

# Epistemologi dalam Natural Language Processing

**Danang Arbian Sulistyono\*, Fadhli Almu'iini Ahda, Vivi Aida Fitria**

Universitas Negeri Malang, Jl. Semarang No. 5 Malang, Jawa Timur, Indonesia

Institut Teknologi dan Bisnis Asia, Jl. Soekarno-Hatta, Rembeksari 1A, Malang, Jawa Timur

\*Penulis korespondensi, Surel: danang.arbian@asia.ac.id/

danang.arbian.2205349@students.um.ac.id

Paper received: Paper received: 01-9-2021; revised: 11-9-2021; accepted: 17-9-2021

## Abstract

How to obtain the truth about knowledge by considering the axiology and anthology aspects of knowledge is the challenge that epistemology must solve. While in scientific epistemology, the accumulation of information that is true will affect how inquiries about the universe are answered heuristically and how natural occurrences are predicted. The primary goal and aim of epistemology, a subfield of philosophy of science, is to investigate and ascertain the nature of knowledge. As such, it examines the origin, sources, and importance of validity from knowledge in addition to discussing the extent and veracity of science. The goal of NLP, a branch of artificial intelligence (AI), is to enable computers to comprehend human language. For instance, text and voice, which people frequently utilize in casual discussions. Integrating computational linguistics with predictive methods led to the development of NLP. NLP has so far done well with text and audio data. There are still others who believe that NLP is in decline, particularly when it comes to managing idioms and sarcasm in contextual data. Due to the vast number of local languages spoken worldwide, the millions of words they contain, the hundreds of regional accents, and their importance in preventing the extinction of local languages, even machine translation, which was the initial purpose of NLP, may still be investigated further.

**Keywords:** epistemologi; nlp; natural language processing

## Abstrak

Masalah yang harus dihadapi oleh Epistemologi adalah bagaimana mendapatkan kebenaran akan pengetahuan dengan menimbang aspek aksiologi dan aksiologi pada pengetahuan. Sedangkan pada epistemologi ilmiah, penyusunan kebenaran suatu pengetahuan akan berpengaruh untuk menjawab pertanyaan di dunia secara heuristik serta dalam memprediksi fenomena alam yang terjadi. Mempelajari dan menentukan hakikat dari suatu pengetahuan adalah fungsi dan tugas utama epistemologi sebagai salah satu cabang dari filsafat ilmu, maka tidak hanya berbicara tentang kebenaran ilmu pengetahuan dan ruang lingkup pengetahuan, akan tetapi secara luas epistemologi juga mempelajari tentang asal mula, sumber dan juga nilai validitas dari pengetahuan. Pemrosesan bahasa alami, atau NLP, adalah bagian dari kecerdasan buatan (AI) yang berkaitan dengan memberi komputer kemampuan untuk memahami bahasa alami manusia. Misalnya teks dan suara yang sering digunakan manusia dalam percakapan sehari-hari. NLP dibuat dengan menggabungkan linguistik komputasi dengan model statistik. Sampai saat ini NLP memiliki performa yang baik pada data teks dan audio. Namun, masih ada orang yang menilai penurunan dunia NLP, terutama dalam penanganan sarkasme dan idiom dalam data kontekstual. Bahkan terjemahan mesin yang merupakan tujuan awal NLP masih dapat dieksplorasi lebih dalam, karena ada banyak bahasa lokal di dunia, ada jutaan kata, ratusan aksentuasi lokal, dan perannya untuk menyelamatkan Bahasa Lokal dari kepunahan.

**Kata kunci:** epistemologi; nlp; natural language processing

## 1. Pendahuluan

Dalam kaitannya dengan filsafat ilmu, terdapat 3 landasan yang berkaitan ketika kita berbicara tentang pengetahuan, yaitu aksiologi, ontologi dan epistemologi. Berdasarkan landasan ontologi dan aksiologi, pengetahuan dianalogikan sebagai suatu alat untuk

memecahkan masalah dan persoalan yang dialami oleh manusia, dengan memprediksi dan melakukan kontrol terhadap alam, pada dasarnya masalah yang dihadapi oleh manusia akan teratasi. Atas dasar pendekatan antologi dan aksiologi seperti itu maka bagaimana menentukan dan mengembangkan epistemologi untuk pengetahuan tersebut, karena mempelajari dan menentukan hakikat dari suatu pengetahuan adalah fungsi dan tugas utama epistemologi sebagai salah satu cabang dari filsafat ilmu, maka tidak hanya berbicara tentang kebenaran ilmu pengetahuan dan ruang lingkup pengetahuan, akan tetapi secara luas epistemologi juga mempelajari tentang asal mula, sumber dan juga nilai validitas dari pengetahuan (Purwanto Wahyudi 2016).

