U.PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

PSI20 – Forecasting with Time Series Analysis and Deep Learning (LSTM)

Testing the efficient markets hypothesis

João Luís Ferreira Queirós

Dissertation
Master in Economics and Business Administration (MEAE)

Supervised by:
José Abílio de Oliveira Matos

2023

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, José Abílio Matos, for his unwavering support, guidance, and availability throughout the research and writing process. His expertise and insightful feedback have been invaluable in shaping this thesis.

I would also like to extend my heartfelt thanks to my family and friends, especially my wife and son, for their support, understanding, and patience. Their constant encouragement, listening ear, and enduring the long hours of work and occasional moments of stress have been crucial in keeping me motivated.

I would like to acknowledge and express my appreciation to all the researchers who have preceded this work in the field of stock market analysis. Their contributions and insights have provided a solid foundation for this research, shaping the direction of this investigation.

Finally, I would like to thank the Universidade do Porto and the Faculty of Economics and Management (FEP) for providing a conducive environment for learning and research. The resources, facilities, and academic support provided by the institution have been essential to the successful completion of this master's thesis.

ABSTRACT

This dissertation is a contribution to the ongoing global research on the use of Machine Learning and Time Series analysis to better understand and predict the behaviour of financial and stock exchange markets. Its specific contribution is to extend the research to the Portuguese main stock index, the PSI20, where a lack of literature was identified specially on predictions using deep-learning methodology.

This research focus on the application of Long Short-Term Memory (LSTM) models, for forecasting the PSI20 index value, and related variables. By using historical data, including intra-day information, cross-market information from other European and American index, and other features, this study tests the Efficient Market Hypothesis (EMH) applicability to the Portuguese index and aims to uncover potential patterns that may impact stock price movements.

The findings based on models trained with daily data reveal the unpredictable nature of stock markets and the influence of arbitrage effects in the neutralization of anomalies. Most of the models adjust their predictions to the last known value confirming the EMH's claim that the best estimator for tomorrow's price is today's prices.

When the focus shifts to intra-day value, promising levels of accuracy in the LSTM models' predictions are found, outperforming random guesses, and hinting for the presence of temporal anomalies, suggesting potential for designing successful algorithm trading simulators.

This research demonstrates the potential of deep learning techniques for stock market analysis in the context of the PSI20 and contributes to the field of finance by shedding light on the challenges and opportunities of utilizing Artificial Intelligence (AI) for forecasting stock values.

RESUMO

A presente dissertação insere-se no contexto da investigação mundial sobre o recurso a *aprendizagem automática* na análise de séries temporais para a compreensão e previsão do comportamento dos mercados financeiros e da bolsa de valores. A contribuição específica desta tese é alargar o âmbito da pesquisa sobre a aplicação de metodologias de *aprendizagem profunda* ao principal índice de ações português, o PSI20, realizando previsões sobre o seu valor futuro.

A investigação realizada explora a utilização de modelos de *Long Short Term Memory* (LSTM) para prever o valor do índice PSI20 e variáveis relacionadas. Recorrendo a dados históricos, incluindo dados diários e intra-diários, informações de outros índices bolsistas europeus e americanos entre outros elementos, este estudo testa a aplicabilidade da hipótese dos mercados eficientes (EMH) ao índice português e tem como objetivo descobrir possíveis padrões implícitos nas variações do índice.

As conclusões obtidas com base em modelos treinados com dados diários revelam a natureza imprevisível dos mercados bolsistas e a influência dos efeitos de arbitragem na neutralização de anomalias. A maioria dos modelos ajusta suas previsões ao último valor conhecido, confirmando a afirmação da EMH de que o melhor estimador para o preço de amanhã é o preço de hoje.

Quando o foco se centra em dados intra-diários, registam-se níveis consideráveis de precisão nas previsões sobre o sentido da variação do valor do índice. Os valores de precisão obtidos ficam acima dos valores esperados por uma tiragem aleatória, sugerindo assim a presença de anomalias temporais, indicando potencial para a criação de algoritmos de negociação automática bem-sucedidos.

Esta pesquisa demonstra o potencial das técnicas de *aprendizagem profunda* para análise do mercado bolsista, no contexto do PSI20, contribuindo para uma melhor compreensão dos desafios e oportunidades inerentes ao uso da inteligência artificial em previsões do valor de ativos.

AUTHOR AND SUPERVISOR

Author

Name: João Luís Ferreira Queirós – 199900072

Email: queiros.joao@gmail.com

UP email: up199900072@edu.fep.up.pt

Supervisor

Name: José Abílio de Oliveira Matos

Email: jamatos@fep.up.pt

Scientific Grouping: Mathematics and Information Systems

LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Networks
DSM	Data Scaling Methods
EDA	Exploratory Data Analysis
EMH	Efficient Market Hypothesis
HFT	High Frequency Trading
LSTM	Long Short-Term Memory
ML	Machine Learning
MSE	Mean Squared Error
OHE	One Hot Encoded
PSI	Portuguese Stock Index
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
XAI	eXplainable Artificial Intelligence

CONTENT

INTRODUCTION

This chapter will present the primary objectives and rationale for this dissertation.

With the rapid advancement of computational tools applied to stock market research, the application of deep learning techniques in financial forecasting has gained significant attention (Jiang, 2021) from the research community.

This thesis focuses on using Long Short-Term Memory (LSTM), a type of deep learning model, to predict future values of the main Portuguese stock index, the PSI20 (sometimes also referred as PSI-20). While the application of deep learning in financial forecasting has been explored in various markets, there is a limited number of studies employing this methodology for the PSI20. Consequently, this research aims to fill this gap in the literature and contribute to the growing body of knowledge in this area.

Deep learning is a special case of machine learning algorithms that use as the underlying basis Artificial Neural Networks (ANN). Machine Learning is a subset of the more general research area, Artificial Intelligence (AI), that is on the spot due to the increasing visibility of its practical applications.

This thesis aims to test the Efficient Market Hypothesis (EMH) in its weak and semi-strong forms. The weak form of EMH suggests that stock prices fully reflect all historical information, and therefore, analysing past data should have no impact on accurately predicting future price movements (Fama E. F., 1965). By using deep learning techniques, which can capture and learning from historical patterns (Raschka, Liu, Mirjalili, & Dzhulgakov, 2022), this study intents to examine whether the weak form of EMH holds true for the PSI20, through analysis and comparison of LSTM model generated predictions with actual market data. Particularly this research will try to find evidence, and benefit, of the *momentum effect* (Jegadeesh & Titman, 1993), a temporal pattern that refers to the tendency of assets that have performed well in the past to continue performing well in the future, and vice versa.

By incorporating other world indexes to aid in the prediction of future values of the PSI20, which was identified in the existing literature as a promising field of research (Jiang, 2021), this thesis aims to find lagged correlations with other stock indexes and draw conclusions on the application to the PSI20 of the semi-strong form of the EMH, that asserts that all publicly

available information, including information from related markets, is immediately reflected in asset prices (Fama E. F., 1965).

While much of the current research in this domain concentrates on American and Asian stock markets (Ferreira, Gandomi, & Cardoso, 2021), this dissertation seeks to shed some light on the Portuguese stock market, which has received limited research attention in the past. The PSI20, being a relatively small and peripherical stock exchange index, offers an interesting context for deep learning-based financial prediction. As *arbitrage resources are heavily concentrated in the hands of a few investors that are highly specialized in trading a few assets, and are far from diversified* (Shleifer & Vishny, 1997), the PSI20 could be less subject to arbitrage effects that balance the anomalies, making it an interesting case study for deep learning forecasting.

Research Question

Are models generated by machine learning methodologies, particularly with deep learning approaches using LSTM, when trained with historical data of the PSI20 and other world indexes, capable of revealing patterns that challenge the applicability of the EMH to the Portuguese stock index?

Research Goals

- Elaborate forecasting models using LSTM that forecasts the next PSI20 closing value, based on the daily historical data from Euronext Lisbon from the last 10 years.
- Elaborate forecasting models using LSTM that forecasts the next PSI20 closing value, based on the daily historical data from the PSI20 alongside with other World Stock indexes.
- Measure the error of the models and compare with a random walk behaviour and draw consequences.
- Test the EMH hypothesis, in its **weak form** that holds that pri*ces have no memory and yesterday has nothing to do with tomorrow* (Goodman, 1968).
- Test the EMH hypothesis, in its **semi-strong** form that holds that that assumes that current stock prices adjust rapidly to the release of all new public information, in this thesis, the values of other world indexes.
- Confirm the existence of the ***momentum effect*** and assert if an investor can benefit from it.

Structure

The main body of the thesis starts with a revision of the available literature, that will be divided in three areas. First a review of the EMH since its enunciation by Eugene Fama, explanation of the three forms and the debates regarding the different anomalies and their different justifications. Second a review on the application of the computational methods to the financial markets covering some of the most common applications and use of Artificial Intelligence. Finally, a brief overview about the PSI20 index, as part of the Euronext ecosystem, and some of the research on this index using computational methods.

On the second Chapter the methodology used in the research is presented. It starts with an overview of the "research setup" used including the software and libraries used. It is followed by a description of the main methodologies used such as the ANN concepts and application, and the differences with Recurrent Neural Networks (RNN) using LSTM. Finally, the prediction workflow will be presented starting from the input data to the model evaluation.

The third & fourth Chapters show the experiments performed with the LSTM models, how the models were parametrized, trained, tested and for what purpose always with connection to the underlying EMH theory. On Chapter three, forecasts are performed with daily data and on Chapter 4 forecasts with intraday data. The experiments will always be followed by the author's considerations on the results and in the end of each chapter conclusions will be drowned.

The last Chapter of this paper presents this thesis' main conclusions. The main research goals are faced with the experiments' results and future investigation possibilities will be laid as this field of study is dynamic and continually evolving, and new findings can enhance our understanding of the topic.

LITERATURE REVIEW

This Chapter is divided in three parts, a general discussion on EMH, the developments of AI methods to analyse the stock markets and the PSI20 case.

Discussion on the "Efficiency of the Markets"

The Efficient Market Hypothesis has been a subject of significant debate among economists and financial experts. In this section a global perspective will be provided based mostly on *The Efficient Market Hypothesis: A Critical Review of the Literature* (Naseer & Tariq, 2015) and *Mercados, Produtos e Valorimetria de Ativos Financeiros* (Fernandes, Mota, Alves, & Rocha, 2014).

Defining Efficient Markets

The EMH is an investment theory attributed to Eugene Fama's research detailed in his 1970 paper, "Efficient Capital Markets: A Review of Theory and Empirical Work" (Fama E. , 1970). Fama advocated that financial markets are efficient and that asset prices fully reflect all available information at any given time. In other words, it is difficult or impossible to consistently outperform the market through active trading or stock picking.

A market is considered efficient if price reflects all available information regarding assets as prices display no memory. This assumes that all information relevant to stock prices is freely and widely available, "universally shared" among all investors therefore rapidly (instantly) incorporated on the asset price, so stocks always trade at their fair market value. In this scenario, price is consistent with the risk of the asset and identifying and exploiting arbitrage opportunities is not a consistent strategy to beat the market as both historical and fundamental (present) analysis cannot assist the investors to predict future values. According to this premisses, security prices are mostly based on an exponential random walk model, and **the best forecast to future security value is the present-day value, as it already reflects all information past and present** available. The EMH rest in three assumptions: 1) Investors maximize their utility. 2) If investors are not rational, their randomness neutralizes any effect on prices. 3) Arbitrageurs' actions eliminate any temporary deviation of a security value caused by irrational agents.

The Three forms of the EMH

The Efficient Market Hypothesis can be broken down in three forms. The ***weak form*** states that the present asset price reflects all historical price data, therefore, consistently with the random walk hypothesis, successive prices changes are independent random variables. Therefore, it is not possible to anticipate a security price change based on technical analysis, where an analyst, or an IA algorithm, predicts stock prices based on past market information. This concept was first presented by Fama and Blume (Fama & Blume, 1966) when they observed that, several trading filters could not beat the buy-and-hold strategy. The ***semi-strong form*** of the EMH adds to latter, that market adjusts rapidly to any public information such as dividends announcements, news, political events, economical events, or other securities / markets evolution. Finally, the ***strong from*** states that the private information (inside information) is also reflected in the present security's price.

The debate around the universality of the EMH

Many researchers, economists and investors disagree with the **universality** of EMH, arguing that price stocks can be predicted to some extent and that is possible for an investor to beat the market consistently as *The Superinvestors of Graham-and-Doddsville* (Buffett, 1984), defend. Many authors such as Lo and McKinlay (Lo & MacKinlay, 1999) claim to have found temporal patterns that contest the EMH, and the studies on behavioural finance from authors such as Daniel Kahneman and Amos Tversky presented alternative research scope for securities' price variation based on behavioural psychological theories.

EMH Anomalies

The EMH is subject to **anomalies** that are defined as an indiscretion or a deviation from an ordinary or natural order or an extraordinary condition (Frankfurter & McGoun, 2001). Researchers have been trying to not only detect and understand anomalies in securities' price movement but also if it is possible to determine a strategy negotiation that allow an abnormal rentability (Fernandes, Mota, Alves, & Rocha, 2014). The ***size Anomaly*** is probably the strongest pattern. As examined by Keim (Keim, 1983), small firms display higher returns than larger companies after adjusting the risk with the Beta.

Calendar anomalies have also been documented such as the ***January return*** (Haugen & Jorion, 1996) or the ***Monday Return*** (French, 1980). **Price to Earnings (P/E) ratios and**

dividend yield have been extensively researched to measure their predictive power, with contradictory results (Fluck, Malkiel, & Quandt, 1997) (Campbell & Shiller, 1998).

Temporal Patterns

Another group of anomalies detected are temporal patterns. The **momentum effect** in stock prices refers to the tendency of stocks that have performed well in the past (past winners) to continue their upward trend, becoming future winners in the short term while stocks that have performed poorly tend to continue their downward trend. This phenomenon suggests that past price movements can influence future price movements, creating a momentum effect in the market (Jegadeesh & Titman, 1993). Investors and traders often use momentum strategies to identify stocks that have exhibited strong price momentum in the hopes of profiting from the continuation of the trend. This effect has been observed in various financial markets and is believed to be driven by investor behaviour, market psychology, and herding instincts.

On the other hand, the **reverse pattern**, is the tendency of stocks that have experienced significant price increases or decreases to eventually reverse and return to their average levels, which could favour a *contrarian strategy* (Lo & MacKinlay, 1990). This pattern suggests that over time, stock prices tend to fluctuate around their intrinsic values, and extreme deviations from these values are likely to be corrected.

Conclusion on the debate

Although these anomalies have been thorough documented and studied, there is no real consensus if they have enough magnitude to be exploited by an investor to obtain an abnormal return. In most cases the results are marginal and diluted in the brokerage fees (Malkiel, 2003). EMH defenders, advocate that even if anomalies are found, they tend to disappear quickly by the action of the arbitrage.

Computational methods applied to stock markets analysis.

Brief overview

Since mid-90s, as computational methods firmly grasped on the finance world, significant research on applying Artificial Intelligence (AI) to stock market investments has taken place.

The use of computational approaches to automate the investments has several advantages, such as removing the influence of emotions on decision making, uncovering patterns that may be overlooked by humans, and providing real-time analysis of information. (Ferreira, Gandomi, & Cardoso, 2021)

Despite the widespread use of automation in hedge fund trading, a significant portion of these operations are still carried out using hardcoded algorithms (Chang, 2018). This suggests that there is still a significant potential for growth and advancement in the use of artificial intelligence in this area.

AI application areas

AI is commonly used in finance in three main areas (Ferreira, Gandomi, & Cardoso, 2021), ***optimizing financial portfolios***, using as benchmark the Markowitz Modern Portfolio Theory, ***predicting future prices or trends in financial assets*** and ***sentiment analysis of news or social media*** comments related to the assets or companies.

Predicting or forecasting stock market trends using historical time series data has become a common technique used by researchers and investors to gain financial profits in stock trading. These predictions, which were initially done using statistical methods, are now increasingly being done using AI algorithms. As a result, the application of AI in investments has become a rapidly growing field of research with a significant number of publications, as shown in Figure 1.

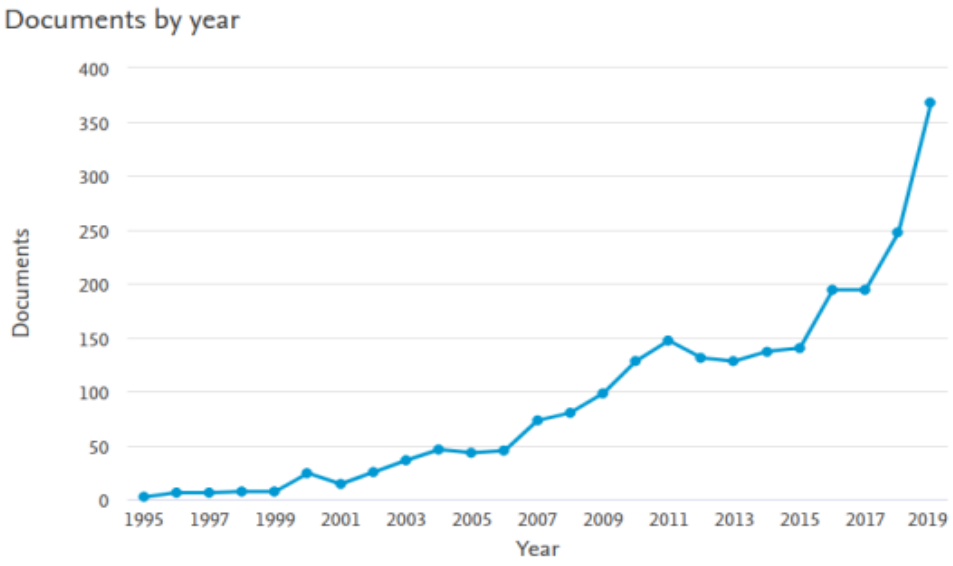


Figure 1 Scientific papers on AI applied to financial investments. Source: (Ferreira, Gandomi, & Cardoso, 2021)

Computational methods evolution

The computational efforts to contest EMH were historically divided in two main branches: 1) Technical analysis, which focuses on analysing historical trends in stock prices to make predictions, and 2) Fundamental analysis, which looks at the socioeconomic factors affecting a company and its operations to forecast future stock prices. (Ferreira, Gandomi, & Cardoso, 2021). For the technical analysis, linear statistical models such as Auto Regressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), gained increased popularity for the technical analysis during this century. Example of this research is the one conducted in 2019 by Idrees et al entitled "A Prediction Approach for Stock Market Volatility Based on Time Series Data" (Idrees, Alam, & Agarwal, 2019).

More recently the focus is shifting to Machine Learning and Deep Learning models. The use of these technics is facilitated by the continuous improvement of programming packages, making easier to implement and deploy new models and by the ever-increasing availability of historical financial data and text data from online news and social networks (Jiang, 2021).

Prediction models available

Within the linear statistics models often available we highlight the mentioned ARIMA and GARCH, the Linear Regression, Logistics Regression, Support Vector Machine and k-Nearest Neighbour. Within the Machine Learning methods there are three categories: 1) Feedforward Neural Network (FFNN), the simplest type of ANN, 2) Convolutional Neural Network and 3) Recurrent Neural Networks (RNN) that include LSTM (Witten, Frank, Hall, & Pal, 2016).

Recent Studies comparing LSTM with SVM, Decision Trees, Bagging, Random Forests and simple ANN show that LSTM algorithm perform consistently better than the others (Vignesh, 2020), (Nabipour, Nayyeri, Jabani, Mosavi, & Salwana, 2020). Following these conclusions from now on this thesis will focus on RNN with LSTM.

Prediction models main issues

Target Output and Frequency

Prediction models can be classified in 4 types of bases on target output and frequency. Target refers to classification (e.g., going up or down), or regression (predicting a specific price). Both targets can result in financial rules to assist investors. According to Liu & Wang, 42% of the

papers are focused on *daily classification*, 44% on *daily regression* and only 6% and 9% on *intra-day regression* and *intra-day classification* (Liu & Wang, 2019).

Prediction Workflow

The Prediction workflow can be divided in 4 steps (Jiang, 2021) & (Ferreira, Gandomi, & Cardoso, 2021): Input data acquisition, Data Processing, Prediction model and evaluation.

Input data acquisition - There are several data types that can be used as for a model such as: Market Data, that can be used as an input and as a prediction target; Text Data, that is the basis for the sentiment analysis, Macroeconomics Data, Fundamental Data, and secondary data types such as Knowledge Graphs, Image Data and Analytics Data. Length of used series is an important factor as there is a trade-off between completeness and computational costs.

Data Processing – For model accuracy Data needs to be treated, transformed, or reduced to solve inconsistencies as missing data issues and remove the noisy information. Other common practice (Witten, Frank, Hall, & Pal, 2016) is the normalization and standardization of the data to control the effect of outliers. The data is then divided in two subsets: 1) the **training set** that will be used to train the model and adjust the hyper-parameters and 2) the **test set**, that will be used to test if the model's performance and avoid over-adjustment to the training set.

Model Training- During training, the model iteratively updates its internal parameters to minimize the discrepancy between the predicted outputs and the actual target values. This optimization process is typically performed using an algorithm like stochastic gradient descent (SGD) or its variants.

