

# CS485 - Project

## P15: Sentiment Classification of Texts

Zervos Spiridon Chrisovalantis – csd4878  
Drakakis Rafail – csd5310

### Selected Dataset

**Name:** IMDb Movie Reviews Dataset

**Source:** <https://ai.stanford.edu/~amaas/data/sentiment/>

### Why this dataset?

This dataset is a widely used benchmark in sentiment analysis. It contains 50,000 highly polarized movie reviews (25,000 for training, 25,000 for testing) labeled as positive or negative, making it ideal for binary sentiment classification tasks. The text is preprocessed and clean, making it easy to work with and suitable for comparing different models.

### Brief Project Description

In this project, we aim to classify the sentiment of text reviews as positive or negative using Natural Language Processing (NLP) techniques. We will explore and compare multiple methods for sentiment classification.

### Expected Outcomes

- A clear comparison of traditional and modern NLP models for sentiment classification.
- Insights into how model complexity affects performance and inference time.
- A practical understanding of text data preprocessing and fine-tuning large language models.