



MAKALAH PERBANDINGAN POS TAGGING NLP TRADISIONAL VS ERA LLM

Disusun guna memenuhi tugas mata kuliah Pemrosesan Bahasa Alami

Dosen Pengampu:

Rizal Setya Perdana, S.Kom., M.Kom., Ph.D.

Oleh:

Pieter Christy Yan Yudhistira

235150201111068

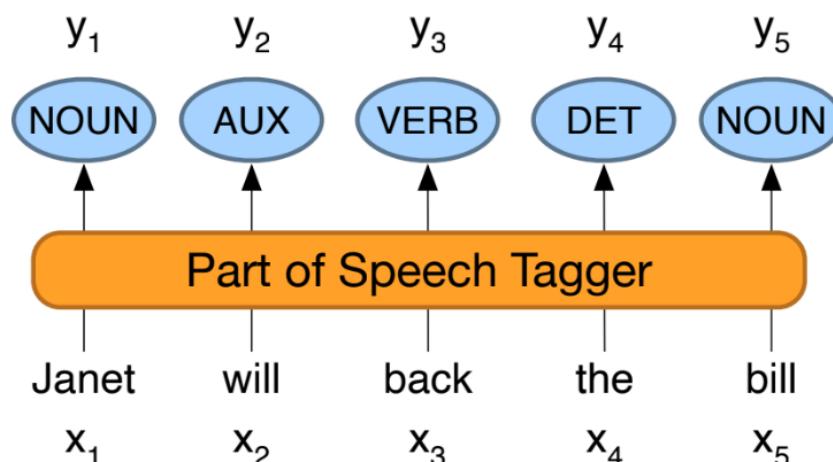
KELAS TIF-PBA-B
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
TAHUN 2025

BAB I

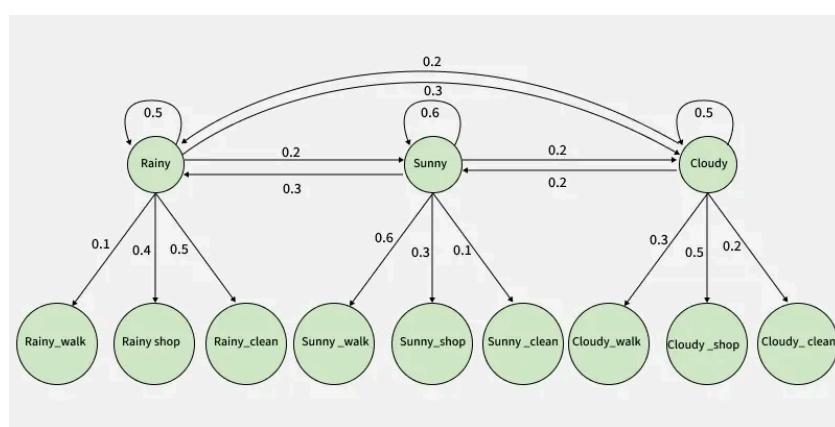
PENDAHULUAN

1.1 Latar Belakang

Part-of-Speech (POS) Tagging merupakan proses penentuan kategori bagian dari kata untuk keseluruhan kata dalam teks (Jurafsky, D., & Martin, J. H., 2025). Secara umum, kategori kata atau token kata dalam teks akan dikategorikan sebagai kata kerja, kata benda, kata sifat atau kata keterangan. Kategori katanya ditandakan dengan tag berupa NOUN, VERB, ADJ dan lain-lainnya. Tujuan dari melakukan POS Tagging sendiri adalah mengurangi ambiguitas kata jika dipisah-pisahkan secara token, misalnya kata *book* bisa diartikan sebagai kata kerja (*book a flight*) atau kata benda (*hand me that book*). Ambiguitas inilah yang diperjelas dengan membedakan *book* kata kerja dan kata benda.



