International Journal of Economics and Financial Research
ISSN(e): 2411-9407, ISSN(p): 2413-8533
Vol. 6, Issue. 7, pp: 170-179, 2020
URL: https://arpgweb.com/journal/journal/5
DOI: https://doi.org/10.32861/ijefr.67.170.179

Academic Research Publishing Group

Original Research **Open Access**

Predicting Intraday Prices in the Frontier Stock Market of Romania Using Machine Learning Algorithms

Dan Gabriel ANGHEL
Department of Money and Banking, The Bucharest University of Economic Studies – Bucharest, Romania
Email: dan.anghel@fin.ase.ro

Article History
Received: June 13, 2020
Revised: July 15, 2020
Accepted: July 24, 2020
Published: July 28, 2020

Abstract

This paper investigates if forecasting models based on Machine Learning (ML) Algorithms are capable to predict intraday prices in the small, frontier stock market of Romania. The results show that this is indeed the case. Moreover, the prediction accuracy of the various models improves as the forecasting horizon increases. Overall, ML forecasting models are superior to the passive buy and hold strategy, as well as to a naïve strategy that always predicts the last known price action will continue. However, we also show that this superior predictive ability cannot be converted into "abnormal", economically significant profits after considering transaction costs. This implies that intraday stock prices incorporate information within the accepted bounds of weak-form market efficiency, and cannot be "timed" even by sophisticated investors equipped with state of the art ML prediction models.

Keywords: Efficient market hypothesis; Intraday prices; Machine learning; Artificial intelligence; Trading strategies; Frontier stock market.

1. Introduction

The topic of market efficiency is one of the most important unsettled debates in financial economics. On the one hand, supporters of the Efficient Market Hypothesis, EMH (Fama, 1970) argue that new information is efficiently incorporated into financial assets' prices. Even if prices can be predicted to some extent, investors should not be able to earn "abnormal", economic returns after transaction costs (Jensen, 1978), risk, and data snooping (Timmermann and Granger, 2004) are considered. On the other hand, detractors argue that stock prices can be systematically predicted and that investors can earn "abnormal" returns by timing the markets. Theoretical arguments (e.g., (Grossman and Stiglitz, 1980) and empirical findings (e.g., (Park and Irwin, 2007) prompted Lo (2004) to propose the Adaptive Market Hypothesis, AMH, as a middle ground between the two, in which markets sporadically deviate from efficiency as a result of external shocks but then quickly recover as a result of the swift action of informed traders. Some empirical findings in favor of the AMH exist (Lim and Brooks, 2011) but the concept has failed to gain significant traction. Recent tests using state of the art methodologies show that the EMH cannot be rejected, at least in developed stock markets and using classical prediction models (e.g., (Taylor, 2014).

The three key questions when investigating market efficiency are: (1) Can financial asset prices be predicted? (2) Is this predictability consistent trough time? (3) Can investors systematically earn "abnormal" returns after adjusting for trading costs, risk, and data snooping? Note that estimating the level of price (return) predictability, i.e. testing the random walk model or the martingale hypothesis, is not sufficient to make relevant inferences about the modern take on EMH (Timmermann and Granger, 2004), as various factors can prevent investors from transforming the predictability into economically significant "abnormal" profits. While the search for abnormal profits has been long and wide, there are still some limitations to our current understanding, especially in the context of the new developments in automated forecasting methods. First, the topic is widely investigated for developed stock markets, but far less so for emerging markets, and even less for the smallest, frontier stock markets of the world. Second, the analysis is largely concentrated on long time-horizons and low data frequencies (at least daily), even though intraday data is increasingly available. Third, the overwhelming majority of papers test simple forecasting methods, mostly from the field of Technical Analysis. Recently, Artificial Intelligence (AI) and Machine Learning (ML) forecasting models have recorded great successes in other scientific fields but have largely been ignored in financial economics for the study of the EMH.

