

---

# SENTIMENT CLASSIFICATION ON IMDB

## FILM REVIEWS

**Team Member:** *Zeja Chen, Francis Lin, Mike Lin*

### **Statement of Research Question:**

The questions that we would love to answer in the final projects are:

### **Methodology:**

- ◆ Are there any different approaches that we can take when doing text classification?
- ◆ Do classification models that we learned from the class also apply to text analysis?  
If so, will they hold still their competitiveness in the realm of NLP?
- ◆ What is the ultimate language model that has the best accuracy in terms of OOS sentiment classification? What are the tradeoffs?

### **Interpretation:**

- ◆ What are some salient characteristics to the topics of positive and negative reviews?
- ◆ What are the words that dictates the model when it separates the classification boundary?

### **Project Overview:**

I have always considered myself a cinephile. Normally on the weekend, I will pick an intriguing film to enjoy with my friends as a weekly ritual. Although my friend and I all have similar film taste, we still spend an incredible amount of time in choosing the film that we ended up watching.

One of the ways that we determine if a film is watchable or not is by looking up its review from various sources. Movie reviews help users decide if the movie is worth their time. A summary of all reviews for a movie can also help users make this decision by not wasting their time reading all reviews.

As a result, I would love to utilize this opportunity to conduct a sentimental analysis on reviews on IMDb with respect to various films and genres. By leverage different classification models, either learned in the class or outside, I am interested in finding the model that can best classify the polarity given a film review. If possible, I

will also generalize characteristics for both positive and negative sentiments in terms of interpreting how the model makes the classification decision.

Sentiment analysis can be very useful in various scenarios. It determines the attitude of critics depending on their reviews. Great Sentiment analysis of a movie review can rate how positive or negative a movie review is and hence the overall rating for a movie.

By evaluating the performance and tradeoffs of different methodologies, I wish to offer a model that can accurately reflect film reviewers' sentiment and offer pertinent guidance for people who are struggling to find the right movie to watch.

## **Proposed Dataset:**

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

Movie-rating websites are often used by critics to post comments and rate movies which help viewers decide if the movie is worth watching. Since our objective is to classify sentiment given a movie review, a comprehensive dataset that includes a range of reviews with respect to different genre films will be perfect for our analysis.

As a result, we chose the IMDB dataset which has large amount of polarized movie reviews for natural language processing or Text analysis. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. It has 50,000 polarized reviews, each labeled with the corresponding sentiments for supervised learning machine learning models.

## **Proposed Methods:**

We propose to apply two approaches learned in this semester for sentiment analysis: logistic regression and sentiment analysis. The first model is chosen for its clear interpretation, and the second one for its predictive accuracy (among shallow models).

We would also compare the two models mentioned above with one or two neural network models. BERT (Bidirectional Encoder Representations from Transformers) is the current state-of-the-art model in NLP sentiment analysis. If given sufficiently

large dataset, BERT is capable of very accurate prediction. The drawback is the difficulty in model interpretation: it is hard to measure the variable importance for the predictions that the model makes. Another choice of neural network is LSTM (Long Short-Term Memory), a recurrent neural network with special gated units. It does not rely on the size of dataset that much, and the size of the model is smaller (which means easier deployment in applications), comparing to BERT, but the accuracy is generally lower than BERT. So, we would also run LSTM if we have time.

### **Project Plan:**

We split the work into data preprocessing, model implementation and paper writing. We plan to finish the first stage by April 23th, the second stage by May 1st and the third stage by May 2nd.

### **Proposed Presentation Format:**

In terms of presentation format, we are planning on writing a research paper-style report to describe our initiative, data & method, and findings.