Episeme yang berarti pengetahuan dan Logos yang berarti pikiran dalam bahasa Yunani adalah kata pembentuk dari Epistemologi yang dalam perkembangannya diartikan sebagai pengetahuan yang benar. Epistemologi adalah salah satu cabang dari filsafat ilmu, maka Epistemologi mempelajari dan menentukan hakikat dari suatu pengetahuan. Epistemologi tidak hanya berbicara tentang kebenaran ilmu pengetahuan dan ruang lingkup pengetahuan, akan tetapi secara luas epistemologi juga mempelajari tentang asal mula, sumber dan juga nilai validitas dari pengetahuan

Natural Language Processing adalah salah satu bagian dari Kecerdasan Buatan (AI) yang kaitannya adalah bagaimana bahasa alami yang digunakan manusia bisa dimengerti dan dipahami oleh komputer, baik berupa text ataupun suara. Untuk membantu sebuah komputer memahami dan mendapatkan makna penuh, lengkap dengan maksud dan setimennya dari text atau suara yang diberikan, maka digunakan gabungan dari beberapa unsur keilmuan, diantaranya adalah komputer linguistik, model statistik, deep learning dan machine learning (Johri, P., Khatri 2021).

## 2. Metode

Sebagian besar pendekatan yang digunakan untuk membuat mesin memahami bahasa masih berdasarkan aturan tertulis tentang bagaimana mesin mencocokkan kata demi kata untuk memahami arti dari sebuah kalimat. Tapi di 1980-an, muncul cabang ilmu komputer lain yang memegang peranan penting di bidang komputasi yaitu Machine Learning, Machine Learning membuat perubahan yang signifikan mengenai aturan tertulis menuju ke pembuatan konsep. Algoritma Machine Learning menyediakan cara yang lebih maju dalam menafsirkan keambiguitas serta dapat memberikan bukti yang bisa dijadikan pertimbangan dalam pengambilan keputusan. Algoritma seperti Decision Trees yang menggunakan aturan If-Then bisa digunakan untuk menyimpulkan hasil dengan optimal, dengan didukung algoritma probabilistik untuk mem-back-up keputusan yang diambil akan memberikan kepercayaan yang cukup baik terhadap kesimpulan dan keputusan yang diambil.

Saat ini, telah terjadi pergeseran ke arah Deep Learning, ini dikarenakan adanya dominasi Deep Learning dalam melakukan tugas-tugas sulit yang tidak bisa dilakukan dengan menggunakan rules dan criteria yang statis. Ketika hal ini dibawa ke dalam ranah NLP, tidak memungkinkan untuk memecahkan masalah keambiguitas bahasa. Dikarenakan satu kata dapat memiliki beberapa arti tergantung kata di sebelahnya dan tidak mungkin untuk menulis aturan atau menggunakan Decision Tree untuk mewakili setiap kemungkinan makna yang ada. Deep Learning memecahkan masalah ini secara efisien, karena tidak membutuhkan programmer untuk memberikan suatu aturan dalam mengambil keputusan, akan tetapi algoritma itu sendiri akan menyimpulkan proses pemetaan dari input sampai ke output.

## 2.1. Data Processing dan Pengembangan

Memproses bahasa manusia adalah salah satu tugas terberat yang harus diselesaikan oleh mesin. Konsep dari Bahasa yang ada bukanlah suatu kalimat yang mempunyai bentuk atau konsep tunggal yang berdiri sendiri, akan tetapi terbentuk dari beberapa hal seperti kata benda, kata kerja, kata sifat, determinan, dll yang saling ber-asosiasi untuk memberikan sebuah arti dalam kalimat tersebut. Disinilah sering muncul nilai-nilai ambigu dari suatu kata yang membentuk kalimat, seperti contoh pada kalimat berikut "he will be running a marathon soon" dan "the software is running faster than expected", dalam kedua kalimat menggunakan frase "Running" yang sama, tapi jika dilihat dari keseluruhan kalimat, maka akan memberikan arti yang berbeda. Arti dari sebuah kata tergantung pada referensi yang digunakan, sedangkan mesin atau komputer jelas akan kesulitan dalam mengurai jenis kalimat yang mengandung ambiguitas seperti ini. Tugas yang dijalankan oleh mesin atau komputer akan jauh lebih kompleks ketika tata bahasa juga turut dipertimbangkan.

