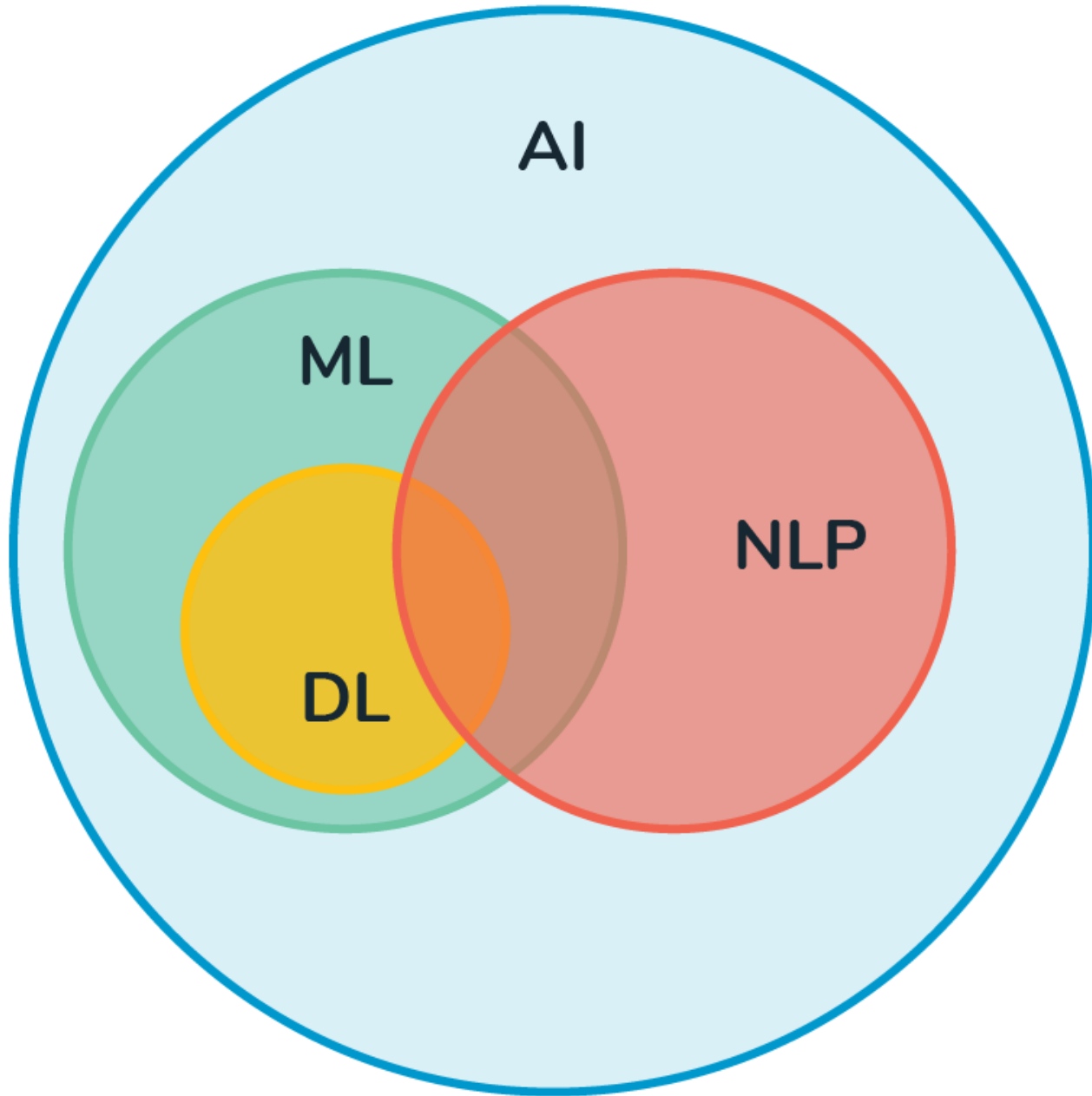




Pengantar Text Mining

Chapter 1

Prodi Sains Data



Definisi

- **Text Mining :**
 - Proses ekstraksi informasi dari teks tidak terstruktur dengan menggunakan teknik statistik, pembelajaran mesin, dan pemrosesan bahasa alami yang bertujuan mengubah teks mentah menjadi data yang dapat digunakan untuk analisis lebih lanjut.
- **Natural Language Processing :**
 - Cabang kecerdasan buatan yang memungkinkan komputer memahami, menginterpretasikan, dan menghasilkan bahasa manusia.
 - NLP menggabungkan linguistik komputasional dengan teknik machine learning untuk mengolah teks dan ucapan secara otomatis

