampung menggunakan *Named Entity Recognition* berbasis *Multi Class Classification*.
2. Menghasilkan model yang dapat melakukan proses *Named Entity Recognition* berbasis *Multi Class Classification* berbahasa Lampung.

1.5. Manfaat Penelitian

Manfaat Penelitian ini yaitu :

1. Menghasilkan sistem yang dapat mengklasifikasikan teks dengan model *Named Entity Recognition* berdasarkan pelabelan entitas bernama yang sudah ditetapkan.
2. Mengetahui nilai evaluasi kinerja model yang dibuat dengan metode yang digunakan.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini tidak terlepas dari penelitian yang sudah dilakukan sebelumnya. Beberapa penelitian terdahulu digunakan sebagai referensi dalam penelitian ini. Daftar penelitian dapat dilihat pada Tabel 1.

Tabel 1. Penelitian Terdahulu

No	Penelitian	Metode	Hasil
1.	Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study: Power Failure in the Special Region of Yogyakarta) (Yanti et al, 2021).	Algoritma Library Spacy	<i>Library</i> spaCy mampu membuat sebuah model NER yang memiliki nilai evaluasi yang diukur dengan Precision, Recall, dan F1-Score dengan nilai masing-masing Precision 95%, Recall 93%, dan F1-Score 94%.

Tabel 1. (lanjutan)

No	Penelitian	Metode	Hasil
2.	Ekstraksi Informasi Berita Online Dengan <i>Named Entity Recognition</i> (NER) Dan <i>Rule-Based</i> Untuk Visualisasi Penyakit Tropis Di Indonesia (Atika, 2021)	<i>Rule based</i>	Sistem yang mampu mengekstraksi informasi dari portal berita online mengenai penyakit tropis di Indonesia.
3.	Text Summarization Using Spacy Algorithm(Vijay M.N et al, 2018)	Algoritma dari library spaCy, Gensim, dan Sumy.	Library spaCy menjadi library yang dapat melakukan tugas text summarization terbaik jika dibandingkan dengan dua library lainnya,
4.	Perbandingan Metode Lstm dan Gru (Rnn) Untuk Klasifikasi Berita Palsu Berbahasa Indonesia (Hanifa et al, 2021).	LSTM dan GRU(RNN)	Nilai akurasi yang didapatkan oleh model LSTM sebesar 73% dan untuk model GRU 64%.
5.	Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo (Kurniawan et al, 2021).	BiLSTM, CNN, dan CRF	Penambahan CNN setelah BiLSTM dapat meningkatkan nilai akurasi, dan diperoleh akurasi sebesar 93%

Tabel 1. (lanjutan)

No	Penelitian	Metode	Hasil
6.	Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network (Maulana dan Rochmawati, 2020).	CNN	Sistem mampu mengklasifikasikan komentar instagram pada tiga kelas yaitu negatif, positif, dan netral. Hasil evaluasi didapatkan pada data <i>testing</i> yaitu presisi 96%, Akurasi 88%, recall 68%.
7.	Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online (Hermanto et al, 2021).	LSTM, dan CNN	Diukur dengan presisi, recall, dan akurasi model CNN-LSTM merupakan model terbaik jika dibandingkan dengan model LSTM, dan LSTM-CNN dengan presisi 66%, recall 76%, dan akurasi 66%
8.	Analisa Learning rate dan Batch size Pada Klasifikasi Covid Menggunakan Deep learning dengan Optimizer Adam (Rochmawati et al, 2021).	Deep Learning optimizer adam	Learning rate yang mendapatkan hasil paling optimal adalah 0.0001, dan 0.00001

Tabel 1. (lanjutan)

No	Penelitian	Metode	Hasil
9.	Pengembangan Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5 (Darmawan dan Iswari, 2022).	Decision Tree c.45	Sistem yang dikembangkan berbasis web mampu mengklasifikasikan data yang dibangun menggunakan bahasa pemrograman python dan microframework python flask.
10.	Pengaruh Fungsi Aktivasi, Optimisasi dan Jumlah Epoch Terhadap Performa Jaringan Saraf Tiruan (Wibawa, 2017).	ANN	Fungsi optimisasi adam memiliki nilai akurasi tertinggi jika dibandingkan dengan fungsi optimisasi atau aktivasi lainnya. Rata-rata nilai epoch yang dibutuhkan untuk mencapai akurasi tertinggi pada setiap data set adalah diatas 200.

2.2. Uraian Tinjauan Pustaka

2.2.1. Bahasa Daerah

Bahasa daerah adalah bukti adanya sebuah peradaban yang berasal dari suatu masyarakat lampau, baik berupa lisan maupun tulisan. Bahasa daerah sendiri dapat dipahami sebagai sistem ilmu pengetahuan dengan nilai-nilai yang dimiliki oleh masyarakat yang mempengaruhi kepribadian masyarakat itu sendiri. Namun demikian, Bahasa daerah memiliki sifat dinamis yaitu mengalami perubahan dan tentunya mengarah terjadinya pergeseran Bahasa (Setyawan, A. 2011).

2.2.2. Bahasa Lampung

Bahasa Lampung adalah bahasa yang digunakan oleh masyarakat/abdi Lampung di Provinsi Lampung, Pantai Barat Banten, dan Selatan Palembang. Bahasa ini masuk kedalam cabang Sundik, dari kelompok bahasa Melayu-Polinesia Barat. Bahasa lampung masih berkerabat dengan bahasa Melayu, bahasa Batak, bahasa Sunda, bahasa Jawa, dan bahasa Bali. Bahasa Lampung memiliki rumpun bahasa sendiri. Rumpun Bahasa Lampung terdiri atas bahasa Lampung Nyo dan bahasa Komerling bahasa Lampung Api. Dialek Nyo yang digunakan oleh masyarakat yang menggunakan adat Lampung Pepadun, sedangkan Dialek Api digunakan oleh masyarakat yang menggunakan adat Lampung Peminggir/Saibatin. Kedua, (Agustina, 2015).

2.2.3. Portal Berita *Online*

Portal berita online adalah tempat penyebaran informasi yang aktual mengenai suatu kejadian, informasi yang didapatkan dari artikel berita online dapat digunakan sebagai sumber informasi dalam mengumpulkan data. Informasi pada suatu artikel berita online mengandung unsur dasar seperti apa (*what*), siapa

(*who*), kapan (*when*), dimana (*where*), mengapa (*why*), dan bagaimana (*how*). Informasi yang ada pada suatu artikel berita online dapat menjadi sumber informasi dalam mendapatkan data (Putra dan Kurniawan, 2020).

2.2.4. Sistem Informasi

Sistem informasi merupakan suatu komponen yang dapat bekerja sama untuk memproses, menyimpan, mengumpulkan, dan mendistribusikan informasi guna membantu dalam membuat suatu keputusan, mengkoordinasikan, dan memberikan suatu gambaran kegiatan dalam suatu organisasi (Sidh, 2013).

Sistem informasi dapat didefinisikan sebagai perangkat entitas yang terdiri dari *brainware*, *software*, dan *hardware* yang bekerja sama untuk menyediakan data yang diproses sehingga dapat berguna untuk pengguna yang akan memakai data tersebut (Herliana dan Rasyid, 2016).

2.2.5. Framework Flask Python

Framework Flask Python merupakan web *framework* berbahasa python dan termasuk dalam jenis *microframework*. Flask bertindak sebagai kerangka aplikasi dan dapat menjadi tampilan dari suatu web. Dengan menggunakan flask dan juga bahasa python, pengembang dapat membuat web yang terstruktur dan mengatur interaksi web dengan mudah (Irsyad, 2018).

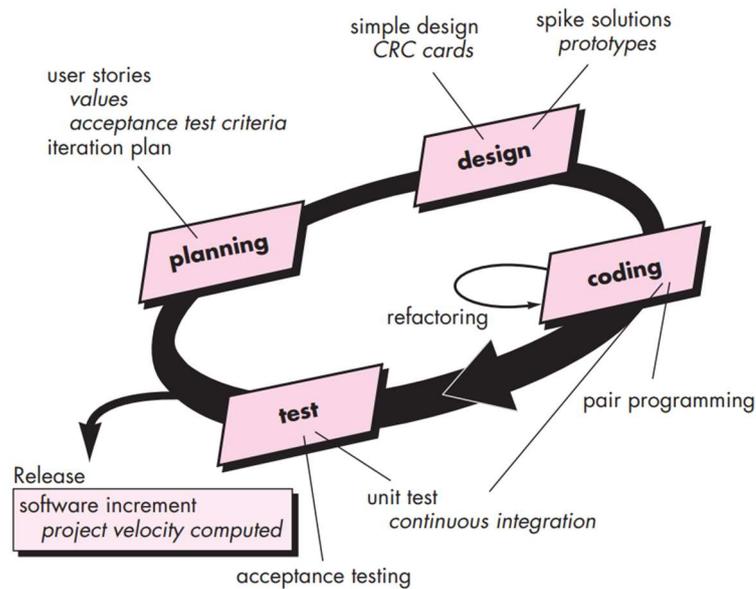
Menurut Yuniuss (2017), Fitur-fitur yang terdapat pada framework flask adalah *debugger* cepat, *build-in development server*, *integrated support unit test*, kompatibel dengan mesin aplikasi *RESTful request dispatching*, Google, Jinja 2 *templating*, berbasis *Unicode*, dan mendukung *secure cookies*.. Flask juga memiliki berbagai kelebihan seperti:

1. Ringan karena mempunyai *core* sederhana dan desain modular.
2. mampu menangani fungsi HTTP *request* dengan lebih mudah.
3. API yang koheren dan baik.
4. Sudah banyaknya dokumentasi dan terstruktur dengan baik, penuh dengan contoh yang dapat digunakan langsung.
5. Mudah untuk di-*deploy* saat produksi.
6. Mudah untuk dikontrol secara menyeluruh.

2.2.6. *Extreme Programming*

Menurut Fathoni dan Dwi (2016), Kent Beck, seorang pakar rekayasa perangkat lunak menciptakan metode *Extreme Programming*. Metode *Extreme Programming* atau biasa disebut dengan metode XP. *Extreme programming* lebih efektif, mudah beradaptasi, dan gesit untuk pengembangan perangkat lunak. Nilai dasar metode extreme programming :

1. *Communication* : Memastikan komunikasi berjalan secara efektif antara programmer dan pengguna, dan juga antar sesama programmer.
2. *Courage* : Pengembang perangkat lunak harus selalu melakukan tugasnya dengan keyakinan diri, keberanian, dan kejujuran.
3. *Simplicity* : Jalankan semua tahap pengembangan dengan sederhana.
4. *Feedback* : Mengandalkan umpan balik dan membutuhkan tim yang berkualitas.
5. *Quality Work* : Perangkat lunak yang baik berasal dari proses yang baik.



