

**FACULTY OF CYBERNETICS, STATISTICS AND INFORMATICS**

**APPLIED STATISTICS AND DATA SCIENCE**

***PROJECT***

***Advanced Time Series &Forecasting***

Coordinator:

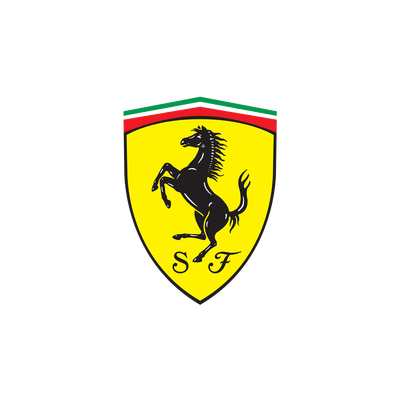
Univ. Prof. Dr. Pele Daniel Traian

Master Student:

Dorobanțu Diana-Mihaela

Bucharest

2022

*****Ferrari N.V.***

***Stock Price Forecast***

***CONTENTS***

1. ***Introduction***
2. ***Dataset Description***
3. ***LSTM Prediction Model***
4. ***Traditional Statistical Tests***
5. ***Bibliography***

***Introduction***

*The most widely studied financial application area is forecasting of a given financial time series, particularly* ***asset price forecasting****. Even though some variations exist, the main focus is on predicting the next movement of the underlying asset. More than half of the existing implementations of DL are focused on this area. Even though there are several subtopics of this general problem, including stock price forecasting, index prediction, forex price prediction, commodity (oil, gold, etc.) price prediction, bond price forecasting, volatility forecasting, cryptocurrency price forecasting, the underlying dynamics are the same in all of these applications.*

*Studies can also be clustered into two main groups based on their expected outputs:* ***price prediction*** *and* ***price movement prediction*** *(trend). Although price forecasting is essentially a* ***regression problem****, in most financial time series forecasting applications, correct price prediction of the price is not perceived to be as important as correctly identifying the directional movement. As a result, researchers consider trend prediction, i.e., forecasting which way the price will change, a more crucial study area compared with exact price prediction. In that sense, trend prediction becomes a* ***classification problem****. In some studies, only up or down movements are taken into consideration (2-class problem), although 3-class problems also exist (up, down, or neutral movements) (Sezer, 2020).*

***Dataset Description***

*In this analysis we use Ferrari N.V.’s stock price data from the period of 2017-2022. More specifically, the data contains stock prices from 8th May 2017 until 8th May 2022. That is a 5 year stock price data. This will act as a training data that will be used to train the LSTM on, whereas the stock price data from 9th May 2022 until 8th June 2022 will be used as a test data. The dataset contains the following data fields incorporated in Table 1. In this analysis we will use the* ***Adjusted Close Prices****, which represents the prices reported at the end of the trading day adjusted for splits and dividends.*

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Date | The date of the reported stock price |
| Open | The stock price at the beginning of the financial day |
| High | The highest price that the stock reached at that date |
| Low | The lowest price that the stock reached at that day |
| Close | The stock price at the end of the financial day |
| Adj Closed | Adjusted close price adjusted for splits and dividend and/or capital gain distributions |
| Volume | The number of shares transacted that day |

*Table 1: Dataset Variables*



*Figure 1: Ferrari N.V. Adjusted Close Price Variable 2017-2022*

***LSTM Prediction Model***

*LSTM and its variations along with some hybrid models dominate the financial time series forecasting domain. LSTM, by its nature, utilizes the temporal characteristics of any time series signal; hence, forecasting financial time series is a well-studied and successful implementation of LSTM. However, some researchers prefer to either extract appropriate features from the time series or transform the time series such that the resulting financial data become stationary from a temporal perspective, meaning even if we shuffle the data order, we will still be able to properly train the model and achieve successful out-of-sample test performance. For those implementations, CNN and Deep Feedforward Neural Network (DFNN) are the most commonly chosen DL models (Sezer, 2020).*

***Traditional Statistical Tests***

*From the graph above we can see that the….* Metoda STL decomposition a returnat analiza componentelor seriei, precum și o nouă serie fără trend. În figura de mai sus se pot observa graficele pentru fiecare componentă a seriei: trend, componentă sezonieră.*.*



*Figure 1: STL Decomposition*

*We check next if the time series is stationary with ADF test.*



*P < 0.05 => TS is not stationary*

*Hegy Test for Seasonality Testing*

Graphical user interface

Description automatically generated

*Difference*



* After checking the ADF results, we can confirm that our variable is **nonstationary** in levels, but **stationary in differences**

**Estimation of the Mean Equation**

**ARIMA (p,d,q)**

**Table

Description automatically generated**



**Step 2: Check existence of ARCH effects**

**1lag**





Graphical user interface, text

Description automatically generated

Text

Description automatically generated



Table

Description automatically generated

ARCH 2 lags



Table

Description automatically generated with low confidence



ARCH Model FSP = 0.000564 -0.152010\*FSP t-1 +0.162992εt-1 + εt

Variance ht=0.000207 +0.324555\*h^2t-1- +0.141776\*h^2t-2

**Model Diagnostics**



Since p >0.05 => there is no heteroskedasticity





Since p>0.05 for every lag => there is no serial correlation

(this residuals are WN)



GARCH



***Bibliography***