Machine Learning for Exchange Traded Fund Price Predictions

JONAS TER HORST, University of Twente, The Netherlands

The financial industry has long made use of machine learning techniques to predict stock prices or price trends. This research paper will focus specifically on exchange traded funds (ETFs) and the usage of machine learning techniques, like feed-forward neural networks, to predict their prices. A machine learning model will be developed to predict if a specific ETF price will go up or down and then predict the actual price. The model is a long short-term memory (LSTM) network, which is a type of feed forward neural network that is often applied to sequence prediction problems, like stock market forecasts. This research is new and relevant to the field because data and information from other ETFs are used to predict ETF prices and comes to the conclusion that there is not one universal set of ETFs that can be included to achieve better forecasting results. There are business sector-specific differences in the accuracy of the results. The results of this model can be used in a variety of different fields, like consultancy firms, but also by private investors.

Additional Key Words and Phrases: Exchange Traded Funds, Machine Learning, Neural Networks, LSTM

1 INTRODUCTION

Both exchange traded funds (ETFs) and machine learning technology constantly become more relevant in the world of finance. It is long known that machine learning techniques can predict and forecast stock returns and many more variables and factors. These predictions heavily influence the decisions of financial traders. [1] However, because of the non-linearity of the stock market, the wide range of features and variables that influence the price of any stock or ETF, which could be political, economic, or natural, and the complex relationships between all of these factors, predicting any value regarding individual stocks, bonds, or ETFs is complex.[1]

An ETF is a pooled investment, which typically tracks one specific business sector. Because a variety of stocks can be included in such an ETF, it decreases investors' risk and diversifies their portfolios. ETFs have grown in popularity over the last few decades. As of 2019, the total global amount of investment in ETFs is at about $4 trillion. [2] Many investors see trading of ETFs as more convenient than trading traditional equity mutual funds. There are multiple reasons for this. For instance, ETFs can always be traded and their prices fluctuate constantly, whereas equity mutual funds can only be traded at the end of a trading day when their prices are determined. Also, ETFs offer low expense ratios since most ETFs track some underlying market index, like the S&P 500, and the ETF structure results in low trading costs. [2]

Machine learning can be considered a buzzword. There are a lot of different methods, techniques, and algorithms that fall under the term 'machine learning'. This paper will use and analyze one machine learning technique to predict the development of ETF prices. In particular, the usage of a Long Short-Term Memory (LSTM) neural network for price predictions or trend analysis will be analyzed. A LSTM is a type of feed-forward neural network.

The goal of this research paper is to find the relation between ETF data and information of different ETFs and business sectors, during price or trend predictions, using the aforementioned machine learning technique. More about the specific research questions can be found in the 'Research Questions'- and 'Methodology' chapters.

2 EXISTING LITERATURE

The existing literature will focus on 4 specific topics. The topics are (1) History of prediction algorithms and methods, (2) algorithms / methods / machine learning techniques for price predictions, (3) Relevant data / features for prediction algorithms, and (4) Price predictability of index mutual funds and ETFs.

2.1 Historical Development of Prediction Algorithms and Methods

Predicting stock prices, especially of large market indices, has been done for a lot of years already. In 1996, Wittkemper and Steiner were one of the first researchers, that compared different forecasting methods to predict the systematic risk, also called beta, for a variety of German stocks. [3] They compared classic statical methods, with, at that time, modern artificial neural network (ANN) methods and concluded that the performance of the individual models heavily depends on the selection of variables fed into the models. [3] The analysis of Atsalakis and Valavanis (2009) even included a research paper about forecasting the Tokyo stock exchange, using ANNs, in the year 1990. [4]

In 2005, Klassen states, that before computers had enough computing power, the predictions were calculated based on standard statistical methods. [5] However, at that time, computers had enough computing power, and Klassen thus used the Levenberg-Marquardt Algorithm, a neural network using back propagation, to predict the prices of the NASDAQ- and the Dow stock index. [5] In 2009, the technology was already advanced enough to analyze all kinds of different techniques and identify the most powerful input variables/features for the different methods. Atsalakis and Valavanis conducted this analysis by reviewing over 100 scientific articles. An overview of the surveyed stock markets, the input variable choices, the modeling techniques, and the performance measures was created. [4] This research paper provided evidence that artificial neural networks are suitable for stock market forecasting, and soft computing techniques would often outperform conventional models. [4]

2.2 Closely Related Topics

There already exist a lot of research papers about the predictability of the general stock market, as well as about the predictability of individual stocks or market indices / ETFs.

