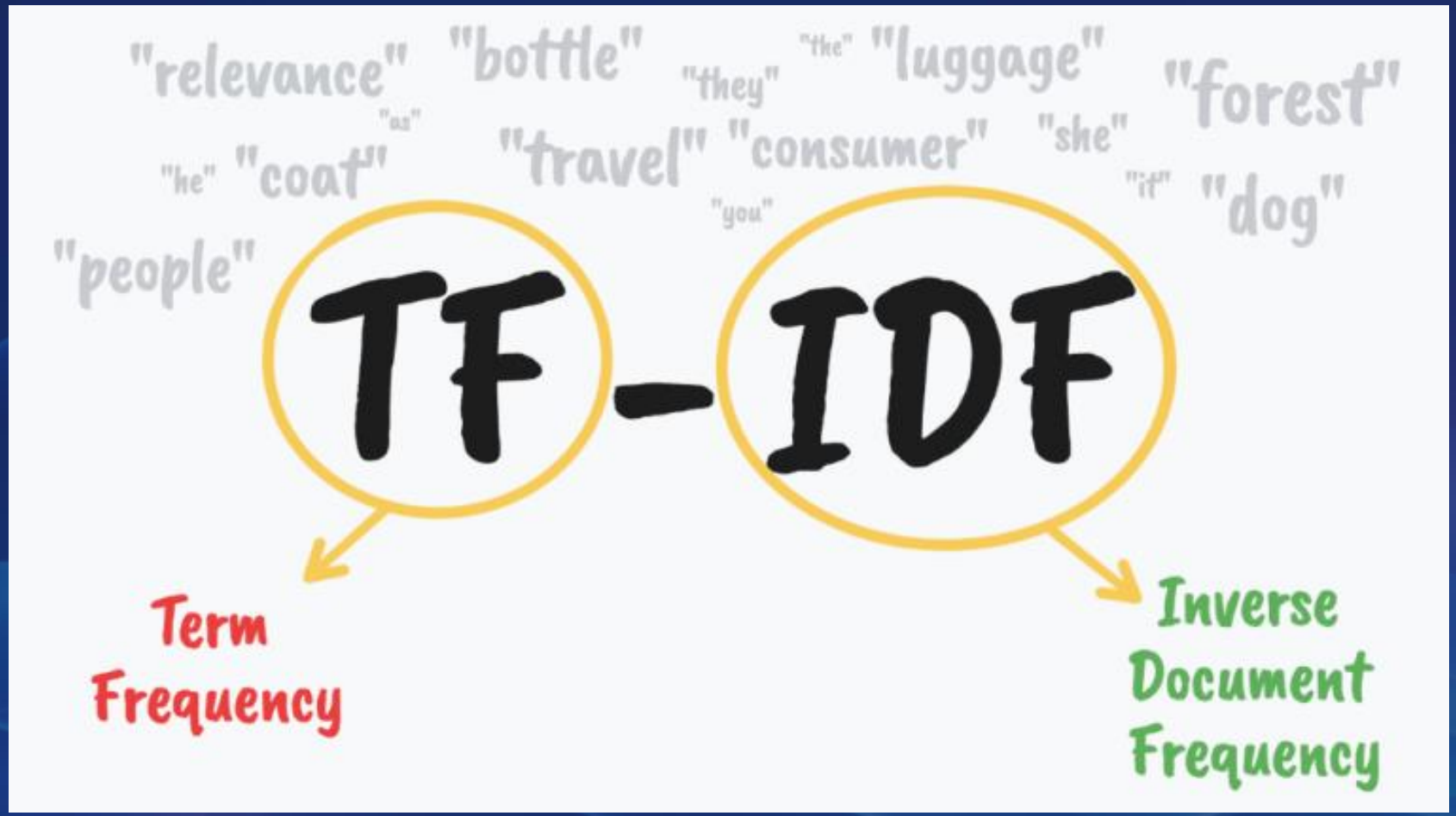
Natural Language Processing
NLP_CLT_1st_May_11th_2025

Eng. Maytham Ghanoum

Artificial Intelligence & Deep Learning Specialist
MTN Syria – SCS – SVU CLT
+963947222064 - +963982018359
https://www.linkedin.com/in/maytham-ghanoum-697aa5207/
https://www.facebook.com/maytham.ghanoum

TF-IDF Representation



TF-IDF Representation

TF-IDF stands for **Term Frequency-Inverse Document Frequency**. It is a statistical measure used to evaluate the importance of a word in a document relative to a collection of documents (corpus).

Unlike the Bag-of-Words (BoW) model, which uses raw frequency counts, TF-IDF reflects how important a word is to a document in the context of the entire corpus.

The TF-IDF score for a word increases proportionally with the number of times the word appears in the document but is offset by the frequency of the word in the corpus. This helps to adjust for the fact that some words are generally more common than others.

TF-IDF Representation

Components of TF-IDF:

1- Term Frequency (TF): Measures how frequently a word appears in a document.

$$TF(t,d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d}$$

TF-IDF Representation

Components of TF-IDF:

2- Inverse Document Frequency (IDF): Measures how important a word is across the corpus.

$$IDF(t, D) = \log\left(\frac{\text{Total number of documents } (N)}{\text{Number of documents with term } t \text{ in it}(df(t))}\right)$$

3- **TF-IDF Score**: The product of TF and IDF.

$$TF - IDF(t, d, D) = TF(t, d)\ X\ IDF(t, D)$$

TF-IDF Representation

Benefits of TF-IDF

1.Relevance Measurement: TF-IDF highlights the words that are more informative for a document, reducing the impact of commonly used words.

2.Improved Accuracy: TF-IDF tends to improve the accuracy of text classification and retrieval systems by focusing on significant terms.

3.Dimensionality Reduction: By emphasizing significant terms, TF-IDF helps in reducing the dimensionality of text data.

TF-IDF Representation

Use Cases of TF-IDF:

1- Text Classification: Sentiment analysis, spam detection, topic categorization.

2- Information Retrieval: Search engines, document clustering, query expansion.

3- Recommender Systems: Content-based filtering, document similarity computation.