Gambar 1. Ilustrasi POS Tagging (Jurafsky, D., & Martin, J. H., 2025)



Gambar 2. Ilustrasi POS Tagging dengan HMM (Hidden Markov Model)

Tradisionalnya, metode POS tagging menggunakan aturan linguistik, model statistik (HMM, CRF), atau model pembelajaran mesin berbasis fitur (SVM, MEMM) untuk melakukan klasifikasi kategori kata. Perkembangan Artificial Intelligence (AI) berupa Large

Language Models (LLMs) seperti GPT, BERT, dan varian autoregresif/transformer, adanya pendekatan dari POS Tagging berubah. POS Tagging menggunakan pendekatan LLM menjadi lebih banyak digunakan dan dikembangkan, baik secara few-shot/zero-shot, atau melalui fine-tuning/adapter untuk tugas terstruktur

1.2 Pembahasan Tugas

Tujuan dari penggerjaan dan eksplorasi tugas ini adalah membandingkan pendekatan POS Tagging tradisional dengan pendekatan berbasis Large Language Model (LLM) untuk memahami evolusi teknik NLP dan relevansinya saat ini. Tugasnya adalah melakukan perbandingan komprehensif antara metode POS Tagging tradisional (yang telah dipelajari) dengan pendekatan modern berbasis LLM, kemudian analisis kelebihan, kekurangan, dan aplikabilitas masing-masing. Tugas ini terdiri dari:

- Pendahuluan (background singkat POS Tagging dan bagaimana LLM mengubah landscape POS tagging)
- Buat tabel perbandingan dari berbagai metode tradisional vs LLM
- Berikan contoh Case Study POS Tagging
- Analisis dan kesimpulan

BAB II

HASIL EKSPLORASI

2.1 Metode Tradisional POS Tagging

POS Tagging memiliki beberapa jenis berdasarkan pendekatan kategorisasinya, yaitu rule based tagging, transformation based tagging dan statistical POS tagging. Metode yang umum adalah menggunakan POS Tagging dengan pendekatan statistik seperti model Hidden Markov Model (HMM) dan Conditional Random Fields (CRF). Pendekatan ini memanfaatkan model probabilistik untuk menetapkan kategori gramatis berdasarkan pola dari corpora menggunakan pembelajaran mesin.

2.2 POS Tagging Pendekatan Modern dengan LLM

Perkembangan Large Language Model (LLM) mendorong pendekatan pemrosesan teks dan performanya berpotensi mendahului metode tradisional. Anotasi POS Tagging saat ini paling umum memanfaatkan model Transformer atau LLM sebagai algoritma klasifikasi kategori suatu kata atau token. Model seperti BERT telah mampu melakukan pendekatan POS Tagging seperti Named Entity Recognition (NER) yang jauh lebih robust dibandingkan metode tradisional dikarenakan kemampuannya untuk memahami konteks dan belajar dari corpora lebih kompleks. Namun, pendekatan LLM tidak selalu meningkatkan performa atau hasil yang dikeluarkan dan tergantung konteks yang ada (Adelani, David Ifeoluwa, et al., 2024).

2.3 Metode Tradisional Vs LLM

Secara sederhana, metode POS Tagging tradisional dengan pendekatan LLM dapat dilihat dari ciri-cirinya. Berikut saya tampilkan perbedaan dari kedua jenis pendekatan POS Tagging.

Tabel 1. Perbedaan POS tagging pendekatan tradisional vs pendekatan modern dengan LLM

Aspek	Metode tradisional (HMM, CRF dkk)	Metode LLM (BERT, dkk)
Pendekatan	Memanfaatkan model probabilistik, feature engineering atau rule based linguistik untuk melakukan kategorisasi	Menggunakan deep contextual embeddings yang sudah belajar dari corpora besar dan umumnya menerapkan transformers.
Data & Label	Memerlukan fitur yang harus dibuat (feature engineering) misalnya menangani prefix/suffix, kapital, hingga tag yang	Tidak perlu pelabelan, mampu melakukan pemahaman berdasarkan konteks.

	bertetangga.	
Konteks	Terbatas (terkadang hanya fixed size context windows)	Mampu melihat konteks dan dependensi yang lebih jauh, dapat lebih menyesuaikan dengan atensi (attention) dari transformer.
Akurasi	~90 - 95% akurasi dengan evaluasi data bahasa Inggris.	Bisa mencapai lebih dari 97-98% akurasi dikarenakan robustnya model
Syarat pelatihan	Memerlukan dataset yang telah terlabel POS tagging, dengan ukuran kecil hingga sedang.	Umumnya pre trained dari kumpulan corpora yang besar, ada juga yang melakukan fine tuning menggunakan data berlabel POS juga
Computational Cost	Rendah, karena metode tradisional hanya memanfaatkan model probabilistik dan cukup efisien	Pelatihan butuh infrastruktur seperti GPU untuk mempercepat dan mempermudah pelatihan, terutama ketika menangani data yang sangat besar seperti corpora
Bahasa	Umumnya bahasanya spesifik, jika dipelajari bahasa Indonesia maka bahasa Inggris atau Malaysia akan mengalami penurunan.	Sekarang telah ada pendekatan multi bahasa, sehingga model mampu mengenali bahasa dan generalisasi antar bahasa.
Skalabilitas	Mudah untuk diimplementasi dalam sistem dan tidak membutuhkan cost yang sangat tinggi.	Sulit, terutama jika infrastruktur tidak mendukung waktu dan cost komputasinya.

