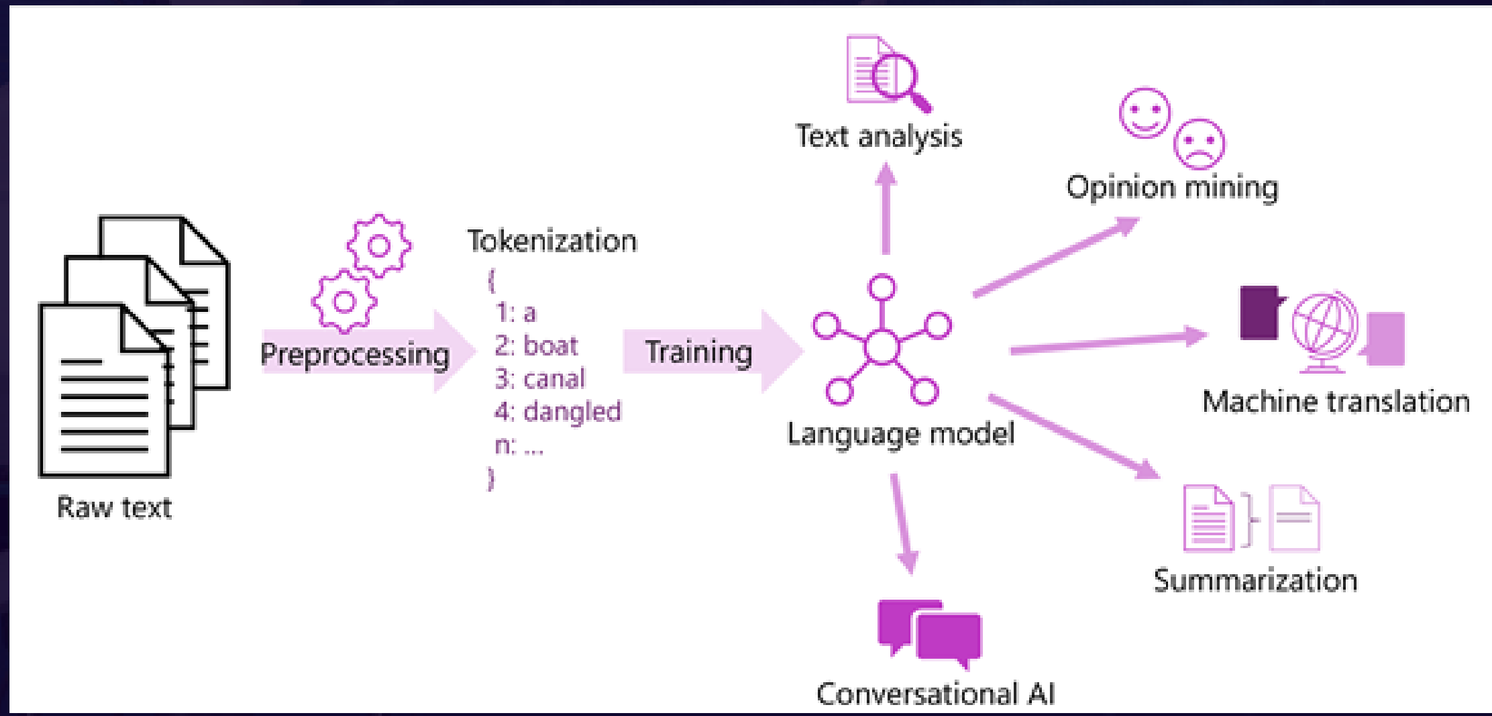
Natural Language Processing

Presented by Group 5

Lidia
Yusuf
Laher
Kownain
Komal
Leon
Kenil

Introduction

- Understand Text Analytics
  1. Tokenization
  2. Frequency analysis
- Machine learning for text classification
- Semantic language models



Tokenization

Is a process of breaking the corpus into smaller pieces.

