Università Bocconi
MILANO

2023-2024

Complex Deep Learning Architectures For Predicting Stock Prices

FINANCE WITH BIG DATA

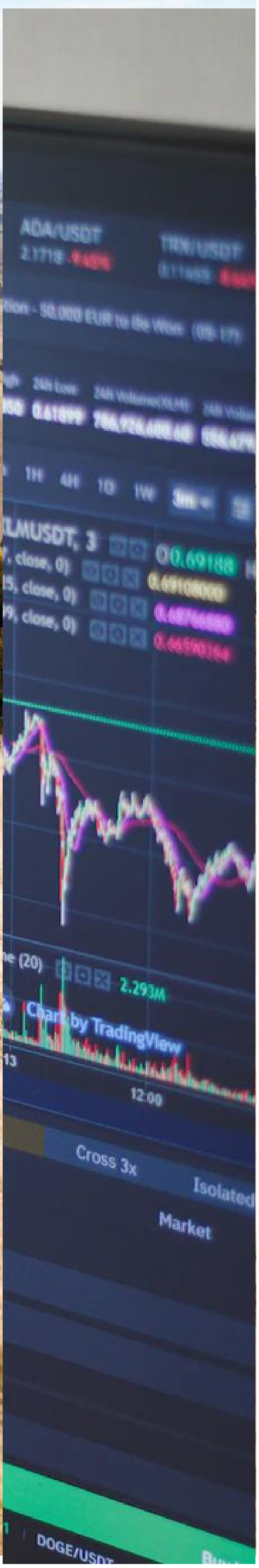
ANTONIO PREITI
SAMUELE COMPAGNONI

COMPLEX DEEP LEARNING ARCHITECTURES FOR PREDICTING STOCK PRICES

INDEX

FIGURES

INDEX



01. INTRODUCTION

This research explores the underlying mechanisms through which advanced deep learning models augment predictive performance in forecasting **S&P 500 stock closing prices**. Therefore, our primary objective is to see empirically the difference in performance of the models and try to understand the factors contributing to the superior efficacy of state-of-the-art models compared to a conventional feed-forward neural network **(ANN)**.

The models considered are **Long-Short-Term-Memory (LSTM)** and **Temporal Fusion Transformer (TFT)**. The **Artificial Neural Network (ANN)** is regarded as the **baseline** due to its comparatively simpler architecture among deep learning models.

Each model's predictive framework involves forecasting the closing price of the S&P 500 by utilizing **historical closing prices from previous days.**

The analysis considers two specific timeframes: the preceding **5 days and 20 days**. This exploration is crucial for comprehending how the predictive performance of each model varies with higher or lower numbers of lagged features. The metrics used for evaluating each model are the **Mean Absolute Error (MAE)**, the **Root Mean Square Error (RMSE)**, and the **Mean Absolute Percentage Error (MAPE)**.

The empirical results of our study highlight:

- Certain models exhibit improved performance with a reduced number of lagged values, while others demonstrate a preference for a higher series of lagged values. In the subsequent paragraphs, we aim to explain this observation by going through the architecture of each model.

- The superior predictive capabilities of advanced deep learning models, **LSTM** and **TFT**, in capturing the intricate dynamics inherent in stock prices.

02. DATA

Our primary dataset consists of daily closing prices of the S&P 500 index. The data were collected from **Yahoo Finance**. The temporal segmentation for the training, validation, and test sets adheres to a systematic approach.

The **training set** spans from January 1, 2010, to December 1, 2018.

The **validation set,** crucial for fine-tuning model parameters and preventing overfitting, is delineated from December 1, 2018, to January 1, 2019.

Finally, the **test set**, designed to assess the generalization performance of the developed model, covers the period from January 1, 2019, to February 28, 2019.

03. ANN MODEL - BASELINE

ARCHITECTURE STRUCTURE

The architecture of the **Artificial Neural Network (ANN)** we designed for stock prices forecasting includes an input layer, two fully connected layers, and an output layer. The input layer represents the features or variables used for prediction, the historical stock close prices. Neurons in the input layer pass information to neurons in the hidden layers through weighted connections.

