



**Indonesia AI**  
AI for Everyone, AI for Indonesia

**# 2023JAGO**  
**TEKNOLOGIAI**



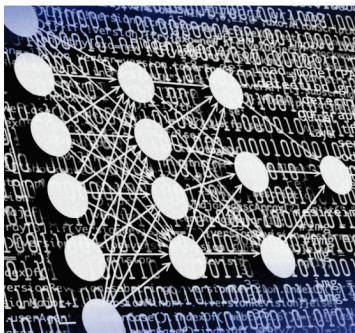
**AI Career Bootcamp**

# **Prediction-Based Text Vectorization**

## 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



Prediction-Based Text Vectorization adalah teknik dalam NLP yang menggunakan model prediksi untuk menghasilkan vektor representasi numerik dari teks. Metode ini melibatkan pelatihan model pembelajaran mesin, seperti model bahasa berbasis neural networks, untuk mempelajari pola dan hubungan antara kata-kata dalam teks. Model ini diajarkan untuk memprediksi kata berikutnya berdasarkan konteks teks sebelumnya.

Setelah melatih model, vektor representasi kata-kata atau frasa dapat diperoleh dengan menggunakan layer atau bobot internal dari model tersebut. Representasi vektor ini mencerminkan pemahaman model tentang kata-kata dan hubungan antara mereka dalam konteks teks. Dengan demikian, setiap kata atau frasa akan diwakili oleh vektor numerik yang merepresentasikan pemodelan prediksi dari teks tersebut.

Kelebihan utama dari Prediction-Based Text Vectorization adalah kemampuannya untuk menangkap konteks dan hubungan yang lebih kompleks antara kata-kata dalam teks. Metode ini memanfaatkan pemodelan prediksi yang canggih untuk menghasilkan vektor representasi yang mengandung informasi yang lebih kaya tentang teks. Representasi ini dapat digunakan untuk berbagai tugas NLP seperti analisis sentiment, klasifikasi teks, atau pemodelan topik dengan akurasi yang lebih tinggi.

Namun, ada beberapa tantangan yang terkait dengan Prediction-Based Text Vectorization. Pelatihan model prediksi yang kompleks dan memakan waktu dapat menjadi kendala dalam implementasi. Selain itu, representasi vektor yang dihasilkan oleh metode ini dapat memiliki dimensi yang besar, yang memerlukan sumber daya komputasi yang lebih tinggi untuk pengolahan dan analisis lebih lanjut.

## Prediction-Based vs Frequency-Based

Prediction-Based Text Vectorization memiliki keunggulan yang signifikan dibandingkan dengan Frequency-Based Text Vectorization. Salah satu keunggulan utamanya adalah kemampuannya untuk menangkap konteks yang lebih kompleks dalam teks. Dengan menggunakan model prediksi, vektor representasi kata-kata atau frasa dapat mencerminkan pemahaman yang lebih dalam tentang hubungan antara kata-kata dalam konteks teks. Ini memungkinkan representasi yang lebih kaya dalam memahami makna kata yang bergantung pada konteks dan memperhitungkan urutan kata-kata yang penting. Sebagai contoh, Prediction-Based Text Vectorization dapat mengenali perbedaan dalam makna kata "bank" sebagai entitas keuangan atau fisik berdasarkan konteks yang lebih luas.

Selain itu, Prediction-Based Text Vectorization juga memiliki keunggulan dalam menangani kata-kata jarang atau belum pernah ditemui. Melalui pelatihan model prediksi, representasi vektor dapat mempelajari pola kata-kata yang jarang muncul dan mengaitkannya dengan konteks yang relevan. Dengan demikian, metode ini dapat memahami kata-kata yang tidak umum dengan lebih baik dan memberikan representasi vektor yang lebih informatif. Ini memungkinkan analisis teks yang lebih akurat dan bermakna, terutama dalam tugas-tugas seperti identifikasi topik, analisis sentiment, atau pemodelan bahasa yang memerlukan pemahaman yang lebih dalam terhadap kata-kata dalam konteks.