Hello everyone, it is a pleasure to be able to share our project with you all.

To start, our aim of the project was to predict the change in stock price based on news headlines. We chose this project in the hopes of helping people of any age and background to get started in understanding the market better. By using an example of an AI reading the market trend from headlines, it could help users to get a better grasp of the stock market.

So what is Sentiment Analysis? Sentiment Analysis is a use case natural language processing and deep learning, where text is classified into categories (e.g. a Positive/Negative review or a Happy/Sad tweet) By classifying if headlines were going to have a positive impact or a negative impact to the stock price, we could predict how the stock price would change. In the past, teams of human experts would read through text to tell what the effect would be on the stock price, but the amount of information available has exploded in recent years. Hence, we train a machine to learn to predict these for us.

To train the machine, we used a dataset from the Inspirit AI program’s Natural Language Processing for Finance project. The dataset had in total 2264 sentences that were labelled to have either “positive, negative or neutral” impact on the price of the company’s stocks.

For the model we used an architecture named Long short term memory, or LSTM for short. This architecture was chosen because it can process sequential data that depend on its order, such as text. The order of the words in a sentence is crucial to understanding its meaning. Additionally, the architecture can take the context into account when processing words, as it can choose to remember certain aspects about past sentences. The model output consists of 3 probabilities that belong to being negative, neutral or positive respectively.

Finally, to gather headlines, we used python’s beautiful soup library to crawl BBC Companies News. These headlines were fed to the model for it to predict the impact on stock prices.

When training the model we tuned some hyperparameters, but due to time limitations we did not optimize the hyperparameters fully.

With the outcome we were not surprised to find that most of the headlines were predicted to have neutral or no impact, since our dataset already had skewed data that were mostly neutral. This kind of bias could be eliminated in the future with better datasets.

In conclusion, the model had more True positives and True Negatives than False positives or False negatives, since they were all aggregated into neutral. With better datasets and longer training, it could be improved in the future.

Thank you for listening.