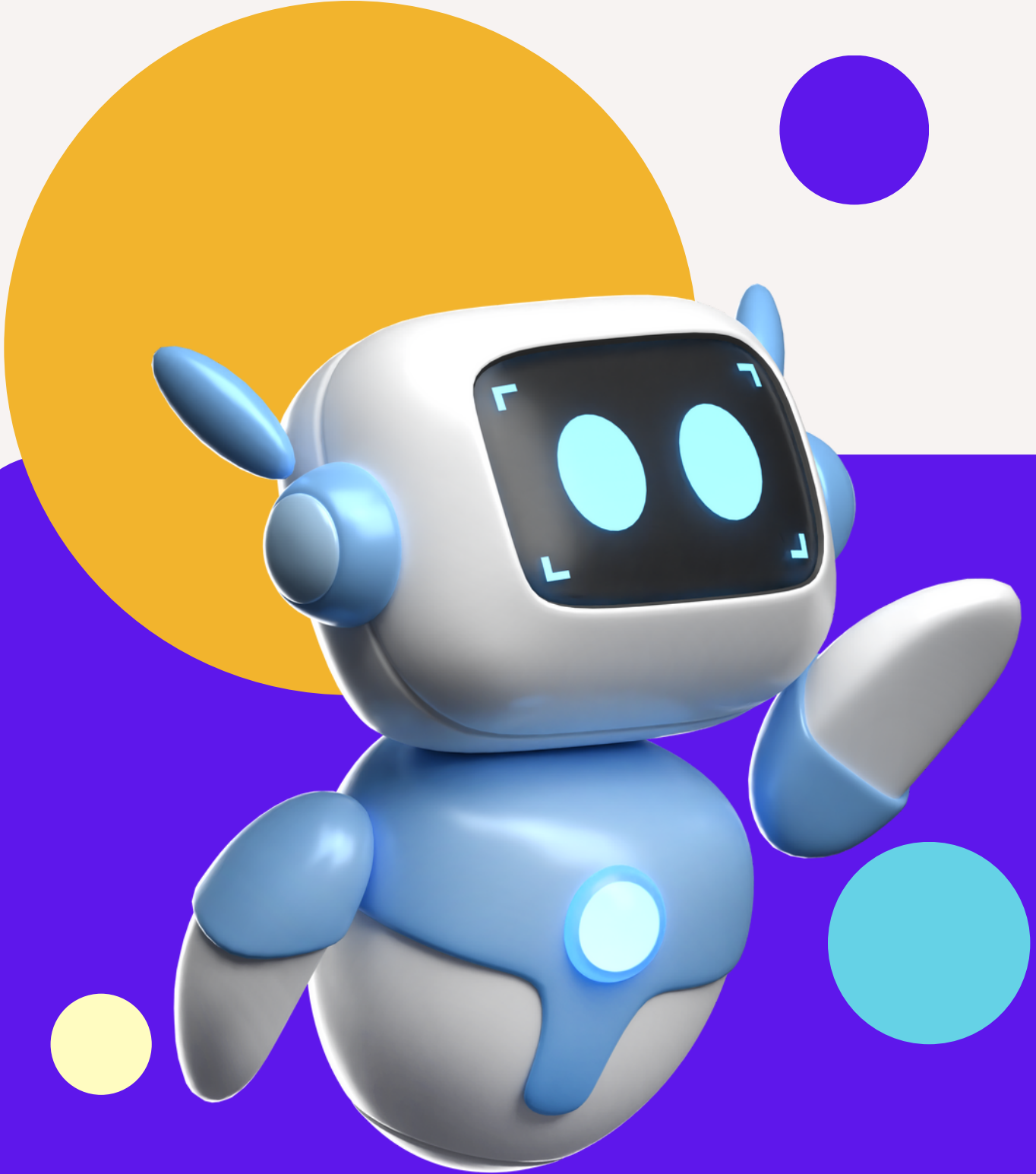




IFB 304 - Sistem Pakar & Bahasa Alamiah

Stemming & Lemmatization

Chalifa Chazar

Course Outline

1

Dasar-dasar sistem pakar dan karakteristik utama sistem pakar.

