Proceedings of the 2nd GCC International Conference on Industrial Engineering and Operations Management Muscat, Oman, December 1-3, 2024

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/GC02.20240125
Published: December 01, 2024

AutoML Driven LSTM and Stock Price Prediction Effectiveness: A New Frontier for Volatile Stock Markets.

Mariam Ait AL, Said Achchab
ENSIAS, Mohammed V University, Rabat, Morocco
mariam aital@um5.ac.ma , s.achchab@um5s.net.ma
Younes Lahrichi
ISCAE, Casablanca, Morocco
ylahrichi@groupeiscae.ma

Abstract

Conventional prediction methods, whether derived from deep learning or statistical techniques, typically demand significant manual intervention, making full process automation challenging. This study investigates the comparative robustness of traditional Long Short-Term Memory (LSTM) models and automated machine learning (AutoML) techniques, specifically TPOT, in predicting stock market behavior. By conducting experiments on the Moroccan stock market, the results demonstrate that AutoML approaches not only enhance prediction accuracy but also simplify the modeling process, making advanced prediction techniques more accessible and effective in dynamic financial markets.

Keywords
Financial time series, stock price prediction, market volatility, LSTM, AutoML.

1. Introduction
Predicting stock market's price fluctuations represents a core challenge in the financial field given its strategic importance for researchers but also for investors and regulators. Conventional statistical techniques and fundamental machine learning approaches frequently find it challenging to address the intricate and ever-changing characteristics of financial markets. In this context, Deep Learning, especially LSTM networks, has proven to be an effective instrument with respect to their capacity to capture temporal dependencies and non-linear patterns in time-series data.

LSTM models, a type of Recurrent Neural Network (RNN), are explicitly designed to address and mitigate the issues of gradient vanishing and exploding, common in standard RNNs. This capability makes LSTMs ideally suited for modeling sequences and financial time series, which are characterized by long-term dependencies and periodic volatility. Despite their advantages, the design and tuning of LSTM networks require significant expertise and are computationally intensive, involving extensive trial-and-error to identify optimal configurations.

Automated Machine Learning (AutoML) constitutes a promising approach to address the above mentioned challenges. AutoML seeks to streamline the complete process of implementing machine learning solutions in order to address practical challenges. Through the automation of model selection, composition and parameterization, AutoML significantly speeds up the model development process. Moreover, it makes available to non-specialists advanced machine learning techniques.

The potential of AutoML to improve LSTM models for stock market's price prediction is notably intriguing, as it may result in an improved accuracy in forecasts and more effective decision-making resources for investors and analysts navigating volatile markets.

Traditional financial forecasting tools heavily rely on human adjustments, requiring substantial expertise and effort. The amalgamation of AutoML with LSTM models offers a distinct benefit by automating hyperparameter adjustment, hence improving prediction accuracy and democratizing access to high-performing models for users with constrained technical expertise. In contrast to conventional LSTM models or statistical techniques, AutoML-driven methodologies such as TPOT methodically investigate model configurations, enabling a more extensive and efficient search for optimum parameters, particularly advantageous in highly volatile markets.

This paper aims to explore and empirically compare the performance of traditional LSTM models with those optimized through AutoML techniques. By conducting a series of experiments on real-world stock market data characterized by high volatility, this study evaluates whether AutoML-enhanced LSTMs can outperform their manually tuned counterparts in terms of prediction accuracy and computational efficiency. Through this research, we seek to contribute to the financial technology field by providing insights into the robustness of AutoML in enhancing predictive models and by offering a robust framework for deploying these advanced techniques in the analysis of volatile financial markets, such as the Moroccan market

This research conducts a comparative study of the robustness of AutoML versus traditional LSTM models, focusing specifically on their capability to predict stock market volatility within the context of the Moroccan stock market. The research seeks to illustrate that AutoML, utilizing the TPOT framework, possesses a superior capacity to navigate market volatility compared to manually optimized LSTM models. Through the automation of hyperparameter optimization, AutoML possesses the capacity to improve predictive accuracy while simultaneously mitigating computational inefficiencies, especially within dynamic and volatile market contexts. An essential aim is to perform a comprehensive analysis of the two methodologies, emphasizing the benefits of AutoML in effectively addressing swift market changes that are frequently overlooked by conventional LSTM models. Furthermore, the research aims to substantiate these conclusions through the utilization of actual stock data from the Casablanca Stock Exchange, thereby establishing a comprehensive framework applicable to the forecasting of financial time series. The research aspires to enhance the expanding domain of financial technology by illustrating that AutoML provides a superior and more scalable approach for forecasting volatile markets in contrast to conventional models.

