



# یادگیری عمیق (Deep Learning)

سعید محقق / زمستان 1400

NLP: Natural Language Processing

مقدمه چرا NLP ؟

چالش های داده متنی

آماده سازی داده متنی

کدنویسی

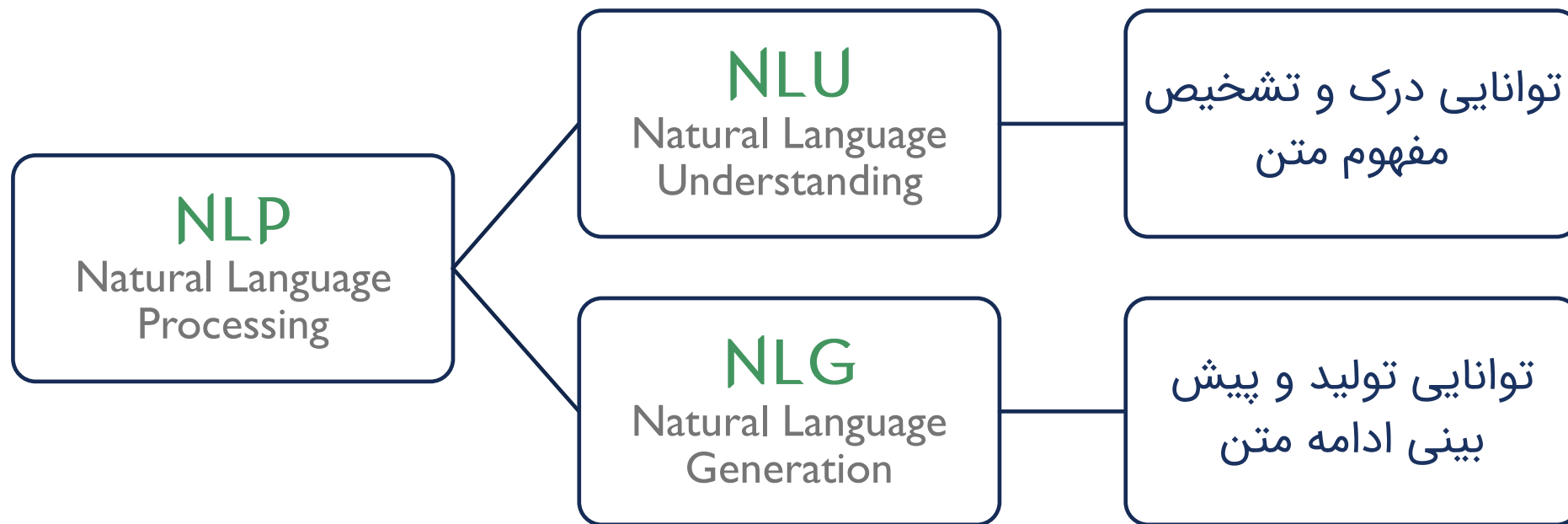
1- تبدیل کلمات به اعداد (Tokenization)