2

Knowledge Engineering

3

Model Inferensi

4

Faktor Ketidakpastian (Uncertainty factor) dan kepastian (Certainty factor)

5

Perangkat Lunak (Tools) untuk sistem pakar dan proyek akhir sistem pakar

6

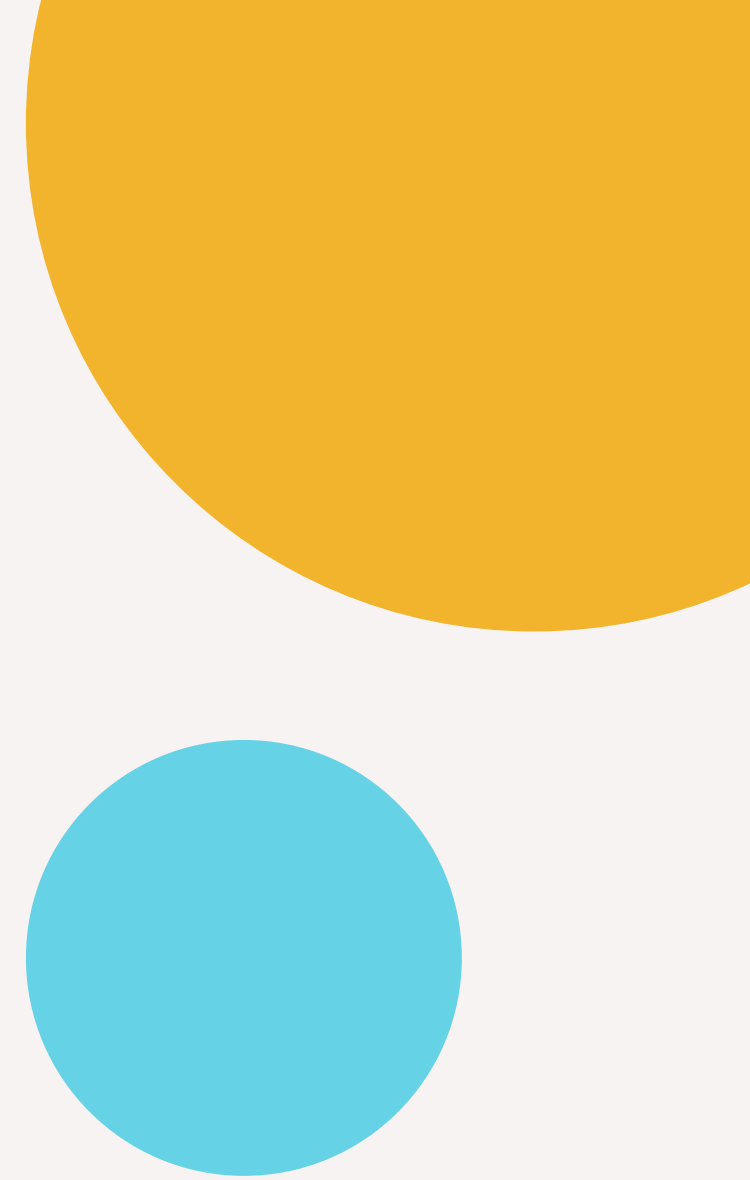
Pemrosesan sintaks dan Interpretasi atau penterjemahan semantik

7

Pembuatan perangkat lunak Bahasa Alamiah untuk proyek akhir bahasa alamiah

Stemming

- Teknik dalam pemrosesan bahasa alami (NLP) yang digunakan untuk mengubah kata-kata menjadi bentuk dasarnya atau akar kata (stem)
- Proses pemotongan akhiran atau imbuhan (prefiks, infix dan suffix) dari sebuah kata untuk mendapatkan bentuk dasar kata tersebut
- Tujuannya adalah untuk menyederhanakan kata-kata yang memiliki makna serupa agar dapat diproses lebih lanjut dalam aplikasi NLP seperti pencarian informasi, analisis sentimen, dan klasifikasi teks



Contoh Stemming

Bahasa Indonesia

- mempelajari → pelajar
- berjalan → jalan
- pendidikan → pendidik

Bahasa Inggris

- running → run
- doing → do
- happiness → happy

Penerapan Stemming dalam NLP

Pencarian Informasi (Information Retrieval)

- Stemming digunakan untuk menyamakan kata-kata yang memiliki akar yang sama agar pencarian lebih efektif.
- Misalnya, pencarian dengan kata "berlari" akan menemukan dokumen dengan kata "lari."