Model Evaluation – The evaluation of the models is usually grouped in 3 main types: 1) *Classification Metrics* that measure the model's performance on movement predictions (going up or down). Examples of this are *Accuracy* and *F1 Score*. 2) *Regression Metrics* that compare the model with the actual stock price performance. Examples are Mean Absolut Error or Root Mean Squared Error (RMSE). 3) *Profit Analysis* that evaluates if the predicted-based trading strategy can bring profit or not. For correct evaluation it is usual to compare with a baseline that can be generated by a *buy-and-hold* strategy or a *momentum* strategy. This can be complemented with algorithmic trading.

Technologies review

Hatcher & Yu (Hatcher & Yu, 2018) make a comprehensive introduction of deep learning tools and they point up Keras and TensorFlow as the dominant frameworks for deep-learning stock prediction. Other packages used are PyTorch, scikit-learn and sktime (Raschka, Liu, Mirjalili, & Dzhulgakov, 2022). For data manipulation the Python libraries Pandas, NumPy, Matplotlib among others, are considered standard.

Yadav et al in their paper "Optimizing LSTM for time series prediction in Indian stock marke**t"** show a straight-forward implementation, with setup details, of Deep-learning technics to the Indian Market, with focus on the influence of the ***Hyper-parameters*** (Yadav, Jhaa, & Sharanb, 2020). Among other details they show the activation function (sigmoid, tanh, etc.), the best optimizers such as Adam and Adadelta, (which are subtypes of SGD), the batch size the importance of number of epochs and number of layers and units per layer.

Future research directions

Future research direction identified on literature (Jiang, 2021) point to: 1) New models, as some neural networks are not fully deployed for market predictions, 2) using Multiple Data sources, as it is wiser to find new data sources that try to out-perform existent studies, 3) Cross Market-analysis, as most of the studies focused mainly in one market., 4) Algorithmic trading, as the prediction is only the beginning of the story.

Portuguese Stock Index - PSI20

Euronext was founded on September 2000 through the merger of the stock exchanges of Paris, Amsterdam, and Brussels, benefiting from the harmonization of financial markets in the European Union. The group expanded its presence in Europe in 2002 with the entry of the Lisbon and Porto stock exchange, in 2018 joined the Irish stock exchange and in 2019 joined the Oslo stock exchange. The Euronext stock market reached a capitalization of EUR 4,409,881 million in 2020, making it one of the largest global markets (Gomes L. M., Soares, Gama, & Matos, 2022).

Some of the literature regarding Research in Portuguese stock PSI-20 / Euronext – Lisbon, includes (Matos, Gama, Ruskin, & Duarte, 2004), with focus on an *econophysics* approach

demonstrating that the Portuguese stock behaviour is getting closer to a developed market stock.

Other Portuguese studies (Gomes L. , Soares, Gama, & Matos, 2018) show evidence that the PSI20 indicate signs of long-term memory in the form of persistence. The structure found suggests that the market is subject to greater predictability than expected by the EMH fundaments.

(Sobreira & Louro, 2020), use linear regression methods such as GARCH to estimate the Value-at-Risk at the Portuguese stock market. Empiric studies using deep-learning methods with LSTM to predict the PSI20 are scarce.

METHODOLOGY

The literature review points to Deep Learning (particularly LSTM) as a promising field for making accurate previsions on stock markets. Models with this methodology will be developed using as an input the PSI20 market history and combining it with other stock markets around Europe and the world, among other variables.

Software

This project will run on Python programming language, based on the Anaconda distribution, one of the world's most popular data science platforms. Standard python libraries for data science such as Pandas and NumPy and Matplotlib will be used.

For deep learning models, choice will be with the industry standard **Keras**, working on top of **TensorFlow**. Keras, developed by Google, is a human-centric deep learning API (Application Programming Interface) that simplifies complex tasks with and simple interfaces that minimizes user actions.

Deep Learning

Deep learning is a subset of machine learning that focuses on building models that self-improve by analysing data (training). Unlike traditional machine learning, which employs simpler concepts, deep learning uses artificial neural networks to try to mimic human thinking and learning processes replying the "neuron" concept (Salem, 2022). In the past, limitations in computing power constrained the complexity of neural networks. However, advancements in hardware and high-level libraries have facilitated larger and more sophisticated networks, enabling models to rapidly observe, learn, and respond to complex situations and surpassing human capabilities in some scenarios.

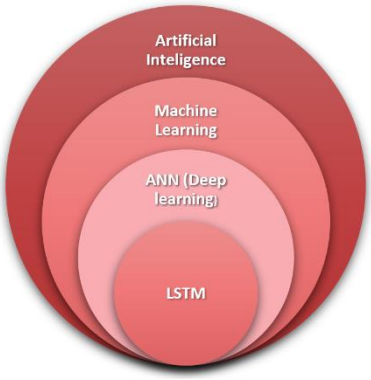


Figure 2 Artificial Intelligence Diagram

Artificial Neural Networks (ANN)

An ANN consist of interconnected layers of nodes, similar to how neurons form the human brain. Just like a single neuron in the human brain receives signals from numerous other neurons, artificial neural networks transmit signals between nodes while assigning weights to them. The weighted inputs are combined in the final layer to generate an output.

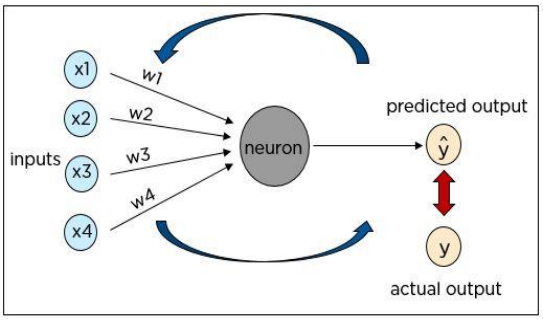


Figure 3 ANN flowchart – Source: (Simplilearn, 2023)

During an ANN training, the goal is to minimize the overall loss, through a process of *backpropagation*. This involves feeding input data through the ANN, comparing the outputs to the target values, and calculating gradients using backpropagation through time (Witten, Frank, Hall, & Pal, 2016). The gradients are then used to update the model parameters, aiming to iteratively minimize the loss. Despite advanced hardware, some neural network can still take weeks to train.

The Recurrent Neural Networks (RNN) are a subtype of ANN suitable for time-series analysis. Unlike the ANNs, they can receive an array of data of unknown size as an input. They are capable of recognize patterns in data sequences, such as those that may appear in stock prices given that, in addition to the actual value, they also include its position in the sequence in the prediction.

The RNN limitations and the Long Short-Term Memory (LSTM)

LSTM model is a subtype of a RNN (Salem, 2022). The problem with RNN is that they have a short-term memory to retain previous information in the current neuron. However, this ability decreases/or increases very quickly for longer sequences, generating the vanishing/exploding gradient problem, making the model advance so slowly for the best adjustment that it never gets there, or making leaps so large that it keeps passing he point where the loss is minimized.

LSTM addresses those issues by introducing three essential gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information within the LSTM cell.

Input Gate: The input gate determines how much new information should be stored in the cell state.

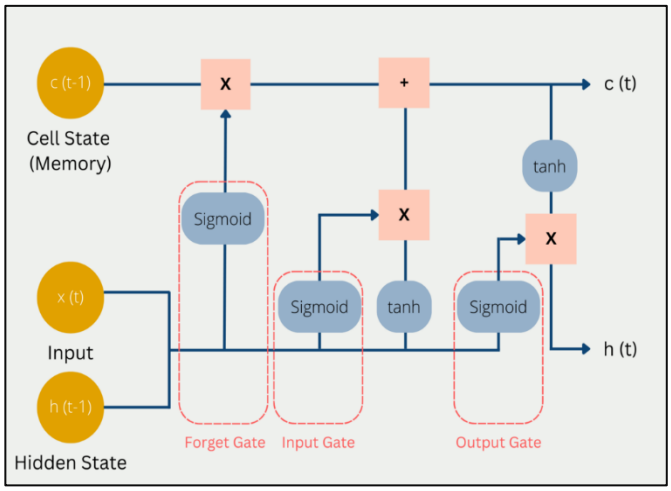


Figure 4 LSTM Architecture - Source: (DATA BASE CAMP, 2022)

Forget Gate: The forget gate controls the extent to which previous information in the cell state should be discarded. It takes the current input and the previous hidden state as inputs and produces values between 0 and 1 using a sigmoid activation function. A value of 0 means "forget all previous information," while a value of 1 means "keep all previous information."

Output Gate: The output gate determines the amount of information to be outputted from the cell state.

By using these gates, an LSTM can selectively store, forget, and output information, enabling it to effectively capture and retain relevant long-term dependencies in sequential data and mitigate the vanishing gradient problem commonly found in traditional RNN (Salem, 2022).

Prediction Workflow

This section describes the prediction workflow of the experiments/ tests performed, starting with the input data acquisition, data processing, parameter optimization, model definition and training and the evaluation of the model's prediction.

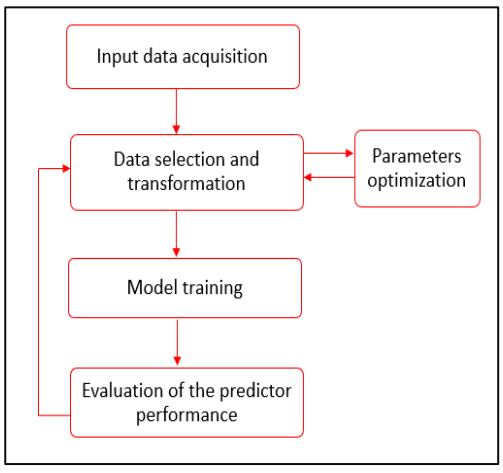


Figure 5 Flowchart for financial forecast with AI

Input data acquisition and data processing

Yahoo finance API (Application Programming Interface) provides the necessary data for the

model, in a data set containing all daily stock prices for the PSI20 and other world stock indexes from the last 10 years, until February 2nd, 2023.

Processing involves striping down unwanted columns from the dataset and rearrange the data in two arrays. The target output (Y_Data) is a simple one-dimension array if the target output is a numeric value, such as a price prediction, and a *one-hot-encoded* two-dimensional array if the target output is a classification. The training labels (X_Data) is two-dimensional, as the model will consider a sliding window (or rolling window) of "N" values to predict the next day Y value.

It is also necessary to create an X_Test array, with supplementary training labels to perform predictions with the models and compare them with the actual observed values, allowing to draw conclusions. The rearranging is done with Python using the Pandas and NumPy libraries.

Y_Data (INDEX VALUE)	Y_Data (Classification)
0.218879	1 (Stable)
0.214028	1 (Stable)
0.217051	1 (Stable)
0.255591	2 (Value increase)
0.236421	0 (Small decrease)
0.283769	2 (Value increase)
0.233263	0 (Value decrease)

Table 1 Possible Target Outputs

Data Scaling Methods (DSM)

Two DSM are using this thesis the first is scaling between 0 and 1 as the ANN Activation functions are generally designed to work well within this range.

In other experiments the log return is used. The log return is the logarithmic difference between the price of an asset at two different times. It can be expressed, for a period *t*, as:

$$log\_return\_t = log(price\_t / price\_\{t-1\})$$

Target output and frequency

Frequency is the interval between observation that will be used as train data to our model. This investigation will start with daily data analysis as it is easily accessible, and lighter on computation. On a second phase intraday data will also be used.

Data for training Deep Learning models is divided in 1) training samples, or 'X' data, for the input data and 2) training labels, 'Y' Data for the output data. This output can have different formats either a numerical value, representing a price prediction or a classification output such as: Large Increase / Small increase/ etc, as can be observed in below table.

Target Output	Frequency
Prices regression	Daily
Classification	Intraday
Negotiation Rules	

Table 2 Target output and data frequency.

The definition of the classifications will depend on the percentual variation of the price in comparison with the day before. This approach will allow the generation of a negotiation rule, such as: "If *Large Price Increase*, then Buy stocks", and through this negotiation rules generate a simulation within a fixed period that can compare itself with a *buy-and-hold* strategy during the same period.

Hyperparameters

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning. The prefix 'hyper' suggests that they are 'top-level' parameters that control the learning process and the model parameters that result from it. These parameters are hard to adjust so this investigation will take as starting point other investigations such as: (Bhandari, et al., 2022), and will further attempt to tune them better as the investigation proceeds.

Some examples are the Number of iterations (epochs), the number of layers, the size of the sliding window (Look_Back), the number of neurons in each layer, the train-validation split ratio, the learning rate, the optimization algorithm, the loss function, the drop-out rate, and the batch size.

The size of the sliding window is the number of observations the model considers for making a prediction and is during this thesis used interchangeably with the name of the variable Look_Back.

Model definition and training

Although the complexity of the LSTM the Keras-TensorFlow API provides an accessible interface to design RNN with *LSTM* layers and a final *DENSE* layer that capture the model output. Below there is an example of an implementation for a LSTM model in Keras.

```
lstm_model = Sequential()
lstm_model.add(LSTM(units=LOOK_BACK, return_sequences=True,
               input_shape=input_shape))
lstm_model.add(LSTM(units=50)
lstm_model.add(LSTM(units=25))
lstm_model.add(Dense(units=1))

# Compiles the model
lstm_model.compile(loss='mse,optimizer=Adam(learning_rate=LEARN_RATE))
```

The code starts by initializing a sequential model a linear stack of layers, where each layer is added one by one. Then adds three LSTM layers to the model. Each LSTM layer processes sequences of data and has a certain number of memory units (or neurons) specified by the `units` parameter. The first LSTM layer also receives the `input_shape` parameter, which defines the shape of the input data. After the LSTM layers, a DENSE layer with a single unit/neuron captures the model prediction (continuous). It the model output is categorical (discrete), the DENSE layer has one unit for each possible output, providing the "probability" of each one.

Finally, the code compiles the LSTM model. The `loss` parameter is set to 'mse' (mean squared error), which is a common loss function used for regression problems. For a categorical output model, the loss function used is the 'Categorical Cross Entropy'.

The `optimizer` parameter is set to Adam, which is a stochastic gradient descent (SGD) algorithm, and the learning rate is specified using the `learning_rate` parameter.

The next step is the "training" of the model with the below code:

```
def trainLstmModel (model, x_train, y_train, epoch=EPOCHS):
    return model.fit(x_train,
                y_train,
                batch_size=BATCH_SIZE,
                epochs=epoch,
                validation_split=VALIDATION_SPLIT,
                verbose='auto')
```

The above function takes as parameter a `model` (the LSTM model), `x_train` (input training data), `y_train` (target output), and as hyperparameter the number of training epochs.

Inside the function, the model is trained using the `fit` method. It takes in the training data (`x_train` and `y_train`), as well as other parameters such as `batch_size` (size of each training batch), `epochs` (number of training iterations) and `validation_split` (fraction of the training data to be used for validation).

The function returns the result of the fit method, which contains information about the training process and performance and allow us to make predictions based on a `x_test` array:

```
predicted_stock_price= lstm_model.predict(X_test)
```

Model evaluation

After training the models will be presented with a test set, that will allow to model to typically make predictions for 50 periods (days), based on a sliding window of N values for each predicted output.

Those predictions are compared with the actual values and evaluated based on two metrics. The RMSE is used when the models predict a continuous value such as the index value and the "accuracy" is used for the discrete predictions when models predict a classification output. The results *per se* of metrics either RMSE or Accuracy, are not self-explanatory, so they are compared with a benchmark, or a baseline value based on the EMH fundaments.

According to the EMH the index values follow a Markov process just as the exponential random walk distribution (Kijima, 2002) so the best estimator for tomorrow's stock value is today's stock value. When the output of the models is the predicted index value it will be compared with the previous day. if the model consistently provides a better adjustment than the one provided by the "previous day", it means that a significant prediction would be achieved, challenging the EMH applicability for this market. In models that predict a classification, the results will be compared with an average "random" picking.

The models, and the experiments, are re-adjusted as results are recorded and conclusions are drawn.

FORECASTS WITH DAILY DATA

In this Chapter experiments performed with daily data from the last 10 years will be presented in detail. It starts with an Exploratory data analysis of the PSI20, then three groups of experiments are presented:

Group 1 – Experiments where only the PSI20 historical data is used to try to forecast the next closing value(s). Mostly the aim is to use deep learning algorithms to try to find temporal patterns in the shape of a *momentum effect* that allow a significant prediction that could challenge the EMH in its weak form.

Group 2 – Experiments where the PSI20 historical data is used alongside other variables such as the trading volume and the weekday, with same purpose of making a significant prediction.

Group 3 – Experiments where the focus shift to a cross-market analysis. Prediction models are trained with historical data from other stock market indexes to try to find delays on the incorporation of information that could allow an investor to have an abnormal positive return.

Exploratory Data Analysis

In time-series data, there are some important things to consider before fitting machine-learning models such as trends in the data, seasonality, long-run cycle, constant variance, or abrupt changes.

PSI20 Close	
count	2487
Mean	5357.71
std	681.55
min	3596.08
25%	4933.60
50%	5283.65
75%	5705.21
max	7734.95

Table 3 - PSI20 Stats

Figure 6 PSI20 last 10 years closing value.

The above chart represents 2487 observations of the closing value of the PSI20 from 30/04/2013 to 01/02/2023. We can observe a few abrupt changes coincident with important events in Portugal and in the world such as a sudden drop in 2014 caused by the resolution of one of Portugal biggest banks, the "BES" (Antunes, 2014), and the March 2020 abrupt descent leading the Portuguese stock index to a 10-year minimum of 3596.08, originated by the Covid-19 Pandemics. Trends in the data, seasonality or long-run cycle are not observable on first-glance observation.

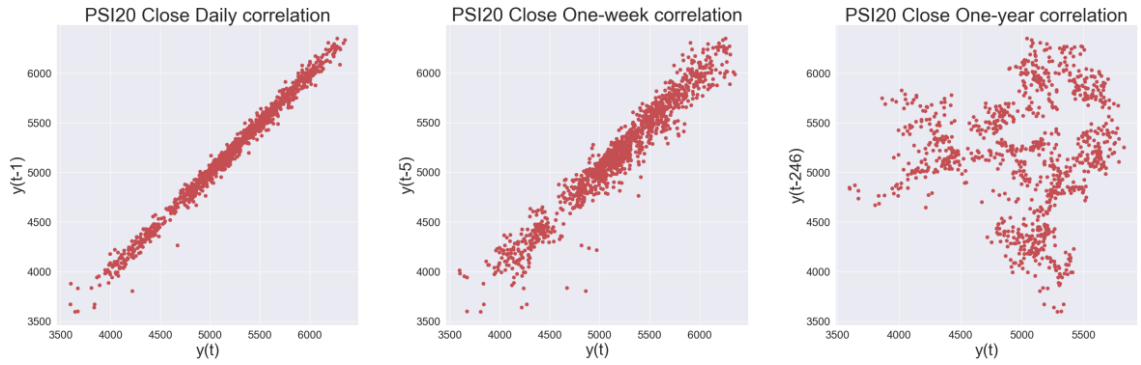


Figure 7 PSI20 Closing value correlation between y(t) and periods y(t-1), y(t-5) and y(t-246).

The above plots show the correlation between the day's closing value and the closing value of the previous day, week, and year respectively. We can see the daily and weekly charts are almost linear. It means there is a correlation between the closing values. It is expected that the price of the stock would not change much in a week and even less on a day. In the yearly scatter graphic, the graphic seems much more random.

The observations line up for the daily data, so it suggests a non-random pattern (Graphic is positively linear). Non-randomness in the data reveals that an autoregressive model can be used.

Group of experiments #1 – Predicting PSI20 based on its close value history.

On this first group of experiments the PSI20 historical data will be used to train models that will aim to predict the PSI20 next daily closing value. These models will try to benefit from the documented "momentum effect" of stock prices to try to elaborate a relevant prediction.

Experiment #1A – Using absolute values to predict next daily closing value.

The first experiment performed aims to determine if the knowledge of the previous 90 trading days of market data can contribute to predict tomorrow's PSI-20 value. The algorithm is trained with the knowledge of the last 5 or 10 years of data, based on a sliding window (Look_Back) of 90 days of trading for each target output.

Data processing

The Training samples array or the X_Data is a multidimensional array with the scaled previous 90 days of data for each timestep. The Y_Data is the target output of the model. The model will try to fit each line on the X_Data array to the value on the Y_Data array. Sample of the data, already processed for the model training, is observable in below Tables 3 and 4.

X_DATA (Training Labels)

0	1	2	3	4	5	…	90
0.1845	0.1411	0.0802	0.1216	0.0858	0.0829	…	0.2067
0.1411	0.0802	0.1216	0.0858	0.0829	0.0903	…	0.2188
0.0802	0.1216	0.0858	0.0829	0.0903	0.1004	…	0.2140
0.1216	0.0858	0.0829	0.0903	0.1004	0.0837	…	0.2170
0.08580	0.0829	0.0903	0.1004	0.0837	0.1134	…	0.2555

Table 4 Sample of the 'X' Data for experiment #1A

Y_Data (Target Outuput)
0.218879
0.214028
0.217051
0.255591
0.236421

Table 5 Sample of the target output

Hyperparameters

For this first test several different combinations of hyperparameters are tested. Results of each combination can be observed in the table with test results. After some preliminary tests the ADAM optimizer and a batch size of 50 were used as a base for the remaining combinations. The data is divided in training data and validation data using a validation split of 0.1, to detect the existence of overfitting. The Loss function is the MSE (Mean Squared Error). The model definition is presented in the below table.