The subsequent sections of this study are organized as follows: Initially, we will examine significant literature about the position of using LSTM and AutoML for stock market forecasting. Subsequently, we will examine a comparison of LSTM and AutoML, particularly within the context of the volatile Moroccan market. This will provide a thorough conclusion of AutoML's impact on the automation of stock market forecasting and its capacity to improve financial decision-making.

2. Literature Review

The quest to predict stock market movements using machine learning techniques has intensified in recent years, leveraging advances in computational power and algorithmic innovations. The focus of recent research has often centered around the integration of deep learning models, particularly LSTM networks, and the application of AutoML to enhance these models' performance in volatile financial environments.

The accurate prediction of financial markets, particularly in volatile environments, has long been a critical challenge in the field of time series forecasting. As market conditions fluctuate rapidly, the need for sophisticated models capable of capturing complex temporal dependencies becomes increasingly important. LSTM networks have emerged as one of the leading approaches to tackle this challenge due to their ability to maintain and process information over long sequences, effectively handling the non-linear and chaotic nature of financial time series data (Fischer and Krauss 2018, Wen and Li 2023, Gers et al. 2002). However, LSTM model performance heavily depends on precise hyperparameter tuning and architecture design, which traditionally demand significant expertise and time.

The introduction of AutoML frameworks presents a promising solution to these challenges by automating the process of model development. AutoML can optimize hyperparameters, select appropriate model architectures, and even combine multiple models to enhance performance (Su et al. 2024, Conrad et al. 2024). This automation not only reduces the time and expertise required for model tuning but also improves the robustness and accuracy of the predictions made by LSTM networks (He et al. 2021, Salehin et al. 2024, Westergaard et al. 2024). The potential of AutoML to transform the application of LSTM in financial forecasting, particularly in volatile markets, is therefore a subject of growing interest.

Given the importance of accurate forecasting in volatile markets, where minor errors can lead to significant financial losses, it is crucial to understand how the integration of AutoML with LSTM networks affects predictive performance. While LSTM models have already demonstrated their superiority over traditional statistical methods in stable market conditions, their effectiveness in volatile environments, with and without the assistance of AutoML, remains to be thoroughly investigated.

By examining the performance gaps, this research tries to contribute to a deeper understanding of the benefits and limitations of AutoML-enhanced LSTM models, especially in the context of high volatility financial markets. The findings could provide insights for both academic research and practical applications in the field of financial forecasting.

3. Data Collection

The FTSE Index Dataset used in our study originated from the Casablanca Stock Exchange (CSE) official website, encompassing detailed trading metrics such as closing values, daily highs, and lows., which ensures accuracy and relevance for financial analysis, particularly for modeling the Moroccan stock market dynamics.

These data are processed to ensure robust analysis and model training. Initial cleaning steps include removing any entries with missing values to maintain data integrity and consistency. Financial metrics critical for analysis, such as the bid-ask spread, are computed from market liquidity measurement, which serves as a proxy for market liquidity. Additionally, volatility is calculated using the logarithmic returns of the closing prices, offering insights into market stability and risk. To maintain the integrity of time-series data, the dataset is chronologically sorted by the date of trading sessions. Each entry is timestamped and indexed by its corre-sponding session date, ensuring that all financial analytics and subsequent predictive modeling respect the sequential nature of stock market data. This meticulous preprocessing not only prepares the dataset for effective modeling but also aligns it with the analytical needs of forecasting market movements.

Figure 1. serves as a pivotal analytical tool, offering deep insights into the market's behavior over the observed period. By charting the volatility derived from the annualized standard deviation of the logarithmic returns of the index, we gain a clear visual representation of the temporal dynamics of market uncertainty and risk.