Klasifikasi Teks

- Menggunakan stemming untuk mengurangi dimensi data dan meningkatkan akurasi model klasifikasi.

Penerapan Stemming dalam NLP

Analisis Sentimen

- Dengan mengurangi kata-kata menjadi bentuk dasar, analisis sentimen akan lebih fokus pada makna inti dari kata tersebut.

Penerjemahan Mesin (Machine Translation)

- Stemming membantu dalam menyederhanakan bentuk kata sehingga model dapat memproses teks lebih efisien.

Algoritma Stemming

- Porter Stemmer
- Snowball Stemmer
- Lancaster Stemmer



Porter Stemmer

- Porter Stemmer adalah algoritma stemming yang paling dikenal dan digunakan untuk Bahasa Inggris
- Algoritma ini mengurangi kata-kata berdasarkan serangkaian aturan yang dirancang untuk memotong akhiran kata

Contoh:

- running → run
- happiness → happi
- played → play
- doing → do

Snowball Stemmer

- Snowball Stemmer adalah versi yang lebih baru dan lebih canggih dari Porter Stemmer
- Snowball Stemmer mendukung banyak bahasa, termasuk Bahasa Inggris dan Bahasa Indonesia
- Snowball Stemmer juga lebih ringan dan lebih cepat dibandingkan Porter Stemmer dalam beberapa kasus

Contoh:

- running → run
- happiness → happi
- playing → play
- tried → tri

Lancaster Stemmer

- Lancaster Stemmer adalah algoritma stemming yang lebih agresif dibandingkan Porter atau Snowball Stemmer
- Algoritma ini mengurangi kata-kata ke bentuk yang lebih singkat, dan terkadang hasil stemming menjadi tidak terduga

Contoh:

- running → run
- happiness → happi
- playing → play
- better → bet
- relational → relat