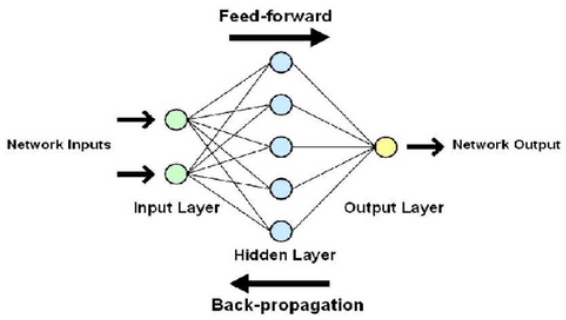


Fig.1: Typical architecture structure of an Artificial Neural Network

The hidden layers employ **non-linear activation functions** like **sigmoid** or rectified linear units **(ReLU)**. This is critical to extract complex non linear patterns and relationships from the input data.

The output layer produces the predicted stock prices based on the learned patterns. The model training adjusts the weights through **backpropagation**, minimizing the difference between predicted and actual close prices.

Our **ANN** is composed of two fully connected layers, containing **64 and 32 neurons**, respectively. To mitigate overfitting, two Dropout layers are incorporated. The network, comprising **2497 parameters**.

The below plots illustrate the predicted values versus the actual values for these two distinct time horizons.

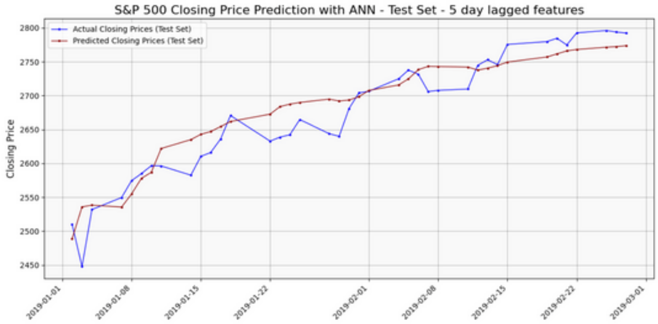


Fig. 2: ANN Prediction vs. Actual Values (Input: Previous 5 Days)

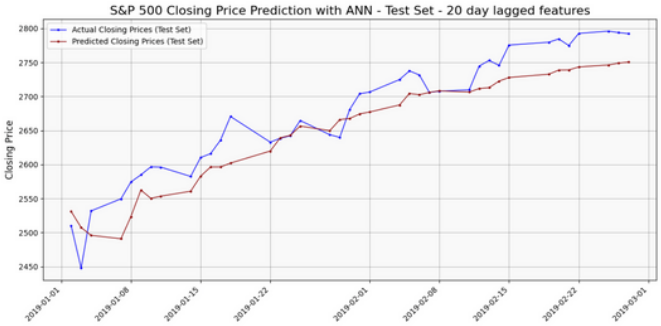


Fig. 3: ANN Prediction vs. Actual Values (Input: Previous 20 Days)

PERFORMANCE RESULTS & EMPIRICAL EVIDENCE

The ANN obtained a Mean Absolute Error **(MAE)** of **24.028** on the **test set** when utilizing the closing prices from the **preceding 5 days** as input. Alternatively, employing the closing prices of the **previous 20 days** results in an **MAE** of **30.813** on the test set.

As is possible to see from the above plot, when the model uses **20 lagged features** the predicted trend seems to be **flatter**. This happens because the **ANN processes the entire input sequence uniformly**. In the paragraph of the **Temporal Fusion Transformer** we describe that a way to achieve better performance is to **differentiate the weights of the input sequence**.

04. LONG-SHORT-TERM MEMORY (LSTM)

ARCHITECTURE STRUCTURE

The **Long Short-Term Memory (LSTM)** architecture, tailored for stock price prediction, stands out from **Artificial Neural Network (ANN)** due to its sequential memory and gating mechanisms. **LSTM** incorporates memory cells, three crucial gates **(forget, input, and output gates)**, and a unique ability to retain information over sequential steps.

Unlike **ANN**, LSTM captures long-term dependencies within stock price data, addressing challenges posed by non-linear patterns. The **forget gate** enables LSTM to selectively discard irrelevant historical information, while the **input gate** facilitates the assimilation of pertinent new data.

