



# AI Career Bootcamp

# Intro to

# Natural Language

# Processing

## Pembukaan

Guide Book ini memuat beberapa informasi-informasi utama yang akan disampaikan oleh mentor di program AI Career Bootcamp yang bisa dijadikan pegangan para students untuk mempersiapkan diri sebelum sesi Live Class berlangsung.

## Deskripsi

Natural Language Processing (NLP) adalah cabang ilmu komputer yang berfokus pada interaksi antara manusia dan komputer melalui bahasa manusia alami. NLP menggabungkan prinsip-prinsip dari linguistik komputasional, pengolahan bahasa alami, dan pembelajaran mesin untuk memungkinkan komputer memahami, memproses, dan menghasilkan teks atau ucapan yang serupa dengan cara manusia.



Tujuan utama NLP adalah untuk mengatasi kompleksitas bahasa manusia dan mengubah teks atau ucapan yang tidak terstruktur menjadi representasi yang dapat diolah oleh komputer. Hal ini melibatkan serangkaian tugas, seperti pemrosesan teks, pemahaman bahasa alami, pengenalan entitas, pemodelan bahasa, terjemahan mesin, dan analisis sentimen.

Pada dasarnya, NLP bekerja dengan memanfaatkan algoritma dan model statistik untuk menganalisis struktur dan makna dalam teks atau ucapan. Metode yang umum digunakan dalam NLP meliputi tokenisasi (pemisahan teks menjadi unit-unit yang lebih kecil, seperti kata atau frasa), analisis sintaksis (menganalisis struktur tata bahasa), analisis semantik (mengidentifikasi makna kata dan kalimat), dan pemodelan statistik (menggunakan data pelatihan untuk memprediksi dan memahami teks baru).

NLP telah diterapkan dalam berbagai bidang, termasuk pencarian web, asisten virtual, analisis sentimen media sosial, penerjemahan otomatis, pengolahan dokumen, dan banyak lagi. Dengan kemajuan teknologi dan pengembangan model pembelajaran mesin yang lebih canggih, NLP terus berkembang dan memberikan kontribusi signifikan dalam meningkatkan interaksi antara manusia dan komputer melalui bahasa manusia alami.

## Sejarah Singkat

Sejarah Natural Language Processing (NLP) dimulai pada tahun 1950-an dan telah mengalami perkembangan yang signifikan sejak saat itu. Berikut adalah rangkuman sejarah utama NLP:

**1** Awal Mula: Era Linguistik Komputasional (1950-an hingga 1960-an)

Pada awalnya, NLP berkaitan erat dengan bidang linguistik komputasional. Pada tahun 1950-an, Alan Turing mengusulkan "Tes Turing" yang menantang komputer untuk meniru perilaku manusia dalam percakapan. Pada dekade yang sama, George Zipf menerbitkan karya pentingnya tentang hukum Zipf, yang menyoroti pola frekuensi kata dalam bahasa alami.

**2** Era Tata Bahasa dan Teks (1960-an hingga 1970-an)

Pada periode ini, perhatian beralih ke analisis tata bahasa dan pemrosesan teks. Proyek-proyek seperti Linguistic String Project dan program komputer SHRDLU oleh Terry Winograd menunjukkan kemampuan komputer dalam memahami tata bahasa dan menjawab pertanyaan.

**3** Era Statistik (1980-an hingga 1990-an)

Pada tahun 1980-an, metode statistik mulai mendominasi NLP. Penerapan model probabilistik dan pembelajaran mesin dalam NLP semakin populer. Teknik seperti n-gram, pemodelan bahasa, dan parsing probabilistik dikembangkan dan digunakan dalam berbagai aplikasi NLP.

**4** Era Web dan Pertumbuhan Data (2000-an)

Pada tahun 2000-an, munculnya World Wide Web dan ledakan data digital mengubah bidang NLP. Perkembangan teknik pemrosesan teks massal, seperti pengindeksan web dan pencarian informasi, menghadirkan tantangan baru dalam pemahaman bahasa manusia alami dalam konteks yang lebih luas.

## 5 Era Deep Learning (2010-an hingga sekarang)

Perkembangan dalam bidang pembelajaran mesin dan khususnya deep learning menghasilkan kemajuan besar dalam NLP. Dengan menggunakan arsitektur jaringan saraf yang dalam, seperti rekurensi dan transformasi, model-model seperti Word2Vec, BERT, GPT, dan Transformer telah mengubah cara komputer memahami dan menghasilkan teks.

Seiring berjalannya waktu, NLP semakin berkembang dan mencapai pencapaian yang luar biasa. Kemampuan komputer dalam menerjemahkan bahasa, mengenali entitas, menganalisis sentimen, dan berinteraksi dengan manusia melalui bahasa alami terus meningkat. Terobosan baru terus muncul, memperluas batas-batas apa yang bisa dicapai oleh NLP dalam berbagai domain dan aplikasi.

## Aplikasi

Natural Language Processing (NLP) adalah bidang dalam kecerdasan buatan yang bertujuan untuk memungkinkan komputer memahami dan berinteraksi dengan bahasa manusia alami. Aplikasi NLP mencakup berbagai teknik dan algoritma yang digunakan untuk memproses, menganalisis, dan menghasilkan bahasa manusia secara otomatis. Contohnya, NLP dapat digunakan untuk pemrosesan bahasa alami, seperti memahami teks, ekstraksi informasi, analisis sentimen, dan klasifikasi teks. Selain itu, aplikasi NLP juga melibatkan pengembangan chatbot dan asisten virtual yang dapat berkomunikasi dengan pengguna menggunakan bahasa alami, sistem terjemahan mesin untuk menerjemahkan teks dari satu bahasa ke bahasa lain, serta analisis media sosial untuk memantau merek dan mengevaluasi sentimen publik. Kemajuan dalam NLP terus menghadirkan inovasi baru, memperluas kemampuan komputer dalam memahami dan menggunakan bahasa manusia secara lebih efektif.

Penggunaan NLP tidak terbatas pada domain tertentu saja, melainkan melibatkan beragam industri dan sektor. Misalnya, dalam industri e-niaga, NLP dapat digunakan untuk menganalisis ulasan pelanggan, memahami preferensi pembeli, dan meningkatkan pengalaman pengguna. Di bidang kesehatan, aplikasi NLP dapat membantu mengidentifikasi gejala penyakit berdasarkan riwayat medis pasien dan mencari solusi terbaik. Di sektor perbankan dan keuangan, NLP dapat digunakan untuk melakukan analisis teks pada laporan keuangan, mendeteksi penipuan, serta memberikan pelayanan pelanggan yang lebih personal melalui chatbot. Melalui penggabungan kecerdasan

buatan dengan kemampuan bahasa manusia alami, aplikasi NLP terus berkembang dan berperan penting dalam memfasilitasi komunikasi dan interaksi yang lebih efektif antara manusia dan komputer.