Text Normalization

Removing punctuation
&
Upper case

Stop Word Removal

Removing Little semantic meaning words
“the, a, it ”

N-grams

Number of phrases in the token
“I” unigram
“I am” bi-gram
“in the morning” tri-gram

Steaming Words

Combining the same words forms one:
“power,
powered,
powering”

Frequency Analysis

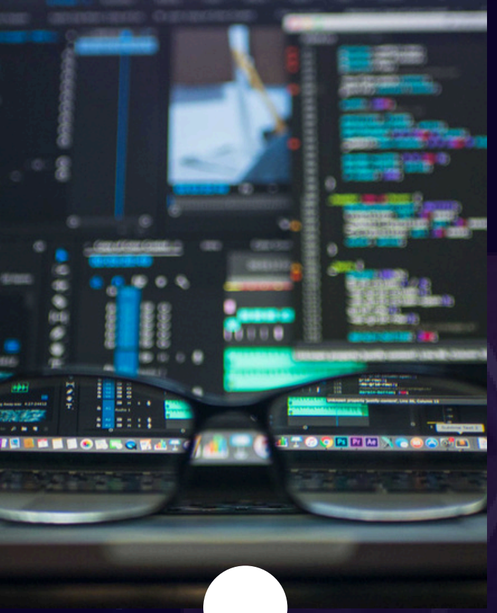
Count the number of occurrences of each token

Term frequency-inverse document frequency (TF-IDF)

Machine Learning for Text Classification

What is **Text Classification**?

- Fundamental task in natural language processing (NLP)
- Involves assigning predefined categories to text data



Data Collection

Gather relevant data from diverse sources.

Preprocessing

Clean, normalize, and engineer data.

Model Selection

Choose and train ML/DL algorithms for peak performance.