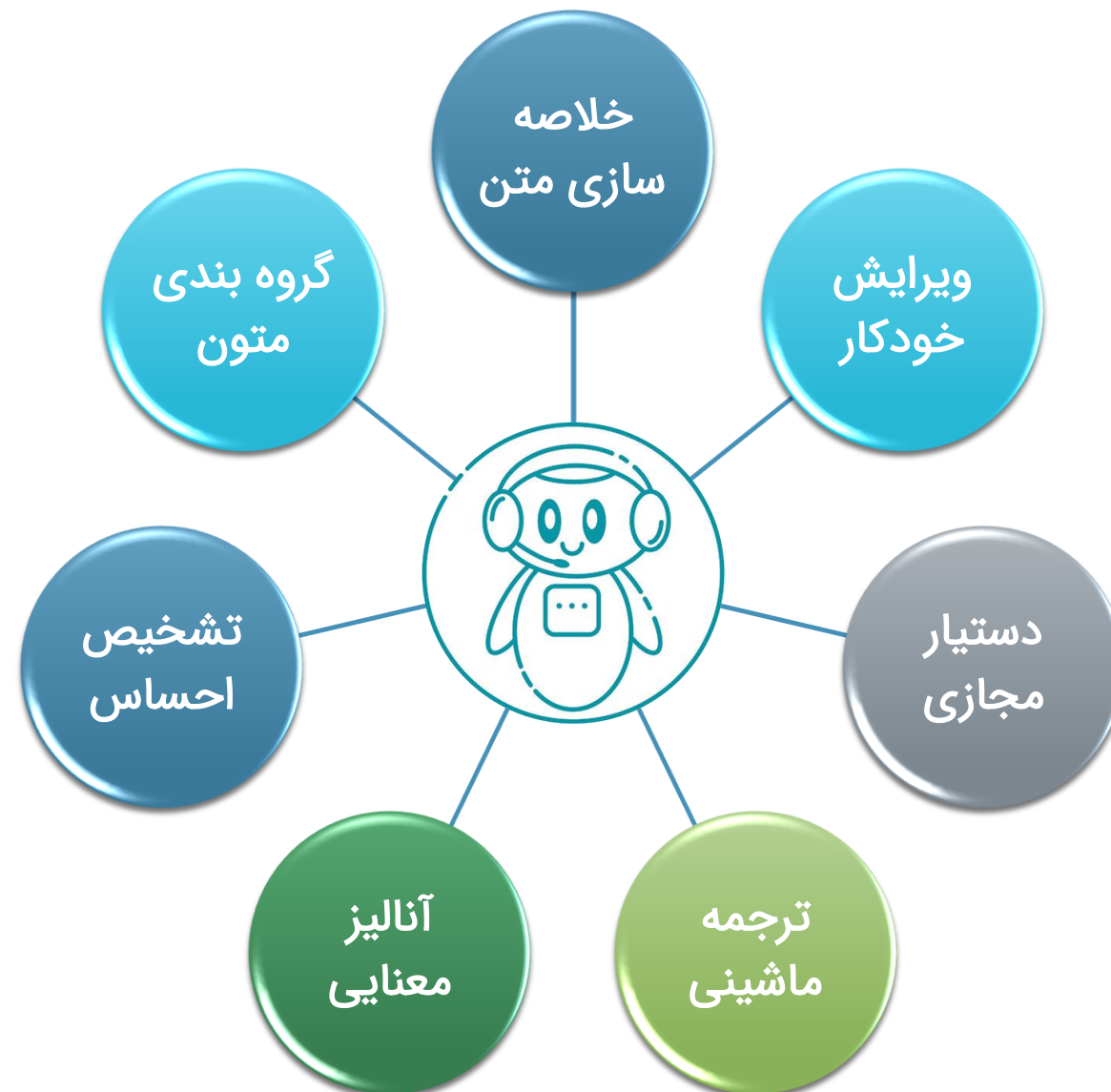
2- تبدیل جملات به بردارهای عددی (Sequencing)

3- آموزش مدل برای تشخیص کلاس (Training)

# 1. مقدمه و معرفی

شناخت داده های متنی و چالش ها





### قوانین و قواعد ثابت و مشخص نداریم

- قوانین سطح پایین: نقطه، علامت، فعل، سوال و ...
- قوانین سطح بالا: طنز، کنایه، اشاره و ...

### لغات دارای معانی متعدد و متفاوت

- مثلاً لغت match را در دیکشنری ببینید !

### شکل های مختلف متن

- ایمیل، مقاله، نظر، محاوره ای، کتابی و ...

### داده های غیر عددی و بدون ساختار

- مدل یادگیری عمیق فقط با داده عددی آموزش می بیند

## پیش پردازش متن

- استانداردسازی / ریشه یابی / گروه بندی / ...