Layer (type)	Output Shape	Param #
Lstm (LSTM)	90, 50	10400
Lstm_1 (LSTM)	25	7600
Dense (Dense)	1	26
Total params: 18,026 Trainable params: 18,026 Non-trainable params: 0		

Table 6 Experiment #1A model summary

Model evaluation

For evaluating the model, the chosen metric was the Root Mean Squared Error (RMSE). The value will be compared with the RMSE provided by the "previous day". The value of the RMSE for the "previous day", for the last 5 years of data, was on February 2nd, 2023, of 42.33.

Test results

Below are some of the most relevant test results. Full table can be found in the appendix.

Test#	Years Data	Epochs	Learn Rate (10^{-4})	Nlayers	Units per layer	Prediction RMSE	Training MSE (10^{-4})	Validation MSE(10^{-4})
A2	5	200	5	2	50;25	**42,27**	4.4	n/a
A3	5	200	5	3	50;30;10	57,47	4.2	n/a
A13	5	200	5	2	50;25	43,45	4.4	5.9
A14	10	100	5	2	50;25	43,94	2.3	2.6
A15*	5	150	5	2	50;25	43.01	4.2	7.2

Table 7 Experiment #1A test results

(*) On test A15, the sliding window as a size of 5 observations.

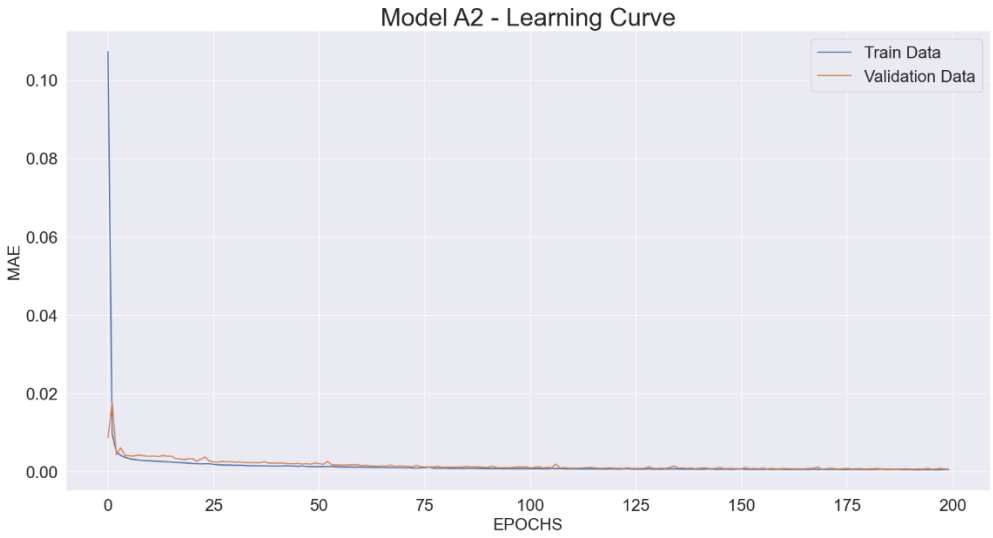


Figure 8 Model A2 - Learning Curve

Observations

There are a few conclusions that can be drawn from the above tests:

- The best RMSE obtained on test #A2 – 42.27 - is slightly lower than the RMSE of the "previous day". Test #A13 is a repetition of test #A2 and the RMSE is of 43.45, slightly higher than the RMSE of the "previous day".
- Models with 3 layers perform significantly worst that models with 1 or 2 layers, which concurs with the literature (Bhandari, et al., 2022).
- Models with two layers provide slightly better adjustment than single layer models. This is not consistent with results of Bhandari et al, and it can derive from using more EPOCHS in the present situation (200) than the mentioned research where only 25 epochs were used.
- The number of epochs shows evidence of relevancy on the quality of the adjustment, as can be observed in the "learning curve" graphic.
- The model does not display signs of overfitting, as the Mean Absolute Error of the training data is like the Mean Absolute Error of the validation data. (As observed in the *Figure 8*.)

It is also worth mentioning the following observations which seem to <u>confirm the EMH</u>:

Comparing the test #A13 with #A14 where the only difference was the use of 10 years of data instead of 5 years, we observe that the latter does not provide a better RMSE, which is

consistent with the EMH, as including more data does not improve the predictions, which reaffirms the notion that prices exhibit no memory.

Model #A15 differs from the other as the LOOK_BACK value is of only 5 observations (All others are 90 observations). The observed RMSE of this model's prediction is in line with the best results of the other models. This suggests that the size of the sliding window used to predict tomorrow's closing value has no impact on the quality of the prediction.

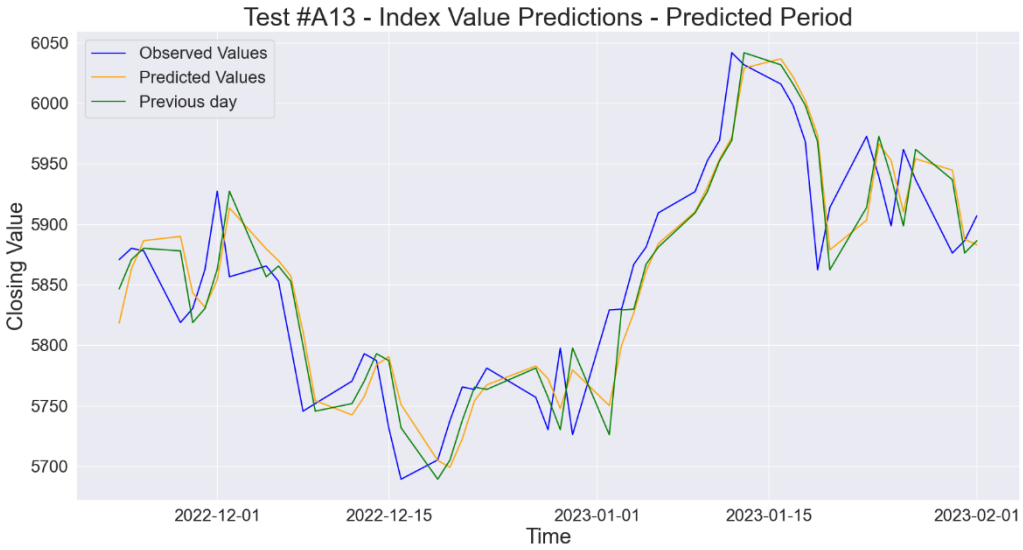


Figure 9 Model generated predictions of the PSI20 closing value compared with the observed values and the previous day.

A third, and more striking conclusion, comes from the analysis of the Index Value Predictions graphic. The "Previous Day" value is used as a benchmark as, according to the EMH, it is the best forecast to today's value. The observation of the graphic shows that the LSTM generated models **seem to try to adjust their predictions to the previous day, as it is the point where loss is minimized**. Therefore, the predictions plot line (yellow) and the Previous Day (Green) plot line almost **overlap**.

Experiment #1B – Using log return to predict price movement classification.

Description

Unlike the previous model the objective of this experience is to elaborate a model that classifies the day's closing value of the PSI-20 by its variation in respect to the previous day closing value. Three target outputs were defined, the index value can either go up, go down or remain stable.

Data Processing

The variation in respect to the previous day was obtained using the log return. These values were processed in an X_DATA in a sliding window of 90 days for each target output.

The target output consists in one of three values as shown in below table. The Lower 33% were marked as "Value decrease", the higher 33% as "Price Increase" and the remaining as "Stable". This method guarantees that all the 3 target categories have the same number of observations, to avoid the possibility of the model to just choose the target with higher number of observations,

A special care was brought on having the exact number of observations for each category, otherwise the model could adjust to always predicting the category with more observations.

Target Output	Number of Obs (for 10 years)
0 - Value decrease (Lower 33% values)	782
1 - Stable	782
2 - Value Increase (Higher 33% values)	782

Table 8 Target Output for experiment #1B

Model definition

The LSTM Model used in this experiment is a classification model, therefore will try to learn which of the three outputs is more likely to occur. The final layer has 3 neurons, that capture the "probability" of each of the outcomes. The other significant difference is the use of "categorical crossentropy" as the loss function instead of the MSE.

Layer (type)	Output Shape	Param #
Lstm_4 (LSTM)	90, 50	10400
Lstm_5 (LSTM)	25	7600
Dense_2 (Dense)	**3**	48
Total params: 18,078 Trainable params: 18,078 Non-trainable params: 0		

Table 9 Experiment #1B Model Summary

Model evaluation

The metric for evaluating the quality of the adjustment is the *accuracy*. As the samples are divided in three equal groups an *accuracy* value above 33% will outperform the expected random probability.

Test Results

Test#	Years of data	Training cat. crossentropy	Validation cat. crossentropy	Training Accuracy	Validation Accuracy	Test accuracy	Last observation
C1	5	1.049	1.0565	43.05%	44.04%	32.65%	01/02/2023
C2	10	1.060	1.1371	42.40%	32.34%	32.65%	01/02/2023
C3	10	1.053	1.1291	42.42%	33.01%	34.69%	01/02/2022

Table 10 Test Results of experiment #1C

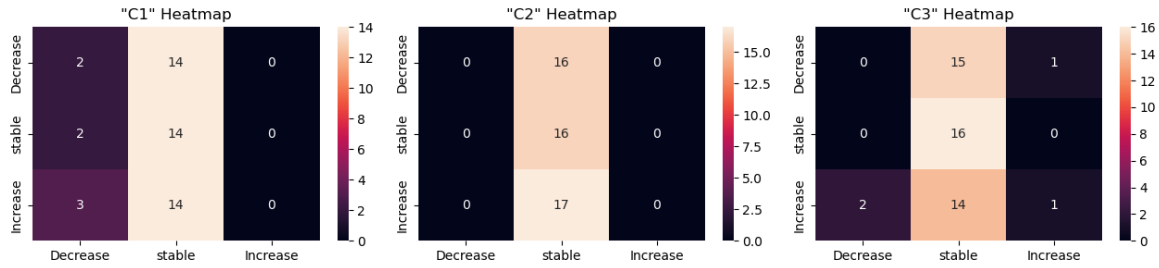


Figure 10 Heatmaps of the models' predictions compared to the actual observations.

Observations

The training output of the C1 test shows an accuracy of over 43% in training data as in Validation data, which is well above the 33% threshold, which gives a good indication in regards of the quality of this model. However, the subsequent tests using 10 years of data and a different end date on test C3, did not confirm the indications of the first test, as the Validation accuracies are around the 33% threshold.

When the trained models are fed with test data (not used in the training) models predict almost always a "stable" outcome. This is a disappointing outcome as it was expected that the model would provide different predictions based on the data.

The model fails to provide an accurate forecast of the PSI20 fluctuations. Mostly it predicts a central value with a accuracy close to 33% which is not significantly better than a random prediction.

Experiment #1C – Using log return to predict PSI20 absolute value.

Description

This experiment combines the approaches of the two previous experiments. It uses the log return as training labels and as target output. Using the cumulative sum, it is possible to generate predictions to get the absolute value of PSI20 next day's value. Information on data processing, hyperparameters and model evaluation are detailed on the appendix.

Test Results

Test#	Years of data	Epochs	Prediction RMSE	Previous day RMSE	End date	Look_back (days)
F1	5	150	42.09	42.33	01-02-2023	15
F2	5	150	58.18	58.48	01-04-2021	15
F3	5	150	59.44	60.21	01-03-2022	15
F4	5	150	128.10	126.71	05-05-2020	15

Table 11 Test Results of experiement #1C

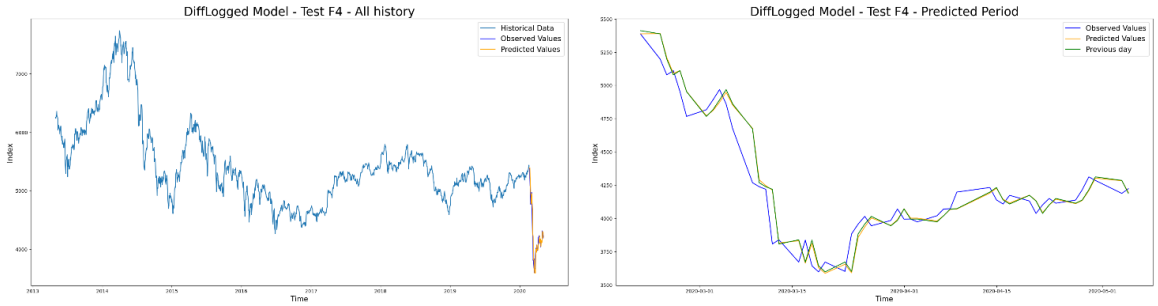


Figure 11 Model F4 generated predictions of the PSI20 closing value compared with the observed values and the previous day.

Observations

This model predicts the variations in regards of the previous day in the form of a difference of logarithms. All the predictions generated are very close to 0 (zero). Therefore, when the cumulative sum is performed to obtain the absolute value, the predicted value is very close to the previous day value.

Trying to predict the value of the PSI20 next day based on the daily variations provides a similar result of the experiment #1A where the absolute values of the PSI20 are predicted directly. As in experiment #1A, the predictions curve almost overlaps the previous day curve.

Test F4 is performed in the Covid-19 pandemic crisis, where high volatility is observed, and showed no significative difference in the behaviour of the model.

Experiment #1D – Expanding the time horizon of the predictions.

Description.

This experiment is based on the same principle of previous experiment, but this model will attempt to predict a broader time horizon. The goal is to observe if the model deviates from

the previous day's value by extending the time horizon. Additional information such as data processing, hyperparameters and model evaluation are detailed on the appendix.

Test results

Test#	Years of data	Epochs	Learn rate (10^{-4})	Nlayers	Units per layer	End date	Look_back (days)	Horizon (days)
G1	5	150	1	2	50;25	01-02-2023	60	10
G2	5	150	1	2	50;25	01-04-2022	60	10

Figure 12 Test Results of Experiment #1D

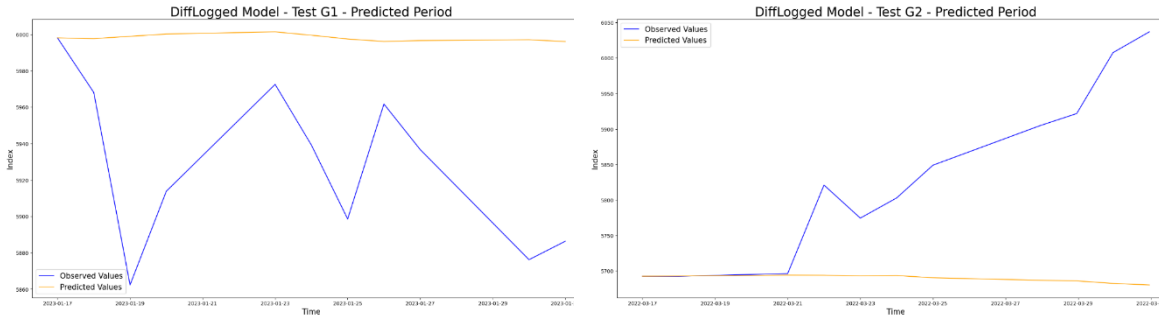


Figure 13 Comparison between the actual values predicted values in models G1 & G2 with 10 days horizon.

Observations

With this experiment the goal was to find if when predicting a broader time horizon, the model would drift from the previous day value and make a relevant prediction. The observation of the above Figures show disappointing results as the model is keeps predicting variations close to 0 (zero). Therefore, after converting variations to absolute value with cumulative sum, the prediction resembles a horizontal line based on the last known value.

Conclusions of Group of experiments #1

This set of experiments was performed to try to find evidence of a *momentum* effect that could allow a prediction able to outperform a benchmark of the previous day values. The results of the experiments show that the adjustment provided by the models is always very close to the previous day, as it is the point where loss is minimized.

When the target output is the variation, captured by the log return, the models predict a value close to 0 (zero). When the absolute value is then calculated using the cumulative sum, the value obtained is again very close to the previous day, as this value is obtained by applying the variation to the previous day.

When the prediction horizon is expanded to more than one day, the model predicts a horizontal line with the value of the last known value.

Increasing the amount of data or the size of the sliding window used to train the model, also does not seem to have a significant impact on the quality of the predictions, again consistent with the concepts of the EMH.

All the observation recorded on this set of experiments, do not confirm the existence of a *momentum effect* of the stock prices. The results confirm the EMH idea that prices hold no memory, and that present price already includes all the relevant information, and that the best estimator for tomorrow's prices is today's prices.

Group of Experiments #2 – Using PSI20 and a second variable

This second group of experiments is based on bivariate models that includes, besides the historical PSI20 values, a second variable that is related with the PSI-20, be it the trading volume or the weekday.

These experiments will allow to look for evidence of *Calendar anomalies* based on the day of the week, and for anomalies caused by technical indicators particularly the volume of trading.

Experiment #2A – Using volume alongside the closing value to predict PSI20.

Description

This is a bivariate model that includes, besides the historical PSI20 values, the Portuguese stock market daily trading volume in the same period.

Test results

Test#	Years of data	Epochs	Learn rate(-04)	N layers	Units per layer	Prediction RMSE	Training MSE (-04)	Validation MSE (-04)	End date
B1	5	200	5	2	50;25	43.15	4.73	6.80	01-02-2023
B2	10	200	5	2	50;25	44.67	2.24	2.58	01-02-2023

Table 12 Experiment #2A Test Results

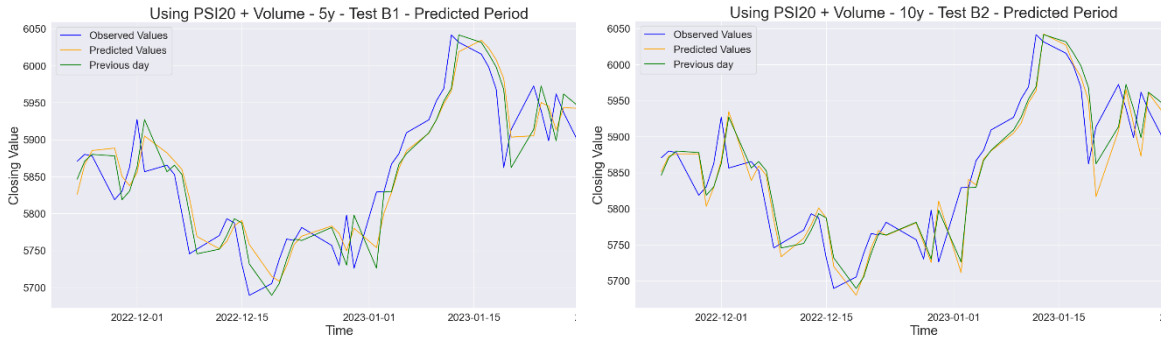


Figure 14 Predictions generated in tests B1 and B2 compared to previous day and to actual values.

Observations

This model displays an adjustment close to the one provided by the univariate model from Experiment #1A. The RMSE value in test B1 is 43.15, and of B2 is 44.67 which are rather close to the previous day's benchmark that is 42.33.

Training MSE and Validation MSE are close enough to conclude for reduced overfitting.

Observing the Figure 15, on the B1 model, the prediction yellow line "drifts" away from the previous day green line more than in the Experiment #1A, while keeping a similar RMSE. It is a more interesting prediction, as the model does not simply mimic the previous day line.

The use of 10 years of data for training the model (instead of 5) did not improve the quality of the prediction, same as in the first group of experiments.

Experiment #2B – Using the weekday alongside with PSI20 historical data.

Description

With this experiment the objective finding fluctuations of price that could be based on the day of the week. For that purpose, the model is trained with a sliding window of the last 60 trading days and the weekday information to try to access if the predictions improve with the inclusion of the day of the week information. A good adjustment would be strong evidence for the presence of calendar anomalies. Further details, including the method to encode the weekday can be consulted in the appendix.

Test Results

Test#	Years of data	Epochs	Learn rate	Units per layer	Predicti. RMSE	Previous Day RMSE	End date	Weekday encoding
I11	5	150	0.001	50;25	43.77	42.33	01-01-2023	TS (0-4)
I12	5	150	0.001	50;25	42.18	42.33	01-01-2023	Last day
I13	5	150	0.001	50;25	41.92	42.33	01-01-2023	Control
I12	5	150	0.001	50;25	44.89	42.33	01-01-2023	OHE

Figure 15 Experiment #2B test results.

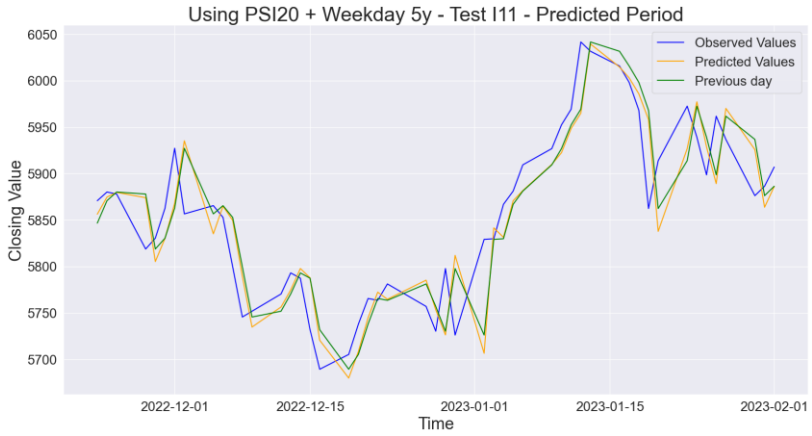


Figure 16 Predictions generated in test I11 compared to previous day and to actual values.

