# Executive Summary

## Stanford Sentiment Treebank with 5 labels

## Introduction

The SST-5, or Stanford Sentiment Treebank with 5 labels, is a dataset utilized for sentiment analysis. It contains 11,855 individual sentences sourced from movie reviews, along with 215,154 unique phrases from parse trees. These phrases are annotated by three human judges and are categorized as negative, somewhat negative, neutral, somewhat positive, or positive. This fine-grained labeling is what gives the dataset its name, SST-5. According to the leader board, the highest accuracy on the test set is 59.8, but more interestingly, the model that obtained 5th rank with accuracy of 55.5 only used BERT Large model with dropouts. The purpose of our project is to see if we can achieve to be in top 5 of the leader board by hyperparameter tuning (on learning rate and hyperparameters of Adam optimizer) and fine-tuning.

---

## Hyperparameter Tuning (FNN on GTE)

We tried optimizing various hyperparameters including optimizer, learning rate, batch size, the number of layers, and the number of neurons with regularization and dropout. We observed that lower learning rate, shallow network, and small number of neurons with dropout improves performance. This is most likely because our training data is small. As we increase the number of epochs, another common issue was overfitting. Possible solutions to this are feature reduction, data augmentation, and fine-tuning pretrained sentence transformer models. We mostly focused on the fine-tuning option for this project.

---

## Conclusion

- With a fine-tuned lightweight sentence transformer and a shallow neural network, we achieved 56.9% accuracy (better than 4th on leaderboard).
- With a heavier sentence transformer and fine-tuning, we achieved 58.6% accuracy (better than 3rd on the leaderboard).

---

## Future Directions

- Ordinal regression
- Feature reduction
- Data augmentation
- Different fine-tuning (e.g., triples and loss functions)