

MSc. in Computing Practicum Approval Form

Section 1: Student Details

Project Title:	Univariate Time Series: Air Traffic Forecasting using Motifs and LSTM
Student ID:	20211232, 20210360
Student name:	Gayathri Sridhar, Meenakshi Srinivasan
Student email	gayathri.sridhar2@mail.dcu.ie , meenakshi.srinivasan2@mail.dcu.ie
Chosen major:	Data Analytics
Supervisor	Dr. Martin Crane
Date of Submission	03-12-2020

Section 2: About your Practicum

What is the topic of your proposed practicum?

Proposed topic: Univariate Time Series: Air Traffic Forecasting using Motifs and LSTM

Univariate Time Series is the collection of observations, of a specific variable over a period of time. Time series Forecasting has proven to be successful in vast domains such as Finance, Health sector, Weather forecasting, Employment growth Et cetera. Air traffic has increased significantly over the last few years. The aim of this project is to discover motifs (recurring patterns) and feed them to Long Short-Term Memory(LSTM) Neural Network model to forecast the air traffic trends. The STUMPY algorithm is used to discover the motifs. Various Airline companies operating all around the world, can plan their operating schedule based on demand through these results and stay ahead in this competitive business.

Please provide details of the papers you have read on this topic.

1. Tularam, A. and Saeed, T. (2016) "Oil-Price Forecasting Based on Various Univariate Time-Series Models", *American Journal of Operations Research*, 06, pp. 226-235, doi: 10.4236/ajor.2016.63023.
2. Deetchiga, S., Harini, U. K., Marimuthu, M. and Radhika, J. (2018) "Prediction of Passenger Traffic for Global Airport using Holt's Winter Method in Time Series Analysis," *2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW)*, Erode, India, 2018, pp. 165-169, doi: 10.1109/I2C2SW45816.2018.8997519.
3. Law, S. (2019) "STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining", *Journal of Open Source Software*, 4(39), 1504, doi: 10.21105/joss.01504.

4. Liu, Y., Su, Z., Li, H. and Zhang, Y. (2019) "An LSTM based classification method for time series trend forecasting," *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, 2019, pp. 402-406, doi: 10.1109/ICIEA.2019.8833725.
5. Barry, B. and Crane, M. (2019) "Analysis of Cryptocurrency Commodities with Motifs and LSTM", *AICS 2019:27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, National University of Ireland, Galway, 5-6 December 2019. pp. 28-39. Available at: http://ceur-ws.org/Vol-2563/aics_5.pdf

How does your proposal relate to existing work on this topic described in these papers?

Tularam, A. and Saeed, T. (2016) [1], implemented various Univariate Time Series models such as ARIMA and Exponential Smoothing for forecasting. Deetchiga, S. et al., (2018) [2] explored the future trends of air travel using Holt-Winters Exponential Smoothing(HWES) model in time series, as it involved seasonality and trends.

Matrix profile algorithm is one of the efficient algorithms to discover motifs. Eamonn Keogh, the pioneer of matrix profile developed various algorithms such as STAMP, STOMP and SCRIMP++. As these algorithms required a lot of computational time, Law, S. (2019) [3] came up with STUMPY algorithm which is a powerful and scalable library that computes matrix profile efficiently to discover motifs. STUMPY makes use of parallel processing which does ordered search for sub sequences in a time series to minimise the runtime. Long Short-Term Model(LSTM) is one of the Recurrent Neural Networks used to retrieve information from sequential data. Liu, Y., et al., (2019) [4] implemented LSTM on time series and found that it outperformed traditional Auto regression models.

Barry, B. and Crane, M. (2019) [5] have incorporated motifs in the LSTM neural network and compared the forecast accuracy of the LSTM model with motifs to standard LSTM models. They have proved that the LSTM model with motifs have reduced the RSME value by 8% than that of the standard one.

What are the research questions that you will attempt to answer?

Can Motifs aid in improving the predictions of Long Short-Term Memory(LSTM) model for Air Traffic forecasting?

How will you explore these questions?

- What software and programming environment will you use?

- Anaconda, which has inbuilt Jupiter notebooks will be used for development and training of models. It supports python packages that comprises of statistics and advanced mathematical methods.

- What coding/development will you do?

- Motifs are discovered using STUMPY algorithm which is implemented in Python. Then, the motifs are fed into the LSTM model and trained to forecast the passenger count.

- What data will be used for your investigations?

- Air traffic dataset is used and it contains nearly 20,000 observations of Air traffic, globally from the year 2005 to 2016.

- Is this data currently available, if not, where will it come from?

- Yes. The data is currently available and it is retrieved from Kaggle.

- What experiments do you expect to run?

- The motifs are discovered and the dataset is split into training and test data. Then, the LSTM model is trained for forecasting.

- What output do you expect to gather?

- The expected output is identifying motif pairs, predicting air traffic trend for the successive years and checking whether the accuracy is improved after applying motifs in LSTM model.

- How will the results be evaluated?

- The results will be evaluated using standard metrics in loss function like Root Mean Squared Error (RMSE).

Roles & Responsibilities

S.No	Task	Name
1	Data Acquisition	Meenakshi S
2	Data cleaning and pre-processing required for time series analysis	Gayathri S
3	Finding and analysing motifs in the dataset	Meenakshi S
4	Training LSTM model using different parameters	Gayathri S
5	Testing and evaluating LSTM model using standard metrics	Meenakshi S & Gayathri S
6	Documentation	Meenakshi S & Gayathri S