

# Week 16.09 – 22.09:

## Word Embeddings & Sentiment Classification

### Objective:

By the end of this week, you will have implemented word embeddings and built a simple text classifier (e.g., sentiment analysis) using your cleaned dataset from Week 1.

### Key Concepts to Explore:

#### 1. Word Embeddings:

- Understand the concept of word embeddings and why they are useful.
- Learn the difference between traditional methods (e.g., TF-IDF) and dense word embeddings (e.g., Word2Vec, GloVe).

#### 2. Text Classification:

- Explore basic machine learning models (e.g., Naive Bayes, Logistic Regression, or SVM) to classify text.

#### 3. Evaluation Metrics:

- Recap about accuracy, precision, recall, and F1-score for evaluating classification models.

### Practical Task:

#### 1. Dataset:

- Use the IMDB Movie Reviews dataset from Week 1.

#### 2. Text Processing Steps:

- Apply your preprocessing pipeline from Week 1 to clean the dataset.
- Implement word embeddings (Using **TF-IDF**)

#### 3. Text Classification Model:

- Build and train a **Logistic Regression** classifier on the TF-IDF features..
- Train the classifier to predict sentiment (positive/negative) from the reviews.

#### 4. Model Evaluation:

- Evaluate your model's performance using accuracy, precision, recall, and F1-score.
- Display the confusion matrix to see where the model struggles (e.g., false positives vs. false negatives).

**Goal: Train a Logistic Regression classifier to predict sentiment in movie reviews and evaluate it with standard classification metrics.**