

Project: Financial time series forecasting

V. 1.0 4.9.2020

The following project is an individual piece of work and you not allowed to share your solution with anyone else. Read below carefully!!

- The aim of the project is that you learn how to set up an analytics project end-to-end. A secondary aim is that you understand how to work with a time series data set and forecast based on such data. Third aim is that you gain an insight into how to interpret data and results.
- The solution must address each grade in the written order. Therefore, to complete grade five you must have completed the other grades first.
- Unintentionally, there may be information missing in the description, please go through the description early in advance so that you have time to ask for assistance.

Project description:

Grade 1 – Implement a complete process for forecasting a single stock.

You should do the following steps:

- Use the EURUSD data set
- Calculate two features from the paper [1] that are not simple moving averages. Note you need to implement your own calculation of each feature.
- Normalize or standardize data depending on which model is used
- Create a label for your forecast, by shifting the close value n steps
- Drop other data than the close and the two features for inference
- Split data into 80%/20% (train/test)
- Set up a linear model
- Fit/train the linear model to the training data
- Forecast based on test data
- Calculate the R^2 error on both the training data set and the test
- Compare the errors and explain the outcome

Grade 2 – Illustrate data using Plotly/other library

- First create a figure based on OHLC candles
- Second do a line chart that illustrates the label (actual data) and the forecast in the same line figure. The lines should have different colors and include names of series.
- What patterns can you observe from the line figure?

Grade 3 – Create sliding windows for the input data, e.g. the window length of 10, 5, and 2 samples.

You will extract data for the window length n (rows), and turn the data from a matrix (2D) form into a vector (1D) form of the size (i.e. number of rows * number of columns).

- Create a function that return a 1d vector (array, tuple) of n window length from raw data
- Create a new data frame where you store the vector as a new row (see lecture notes)
- Fit and Run three different models and that test three different lengths
- Summarize and Compare their R^2 error measures which is better. Is anyone better than the original model without a window length?

Grade 4 – Implement an investment decision to either buy or sell based on some signals which you choose to detect.

See the paper [1] for how this can be done, the easiest solution is to hand-craft this decision to either buy or sell.

- Compare the regression forecast with the underlying close price (Note the last close input price).
- Once the close price go above the forecast you can use it to define a buy opportunity
- Once the close price go below the forecast you can use it to define a sell opportunity
- Calculate the hit ratio of your investment decision for each of the windows
- Which setup was the best, and why was that?

Grade 5 – Use a more advanced model, e.g. LSTM and compare all previous tests for this model.

- Train and run a model for no window, and 2, 5, and 10 window size.
- Visualize all LSTM forecasts and the label in the same graph.
- Compare the R^2 errors in a table format

Model	Window length	R^2
Linear	1	x.x

Reference:

[1] Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4873195/>