Gambar 1. Fase Pada Extreme Programming (Pressman, 2005)

Menurut Carolinna dan Supriyatna (2019), Terdapat 4 tahap yang harus dilakukan pada metode *extreme programming*(xp) yaitu:

1. *Planning* (Perencanaan). *Planning* adalah langkah pertama dalam pengembangan sistem. Pada tahap ini dilakukan sejumlah kegiatan perencanaan, antara lain mencari tahu apa yang perlu dilakukan dan mencari tahu kapan sistem akan di implementasi.
2. *Design* (Perancangan). *Design* merupakan tahap selanjutnya, dimana kegiatan pemodelan seperti pemodelan sistem dan pemodelan arsitektur dilakukan. Diagram dalam *Unified Modeling Language* (UML) digunakan dalam pemodelan sistem dan arsitektur.
3. *Coding* (Pengkodean). *Coding* merupakan tahap penerapan pemodelan untuk membuat antarmuka pengguna bersama dengan logika menggunakan bahasa pemrograman.
4. *Testing* (Pengujian). Setelah tahapan pengkodean selesai, tahap pengujian sistem dilakukan untuk mengidentifikasi kesalahan yang terjadi selama aplikasi berjalan dan untuk menentukan apakah sistem yang dibangun memenuhi kebutuhan pengguna.

2.2.7. *Black Box Testing*

Pengujian yang berfokus pada spesifikasi fungsional perangkat lunak dikenal juga sebagai *Black-Box Testing*. Penguji dapat menentukan rangkaian kondisi masukan dan menjalankan pengujian berbasis spesifikasi program fungsional (Hidayat dan Muttaqin, 2019). Pengujian Black-box biasanya mengidentifikasi berbagai masalah, termasuk kesalahan inisialisasi dan terminasi, kesalahan struktur data, kesalahan antarmuka, masalah kinerja program, dan kebenaran atau kesalahan fungsional (Pratama dan Junianto, 2015). Metode *Equivalence Partitions* adalah salah satu dari beberapa pendekatan pengujian metode *Black Box Testing*. *Equivalence Partitions* adalah pengujian yang menggunakan *input* data pada setiap formulir dalam sistem pemilihan penjualan terbaik. Dalam pengujian ini, setiap menu masukan akan diuji dan dikelompokkan sesuai dengan cara kerjanya, apakah sesuai dengan yang telah ditetapkan atau tidak (Hidayat dan Muttaqin, 2019). Menurut Jaya (2018), Keuntungan menggunakan metode black-box testing adalah:

1. Penguji tidak harus mengetahui bahasa pemrograman tertentu,
2. Karena pengujian dilakukan dari sudut pandang pengguna, spesifikasi kebutuhan dibuat lebih jelas dari ambiguitas dan ketidakkonsistenan.

2.2.8. *HTML (Hypertext Markup Language)*

HTML, juga dikenal sebagai hypertext markup language, adalah suatu Bahasa dari *website* (WWW) yang digunakan untuk menyusun dan memformat dokumen agar dapat ditampilkan oleh aplikasi browser. Bahasa atau aturan untuk menampilkan teks, audio, video, dan gambar pada halaman website dikenal dengan HTML. Menggunakan *tag* yang telah dideklarasikan pada halaman *notepad*, HTML biasanya digunakan untuk membuat halaman web. *Tag* ini dapat menghubungkan *file* dan dokumen HTML satu sama lain (Batubara *et al*, 2021).

Menurut Arisanti (2014), Dokumen HTML memiliki tiga buah tag utama untuk membuat struktur dari dokumen HTML, yaitu :

1. <html> berfungsi untuk mendeklarasikan suatu dokumen HTML,
2. <Head> berfungsi untuk memberikan informasi tentang dokumen HTML,
3. body> berfungsi untuk menyimpan data atau informasi untuk digunakan dalam dokumen HTML.

2.2.9. Python

Guido van Rossum mengembangkan bahasa pemrograman Python pertama kali di Belanda pada tahun 1990. Bahasa pemrograman Python bersifat gratis, dapat dikembangkan oleh siapa saja, dan dapat digunakan tanpa lisensi. Ini juga dikenal sebagai *open source*, atau pemrograman sumber terbuka (Saptono *et al*, 2013).

Menurut Ginting *et al* (2020), Python adalah bahasa yang terkenal dan memiliki banyak keuntungan untuk membantu pemrograman berbasis objek dan dapat berjalan pada berbagai tahap kerangka kerja seperti Mac, Linux, dan Windows. Beberapa kelebihan dari Bahasa python adalah :

1. Pengembangan program dapat selesai lebih cepat dan dengan lebih sedikit baris kode,
2. Mendukung multi-platform,
3. Memiliki sistem untuk mengelola memori secara otomatis,
4. Python menggunakan pemrograman berbasis objek atau *OOD*.

2.2.10. UML (Unified Modeling Language)

Bahasa berbasis grafik atau gambar dikenal sebagai UML (Unified Modeling Language) yang digunakan untuk memvisualisasikan, menentukan, membuat, dan mendokumentasikan sistem pengembangan perangkat lunak berbasis OO (*Object-Oriented*). Konsep proses bisnis, deskripsi kelas dalam bahasa pemrograman

tertentu, skema *database*, dan komponen yang diperlukan untuk sistem perangkat lunak semuanya termasuk dalam standar UML untuk cetak biru sistem (Mubarak, 2019).

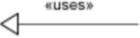
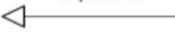
Menurut Ardian *et al* (2018), UML mendefinisikan diagram-diagram berikut:

a. *Use case diagram*

Use case diagram menampilkan fungsionalitas yang diharapkan dari sebuah sistem. Fokusnya bukan pada "bagaimana", tetapi "apa" yang dilakukan sistem.

Use case menunjukkan bagaimana aktor dan sistem berinteraksi satu sama lain.

Adapun notasi pada *use case diagram* dijelaskan pada gambar 2.

Simbol	Keterangan
	<i>Actor</i>
	<i>Use case</i>
	<i>Association</i>
	<i>Extends</i>
	<i>Uses (includes)</i>
	<i>Depends on</i>
	<i>Inheritance</i>

Gambar 2. Notasi *Usecase Diagram* (Ardian et al, 2018).

b. *Activity diagram*

Activity diagram mengilustrasikan berbagai aliran aktivitas sistem, titik awal dan akhirnya, dan pilihan yang dapat dibuat. Proses paralel yang dapat terjadi dalam beberapa proses juga dapat dijelaskan dengan menggunakan *activity diagram*.

Notasi pada *activity diagram* dijelaskan pada gambar 3.

Simbol	Keterangan
	<i>Activity</i>
	<i>Initiate Activities</i>
	<i>Start of the Process</i>
	<i>Termination of the Process</i>
	<i>Synchronization Bar</i>
	<i>Decision Activity</i>

Gambar 3. Notasi Pada *Activity* Diagram (Ardian et al, 2018).

2.2.11. Wireframe

Sebelum memulai proses perancangan dan pengkodean situs pada website, kerangka kerja disebut *wireframe* digunakan untuk menyesuaikan tata letak pada dengan desain antar muka. Proses *wireframe* dapat dilakukan secara manual atau dengan bantuan perangkat lunak khusus. *Header, body atau content, footer, sidebar*, dan sejumlah komponen lainnya adalah beberapa bahan yang biasa digunakan dalam proses pembuatan model *wireframe*. (Rudiana dan Sutisna, 2021).

2.2.12. Natural Language Processing

Natural language processing atau pemrosesan bahasa alami yang sering disingkat sebagai NLP, adalah kecakapan teknologi komputasi untuk memproses Bahasa alami manusia. Ilmu komputer, linguistik komputasi, dan kecerdasan buatan bergabung dalam bidang pemrosesan bahasa alami untuk mempelajari bagaimana komputer dan bahasa manusia berinteraksi satu sama lain. Pemrosesan otomatis bahasa alami yang diucapkan oleh manusia disebut sebagai pemrosesan bahasa alami, dan penerapannya meliputi pengembangan perangkat lunak web, analisis pasar dan tugas pemrosesan bahasa alami lainnya (Fahmi *et al*, 2020).

2.2.13. Text Mining

Text mining adalah salah satu teknik dari *data mining* yang bertujuan untuk menemukan pola-pola menarik dalam kumpulan besar data teks. *Text mining* adalah salah satu metode yang digunakan untuk melakukan tugas klasifikasi (Ratniasih *et al*, 2017).

Text mining merupakan proses penambangan dengan menggunakan data teks. Dalam *text mining*, sumber data diperoleh dari dokumen dengan maksud menemukan kata yang mewakili isi dokumen. *Text mining* adalah salah satu varian data mining yang bertujuan untuk menemukan pola-pola menarik dalam kumpulan besar data teks untuk melakukan analisis konektivitas antara satu dokumen dengan dokumen lain yang digunakan untuk mengklasifikasikan (Aditya, 2015).

2.2.14. Information Extraction

Information extraction atau ekstraksi informasi adalah proses menemukan informasi terstruktur dari teks tidak terstruktur atau semi terstruktur. Ekstraksi informasi adalah tugas penting dalam *text mining* dan telah dipelajari di berbagai bidang seperti *natural language processing*, *information retrieval* dan *web mining*. Dua tugas dasar ekstraksi informasi adalah *named entity recognition* dan *relation extraction*. *Named entity recognition* mengacu pada penemuan entitas bernama seperti orang, organisasi, dan tempat. Sedangkan *relation extraction* mengacu pada menemukan hubungan semantik antara entitas (Aggarwal dan Zai, 2015).

Menurut Maryanto (2017), secara umum data dibagi menjadi 2 kelompok yaitu :

1. Data terstruktur/*structured* data

Structured data atau data terstruktur merupakan kumpulan data dengan tipe, format, dan struktur data tertentu. Data transaksional. Data OLAP, RDBMS tradisional, file CSV, dan spreadsheet adalah semua sumber data yang memungkinkan.

2. Data tidak terstruktur/*unstructured* data

Unstructured data atau data tidak terstruktur adalah data tekstual yang belum memiliki format. Aplikasi-aplikasi pada internet, seperti data URL log, audio, video, blog, email, dan media sosial menghasilkan data yang tidak terstruktur.

2.2.15. Named Entity Recognition

Named entity recognition merupakan salah satu tugas ekstraksi informasi yang berpusat pada mengidentifikasi dan mengenali suatu entitas bernama seperti nama, lokasi, organisasi, hari, atau hal khusus seperti jenis tanaman, penyakit, maupun jenis-jenis obat. Klasifikasi entitas bernama mungkin berbeda tergantung pada tujuan NER itu sendiri (Saputro dan Hidayatullah, 2021).