Training and Evaluation

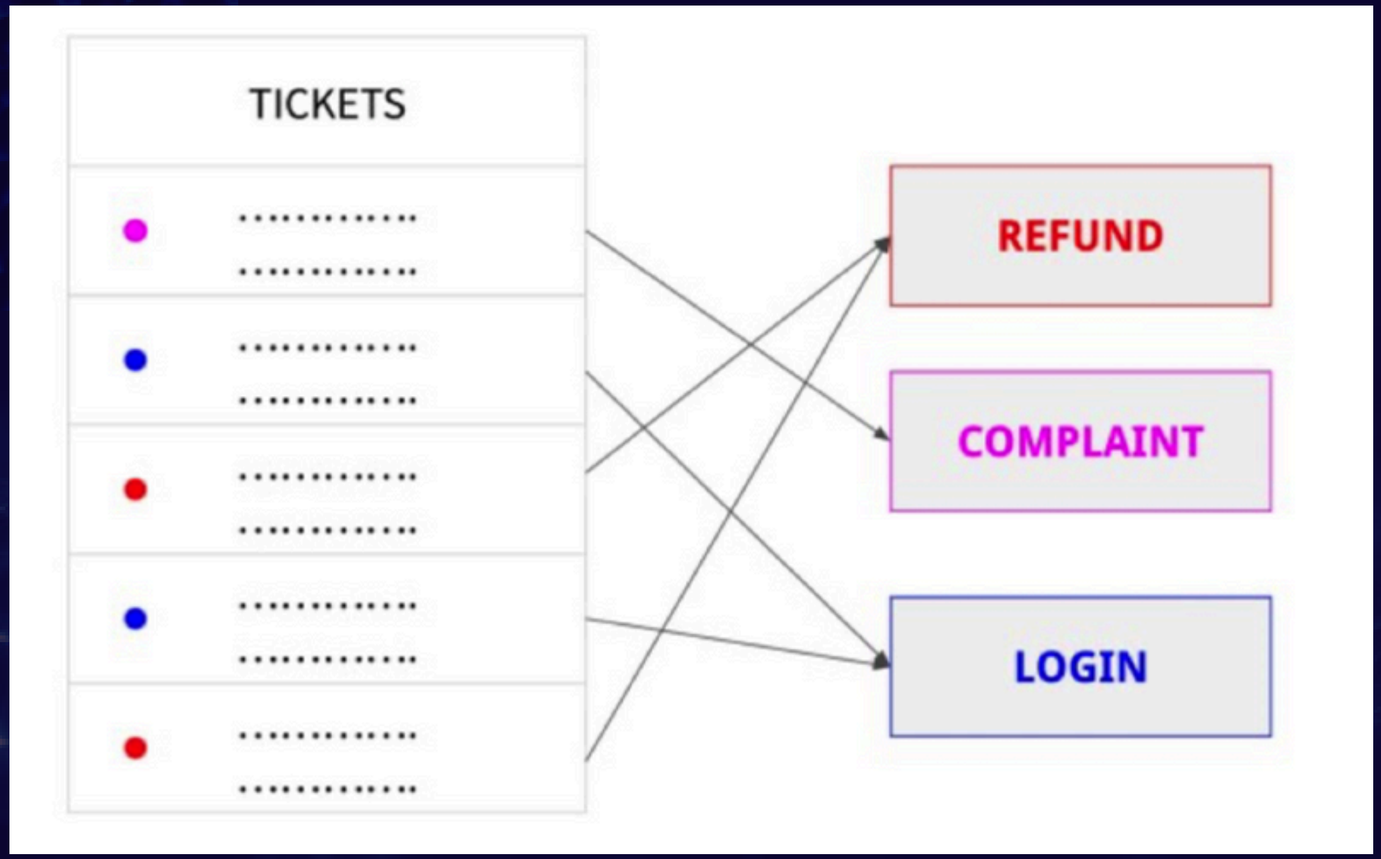
Analyzing training programs and results

Deployment and Monitoring

Monitoring your ML models

Methodology

Use case



TICKET ROUTING

Other Use Cases

- Spam Detection
- Sentiment Analysis
- Topic Classification
- Document Classification
- Ticket Routing
- Intent Recognition
- Survey Response Analysis
- Content Moderation
- Chatbots