TScIT 37, July 8, 2022, Enschede, The Netherlands

For instance, Hajek et al., predicted stock prices of U.S. firms using neural networks and support vector machines [6], Andrés et al. used deep learning neural networks to predict market indices for developing and emerging economies [7], and Ciner used the random forest method to forecast individual market index returns.[8]

However, this research is specifically focused on the predictability of ETFs and the relevant variables/features. Baek et al. were tackling the research question on how to efficiently predict and detect momentum patterns of ETF asset prices.[2] For this purpose, a predictive support vector machine (SVM) was developed and consisted of three risk factors; (1) systematic risk factors, (2) credit risk factors, and (3) market fear factor, which were extracted from 16 financial risk indicators by principal component analysis (PCA). [2]

Malinda and Chen used the Grey Relational Analysis (GRA) and ANNs to predict the volatility of consumer ETF returns.[9] They identified four main variables affecting consumer ETFs. These variables are (1) the NYSE Composite Index, (2) the CRB Index, (3) the USD/EUR Exchange Rates, and (4) the Put-Call Ratio (PCR). [9] Additionally, Malinda and Chen identified the Back Propagation Neural Network (BPNN) and the Recurrent Neural Network (RNN) to be the most consistent and precise models. [9]

Sismanoglu et al. identified the long short-term memory (LSTM) neural network to be the most successful recurrent neural network architecture. [10] As for their features, they only used date, opening price, closing price, highest price per day, and lowest price per day. [10] Ahmed et al. executed a very similar research project. They also use a LSTM model, with the same features. In this case, the to-be-predicted ETF was the S&P500. To train and test the data, they included daily feature values from 2010 until 2021. [11] Vijh et al. used the same set of features, and additionally the 3-, 7-, 14-, and 21-day moving average of the individual stocks that were to be predicted. [12] Niaki and Hoseinzade were forecasting the S&P500 using artificial neural networks. They mention a big advantage for ANNs is the fact that they recognize nonlinear relations between the individual features given to the model and the results, or the output. [13] Also, ANNs can generalize and do not need assumptions on the distribution of the given data, statistical methods do. [13] This is especially relevant in the case of the finance sector since ETF prices are not linear and do not have a standard distribution. However, Niaki and Hoseinzade also stated some negative arguments regarding ANNs. Individual factors, like learning rate or the number of layers and nodes, are difficult to determine, but heavily affect the results. Also, identifying relevant features might be complicated, as well as the fact that a great volume of data is needed to create an accurate model that produces acceptable predictions. [13] The results of the Asian stock market prediction of Chou confirmed the statement that Neural Networks are the most accurate prediction method. Chou compared naive Bayes, a decision tree, and a neural network. With a value of 0.46, the neural network performed the best accuracy-wise. [14]

Gondkar et al. were trying to identify the most effective neural network for stock price predictions and stated that the conventional LSTM neural network works best for sequential data, like stock price predictions. [15] To be more specific, the 1D-Conv-LSTM and the GRU-LSTM train and converge the quickest. [15] They include more details of the exact model parameters that were used to predict the individual stock. For instance, a window size of 20 resulted in the most accurate predictions for two specific Indian stocks, however, the window size of 60 was the most accurate for another Indian stock. [15] This shows that the parameters of the model heavily influence the accuracy of the outcome.

All of this related work shows the diversity in techniques that can be used to predict stock prices. It is not in the scope of this research project to focus on more than one, however, it is important to note, that multiple techniques are already proven to be effective when it comes to stock price predictions.

2.3 Research Gap

A lot of research in this field has already been done for individual research questions and topics, for instance, the research of Hajek et al. [6], Andrés et al. [7], Ciner [8], Baek et al. [2], and Malinda & Chen [9] is related to the predictability of index mutual funds or ETFs specifically. Pyo et al. [16], Zhao et al. [17], Ahn et al. [18], and Rodríguez-Gonzàles et al. [19], all provide relevant information about the kind of algorithms that are most efficient during price predictions of ETFs and market index funds. Sismanoglu et al. [10], Niaki & Hoseinzade [13], and Aloud [20] all identified some features and variables for accurate forecasts and predictions.

Even though a lot of different research regarding price prediction and predictability of stock market index funds and ETFs has already been done, there is still a research gap, that this research paper will solve. The question is whether including the data of other ETFs can produce more accurate price predictions and forecasts than data of only the to-be-researched ETF.

3 RESEARCH QUESTIONS

3.1 RQ 1: How does data and information from other ETFs influence the accuracy of ETF price predictions using machine learning algorithms?

The goal of this research question is to find out if price predictions of ETFs become more precise or accurate when you include data and information from other ETFs. This question can then also be divided into two subquestions, one considering data and information of other ETFs which are in the same sector, and one considering data and information for other ETFs which are more general, and do not focus on one specific sector (see subquestions 1 and 2). A sector could for instance be the tech sector, with stocks like Amazon (AMZN), Apple (AAPL), Facebook (FB), or Alphabet (GOOG), and ETFs like the Vanguard Information Technology ETF (VGT), the ARK Innovation ETF (ARKK), or the State Street Technology Select Sector SPDR Fund (XLK).

3.1.1 Subquestion 1: How does data from other ETFs, but the same general sector, influence the accuracy of ETF price predictions using machine learning algorithms?

3.1.2 Subquestion 2: How does data from other ETFs of different sectors influence the accuracy of ETF price predictions using machine learning algorithms?

3.2 RQ2: Which sectors entail the most predictable ETFs?

Considering the individual sectors of all the analyzed ETFs, the goal of this research question is to identify which sector contains the most predictable ETFs. This research question has already been partly answered in other research, however, the features used to produce the outcome will be new, since other ETFs data and information will be integrated.

4 METHODOLOGY

The methodology can be divided into two sections: (1) the theoretical methodology for the research paper and (2) the technical methodology for the data preparation and analysis.

4.1 Theoretical Methodology

The goal of this research is to find relevant information in four different topic areas, namely, (1) history of prediction algorithms, (2) price predictability of index mutual funds and ETFs, (3) algorithms / methods for price prediction, and (4) relevant variables / features for machine learning algorithms in the topic of stock market analysis.

