# FTD\_Project\_3

This project is a continuation of FTD\_Project\_2. FTD\_Project\_2 aimed to discover whether or not Fail-to-Deliver (FTD), Short Interest, and historical price data could be used alongside machine learning in order to make predictions on a stock’s future price movements, predictions which could then be used to make real-life trading decisions with.

The answer appeared to be yes, machine learning models can be created using FTD and Short Interest data as features in order to track and predict price movement. However, there two main issues to overcome before the models could really be used to make real-life trading decisions.

FTD\_Project\_3 aims to solve these problems, and either present a working, complete and tradeable model, or will present a pathway to further completing the models in order to truly make them worth trading with.

The first problem: the FTD data from the SEC, which is permanently delayed by 10 trading-days (2 weeks) or more, is the most significant issue, because it means the models would be missing the most recent data while attempting to make price predictions.

The second problem was with the models themselves – while they were able to quite accurately track the same-day closing price using our provided dataset, they were not ready or set up to accurately make future predictions using the data.

By using our models and data from FTD\_Project\_2, we can identify the most accurate models and stock symbols using our current data. And then supplement those models with LSTM models that attempt to predict FTD data for the missing 15 days.

Afterwards, the models FTD\_Project\_2 can be used with the new column of predicted FTD data, alongside real, up to date EOD-data that was not present for training, in order to attempt to predict future price movements of our selected stocks.

Additionally - control models (models without FTD data) will be created and evaluated to determine the exact effect FTD data is having on this Machine Learning Project, in comparison to the short interest data, techinical indicator data, and raw historical data also being fed into the models.

At the moment - FTD data is available up to the end of 2021 (Dec 31st). On February 1st, FTD data for Jan 1st until Jan 15th will be released. These 15 days of FTD data will be compared against the FTD\_LSTM models, to see if they are capable of accurately predicting FTD data ahead of when FTD released.

If the data is accurate? Then we can use the predicted FTD data to fill in 15 day-delay in reporting, and then predict the following (Jan 16th-Jan 31st) price movements, and compare them to real-life data.

If this works? Then this model can be used to make price predictions, every 15 days (on the 1st and ~15th of every month) to attempt to predict the next 15 days of trading movements. And these predictions can then be used to make trading decisions on the stock symbols that come out of this.

If time allows - more than LSTM will be used for FTD prediction. But for now, a LSTM model will be contructed to suplement the existing neural network models that are accurately tracking price movement of approximately ~250 different stock symbols.

There are five folders with useful model data that will be used for the project presentation

All located in '../Model\_Data/'

There are two sets of Neural Network (NN) models which are direct improvements upon the original machine-learning models from FTD\_Project\_2. The difference between each set, which should be obvious from the title path, is one set of models was trained with and including the SEC's FTD data, while the second set had the FTD data omitted from the models. Each model is improved relative to FTD\_Project\_2, because now they not only attempt to predict the close price with the data provided, but now they attempt to forecast the close price 1, 2, 5, or 10 days in advance.

Model locations:

Model\_1:

path\_1 = Path('../Model\_Data/ Date\_Test\_NN\_w\_FTD\_all /') Contains models trained with FTD data

Model\_2:

path\_2 = Path('../Model\_Data/Date\_Test\_NN\_noFTD\_all /') Contains models trained without FTD data

The third set of models shares a similarity with the previous two NN models - it used the same dataset (up to Dec31st) to train, however, instead of a neural network (NN) model, it is an LSTM model designed to supplement the performance of the NN models and compare its results alongside. This third model does not predict/forecast data in quite the same way as the NN models - it should be used as a comparison or addition to the 1-day NN models listed above. It also acts as demonstration that LSTM-models can also be used, in addition to NN, to predict/track the closing price using our FTD, Short Interest, and Historical Pricing dataset.

Model Location:

Model\_3:

path\_3 = Path('../Model\_Data/LSTM\_Model\_Data\_1/') ## Contains LSTM models trained with FTD data, up to Dec31st.

The fourth set of models were designed after early evaluations of models \_1 and \_2. The purpose of the first two models was to compare the effect of removing the FTD Data from the features of the machine learning model, and early evaluations showed that, while definitely improving the accuracy of the models when included, the FTD data was not necessary or required in order to achieve accurate price tracking / predictions. So, the fourth (and fifth) models were designed as concept of how these models could genuinely be used to make predictions, select stock symbols, plan a buying strategy, and actually trade using these models that have been created over the course of this project.

Model Location:

Model\_4:

path\_4 = Path('../Model\_Data/Feb2022\_NN\_models/') Contains NN models, capable of taking January's data and making predictions 1, 2, 5, or 10 trading days into February.

The fifth set of models were created for the same purpose as the fourth and perform a 10-day forecast of close price for a given stock. Instead of a NN model, this is an improved-upon (relative to Model\_3) LSTM model that alongside our NN can be used to forecast stock price predictions, with enough accuracy that one could consider trading off these models.

Model Location:

Model\_5:

path\_5 = Path('../Model\_Data/Feb2022\_LSTM\_models\_1/'

Code is located in their respective Functions\_ Folder

## Machine Learning Folders

Model locations:

Model\_1:

path\_1 = Path('../Model\_Data/Date\_Test\_NN\_noFTD\_all/') Contains models trained without FTD data

Model\_2:

path\_2 = Path('../Model\_Data/Date\_Test\_NN\_w\_FTD\_all/') Contains models trained with FTD data

Model Location:

Model\_3:

path\_3 = Path('../Model\_Data/LSTM\_Model\_Data\_1/') ## Contains LSTM models trained with FTD data, up to Dec31st.

Model Location:

Model\_4:

path\_4 = Path('../Model\_Data/Feb2022\_NN\_models/') Contains NN models, capable of taking January's data and making predictions 1, 2, 5, or 10 trading days into February.

Model Location:

Model\_5:

path\_5 = Path('../Model\_Data/Feb2022\_LSTM\_models\_1/'

## Updates

Models are finished. For now. Evaluation and comparison on completed models needs to be performed.

## Next steps -

Evaluate and compare models. Prepare data for presentation.

Find out status of "front-end" development from other group members.