Tools and Libraries

Pattern

Polyglot

Google Cloud Natural Language API

NLTK

Apache OpenNLP

PyTorch

SpaCy

JaTeCS

MonkeyLearn

Hugging Face Transformers

Mallet

Scikit-Learn

Gensim

Orange

Naive Bayes

TensorFlow

CoreNLP

TextBlob

OpenNLP

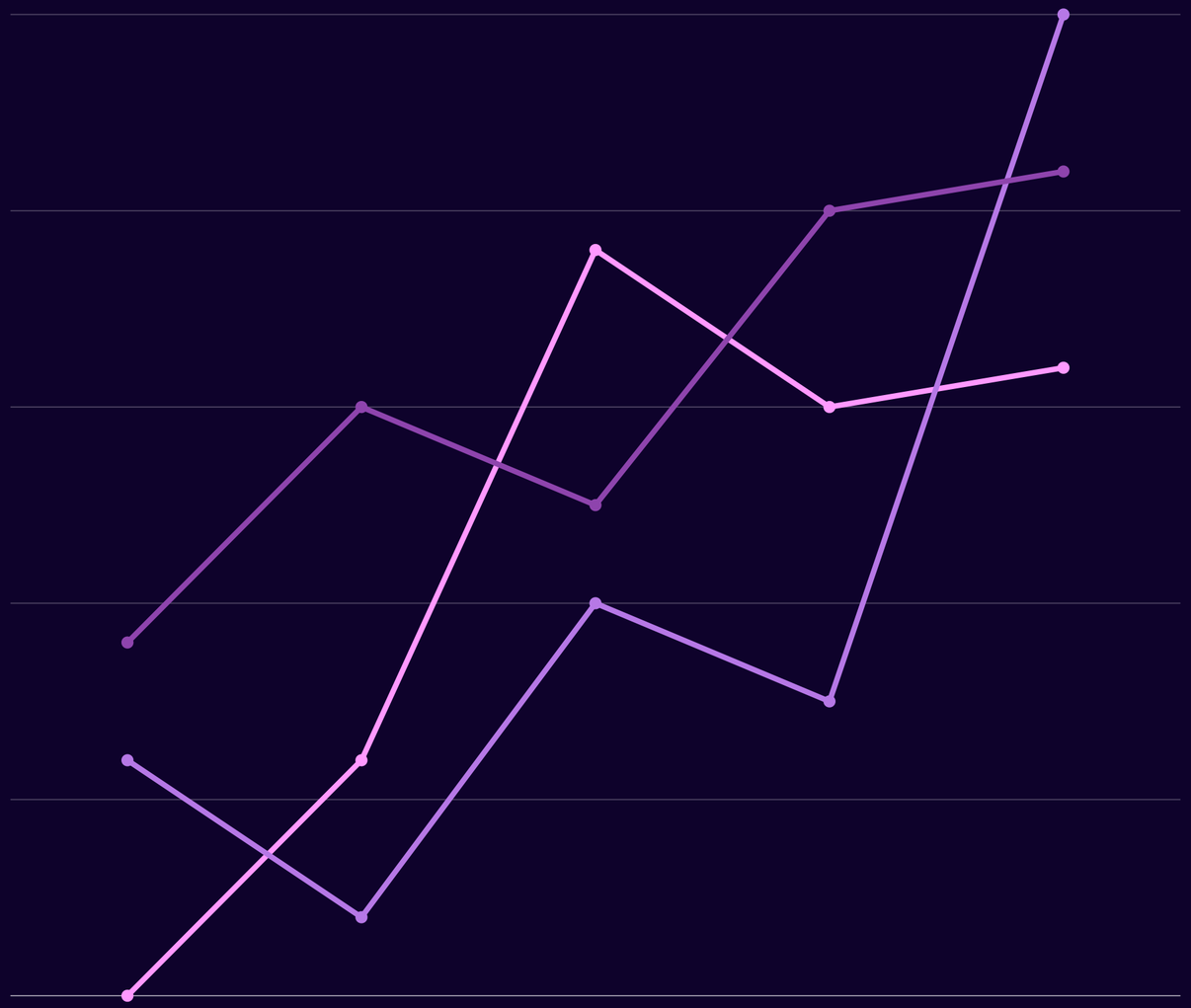
Caret for R

Amazon Comprehend

Lexalytics

AYLIEN

Semantic language models



- What is Semantics ?
- What is language models ?

EXAMPLE

"DESTINATION" = "LAST STOP"

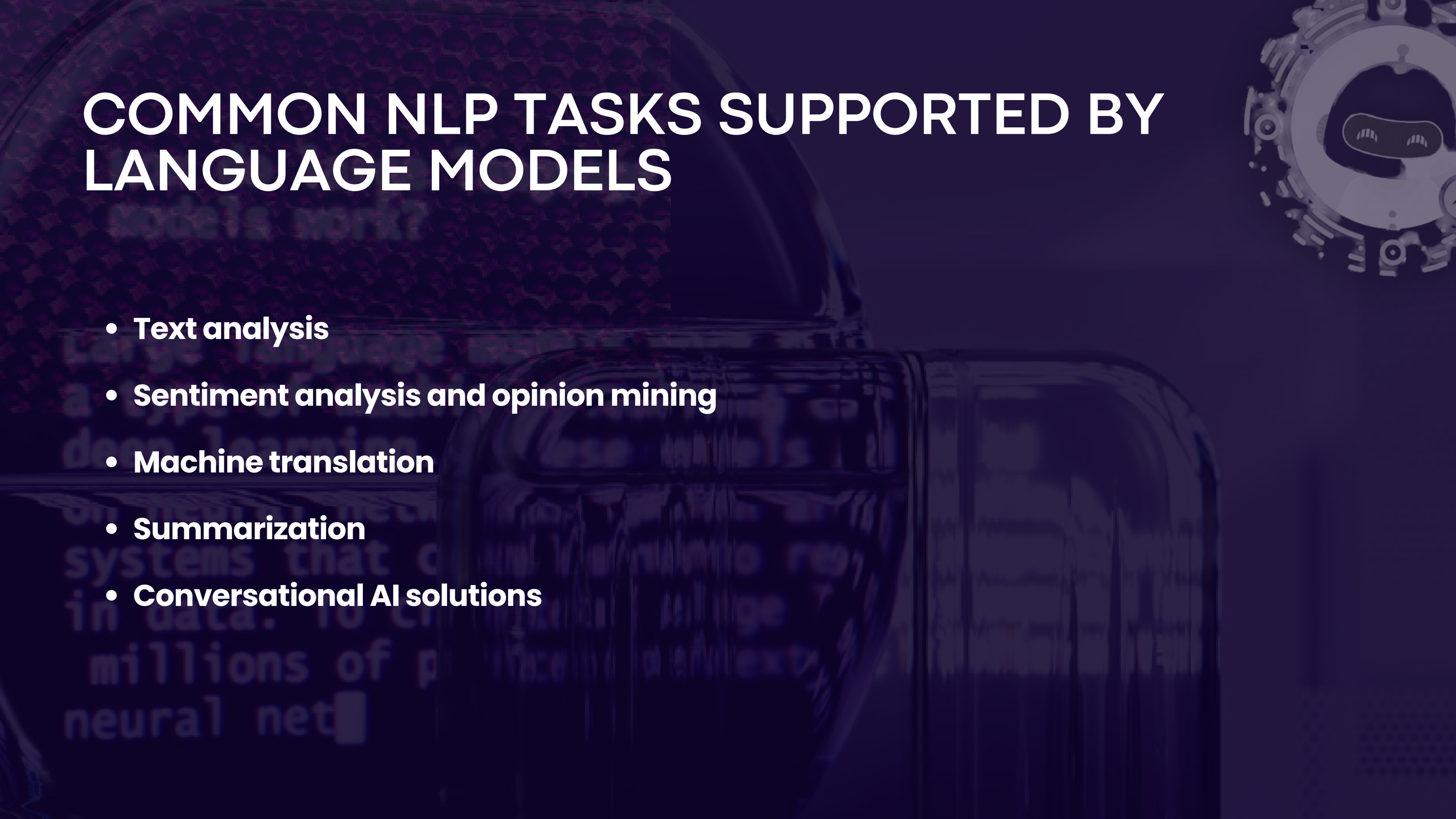
"'crash'= 'accident' = 'drop in the stock market' = 'attending a party without an invitation'