After the literature research, relevant sectors and ETFs will be selected for analysis. The selection will be based on a few criteria. First, a diverse selection of business sectors will be included. These sectors could for instance be tech, finance, real estate, and food. Since the research is based on different business sectors, the ETFs that are to-be-predicted should all come from a similar source, this will reduce the bias of the outcome. However, the additional ETFs will also include at least one international ETF, that is not focused on one specific country.

After the sectors and ETFs have been found, relevant features/variables (forth on they will be called features) will be identified. This will be done using the results of the literature research. The features will be the same for all of the different sets of ETFs and all of the business sectors.

Once relevant sectors, ETFs, and features have been chosen, the data source needs to be selected. As a primary data source, Yahoo Finance will be used since this website offers the functionality to download different datasets for most of the relevant ETFs.

4.2 Technical Methodology

Before explaining the concrete methodology of the technical part, some relevant information need to be included. First of all, the programming language for the analysis will be Python. The reason for this is the diverse selection of free libraries and frameworks. Depending on the purpose, this analysis will use TensorFlow, NumPy, Pandas, Matplotlib, Keras, and SciKit-Learn. All of these already existing libraries provide functionalities for data gathering, data analysis, neural network creation, and data visualization.

The whole development process is done in Google Collaboratory, or Colab, which is a Google Research product, allowing to create Jupyter Notebooks. The main advantage is the free usage of external GPU power. This is a reason, why the Google Colab environment is often used for machine learning processes since the usage of external GPUs enhances the model speed and decreases the requirements of the researcher's hardware.

The technical part will start by importing all the relevant data to the workspace. This is being done using the API of Yahoo Finance, from which data can be imported within the script. There is no need to locally store the data.

The imported data will then be scaled for performance reasons. All of the individual features will be normalized. How this is being done, and which formula is being used will be explained in the "Data" chapter of this paper. After the data has been scaled, it will be divided into training-, validation- and testing data. The split will be about 80-10-10, with 80% training-, 10% validation-, and 10% testing-data. With this data, different models can be executed. The models will be thought using the training-, and validation data, and then tested on the testing data, which it has never seen before. One very important aspect of splitting the dataset is keeping in mind these experiments are based on sequential data. This means, the program cannot randomly select 80% as training data, 10% as validation data, and 10% as testing data, but it has to use the first 80% of the given 10 years as training data (2009 trading days), the next 10% of the given 10 years as validation data (251 trading days), and the last 10% as testing data (251 trading days).

There will be a total of 15 individual experiments conducted. For each of the five selected business sectors, there will be three models created. (1) Using only the data of the to-be-predicted ETF, (2) using the data of the to-be-predicted ETF and data of three external ETFs of the same business sector, and (3) using the data of the to be predicted ETF and data of three more general ETFs, that are not business sector specific.

Generally, the models makes two predictions. Firstly, the model will predict if the price of the ETF will rise or fall. The second prediction is a precise price prediction.

5 DATA

5.1 Data Selection

There are different choices to be made regarding the data involved during the individual stages of the model creation. First of all, five general business sectors need to be identified. Because bias should be reduced in this research, the business sectors should be divers and not much related to each other. The five business sectors will be (1) energy, (2) finance, (3) industry, (4) information technology, and (5) real estate.

After that, one specific ETF of each of the five business sectors needs to be selected to be the main research object, hence, the ETF, whose price is to be predicted. In addition to these 5 sector specific ETFs, 3 additional ETFs need to be selected for each business sector. The data of these additional ETFs will be given to the LSTM model to answer subquestion 1. Then, three more general, hence, not sector specific ETFs need to be selected. These will be given to the LSTM model to answer subquestion 2. There are a few criteria for the selected ETFs. First of all, each of the ETFs needs to be at least 10 years old, because the model is getting daily data from the past 10 years. To be exact, from 29.05.2012 until 31.05.2022. The complete list of selected ETFs can be found in Table 1, in the "Data Preparation" chapter.

Finally, the features that will be given to the LSTM model need to be selected. Because of the similarity of this research to the work of

Ahmed et al., the features will be the same as in their research. [11] Ahmet et al. also predicted an ETF over the timeframe of 11 years and used an LSTM neural network to make the predictions. [11] The features, or data entries, for each model will, be (1) opening price, (2) closing price, (3) highest price per day, (4) lowest price per day, and (5) total volume traded within the ETF.

5.2 Data Preparation

The individual model parameters will be explained in the 'Model Information' chapter. However, before anything can be added to the model, the data needs to be prepared, so that the model can perform optimally.

First of all, the individual ETFs need to be concatenated into 10 individual lists. Five lists containing the sector specific ETFs and the other ETFs of the same sector, and five other lists containing the same sector specific ETFs, but with the 3 general ETFs. These are the selected ETFs and the aforementioned lists:

Table 1. Selected ETFs and respective ETF sets

Nr	Main ETF	Extra 1	Extra 2	Extra 3
1	VDE	DBE	PXI	FAN
2	VFH	XLF	PSCF	IXG
3	VIS	RGI	XLI	EXI
4	VGT	RYT	XSW	IXN
5	VNQ	SCHH	REM	RWO
6	VDE	SPY	IVV	VTI
7	VFH	SPY	IVV	VTI
8	VIX	SPY	IVV	VTI
9	VGT	SPY	IVV	VTI
10	VNQ	SPY	IVV	VTI

After these lists are created, they need to be formatted for the model to use them. For this, it is important to understand on what basis the model is going to make price predictions. During the training process, the model will get the daily feature values for first 5 days. On this basis, the model tries to predict the opening price of the 6th day. After that, it gets the same data from the second until the 6th day and tries to predict the opening price of the 7th day. This continues until the end of the training dataset has been reached. The value to be predicted is always the opening price of the next day.