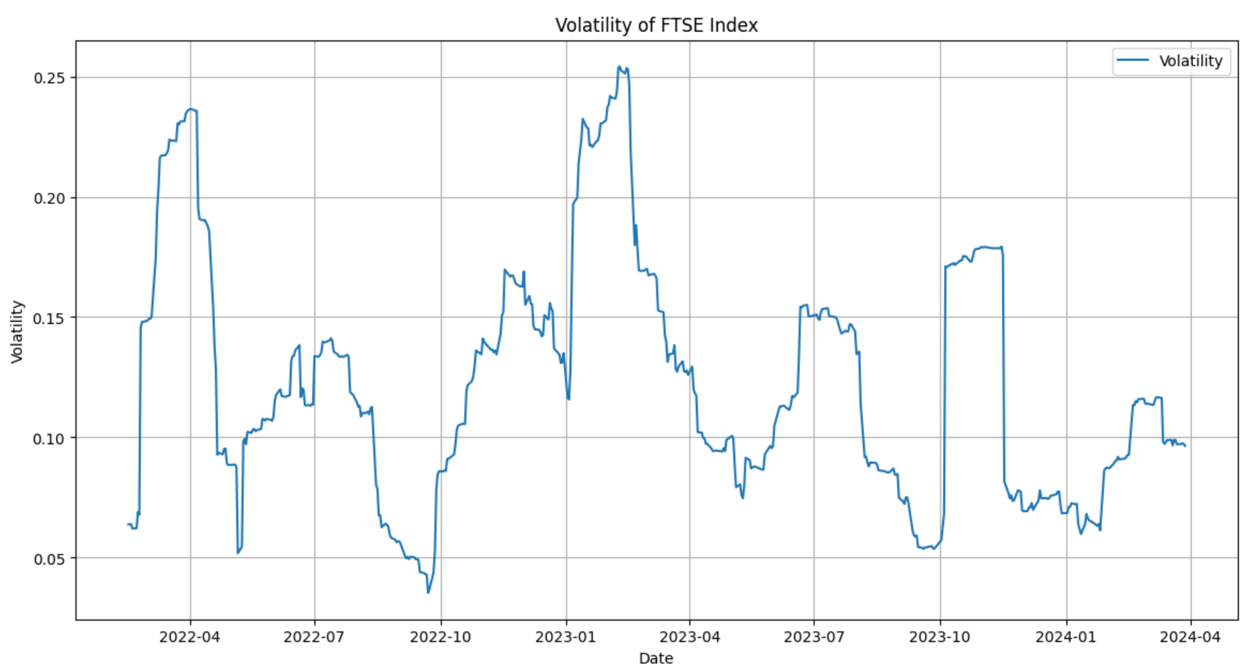

Figure 1. The volatility in the FTSE Index dataset

Volatility, as depicted in the plot, is not constant but varies significantly over time, demonstrating periods of relative calm interspersed with spikes that indicate heightened market activity or stress. These fluctuations are essential for understanding the risk profile of the index and can inform investment strategies and risk management decisions. The plot typically shows a baseline level of volatility, reflecting the inherent market conditions and investor sentiment under normal circumstances. However, the peaks in the plot reveal instances of increased volatility, which often correspond to external market shocks, economic announcements, or significant geopolitical events.

Each spike in the volatility plot can be analyzed to correlate with specific historical events, providing a narrative of how external factors influence market behavior. For example, a sharp increase in volatility might coincide with political instability, economic policy changes, or major corporate announcements within the companies that constitute the FTSE Index. These insights are invaluable for traders, investors, and financial analysts as they attempt to predict future market movements based on past patterns.

Furthermore, the volatility plot is not just a passive reflection of past market conditions but can be actively used in predictive analytics. By employing statistical and machine learning models, analysts can use the patterns observed in the plot to forecast future volatility. This predictive aspect is crucial for derivative pricing, portfolio optimization, and risk management strategies, where anticipating market volatility is as essential as reacting to it. These fluctuations underscore the need for forecasting models that can adapt to and learn from complex patterns in data.

4. Models

4.1 Long Short-Term Memory (LSTM)

LSTM networks, a special class of recurrent neural networks, are particularly well-suited for this task. Their ability to process entire sequences of data and remember information over long periods makes them ideal for time-series data like stock prices, where the future value may depend on a lengthy history of past values (Koo et al. 2023, Razavi et al. 2019).

The implementation of an LSTM model for volatility forecasting begins with meticulous data preprocessing. The dataset is first cleaned by converting date columns to date-time objects and removing any entries with missing values, ensuring that subsequent analyses rest on a solid foundation of complete and accurate data.