Tata bahasa tersebut dapat digunakan untuk menangani beberapa sub kategori dengan menambahkan aturan ataupun batasan tambahan (misalnya sebuah kalimat akan dikategorikan ke dalam kalimat bersyarat dengan klausa yang dimulai dengan kata "jika"). Meskipun begitu, penggunaan pendekatan berbasis aturan/rule masih belum memberikan hasil yang efisien dikarenakan ambiguitas dan kompleksitas bahasa, kedua masalah tersebut menjadi fokus penting untuk diselesaikan dan oleh sebab itu maka metode statistik untuk NLP mulai mendapatkan tempat sebagai metode dalam penyelesaian masalah di NLP. Beberapa rule/aturan dalam pemecahan kalimat dapat ditangani dengan baik dengan adanya pendekatan probabilistik untuk kemudian diuraikan secara statistik (Jurafsky, D., & Martin, J. 2008).

Dengan mengumpulkan rule secara terperinci dan mendefinisikan rule secara lebih luas maka algoritma Machine Learning bisa menentukan tingkat peluang yang terjadi dan dimasukkan ke dalam rule secara terpisah. Pendefinisian yang lebih luas terhadap aturan/rule dapat membantu untuk membangun sebuah pohon keputusan/decision tree sehingga secara statistik, parser akan dapat mengurai kalimat dengan kemungkinan maksimum. Dengan metode seperti ini, dinilai pendekatan secara statistik menghasilkan hasil yang lebih baik dari metode sebelumnya

NLP terdiri dari lima komponen utama: analisis morfologi, analisis sintaksis, analisis semantik, analisis wacana, dan analisis pragmatis. Parsing adalah langkah utama dalam melakukan analisis sintaksis, proses ini menganalisis kata-kata penyusun dalam sebuah teks berdasarkan tata bahasa yang mendasarinya untuk menentukan struktur sintaksisnya (Mallamma V. Reddy 2014). Lima komponen tersebut diatas mencakup beberapa sub-tugas seperti: dekomposisi teks, ejaan, penguraian secara morfologis, stemming. Dari proses parsing/penguraian yang dilakukan terhadap suatu kalimat akan menghasilkan parse tree yang memiliki akar kata, kata benda, kata kerja, dll baik sebagai perantara ataupun kata-kata individu yang berdiri sendiri.

Proses Semantik pada NLP adalah proses untuk menyimpulkan arti dan makna dari teks, didalamnya termasuk beberapa proses seperti analisis semantik (memeriksa kebermaknaan teks), disambiguasi arti kata (menentukan arti kata yang benar, jika terjadi ambiguitas dalam makna), semantik leksikal (memeriksa sinonim, antonim, homonim, dll). Sedangkan proses pragmatis berkaitan dengan ekstraksi informasi dari sepotong teks yang selanjutnya dibagi

menjadi tiga sub-bidang: resolusi referensi (mendeteksi referensi) analisis wacana (menentukan struktur teks), dialog interpretasi (menafsirkan informasi dari teks) (Pustejovsky, J., Stubbs A. 2012).

Fokus dari NLP adalah bagaimana membuat komputer atau mesin untuk bisa menafsirkan, memahami, dan memanipulasi bahasa manusia. Secara tradisional, interaksi manusia dengan komputer dilakukan melalui bahasa pemrograman. Namun, ketika berbicara tentang interaksi bahasa antara manusia dengan mesin, maka pencapaiannya akan lebih menantang, dikarenakan bahasa manusia sangat ambigu dan kompleks, yang didalam strukturnya terkadang termasuk bahasa gaul/slang yang mempunyai makna yang tidak biasa dan terdapat konteks sosial didalamnya. Penelitian di bidang NLP menjadi lebih menantang ketika memperhitungkan masalah aksen bahasa, karena orang-orang dari berbagai daerah memiliki aksen yang berbeda pula.

## **2.2. Data Processing dan Pengembangan**

NLP menggabungkan dua tugas utama: analisis sintaks dan analisis semantik. Analisis Sintaks digunakan untuk membuat susunan kata-kata dalam kalimat secara sedemikian rupa agar dapat dipahami secara tata bahasa. Ini membantu NLP untuk menyimpulkan arti dari suatu kalimat berdasarkan kaidah tata bahasa. Analisis semantik dilakukan untuk menemukan makna di balik kata-kata dan penggunaannya dalam sebuah kalimat. Ini diterapkan oleh NLP untuk memahami struktur dan makna kalimat (trivusi.web.id 2012).

Sebelum dilakukannya tugas utama NLP yaitu analisis sintaks dan semantik, terlebih dahulu dilakukan dua fase utama pada NLP yaitu data preprocessing dan pengembangan algoritma (NLPlanet, Medium.com 2021). Komputer perlu melakukan persiapan dan pembersihan data agar text yang diberikan bisa dianalisis, untuk kemudian dilakukan pengambilan fitur yang ada pada teks dan diterapkan dan dikerjakan dalam suatu algoritma. Penerapan teknik pada data processing bisa berupa:

Tokenization: Proses membagi teks menjadi beberapa kata yang lebih kecil untuk memudahkan kinerja sistem.

