



Project Number: 956086

Project Acronym: SMARTHEP

Project title: Synergies between MACHine learning, Real Time analysis and Hybrid architectures
for efficient Event Processing and decision making

DELIVERABLE 5.2 – LIGHTBOX Software & associated publications

DOCUMENT AND DELIVERABLE DESCRIPTION

Work Package	WP5 – Hybrid Architectures
Deliverable	D5.2 – LIGHTBOX Software & associated publications
Document Name	LIGHTBOX software to optimize parallelization of financial transactions, associated publications
Submission Date	
Revision Date	
Dissemination Level	C – Confidential
Deliverable Lead	Sorbonne University
Author(s) and contact	Leon Bozianu, Sofia Cella
Discussion and review	N/A
Approval	N/A

Abstract	LIGHTBOX software to optimize parallelization of financial transactions, associated publications
Keywords	Hybrid Architectures, WP5
Statement of originality	This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Financial time series analysis at Lightbox

Project description

The code in this repository is the result of secondment work at Lightbox Technologies SA as part of the [SMARTHEP](#) European Training Network (ETN).

Secondment summary

Lightbox provides consultancy, research and development of hardware and software infrastructures, algorithms, mathematical and statistical models, aimed at the acquisition, maintenance, processing and analysis of massive volumes of data, and the optimization of decision-making and production processes of third-party entities. The work contained in this repository focuses on the application of machine learning (ML) tools to financial time-series data through the development of systems for predictive analysis using historical market data time series for a multitude of financial instruments.

Algorithmic biases were prevented by dividing the time series in training and testing periods. A Transfer Learning technique was employed in order to improve the training of the model by combining data from different securities to compensate for the limited availability of historical data. The training sample was identified by selecting datasets showing similar statistical features. The various models provide future predictions in unseen time-series data along with associated uncertainties.

This repository contains the codebase for a number of ML models ranging from the simple to the complex. We utilise and show the training and inference from classical boosted decision trees to Bayesian and recurrent neural networks. Each network architecture informed subsequent designs. An important part of the problem was the model adapting to changing and often turbulent time series data. In comparison to other ML applications the models are susceptible to going “out-of-date” (out-of-sample) very rapidly following training - thus presenting challenges to our choice of training/validation data.

This repository contains part of the results produced during this project.

<https://github.com/BozianuLeon/Lightbox/tree/main> [github.com].

Code Usage

In order to start working with the code contained in this repo use the `install.sh` (first time) and `setup.sh` scripts found in the `setup` directory. This will create a conda/miniconda environment with the necessary python libraries installed to reproduce our base results.

The code contained in the `src` directory can be used to train and save particular models before inferring on unseen test data using the pre-trained model weights. A simplified workflow is described below.