Penggunaan *Named entity recognition* dipengaruhi oleh jenis Bahasa yang digunakan, karena Bahasa yang berbeda akan diperlakukan berbeda, sehingga algoritma yang akan digunakan dalam NER juga akan berbeda. NER sendiri memiliki dua tugas utama, yang pertama adalah mengidentifikasi entitas yang benar, dan yang kedua adalah mengklasifikasikan entitas tersebut ke dalam kategori entitas yang telah ditentukan. NER dapat dikategorikan menjadi 3 kelas, yaitu *Machine learning-based* NER, *namely rule-based* NER dan hybrid NER (Wu *et al*, 2006). NER Berbasis *rule-based* berfokus pada pencarian nama entitas menggunakan aturan buatan yang telah ditentukan sebelumnya dan terdiri dari serangkaian pola yang menggunakan tata bahasam sintaksis, dan ejaan. Kemudian NER berbasis *machine learning* menggunakan metode untuk menyelesaikannya dengan mengubah tugas identifikasi menjadi tugas klasifikasi menggunakan model klasifikasi statistik. Pada metode NER berbasis *machine learning*, menggunakan model statistik dan algoritma pembelajaran mesin, sistem membuat model dengan mencari pola dan hubungan dalam teks. Kemudian, NER hybrid menggunakan metode yang menggabungkan NER *rule-based* dan metode NER berbasis *machine learning*, dan menggunakan kekuatan masing-masing metode tersebut untuk membuat metode baru (Yanti *et al*, 2021).

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported ORG by F.B.I. Agent Peter Strzok PERSON ,
 Who Criticized Trump PERSON in Texts, Is FiredImagePeter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel
 investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. CreditT.J. Kirkpatrick PERSON for The New York
 TimesBy Adam Goldman ORG and Michael S. SchmidtAug PERSON . 13 CARDINAL , 2018WASHINGTON CARDINAL — Peter Strzok
 PERSON , the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped
 oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON 's lawyer
 said Monday DATE . Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer,
 Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate "witch hunt." Mr. Strzok PERSON , who rose over 20 years
 DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the
 inquiry. Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account. The
 F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON , who was removed last summer
 DATE from the staff of the special counsel, Robert S. Mueller III PERSON . The president has repeatedly denounced Mr. Strzok PERSON in posts on

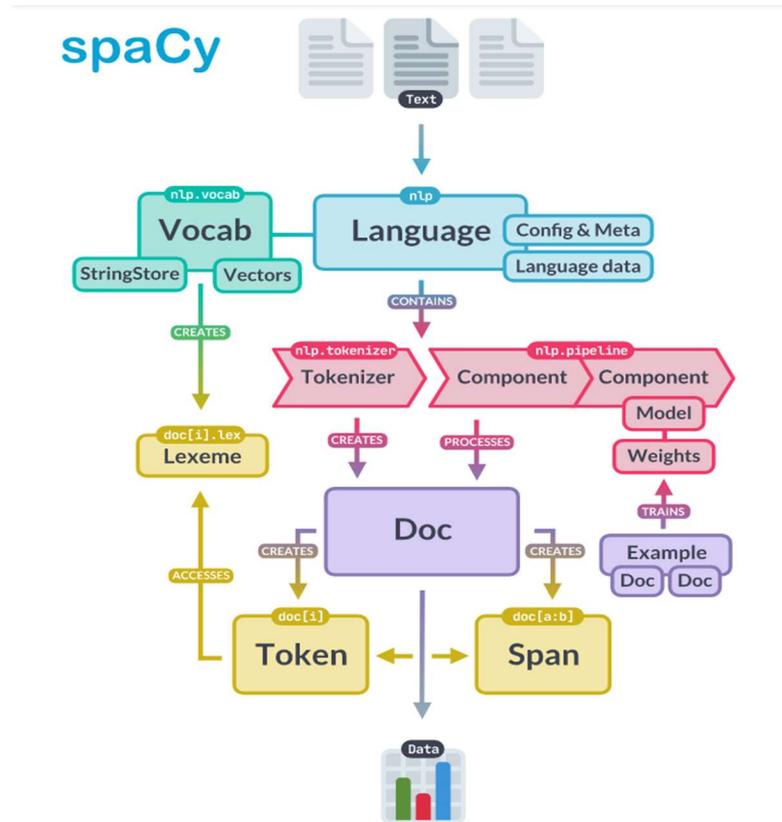
Gambar 4. Contoh *Named Entity Recognition* (Brown , 2020).

2.2.16. SpaCy

SpaCy adalah *open source library* yang dimiliki oleh *Python* untuk *natural language processing*. SpaCy dapat melakukan tugas seperti *POS tagging*, *Named Entity Recognition*, *Dependency Parsing*, dan sebagainya. Setiap keputusan yang dibuat oleh spaCy adalah prediksi, dan modelnya bersifat statistik. *Library* spaCy tidak menawarkan *pre-trained model* untuk Bahasa Indonesia, tetapi memberikan kesempatan untuk melakukan pelatihan dan mendapatkan model sendiri. (Yanti et al, 2021).

Library spaCy menyediakan berbagai alat untuk melakukan tugas *natural language processing* secara cepat dan dikembangkan menggunakan *pipeline* yang ada pada spaCy. Penggunaan *pipeline* pada spaCy dapat dibuat secara khusus untuk tugas-tugas tertentu sesuai kebutuhan. Komponen *named entity recognition* pada spaCy menggunakan *long short term memory* (LSTM) dan *Convolutional neural network* (CNN) untuk arsitektur modelnya. SpaCy menyediakan model NER yang dapat di kustomisasi entitasnya dengan cara memulai model kelas dengan kelas kosong, yang tidak memiliki entitas yang sudah dilatih, atau bisa juga menggunakan model yang sudah tersedia pada spaCy (Tarcar et al, 2020).

Struktur data pada spaCy adalah *language class*, *vocab*, dan *doc object*. *Language class* digunakan untuk memproses teks dan mengubahnya menjadi *doc object*, *doc object* biasanya disimpan sebagai variabel yang disebut *nlp*. *Doc object* memiliki urutan token dan semua anotasinya, *doc object* dibangun oleh *tokenizer*, dan kemudian dimodifikasi oleh komponen *pipeline*. Dengan memusatkan *string*, *word vector*, dan *lexical attributes* pada *vocab*, dapat menghindari dari salinan data dengan penyimpanan yang banyak, itu juga dapat menghemat memori (SpaCy, 2022). Berikut ilustrasi dari model arsitektur spaCy.



Gambar 5. Arsitektur spaCy (SpaCy, 2022).

2.2.17. *Multi Class Classification*

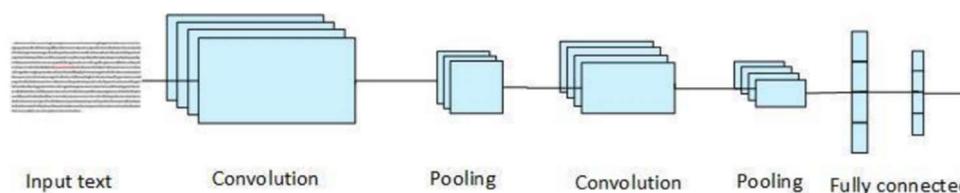
Klasifikasi adalah teknis penambangan data sehingga menentukan kelas ke satu set data untuk membantu prediksi dan analisis. Klasifikasi merupakan fungsi untuk mengekstrak data dalam suatu kelompok ke kelas dasar atau kelompok. *Multi class classification* yaitu mengklasifikasi data menjadi beberapa kelompok yang memiliki lebih dari dua kelas dalam variabel target (Mutlag et al, 2019).

Tabel 2. Contoh *Multi Class Classification*

Nama	Tempat	Organisasi	Tanggal
Susilo	Pasar Pagi	PSSI	28/03/2022
Bambang	Terminal Rajabasa	KPK	20 Febuari 2021
Yudoyono	Gudang Pasar	Himpunan Mahasiswa Jurusan Ilmu Komputer	28-Oktober-2020

2.2.18. *Convolutional Neural Network*

Convolutional neural network atau biasa dikenal dengan CNN adalah salah satu tipe jaringan syaraf tiruan yang mempunyai lapisan konvolusi yang berfungsi untuk mengekstraksi fitur untuk mendapatkan informasi sebanyak mungkin., dimana proses tersebut berfungsi untuk memperoleh informasi secara maksimal. Pemrosesan bahasa lami atau NLP adalah salah satu pengaplikasian umum untuk CNN, meskipun digunakan secara luas untuk pemrosesan data gambar. Cnn juga dapat digunakan untuk teks yaitu menggunakan CNN satu dimensi(Kurniawan et al, 2021). ilustrasi arsitektur CNN 1 dimensi ditunjukkan pada gambar 6 .



Gambar 6. Arsitektur CNN (Yahia dan Abdulazeez, 2021).

Menurut Maulana dan Rochmawati (2020), CNN pada umumnya terdiri dari berbagai lapisan penyusun, lapisan tersebut saling terhubung untuk memproses data yang di *input*, lapisan-lapisan tersebut adalah *convolution layer*, *pooling layer*, dan *fully connected*.

- *Convolution layer*

Fitur dan ciri utama dari algoritma CNN adalah *convolution layer*. Untuk mendapatkan *feature map* atau *activation map*, *Convolution layer* akan melakukan konvolusi terhadap data masukan atau data *output* dari lapisan sebelumnya dengan menggeser sebuah filter.

- *Pooling layer*

Salah satu cara untuk mereduksi atau meringkas *feature map* atau input data yang masih cukup besar menjadi ukuran yang lebih kecil adalah melalui *pooling layer*. Terdapat 2 jenis *layer pooling* yang dapat dilakukan, yaitu *average pooling*, dan *max pooling*. *Average pooling* akan menghitung nilai rata-rata setiap *feature map*

pada setiap pergeseran *window pooling*, sedangkan *max pooling* akan meringkas *input* data atau *feature map* dengan cara mencari nilai terbesar dari *feature map* berdasarkan pergeseran *window pooling*.

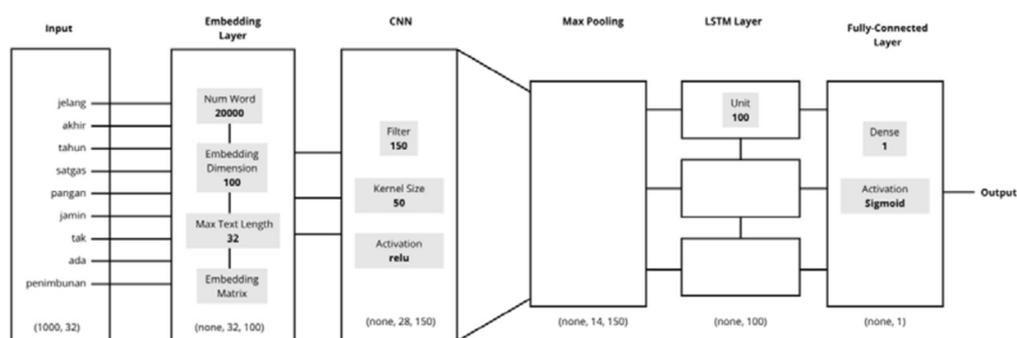
- *Fully connected*

Fitur-fitur yang telah dibuat pada langkah sebelumnya kemudian akan diklasifikasikan. Prosedur akan dilakukan dengan pada layer dengan semua koneksi. Proses akan diproses di lapisan *fully connected*. Pada lapisan *fully connected* ada beberapa lapisan yaitu lapisan *input*, lapisan *output* yang mirip pada *multi layer perceptron* (MLP), dan lapisan tersembunyi setiap lapisan yang terdapat pada lapisan *fully connected* memiliki *neuron* yang saling terhubung pada setiap lapisan yang terhubung satu sama lain. *Neuron* ini memiliki bobot yang membantu mengelompokan fitur ke dalam kategori yang sesuai.