Observations

Including the day of the week did not have any impact on the quality of adjustment as the model that was fed with blank data on the weekday (control) was not outperformed by the models that had information regarding the day of the week.

Conclusions of Group of experiments #2

These experiments allow to check for influence of technical indicators, in this case the volume, and of calendar events (weekday), on the Portuguese stock index closing value. The models were presented with PSI20 data, same as in group of experiments #1, and a second time series of: "Trading volume", in experiment #2A and weekday in experiment #2B.

The results for the training volume are not better than results without the "trading volume" in terms of prediction error. However, when the graphics are observed, we see that the prediction line drifts apart from the previous day, more than on models that not include the "trading volume", while maintaining a similar loss. This suggests that further research is necessary to clarify the relevance of including trading volume on predictions models.

Including the weekday on the model, with different strategies, did not have any impact on the quality of the adjustment. In fact, the model that had better adjustment was the one where, for control purposes, the model was fed with "blank" data on the weekday. Therefore, no evidence of a calendar anomaly based on weekday was found in this experiment.

Group of Experiments #3 – Using a cross-market approach

On this third set of experiments, a cross-market approach is adopted. This was pointed out by literature as a promising and understudied field of computational methods applied to stock markets predictions (Jiang, 2021).

The semi-strong form of the EMH will be tested on the PSI20 context. A model able to outperform a benchmark constitutes evidence that information, in this case other world stock indexes value, is not instantly being incorporated in the Portuguese index value.

For this group of experiments, the stock indexes of Spain (IBEX), France (CAC) will be used, as these countries are Portugal's exports main customers and USA (NYSE) as it is UE's main commercial partner and represents one of the biggest economies in the world.

Exploratory Data Analysis

The following Graphics show the Portuguese stock market index and the other world indexes:

Figure 17 Charts comparing the values of PSI20 and IBEX35. On the left absolute values and on the right, scaled data.

Comparing with the IBEX35 a strong correlation is observed, except for the 2014 drop in the Portuguese stocks which, as previously explained in this paper, is due to the resolution of a main Portuguese bank. Other effects, such as the Covid-19 pandemic and the recovery from it and are especially visible in the scaled data graphic.

Figure 18 Charts comparing the values of PSI20 the NYSE. On the left absolute values and on the right, scaled data.

NYSE is considered be the main stock exchange in the world, so it is expected to have an influence on the PSI20 but, is there a time delay for that influence? The observation of the above Graphics of the last 10 years of data shows the movements of the two indexes reveal correlation. An interesting difference is how the NYSE has bounced back much more energetically form the pandemics in part due to the raise of the "retail traders" (Ozik, Sadka, & Shen, 2021).

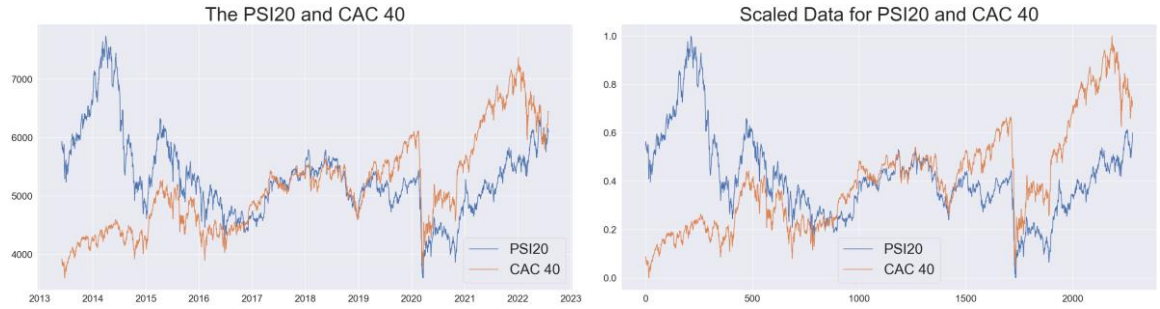


Figure 19 Charts comparing the values of PSI20 the CAC40. On the left absolute values and on the right, scaled data.

The same correlation between the PSI20 and the France Main Stock exchange index, the CAC40 is observed. Again we observe a faster bounce from the covid-19 lockdown and again the difficulty of the portuguese stock index to recover from the "BES" resolution.

Experiment #3A – Using other world indexes to predict PSI20.

Description

This first experiment tests if the knowledge of the last 90 values of another stock index can be used to train a model that can make an accurate prediction of the PSI20.

The training labels consist in the index values of other country's stock market, and the target output the Portuguese stock market main index value, both scaled between 0 and 1. Data processing, hyperparameters and model evaluation strategy can be consulted on the appendix.

Test Results

Test#	Years of data	Prediction RMSE	Day before RMSE	Training MSE (-04)	Validation MSE (-04)	End date	World index
D1	10	963.27	42.33	20	1015	01-02-2023	IBEX
D11	5	1717.15	75.43	52	118	01-18-2022	NYSE
D21	5	1651.22	74.23	37	262	01-08-2022	CAC

Table 13 Experiment #3A test results.

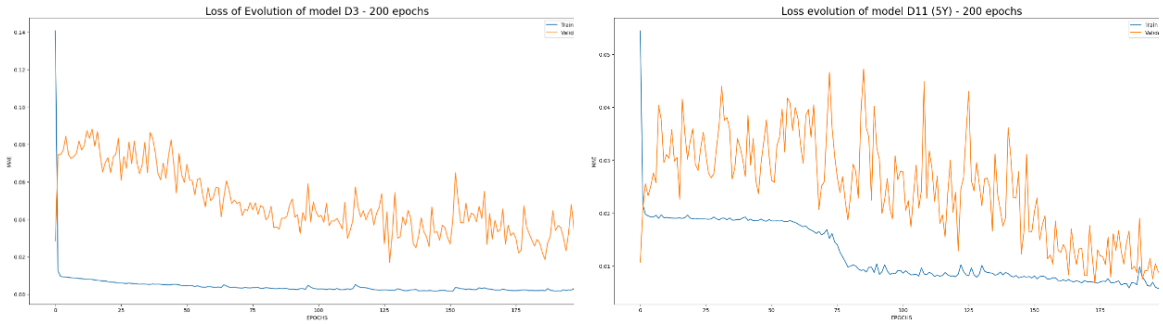


Table 14 History of learning progress of Tests D3 and D11

Observations

Observing the loss evolution of the model it can be observed that the model learns, as the loss tends to descend, however some strong signs of overfitting are visible, as the validation MSE is considerable higher than the training MSE, as observable in table 13.

Test #D1 provided a poor adjustment, with very high RMSE that could be in part explained by a divergence in the PSI20 and IBEX35 in the test period. On tests D2, D3 and D4 (see appendix) the data was truncated to end in a different date, but results show the same significative overfitting and higher RMSE than our benchmark, the "day before" RMSE. When considering CAC and NYSE, the results are similar with very high RMSE and predictions that are not accurate, as it can be observed on the below charts:

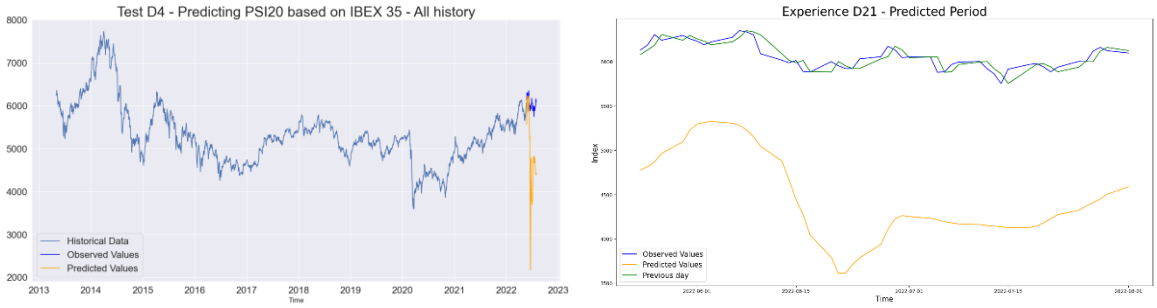


Table 15 Graphics of the predicted close values generated by models. On the left 'IBEX' trained models and on the right the 'CAC'

We conclude that no evidence was found that relates the last 90 days of data of a foreign stock index with the next value of the PSI20 value. On the next experiments, the foreign stock index will be used alongside with the PSI20 historical data in an effort to improve the accuracy.

Experiment #3B – Using PSI20 and another world index.

Description

This experiment combines the approaches of the experiment #1A and experiment #3A, as the PSI20 historical data will be used alongside another world stock exchange Index (NYSE, CAC30, IBEX35) as training labels and the PSI20 closing value of the following day is the target output. Two variations were tested, one with data scaled between 0 and 1 and a second with the log-returns (that capture the index variations), on both training labels and target output.

Test Results (with data scaled 0-1)

Test#	Years data	Epochs	Learn rate	Units per layer	Prediction RMSE	Previous Day RMSE	End date	Other index
E1	10	150	0.001	50;25	47.81	46.32	01-01-2023	IBEX35
E4	5	150	0.001	50;25	46.94	46.32	01-01-2023	CAC
E5	5	150	0.001	50;25	71.94	67.93	01-01-2023	NYSE

Table 16 Test Results of Experiment #3B using absolute values.

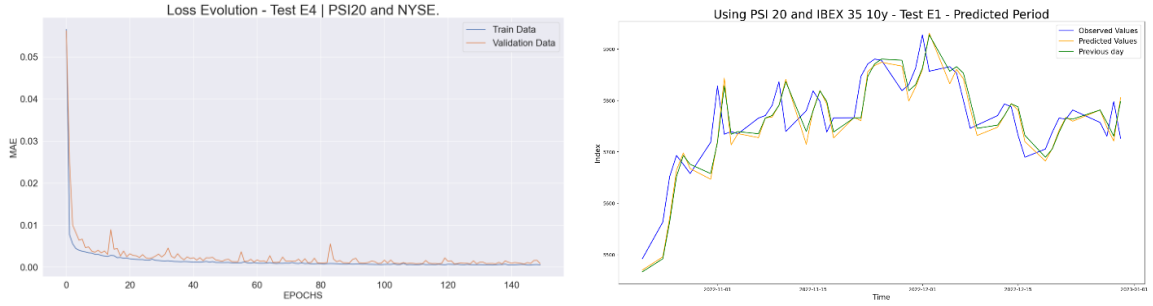


Figure 20 Loss Evolution of test E4, and model generated prediction from test E1 (PSI20 + IBEX).

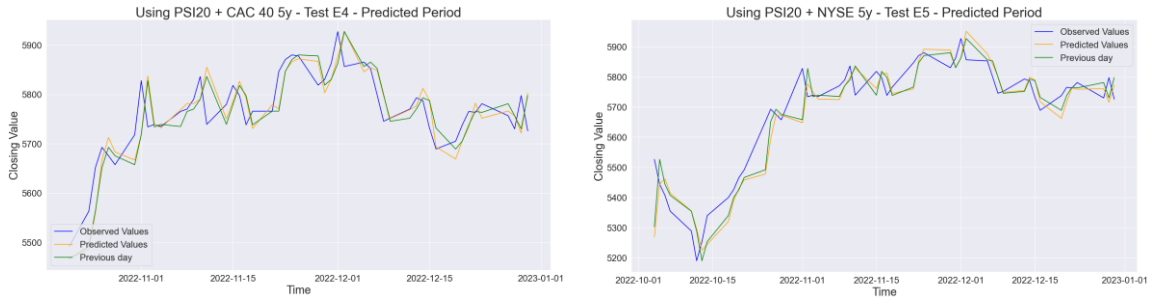


Figure 21 Model Generated Predictions for models E4 (CAC) and E5 (NYSE)

Test results (with log returns)

Test#	Years data	Epochs	Learn rate	Units per layer	Prediction RMSE	Previous day RMSE	End date	Other index
H1	5	150	0.001	50;25	45.42	46.32	01-01-2023	IBEX35
H2	5	150	0.001	50;25	50.60	49.91	01-01-2022	IBEX35
H4	5	150	0.001	50;25	61.26	67.93	01-01-2023	NYSE
H6	5	150	0.001	50;25	46.25	46.32	01-01-2023	CAC30

Table 17 Test Results of Experiment #3B using variations.

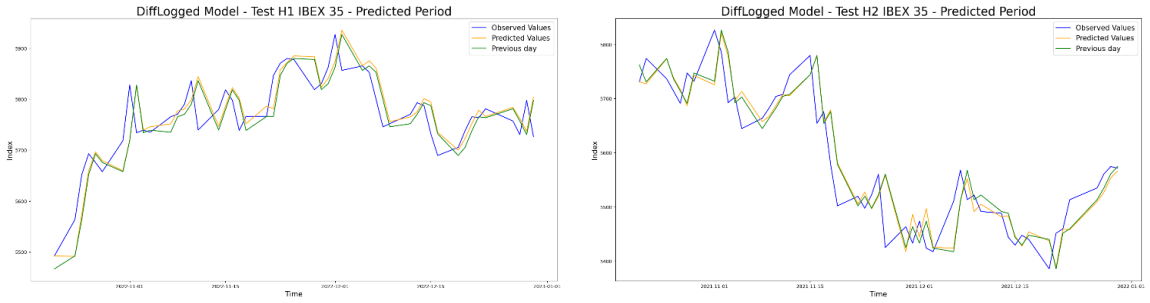


Figure 22 Model generated predictions of tests H1 and H2, using PSI + IBEX variation.

Observations

With this experiment it was aimed to test the EMH on its semi-strong effect, testing if the knowledge of the previous closing value of other world indexes closely connected to the PSI20, alongside the Portuguese index historical data, could improve the prediction for tomorrow's closing value.

Two strategies were used. One was to feed the model with absolute values (scaled between 0 and 1) and other with the variation regarding the previous day, captured by the log return. The results of the experiments with both strategies do not rule out that including another index together with the PSI20 has a positive impact on the quality of the regression, but do not confirm it either. The main trend observable is that the model still tries to adjust to the latest known value either by predicting it directly or by predicting a very small variation. For that reason, the predictions RMSE is always very close to the previous day RMSE, a recurrent pattern also observed in several of the previous experiments.

Experiment #3C – Multi Index approach

Description

This model is an expansion of the one trained on experiment #3B as it uses a multi-index approach using a sliding window of the last N closing values from PSI20, CAC40, IBEX35 and NYSE, to anticipate the closing value of the next PSI20 day. Again, two data scaling methods (DSM) are used: one with the scaled values and the second with the log returns for both training labels and target output.

As the variation relative to the last day can either be positive or negative, an accuracy test will be performed on the log return trained models, to check the accuracy rate of the model, and a value above 50% outperforms a pure "random" guess.

Test Results (with data scaled 0-1)

Test#	Years of data	Epochs	Prediction RMSE	Previous day RMSE	End date	Look_back	DSM
J1	5	150	71.33	57.62	01-02-2023	60	Scaled (0-1)
J2	5	150	89.17	76.91	01-06-2022	60	Scaled (0-1)

Figure 23 Test Results of Experiment #3C using absolute values.

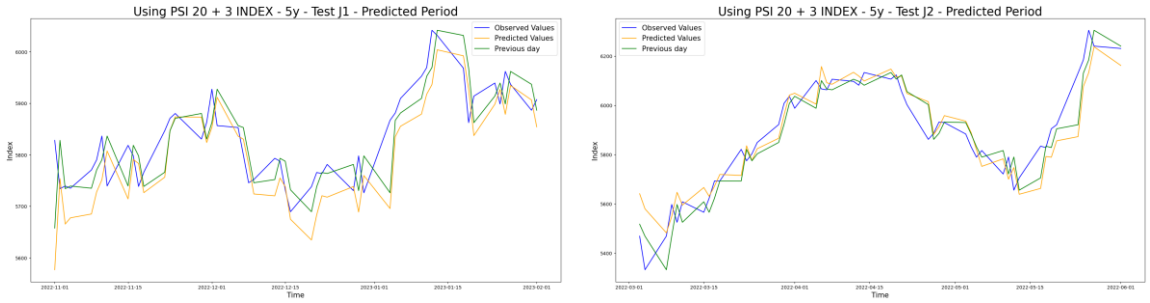


Figure 24 Model generated predictions of tests J1 and J2

Test Results (with log returns)

Test#	Years data	Epochs	Prediction RMSE	Previous Day RMSE	End date	Accuracy	DSM
K1	5	150	51.63	57.62	01-02-2023	57.14%	Log return
K2	5	150	67.71	65.39	01-03-2022	42.86%	Log return
K3	10	150	50.30	51.92	01-01-2022	53.06%	Log return
K4	10	150	81.44	76.99	01-06-2022	51.02%	Log return

Table 18 Test Results of Experiment #3C performed with variation-based data.

Observations

The tests performed with absolute values provided RMSEs above the "previous day". The models trained with the log returns show on test 'k1' promising results with a RMSE that is well under the benchmark RMSE and an accuracy above 57%.

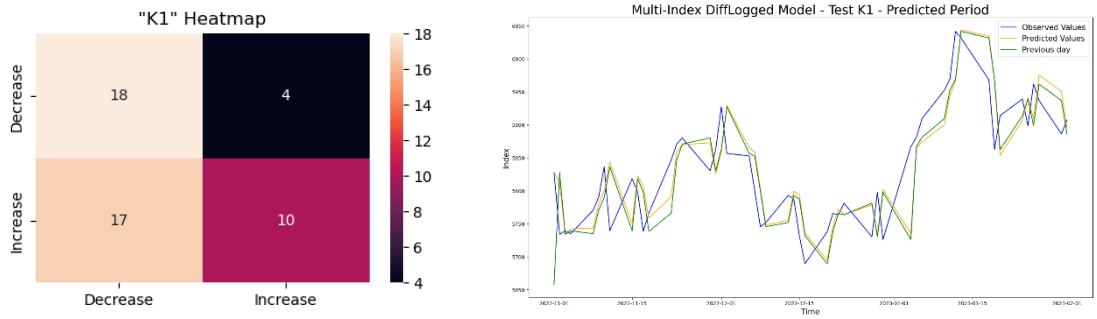


Figure 25 "Heat Map" and graphic generated by test K1 model

To confirm the results a second test ('k2') was performed for a different peroid, to check if the results obtained in 'k1' were consistent or just a particularity for that exact data set. The results on 'k2' did not confirm the results as the RMSE is again above the benchmark and the accuracy is below 50%.

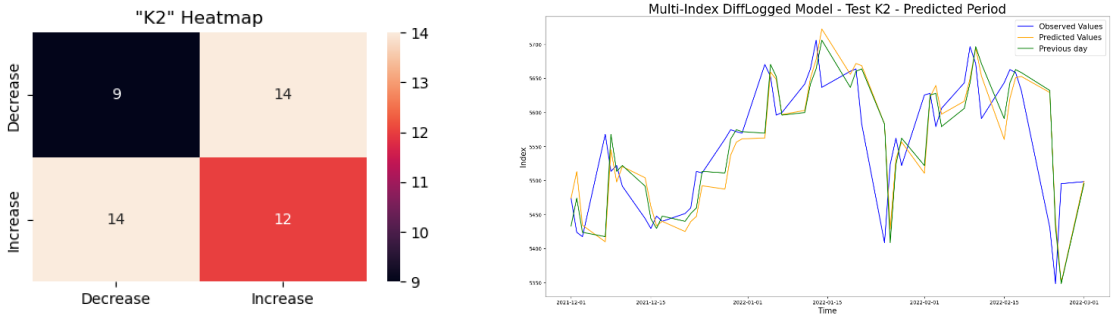


Figure 26 "Heat Map" and graphic generated by test K2 model

Finally, the tests performed with 10 years of data show RMSE, close to the previous day but the heatmaps reveal that the models make more conservative predictions, almost always predicting an increase on the value.

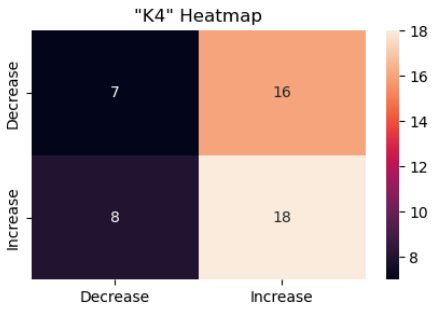


Figure 27 Test K4 with 10y of data heat map.

Conclusions on group of experiments #3

This group of experiments aimed to assess if the inclusion of three world stock exchanges (NYSE/CAC/IBEX), alongside with PSI20 historical data, could contribute to a better prediction than the one provided by the PSI20 historical data alone.

Models trained with this additional data that perform better than the model with historical data alone would hint that for some delay on the integration of the information regarding other stock indexes in the PSI20. That would be evidence that would challenge the application of the EMH in its semi-strong form (that holds that prices immediately reflect all available information), to the Portuguese stock market.

Three sets of tests were performed, the first using a foreign index as standalone training labels (X_Data) for the model, the second using PSI20 historical data alongside with another index and a third test using PSI20 historical data and the historical data of three additional indexes (CAC/IBEX/NYSE).

The standalone foreign index experiment did not perform well giving an indication of small predictive relevance for these foreign indexes. The tests from experiments #3B and #3C, performed significantly better. Some tests showed promising results namely when the models tried to predict if the PSI20 variation would be positive or negative. However, the global analysis does not provide consistent evidence for a better adjustment than the PSI20 historical data alone.