Finally, once the data has the correct format, all of the values need to be normalized. During data normalization, all of the values are scaled to be between 0 and 1, where 0 is the minimum of a specific column in an individual ETF for all of the timestamps, and 1 is the maximum respectively. The formula for normalization is as follows:

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

, where $\bar{x}$ is the normalized value, x is the current data point, and x_{min}&x_{max} are the minimum and maximum values of the data columns respectively.

5.3 Evaluation Metrics

For each of the 15 individual executed experiments, there are 4 values that are relevant for answering the aforementioned research-, and sub-questions. These indicators are the Root Mean Squared Error (RMSE) values of the train-, validation-, and test datasets, and the accuracy up-down accuracy.

The RSME is a commonly used measurement for evaluating the quality of predictions made by supervised learning applications. The RSME is applicable in this case since the data is normalized and scaled. This is also the reason why the value can be used to compare the performance of the different models. All of the prices are normalized to a value between 0 and 1, where 0 is the minimum, and 1 is the maximum of the respective feature over the respective timeframe. One downside of using the RSME as performance indicator is outliers heavily affect the value of the RSME. However, since the predictions are based on timeseries data, all of the values are relevant and based on the previous values. The RSME is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} ||y(i) - \bar{y}(i)||^2}{N}}$$

, where N is the number of data points, y(i) is the i-th datapoint/measurement, and $\bar{y}$ (i) is the corresponding prediction value.

The other value used to evaluate the performance of the different models is the up-down accuracy. Since the model predicts precise prices, it is possible to predict if the price goes up or down. So, by comparing the predicted price with the price of the day before, the model also predicts if the price rises or falls. This value can then again be compared to the actual prices and whether they rise or fall. The up-down accuracy is the percentage of correct predictions made by the model, in terms of whether the model correctly predicted the price to rise or fall. It is calculated as follows:

$$Up - DownAccuracy = \frac{x}{N}$$

, where x is the number of correct predictions in terms of if the price rises or falls, and N is the total number of predictions made.

6 MODEL INFORMATION

The model is based on a variety of different parameters. All of these parameters heavily affect the outcome and the accuracy of the results. The parameters used for the final results will now be presented. However, it is noteworthy, that these parameters were the best of many. During a trial-and-error process, these parameters resulted in the best models. Also, it is not in the scope of this research project to explain what all of these parameters do and how they affect the results.

In the beginning, the model is based on a 64-layer LSTM network, with 'relu' as activation function. Then an 8-layer Dense level is introduced, also with 'relu' as activation function. Finally, a 1-layer Dense level is introduced, with a 'linear' activation function. This is then used to predict precise prices.

The input shape depends on the experiment executed. If the model only uses the data of one single ETF, the input shape would be (5, 5), where the first five is the number of timestamps/entries, which would be one workweek, and the second 5 is the number of features passed to the model. These would be daily values for the opening price, closing price, highest price, lowest price, and volume traded. If one predicts based on the set of 4 ETFs, the input shape would be (5,20), where the first 5 is again the number of timestamps/entries and the 20 is the number of features, which are the aforementioned five features for every ETF of the set.

The learning rate is set to be 0.0001, the optimizer is 'Adam', the loss is computed by the MeanSquaredError() function, and the model executes each experiment for 100 epochs. Also, the save_-best_only function is set to true, which means that only the best models are saved. Thus, if the best-performing model was reached after 60 epochs, the next ones will not be saved. Finally, shuffle is set false, because the model uses sequential data that cannot be shuffled to train the model randomly.

7 RESULTS

The results of the explained experiments will now be presented.

Table 2 includes the results of the15 individual experiments. The experiments were, as already stated, divided into 5 business sectors and three sets of ETFs for each business sector.

In addition to table 2, all of the results were also plotted and visualized for better understanding.

Now, to answer all of the relevant research questions and sub-questions, one can use the results of table 2 and of the different visuals.

7.1 Subquestion 1

Table 2 shows, that the training-RMSE value always performs better than the testing-RMSE value, of the 'ETF set (same sector)' rows. This is to be expected since the testing values are completely new to the model, however, within the RMSE values of the testing set, one can see large differences in performance. For instance, the testing-RMSE for the information technology sector is by far the highest, with a value of 48.61, whereas the same value for the real estate sector is 3.2. This would already indicate that the performance of the model is heavily based on the individual sectors.

Now, the focus will be on the testing-RMSE and the up-down accuracy since both of these values indicate the performance of a dataset that the model has not seen before. Considering the testing-RMSE values, the model has always performed better with additional data of ETFs of the same sector, in comparison to only data of the to be predicted ETF, except for the energy sector, where the RMSE values for the testing data of only the ETF and the set of ETFs of the same sector were 5.29 and 13.56 respectively. The best overall performance is in the real estate sector, where the model could achieve a testing-RMSE value of 3.2 with the set of ETFs of the same sector.