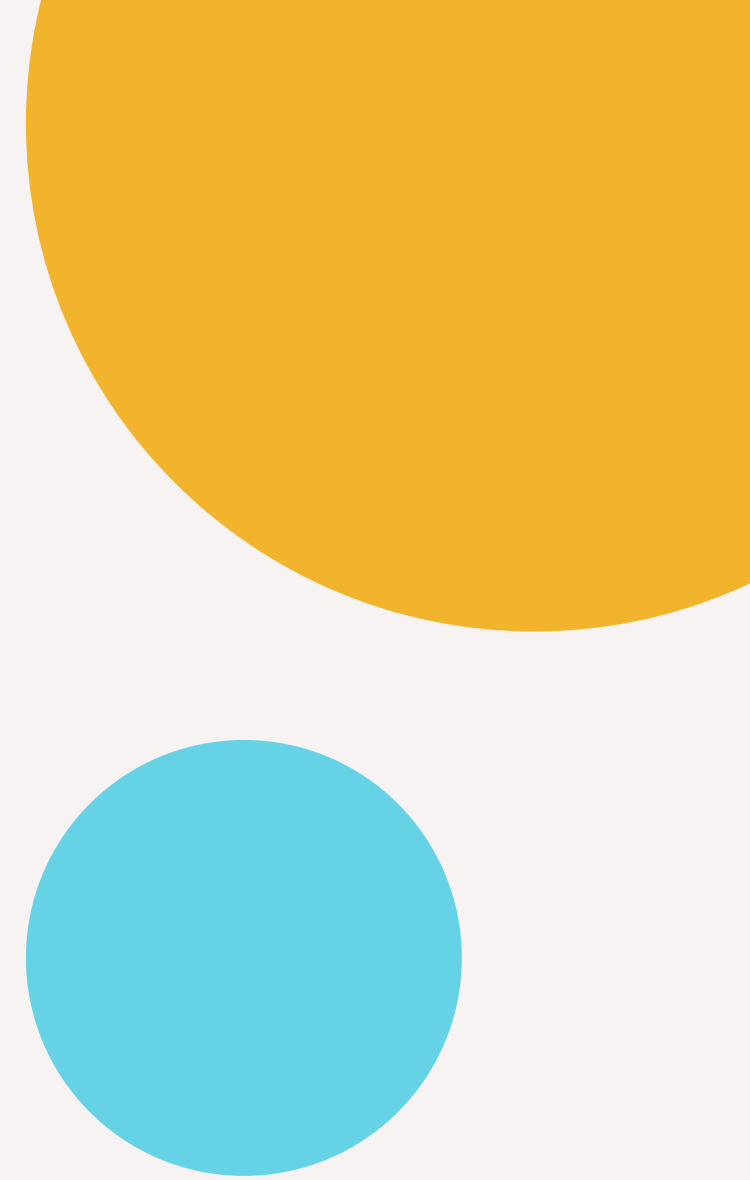
Kelebihan dan Kekurangan Stemming

Kelebihan:

- **Efisiensi:** Mengurangi variasi kata, sehingga memudahkan pemrosesan teks.
- **Sederhana:** Algoritma stemming umumnya lebih sederhana dan lebih cepat dibandingkan lemmatization.

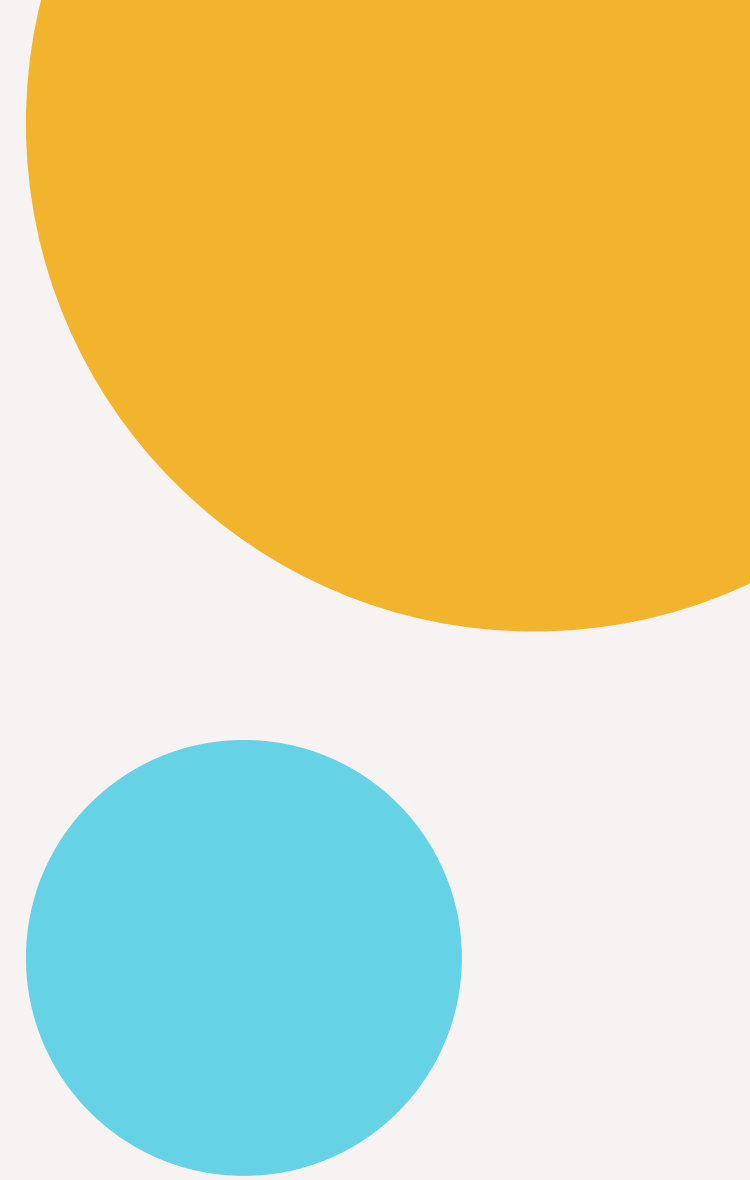
Kekurangan:

- **Tidak selalu akurat:** Stemming dapat menghasilkan bentuk kata yang tidak sesuai dengan tata bahasa.
- **Mengurangi konteks:** Proses stemming bisa menghilangkan informasi yang relevan mengenai bentuk kata.