A plausible cause for the absence of improvement observed, when the other stock indexes are included, is the arbitrage effect (Fama E. F., 1998). On the following group of experiments intraday data will be analysed and used to train models in an attempt to reduce the impact of that effect.

FORECASTS WITH INTRADAY DATA

This group of Experiments is designed to test the momentum effect within a shorter time frame. Instead of considering the daily values, the models are trained with intraday data with a one-minute granularity.

The aim is to try to overcome the arbitrage effect that could be a relevant factor for the quality of predictions and access if an investor could have an abnormal positive return, by using a High Frequency Trading (HTF) strategy, based on models trained with PSI20 and other world indexes one-minute data.

The NYSE was replaced by the German stock index because its open/close hours are much closer to the PSI20. The relevance of using the German index is also consubstantiated by the fact that Germany is one of Portugal's main commercial partners along with Spain and France (Workman, 2022).

Exploratory Data Analysis

The available data for intraday analysis with one-minute delay is limited. We have access to the last 20 days of trading of intraday data from Yahoo API, fetchable in 5 days chunks. The source of the data is the Yahoo API, same as for daily data.

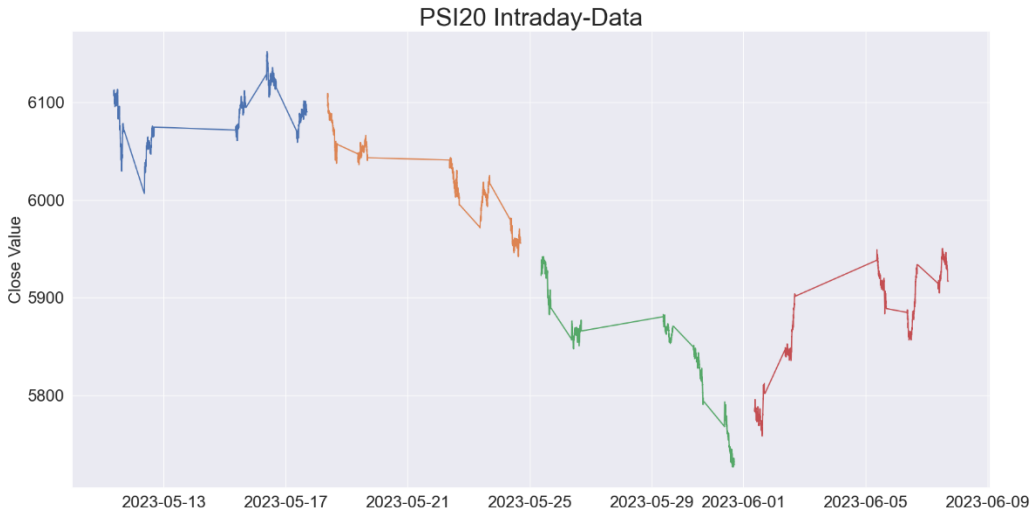


Figure 28 Graphic of the PSI20 index intraday (one-minute frequency) value.

There is a global downward trend in the global period, a maximum of 6151.50 in the first week and a minimum of 5726.81 in the end of week number 3. The data starts in 11-5-2023 at 9:00 and the values are available until 17:05 each day. There are no available data from the weekends. The data starts in 11-05-2023 and finishes on 18-06-2023.

PSI20 intraday	18-5-2023	25-05-2023	01-06-2023	08-06-2023	Global
count	2408.00	2405.00	2395.00	2402.00	9610.00
Mean	6087.01	6021.19	5845.95	5874.07	5957.24
std	26.20	41.40	56.45	54.81	110.48
min	6006.54	5942.35	5726.81	5758.63	5726.81
25%	6068.38	5995.73	5819.23	5843.07	5866.03
50%	6089.78	6021.46	5860.12	5886.68	5947.44
75%	6105.06	6052.54	5872.14	5919.16	6058.93
max	6151.50	6108.92	5941.95	5950.34	6151.50

Table 19 PSI20 intraday data from 11/5 to 8/6 main statistics

The other indexes are available with same format as the PSI20, and the main statistics are as follows:

	IBEX (Spain)	CAC (France)	DAX (Germany)
Count	9589.00	9583.00	9558.00
Mean	9223.83	7320.18	15979.75
Std	71.48	113.65	141.34
min	9047.30	7083.60	15646.93
25%	9167.30	7227.13	15883.91
50%	9220.60	7298.42	15947.94
75%	9272.00	7419.73	16048.75
Max	9401.80	7523.24	16330.78

Table 20 IBEX, CAC, and DAX intraday data from 11/5 to 8/6 main statistics

By observing the graphics there and stats we can conclude that there are no extreme outliers, and some patterns seem to be common to all stocks, for example all indexes have their minimum around June 1st and a local maximum around May 17th.

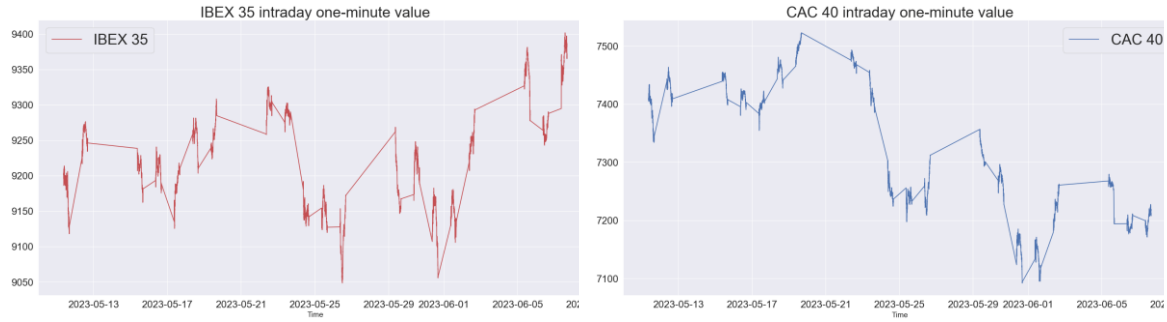


Figure 29 Graphic of IBEX and CAC in the 11/5 to 8/6 period (intraday)

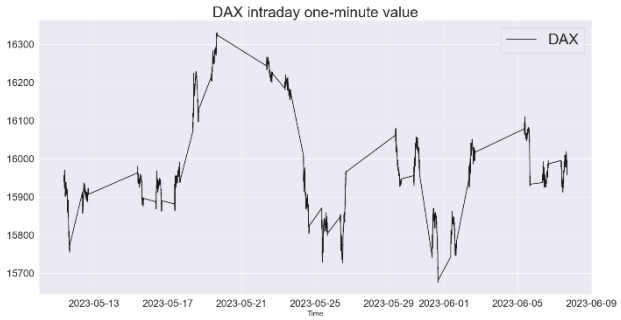


Figure 30 Graphic of the DAX in the 11/5 to 8/6 period (intraday)

The Exploratory Data analysis suggest some correlation evidence on the four stock indexes, therefore the research will proceed with models that include training data for all four indexes to predict PSI20 next minute value.

There is some discrepancy on the last hour of data on the data retrieved with the YAHOO API. Although the Euronext-Lisbon trades until 17:30 (CET), the data is only available until 17:05, and after 16:30, the values do not change, are always the same.

For data consistence it was decided to remove all observations after 16:30 and to drop all lines where one of the stocks did not have value available. The final count of observations after this modification is 8935.

Group of Experiments #4 – Intraday values with one minute frequency

This section describes the experiments performed using intraday data. On the first experiment the models are trained with intraday data from all 4 indexes (PSI20, CAC, IBEX and DAX) to predict the next minute PSI20 value.

The second experiment follows a different methodology: 30 models are trained with the same data and then compared. The models follow the same principles of the first experiment of this group using all 4 indexes as training labels.

On the third experiment the models will be trained only with PSI20 data, to compare with the results of the second experiment.

Experiment #4A – Training a model with intraday multi-index data.

Description

A multi-index approach will be used with a sliding window with the values of the last N minutes of the PSi20, CAC40, IBEX35 and DAX, to predict the PSI20 next minute value. Four tests will be performed for each of the last weeks and a one additional test after concatenation of the data of the 4 weeks.

Data Processing

The X_Data (training labels) is a multidimensional array with the scaled previous N minutes closing values of PSI20 and the other stock index (CAC40, IBEX35, CAX) for each timestep. The target output (Y_Data) is the value of the PSI20 index in the next minute. All values are scaled between 0 and 1, where 0 is the minimum value recorded for each index and 1, the maximum value. Below is an example of the data set before being scaled:

Date time	PSI20	CAC 40	IBEX 35	DAX
2023-05-24 09:00:00	5979.07	9183.79	7308.58	16019.35
2023-05-24 09:01:00	5974.82	9178.50	7303.06	16012.22
2023-05-24 09:02:00	5976.56	9181.29	7302.87	16009.40
2023-05-24 09:03:00	5976.12	9185.00	7292.02	16008.29

Table 21 Sample of the data set with intraday data from PSI20, IBEX, CAC, and DAX.

Hyperparameters and model evaluation

Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.001 learning rate for 150 epochs, and a validation data split of 20% and a batch size of 8.

As usual, the Root Mean Squared Error (RMSE) is the chosen metric, using the day before as a Benchmark.

A second evaluation method is proposed using accuracy to calculate the percentage of times the model gets the variation right. The variation is obtained by calculating the difference of the logarithms, and the accuracy will be measured in two respects:

1. A first evaluation is based on the signal of the variation, either positive or negative, meaning a value above 50% beats a random guess.
2. A second evaluation considers 3 possibilities, "decrease"/"stable"/" increase". The value is declared stable when the variation is between a threshold defined based on the maximum variation observed. In this case it is not only important to measure the accuracy but is also important to observe how the model fails, as predicting an "increase" and the observation returns a "decrease" is a worst result than if the observed outcome is "stable", especially if we are considering using the results for an algorithmic training simulation.

Test Results

Test#	Days of data	Epochs	Learn rate (-03)	Prediction RMSE	Previous Minute RMSE	End date	Accuracy	Accuracy (3 outputs)
L1	5	150	1	1.38	1.35	25-05-2023	57.14%	34.69%
L2	5	150	1	1.39	1.35	01-06-2023	53.06%	30.61%
L3	5	150	1	1.13	1.04	08-06-2023	57.14%	34.69%
L4	5	150	1	1.32	1.21	18-05-2023	51.02%	34.69%
L5	20	150	1	1.02	1.04	08-06-2023	51.02%	26.53%

Table 22 Test results of experiment 4A - Intraday Data

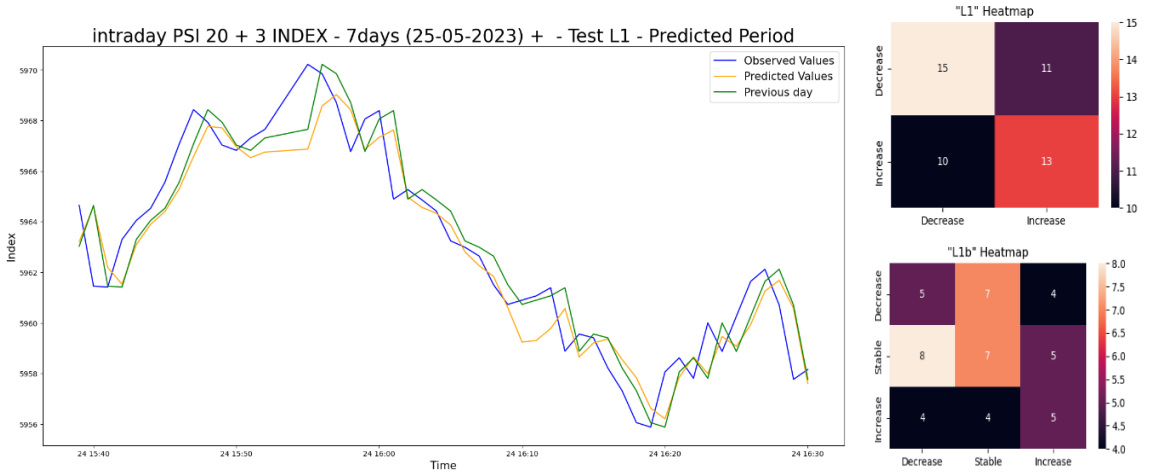


Figure 31 Test L1 Model generated prediction and heatmaps with 2 and 3 possible outcomes.

Observations

The accuracy results observed are promising, with accuracies above a random scenario in most tests. To check the consistence of this results an extensive test with different methodology will be presented in next experiment.

Experiment #4B – Training 30 models wilt multi-index intraday data.

Description

Following the encouraging results on the accuracy metric it was decided to try a different methodology in this thesis based on the one suggested by Yang et al (Yang, Wang, & Li, 2022), where 30 models are trained with the same data and the results of each model are compared.

Two sets of models are trained, one with the data from the first week of observations and a second with the data of the second week.

Two predictions will be made by each set of models. On the first, the model uses the test data of the corresponding week. On the second, the test data of the other week is used, to verify if the models are capable of producing a good adjustment with different, non-consecutive data.

Model evaluation

The 30 models will be compared and the results of the model with the BEST metrics are presented. The average of the 30 models is also considered and displayed on the below Table.

Test Results

Test#	Days	Best RMSE	Previous minute RMSE	Training week	Test Week	Best Accuracy	Best accuracy 3 outputs	Accuracy average	Accuracy average 3 outputs
M1	5	1.21	1.21	18-05	1	61.22%	40.82%	56.3%	34.9%
M2	5	1.33	1.35	25-05	2	59.18%	38.78%	57.4%	35.4%
M3	5	1.39	1.33	18-05	2	65.31%	40.82%	60.3%	35.2%
M4	5	1.20	1.21	25-05	1	57.14%	36.76%	54.5%	33.27%

Table 23 Test results of experiment #4B.

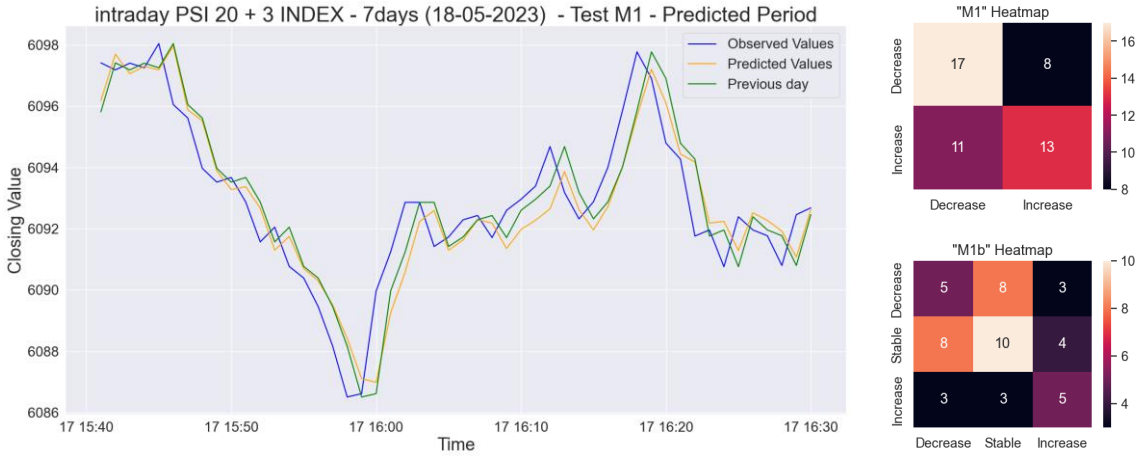


Figure 32 Predicted value and accuracy Heat Maps for the BEST model generated in test M1.

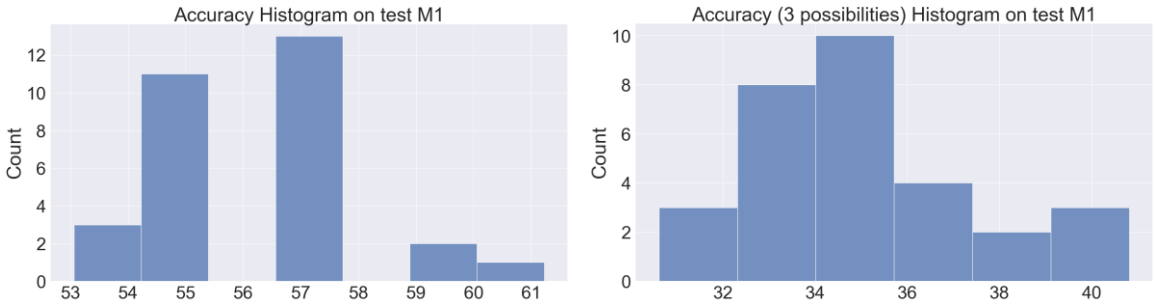


Figure 33 Histogram with the accuracies (2 and 3 outputs) recorded by the 30 trained models in test M1.

Observations

The observation of the results on the above Figure (and the ones on the appendix) confirms what was hinted in previous experiment. The accuracy of the models is surprisingly high. Both models with different test sets have average accuracies over 54% when given two possibilities (positive or negative variation). The histograms show that in all four tests the worst of the 30 models never provides an accuracy below 50%, which gives good indications for the reliability of the models.

When given three possibilities, the average accuracy is above the 33,3% guest in all tests except M4. Focusing on the heat maps, we observe that the models usually fail "graciously", meaning that when the model predicts an increase, and it fails, in most cases the observed outcome will be "stable", which can proof crucial for a trading algorithm based on this model.

Experiment #4C – Training 30 models wilt PSI20 intraday data.

Description

Following the results presented by intraday models' accuracy, it is important to verify if those results are obtained due to the combination of the PSI20 historical data with the other stock indexes or if these results could be obtained if only the PSI20 data was used.

Test Results

Test#	Days Data	Best RMSE	Previous Minute RMSE	Training Week	Test Week	Best Accuracy	Best Accuracy 3 output	Accuracy Average	Accuracy Average 3 output
N1	5	1.16	1.21	18-05	1	59.18%	38.78%	54.08%	37.6%
N3	5	1.33	1.35	18-05	2	59.18%	34.69%	60.48%	31.6%

Table 24 Test results of experiment #4C.

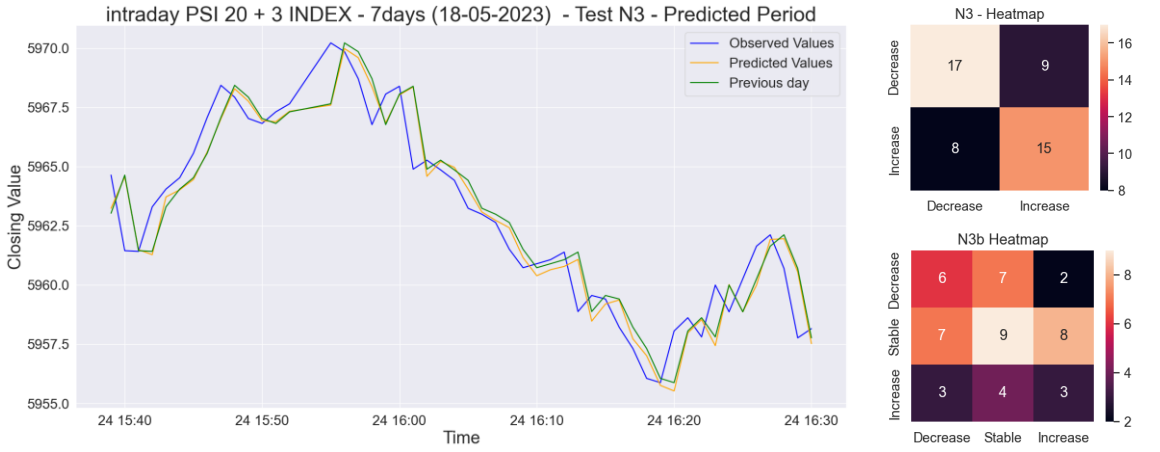


Figure 34 Predicted value and accuracy Heat Maps for the BEST models generated in test N3.

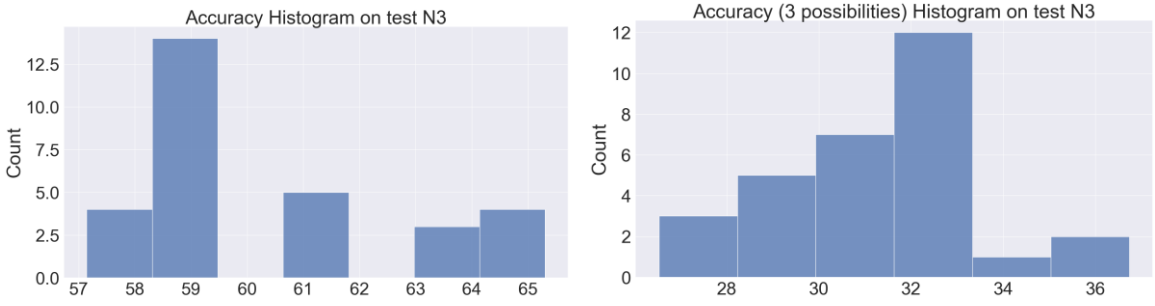


Figure 35 Histogram with the accuracies (2 and 3 outputs) recorded by the 30 trained models in test N3 (with test data from week 2).

Observations

The intraday models train with only the PSI20 one minute frequency data have a very good performance in terms of RMSE with excellent adjustment to the previous minute.