Oleh karena itu, perbedaan ini dapat disimpulkan dimana metode tradisional

BAB III

ANALISIS DAN KESIMPULAN

3.1 Teks untuk Studi Kasus

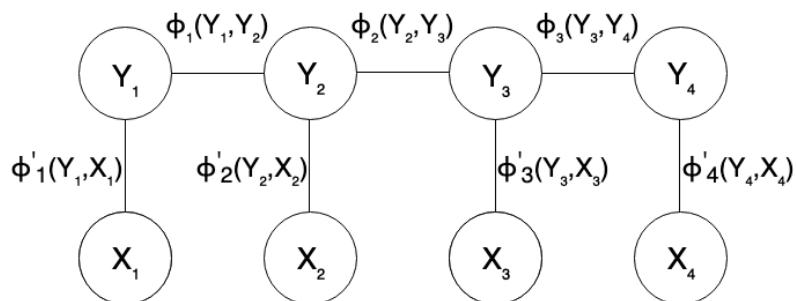
Dalam eksperimen studi kasus yang saya jalankan, saya menggunakan tiga jenis teks untuk mengidentifikasi berbagai metode POS Tagging melakukan suatu anotasi kategori kata. Untuk studi kasus ini, saya menggunakan teks yang keseluruhannya Bahasa Inggris dikarenakan Bahasa Inggris adalah bahasa yang memiliki resource yang banyak (High-source Language). Dalam teks ketiga, saya menggunakan paragraf yang berasal dari buku lama, Metamorphosis oleh Ovid, untuk melihat bagaimana model menangani teks yang pembahasannya lebih kompleks. Teks yang saya gunakan bisa dilihat dibawah ini:

Tabel 2. Teks studi kasus POS Tagging

1	The quick brown fox jumps over the lazy dog.
2	The young researcher presented her findings at the international conference. Later, she answered questions confidently from the audience.
3	“They took the upward path, through the still silence, steep and dark, shadowy with dense fog, drawing near to the threshold of the upper world. Afraid she was no longer there, and eager to see her, the lover turned his eyes. In an instant she dropped back, and he, unhappy man, stretching out his arms to hold her and be held, clutched at nothing but the receding air. Dying a second time, now, there was no complaint to her husband (what, then, could she complain of, except that she had been loved?). She spoke a last ‘farewell’ that, now, scarcely reached his ears, and turned again towards that same place.” Sumber: Orpheus and Eurydice, Metamorphosis, Ovid.