"Jenny dropped by the office for the keys so I gave them to [...],"

CHATGPT- LANGUAGE MODEL

- AI language model developed by OpenAI.
- Designed to understand and generate human-like text based on the input
- Training Data, Knowledge Base, Text Generation
- Input and Output>Natural Language Processing (NLP)>Interactivity

COMMON NLP TASKS SUPPORTED BY LANGUAGE MODELS

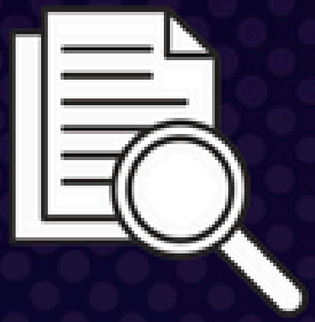


- Text analysis
- Sentiment analysis and opinion mining
- Machine translation
- Summarization
- Conversational AI solutions

Get started with text analysis

Azure AI Language's text analysis features include:

- **Named entity recognition**
- **Entity linking**
- **Personal identifying information (PII) detection**
- **Language detection**
- **Sentiment analysis and opinion mining**
- **Summarization**
- **Key phrase extraction**



Get started with text analysis

Entity recognition and linking

An example: *"We went to Seattle last week."*

Entity	Type	SubType	Wikipedia URL
Seattle	Location		https://en.wikipedia.org/wiki/Seattle
last week	DateTime	DateRange	

Type	SubType	Example
Person		"Bill Gates", "John"
Location		"Paris", "New York"
Organization		"Microsoft"
Quantity	Number	"6" or "six"
Quantity	Percentage	"25%" or "fifty percent"
Quantity	Ordinal	"1st" or "first"
Quantity	Age	"90 day old" or "30 years old"
Quantity	Currency	"10.99"
Quantity	Dimension	"10 miles", "40 cm"
Quantity	Temperature	"45 degrees"
DateTime		"6:30PM February 4, 2012"
DateTime	Date	"May 2nd, 2017" or "05/02/2017"
DateTime	Time	"8am" or "8:00"
DateTime	DateRange	"May 2nd to May 5th"
DateTime	TimeRange	"4pm to 7pm"
DateTime	Duration	"1 minute and 45 seconds"
DateTime	Set	"every Tuesday"
URL		"https://www.bing.com"
Email		"support@microsoft.com"
US-based Phone Number		"(312) 555-0176"
IP Address		"10.0.1.123"