On the accuracy ratings, regarding the two possibilities scenario, the models perform as well as the models in experiment 4B. However, in the 3 possibilities scenario these models loose some consistency as when we the model made prediction with test data from the week 2, the average is below 33% and the histogram show many models giving a below 33% accuracy.

Conclusions on group of experiments #4

This group of Experiments was designed to find evidence of a *momentum effect* using a shorter time frame of intraday data with a 1-minute granularity.

The aim was to overcome arbitrage effects and train models that could assist an investor using HTF to have an abnormal positive return.

On the first experience the models trained with PSI20 and other three stock indexes revealed accuracy results that encouraged further investigation, while showing the usual adjustment of the predictions to the previous known value.

Experiment #4B deepened the analysis using a different strategy. 30 models were trained with similar data and their predictions compared. The results were promising as we observed a consistent above 50% accuracy when model had to predict if the value should increase or decrease and consistent above 33.33% accuracy, when the models was predicting one of three outcomes ("decrease"/" stable"/" increase"). Moreover, when the model failed it predominantly failed to the closest value, giving good indication that this model can perform well in a trading algorithm.

Finally in experiment #4B we tried to isolate the effects responsible for the accuracy scores, training the model with only the PSI20 Data. The results are not conclusive but suggest that including the data of the other three indexes (IBEX/CAC/DAX) provides greater consistency in the accuracy ratings.

CONCLUSION

This chapter will present the dissertation main conclusions and point out future lines of research.

Overview of the research

This thesis has addressed the research gap in the application of deep learning techniques for financial forecasting of the PSI20, the main Portuguese stock index. By utilizing LSTM models, this study has contributed to the growing body of knowledge in this area and examined the applicability of the EMH in its weak and semi-strong forms.

The research started by trying to determine whether historical information holds predictive power for the PSI20 and tested the possibility of momentum effect being an important element in the formation of prices. The **group of experiments #1** relayed on the PSI20 historical data alone to train models that attempted to predict the PSI20 next daily closing value. These models aimed to benefit from the documented "momentum effect" (the tendency of stocks that have performed well in the past to continue their upward trend, and of stocks that have performed poorly tend to continue their downward trend) of stock prices to try to elaborate a relevant prediction.

The results of this first group of experiments confirmed the EMH prepositions as results show that models adjusted their predictions to the previous day value, as it is the point where loss is minimized. When the focus shifted to the index variation using the log return, the predicted variation was always close to 0 (zero). Expanding the time horizon of the prediction to 10 days, produced a horizontal line of values, based on the last known value. These results are strong evidence that the index values follow a Markov process and the best estimator for tomorrow's stock value is today's stock value, and that the knowledge of previous days of trading has no relevance in making a prediction.

On the **group of experiments #2** this thesis examined the impact of temporal patterns on the accuracy of predictions by incorporating weekdays alongside the historical data of the PSI20, and explored the significance of incorporating trading volume, as a technical indicator, alongside the PSI20 data.

By considering the influence of weekdays, the study aimed to assess whether certain days of the week exhibit consistent patterns, such as the "Monday effect" that could be leveraged for improved forecasting. However, the results revealed no significant influence of weekdays on the predictions, indicating that the predictive power of the LSTM models remained unchanged after including the weekday.

By integrating trading volume as a feature, the study aimed to determine whether it contributes to the quality of predictions. Although the adjustment of the models using trading volume yielded similar results to models without it in terms of the metric used (the RMSE), the behaviour of the predictions observed in the graphics, suggested that further research is necessary to draw definitive conclusions.

On the **group of experiments #3** this research has adopted a cross-market approach, incorporating other world indexes to enhance the prediction of PSI20 values. This approach aimed to investigate potential lagged correlations and testing the applicability of the semi-strong form of EMH to the Portuguese stock index.

Despite the incorporation of other prominent stock indexes, namely the IBEX, CAC, and DAX, to assist in predicting the values of the PSI20, the quality of the predictions did not show significant improvement when compared to models that solely relied on PSI20 historical data. The absence of substantial improvement suggests that the inclusion of additional indexes may not provide substantial benefits due to the influence of arbitrage. This observation rose the question of whether training the models with intra-day data, where the effect of arbitrage may be less pronounced, would yield more promising results and potentially offer a solution to address the limitations observed on this group of experiments.

In the **group of experiments #4**, this study captured finer-grained market dynamics as it looked to benefit from intraday anomalies to generate LSTM with improved predictive accuracy.

After a preliminary test that revealed promising accuracy rates when predicting the direction of the variation, a different methodology was used as thirty models were trained using the same intra-day data, and then compared. The results showed interesting levels of accuracy, surpassing random guesses, and indicating the potential of these models in capturing short-term market dynamics. The models provided accuracies consistently above 55% when the models had two possibilities and above 33% when having three possibilities.

This pattern suggests the presence of temporal anomalies and hints that these models could be used to design algorithmic trading simulations that could allow an investor to have an abnormal positive return by using a high frequency trading strategy.

Future work

The possibilities for utilizing artificial intelligence (AI) in stock market analysis are vast, and choices had to be made in this thesis to narrow down the scope. The selected focus of this research was on predictions, which has garnered significant attention from market researchers and traders alike. However, it is important to acknowledge that prediction is also one of the most challenging aspects of stock market analysis.

In the future work, two sections can be delineated: enhancements to the research conducted in this thesis and exploring other scopes of analysis.

Enhancements to the research

Further testing on the influence of volume and other technical indicators: The experiments using volume as a training label displayed an interesting behaviour of the model's predictions, so the use of volume, and other technical indicatorsm deserves more attention in future research.

Designing a trading simulation algorithm: To further validate the practicality and effectiveness of the forecasting models developed in this thesis, a trading simulation algorithm should be designed. This algorithm would assess whether it is possible to achieve abnormally positive returns by utilizing a strategy based on the forecasting models, surpassing the performance of a baseline buy-and-hold strategy. This simulation would account for transaction costs, market liquidity, and other relevant factors, providing a more comprehensive evaluation of the models' real-world performance.

Incorporating sentiment analysis: The inclusion of sentiment analysis as an additional feature in the predictive models would offer a deeper understanding of market dynamics. By integrating sentiment analysis techniques, such as natural language processing and sentiment scoring algorithms, the models can factor in the influence of news, social media sentiment, and other

qualitative factors on stock prices. This enhancement would enable a more comprehensive analysis and potentially improve the accuracy of predictions.

Further exploring log-returns analysis: Repeating some of the analyses using log-returns instead of absolute values would provide valuable insights. By focusing solely on variations rather than absolute values, log-returns eliminate the influence of the initial value and allow for a more accurate comparison of results. This approach can further our understanding of the models' performance, particularly in terms of capturing relative price movements and volatility.

Using explainable AI (xAI): Incorporating xAI techniques into the models can be ideal for decision-makers who require transparency and interpretability. By adopting methods that provide simple explanations into the reasons behind the predictions, decision-makers can better understand and trust the models' outputs. However, the LSTM models are naturally complex, so the challenge of applying XAI is huge, as "*everything should be made as simple as possible, but not simpler"* (Albert Einstein).

Exploring Other Scopes of Analysis

Beyond forecasting, there are other scopes of analysis that warrant exploration. Applying other machine learning methods to time series data can yield valuable insights. Techniques such as time series classification, regression, and clustering can be employed both within each of the studied indexes and across multivariate time series data. This extension would enable a more comprehensive analysis, to better grasp patterns that transcend a single market and providing a deeper understanding of complex relationships.

Final notes

In the realm of stock market analysis, the pursuit of accurate predictions can often resemble the alchemist's pursuit to transform base metals into gold. Researchers embark on a quest, starting with freely available raw data and attempting to transform it into the golden outcome of relevant and accurate predictions that outperform random guesses and challenge the EMH. However, achieving meaningful results proves to be an arduous task due to the unpredictable nature of stock markets and the balancing effects of arbitrage that often neutralize potential anomalies.

Nevertheless, like alchemy, the value of the process sometimes surpasses the final results. Although the alchemists are not reported to have obtained success converting iron into gold, their experiments and relentless pursuit of knowledge contributed to the development of laboratory techniques, apparatus, and chemical processes that laid foundations for the emergence of early modern science.

Conversely, in the pursuit of forecasting stock values, researchers acquire a broader knowledge of the intricacies of the stock market. This knowledge, gained through the exploration of various models and methodologies, proves invaluable not only to the researchers themselves but also to the global community. The process of attempting to forecast stock market behaviour leads to deeper insights, a better understanding of market dynamics, and the generation of new ideas and perspectives. Ultimately, this collective knowledge contributes to the advancement of financial research and decision-making in the face of the complex and ever-changing stock market landscape.

References

Antunes, S. (31 de 12 de 2014). Obtido de Jornal de Negócios: https://www.jornaldenegocios.pt/mercados/bolsa/detalhe/bolsa_nacional_perde_mais_de_um_quarto_do_seu_valor_em_2014

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM.

Buffett, W. E. (1984). The Superinvestors of Graham-and-Doddsville. *Hermes, Columbia Business School magazine.*

Campbell, J. Y., & Shiller, R. J. (1998). Stock Prices, Earnings and Expected Dividend",. *Journal of Finance*, pp. Vol. 43, pp. 661-76.

Chang, M. (2018). How A.I. Traders Will Dominate Hedge Fund Industry.

Consoli, Negri, Tebbifakhr, Tosetti, & Turchi. (2021). Forecasting the IBEX-35 Stock Index Using Deep Learning and News Emotions. *, et al. Machine Learning, Optimization, and Data Science. LOD 2021. Lecture Notes in Computer Science(), vol 13163.* . Springer, Cham.

DATA BASE CAMP. (4th de June de 2022). *Long Short-Term Memory Networks (LSTM)- simply explained!* Obtido de DATA BASE CAMP: https://databasecamp.de/en/ml/lstms

Fama, E. (May de 1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, Vol. 25, No. 2,*, pp. 383-417 (35 pages).

Fama, E. F. (January de 1965). The behavior of stock-market prices. *The Journal of Business, Vol. 38, No. 1*, pp. 34-105.

Fama, E. F. (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *Journal of Financial Economics, Volume 49, Issue 3*, pp. 283 - 306.

Fama, E. F., & Blume, M. E. (1966). Filter rules and stock market trading profits. *Journal of Business 39, Issue 1 , Part 2*, pp. 226-241.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, Volume 33, Issue 1,*, pp. 3-56.

Fernandes, A., Mota, P. R., Alves, C. F., & Rocha, M. D. (2014). Mercados, Produtos e Valorimetria de Ativos Financeiros.

Ferreira, F. G., Gandomi, A. H., & Cardoso, R. T. (January de 2021). Artificial Intelligence Applied to Stock Market Trading: A Review. *IEEE Access, vol. 9*, pp. 30898-30917. doi:10.1109/ACCESS.2021.3058133.

Fluck, Z., Malkiel, B., & Quandt, R. (1997). The Predictability of Stock Returns: A CrossSectional Simulation. *Review of Economics and Statistics*, pp. 176-183.

Frankfurter, G. M., & McGoun, E. G. (2001). Anomalies in finance: What are they and what are they good for? *International Review of Financial Analysis, Volume 10, Issue 4*, pp. 407-429.

French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics, Volume 8, Issue 1*, pp. 55,69.

Gomes, L. M., Soares, V. J., Gama, S. M., & Matos, J. A. (2022). Efficiency Drifts in Euronext Stock Indexes Returns. *International Journal of Business , Vol 27, issue 2*, pp. pl-17.

Gomes, L., Soares, V., Gama, S., & Matos, J. (2018). Estimating and testing long memory in Portuguese stock market returns. *Proceedings of the 32nd International Business Information Management Association Conference, IBIMA 2018 - Vision 2020: Sustainable Economic Development and Application of Innovation Management from Regional expansion to Global Growth* (pp. 2846-2857). IBIMA: Scopus.

Goodman, G. (1968). *The Money Game - Chapter 11, What The Hell Is A Random Walk?* Michael Joseph.

Hatcher, W. G., & Yu, W. (2018). A survey of deep learning: Platforms, applications and emerging research trends. *IEEE Access, vol. 6*, pp. 24411-24432.

Haugen, R. A., & Jorion, P. (Jan-Feb de 1996). The January Effect: Still There after All These Years. *Financial Analysts Journal, Vol. 52, No. 1*, pp. 27-31.

Idrees, S. M., Alam, M. A., & Agarwal, P. (2019). A Prediction Approach for Stock Market Volatility Based on Time Series Data". *IEEE Access, vol. 7*, pp. 17287-17298.

Jegadeesh, N., & Titman, S. (March de 1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, Vol. 48, No. 1*, pp. 65-91.

Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications, Volume 184*.

Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics, Vol 12, issue 1*, pp. 13-32.

Kijima, M. (2002). *Stochastic Processes with Applications to Finance.* Chapman & Hall/CRC.

Liu, G., & Wang, X. (2019). A numerical-based attention method for stock market. *IEEE Access, vol. 7* , pp. 7357-7367.

Lo, A. W., & MacKinlay. (1999). *A Non-Random Walk Down Wall Street.* Princeton University Press.

Lo, A. W., & MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *The Review of Financial Studies, 3(2)*, pp. 175-205.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives, 17(1)*, pp. 59 - 82.

Markowitz, H. (1952). Portfolio selection. *J. Finance*, pp. vol. 7, no. 1.

Matos, J. A., Gama, S. M., Ruskin, H. J., & Duarte, J. A. (2004). An econophysics approach to the Portuguese Stock Index—PSI-20,. *Physica A: Statistical Mechanics and its Applications, Volume 342, Issues 3–4,*, pp. 665-676.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (Jul de 2020). Deep learning for stock market prediction. *Entropy, vol. 22, no. 8*, p. 840.

Naseer, M., & Tariq, Y. B. (December de 2015). The Efficient Market Hypothesis: A Critical Review of the Literature. *The IUP Journal of Financial Risk Management, Vol. XII, No. 4*, pp. 48-63.

Ozik, G., Sadka, R., & Shen, S. (13 de 5 de 2021). Flattening the Illiquidity Curve: Retail Trading During the COVID-19 Lockdown. *Journal of Financial and Quantitative Analysis, forthcoming*.

Raschka, S., Liu, Y., Mirjalili, V., & Dzhulgakov, D. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python.* Birmingham: Packt Publishing Ltd.

Salem, F. M. (2022). *Recurrent Neural Networks: From Simple to Gated Architectures by* (1st Edition ed.). Springer.

Shleifer, A., & Vishny, R. W. (March de 1997). The Limits of Arbitrage. *THE JOURNAL OF FINANCE * VOL. LII, NO. 1*, pp. 35-55.

Simplilearn. (21st de February de 2023). *The Best Introduction to Deep Learning - A Step by Step Guide*. Obtido de Simplilearn: https://www.simplilearn.com/tutorials/deep-learning-tutorial/introduction-to-deep-learning

Sobreira, N., & Louro, R. (2020). Evaluation of volatility models for forecasting Value-at-Risk and Expected Shortfall in the Portuguese stock market. *Finance Research Letters, Vol 32*.

Vignesh, C. K. (April de 2020). Applying machine learning models in stock market prediction. *EPRA International Journal of Research and Development (IJRD) - Vol. 5, no. 4,* , pp. 395–398.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining - Practical Machine Learning Tools and Techniques* ((4th Edition) ed.). Morgan Kaufmann.

Workman, D. (2022). *Portugal's Top Trading Partners*. Obtido de World's Top Exports: https://www.worldstopexports.com/portugals-top-import-partners/

Yadav, A., Jhaa, C. K., & Sharanb, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science, Volume 167*, pp. 2091-2100.

Yang, J., Wang, Y., & Li, X. (16 de November de 2022). Prediction of stock price direction using the LASSO-LSTM model combines technical indicators and financial sentiment analysis. *PeerJ Computer Science 8:e1148*. Obtido de https://doi.org/10.7717/peerj-cs.1148

ERRORS

- On the forecasts with intraday data, where the graphics legends show “previous day” it should show “previous minute” as the data is intraday with one-minute frequency. Pages 47, 48 and 49 on figures 32, 33 and 35.

- In the same Graphics where it reads 7 days, it should read 5 days, as the markets only work from Monday to Friday.

APPENDIX

In this section additional information regarding the forecasts is displayed, including complete test results, information on data processing, additional charts and model definitions.

Code repository

In the pursuit of fostering transparency and scientific rigor, the code utilized in this dissertation can be found at the following GitHub address: **https://github.com/zohanny/LSTM-Euronext**. By making the code openly available to the research community we enable the replication and validation of our findings, allowing for an assessment by peers. Moreover, the accessibility of this code serves as a resource for future research endeavours, encouraging the advancement and refinement of methodologies in the field while contributing to a culture of collaboration within the scientific community.

Forecasts with Daily Data

Group of experiments #1

Experiment #1A

Complete Test results

Test#	Years Data	Epochs	Learn Rate (10^{-4})	Nlayers	Units per layer	Prediction RMSE	Training MSE (10^{-4})	Validation MSE(10^{-4})
A1	5	200	7	1	50;	48,26	6.8	n/a
A2	5	200	5	2	50;25	**42,27**	4.4	n/a
A3	5	200	5	3	50;30;10	57,47	4.2	n/a
A4	5	150	3	2	50;30	49,93	7.1	n/a
A5	10	200	5	2	50;30	44,16	2.4	2.5
A6	10	200	5	1	100;	52,57	2.2	3.0
A7	10	200	5	1	150;	44,90	2.2	2.5
A8	10	200	5	1	200;	44,33	2.2	2.5
A9	10	200	5	2	150;75	43,87	2.2	2.5
A10	10	200	5	2	200;100	71,62	2.5	4.3
A11	10	200	5	3	100;50;25	48,75	2.4	2.7
A12	10	100	5	2	50;25	43,61	2.3	2.5

A13	5	200	5	2	50;25	43,45	4.4	5.9
A14	10	100	5	2	50;25	43,94	2.3	2.6
A15*	5	150	5	2	50;25	43.01	4.2	7.2

App_Table 1 Experiment #1A test results

Experiment #1B

Hyperparameters

Other parameters are the same as in previous experiments: Epochs = 200, an Adam optimizer with a 0,0005-learning rate and a two LSTM layers one with 50 neurons other with 25 neurons.

Test Results

Although some overfitting is observed on experiment, the learning curve shows the model is learning and reducing the loss as the epochs go by, however when the last day of data is moved to 1/2/2022, the overfitting shows up, as the training data loss reduces but the validation data loss raises. Below figure shows the *loss* on the Y axis and the 'Epochs' on the X axis. In orange is the loss of the Validation Data and in Blue the loss of the Training Data.

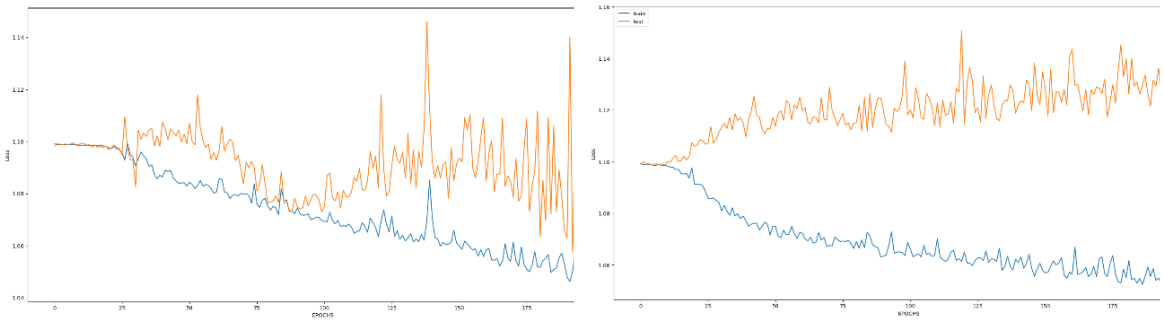


Figure 36 Loss Evolution in models C1 and C3

Experiment #1C

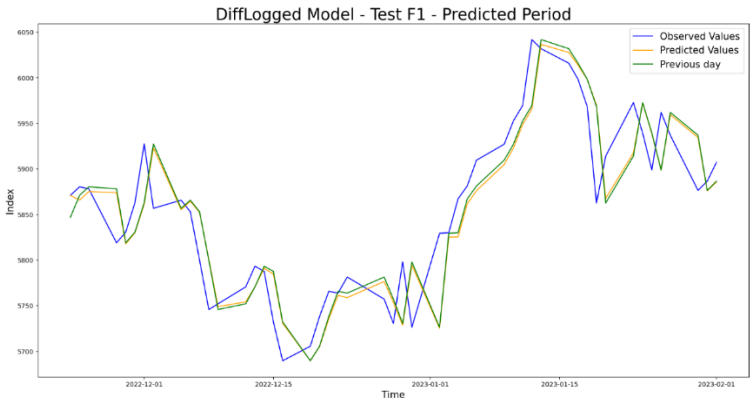
Data Processing

The variation in respect to the previous day was obtained using the log return. These values were processed in an X_Data in a sliding window of N days corresponding to one target Output.

The target output (Y_Data) is also the log return for the predicted day. After the prediction is made, the reverse operation is performed with the *cumulative sum*, generating predictions to get the absolute value of PSI20 next day's value.

Hyperparameters and model evaluation

Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.0001 learning rate for 150 epochs, and a validation data split of 20%. A shorter sliding window (Look_Back) was used as its size seems to have small impact on the predictions. The RMSE is the chosen metric, using the previous day as a Benchmark.



