# Homonyms in Sentiment Analysis — Final Report

## Introduction

This project explores the challenge of **homonyms** in sentiment analysis. Homonyms are words with the same surface form but different meanings depending on context. We assess four different NLP classification techniques to determine which method handles this problem most effectively.

## Dataset Description

* **Dataset**: rotten\_tomatoes (from Hugging Face)
* **Instances**:
  + Train: ~8,000
  + Test: ~1,000
* **Labels**: Binary sentiment (0 = Negative, 1 = Positive)
* **Format**: Movie review texts and sentiment labels

## Experiments & Results

### 🧪 Experiment 1: Pre-trained Text Classification (Roberta-base)

* **Model**: cardiffnlp/twitter-roberta-base-sentiment-latest
* **Method**: Sentiment prediction using a pre-trained Roberta model via Hugging Face pipeline.
* **Result**:

precision recall f1-score support

Negative Review 0.76 0.88 0.81 533

Positive Review 0.86 0.72 0.78 533

accuracy 0.80 1066

macro avg 0.81 0.80 0.80 1066

weighted avg 0.81 0.80 0.80 1066

### 🧪 Experiment 2: Supervised Classification Using Sentence Embeddings

* **Model**: sentence-transformers/all-mpnet-base-v2 + Logistic Regression
* **Method**: Sentence embeddings used to train a supervised classifier.
* **Result**:

precision recall f1-score support

Negative Review 0.85 0.86 0.85 533

Positive Review 0.86 0.85 0.85 533

accuracy 0.85 1066

macro avg 0.85 0.85 0.85 1066

weighted avg 0.85 0.85 0.85 1066

### 🧪 Experiment 3: Prototype-Based Classification (Average + Cosine Similarity)

* **Method**: Average embedding per class → cosine similarity to classify.
* **Result**:

precision recall f1-score support

Negative Review 0.85 0.84 0.84 533

Positive Review 0.84 0.85 0.84 533

accuracy 0.84 1066

macro avg 0.84 0.84 0.84 1066

weighted avg 0.84 0.84 0.84 1066

### 🧪 Experiment 4: Zero-Shot Classification

* **Method**: Cosine similarity between sentence embedding and label embedding (zero-shot).
* **Result**:

precision recall f1-score support

Negative Review 0.78 0.77 0.78 533

Positive Review 0.77 0.79 0.78 533

accuracy 0.78 1066

macro avg 0.78 0.78 0.78 1066

weighted avg 0.78 0.78 0.78 1066

## Tools Used

* Python 3
* Transformers (Hugging Face)
* Sentence-Transformers
* scikit-learn
* pandas
* numpy
* datasets
* torch

## External Resources

* [Hugging Face model hub](https://huggingface.co/models)
* Rotten Tomatoes dataset via datasets
* scikit-learn documentation

## Reflection Questions

**Q1: What was the biggest challenge?**  
Disambiguating homonyms using models that don’t inherently understand context.

**Q2: What did you learn from the project?**  
That deep contextual models are significantly more capable of handling complex semantic tasks involving homonyms than traditional approaches.