Word2Vec:
https://nlp.stanford.edu/projects/glove/
https://nlp.stanford.edu/data/glove.6B.zip

Fine tune the model on our financial data.

API Details:
https://documenter.getpostman.com/view/23904979/2s847CyFyY

API Token access
In order to get access token, we need to register developers page on stocktwits:
https://api.stocktwits.com/developers

Then we get access tokens...

And Public API is https://api.stocktwits.com/api/2/streams/symbol/SYMBOL_NAME.json

Hugging Face Link:
https://huggingface.co/tarnformnet/Stock-Sentiment-Bert

I have divided whole dataset into 2 parts, i.e. Training Set and Validation Set. In the training set I have taken 90 percent of total records from the whole data set.
In data set we have 2 column i.e Textual tweet and Label of respective tweet(Bullish - 0, Bearish - 1)

Accuracy - Out of total tweets from validation set, percentage correctly predicted tweets.
Example: Let's assume that we have 1000 tweets in the validation set, then out of these 1000, 850 tweets are correctly predicted by our model then accuracy of our model will be 85% (850/1000)

Model precision: Out of total positive predicted(specific class) classes, percentage of correctly classified.
Example: Let assume that our model predicts 100 Bearish classes then out of these 100, precision explains how many are actually bearish. Let’s say 70 examples are belongs to bearish then precision will be 70% (70/100)

Model precision is a measure of how accurately a binary classifier model predicts the positive class. It is calculated as the number of true positive predictions made by the model divided by the total number of positive predictions made by the model. For example, if a model makes 100 predictions and 80 of them are correct, the model has a precision of 80%.

To illustrate this further, suppose a medical test is designed to identify patients with a certain disease. The test is applied to a group of 100 patients, and the results show that 30 of them have the disease. If the model correctly identifies 28 of the 30 patients with the disease, the model has a precision of 28/30 = 93.3%. This means that when the model predicts that a patient has the disease, there is a 93.3% chance that the prediction is correct. On the other hand, if the model only correctly identifies 20 of the 30 patients with the disease, the model has a precision of 20/30 = 66.7%. In this case, when the model predicts that a patient has the disease, there is only a 66.7% chance that the prediction is correct.

Model Recall: Out of total actual positive classes, percentage of correctly classified.

Example: Let assume that our validation set has 900 Bullish classes then out of these 900, recall explains how many are correctly predicted by our model. Let's say 700 examples are predicted bullish by our model then 77.77% (700/900)

Model recall is a measure of a model's ability to correctly identify all relevant instances of a given class. It is also known as the true positive rate or sensitivity. For example, suppose you have a model that is used to detect cancer in patients. The model's recall would be the percentage of patients with cancer that the model correctly identified. High recall means that the model is good at correctly identifying all relevant instances of the class, while low recall means that the model is missing some relevant instances.

Model F1-Score: It is an average of F1-Score and In place of average, it uses harmonic mean.

Explain Model F1-Score with example. Model F1-Score is a metric that is used to measure the accuracy of a classification model. It takes into account both precision and recall in order to calculate the score. Precision is the ratio of true positives to the sum of true positives and false positives, while recall is the ratio of true positives to the sum of true positives and false negatives. The F1-Score is calculated by taking the harmonic mean of precision and recall. For example, say you have a model that predicts whether or not an email is spam. You could use the F1-Score to measure the accuracy of the model. You would calculate the precision by dividing the number of true positives (emails correctly identified as spam) by the sum of true positives and false positives (emails incorrectly identified as spam). You would then calculate the recall by dividing the number of true positives by the sum of true positives and false negatives (emails incorrectly identified as not spam). Finally, you would take the harmonic mean of these two values to get your F1-Score.

ROC Curve: It is a curve between True Positive rate and False Positive rate.

A receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as the discrimination threshold is varied. The plot is generated by plotting the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. The ROC curve is a useful tool for evaluating the performance of a classifier, as it shows the trade-off between the true positive rate and the false positive rate for a given classifier. A classifier with a higher true positive rate is said to be more sensitive, while a classifier with a lower false positive rate is said to be more specific. An ideal classifier would have a high true positive rate and a low false positive rate, resulting in a point on the upper left corner of the ROC curve.

AUC Score: It refers to the area under the roc curve.
Example: Curve cover 74% area under by roc curve then auc-score will be 0.74

Explain AUC score with an example. AUC (Area Under the Curve) is a measure of the overall performance of a classification model. It is a metric that measures the ability of a model to distinguish between classes. It is calculated by plotting the true positive rate (TPR) against the false positive rate (FPR) for all possible thresholds of a binary classifier. The AUC score is the area under the ROC curve, which ranges from 0 to 1, where a higher score indicates a better model performance. For example, let's say that we have a model that is being trained to predict whether a customer will purchase a product or not. We can measure the model's performance by using its AUC score. If the AUC score is 0.90, then it means that the model can accurately distinguish between those who will purchase the product and those who will not with 90% accuracy.

The true positive rate, also known as sensitivity, is the proportion of positive cases that are correctly identified as such by a binary classification test. For example, if a medical test for a certain disease has a true positive rate of 90%, it means that out of 100 people who have the disease, the test will correctly identify 90 of them as positive.

What is model loss converge?
Model loss convergence is the process of minimizing the loss of a model over a period of time. It occurs when the model is able to accurately predict the output for a given input, or when the model is able to minimize the error with each iteration of training. It is a measure of how well the model is able to learn from the data it is given.

Explain sparse_categorical_accuracy with example?

Sparse categorical accuracy is a metric used in Keras to evaluate the accuracy of predictions made by a neural network on a classification problem with a single target variable. It compares the predicted category against the true category and returns the percentage of predictions that were correct. For example, if a neural network is used to classify images of cats and dogs, sparse categorical accuracy can be used to evaluate the accuracy of the model's predictions. If the model correctly predicted that an image was a cat 80% of the time, then the sparse categorical accuracy would be 80%.

Explain val_sparse_categorical_accuracy with example?
Val_sparse_categorical_accuracy is a type of metric that is used to evaluate the accuracy of a model in a classification problem. It is similar to accuracy, but it is calculated by comparing the predictions of the model to the true labels (sparsely encoded). For example, let's say you are training a model to classify images of cats and dogs. The true labels of the images will be encoded as 0 (for cats) and 1 (for dogs). The model's predictions will also be encoded as 0 or 1, and the val_sparse_categorical_accuracy will be calculated as the percentage of images that were correctly classified.