App_Figure 1 Model F1 generated predictions of the PSI20 closing value compared with the observed values and the previous day.

Experiment #1D

Data Processing

The variation in respect to the previous day was obtained using the log returns. These values were processed in an X_DATA in a sliding window of N days corresponding to one target Output.

The target output is an array of the difference of logarithms between the target day and its previous day. The size of this array is the number of days (time horizon) that the model will predict.

Hyperparameters and model evaluation

Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.0001 learning rate for 150 epochs, and a validation data split of 20%. A longer sliding window (Look_Back) was used to frame an also longer horizon of prediction.

Group of experiments #2

Experiment #2A

Data processing

Same data processing technics of previous experiments was applied to the PSI20 closing value data, but now the X_Data has the 90 days of the PSI-20 closing values and the previous 90 days of the PSI-20 trading volume (scaled between 0 and 1). The target data (Y_Data) remains the same, the scaled next PSI20 closing value.

Hyperparameters

Similar hyperparameters were used in this experiment, with two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.0005 learning rate. The model summary in on the appendix on app_table 1.

Model evaluation

Same strategy for evaluating the model as the univariate model. The Root Mean Squared Error (RMSE) is observed and compared with the RMSE of the "day before" that, as exposed before is 42.33.

Model Definiton

Layer (type)	Output Shape	Param #
Lstm _2 (LSTM)	50	10400
Lstm_3 (LSTM)	25	7600
Dense_2 (Dense)	1	26
Total params: 18,026 Trainable params: 18,026 Non-trainable params: 0		

App_Table 2 Experiment #2A Model Summary

Experiment #2B

Data Processing

The X_Data is a multidimensional array with the scaled previous 60 day's closing values of PSI20 for each timestep, and information of the day of the week. The Y_Data is the scaled closing value for each day. The weekdays are included in the X_Data in four different ways:

1. As a Time, Series of a value between 0 and 4.
2. Only the value of the last day.
3. A "Control" test with all weekdays set to "0".

4. As "One hot encoded" (OHE) time series with a binary feature for each day of the week.

Hyperparameters and model evaluation

Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.001 learning rate for 150 epochs, and a validation data split of 20% and a batch size of 8.

As usual, the Root Mean Squared Error (RMSE) is the chosen metric, using the day before as a Benchmark.

Group of experiments #3

Experience #3A

Data Processing

Same methodology of Experience #1A, except that the X_Data will derive from other stock markets other than the Portuguese. Also, it is necessary to remove the days where only one of the stock markets operated such as national holidays. The PSI20 data in this experiment will not be included in the training labels.

Hyperparameters

Two LSTM layers, one of 50 neurons and other of 25 neurons, and an ADAM predictor with a 0.0005 learning rate for 200 epochs. Model description is on the appendix.

Model Evaluation

The Root Mean Squared Error (RMSE), using the day before as a Benchmark.

Model Definition

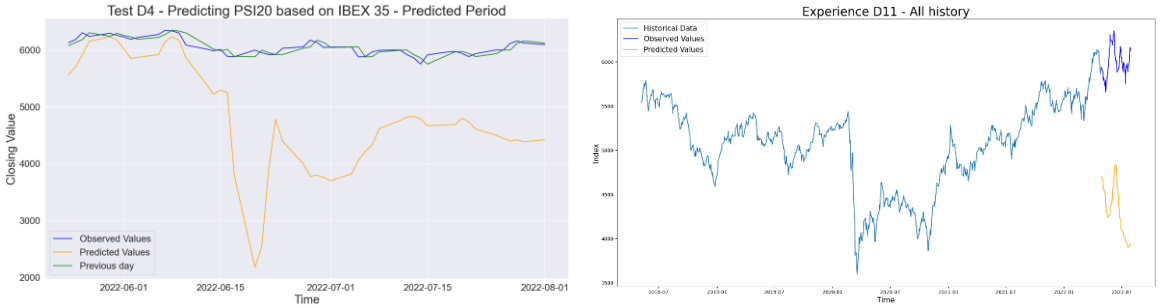
Layer (type)	Output Shape	Param #
Lstm _8 (LSTM)	50	10400
Lstm_9 (LSTM)	25	7600
Dense_4 (Dense)	1	26
Total params: 18,026 Trainable params: 18,026 Non-trainable params: 0		

App_Table 3 Experiment #3A Model Summary

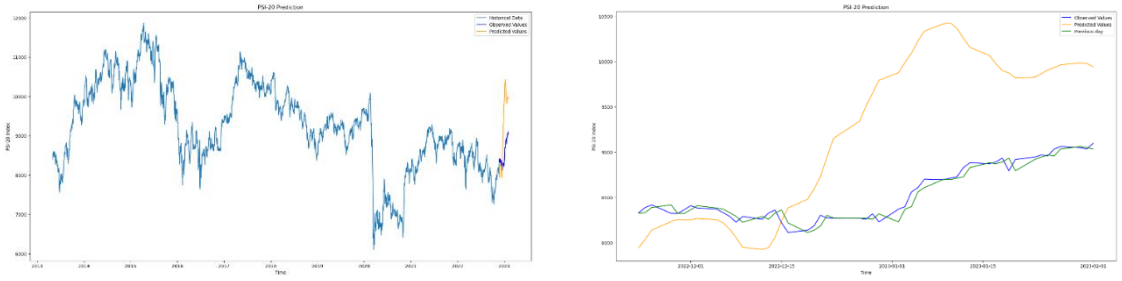
Complete Test Results

Test#	Years of data	Prediction RMSE	Day before RMSE	Training MSE (-04)	Validation MSE (-04)	End date	World index
D1	10	963.27	42.33	20	1015	01-02-2023	IBEX
D2	10	968.82	63.29	13	253	01-05-2022	IBEX
D3	5	707.80	63.29	17	220	01-05-2022	IBEX
D4	10	1261.65	75.02	21	436	01-08-2022	IBEX
D11	5	1717.15	75.43	52	118	01-18-2022	NYSE
D21	5	1651.22	74.23	37	262	01-08-2022	CAC

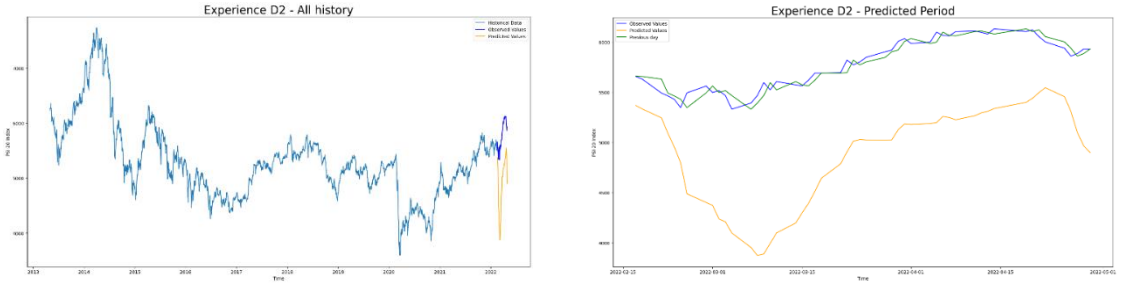
App_Table 4 Experiment #3A complete test results



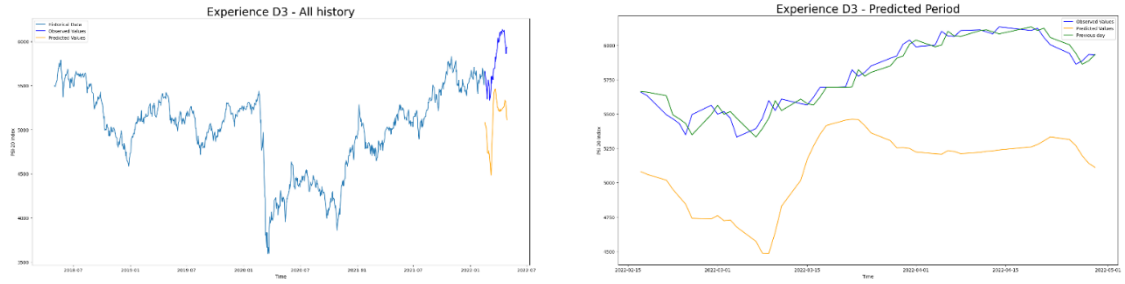
App_Figure 2 D4 and D11 Historical data and model generated predictions



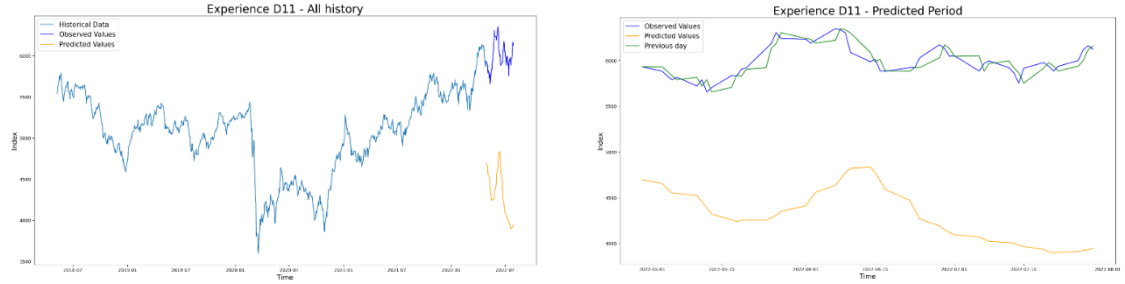
App_Figure 3 D1 Historical data and model generated predictions



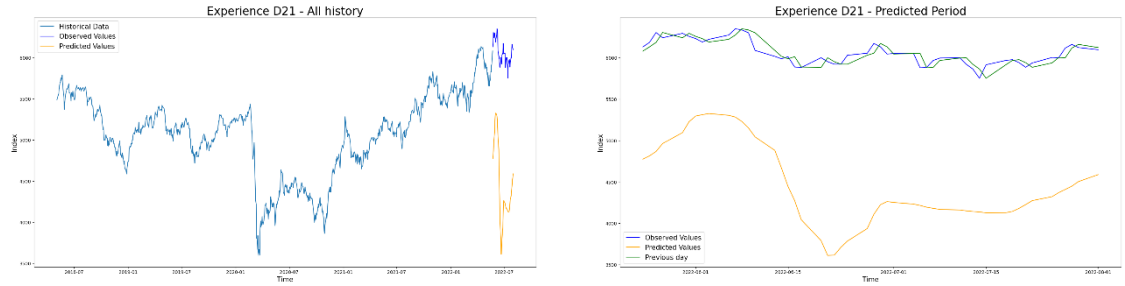
App_Figure 4 D1 Historical data and model generated predictions



App_Figure 5 Test D3 Historical data and model generated predictions



App_Figure 6 Test D11 Historical data and model generated predictions



App_Figure 7 Test D21 Historical data and model generated predictions

Experience #3B

Data Processing

The X_Data is a multi-dimensional array. For each value of the Y_Data the previous 90 days of the PSI20 closing values and the previous 90 days of a foreign stock exchange closing value are considered. The Y_Data is the PSI-20 index closing value of the 91st day and is the value that the model will try to predict.

A second variation of this test will be performed with the log return of each value, to train the model with the variations instead of the absolute index closing value. The Y_Data is the difference between the logarithm of the 91st and 90th observation.

Complete Test Results (with data scaled 0-1)

Test#	Years data	Epochs	Learn rate	Units per layer	Prediction RMSE	Previous Day RMSE	End date	Other index
E1	10	150	0.001	50;25	47.81	46.32	01-01-2023	IBEX35
E2	5	150	0.001	50;25	49.29	46.32	01-01-2023	IBEX35
E3	5	150	0.001	50;25	69.05	64.31	01-06-2022	IBEX35
E4	5	150	0.001	50;25	46.94	46.32	01-01-2023	CAC
E5	5	150	0.001	50;25	71.94	67.93	01-01-2023	NYSE

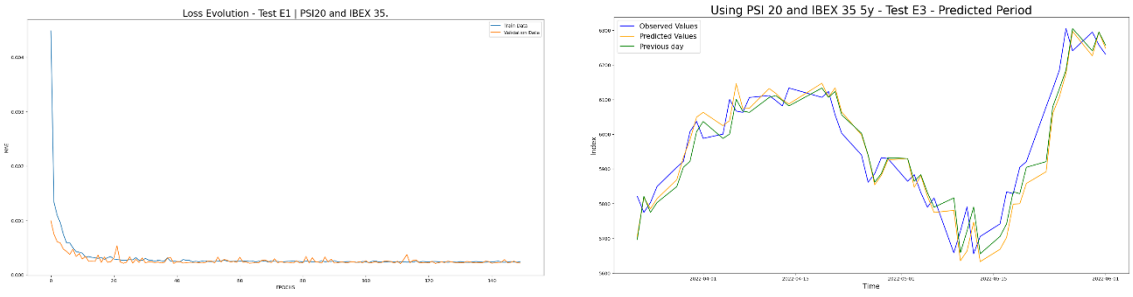
App_Table 5 - Experience #3B Complete Test Results (with data scaled 0-1)

Experiment variation

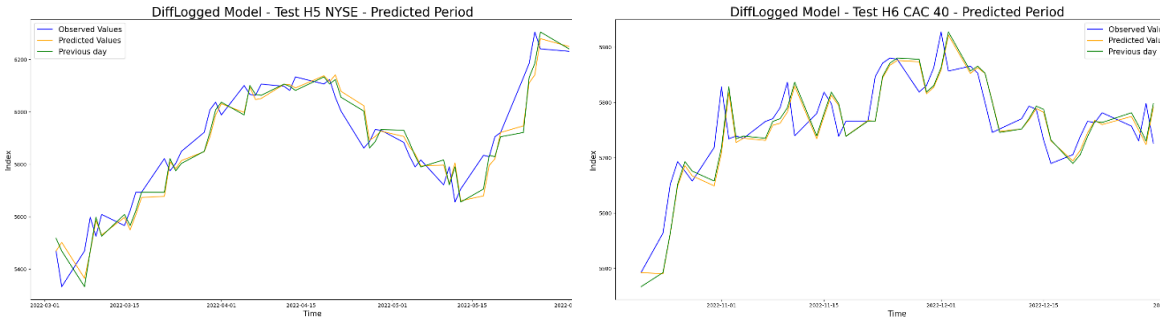
As mentioned on this experiment *data processing,* a second strategy was used based on the variations in regards of the last day of other world index (using the log return) and use it alongside the variations of the PSI20 to predict the log return of the PSI20 next day and extrapolate to the Portuguese index closing value using the cumulative sum.

Hyperparameters Model Evaluation

The number of epochs will be reduced to 150, and the RMSE will be used using as benchmark the previous day, as usual in this thesis.



App_Figure 8 Model E2 learning history and model E3 generated predictions scaled values of PSI + IBEX.



App_Figure 9 Model generated predictions of tests H5 and H6, using log-return of PSI + NYSE (Left) and PSI + CAC (right).

Test results (with log returns)

Test#	Years data	Epochs	Learn rate	Units per layer	Prediction RMSE	Previous day RMSE	End date	Other index
H1	5	150	0.001	50;25	45.42	46.32	01-01-2023	IBEX35
H2	5	150	0.001	50;25	50.60	49.91	01-01-2022	IBEX35
H3	10	150	0.001	50;25	48.52	46.32	01-01-2023	IBEX35
H4	5	150	0.001	50;25	61.26	67.93	01-01-2023	NYSE
H5	5	150	0.001	50;25	80.30	76.91	01-06-2022	NYSE
H6	5	150	0.001	50;25	46.25	46.32	01-01-2023	CAC30

App_Table 6 Experiment #3B complete test results (with log returns)

Experience #3C

Data Processing

The training labels (X_Data) is a multidimensional array with the scaled previous N days closing values of PSI20 and the previous N days of the other stock index (CAC40, IBEX35, NYSE) for each timestep. The target output (Y_Data) is the closing value for each day. All values are scaled between 0 and 1, where 0 is the minimum value recorded for each index and 1, the maximum value.

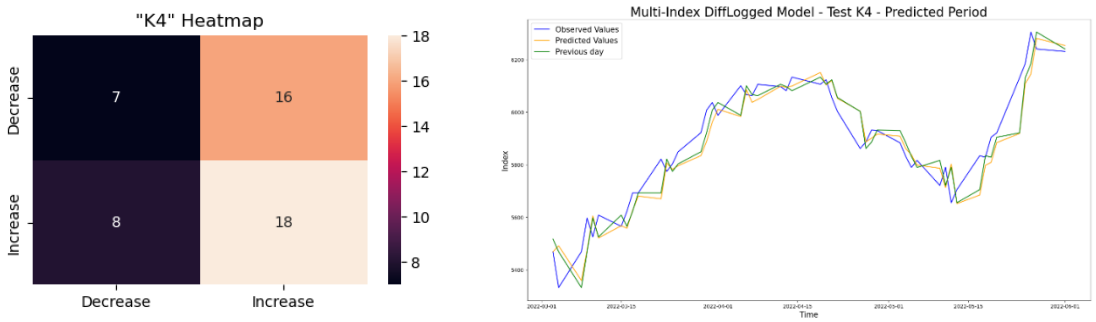
Two data scaling methods (DSM) are again used: one with the scaled values and the second with the log returns for both training labels and target output.

Hyperparameters and model evaluation

Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.001 learning rate for 150 epochs, and a validation data split of 20% and a batch size of 8.

As usual, the Root Mean Squared Error (RMSE) is the chosen metric, using the day before as a Benchmark.

For the data based on variation an accuracy metric is also used. The percentage of times the model gets the signal of the variation right is calculated and a value above 50% outperforms a pure "random" guess.