Menurut Hermanto et al (2021), ada dua macam kombinasi antara gabungan dua metode CNN dan LSTM, yaitu CNN-LSTM, dan LSTM-CNN.

- CNN-LSTM

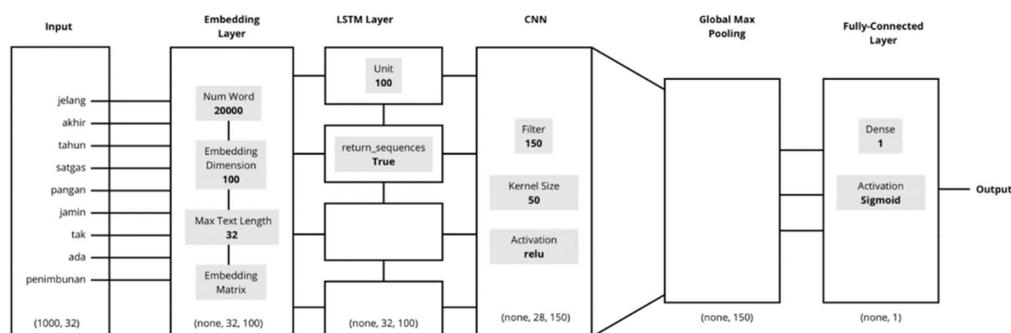
Lapisan *convolutional* awal model CNN-LSTM akan menerima embeddings kata untuk token yang dimasukkan. *Output* lapisan CNN akan digabungkan menjadi dimensi yang lebih kecil sebelum ditransfer ke lapisan LSTM, dimana diharapkan fitur lokal akan diekstraks fiturnya. Arsitektur CNN-LSTM dapat dilihat pada gambar 7.



Gambar 7. Arsitektur Model CNN-LSTM (Hermanto et al, 2021).

- LSTM-CNN

Model LSTM-CNN terdiri dari lapisan LSTM awal yang akan menerima *embeddings* kata untuk setiap token yang dimasukkan. Intuisi merupakan token *output* akan menyimpan data tidak hanya dari token awal, tapi juga dari token sebelumnya. Lapisan LSTM menghasilkan kode baru untuk input asli. *Output* dari *layer* LSTM kemudian dilanjutkan kedalam lapisan *convolution* yang nantinya akan mengekstrak fitur lokal. Akhirnya *output* lapisan *convolution* akan disatukan kedalam kelompok dimensi yang lebih kecil. Arsitektur model LSTM-CNN dapat dilihat pada gambar 8.



Gambar 8. Arsitektur Model LSTM-CNN (Hermanto et al, 2021).

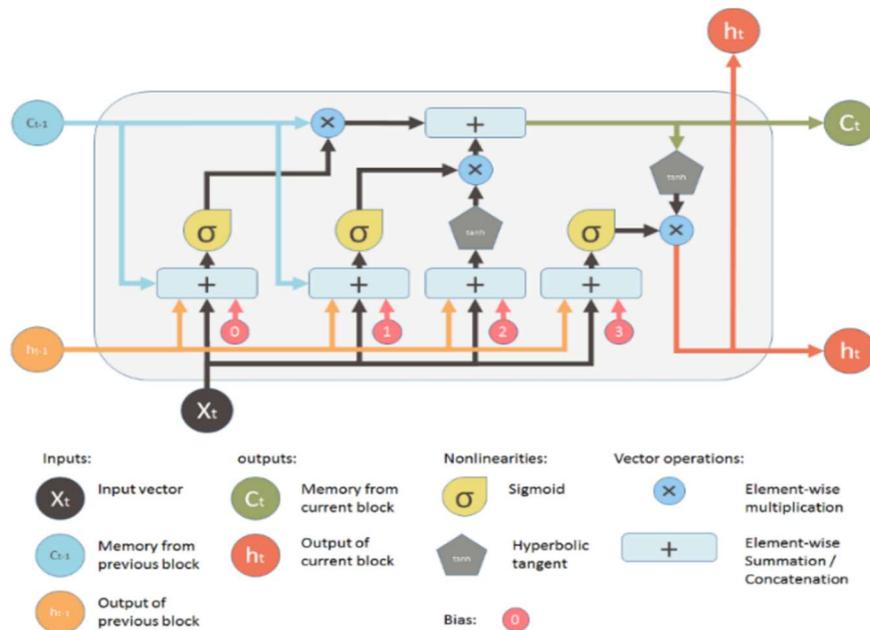
2.2.19. Long Short Term Memory

Menurut Hanifa et al (2021), LSTM merupakan salah satu pengembangan dari algoritma RNN atau *recurrent neural network* dengan menambahkan *memory cell* yang dapat menyimpan sebuah informasi dalam jangka Panjang. Tujuan LSTM dibentuk ialah agar dapat menangani persoalan *vanishing gradient* pada RNN saat memproses data *sequential* yang cukup Panjang, persoalan tersebut menyebabkan RNN tidak dapat menangkap *long term dependencies* sehingga akurasi yang didapatkan pada prediksi menggunakan metode RNN berkurang . LSTM memiliki 3 *gate* yang memiliki dimensi yang serupa dengan *dimension hidden state* sebagai ukuran pada *hidden state* 3 *gate* diantaranya yaitu

- *input gate (it)* berfungsi untuk memutuskan nilai mana saja yang diperbarui.

- *forget gate* (f_t) berfungsi untuk menentukan informasi apa saja yang akan dihapus.
- *output gate* (o_t) berfungsi untuk menentukan bagian mana dari konteks yang akan didapatkan.

Menurut Rizki et all (2020), rumus LSTM dibagi menjadi 4, seperti pada



Gambar 9. Ilustrasi LSTM (Rizki et all, 2020)

persamaan berikut:

- *Forget gate*

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f)$$

- *Input gate*

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i)$$

- *Memory update*

$$c_t = f_t \circ c_{t-1} + i_t \circ \phi(W_{cx}x_t + W_{ch}h_{t-1} + b_c)$$

- *Output gate*

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o)$$

$$h_t = o_t \circ \phi(c_t)$$

2.2.20. Hyperparameter

Hyperparameter merupakan suatu konfigurasi yang berada diluar model. Hyperparameter membantu dalam menenmukan parameter model yang tidak bergantung pada data training (Hutter, Lucke, & Schmidt-Thieme, 2015). Hyperparameter yang biasa digunakan yaitu *epoch*, *optimizer*, *batch size*, dan *learning rate*.

A. Epoch

Epoch Merupakan Hyperparameter yang menentukan berapa kali proses yang akan dilakukan pada tahap training/pelatihan(Wibawa, 2017).

B. Optimizer Adam

Oprimizer adam Adalah Algoritma Stokastik berdasarkan perkiraan adaptif dari order rendah. *Optimizer* adam cocok digunakan pada permasalahan data dengan gradient yang menyebar (Wibawa, 2017).

C. Batch Size

Batch size adalah jumlah training sample yang dipakai untuk suatu iterasi. *Batch size* digunakan dalam proses training untuk menentukan berapa jumlah data yang digunakan untuk training (Rochmawati et al, 2021).

D. Learning Rate

Learning rate merupakan salah satu *Hyperparameter training* yang diatur untuk menghitung nilai pembobotan/nilai koreksi bobot pada proses *training*(Rochmawati et al, 2021).

III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat

3.1.1. Waktu

Penelitian dilaksanakan dari semester genap sampai semester ganjil tahun ajaran 2022/2023. Penjelasan detail kegiatan dapat dilihat pada Tabel 3 berikut.

3.1.2. Tempat

Penelitian dilaksanakan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan juga kediaman peneliti.

3.2. Data dan alat

3.2.1. Data

Dalam penelitian ini data yang digunakan adalah data dari artikel dalam portal berita berbahasa Indonesia yang nantinya akan diterjemahkan setiap katanya menjadi Bahasa Lampung saat kata pada artikel tersebut sudah dikelompokkan. Data didapatkan dari hasil proses *scrapping* berita pada portal berita *online*. Pada penelitian data didapatkan dari situs berita *online* yang mencakup topik pada bidang pertanian. Data didapatkan dari beberapa situs berita *online* seperti Kompas.com, Republika.co.id, Pertanian.go.id, dan beberapa situs berita online yang berisi topik tentang bidang pertanian. Setelah data didapatkan, data akan dilakukan proses seperti *case folding*, dan *tagging*. Baru setelah itu data akan dilatih dan disimpan menjadi sebuah model NER Spacy dan dievaluasi dengan fungsi *evaluate* pada *library spaCy*. Data portal berita itu sendiri nantinya akan dibagi atau dikelompokkan menjadi 9 entitas. Setiap token pada data yang didapatkan akan dilakukan proses pelabelan secara manual menjadi 9 kelas. Berikut penjelasan 9 kelas dijelaskan pada tabel 4.

Tabel 4. Kelompok Entitas

Kelompok/Entitas	Deskripsi	Contoh
Nama_orang	Merupakan entitas yang menunjukkan nama seseorang yang terdapat pada berita.	Muhammad Azriel
Hari	Merupakan entitas yang menunjukkan kata hari yang terdapat pada berita.	Selasa
Tanggal	Merupakan entitas yang menunjukkan tanggal yang terdapat pada berita.	19-03-2022

Tabel 4. (lanjutan)

Lokasi	Merupakan entitas yang menunjukkan lokasi yang tertera pada isi berita.	Bandung, Jawa Barat
Organisasi	Merupakan entitas yang menunjukkan nama organisasi yang terdapat pada berita.	Kementrian Pertanian Republik Indonesia
Musim	Merupakan entitas yang menunjukkan jenis musim yang tertera pada isi berita.	Musim Hujan
Produk Pertanian	Merupakan entitas yang menunjukkan contoh dari produk pertanian yang terdapat pada isi berita.	Tomat, Cabai, dan Jagung
Penyakit	Merupakan entitas yang menunjukkan jenis penyakit, wabah, ataupun hama yang tertera pada isi berita.	Jamur tepung, Tikus padi
Harga	Merupakan entitas yang menjelaskan mengenai jumlah nominal harga produk pertanian yang tertera pada isi berita.	Rp 100.000,00

3.2.2. Alat

3.2.2.1. Perangkat Keras

Perangkat keras yang digunakan dalam penelitian adalah:

1. System Manufacturer : Acer,
2. System model : Nitro AN515-52,
3. Processor : Intel Core i7-8750H,
4. RAM : 16 GB

3.2.2.2. Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian adalah:

1. Sistem Operasi Windows 10,
2. Visual Studio Code,
3. StarUML,
4. Anaconda3,
5. Jupyter Notebook,
6. Browser Microsoft Edge.
7. *Packages* Python:

- a. *Library* Request

Library Request berfungsi untuk membuat permintaan HTTP seperti GET, POST, dan lain sebagainya. Request HTML digunakan untuk mengumpulkan data dari halaman website dengan mengirimkan *request* ke *server* situs web.