Stopword removal: mengeliminasi kata yang dinilai terlalu umum sehingga menyisakan kata yang menunjukkan informasi yang bisa diambil dari teks.

Lemmatization dan stemming: Proses pengembalian kata ke bentuk dasar, menghilangkan imbuhan untuk mendapatkan kata dasarnya.

Part-of-speech tagging: Proses penandaan atau pengenalan kata berdasarkan pengucapannya (kata benda, dasar dan sifat).

Setelah melewati proses diatas, maka akan dilanjutkan ke pengembangan algoritma yang akan dipakai (P. Withanage 2018). Beberapa algoritma yang familiar diantaranya adalah:

### **1. Rules Based Algorithm**

Sebagai salah satu pendekatan dalam NLP yang masih digunakan sampai sekarang , Ruled Based Algoritim mengandalkan kecermatan dalam merancang aturan linguistik yang diterapkan

### **2.2.1. Machine Learning Based Algoritim**

Dengan menggunakan model statistik dan melakukan training terhadap data yang diberikan, menjadikan algoritma ini dianggap algoritma yang baik dalam NLP. Metodenya bisa disesuaikan terhadap banyaknya data yang di training. Algoritma ini dapat menentukan aturannya sendiri melalui pembelajaran dan training secara berulang dengan menggunakan kombinasi dari machine learning, Neural network dan juga deep learning.

Tahapan selanjutnya adalah dilakukannya tugas utama dari NLP, yaitu analisis sintaks/sintaksis dan analisis semantic. Analisis sintaksis adalah bagaimana membuat arti gramatikal dalam suatu kalimat berdasarkan susunan kata yang ada. Aturan baku dari tata bahasa digunakan oleh analisis sintaksis untuk menilai arti dan makna dari sebuah kalimat. Beberapa teknik yang digunakan antara lain adalah:

#### **2.2.1.1. Parsing**

digunakan dalam gramatical analysis untuk sebuah kalimat. Contohnya dalam kalimat "Ayam Berkokok", maka jika di parsing akan menjadi 2 kata, yaitu Ayam sebagai kata benda dan Berkokok sebagai kata kerja. Kedepannya, dalam melakukan pemrosesan lebih lanjut, proses ini akan sangat berguna.

#### **2.2.1.2. Word segmentation**

adalah proses pengambilan string dari suatu teks, untuk kemudian diturunkan ke dalam bentuk katanya. Contohnya adalah ketika kita melakukan scan terhadap tulisan tangan. Algoritma yang dipakai akan melakukan analisis tiap kata dengan dibagi spasinya.

#### **2.2.1.3. Sentence breaking**

adalah penempatan batas dari kalimat dalam suatu teks. Contohnya adalah ketika memasukkan sebuah kalimat "Ayam berkokok, saya bangun", maka komputer akan mengenali bahwa tanda koma tersebut digunakan untuk memecah kalimat menjadi beberapa kalimat.

#### **2.2.1.4. Morphological segmentation**

adalah pembagian kata menjadi bagian yang lebih kecil atau morfem. Contoh dalam kata "keingintahuan" akan dipecah menjadi [[ke][ingin][tahu][an]], dimana nantinya komputer akan mengenali "ke", "ingin", "tahu" dan "a" sebagai sebuah morfem. Hal ini sangat berguna dalam pengenalan ucapan ataupun aplikasi penerjemah.

#### **2.2.1.5. Stemming**

adalah pembagian kata yang mempunyai imbuhan dan dirubah menjadi kata dasarnya. Contoh "Ayam Berkokok", maka komputer akan mengenali kata dasar dari "berkokok" adalah "kokok".

Analisis kedua adalah berdasarkan semantik, dimana dalam hal ini melibatkan makna dan fungsi dari sebuah kata dalam kalimat. Penerapan algoritma digunakan NLP dalam memahami makna dan struktur dalam sebuah kalimat. Beberapa teknik analisis semantik adalah:

#### **2.2.1.5.1. Word sense disambiguation**

Menerjemahkan arti dari sebuah kata berdasarkan konteksnya. Contoh: "Bisa ular dapat berakibat fatal bagi korban jika tidak ditangani dengan tepat". Pada kalimat tersebut, kata 'bisa' memiliki arti racun.

#### **2.2.1.5.2. Named entity recognition.**

Melakukan pengkategorian kata-kata dalam beberapa kelompok kata, digunakan untuk mempelajari arti dari sebuah berita untuk kemudian mengidentifikasi penyebutan sebuah instansi atau produk, analisis ini bisa membedakan entitas yang sama secara visual.

#### **2.2.1.5.3. Natural language generation**

Natural Language Generation atau NLG menggunakan sebuah database untuk melakukan analisis semantik di balik kata sehingga bisa membuat teks baru. Layaknya manusia, NLG dapat menggambarkan, meringkas dan menjelaskan input data terstruktur dalam bentuk narasi dengan sangat cepat.