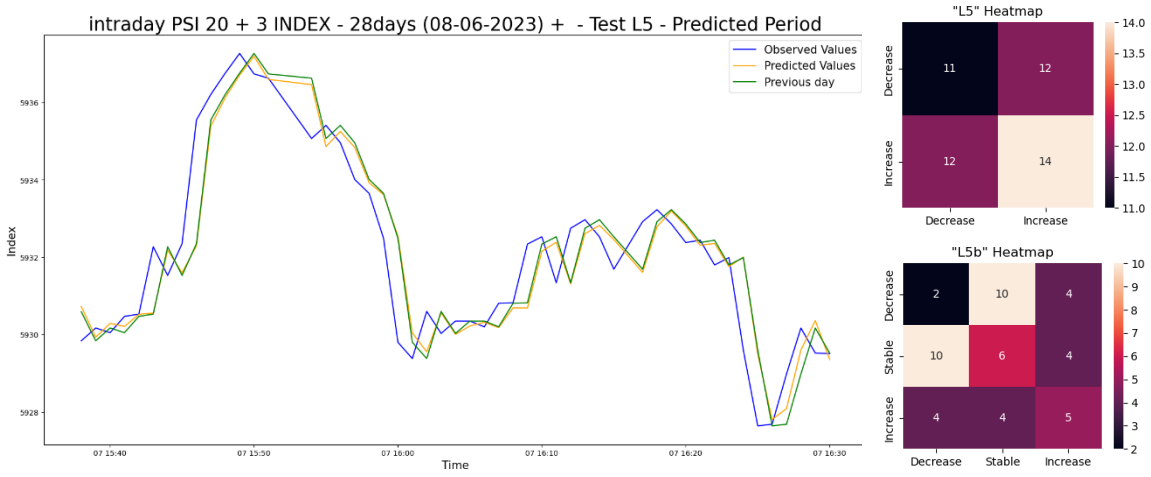


App_Figure 10 "Heat Map" and graphic generated by test K4 model

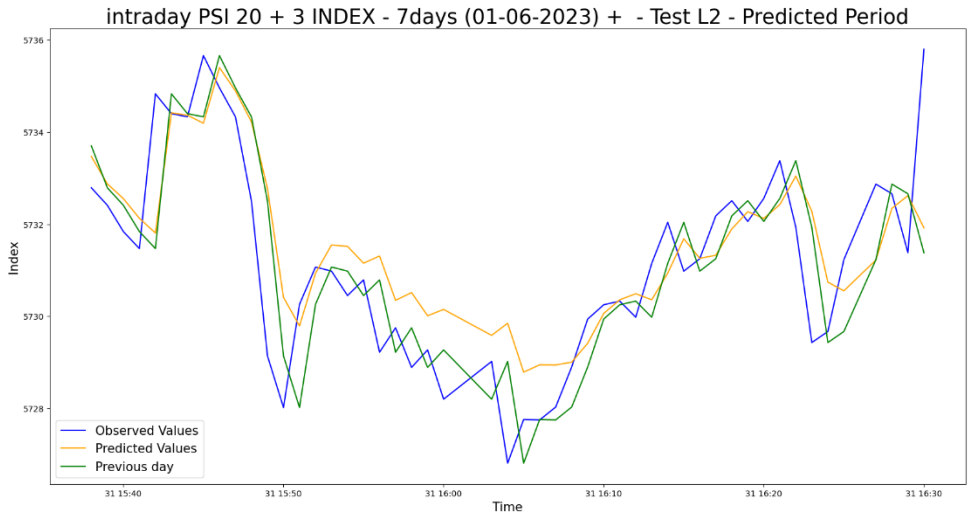
Forecasts with intra-day data

Group of experiments #4

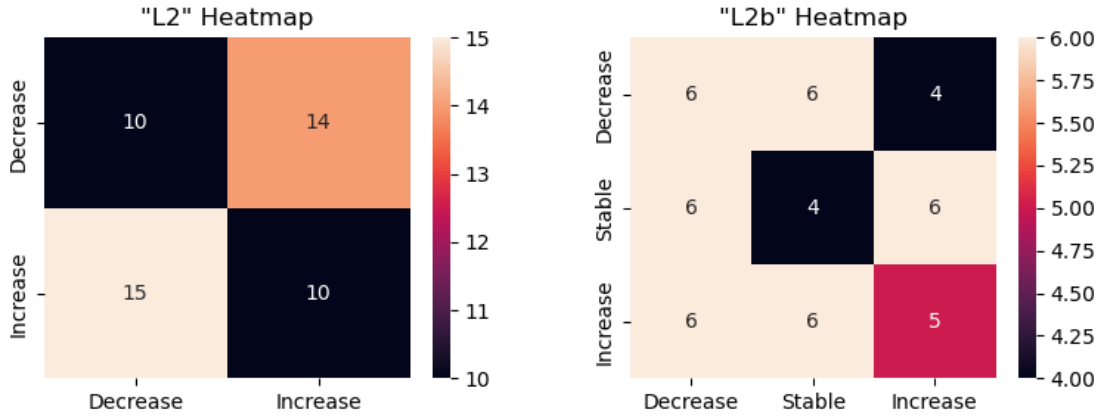
Experience #4A



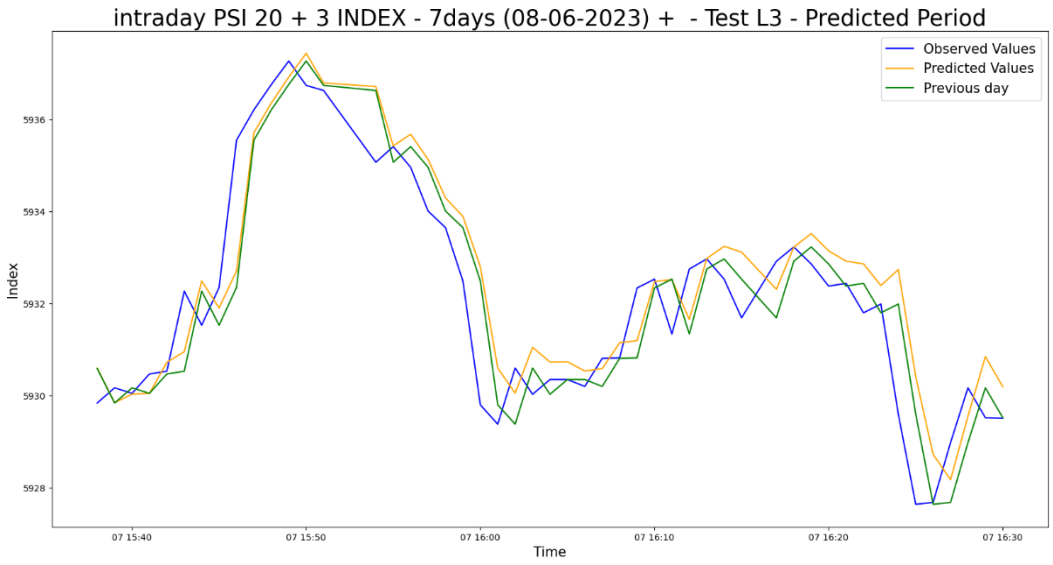
App_Figure 11 Test L5 Model generated prediction and heatmaps with 2 and 3 possible outcomes.



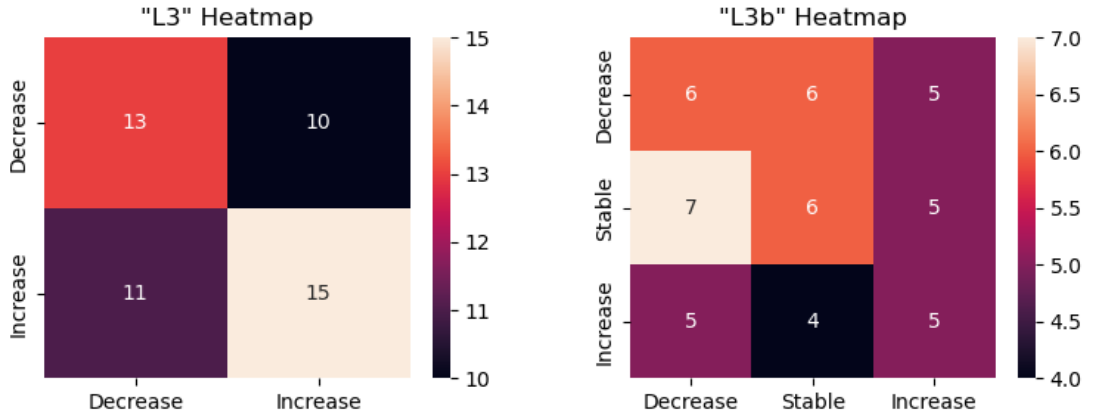
App_Figure 12 Test L2 Graphic



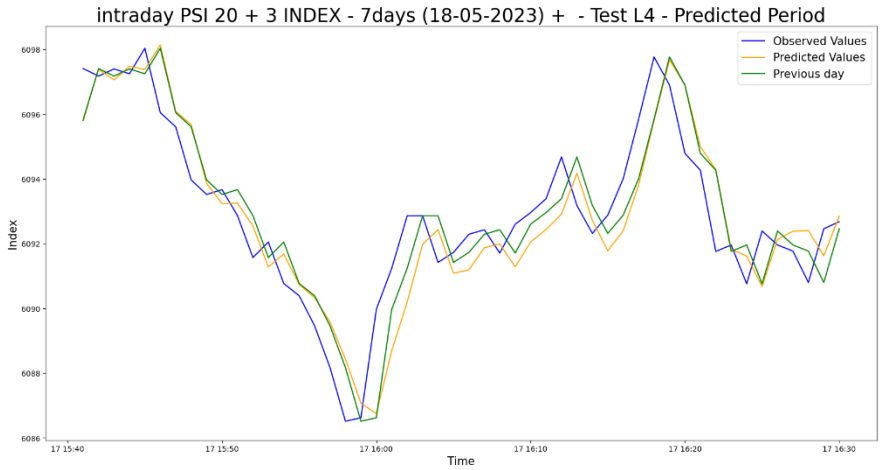
App_Figure 13 Test L2 Heatmaps



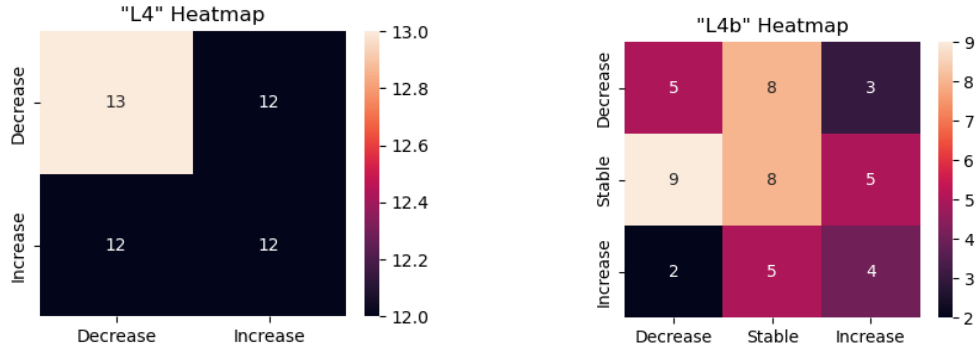
App_Figure 14 Test L3 Graphic



App_Figure 15 Test L3 Heatmap



App_Figure 16 Test L4 Graphic



App_Figure 17 Test L4 heatmaps

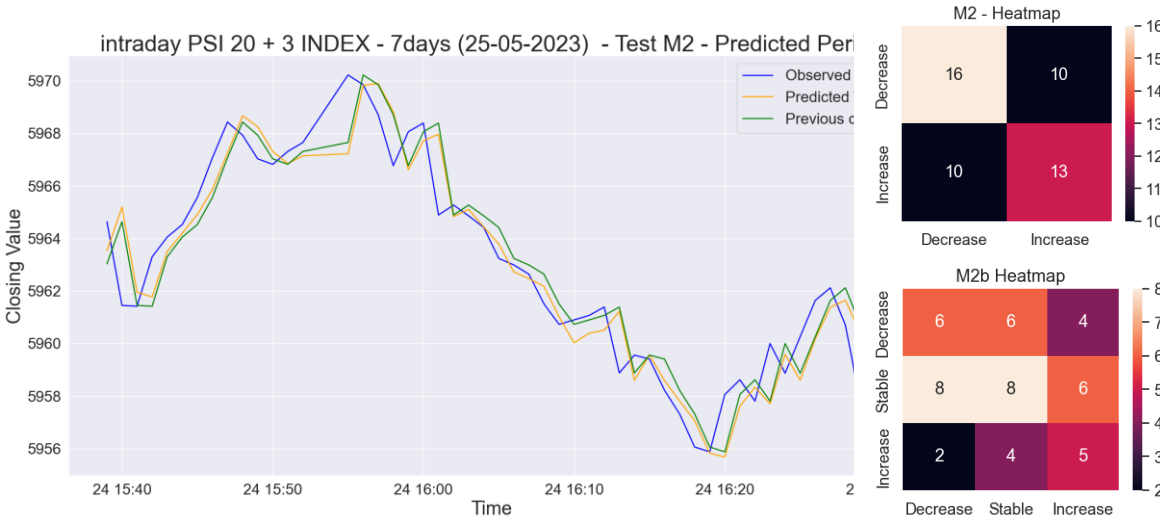
Experiment #4B

Data Processing

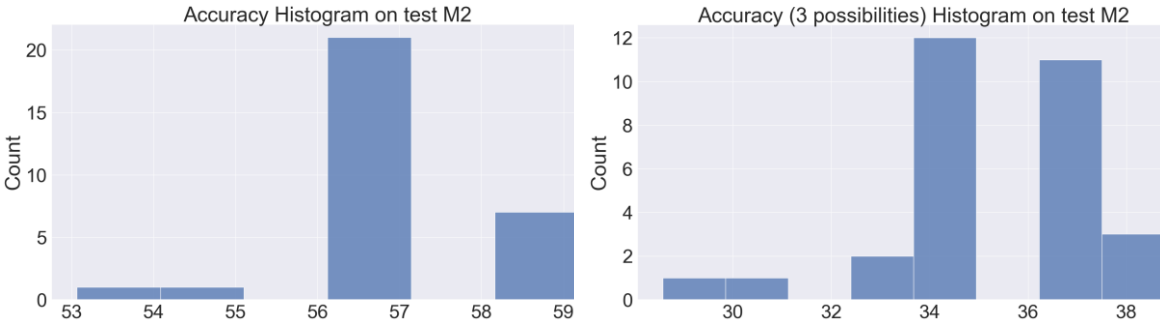
Exactly the same as in in previous experiment. Models will use for training a factor with a sliding window of the last 60 minutes of trading of the PSI20, IBEX, CAC and DAX stock indexes to try to antecipate PSI20 value in the next minute.

Hyperparameters

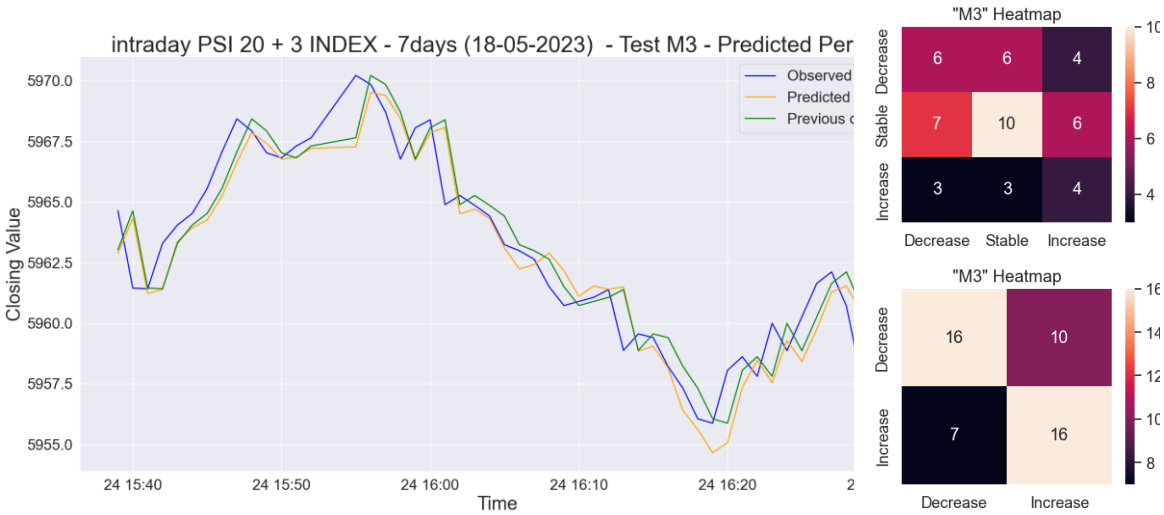
Two LSTM layers, one of 50 neurons and other of 25 neurons, and the ADAM predictor with a 0.001 learning rate for 150 epochs, and a validation data split of 20% and a batch size of 8.



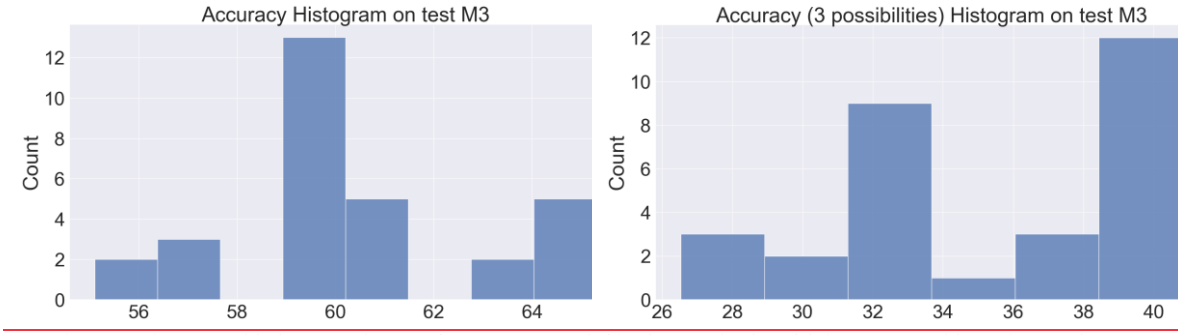
App_Figure 18 Predicted value and accuracy Heat Maps for the BEST models generated in test M2.



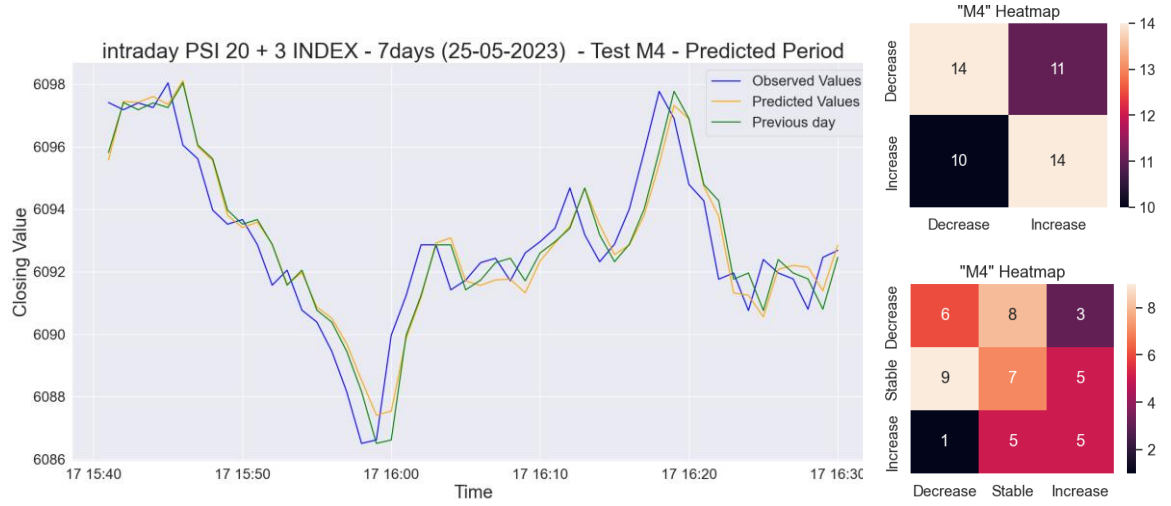
App_Figure 19 Histogram with the accuracies (2 and 3 outputs) recorded by the 30 trained models in test M2.



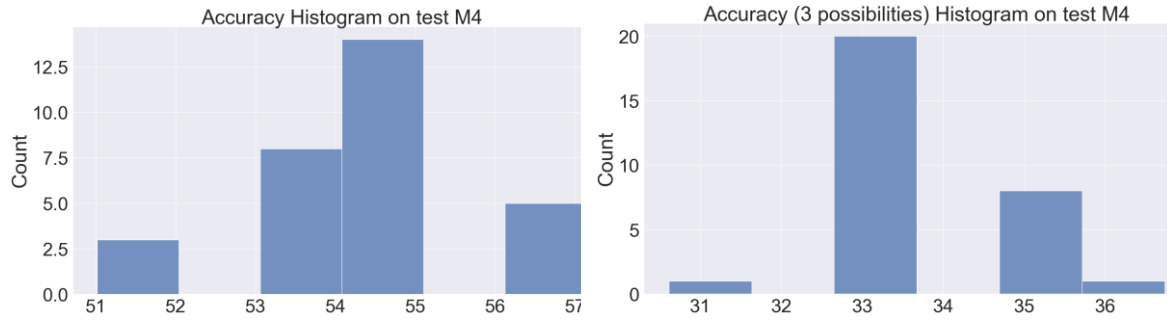
App_Figure 20 Predicted value and accuracy Heat Maps for the BEST models generated in test M3. (With test data from Week 2)



App_Figure 21 Histogram with the accuracies (2 and 3 outputs) recorded by the 30 trained models in test M3. (With Week 2 test data)

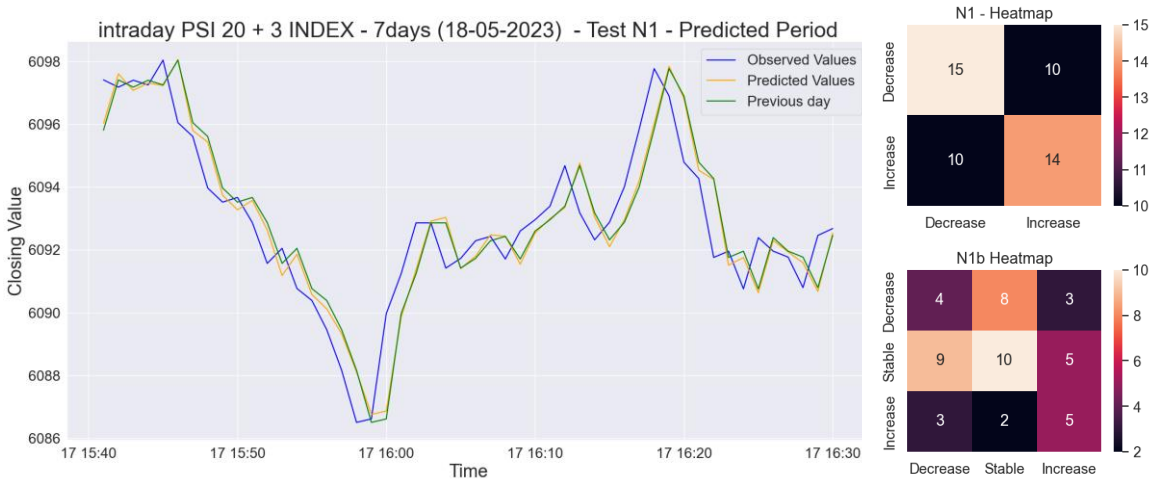


App_Figure 22 Predicted value and accuracy Heat Maps for the BEST models generated in test M4.

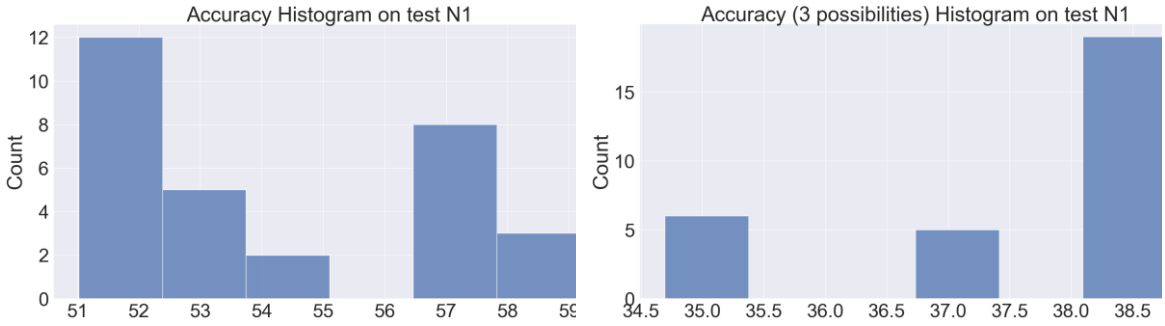


App_Figure 23 -histograms of test M4.

Experiment #4C



App_Figure 24 Predicted value and accuracy Heat Maps for the BEST models generated in test N1.



App_Figure 25 Histogram with the accuracies (2 and 3 outputs) recorded by the 30 trained models in test N1