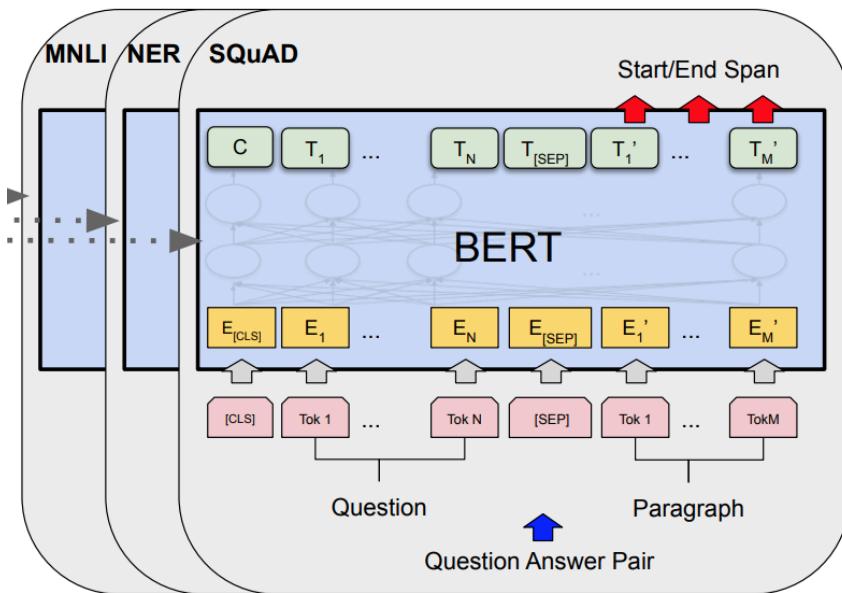
3.2 Algoritma dan Source Code

Untuk algoritma dan model yang saya gunakan, saya menggunakan dua model tradisional dan dua model pendekatan transformer (LLM). Untuk model tradisional, saya menggunakan model **Hidden Markov Model (HMM)** dan **Conditional Random Fields (CRF)**. Kedua model ini saya lakukan pelatihan menggunakan treebank dan memastikan menggunakan universal_tagset, sesuai dengan library nltk.



Gambar 3. Arsitektur model CRF

Untuk model modern, karena keterbatasan infrastruktur komputasi dan waktu untuk melakukan pelatihan dari awal, saya menggunakan model yang sudah pre trained dan fine tune. Dalam eksperimen ini, saya menggunakan model berbasis BERT dan ROBERTA. Kedua model tersebut sudah mendapatkan pendekatan transformer dan fine tuning untuk melakukan kategorisasi layaknya POS Tagging.



Gambar 4. Arsitektur model BERT

3.3 Hasil dan Analisis Eksperimen

Berdasarkan analisis kategorisasi POS tagging dari teks terlampir, saya bisa menyimpulkan bahwa untuk teks sederhana (1-2 kalimat), model tradisional sudah mampu melakukan klasifikasi POS Tagging. Namun, akurasinya belum mampu mencapai model transformer. Hal ini didapat dari hasil prediksi yang ketika konteksnya sudah panjang, maka kata tersebut akan dikategorikan NOUN. Hal ini bisa saja dipengaruhi dari data apa yang dipelajari oleh model (treebank dalam kasus ini). Hal ini dimitigasi dalam model CRF yang lebih mampu mendeteksi ADP dan DET dalam kalimat, namun tidak bisa semaksimal model BERT dan ROBERTA.

Tabel 3. Analisis teks pertama (1) metode tradisional vs LLM

Text	HMM	CRF	BERT	ROBERTA
The	DET	DET	DET	DET
quick	ADJ	ADJ	ADJ	ADJ
brown	NOUN	NOUN		

fox	NOUN	NOUN	NOUN	NOUN
jumps	NOUN	NOUN	VERB	VERB
over	NOUN	ADP	ADP	ADP
the	NOUN	DET	DET	DET
lazy	NOUN	NOUN	ADJ	ADJ
dog	NOUN	NOUN	NOUN	NOUN
.	NOUN	.	PUNCT	PUNCT

Dalam hasil eksperimen teks kedua, saya menemukan bahwa model tradisional CRF mampu lebih mengenali teks tersebut dibandingkan model HMM, meskipun keduanya belajar dari data yang sama. Hal ini dapat disimpulkan dari model mampu menganalisis konteks. Dalam eksperimen ini, model transformer mampu mendeteksi apakah adanya tanda baca (punctuation) yang dimana memisahkan teks sebelumnya dan tanda baca. Kategori kata seperti ADJ, ADP, dan ADV terlihat masih sulit diklasifikasi untuk model tradisional, hal ini dikarenakan masih ada kesalahan pemahaman model dan mengelompokkan menjadi NOUN.

Tabel 4. Analisis teks kedua (2) metode tradisional vs LLM

The	DET	DET	DET	DET
young	ADJ	ADJ	ADJ	ADJ
researcher	NOUN	NOUN	NOUN	NOUN
presented	VERB	VERB	VERB	VERB
her	PRON	PRON	PRON	PRON
findings	NOUN	NOUN	NOUN	NOUN
at	ADP	ADP	ADP	ADP
the	DET	ADJ	DET	DET
international	ADJ	NOUN	ADJ	PROPN
conference.	NOUN	NOUN	NOUN	NOUN
Later,	NOUN	NOUN	ADV	ADV
			PUNCT	PUNCT
she	NOUN	NOUN	PRON	PRON
answered	NOUN	VERB	VERB	VERB
questions	NOUN	ADJ	NOUN	NOUN
confidently	NOUN	NOUN	ADV	ADV
from	NOUN	ADP	ADP	ADP
the	NOUN	DET	DET	DET
audience.	NOUN	NOUN	NOUN	NOUN
			PUNCT	PUNCT

Dalam eksperimen teks ketiga (paragraf), saya sengaja tidak menampilkan disini dikarenakan cukup panjang untuk membandingkan semua hasilnya. Saya lampirkan potongan kecil dari hasil prediksi. Saya mendapati bahwa sekali lagi, model tradisional dengan limitasi data (treebank) hanya dapat melakukan klasifikasi kata. HMM cenderung mengkategorikan kata sebagai NOUN, CRF meskipun sudah cukup baik mengkategorikan yang lebih beragam, belum semaksimal model LLM.

