



Indonesia AI
AI for Everyone, AI for Indonesia

2023JAGO
TEKNOLOGIAI



AI Career Bootcamp

Text Preprocessing

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

Text processing pada Natural Language Processing (NLP) adalah suatu proses pengolahan teks yang bertujuan untuk memahami dan menganalisis bahasa manusia secara komputasional. Dalam tahap pertama, teks mentah yang diperoleh dari berbagai sumber, seperti dokumen, artikel, atau percakapan, akan diubah menjadi bentuk yang lebih terstruktur dan dapat dipahami oleh komputer. Proses ini melibatkan beberapa teknik, seperti tokenisasi, yaitu pemisahan teks menjadi unit-unit yang lebih kecil, seperti kata-kata atau kalimat-kalimat. Selain itu, langkah-langkah seperti penghapusan tanda baca, stemming, dan lematisasi juga dilakukan untuk mengurangi variasi kata yang serupa, sehingga mempermudah analisis lebih lanjut.



Setelah teks mentah diolah menjadi bentuk yang lebih terstruktur, tahap berikutnya dalam text processing adalah pemodelan representasi teks. Pada tahap ini, tujuannya adalah mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Salah satu teknik yang sering digunakan adalah metode vektorisasi, seperti TF-IDF (Term Frequency-Inverse Document Frequency) atau word embedding, seperti Word2Vec atau GloVe. Representasi numerik ini memungkinkan komputer untuk mengenali pola dan hubungan dalam teks, serta melakukan berbagai tugas NLP, seperti klasifikasi teks, penerjemahan mesin, atau pengenalan entitas.

Selain itu, text processing dalam NLP juga melibatkan analisis linguistik yang lebih mendalam. Teknik-teknik seperti pengenalan entitas bernama, parsing sintaksis, dan analisis semantik digunakan untuk memahami struktur dan makna dari teks. Dalam tahap ini, komputer mencoba mengenali entitas, seperti orang, tempat, tanggal, dan koneksi antar kalimat atau frasa dalam teks. Hal ini memungkinkan komputer untuk memahami konteks dan menghasilkan informasi yang lebih kaya dari teks yang diproses.

Mengapa Text Preprocessing Penting?

Text processing dalam NLP sangat penting dilakukan karena alasan-alasan berikut:

1 Memahami Bahasa Manusia

Manusia berkomunikasi melalui bahasa manusia yang kompleks dan bervariasi. Text processing memungkinkan komputer untuk memahami dan menganalisis bahasa manusia dengan cara yang mirip seperti manusia. Ini memungkinkan komputer untuk memproses teks mentah dan mengenali struktur, makna, dan konteks dari teks tersebut.

2 Pengolahan Informasi yang Efisien

Dalam era digital yang penuh dengan informasi, text processing membantu dalam pengolahan informasi yang efisien. Dengan teknik-teknik seperti tokenisasi dan penghapusan tanda baca, teks dapat diubah menjadi bentuk yang lebih terstruktur, memungkinkan komputer untuk dengan mudah mencari, menyaring, dan mengambil informasi yang relevan.

3 Analisis dan Pemahaman Teks yang Lebih Mendalam

Text processing memungkinkan analisis yang lebih mendalam terhadap teks. Dengan menggunakan teknik seperti analisis semantik, pengenalan entitas bernama, dan parsing sintaksis, komputer dapat memahami struktur dan makna dari teks tersebut. Hal ini membuka peluang untuk melakukan analisis sentiment, pengenalan topik, ekstraksi informasi, dan tugas-tugas NLP lainnya.

4 Pengembangan Aplikasi NLP yang Lebih Canggih

Text processing merupakan langkah penting dalam pengembangan aplikasi dan sistem yang berbasis NLP. Dengan memahami dan memproses teks dengan baik, komputer dapat berinteraksi dengan bahasa manusia secara efektif. Contohnya termasuk asisten virtual yang dapat memahami perintah lisan, sistem penerjemahan mesin yang dapat menerjemahkan teks antar bahasa, atau sistem analisis sentimen yang dapat menganalisis opini dari ulasan pengguna.

Dengan demikian, text processing penting dilakukan karena memungkinkan komputer untuk memahami, menganalisis, dan mengolah teks secara komputasional. Hal ini membuka peluang

untuk berbagai aplikasi dan sistem yang dapat berinteraksi dengan bahasa manusia dengan lebih baik, meningkatkan efisiensi pengolahan informasi, dan memungkinkan analisis yang lebih mendalam terhadap teks.