### **3. Hasil dan Pembahasan**

Seperti disebutkan di awal, bahwa tugas dari NLP di awal perkembangannya adalah sebagai mesin penerjemah, akan tetapi dengan semakin banyaknya software dan penelitian di bidang NLP, tugas dari NLP pun semakin besar dan kompleks (NLPlanet, Medium.com 2021).

Tugas dan wilayah kerja NLP pun bertambah seperti terlihat pada gambar.1, diantaranya adalah:

#### **3.1. Classification**

Text Classification: menetapkan kategori ke kalimat atau dokumen (misalnya pemfilteran spam).

Sentiment Analysis: mengidentifikasi polaritas sepotong teks.

#### **3.2. Information Retrieval and Document Ranking**

Sentence/document similarity: menentukan tingkat kemiripan dari dua atau lebih dokumen

Question Answering: menjawab pertanyaan dalam bahasa alami

#### **3.3. Text-to-Text Generation**

Machine Translation: menerjemahkan dari satu bahasa ke bahasa lain.

Text Generation: membuat teks yang tampak tidak dapat dibedakan dari teks tulisan manusia.

Text Summarization: membuat versi singkat dari beberapa dokumen yang mempertahankan sebagian besar maknanya.

Text Simplification: membuat teks lebih mudah dibaca dan dipahami, sambil mempertahankan gagasan utama dan makna aslinya.

Lexical Normalization: menerjemahkan/mengubah teks non-standar ke register standar.

Paraphrase Generation: membuat kalimat keluaran yang mempertahankan makna masukan tetapi mencakup variasi dalam pilihan kata dan tata bahasa.



Gambar 1. Perkembangan Tugas NLP (NLPlanet, Medium.com 2021).

### 3.4. Knowledge bases, entities, and relations

Relation extraction: mengekstraksi relasi semantik dari sebuah teks. Hubungan yang diekstraksi biasanya terjadi antara dua atau lebih entitas dan termasuk dalam kategori semantik tertentu (misalnya tinggal di, saudara perempuan dari, dll).

Relation prediction: mengidentifikasi relasi bernama antara dua entitas semantik bernama.

Named Entity Recognition: menandai entitas dalam teks dengan jenis yang sesuai, biasanya dalam notasi BIO.

Entity Linking: mengenali dan membedakan entitas bernama ke basis pengetahuan (biasanya Wikidata).

### 3.5. Topics and Keywords

Topic Modeling: mengidentifikasi "topik" abstrak yang mendasari kumpulan dokumen.

Keyword Extraction: mengidentifikasi istilah yang paling relevan untuk menggambarkan subjek dokumen

### **3.6. Chatbot**

Intent Detection: menangkap semantik di balik pesan dari pengguna dan menetapkannya ke label yang benar.

Slot Filling: bertujuan untuk mengekstrak nilai dari jenis atribut tertentu (atau slot, seperti kota atau tanggal) untuk entitas tertentu dari teks.

Dialog Management: mengelola keadaan dan aliran percakapan.

### **3.7. Penalaran Teks**

Common Sense Reasoning: penggunaan "akal sehat" atau pengetahuan dunia untuk membuat kesimpulan.

Natural Language Inference: menentukan apakah "hipotesis" itu benar (entailment), salah (kontradiksi), atau tidak ditentukan (netral) dengan "premis".

### **3.8. Fake News and Hate Speech Detection**

Fake News Detection: mendeteksi dan menyaring teks yang mengandung informasi palsu dan menyesatkan.

Stance Detection: menentukan reaksi individu terhadap klaim aktor utama. Ini adalah bagian inti dari serangkaian pendekatan terhadap penilaian berita palsu.

Hate Speech Detection: mendeteksi jika suatu teks mengandung ujaran kebencian.

### **3.9. Text-to-Data and vice-versa**

Text-to-Speech: teknologi yang membacakan teks digital dengan lantang.

Speech-to-Text: menyalin ucapan ke teks.

Text-to-Image: menghasilkan gambar foto-realistic yang secara semantik konsisten dengan deskripsi teks.

Data-to-Text: memproduksi teks dari input non-linguistik, seperti database catatan, spreadsheet, dan basis pengetahuan sistem pakar.

### **3.10. Text Preprocessing**

Resolusi Coreference: mengelompokkan penyebutan dalam teks yang merujuk pada entitas dunia nyata yang sama.

Part Of Speech (POS) tagging: menandai kata dalam teks dengan part of speech-nya. Part of speech adalah kategori kata dengan sifat gramatikal yang mirip, seperti kata benda, kata kerja, kata sifat, kata keterangan, kata ganti, kata depan, kata hubung, dll.