Potongan ini juga menunjukkan bahwa ada perbedaan dari kedua model melakukan tokenisasi yang digunakan dalam POS Tagging. Model BERT dan ROBERTA dapat langsung mengklasifikasi dua atau tiga kata yang jadi satu kesatuan. Hal ini didapatkan dari kata “except that” dan “has been” dikategorikan sebagai SCONJ dan AUX, namun dua kata tersebut adalah satu kesatuan kata, bukan dua kata terpisah yang di label berbeda.

Tabel 5. Analisis teks ketiga (3) metode tradisional vs LLM

Dying	NOUN	NOUN	VERB	VERB
a	NOUN	DET	DET	DET
second	NOUN	ADJ	ADJ	ADJ
time,	NOUN	NOUN	NOUN	NOUN
			PUNCT	PUNCT
now,	NOUN	NOUN	ADV	ADV
			PUNCT	PUNCT
there	NOUN	NOUN	PRON	PRON
was	NOUN	VERB	VERB	VERB
no	NOUN	ADJ	DET	DET
complaint	NOUN	NOUN	NOUN	NOUN
to	NOUN	PRT	ADP	ADP
her	NOUN	PRON	PRON	PRON
husband	NOUN	NOUN	NOUN	NOUN
			PUNCT	PUNCT
(what,	NOUN	NOUN	PRON	PRON
			PUNCT	PUNCT
then,	NOUN	NOUN	ADV	ADV
			PUNCT	PUNCT
could	NOUN	NOUN	AUX	AUX
she	NOUN	VERB	PRON	PRON
complain	NOUN	ADJ	VERB	VERB
of,	NOUN	NOUN	ADP	ADP
			PUNCT	PUNCT
except	NOUN	NOUN	SCONJ	ADP
that	NOUN	ADP		SCONJ
she	NOUN	PRON	PRON	PRON
had	NOUN	VERB	AUX	AUX
been	NOUN	VERB		
loved?).	NOUN	ADJ	VERB	VERB

			<i>PUNCT</i>	<i>PUNCT</i>
--	--	--	--------------	--------------

3.4 Kesimpulan

Berdasarkan hasil eksperimen yang sudah dilakukan, saya menemukan bahwa metode pendekatan LLM atau transformer lebih baik dibandingkan model tradisional. Model yang telah di fine tune ini mampu mengkategorikan lebih baik, mengenal konteks suatu kata dan lebih akurat. Model tradisional HMM cukup terkendala apalagi kompleksitas kata dan panjang kalimat, menyebabkan kata tersebut hanya sebagai NOUN dan model CRF belum mampu mengelompokkan kata dalam kelompok seperti ADV dan AUX. Namun, eksperimen yang saya buat adalah teks bahasa inggris dan hasil bisa saja berbeda dengan studi kasus yang menggunakan bahasa non inggris. Hasil eksperimen ini saya lampirkan dalam kode berikut:

<https://github.com/Pieter414/cif61343-natural-language-processing/tree/main/pos-tagging>

DAFTAR PUSTAKA

Adelani, David Ifeoluwa, et al. "Comparing LLM prompting with Cross-lingual transfer performance on Indigenous and Low-resource Brazilian Languages." arXiv preprint arXiv:2404.18286 (2024).

AndyLiang12. (2025). bert-finetuned-pos. Hugging Face. dari
<https://huggingface.co/AndyLiang12/bert-finetuned-pos>

GeeksforGeeks. (2025, August 19). NLP | Part of Speech (POS) tagging. In GeeksforGeeks. Dari
<https://www.geeksforgeeks.org/nlp/nlp-part-of-speech-default-tagging/>

Jurafsky, D., & Martin, J. H. (2025). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, with language models (3rd ed.).

Wietseveld. (2025). xlm-roberta-base-ft-udpos28-en. Hugging Face. dari
<https://huggingface.co/wietseveld/xlm-roberta-base-ft-udpos28-en>