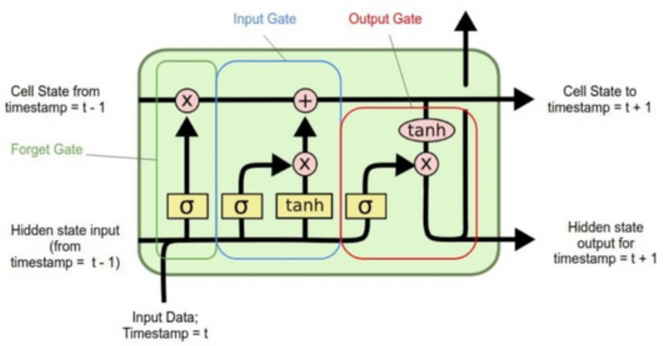


Fig.4 Typical architecture structure of Long-Short-Term-Memory

The main improvement of this network model is the **additional filtering system** of input data through its gates. Each gate operates in the following manner:

- The **forget gate** decides which information from the previous state should be discarded, allowing the **LSTM** to **filter out irrelevant past information.**
- The **input gate** determines which **new information** is important to add to the memory cell.
- The **output gate** controls the exposure of the internal memory content to the next step in the sequence.

PERFORMANCE RESULTS & EMPIRICAL EVIDENCE

The **LSTM** achieved a Mean Absolute Error **(MAE)** of **21.968** on the test set when utilizing the closing prices from the **last 5 days** as input. Conversely, employing the closing prices from the preceding **20 days** yields an **MAE** of **26.65** on the test set.

The result shows an important improvement with respect to **ANN**, this improvement is explained by the architecture structure and the **filtering mechanism** described above. Like the **ANN** the **LSTM** performs better with a **lower number of lag days.**

The following visualizations show the predicted values compared to the actual values for these two different time spans.

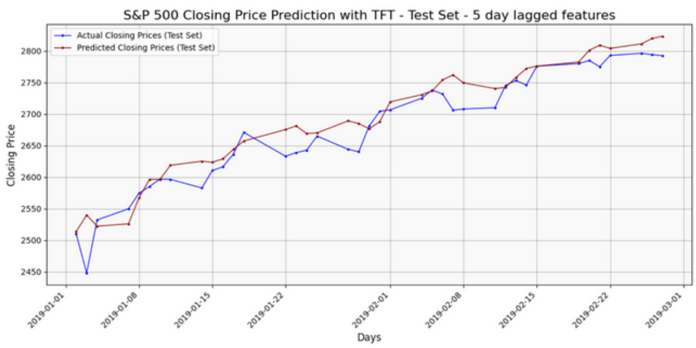


Fig.5: LTSM Prediction vs. Actual Values (Input: Previous 5 Days)

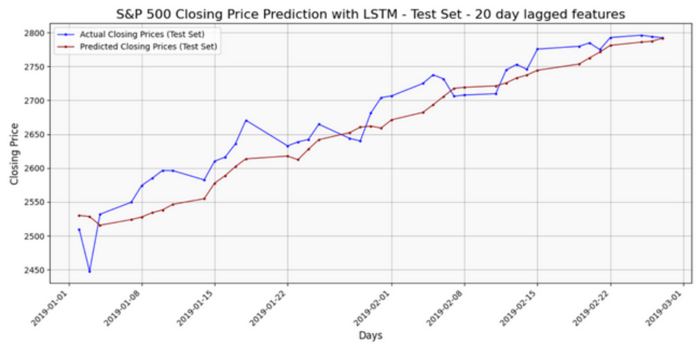


Fig.6: LTSM Prediction vs. Actual Values (Input: Previous 20 Days)

05. TEMPORAL FUSION TRANSFORMER (TFT) MODEL

ARCHITECTURE STRUCTURE

The **Temporal Fusion Transformer (TFT)** is an advanced deep learning model designed for time series forecasting.

The core improvement with respect to **ANN** and **LSTM** is its ability to assign **different weights to the input sequence**. This feature allows the model to understand which of the input values is more important to predict the closing price.