Regarding the up-down accuracy, the models, again, performed differently within each sector. For the energy-, industry-, and real estate sector, the up-down accuracy was better with the data of the ETFs within the same sector. For the finance-, and information technology sectors, the up-down accuracy was better with only the data of the to-be-predicted ETF. However, these accuracies are generally pretty close to each other since there is never a difference of more than 2%.

7.2 Subquestion 2

To answer this subquestion, one can first of all again compare the training-RMSE- and the testing-RMSE values of the individual sectors. By doing so, one can see that the training-RMSE values are always better than the testing-RMSE values, with an exception in the industry sector, where the testing data performed better, than the training data.

Now, to answer this Subquestion, the testing-RMSE values and the up-down accuracy values of the model with only the data of the to-be-predicted ETF and of the model with the data of the ETF and of the more general ETFs need to be compared. The testing-RMSE values for the finance-, information technology-, and industry sectors are better with the additional data of the general ETFs. For the energy- and real estate sector, the model performed better without the additional data. The biggest differences can be found in the energy-, and industry sectors, where the differences within the RMSE values are +6.79 and -10.91 respectively.

The up-down accuracy is in all cases better if the additional general ETFs are also given to the model, except for the energy sector. For the industry sector, the difference in this value is the biggest, at 5%.

7.3 Research Question 1

To answer the first research question, one has to take the answers to Subquestions 1 and 2 into account. The first major argument is the fact that the results not only differ from Subquestion 1 and 2, so if additional data of the same sector or more general sectors has been added, but they also differ from business sector to business sector. Even without the performance indicators, hence the RMSE-, and up-down accuracy values, one can simply see this by comparing the different visualizations of the model performances. The best performance out of all of the 15 experiments was within the real estate sector, with data of ETFs of the same sector. The testing-RMSE was a 3.2. The second and third best performances were in the finance and energy sectors, with testing-RMSE values of 3.3 and 5.29 respectively. However, in the finance sector, the model with data of the more general ETFs performed best, and in the energy sector the model with only the data of the to-be-predicted ETF performed best. This means that the top three performances, based on the testing-RMSE value, of all the conducted experiments, are in different sectors and have different datatypes included in the training of the model.

To come back to the original research question, this means that data of other ETFs does influence the accuracy of ETF price predictions using machine learning models, however, it leads to different outcomes for different business sectors.

Next to this, it is also notable that the information technology sector performed by far the worst for all three experiment variations conducted. The training-RMSE values were already higher than for the other sectors, but the testing-RMSE values are around 50. This means that the best testing-RMSE value of the information technology sector is still 32.85 higher than the worst testing-RMSE

Table 2. Selected ETFs and respective ETF sets

ETF sets and sectors	Training-RMSE	Validation-RMSE	Testing-RMSE	Up- Down-Accuracy
Energy				
Only ETF	5.15	5.14	5.29	0.508
ETF set (same sector)	2.94	13.42	13.56	0.524
ETF set (general)	3.56	2.7	12.08	0.464
Finance				
Only ETF	1.5	2.3	5.7	0.496
ETF set (same sector)	2.28	3.0	3.66	0.492
ETF set (general)	1.5	3.0	3.3	0.524
Industry				
Only ETF	3.72	7.28	15.76	0.456
ETF set (same sector)	4.25	4.24	12.67	0.476
ETF set (general)	9.47	5.79	4.85	0.516
Information Technology				
Only ETF	5.78	22.29	60.21	0.472
ETF set (same sector)	7.43	22.69	48.61	0.456
ETF set (general)	7.68	24.47	58.64	0.504
Real Estate				
Only ETF	2.19	2.01	5.97	0.5
ETF set (same sector)	2.03	4.96	3.2	0.508
ETF set (general)	2.21	2.85	7.76	0.524

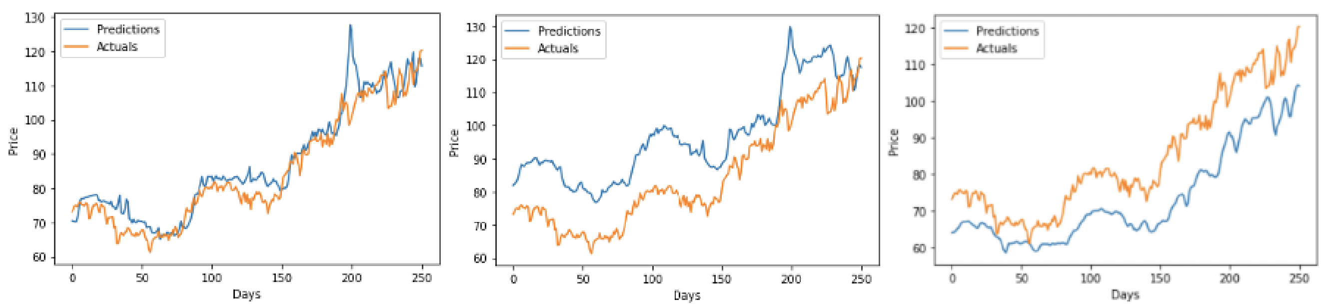

Fig. 1. Plotted results of the Energy sector (only ETF / ETF + ETFs same sector / ETF + ETFs general sector)