- b. *Library* HTML5Lib

Library HTML5Lib digunakan untuk mem-parsing data dari HTML.

- c. *Library* bs4(BeautifulSoup)

Library BeautifulSoup digunakan untuk menarik/mengambil data dari file HTML dan XML.

- d. *Library* TQDM

Library tqdm digunakan untuk menampilkan sebuah progres bar untuk mengetahui proses yang dilakukan

e. *Library* JSON

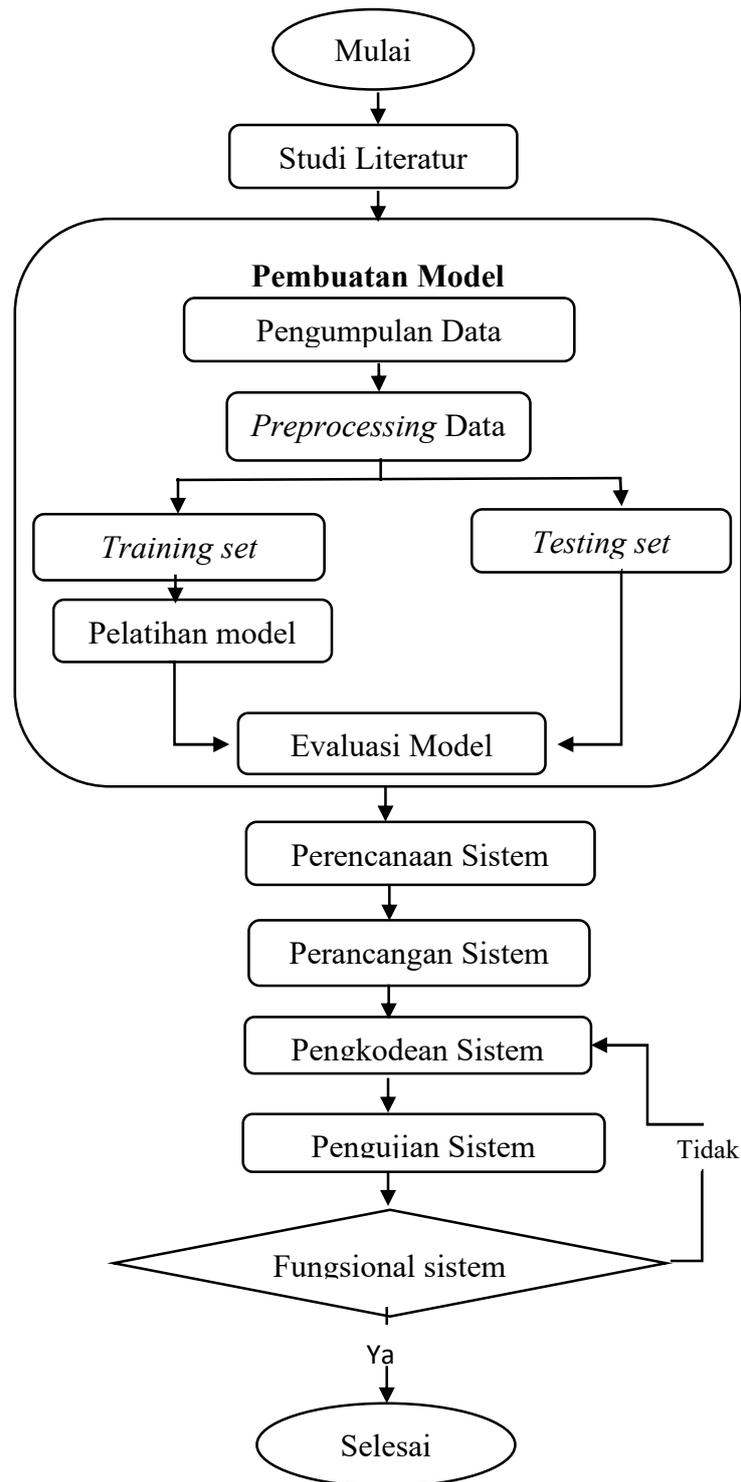
Library JSON digunakan untuk mengelola data yang berekstensi `.json`

f. *Library* SpaCy

Library SpaCy digunakan untuk membantu proses pemrosesan bahasa alami, *Library* SpaCy digunakan juga untuk melakukan pengaturan pada model dan menampilkan hasil ekstraksi informasi.

3.3. Alur Kerja Penelitian

Penelitian ini dilakukan beberapa tahapan seperti studi literatur, pembuatan model, perencanaan sistem, perancangan sistem, dan pengujian sistem. Pada pembuatan model dilakukan pengumpulan data, *preprocessing* data, data dibagi menjadi dua bagian yaitu data set untuk *training* data, dan juga *testing* data, lalu *training* data dilatih untuk mendapatkan model yang terbaik. Setelah data selesai dilatih, data dilakukan evaluasi dengan data set *testing* data untuk evaluasinya. Alur penelitian bisa dilihat pada gambar 10.



Gambar 10. Diagram Alur Penelitian

3.3.1. Studi Literatur

Studi literatur dilaksanakan dengan cara mengumpulkan data yang diperlukan oleh penelitian dan sistem. Data tadi dihasilkan berdasarkan aneka macam sumber, yaitu buku, internet, dan diskusi dengan nakan dosen pembimbing.

3.3.2. Pembuatan Model

Pembuatan model untuk penelitian ini meliputi beberapa tahap, yaitu:

3.3.2.1. Pengumpulan Data

Data dikumpulkan dengan Teknik *web scrapping* di portal berita *online*. *Web scrapping* adalah Teknik yang digunakan untuk mengesktrak sejumlah data yang besar dari situs web di mana data diekstraksi dan disimpan ke dalam komputer. Dalam melakukan *scrapping* digunakan *library* BeautifulSoup, menggunakan *request library* untuk mengirim *request* ke situs web dan mendapatkan respons untuk mendapatkan konten HTML dan meneruskannya ke objek BeautifulSoup untuk *parsing*. Fungsi *library* BeautifulSoup yaitu mengambil data *string* HTML dan menghasilkan struktur dokumen HTML untuk memanipulasi konten teks melakukan pencarian, proses navigasi, dan juga modifikasi struktur data situs website yang di *scrapping*.

Setelah data didapatkan dari sebuah web yang berupa string, data yang berupa teks berbahasa Indonesia di terjemahkan menjadi bahasa lampung berdialek Api.

3.3.2.2. Preprocessing data

Preprocessing data adalah proses mengubah data menjadi sesuatu yang bisa dipahami komputer. Dataset yang didapatkan akan melakukan beberapa tahapan *preprocessing*, yaitu:

a. Case folding

Case folding merupakan proses penghapusan perbedaan huruf besar yang terdapat dalam *string*. Metode ini digunakan guna melakukan perbandingan *case-sensitive* antara 2 *string*. Proses ini dilakukan dengan mengubah setiap huruf kapital menjadi huruf kecil.

b. Tagging

Tagging merupakan proses pemberian penanda atau label pada setiap kata. Kategori yang ada pada penelitian ini meliputi 'nama', 'tanggal', 'hari', 'lokasi', 'penyakit', 'produkpertanian', 'musim', 'organisasi', 'harga'. Pelabelan pada penelitian ini menggunakan bantuan website *annotator*/pelabelan <https://tecoholic.github.io/ner-annotator/> untuk mempermudah proses pelabelan pada berita.

c. Convert to spaCy format

Setelah data selesai diberi label, data diubah format dari sebelumnya berupa berkas berekstensi .json menjadi *doc object* yang nantinya dapat dipahami oleh *library spaCy*.

3.3.2.3. Pembagian data menjadi *training set* dan *testing set*

Data yang sudah dilakukan *preprocessing* dibagi menjadi 2 bagian, yaitu untuk *training set* dan *testing set*. Untuk penelitian ini dilakukan dua skenario pembagian data yaitu,

1. 70% *training set*, 30% *testing set*
2. 80% *training set*, 20% *testing set*

3.3.2.4. Pelatihan Data

Dilakukan pelatihan data terhadap data yang sudah dikumpulkan dengan skenario-skenario yang ditentukan dengan melihat skenario pada penelitian-penelitian sebelumnya. Perbedaan setiap skenario terdapat pada optimasi *hyperparameter* yang diperlukan untuk proses training data seperti *epoch*, *learning rate*, *batch size*. Skenario pada pelatihan data dapat dilihat pada tabel 5.

Tabel 5. Skenario Pelatihan Data

Skenario	Pembagian data	<i>Hyperparameter</i>		
		<i>Epoch</i>	<i>Batch size</i>	<i>Learning rate</i>
1	70% <i>training set</i>	100	1000	0.0001
2		100	256	0.00005
3	30% <i>training set</i>	50	256	0.0001
4	80% <i>training set</i>	100	1000	0.0001
5		100	256	0.00005
6	20% <i>testing set</i>	50	256	0.0001

3.3.2.5. Evaluasi Model

Untuk mengukur kinerja dari model *named entity recognition* yang sudah dibuat. Pada penelitian ini dilakukan pengukuran tiga perhitungan umum pada evaluasi model, yaitu *precision*, *f1-score*, dan *recall*. Terdapat fungsi pada *library spaCy* yang dapat menghitung kinerja model *named entity recognition* yaitu fungsi *evaluate*. Fungsi tersebut dapat menghitung nilai *precision*, *f1-score*, dan *recall* dari keseluruhan data, dan setiap label masing-masing. Berikut merupakan persamaan perhitungan *precision*, *f1-score*, dan *recall*.

- $Precision = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F1-score = \frac{2*Precision*Recall}{Precision+Recall}$

- True Positive (TP) : Jumlah prediksi dimana pengklasifikasian dilakukan dengan benar memprediksi kelas positif sebagai positif.
- False Positive (FP) : Jumlah prediksi dimana pengklasifikasian salah memprediksi kelas negatif sebagai positif.
- False Negative (FN) : Jumlah Prediksi dimana pengklasifikasian salah memprediksi kelas positif sebagai negatif.