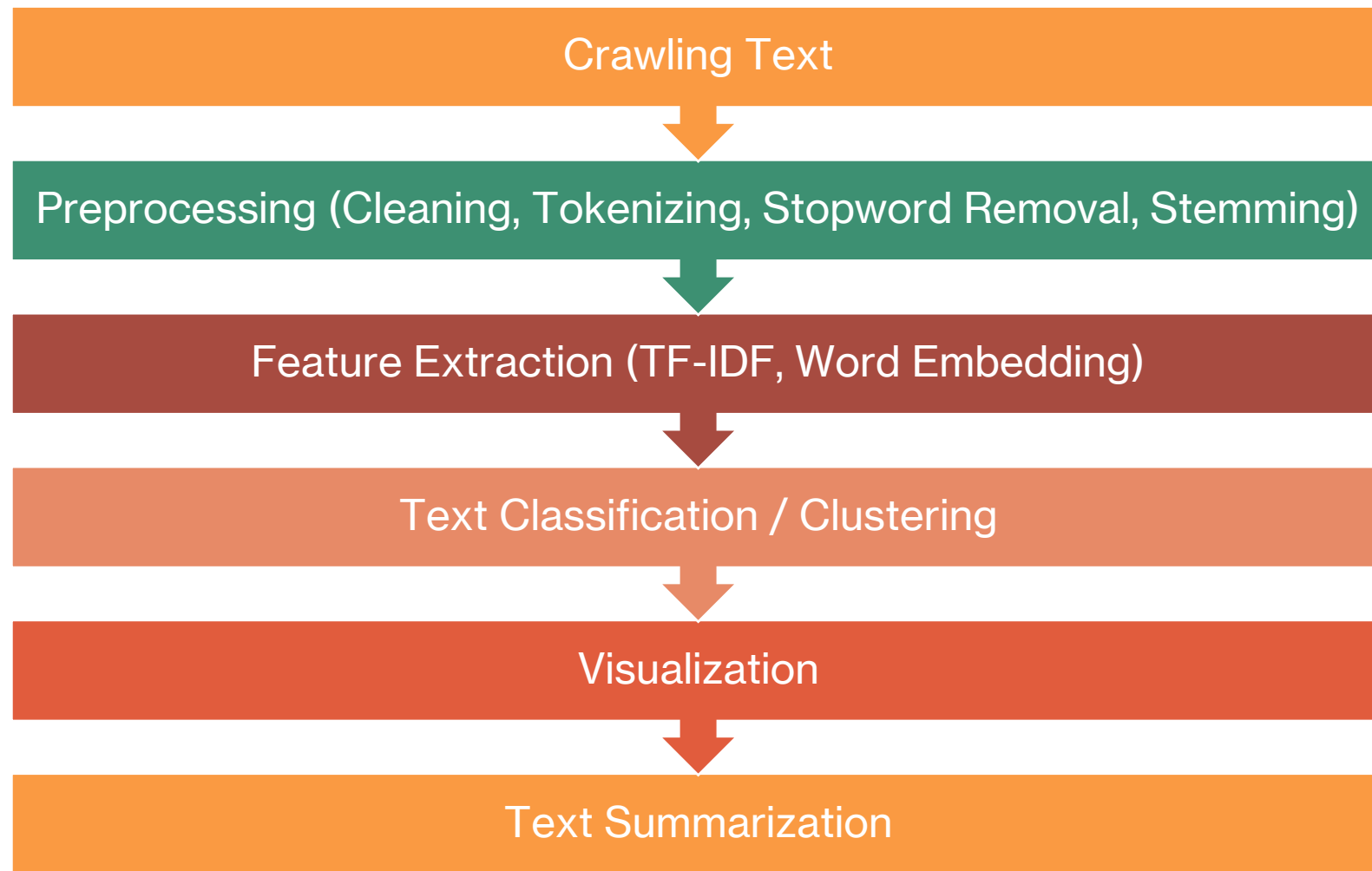
Penggunaan NLP



Penggunaan NLP, Lanjutan



Proses NLP





Crawling Text

CRAWLING TEXT

- PROSES OTOMATIS UNTUK MENGUMPULKAN DATA TEKS DARI BERBAGAI SUMBER SEPERTI SITUS WEB, MEDIA SOSIAL, DAN DOKUMEN DIGITAL.



Preprocessing Text

Tahapan Preprocessing Data Text

Cleaning → Pembersihan karakter khusus, angka, hastag, mention, dll



Tokenizing → Memotong kalimat menjadi kata yang dibuat dalam kolom



Stopword Removal → Menghapus kata yang tidak bermakna



Stemming dan Lematization → Merubah menjadi kata dasar



Terimakasih