Word Sense Disambiguation: mengaitkan kata-kata dalam konteks dengan entri yang paling sesuai dalam inventaris sense yang telah ditentukan sebelumnya (biasanya WordNet).

Grammatical Error Correction: mengoreksi berbagai jenis kesalahan dalam teks seperti kesalahan ejaan, tanda baca, tata bahasa, dan pilihan kata.

Feature Extraction: ekstraksi fitur numerik generik dari teks.

Untuk membantu NLP dalam melaksanakan tugasnya, dan juga dengan disertai perkembangan software bantu yang berkembang secara pesat, maka semakin banyak pula pilihan dalam penggunaan software untuk NLP, diantaranya adalah beberapa software open source berdasarkan bahasa pemrogramannya:

## **3.2. Python Tools**

### **3.2.1. NLTK (Natural Language Tool Kit)**

NLTK adalah salah satu tool kit yang paling lengkap sampai saat ini, hampir semua komponen NLP bisa kita implementasikan, mulai dari klasifikasi, tokenisasi, stemming, tagging, parsing dan semantic reasoning. Meskipun dinilai untuk library nya sedikit lebih lambat dibandingkan dengan tools yang lain, akan tetapi NLTK merupakan toolkit yang baik untuk bereksperimen dan mengeksplorasi kombinasi algoritma untuk NLP.

### **3.2.2. SpaCy**

Sebagai salah satu pesaing dari NLTK, SpaCy dinilai lebih cepat dalam beberapa kasus. Salah satu kelebihan yang dimiliki SpaCy adalah bahwa semuanya direpresentasikan dalam bentuk objek, tidak lagi sebagai string seperti yang dilakukan di NLTK serta ditunjang dengan interface yang lebih sederhana untuk membangun aplikasi

### **3.2.3. TextBlob**

TextBlob adalah salah satu varian ekstensi dari NLTK, TextBlob menyederhanakan cara akses ke dalam semua fungsi dari NLTK, selain itu TextBlob juga menyederhanakan fungsionalitas ke dalam Pattern Library

### **3.2.4. Textacy**

Textacy menggunakan SpaCy dalam fungsi dari NLP, dengan menggunakan Textacy kita dapat menggunakan banyak jenis dari tipe data tanpa harus menggunakan pembantu kode tambahan

### **3.2.5. PyTorch-NLP**

PyTorch-NLP adalah sebuah software baru yang masih berumur beberapa tahun kebelakang, akan tetapi sudah mendapatkan tanggapan yang luar biasa dari komunitas NLP. PyTorch-NLP selalu terupdate dengan penelitian terbaru, dan banyak sekali perusahaan besar atau peneliti yang menggunakannya untuk segala macam pemrosesan seperti transformasi gambar.

### **3.3. Node Tools**

#### **3.3.1. ReText**

Retext adalah bagian dari unified collective. Unified adalah antarmuka yang memungkinkan beberapa alat dan plugin untuk berintegrasi dan bekerja sama secara efektif. Retext tidak menggunakan banyak teknik yang mendasarinya, tetapi menggunakan plugin untuk mencapai hasil dengan NLP. Sangat mudah untuk melakukan hal-hal seperti memeriksa ejaan, memperbaiki tipografi, mendeteksi sentimen, atau memastikan teks dapat dibaca dengan plugin sederhana.

#### **3.3.2. Compromise**

Compromise bukanlah alat yang paling canggih di NLP apalagi jika kita mencari algoritma tercanggih atau sistem terlengkap. Namun, jika kita menginginkan alat berkinerja yang memiliki banyak fitur dan dapat berfungsi di sisi klien, maka Compromise bisa kita gunakan. Secara keseluruhan, Compromise berkompromi pada fungsionalitas dan akurasi dengan berfokus pada fungsionalitas yang jauh lebih spesifik yang diuntungkan oleh pengguna yang lebih memahami konteks seputar penggunaan.

#### **3.3.3. Natural**

Natural mencakup sebagian besar fungsi pada Library dari NLP secara umum. Sebagian besar berfokus pada bahasa Inggris, meskipun beberapa bahasa lain sudah dikontribusikan. Software ini sudah mendukung tokenizing, stemming, klasifikasi, fonetik, WordNet, dan kesamaan string. Software ini dinilai paling mendekati NLTK, karena memasukkan semua teknik NLP dalam satu paket.

#### **3.3.4. Nlp.js**

Nlp.js dibangun di atas beberapa Library NLP, termasuk didalamnya adalah Franc dan Brain.js. Nlp.js menyediakan antarmuka yang baik dan di dalamnya terdapat banyak komponen NLP, seperti klasifikasi, analisis sentimen, stemming, pengenalan entitas bernama, dan pembuatan bahasa alami. Nlp.js juga mendukung beberapa bahasa, yang berguna untuk bekerja dalam suatu proyek selain bahasa Inggris.