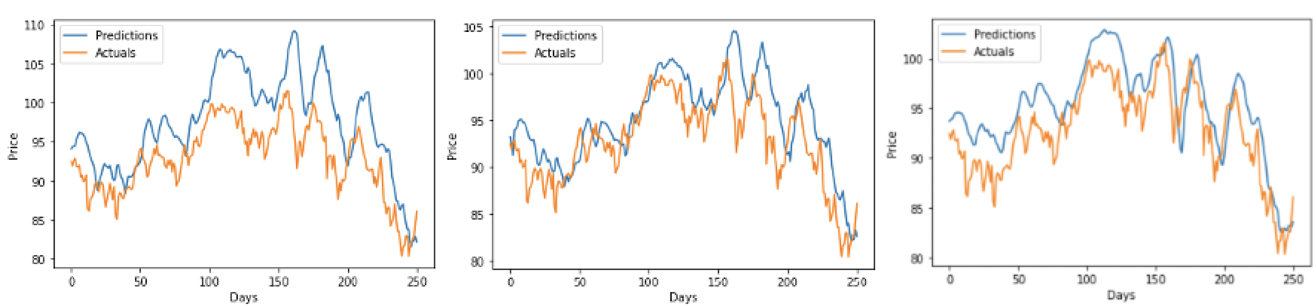

Fig. 2. Plotted results of the Finance sector (only ETF / ETF + ETFs same sector / ETF + ETFs general sector)

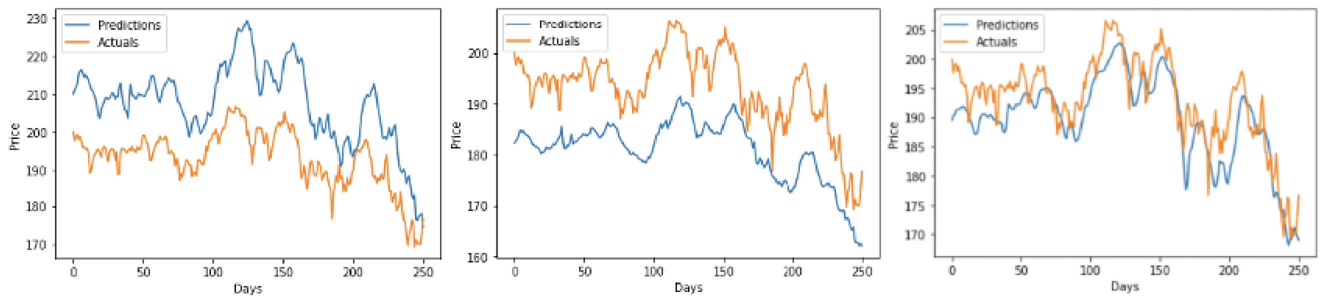

Fig. 3. Plotted results of the Industry sector (only ETF / ETF + ETFs same sector / ETF + ETFs general sector)

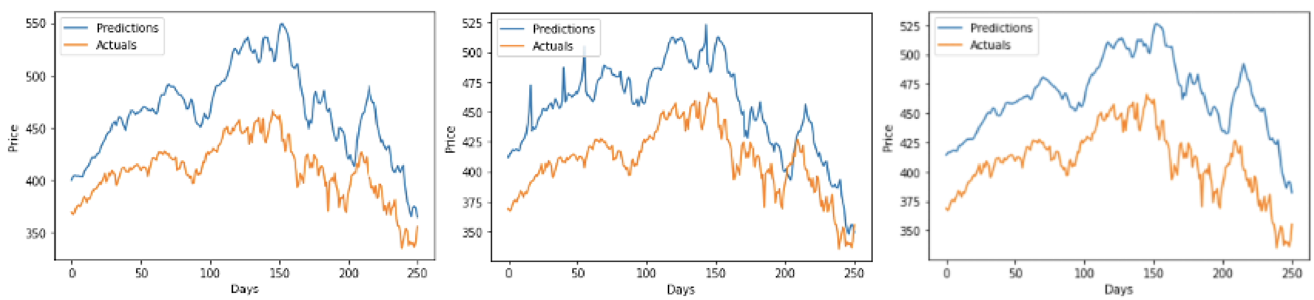

Fig. 4. Plotted results of the Information Technology sector (only ETF / ETF + ETFs same sector / ETF + ETFs general sector)

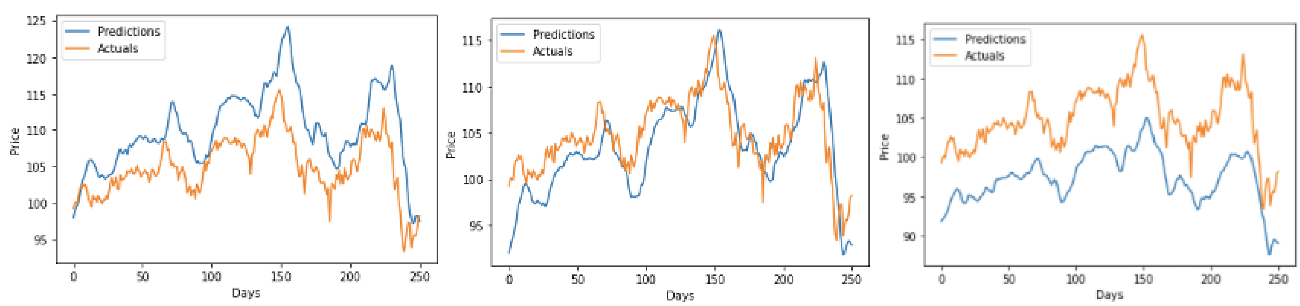

Fig. 5. Plotted results of the Real Estate sector (only ETF / ETF + ETFs same sector / ETF + ETFs general sector)

of all other sectors, which is the model using only the to be predicted ETF of the industry sector. Additionally, the information technology sector also has the lowest up-down accuracy value together with the industry sector.