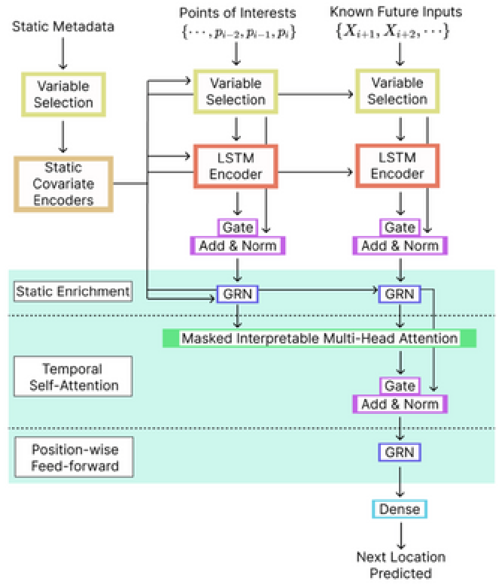


Fig. 7 typical architecture structure of Time series Fusion Transformer

This ability relies on the **Multi-Head Attention mechanism**. This peculiar feature makes the model understand by itself where to focus during the training time by minimizing the selected loss function.

We will see in the result section of this model that this is the **only model that performs better than the others when the number of input lagged days is higher**. This is one of the most important empirical evidences found in this study. The model does not process the input values uniformly, but gives more weight to the ones more important to enhance the prediction performance.

The attention mechanisms are inside the **Transformer Encoder Block**. After this layer a **Dropout** layer is applied to prevent overfitting. Then a Feedforward Layer is employed to further extract and refine features. This layer consists of two **TimeDistributed Dense layers** with **128** and **64** neurons, respectively, and **Rectified Linear Unit (ReLU)** activation functions.

The final output layer produces the time series predictions. To avoid overfitting during training, an **early stopping** mechanism is implemented with a patience of 25 epochs.

The model is based on a total of **10198** parameters.

This model has also the ability to combine time-varying inputs with static features. In our study we didn't add static metadata for the lack of computational resources. This can improve the predictive performance of the model. The holistic structure contributes to **TFT**'s robustness in capturing a broader range of relevant information.

PERFORMANCE RESULTS & EMPIRICAL EVIDENCE

The **TFT** model achieved a **MAE** of **21.153** on the test set with input data from the last **5 days**.

When utilizing closing prices from the preceding **20 days**, the **TFT** demonstrated enhanced accuracy with an even lower **MAE** of **19.708** on the test set. This shows in practice that the model is able to understand by itself what features are important, so adding more lagged value can only enhance the performance.

This is the opposite of what we observed in **ANN** and **LSTM** where the performance decreased when the time horizon was composed of twenty lagged values.

The visualizations further illustrate the TFT's ability to accurately predict S&P 500 closing prices across different time spans.

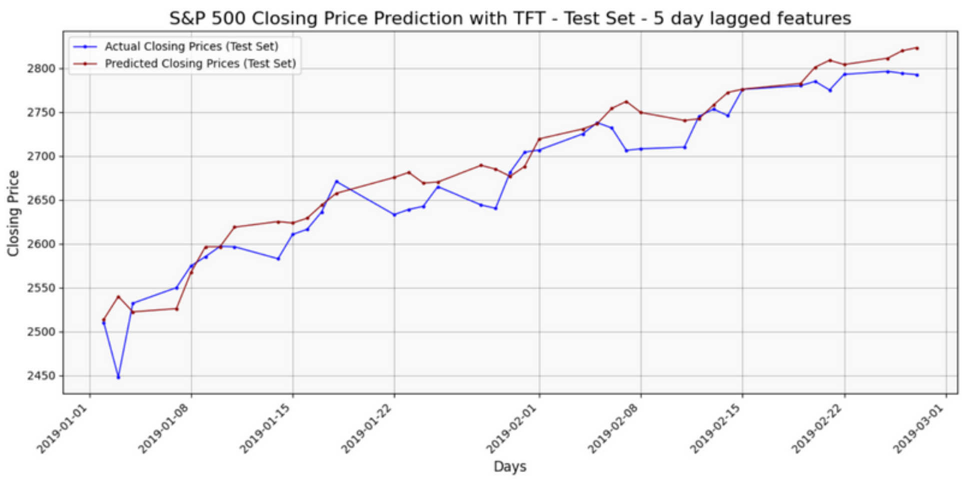


Fig.8: TFT Prediction vs. Actual Values (Input: Previous 5 Days)