3.3.3. Perencanaan sistem

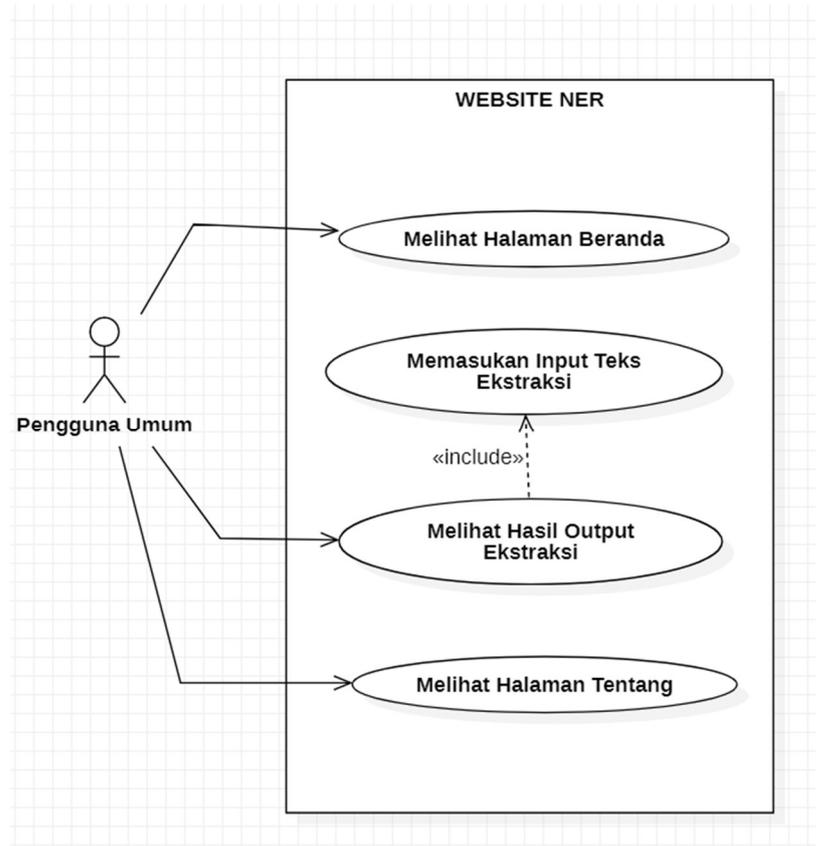
Perencanaan sistem dilakukan dengan cara menganalisis masalah dan kebutuhan apa-apa saja yang nantinya diperlukan untuk keperluan sistem informasi yang akan dibuat seperti, Bahasa yang akan digunakan, *framework* apa yang akan digunakan, dan apa saja fitur yang diperlukan, dan bagaimana sistem itu dapat berjalan

3.3.4. Perancangan sistem

Perancangan sistem adalah tahap pengembangan sistem yang dilakukan dengan membuat rancangan desain sistem. Tujuannya untuk memudahkan pengguna dalam memahami desain dari sistem yang dibuat. Berikut ini adalah berbagai desain diagram sistem dan desain antarmuka sistem.

1. Usecase diagram

Usecase diagram dari sistem *named entity recognition* ini terdapat pada gambar 11.



Gambar 11. Usecase Diagram Sistem

Tabel 6. Deskripsi Usecase Diagram

Aktor	Usecase	Deskripsi usecase
Pengguna Umum	Melihat halaman beranda	Usecase ini menampilkan halaman beranda yang berisikan nama dari website dan informasi singkat mengenai website

Tabel 6. (lanjutan)

Aktor	Usecase	Deskripsi usecase
Pengguna umum	Memasukan <i>Input</i> teks ekstraksi	<i>Usecase</i> ini menampilkan sebuah form yang akan diisi oleh aktor yang akan digunakan untuk proses ekstraksi
	Melihat hasil <i>output</i> ekstraksi	<i>Usecase</i> ini memampikan hasil dari proses ekstraksi informasi dari teks yang sudah diisi sebelumnya
	Melihat halaman tentang	<i>Usecase</i> ini menampilkan informasi singkat mengenai pengembang website

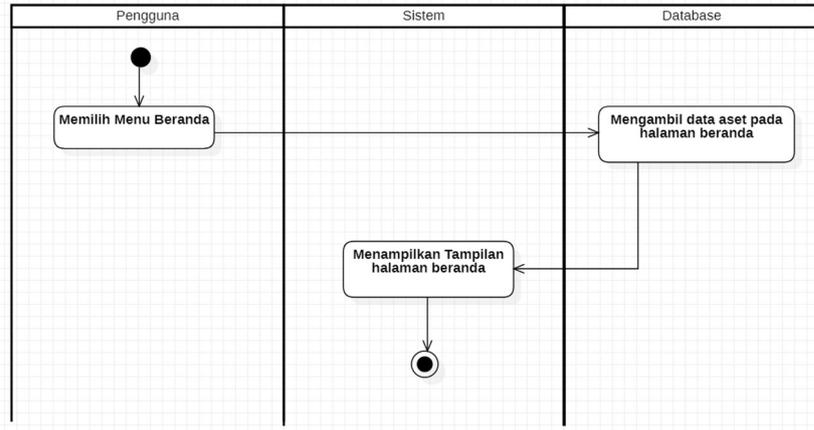
2. Activity diagram

Activity diagram digunakan guna menjelaskan jalannya masing-masing *usecase*.

Berikut *activity* diagram yang telah dibuat.

A. *Activity* diagram melihat halaman beranda

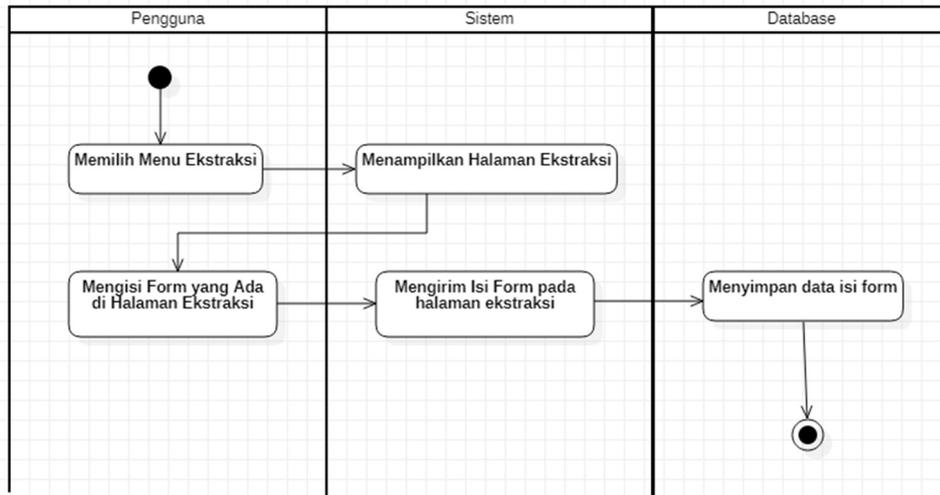
Activity diagram untuk melihat halaman beranda dapat dilihat pada gambar 12. Untuk mengakses menu beranda, pengguna harus memilih menu beranda pada sistem, lalu *database* akan mengirimkan aset yang diperlukan untuk halaman beranda.



Gambar 12. *Activity* Diagram Melihat Halaman Beranda

B. *Activity* diagram memasukan *input* teks ekstraksi

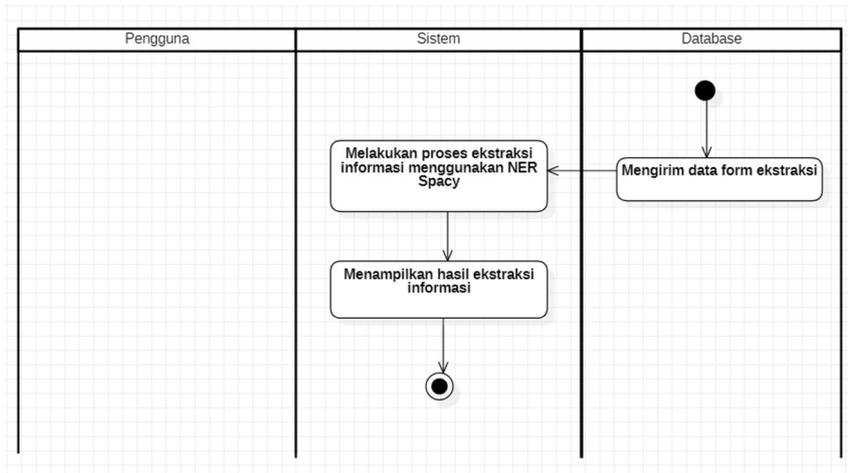
Pengguna yang ingin melakukan proses ekstraksi informasi harus memilih menu ekstraksi terlebih dahulu, setelah memilih menu ekstraksi akan muncul form yang harus diisi oleh pengguna, pengguna harus mengisikan form tersebut berupa teks yang ingin diekstraksi, setelah itu sistem akan menyimpan teks tersebut pada database. *Activity* diagram untuk memasukan *input* teks ekstraksi dapat dilihat pada gambar 13.



Gambar 13. Activity Diagram Memasukkan Input Teks Ekstraksi

C. Activity diagram melihat hasil output ekstraksi

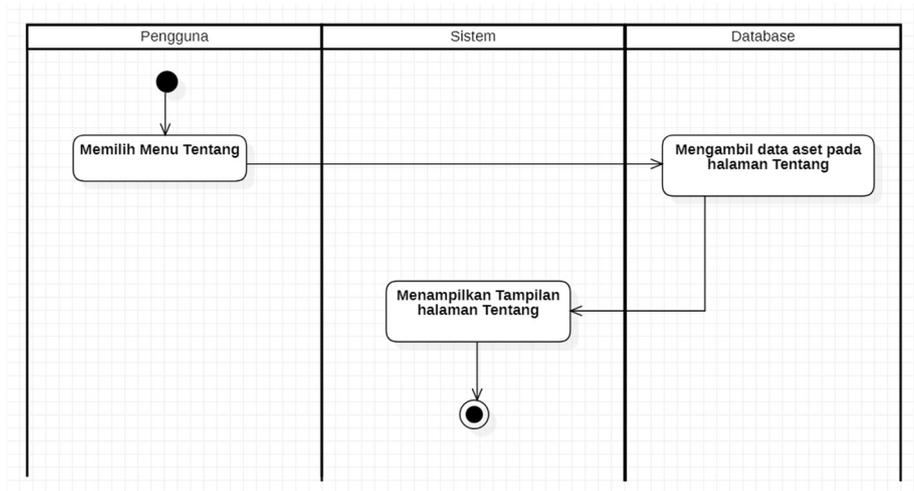
Activity diagram melihat hasil output ekstraksi dapat dilihat pada gambar 14. Data pada form yang sudah diisi sebelumnya langsung di proses oleh sistem untuk dilakukan ekstraksi informasi menggunakan NER, setelah dilakukan proses ekstraksi, sistem akan menampilkan hasil dari ekstraksi informasi tersebut.



Gambar 14. Activity Diagram Melihat Hasil Output Ekstraksi

D. *Activity* Diagram melihat halaman tentang

Untuk mengakses halaman tentang, pengguna harus memilih menu tentang yang ada pada sistem, lalu *database* akan mengirimkan aset yang diperlukan untuk halaman beranda. *Activity* diagram untuk melihat halaman beranda dapat dilihat pada gambar 15.