Feature selection follows, focusing on variables that directly impact or reflect market volatility, such as high and low prices, daily and yearly variations, and the bid-ask spread. These features are normalized using the MinMaxScaler to ensure that the LSTM model receives data within a scale appropriate for neural network training. This normalization not only aids in faster convergence during training but also prevents the model from developing a bias toward variables on larger scales.

The crux of LSTM modeling lies in transforming the time-series data into a format suitable for sequential processing. This is achieved through the creation of sequences and corresponding targets, where each sequence consists of several consecutive days' worth of data, and the target is the volatility at the next time step. This setup captures the temporal dependencies inherent in the data, allowing the LSTM model to make predictions based on the observed patterns.
The architecture of the LSTM model is designed to optimize learning from these sequences. It includes multiple layers of LSTM units, each capable of learning different aspects of the sequence pattern. Dropout layers are interspersed between LSTM layers to reduce overfitting by randomly omitting a proportion of the features during training. This technique enhances the model's generalizability to new, unseen data. The final output of the model is a single dense layer that predicts the volatility value, translating the features learned from the sequences into a practical output. (Wen and Li 2023; Gers et al. 2002; Sherstinsky 2020; Staudemeyer and Morris 2019). Training the LSTM model involves adjusting numerous parameters, such as the number of epochs, batch size, and the architecture of the LSTM layers.

These parameters are critical as they directly influence the model's ability to learn effectively. The model is compiled with an Adam optimizer, that enables tackling effectively the issue of gradient sparsity inherent to noisy information such as stock market financial data of noisy problems and a mean squared error loss function that quantifies the prediction error in terms appropriate for continuous data.

The evaluation of model performance is conducted through Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which provide insights into the average magnitude and the square root of the prediction errors, respectively. These metrics are crucial for understanding the accuracy of the forecasts and for comparing the LSTM model's performance against other models or benchmarks within financial analytics.

Despite their widespread adoption in financial forecasting, LSTMs come with inherent limitations that can restrict their effectiveness in dynamic market environments.

The complexity of LSTM models makes them subject to overfitting, particularly when the available financial data is noisy and non-stationary. Overfitting results in models that excel on training data but struggle to generalize to novel data, rendering them ineffective for forecasting future market conditions. Additionally, the training process of LSTMs is computationally demanding, which can be a significant bottleneck, limiting their scalability and responsiveness—key attributes needed in fast-paced financial environments.

The complexity and "black-box" nature of LSTM models also pose challenges in terms of interpretability. Financial institutions often require clear explanations of the predictions made by models to justify decisions to stakeholders and comply with regulatory requirements (Bokhare et al. 2024, Ren et al. 2019). The intricate internal mechanisms of LSTMs make such explanations non-trivial, often necessitating additional tools and techniques to elucidate how inputs are transformed into outputs.

4.2 Automated Machine Learning (AutoML)

The integration of AutoML with LSTM networks presents a compelling solution to these challenges, pushing the boundaries of what can be achieved with traditional machine learning methodologies in financial forecasting. AutoML aims to automate and optimize the processes of model selection, hyperparameter tuning, and feature engineering, thereby addressing several of the limitations associated with LSTMs.

AutoML significantly simplifies the hyperparameter optimization process through algorithms designed to efficiently search the parameter space, such as Bayesian optimization, genetic algorithms, or even more sophisticated ensemble methods that combine multiple approaches.

When manually tuning LSTM models, practitioners often rely on their experience and intuition to navigate the complex landscape of model architecture and hyperparameter settings. While this approach allows for bespoke model development tailored to specific forecasting tasks, it is inherently limited by the practitioner's understanding and the feasibility of exhaustively exploring the model configuration space.

In contrast, AutoML provides a systematic and data-driven approach to model development. By automating the exploration of model architectures and hyperparameters, AutoML not only uncovers potentially superior models but also does so with greater efficiency and consistency. The ability of AutoML to dynamically adjust to the data ensures that the models it develops are both accurate and adaptable to new market conditions, a crucial advantage in the volatile world of finance.

The integration of AutoML into the LSTM modeling process represents a paradigm shift in financial forecasting. By addressing the core limitations of traditional LSTM networks through automation, AutoML enhances the robustness, scalability, and interpretability of predictive models. This synergy between LSTM's deep learning capabilities and AutoML's optimization prowess offers a potent solution to the challenges of financial time series forecasting, providing analysts and traders with more reliable and actionable insights.