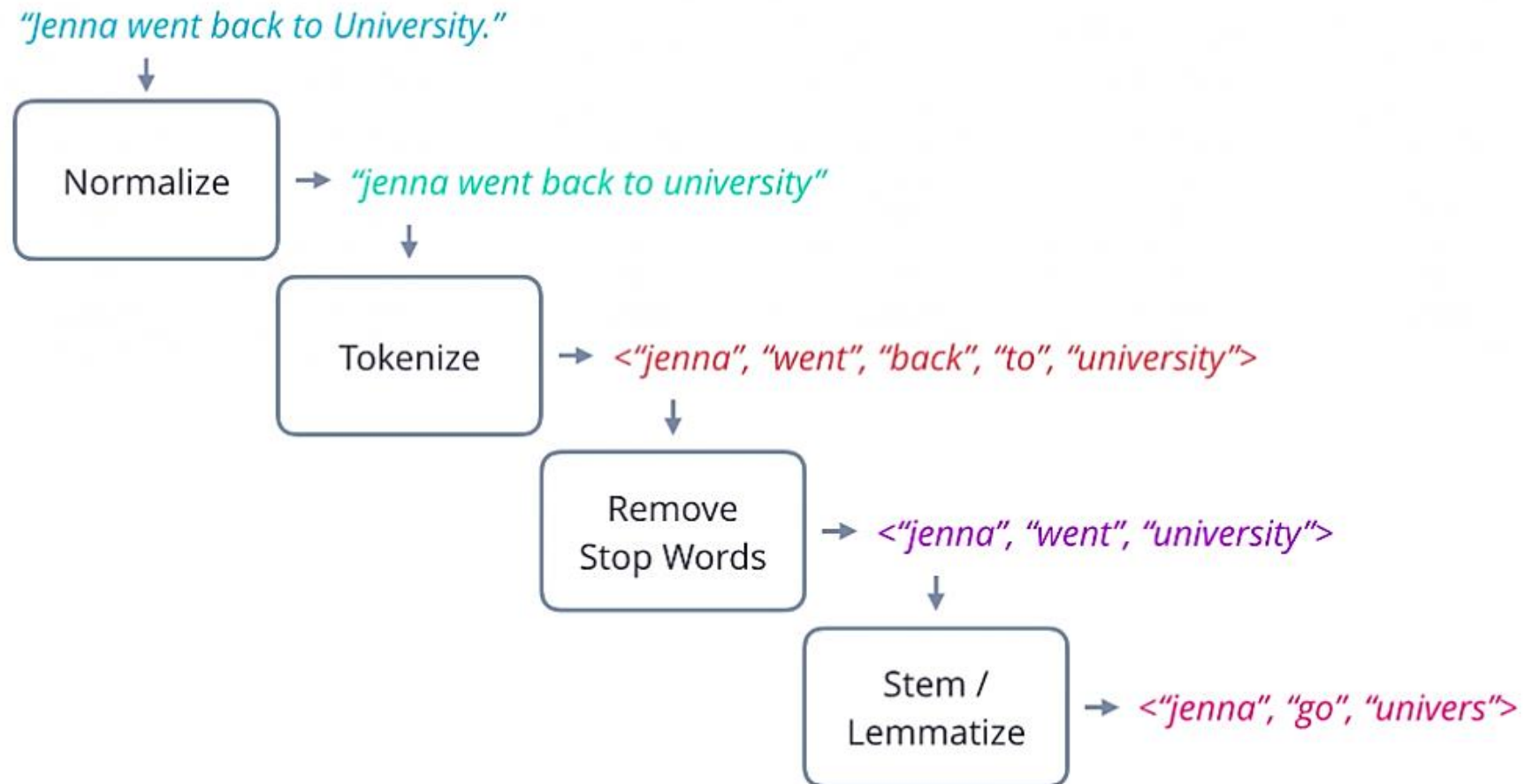
## تبدیل نوع داده

- تبدیل کلمات به اعداد یا بردارهای عددی

## طراحی لایه ها و مدل های متناسب

- لایه ها و مدل های یادگیری عمیق مناسب برای بردارهای متنی

حوزه Data Science





## 2. کدنویسی در عمل

آماده سازی داده و آموزش مدل برای طبقه بندی داده متنی

- اسلایدها و کدهای NLP در گیتهاب: [https://github.com/DaneshJoy/DeepLearning\\_NLP/](https://github.com/DaneshJoy/DeepLearning_NLP/)
- کتاب های یادگیری عمیق در گیتهاب: <https://github.com/daneshjoy/dl-books/>
- مثال های NLP در سایت Keras: <https://keras.io/examples/nlp/>