# Introduction

## Group Members:

Eli Pinkus

Akshay Vegesna

## Group Name

Yasser’s Disciples

## Division of Labour

Each of us implemented models on or own and we generally discussed the merits of models and what was or wasn’t working as a group. Labour was well divided, and although some group members models performed better than others, the overall work load was well distributed.

# Overview

## Models & Techniques

* **Ridge Regression:** Simple model performed poorly out of sample
* **Lasso Regression:** Performed decently and informed feature reduction decisions
* **Logistic Regression:** Performed surprisingly well after fine tuning
* **Shallow Neural Network:** Performed slightly worse than logistic regression. Noticeable problems with overfitting.
* **Decision Tree with Gini Loss:** not great performance
* **Random Forests:**
* **Feature Reduction:** We tried to reduce the dimensionality of the input space based on various criterion with mixed results.
* **K-Fold cross validation:** We used K-Fold cross validation to get a sense of how a given model was performing on unseen data.

# Approach

## Data Processing/Manipulation

The bulk of our data manipulation focused on feature selection with views toward reducing the dimensionality of the input space. We began by considering a restriction on the minimum letters in a word represented in the ‘bag of words’. We do so under the rough assumption that longer words would, in general, be more specifically employed, and thus, more meaningful in sentiment analysis. We also considered more principled approaches to feature reduction and we implemented lasso regression and principle component analysis for certain models as will be discussed more in-depth in discussing the implementation of each model. In general we looked to apply these techniques more aggressively with more complicated models that have high variance in an attempt to reduce the complexity of those models and drive down variance.

# Model Selection

# Conclusion