5. Results and Discussion

The study targeted evaluating the effectiveness of an LSTM model, a type of recurrent neural network (RNN) that is particularly well-suited for time series forecasting, against an optimized model developed using TPOT, a Python AutoML tool that optimizes machine learning pipelines using genetic programming, which is designed to automatically search for the best machine learning pipelines. The comparison focused on how well each model could predict future values based on historical data, with specific attention given to the financial time series FTSE data provided.

The LSTM model was developed using the TensorFlow and Keras libraries. The architecture consisted of two LSTM layers, the first with 50 units and the second also with 50 units, but without returning sequences, followed by two dense layers, the first with 25 units and the final output layer with a single unit to predict the target variable. This architecture was chosen due to LSTM's capability to capture long-term dependencies in sequential data.

The model was compiled using the Adam optimizer and the mean squared error (MSE) loss function. Adam was selected due to its adapted learning rate often results in accelerated convergence and improved efficiency in practice. The model experienced training for 5 epochs, utilizing a batch size of 1. The relatively small number of epochs was chosen to prevent overfitting, given the model's complexity and the potential for the LSTM to memorize the training data if trained for too long.

In parallel to the LSTM model, an AutoML approach was employed using TPOT, a tool that automates the process of model selection and hyperparameter tuning. TPOT uses genetic programming to explore a wide range of machine learning pipelines and configurations, optimizing for the best performance (Kaftantzis et al. 2024). The TPOT model was trained on the same training data, but instead of using sequences as in the LSTM model, the data was reshaped into a flat format suitable for traditional machine learning models.

The TPOT process involved 5 generations with a population size of 20, meaning that TPOT explored 20 different models in each generation, iteratively improving the models over 5 iterations. The final model chosen by TPOT was the one that minimized the mean squared error on the validation set. This approach allows TPOT to discover potentially better-performing models that might not be immediately apparent through manual tuning.

At the conclusion of training process, predictions were produced on the test set for further investigation. The predictions were then inverse processed using the scaler to revert them to the original data scale, facilitating a direct comparison with the actual results. The models' performance was assessed using root mean squared error (RMSE), a prevalent statistic in regression tasks that quantifies the average amount of the discrepancy between predicted and actual values.

Figure 2 plots a clear visual comparison of how closely each model's predictions aligned with the true data points. The LSTM model's predictions, shown in orange, generally followed the trend of the actual values but exhibited some lag during rapid market changes. The TPOT model's predictions, shown in green, appeared to track closer to the actual values, particularly during periods of high volatility.

The comparative analysis of the two models revealed several key insights. Firstly, both models demonstrated a strong ability to capture the overall trend in the financial time series data, indicating that both LSTM and TPOT are viable options for this type of forecasting. However, there were notable differences in performance, particularly in how each model handled periods of rapid market change.

The LSTM model, despite its sophistication and ability to capture long-term dependencies, struggled slightly during periods of high volatility. This is likely due to the inherent difficulty of predicting sudden market shifts, which are often influenced by factors outside the scope of the data provided. In contrast, the TPOT option, by optimizing over a broader range of potential models and hyperparameters, was able to produce predictions that were more responsive to these changes. This suggests that the AutoML approach was effective in identifying a model configuration that better captured the dynamics of the financial market.

Figure 2 illustrates the comparative forecasts of the LSTM and TPOT-optimized models, emphasizing market volatility over training epochs. This investigation evaluates the forecasts of both models during significant market occurrences. Periods of instability and economic announcements are characterized by significant fluctuations in real market prices. The TPOT model exhibits superior alignment with real data during high-volatility periods, underscoring the efficacy of AutoML in adapting to rapid market fluctuations. The LSTM model exhibits a delayed reaction to these fast changes, indicating constraints in its adaptability to sudden external factors.

