# Sentiment Analysis

Ashish Zantye

May 2024

## 1 Summary

The project successfully explored and implemented advanced NLP techniques for sentiment analysis. The LSTM model provided a foundational understanding of sequence modeling and RNNs in sentiment analysis but was surpassed in accuracy by the fine-tuned DistilBERT model. DistilBERT, being a transformer-based model, demonstrated superior performance due to its ability to capture contextual relationships in text effectively.

LSTM (Long Short-Term Memory) Model: Initially I implemented an LSTM neural network with input, output, and forget gates and utilized Backpropagation Through Time (BPTT) for training the many-to-one RNN (Recurrent Neural Network).Later using the NLTK library and WordNet applied tokenization and embedding on LSTM to preprocess movie review data.

DistilBERT Model: I utilized a pre-trained DistilBERT-base model with 6 layers, 768 hidden units, and 12 attention heads, totaling 66 million parameters which has lesser parameters than BERT and thus quicker to train with almost similar accuracy in terms of result.By employing WordPiece segmentation for tokenization, a subword tokenization method and creating embedding , the data was run on LLM , After that the decision of positive or negative was taken after training and testing on logistic regression after running on LLM.

The combination of LLM and logistic regression gave an accuracy of 81.8% indicating it being a robust method for NLP task of sentiment analysis.