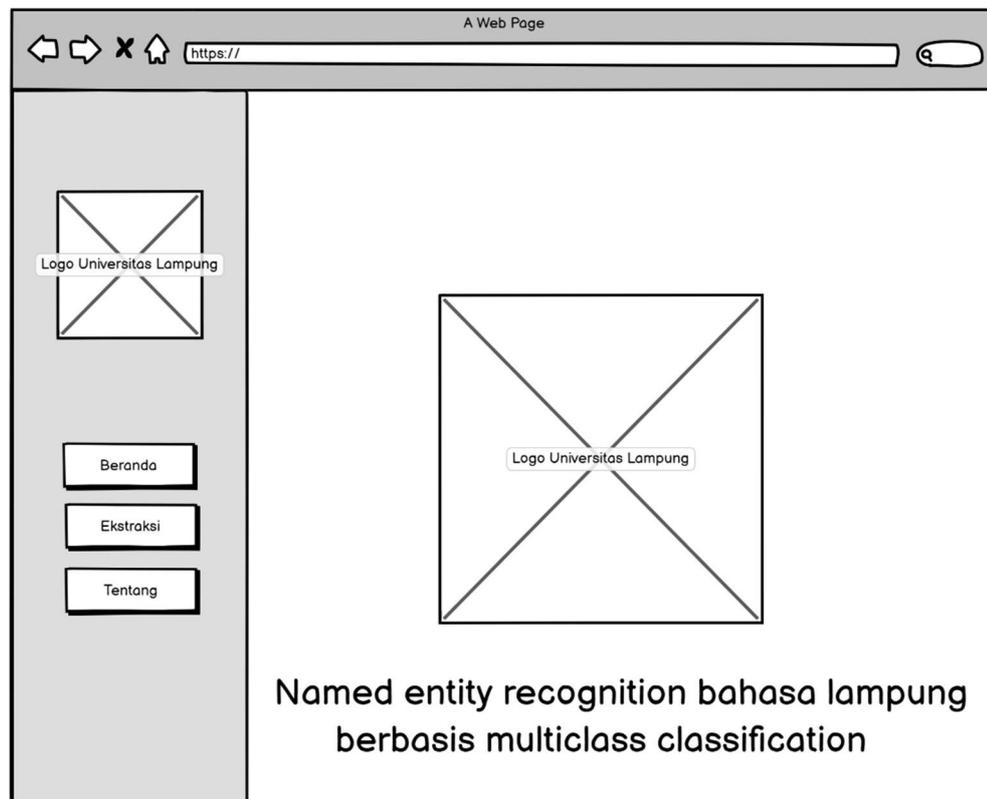


Gambar 15. *Activity* Diagram Melihat Halaman Tentang

3. Rancangan tampilan antarmuka sistem

A. Rancangan tampilan antarmuka sistem beranda

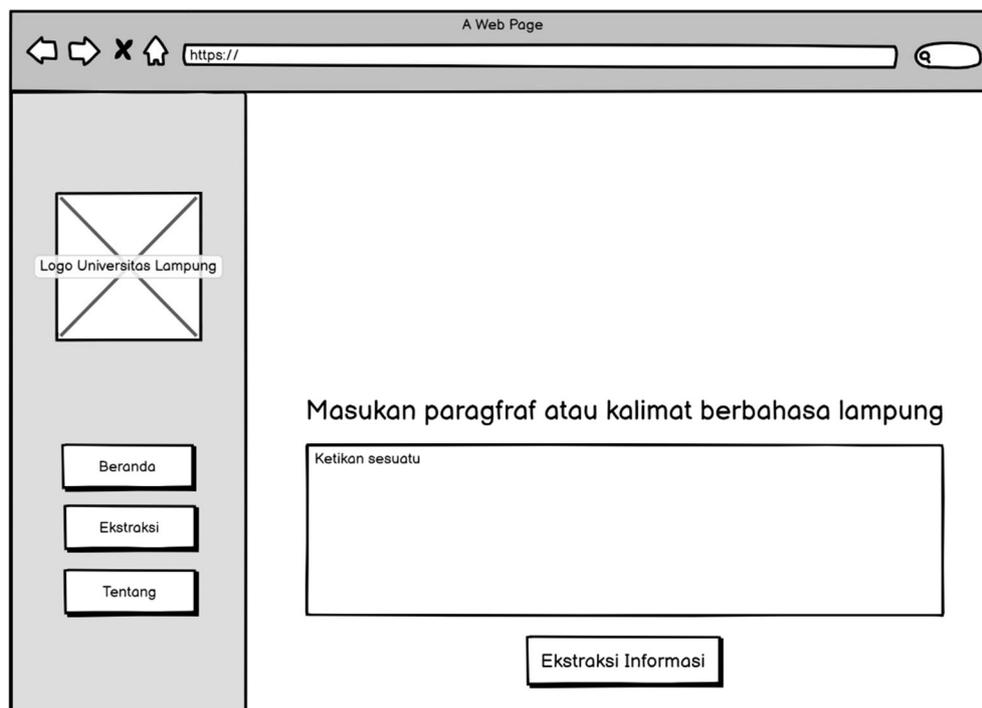
Rancangan tampilan antarmuka sistem beranda akan menampilkan logo universitas lampung dan juga judul dari sistem yang dibuat. Rancangan tampilan antarmuka sistem beranda ditunjukkan pada gambar 16.



Gambar 16. Rancangan Tampilan Antarmuka Sistem Beranda

B. Rancangan tampilan antarmuka sistem input teks ekstraksi

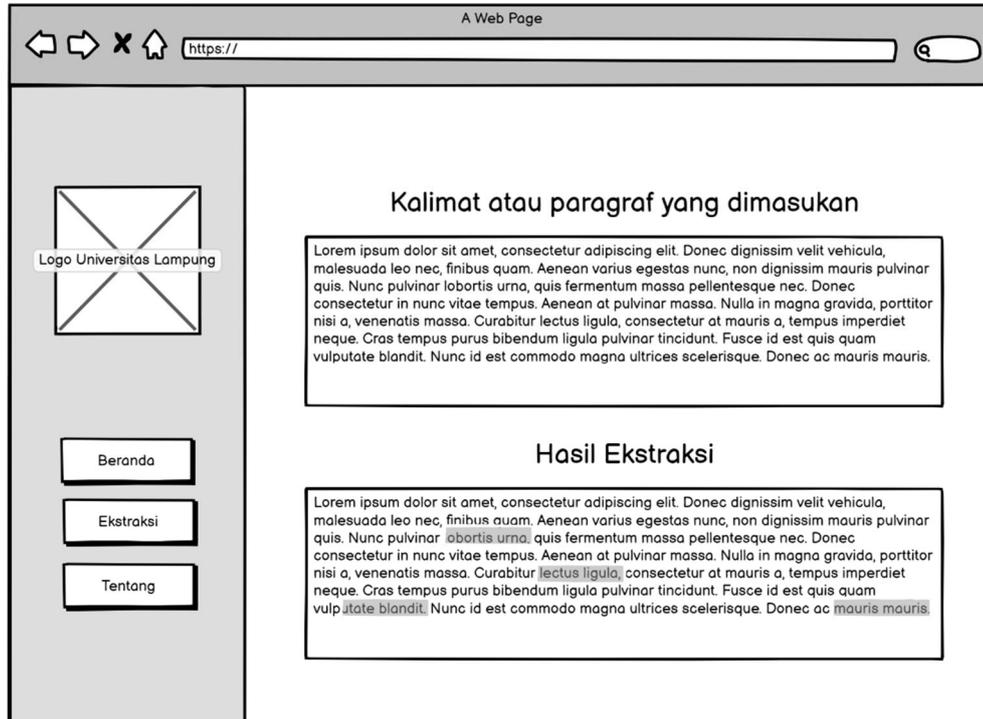
Rancangan tampilan antarmuka sistem input teks ekstraksi ditunjukkan pada gambar 17. Dalam halaman ini akan berisi form yang akan diisi oleh pengguna. Pengguna harus mengisi form dengan teks berupa kalimat atau paragraf yang akan di ekstraksi.



Gambar 17. Rancangan Tampilan Antarmuka Sistem Input Teks Ekstraksi

C. Rancangan tampilan antarmuka sistem hasil output teks ekstraksi

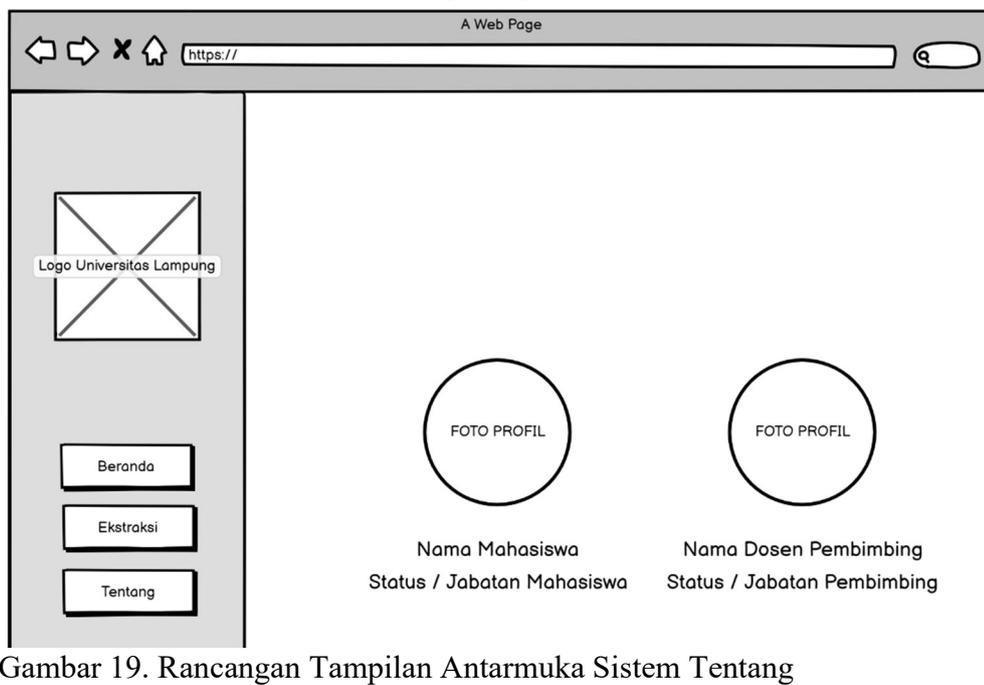
Rancangan tampilan antar muka sistem hasil output teks ekstraksi akan berisikan hasil dari proses ekstraksi yang sudah dilakukan. Tampilan dibagi menjadi dua bagian, tampilan pertama akan menampilkan teks yang *diinputkan* oleh pengguna, dan kedua akan menampilkan hasil dari ekstraksi informasi dari teks yang telah dimasukan sebelumnya. Rancangan tampilan antar muka sistem hasil output teks ekstraksi ditunjukkan pada gambar 18.



Gambar 18. Rancangan Tampilan Antarmuka Sistem Hasil Teks Ekstraksi

D. Rancangan tampilan antarmuka sistem tentang

Pada Rancangan tampilan antarmuka sistem tentang akan menampilkan profil singkat beserta foto mengenai tim pengembang sistem. Rancangan tampilan antarmuka sistem beranda ditunjukkan pada gambar 19.



Gambar 19. Rancangan Tampilan Antarmuka Sistem Tentang

3.3.5. Pengkodean sistem

Pengkodean sistem dilakukan menggunakan Bahasa PHP dan juga dengan bantuan framework flask. Pengkodean akan dilakukan sesuai kebutuhan sistem dan perancangan sistem yang ditentukan sebelumnya.