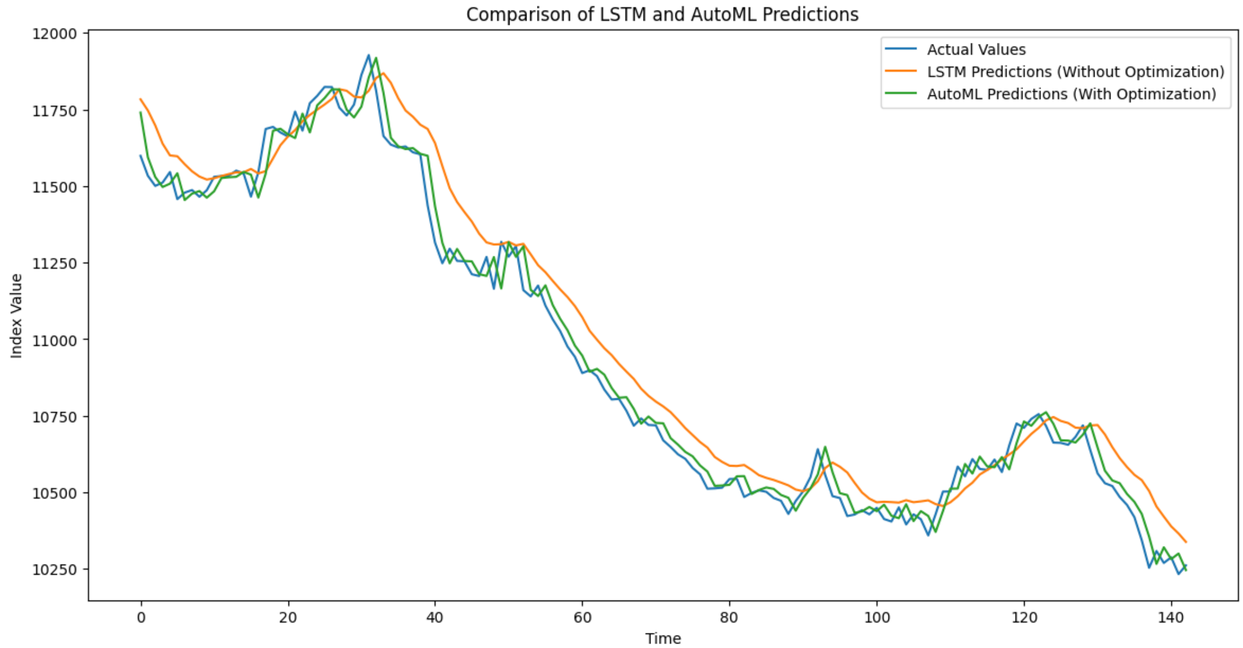

Figure 2. Comparison of Residuals During Training

For practitioners in the field of financial forecasting, these results suggest that while LSTM models are a powerful tool for capturing temporal dependencies, there is value in exploring AutoML approaches like TPOT that can optimize model performance in a more automated and potentially more effective manner. The ability of TPOT to discover better-performing models through genetic programming offers a promising avenue for enhancing the accuracy of financial predictions, particularly in volatile markets.

6. Conclusion

In this study, we investigated the application of Long Short-Term Memory (LSTM) networks and Automated Machine Learning (AutoML), focusing on TPOT for forecasting volatile financial time series. Both models exhibited robust predictive ability. Nevertheless, AutoML surpassed LSTM, particularly in managing market volatility, due to its automated model selection and hyperparameter optimization. Nonetheless, the "black-box" characteristic of AutoML poses transparency issues, limiting understanding of the precise elements influencing its efficacy. Our results underscore AutoML's capacity to improve financial forecasting by providing a scalable, efficient solution that minimizes the need for manual adjustment, hence making it suitable for fluctuating market circumstances. Subsequent study needs to investigate alternative AutoML frameworks and emphasize interpretability to improve comprehension of long-term patterns and infrequent market occurrences, hence augmenting transparency for decision-making processes in finance.

References

Bokhare, A., Rao, M., Oliver, M.P., Rai, R., and Adesara, U. "Modelling Stock Prices Prediction with Long Short-Term Memory (LSTM): A Black Box Approach." In: *Lecture notes in networks and systems*, pp. 65–73, 2024. https://doi.org/10.1007/978-981-99-8476-3_6.

Conrad, F., Mälzer, M., Lange, F., Wiemer, H., and Ihlenfeldt, S. "AutoML Applied to Time Series Analysis Tasks in Production Engineering." *Procedia Computer Science*, vol. 231, pp. 865-872, 2024. https://doi.org/10.1016/j.procs.2024.01.085.

European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.