### **3.4. Java Tools**

#### **3.4.1. OpenNLP**

OpenNLP didukung oleh Apache Foundation, sehingga mudah untuk mengintegrasikannya ke dalam proyek Apache lainnya, seperti Apache Flink, Apache NiFi, dan Apache Spark. OpenNLP adalah software NLP yang mencakup semua komponen pemrosesan umum NLP, dan dapat digunakan dari baris perintah atau dalam aplikasi sebagai Library. OpenNLP juga memiliki dukungan luas untuk berbagai bahasa. Secara keseluruhan, OpenNLP adalah alat yang hebat dengan banyak jika kita menggunakan Java.

#### **3.4.2. StanfordNLP**

Stanford CoreNLP adalah seperangkat alat yang digunakan untuk statistikal NLP, Deep Learning NLP, dan Rule-based NLP. Banyak bahasa pemrograman selain java sudah di binding dengan Stanford CoreNLP sehingga software ini dapat digunakan di luar Java.

### 3.4.3. CogCompNLP

CogCompNLP, yang dikembangkan oleh University of Illinois, juga memiliki pustaka Python dengan fungsi serupa. CogCompNLP dapat digunakan untuk memproses teks, baik secara lokal atau pada sistem jarak jauh, yang dapat menghilangkan beban luar biasa dari perangkat lokal kita. CogCompNLP menyediakan fungsi pemrosesan seperti tokenization, part-of-speech tagging, chunking, named-entity tagging, lemmatization, dependency and constituency parsing, and semantic role labeling.

## 3.5. Implementasi dan aplikasi NLP

NLP memiliki banyak aplikasi di berbagai bidang. Deep Learning telah membuka pintu untuk penelitian di bidang NLP yang lain. NLP banyak digunakan di perangkat IoT. Suara asisten seperti Cortana Alexa, Siri, dan Asisten Google telah membuat pasar mereka sendiri sekarang. Bahkan kita bisa melihat kegunaannya di kantor atau bahkan dirumah kita sendiri. Voice assistant tidak terbatas ke rumah atau kantor saja melainkan sekarang juga sudah mulai diterapkan di dunia otomotif juga [8]. Ada banyak sekali komentar, ide, saran telah diposting di internet dari media sosial. Menjadi sangat sulit bagi suatu perusahaan untuk memantau kinerja bisnis mereka berdasarkan komentar publik. Di situlah analisis sentimental, mengambil peran pentingnya. Sentimen analisis membantu bisnis menyaring informasi penting dari sejumlah besar data teks dan membuat keputusan optimal ke arah yang benar untuk memaksimalkan keuntungan mereka.

NLP dapat digunakan untuk menganalisis sejumlah besar data tekstual dan membuat prosesnya jauh lebih cepat dan efisien. Sejumlah besar informasi disimpan dalam bentuk teks dan akan memakan waktu untuk membaca teks ini secara manual untuk menemukan solusi. Sebelum NLP, semua dilakukan secara manual, akan tetapi dengan munculnya NLP berbasis Deep Learning maka seluruh proses tersebut direvolusi menjadi lebih efektif dan efisien. Contohnya adalah masalah catatan medis pasien, dengan banyaknya laporan medis yang dihasilkan di sebuah rumah sakit maka diperlukan kepakaran atau bahkan dokter dengan spesialisasi khusus untuk memeriksa laporan tersebut dan menyarankan sebuah diagnosis. Jika dilakukan secara manual, maka proses tersebut akan menjadi sangat lambat, akan tetapi dengan menggunakan NLP Teks Analisis maka proses tersebut akan menjadi lebih efisien.

Klasifikasi adalah salah satu teknik yang digunakan secara luas dalam bisnis yang mempunyai pasar yang berbeda. Klasifikasi digunakan untuk menyaring email penting dari email spam. Dengan teknik klasifikasi maka akan mudah untuk mengklasifikasikan sebuah email sebagai spam mail atau bukan karena ukuran datanya yang besar. Untuk data besar, kemungkinan deteksi spam meningkat. Tapi ketika datanya kecil, maka akan sulit untuk membedakan antara yang asli dan yang palsu. Identifikasi dan penghapusan ulasan palsu seperti itu akan membutuhkan otomatisasi karena itu adalah tugas yang sulit bagi manusia untuk dilakukan secara manual. Algoritma machine learning dapat digunakan dengan NLP untuk menghasilkan hasil yang baik (P. Withanage 2018).