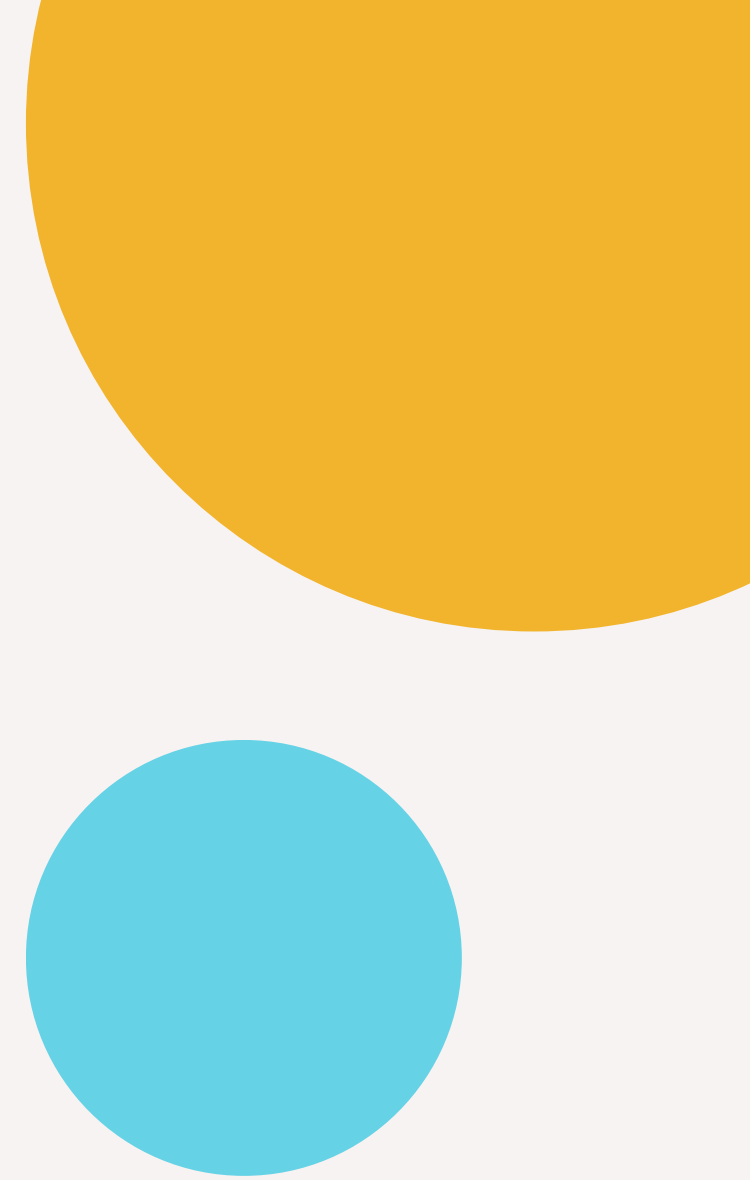
Perbedaan Stemming & Lemmatization

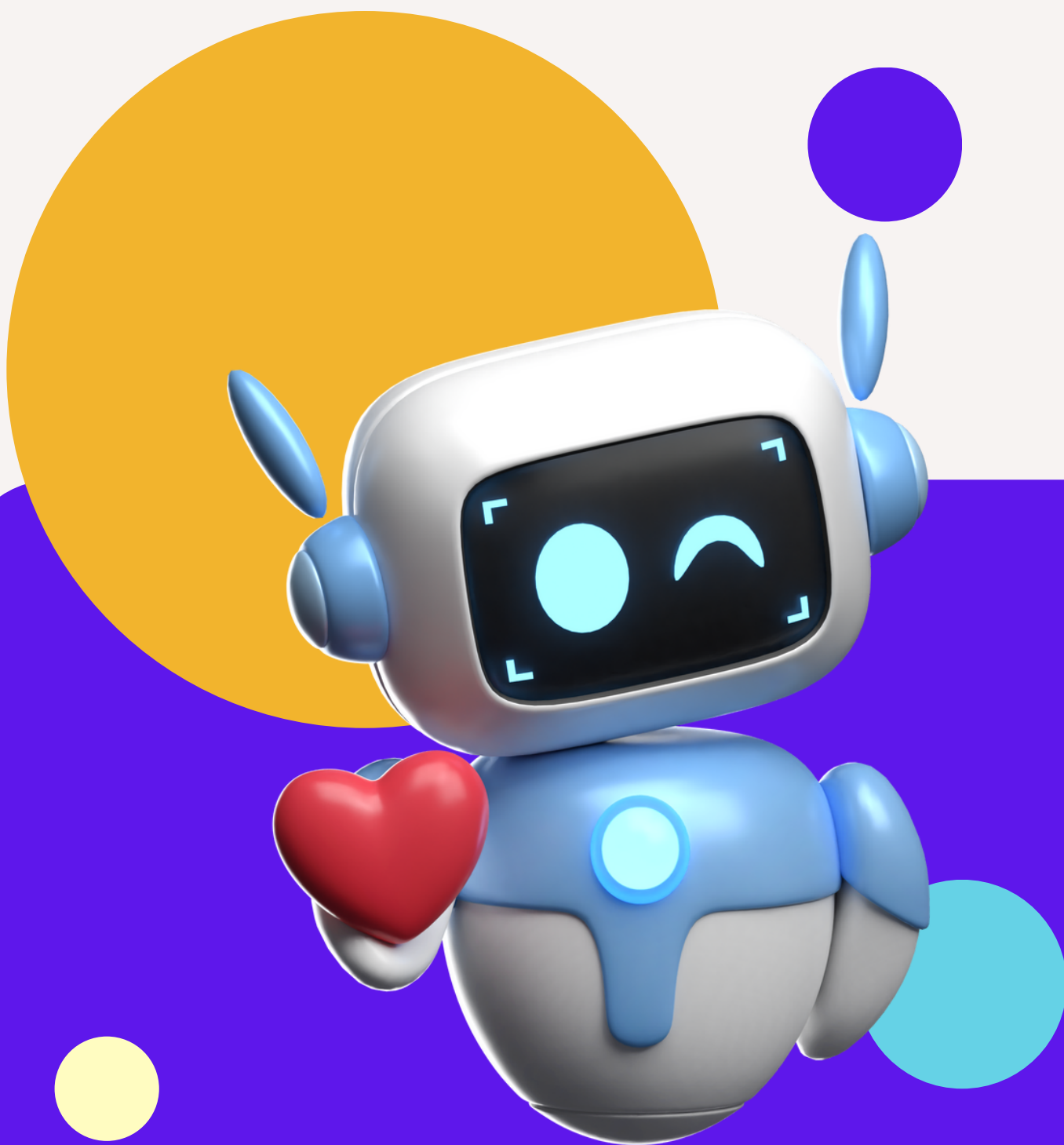
- **Stemming:** Memotong akhiran atau imbuhan dari kata untuk mendapatkan bentuk dasar. Proses ini bisa lebih agresif dan terkadang menghasilkan kata yang tidak sesuai dengan bahasa
- **Lemmatization:** Mengubah kata ke bentuk dasarnya dengan mempertimbangkan konteks kalimat dan aturan tata bahasa. Proses ini lebih kompleks dan akurat dibandingkan stemming



Stemming vs Lemmatization

- Gunakan **Stemming** ketika Anda ingin **meningkatkan efisiensi** pemrosesan teks dan **mengurangi dimensi data**, terutama dalam kasus di mana keakuratan kata tidak terlalu penting. Stemming sering digunakan dalam pencarian informasi, analisis sentimen, dan sistem rekomendasi.
- Gunakan **Lemmatization** ketika **akurasi kata sangat penting**, terutama dalam kasus di mana bentuk kata yang tepat sangat diperlukan (misalnya, dalam analisis teks yang lebih mendalam atau model klasifikasi teks yang memerlukan pemahaman konteks kata).





See u next..

Thanxs!

chalifa.id