3.3.6. Pengujian sistem

Pengujian sistem dilakukan dengan melakukan pengujian secara fungsional dari fitur yang ada pada sistem tanpa harus mengetahui kode program yang terdapat didalamnya. Untuk menjalankan pengujian ini, pengguna harus menguji semua fitur yang ada dan memastikan bahwa semua fitur merespon seperti yang diharapkan oleh pengguna dan memberikan hasil. Pada penelitian sistem diuji menggunakan metode *black box testing*.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang dapat penulis ambil dari penelitian ini berdasarkan perancangan dan hasil implementasi yang dilakukan adalah sebagai berikut :

1. Telah dibuatnya sistem informasi yang dapat melakukan proses *Named Entity Recognition* dari input teks bahasa lampung dan menghasilkan informasi hasil ekstraksi. Sistem juga dapat mengklasifikasikan teks berdasarkan pelabelan entitas yang ditetapkan, dan berdasarkan hasil pengujian *Blackbox Testing*, Sistem yang dikembangkan mendapatkan hasil sesuai dengan yang diharapkan pada tiap skenario uji yang dilakukan.
2. Berdasarkan hasil evaluasi kinerja model yang dilakukan pada penelitian ini, skenario pembagian data 70% data *training*, dan 30% data *testing* dengan *hyperparameter epoch* 100, *batch size* 1000, dan *learn rate* 0.0001 mendapatkan hasil terbaik berdasarkan pengukuran *precision* 99.64%, *recall* 99.64%, dan *f1-score* 99.64%.

5.2. Saran

Berdasarkan perancangan dan hasil implementasi yang dilakukan, maka saran yang perlu diperhatikan dalam pengembangan lebih lanjut penelitian ini adalah sebagai berikut :

1. Perbanyak data yang digunakan agar tidak terjadi *unbalanced* data pada proses pembuatan model.
2. Data yang digunakan dapat menggunakan bahasa lampung yang berdialek nyo/ dialek O, atau juga bahasa daerah yang lainnya.

DAFTAR PUSTAKA

- Aditya, B. R. (2017). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, Vol. 7(2).
- Aggarwal, C., & Zai, C. (2015). *Mining Text Data*. New York: Springer.
- Agustina, E. S. (2015). Pemakaian Bahasa Lampung Di Daerah Rajabasa. *Lokabasa*, Vol. 6(1).
- Alsafy, B. M., Aydam, Z. M., & Mutlag, W. K. (2019). Multiclass Classification Methods : A Review. *International Journal of Advanced Engineering Technology and Innovative Science*, Vol. 5(3).
- Ardian, D., Larasati, D. P., & Irawan, A. (2018). Perancangan Sistem Aplikasi Rental Mobil Menggunakan Java Netbeans dan MySQL Pada Perusahaan Dean's Car Rent. *Applied Information System and Management*, Vol. 1(1).
- Arisanti, A. (2014). Perancangan Sistem Informasi Pendataan Penduduk Berbasis Web Menggunakan Metode Waterfall Pada Desa Bogorejo Kecamatan Gedongtataan. *Jurnal Manajemen Informatika dan Komputer*.
- Atika, D. (2021). Ekstraksi Informasi Berita Online Dengan Named Entity Recognition (NER) dan Rule-Based untuk Visualisasi Penyakit Tropid di Indonesia. *Jurnal Teknologi Informasi*.
- Batubara, N., Antoni, & Prayogi, S. Y. (2021). Perancangan Aplikasi Elearning Berbasis Web di SMP Negeri 1 Saipar Dolok Hole. *Jurnal Teknik Informatika*, Vol. 4(1).
- Bolyki, B., Polonkai, D., & Kovacs, D. (2022). Efficiency Analysis of NNS Methods. *Production System and Information Engginering*, Vol. 9.
- Brown, R. (2020, Feb 17). *How Does Named Entity Recognition Work: NER Methods?* Retrieved from Medium.com:
<https://medium.com/cogitotech/how-does-named-entity-recognition-work-ner-methods-f23201a69648>

- Carolina, I., & Supriyatna, A. (2019). Penerapan Metode Extreme Programming Dalam Perancangan Aplikasi Perhitungan Kuota SKS Mengajar Dosen. *Jurnal Komputer dan Informatika*, Vol. 3(1).
- Darmawan, A. C., & Iswari, L. (2022). Pengembangan Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi. *JURNAL PENDIDIKAN DAN KONSELING*, 5351-5362.
- Devianty, R. (2017). Bahasa Sebagai Cermin Kebudayaan. *Jurnal Tarbiyah*, Vol. 24(2).
- Fahmi, A., Ramadhan, I., & Agussalim. (2020). Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi COVID-19. *Jurnal Informatika dan Sistem Informasi*, Vol. 1(1).
- Fathoni, A., & Dwi, D. (2016). Rancang Bangun Sistem Extreme Programming Sebagai Metodologi Pengembangan Sistem. *Jurnal Prosisko*, Vol. 3(1).
- Ginting, V. S., Kusriani, & Luthfi, E. T. (2020). Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python. *Jurnal Teknologi Informasi*, Vol. 4(1).
- Hanifa, A., Fausan, S. A., Haikal, M., & Ashfiya, M. B. (2021). Perbandingan Metode LSTM dan GRU (RNN) Untuk Klasifikasi Berita Palsu Berbahasa Indonesia. *Jurnal Dinamika Rekayasa*, Vol. 7(1).
- Herliana, A., & Rasyid, P. M. (2016). Sistem Monitoring Pengembangan Software Pada Tahap Development Berbasis Web. *Jurnal Informatika*, Vol. III(1).
- Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online. *Citec Journal*, Vol 8(1).
- Hidayat, T., & Muttaqin, M. (2018). Pengujian Sistem Informasi Pendaftaran dan Pembayaran Wisuda Online menggunakan BlackBox Testing dengan Metode Equivalence Partitioning dan Boundary Value Analysis. *Jurnal Teknik Informatika UNIS*, Vol. 6(1).
- Huttler, F., Lucke, J., & Thieme, L. S. (2015). Beyond Manual Tuning of Hyperparameter. *CrossMark*, Vol.29.
- Irsyad, R. (2018). Penggunaan Python Web Framework Flask Untuk Pemula. *Jurnal Telematika*.
- Jaya, T. S. (2018). Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis (Studi Kasus: Kantor Digital Politeknik Negeri Lampung). *Jurnal Informatika : Jurnal Pengembangan IT*, Vol. 03(02).

- Kurniawan, M., Kusriani, K., & Arief, M. R. (2021). Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo. *Jurnal Eksplorasi Informatika*, Vol 11(1).
- Maryanto, B. (2017). Big Data dan Pemanfaatannya Dalam Berbagai Sektor. *Jurnal Media Informatika*, Vol. 16(2).
- Maulana, A. R., & Rochmawati, N. (2020). Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network. *JINACS*, Vol 02(01).
- Mubarak, A. (2019). Rancang Bangun Aplikasi Web Sekolah Menggunakan UML (Unified Modeling Language) dan Bahasa Pemrograman PHP (PHP Hypertext Preprocessor) Berorientasi Objek. *Jurnal Informatika dan Komputer*, Vol. 02(1).
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 5(5).
- Pratama, Y. A., & Junianto, E. (2015). Sistem Pakar Diagnosa Penyakit Ginjal dan Saluran Kemih Dengan Metode Breadth First Search. *Jurnal Informatika*, Vol. II(1).
- Pressman, PH.D., R. S. (2005). *Software Engineerinh : A Practitioner's Approach*. Ney York: McGraw-Hill.
- Putra, A. A., & Kurniawan, R. (2020). *Bidirectional LSTM-CNNs Untuk Ekstraksi Entity Lokasi Kebakaran Pada Berita Online Berbahasa Indonesia*. Jakarta: Politeknik STIS.
- Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017). Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat. *Jurnal Teknologi Elektro*, Vol. 16(3).
- Rizki, M., Basuki, S., & Azhar, Y. (2022). Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang. *Repositor*, Vol 2(3).
- Rochmawati, N., Hidayati, H. B., & Yamasari, Y. (2021). Analisa Learning rate dan Batch size Pada Klasifikasi Covid Menggunakan Deep learning dengan Optimizer Adam. *Journal Information Engineering and Educational Technology*, Vol.05(02).
- Rudiana, R., & Sutisna, M. A. (2021). Sistem Pengelolaan Administrasi Simpan Pinjam Berbasis Web Pada Koperasi Mitra. *Jurnal Visualika*, Vol.7(2).
- Saptono, D., Sampurna, M. T., R.N, T. W., & Fitrianiingsih. (2013). Implementasi Algoritma Gunning Fog Index Pada Uji Keterbacaan (Readabiliy Test)

Bahasa Indonesia Menggunakan Bahasa Pemrograman Python. *Jurnal Teknologi Informasi & Komunikasi Terapan*.

- Saputro, F. E., & Hidayatullah, A. F. (2021). Tinjauan Literatur : Named Entity Recognition pada Resep Makanan Indonesia. *Automata*, Vol. 2(2).
- 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*, Vol. 1(12).
- Setyawan, A. (2011, Juli 2). Bahasa Daerah Dalam Perspektif Kebudayaan dan Sosiolinguistik : Peran dan Pengaruhnya Dalam Pergeseran dan Pemertahanan Bahasa. *Language Maintance and Shift*, pp. 65-69.
- Sidh, R. (2013). Peranan Brainware Dalam Sistem Informasi Manajemen. *Jurnal Computech & Bisnis*, Vol. 7(1).
- SpaCy. (2022). *SpaCy*. Retrieved from SpaCy.io: SpaCy.io
- Tarcar, A. K., Tiwari, A., Rao, D., Dhaimodker, V. N., Rebelo, P., & Desai, R. (2020). Healthcare NER Model Using Language Model Pretraining Empirical Evaluation of Healthcare NER Model Performance with Limited Training Data. *HDSM*.
- Vignesh, Sivaprasath, Tamilalakan, Indhu, & Vijay. (2020). Text Summarization Using Spacy Algorithm. *International Journal of Future Generation Communication and Networking*, Vol.13(3).
- Wibawa, M. S. (2017). Pengaruh Fungsi Aktivasi, Optimisasi dan Jumlah Epoch Terhadap Performa Jaringan Saraf Tiruan. *JURNAL SISTEM DAN INFORMATIKA*, Vol.11(2).
- Wu, Y. C., Fan, T. K., Lee, Y. S., & Yen, S. J. (2006). *Extracting Named Entities Using Support Vector Machines*. Taoyuan: Springer.
- Yahia, H. S., & Abdulezeez, A. M. (2021). Medical Text Classification Based on Convolutional Neural Network : A Review. *International Journal of Science and Business*, Vol 5(27-41).
- Yanti, M. R., Santoso, I., & Suadaa, L. H. (2021). Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study: Power Failure in the Special Region of Yogyakarta). *Indonesian Journal of Information System*, Vol. 4(1).
- Yunius, Y. R. (2017). Implementasi Algoritma Convolutional Neural Network Dengan Framework Tensorflow Pada Aplikasi Mobile Pendeteksi Penyakit Melanoma Dengan Memanfaatkan Webservice Framework Flask. *Jurnal Sains dan Komputer*, Vol. 1(2).

