

**CS 454 Project Proposal**  
**Sentiment Analysis on Movie Reviews Using Clustering and Neural Networks**

**Project Group:**

Mehmet Fatih Karaosman – S028722

Mordovtsev Roman – S047698

**Problem Definition**

In NLP (Nature Language Processing) one of the biggest problems is the understanding the emotion of the sentences. Sentiment analysis is understanding whether a sentence is positive or negative. In this project I aimed to build a model that analyses whether the movie review is positive or negative.

I will try several algorithms in my project, which are:

1. **K-Means Clustering:** The first step will apply K-Means clustering on the reviews to understand the overall emotion of the sentence. Since K-Means does not use labels, we will map cluster IDs to labels.
2. **Linear Perceptron:** The second step will involve training a linear perceptron classifier on the labeled reviews to distinguish between positive and negative sentiments based on the TF-IDF vectors.
3. **Multi-Layer Perceptron (MLP):** Finally, we will train a multi-layer perception with the same data to assess whether a non-linear model can capture deeper patterns in the text.

After these trials, we will compare performances of all algorithms and choose the best algorithms between the results.

**Dataset**

We will use the **IMDb Movie Reviews Dataset** — a large and widely used sentiment analysis dataset containing 50,000 labeled movie reviews, split evenly between positive and negative sentiments.

For this project, we will randomly sample 10,000 reviews (5,000 positive and 5,000 negative) to meet the class size requirement.

- **Source:**  
Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). "Learning Word Vectors for Sentiment Analysis." Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.  
<https://ai.stanford.edu/~amaas/data/sentiment/>

The dataset will be split into 70% training, 15% validation, and 15% testing.

**References:**

- Maas, A. L., et al. (2011). "Learning Word Vectors for Sentiment Analysis."  
<https://ai.stanford.edu/~amaas/data/sentiment/>