Now, to analyze the broader picture, it is important to note, again considering the testing-RMSE values, that including the data of other ETFs of the same sector is the best solution for the information technology-, and real estate sector, including data from other, more general ETFs is the best solution for the finance-, and industry sector and only including data of the to-be-predicted ETF is the best solution for the energy sector. Hereby it is important to note that these results were only the results of the aforementioned set of ETFs, it was not tested using other ETFs of the same/more general sectors.

One final interesting observation is the fact that the exact price predictions may be more or less correct from model to model, however the average up-down accuracy values represent the same probability as flipping a coin.

7.4 Research Question 2

Keeping in mind that the results of the experiments and the previous research questions show differences from sector to sector, it is safe to say that not one model variation outperformed the other.

What can easily be seen is the information technology sector performed by far the worst, thus, using the parameter used in this model is the least predictable business sector.

Table 3 shows the different averages of the three conducted experiments per business sector for the testing-RMSE and the up-down accuracy values. This table shows that the two best averages of the testing-RMSE values belong to the finance sector and the real estate sector, the same goes for the average up-down accuracy values.

Table 3. Sector specific averages of the testing-RMSE & the up-down accuracy

Business Sectors	Avg. Testing-RMSE	Avg. up-down
Energy	10.31	0.499
Finance	4.22	0.504
Industry	11.09	0.483
IT	55.82	0.477
Real Estate	5.64	0.511

One needs to consider that these averages can also be misleading because one model might have performed very well in comparison to other the other two models, but then the average is still low due to the other two values. However, looking at the results presented in table 2, it is easy to see that the two best performing models considering the testing-RMSE values also belong to the finance-, and real estate sectors. The same holds for the up-down accuracy values, with the exception, that one model of the energy sector also performed as well as the models of the finance-, and the real estate sectors. Here, the best three values were all 0.524.

8 DISCUSSION

There are a lot of different parameters and information that created the final results. Changing some of these parameters into different values might lead to different results. Some of the aspects that might lead to different results will now be discussed.

First of all, maybe the most important parameters chosen, are the to-be-predicted ETFs. All results were considering opening prices of the aforementioned set of five ETFs of the five business sectors. Trying to predict other ETFs might lead to completely different results, which could be more or less accurate than the presented results. However, the choice of the five selected Vanguard ETFs has already been described and reasoned for.

Also, the selected ETF sets can create a difference in the outcome of the experiments. In this case, 3 additional ETFs have been selected, but in reality, it might be more efficient to use more or less ETFs. The fact that all of the sets of ETFs of the same sector are including U.S.-based ETFs and international ETFs lower the bias, since the results are compared to each other in the end, however, the final goal should be to create the most efficient forecasts for each ETF. This might be done by using an entirely different set of ETFs, which might be more international, less international, or differs in other ways to the selected sets. Next to the ETFs, the features that were used as input for the data also heavily affects the accuracy of the result. Some researchers might argue that using returns instead of prices leads to more accurate predictions, however, during my literature research, I came across a lot of researchers, that also used prices and volumes as features for similar models. All of these can be found in the "Existing Literature" chapter of this paper.

Finally, the chosen model and parameters of the model also have an impact on the accuracy of the results. During the experiments a LSTM neural network was used, but no alternatives were executed due to the limited timeframe. It might be the fact that other algorithms or types of neural networks produce better results. However, the reason for choosing the LSTM neural network was, again, based on literature research and can be found in the "Existing Literature" chapter.

9 CONCLUSION

The conducted experiments lead to new insights into the process of predicting price trends and precise price prediction of Exchange Traded Funds. First of all, the most important takeaway is the fact that it is sector specific, whether including external ETFs to the input of a model leads to more accurate price predictions. There are some sectors, where the best model was achieved by only including the data of the to-be-predicted ETF, some sectors, where including data of other ETFs of the same sector has produced the best results, and some sectors, where including more general ETFs have produced the best results.

In terms of validation metrics, such as the Root Mean Squared Error, there are some fluctuations within the same sector, as well as in between the different sectors. The up-down accuracy, so the percentage of correct prediction regarding the fact if the price will go up or down, is between 45.6% and 52.4%.

So, regarding the stated research questions, the data and information of other ETFs does influence the price predictions of specific ETFs, however, there is no common set of ETFs that can universally be included in all models to achieve better predictions. It all heavily depends on the to-be-predicted ETF and the business sector of the ETF. However, during the conducted experiments, the finance sector had the best predictions, and the information technology sector had the worst, based on the evaluation metrics.

10 FUTURE WORK

There are multiple possibilities to continuing the field of study. The first one would be trying to use other ETFs and see if their predictions become more or less accurate in comparison to the results presented here.