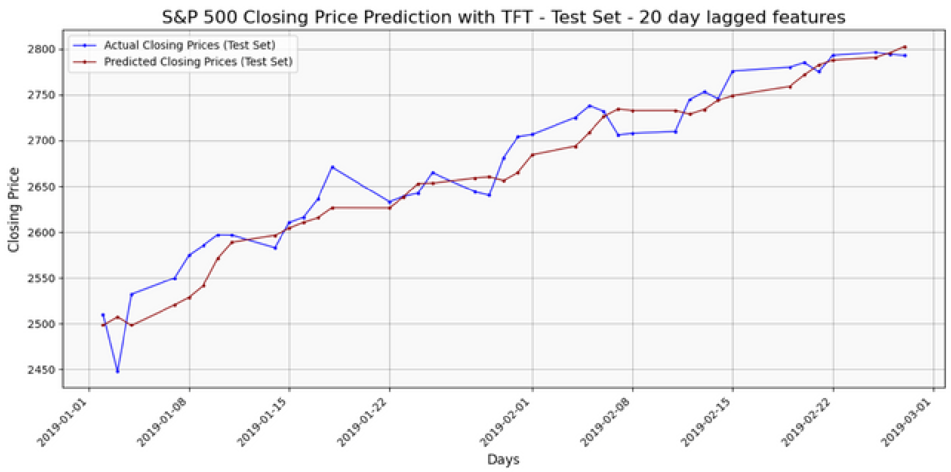


Fig.9: TFT Prediction vs. Actual Values (Input: Previous 20 Days)

06. CONCLUSIONS

In conclusion the main key points of our study are the following:

- The **Long Short-Term Memory (LSTM)** model outperforms the **Artificial Neural Network (ANN)** by leveraging its sequential memory and gating mechanisms.

- The **Temporal Fusion Transformer (TFT)** demonstrates superiority over the **ANN** and **LSTM** thanks to its advanced architecture and the unique **Multi-Head Attention** mechanism. **TFT's** ability to assign varying weights to input sequences allows it to focus on crucial features, improving predictive accuracy.

- **LSTM** and **ANN** prefer short-term predictions **TFT** showcases versatility and enhanced accuracy in **longer time horizons**, such as the preceding 20 days.

The obtained results are summarized in the following table:

Model	MAE Value (5 days)	MAE Value (20 days)
ANN	24.028	30.8133
LSTM	21.9682	26.6514
TFT	21.1534	19.7081

Fig.10: Recap of the model performance over different time horizons

PROJECT

07. BIBLIOGRAPHY

X. Hu, «Stock price prediction based on temporal Fusion Transformer», 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), dic. 2021, doi: 10.1109/mlbdbi54094.2021.00019.

M. Vijh, D. Chandola, V. A. Tikkiwal, y A. Kumar, «Stock closing price prediction using machine learning techniques», Procedia Computer Science, vol. 167, pp. 599-606, ene. 2020, doi: 10.1016/j.procs.2020.03.326.

S. Gu, B. Kelly, y D. Xiu, Empirical asset pricing via machine learning. 2018.

J. R. Coakley y C. E. Brown, «Artificial Neural Networks in Accounting and Finance: Modeling issues», Intelligent Systems in Accounting, Finance and Management, vol. 9, n.o 2, pp. 119-144, ene. 2000, doi: 10.1002/1099-1174(200006)9:2.

Q. Wu, J. Wang, Y. He, Y. Liu, y Q. Jiang, «Quantitative Assessment and Mitigation Strategies of greenhouse gas emissions from rice fields in China: a data-driven approach based on machine learning and statistical modeling», Computers and Electronics in Agriculture, vol. 210, p. 107929, jul. 2023, doi: 10.1016/j.compag.2023.107929.