NLP memudahkan orang untuk menulis esai, email, artikel, dll. Sebagai contoh, perangkat lunak Grammarly dapat mendeteksi sebuah artikel setelah ditulis, perangkat lunak tersebut dapat memeriksa tata bahasa, bahkan dapat menyarankan cara yang lebih baik untuk menulis kalimat. Karena sebagian besar informasi berhubungan dengan dokumen dan kontrak digital, maka diperlukan suatu mekanisme yang melindungi privasi suatu organisasi. Named Entity Recognition dapat digunakan untuk memantau dan mendeteksi privasi pelanggaran dalam kontrak online secara otomatis dengan memantau Informasi Identifikasi Pribadi (Vijayakumar B 2019).

Mesin Pencari seperti Google, Bing, DuckDuckGo, dll. menggunakan mesin penerjemah untuk menerjemahkan konten halaman web ke bahasa lain sesuai dengan makna asli dan pesan yang ingin disampaikan. Banyak organisasi dan perusahaan menggunakan chat-bot di situs web mereka sebagai sumber utama interaksi antara situs dan pengguna. Chat-bot memproses query yang dimasukkan pengguna dan menunjukkan hasil serta saran yang sesuai. Chat-bot dianggap banyak lebih cepat dari manusia untuk memproses informasi dan menghasilkan saran, sehingga mengurangi gesekan antara masalah dan solusi

#### 4. Simpulan

NLP sudah jauh berkembang dalam hal implementasi dan aplikasinya. Dari kamus dwibahasa ke aturan penulisan tangan, lalu dari penataan frasa ke ATN (Augmented Transition Network), dan akhirnya machine learning, NLP telah berkembang dalam berbagai tahap dengan lebih baik dengan kemampuan pemrosesan tingkat lanjut yang lebih baik lagi. Sulit untuk menyangkal dan menolak kemunculan AI di masa depan untuk datang ke dalam kehidupan kita sehari-hari. Ada banyak aplikasi NLP yang berkembang melalui metode deep learning dan aplikasi tersebut sudah bekerja sesuai dengan harapan manusia di beberapa bidang terkait, yang bahkan sekarang mungkin sudah melampaui kinerja dari manusia.

Saat ini, di lingkungan kita, hampir sebagian besar di sekitar kita sudah berjalan secara otomatis. Machine Learning telah membanjiri pasar dengan segala perangkat yang semakin canggih sehingga memudahkan seluruh pekerjaan kita. Sebagian orang berfikir bahwa NLP sudah mencapai puncaknya dan sudah tidak ada pembaharuan lagi di dalamnya, akan tetapi masih banyak hal yang masih belum tersentuh oleh NLP, masih banyak bidang yang masih bisa di eksplorasi lebih dalam oleh NLP dengan batuan machine learning dan deep learning

Sampai saat ini, NLP dinilai sudah memiliki kinerja yang baik dalam data tekstual dan audio, NLP memiliki prestasi yang baik dalam memproses data seperti tersebut. Akan tetapi masih ada yang menilai kemunduran dalam dunia NLP seperti pemrosesan sarkasme, ironi dan idiom dalam data kontekstual. Mesin penerjemah pun yang menjadi tujuan awal dibuatnya NLP pun masih bisa digali lebih dalam, dikarenakan banyaknya bahasa kedaerahan di dunia, dengan jutaan kata, ratusan aksent kedaerahan, yang fungsinya naik sebagai pengangkat dan penyelamat bahasa daerah lokal dari kepunahan.

#### Daftar Rujukan

- Purwanto, W. (2016). Apa yang Dimaksud dengan Epistemologi? <https://spi.uin-alauddin.ac.id/index.php/2016/10/31/apa-yang-dimaksud-dengan-epistemologi/>
- Johri, P., Khatri, S. K., Al-Taani, A. T., Sabharwal, M., Suvanov, S., & Kumar, A. (2021). Natural language processing: History, evolution, application, and future work. In *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020* (pp. 365-375). Springer Singapore.
- Lees, R. B. (1957). Syntactic structures. 33(3):375-408. <https://doi.org/10.2307/411160> JSTOR 411160
- Daniel, J., & James H, M. (2007). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. prentice hall.

Mallamma, V. R., & Hanumanthappa, M. (2014). Semantical and syntactical analysis of nlp. *International Journal of Computer Science and Information Technologies*, 5(3), 3236-3238.

Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc."

<https://www.trivusi.web.id/2022/08/natural-language-processing.html>

<https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b>

Withanage, P., Liyanage, T., Deeyakaduwe, N., Dias, E., & Thelijagoda, S. (2018, December). Voice-based Road navigation system using natural language processing (nlp). In *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)* (pp. 1-6). IEEE.

Vijayakumar, B., & Fuad, M. M. M. (2019). A new method to identify short-text authors using combinations of machine learning and natural language processing techniques. *Procedia Computer Science*, 159, 428-436."