Another possibility would be to introduce new features, sets of ETFs, and business sectors. There exist a huge number of ETFs that can be analyzed and predicted.

To evaluate the model, another continuation would be to trade based on the results of the model. This could be done by developing a bot that buys or sells ETFs based on the up-down predictions made by the model. To not take too much risk, one should start by trading in some virtual portfolio and then evaluate the performance of the model.

11 REFERENCES

[1] Q. Mingyue, L. Cheng, and S. Yu, "Application of the Artifical Neural Network in Predicting the Direction of Stock Market Index," in Proceedings - 2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2016, 2016, pp. 219–223. doi: 10.1109/CISIS.2016.115.

[2] S. Baek, K. Y. Lee, M. Uctum, and S. H. Oh, "Robo-advisors: Machine learning in trend-following ETF investments," Sustainability (Switzerland), vol. 12, no. 16, 2020, doi: 10.3390/SU12166399.

[3] H.-G. Wittkemper and M. Steiner, "Using neural networks to forecast the systematic risk of stocks," European Journal of Operational Research, vol. 90, no. 3, pp. 577–588, 1996, doi: 10.1016/0377-2217(95)00099-2.

[4] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques - Part II: Soft computing methods," Expert Systems with Applications, vol. 36, no. 3 PART 2, pp. 5932–5941, 2009, doi: 10.1016/j.eswa.2008.07.006.

[5] M. Klassen, "Investigation of some technical indexes in stock forecasting using neural networks," in Proceedings - WEC'05: 3rd World Enformatika Conference, 2005, vol. 5, pp. 75–79.

[6] P. Hájek, V. Olej, and R. Myšková, "Forecasting stock prices using sentiment information in annual reports - A neural network and support vector regression approach," WSEAS Transactions on Business and Economics, vol. 10, no. 4, pp. 293–305, 2013.

[7] A. R. Andrés, A. Otero, and V. H. Amavilah, "Using deep learning neural networks to predict the knowledge economy index for developing and emerging economies," Expert Systems with Applications, vol. 184, 2021, doi: 10.1016/j.eswa.2021.115514.

[8] C. Ciner, "Do industry returns predict the stock market? A reprise using the random forest," Quarterly Review of Economics and Finance, vol. 72, pp. 152–158, 2019, doi: 10.1016/j.qref.2018.11.001.

[9] M. Malinda and J.-H. Chen, "The forecasting of consumer exchange-traded funds (ETFs) via grey relational analysis (GRA) and artificial neural network (ANN)," Empirical Economics, vol. 62, no. 2, pp. 779–823, 2022, doi: 10.1007/s00181-021-02039-x.

[10] G. Sismanoglu, M. A. Onde, F. Kocer, and O. K. Sahingoz, "Deep learning based forecasting in stock market with big data analytics," 2019. doi: 10.1109/EBBT.2019.8741818.

[11] S. Ahmed, R. K. Chakrabortty, D. L. Essam, and W. Ding, "Poly-linear regression with augmented long short term memory neural network: Predicting time series data," Information Sciences, vol. 606, pp. 573–600, 2022, doi: 10.1016/j.ins.2022.05.078.

[12] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques," in Procedia Computer Science, 2020, vol. 167, pp. 599–606. doi: 10.1016/j.procs.2020.03.326.

[13] S. T. A. Niaki and S. Hoseinzade, "Forecasting S&P 500 index using artificial neural networks and design of experiments," Journal of Industrial Engineering International, vol. 9, no. 1, 2013, doi: 10.1186/2251-712X-9-1.

[14] T.-N. Chou, "Stock trends prediction using the feature pattern constructed with the panel data of Asian stock markets," in 2017 7th International Workshop on Computer Science and Engineering, WCSE 2017, 2017, pp. 1296–1301.

[15] A. Gondkar, J. Thukrul, R. Bang, S. Rakshe, and S. Sarode, "Stock Market Prediction and Portfolio Optimization," 2021. doi: 10.1109/GCAT52182.2021.9587659.

[16] S. Pyo, J. Lee, M. Cha, and H. Jang, "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets," PLoS ONE, vol. 12, no. 11, 2017, doi: 10.1371/journal.pone.0188107.

[17] J. Zhao, D. Zeng, S. Liang, H. Kang, and Q. Liu, "Prediction model for stock price trend based on recurrent neural network," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 1, pp. 745–753, 2021, doi: 10.1007/s12652-020-02057-0.

[18] J. J. Ahn, H. W. Byun, K. J. Oh, and T. Y. Kim, "Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting," Expert Systems with Applications, vol. 39, no. 9, pp. 8369–8379, 2012, doi: 10.1016/j.eswa.2012.01.183.

[19] A. Rodríguez-González, A. García-Crespo, R. Colomo-Palacios, F. Guldrís Iglesias, and J. M. Gómez-Berbís, "CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator," Expert Systems with Applications, vol. 38, no. 9, pp. 11489–11500, 2011, doi: 10.1016/j.eswa.2011.03.023.

[20] M. E. Aloud, "The role of attribute selection in Deep ANNs learning framework for high-frequency financial trading," Intelligent Systems in Accounting, Finance and Management, vol. 27, no. 2, pp. 43–54, 2020, doi: 10.1002/isaf.1466.