Get started with text analysis

Language detection

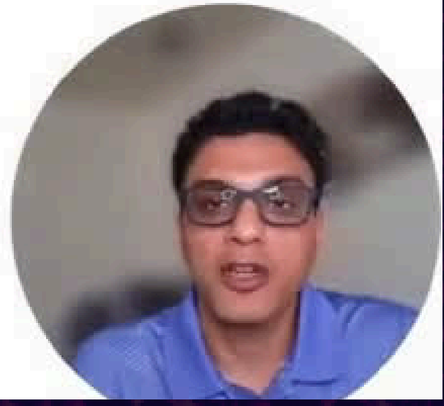
Review 1: *"A fantastic place for dinner. The rice was yummy."*
Review 2: "هذا مكان يحتاج مزيد كرسى!"
Review 3: ":)"

Language Name	ISO 6391 Code	Score
English	en	1.0
Arabic	ar	0.8
unknown	unknown	NaN

Microsoft

How to use Text Analytics for Health

Azure AI Language video series



Sentiment Analysis & Opinion Mining

- Natural Language Processing (NLP) Technique
- Azure AI Language
- Categories: Positive, Neutral, and Negative

Types of Sentiment Analysis

- Graded Sentiment Analysis
- Emotion Detection
- Aspect-based Sentiment Analysis
- Multilingual Sentiment Analysis

Advantages of Sentiment Analysis

- Sorting Data at Scale
- Real-Time Analysis
- Consistent Criteria

Applications of Sentiment Analysis

- Social Media Monitoring
- Customer Support Ticket Analysis
- Brand Monitoring & Reputation Management
- Product Analysis
- Market and Competitive Research

How Sentiment Analysis works

Sentiment analysis algorithms fall into one of three buckets:

- Rule-based
- Automatic
- Hybrid

Key Phrase Extraction

- Finds the most important words and phrases
- Simplifies the task of finding relevant words and phrases within unstructured data

Applications of Key Phrase Extraction

- Social Media / Brand Monitoring
- Customer Service / Feedback
- Business Intelligence
- Search Engine Optimization (SEO)
- Product Analytics
- Knowledge Management

Analyze text with Language Studio

- Create a Language or Azure AI services resource in your Azure subscription
- Configure your resource in Azure AI Language Studio
- Analyze reviews in Language Studio
- Clean up

Understanding Question Answering with Azure AI

Context and Need:

- **24/7 Communication:** Customers expect responses anytime, pressuring organizations to react quickly.
- **Personalized Responses:** Customers prefer personal answers over reading extensive documentation.
- **Overloaded Support Staff:** Multiple queries through various channels often overwhelm support teams.
- **Conversational AI:**
  Definition: Solutions enabling a dialog between an AI agent and a human.
- **Bots:** Generic term for conversational AI agents that interact through web chat, email social media, etc.

Azure AI Language's Question Answering:

- **Feature:** Provides the ability to create automated conversational AI solutions.
- **Capabilities:**

1. Immediate responses.
2. Accurate answers.
3. Natural, multi-turn interactions.

- **Platforms:** Can be implemented on websites, social media, and other platforms.
- **Benefits:**
- **Efficient Query Handling:** Users get answers quickly, reducing wait times.
- **Availability:** Provides answers anytime, beyond typical office hours.

Get started with the Language service and Azure Bot Service

You can easily create a user support bot solution on Microsoft Azure using a combination of two core services:

- **Azure AI Language:** includes a custom question-answering feature that enables you to create a knowledge base of question and answer pairs that can be queried using natural language input.

- **Azure AI Bot Service:** provides a framework for developing, publishing, and managing bots on Azure

Steps for creating a custom question answering knowledge base

Azure AI Language Studio:

- You can create a project, by selecting a **Language** resource in your Azure Subscription.
- **Define** questions and answers
- **Test/Run** the project
- Build a bot with **Azure AI Bot Service**
- **Connect Channels**

How To Use Language Studio To Make Custom QnA Maker

References

- Free Online Sentiment Analysis Tool: https://monkeylearn.com/sentiment-analysis-online/
- Free Online Keyword Extractor Tool: https://monkeylearn.com/keyword-extractor-online/
- https://monkeylearn.com/sentiment-analysis/
- https://monkeylearn.com/keyword-extraction/
- https://monkeylearn.com/blog/text-classification-machine-learning/
- https://levity.ai/blog/text-classifiers-in-machine-learning-a-practical-guide


Thank You!