Fischer, T., and Krauss, C. "Deep learning with long short-term memory networks for financial market predictions."

Gers, F.A., Eck, D., and Schmidhuber, J. "Applying LSTM to Time Series Predictable Through Time-Window Approaches." In: Tagliaferri, R., and Marinaro, M. (eds) *Neural Nets WIRN Vietri-01*. Perspectives in Neural Computing. Springer, London, 2002. https://doi.org/10.1007/978-1-4471-0219-9_20.

He, X., Zhao, K., and Chu, X. "AutoML: A survey of the state-of-the-art." Knowledge-Based Systems, vol. 212, Kaftantzis, S., Bousdekis, A., Theodoropoulou, G., and Miaoulis, G. "Predictive Business Process Monitoring with AutoML for Next Activity Prediction." *Intelligent Decision Technologies*, vol. 18, no. 3, pp. 1965-1980, 2024. https://doi.org/10.3233/IDT-240632.

Koo, E., and Kim, G. "A New Neural Network Approach for Predicting the Volatility of Stock Market." *Computational Economics*, vol. 61, pp. 1665–1679, 2023. https://doi.org/10.1007/s10614-022-10261-7.
pp. 106622, 2021. https://doi.org/10.1016/j.knosys.2020.106622.

Razavi, H., Sarabadani, H., Karimisefat, A., and Lebraty, J.-F. "Profitability Prediction for ATM Transactions Using Artificial Neural Networks: A Data-Driven Analysis." *Proceedings of the 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, pp. 661–665, Tehran, Iran, 2019. https://doi.org/10.1109/KBEI.2019.8735037.

Ren, B., and Duncan, I. "Modeling Oil Saturation Evolution in Residual Oil Zones: Implications for CO2 EOR and Sequestration." *Journal of Petroleum Science and Engineering*, vol. 177, pp. 106682, 2019. https://doi.org/10.1016/j.petrol.2019.106682.

Salehin, I., Islam, M.S., and Saha, P. "AutoML: A Systematic Review on Automated Machine Learning with Neural Architecture Search." *Journal of Information and Intelligent Systems*, vol. 2, no. 1, pp. 52–81, 2024. https://doi.org/10.1016/j.jiixd.2023.10.002.

Sherstinsky, A. "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network." *Physica D Nonlinear Phenomena*, vol. 404, pp. 132306, 2020. https://doi.org/10.1016/j.physd.2019.132306.

Staudemeyer, R.C., and Morris, E.R. "Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks." arXiv (Cornell University), 2019. https://doi.org/10.48550/arxiv.1909.09586.

Su, Y., Wang, M.C., and Liu, S. "Automated Machine Learning Algorithm Using Recurrent Neural Network to Perform Long-Term Time Series Forecasting." *Computers, Materials & Continua*, vol. 78, no. 3, pp. 3529-3549, 2024. https://doi.org/10.32604/cmc.2024.047189.

Wen, X., and Li, W. "Time Series Prediction Based on LSTM-Attention-LSTM Model." IEEE Access, vol. 11, pp. *48322*–48331, 2023. https://doi.org/10.1109/access.2023.3276628.

Westergaard, G., Erden, U., Mateo, O.A., Lampo, S.M., Akinci, T.C., and Topsakal, O. "Time Series Forecasting Utilizing Automated Machine Learning (AutoML): A Comparative Analysis Study on Diverse Datasets." *Information*, vol. 15, no. 1, pp. 39, 2024. https://doi.org/10.3390/info15010039.

Biographies

Mariam Ait Al is currently pursuing Ph.D. research at ENSIAS, Mohammed V University in Rabat, Morocco. She is also a Senior Software Engineer and Project Manager with extensive experience in consultancy for complex technical scenarios. Her professional expertise informs her academic work, focusing on bridging practical industry applications with theoretical advancements in technology. Her research interests include artificial intelligence, deep learning, and their applications in financial markets

Said Achchab Professor of digital finance, artificial intelligence, and risk management at ENSIAS, Mohamed V University, and the coordinator of the "Digital Engineering for Finance" engineering program. He is the founding chairman of the African Fintech Institute

Younes Lahrichi Ph.D, full professor and senior lecturer in Finance at ISCAE Casablanca. Autor of several academic publications, Dr. Prof. Younes Lahrichi is head of finance and accounting department and Master in digital finance program director.