This paper tries to partially fill this gap by investigating the predicting ability of AI and ML models on intraday prices in the frontier stock market of Romania. This is not the first time when the EMH is tested in Romania (e.g., (Dragotă and Oprea, 2014), offer a review of the literature dedicated on this market), neither the first time when AI and ML models are used to investigate important questions about the behavior of local stock prices (e.g., (Ruxanda and Badea, 2014), nor even the first time when intraday data is used (e.g., (Anghel, 2017) previously tested the random walk hypothesis using intraday data; while Anghel *et al.* (2020), investigated intraday patterns in returns). However, this is the first time when all the concepts are combined in a way that enables us to make relevant inferences about the modern-EMH and the AMH on intraday data in a frontier stock market, while using state of the art ML techniques. Also, compared to previous attempts, we additionally test for the optimum prediction horizon–i.e. the look-ahead window on which the best results can be obtained–of stock prices from historical intraday data.

In this endeavor, several contributions can be noted. First, we define and use 7 ML forecasting models, some of which have not been considered so far. Second, we test the prediction ability of all models over 7 different investment horizons. This test for the optimal prediction horizon of AI and ML forecasting models has never been performed before. Third, we measure the predictive accuracy of ML models, compare it to two relevant benchmark models, and test the statistical significance of the results. Finally, we perform a time-series analysis of the predictive ability of AI and ML forecasting models in order to get a better understanding about the time-varying nature of market efficiency.

The remainder of the paper is organized as follows. Section 2 presents the data, the ML forecasting models, and the testing methodology. Section 3 presents and discusses the results. Section 4 concludes.

2. Data and Methodology

2.1. Data

The data sample is collected from Bloomberg and consists of intraday prices for the main market index in Romania, BET, in the interval December 27, 2017–June 4, 2020. We retrieve the data at 1 minute intervals but we aggregate it to 15 minutes in order to alleviate the potential bias associated with low liquidity and microstructural noise such as the bid-ask bounce, which are important in small, frontier markets (see, e.g., (Anghel, 2017). The prices are used to calculate the series of log-returns, using $R_t = \ln(L_t/L_{t-1})$, where L_t is the last price recorded before or at moment t. Table 1 shows some summary statistics for the resulting series of returns.

Table-1. A Summary statistics for the data sample (BET index)

General information		Statistics for returns distribution	
Market Open	09:45:00	Minimum	-6.916238%
Market Close	16:45:00	Maximum	4.439241%
First Observation	12/27/2017 09:45	Median	0.000955%
Last Observation	6/4/2020 14:00	Average	0.000798%
No. Observations	17,017	Standard Deviation	0.204784%
Percent Up	50.40%	Skewness	-5.12
Percent Down	48.80%	Kurtosis	213.72

2.2. Forecasting Models

We construct 7 prediction models inspired from AI and ML forecasting techniques, namely Support Vector Machines; Logistic Regression; a 1000-tree Random Forest; a deep Neural Network (4 hidden layers with 25 artificial neurons each) with a Wide-and-Deep architecture that includes dropout layers (at a rate of 20%); a simple Recurrent Neural Network; a deep Recursive Neural Network (4 hidden layers with 20 artificial neurons each); and a deep Recursive Neural Network (4 hidden layers with 25 artificial neurons each) with a Long-Short Term Memory architecture that includes dropout layers (at a rate of 20%). Table 2 lists the architecture of the different models as implemented in *Python*, using the *scikit-learn* and *Keras* libraries. A detailed discussion on their characteristics, the type and roles of the various hyperparameters, and the way they can be implemented in practice is provided by Géron (2019).

Table-2. Summary of ML prediction models

No.	Abb.	Implementation
1	SVM	SGDClassifier(random_state=7);
2	LOGISTIC	LogisticRegression(max_iter=1000)
3	FOREST	RandomForestClassifier(n_estimators=1000, random_state=7);
4	WDNN	input_ = keras.layers.Input(shape=len(features)) droppout1 = keras.layers.Dropout(rate=0.2)(input_) hidden1 = keras.layers.Dense(25, activation="elu", kernel_initializer="he_normal", kernel_regularizer=keras.regularizers.l1())(droppout1) droppout2 = keras.layers.Dropout(rate=0.2)(hidden1) hidden2 = keras.layers.Dense(25, activation="elu", kernel_initializer="he_normal", kernel_regularizer=keras.regularizers.l1())(droppout2) droppout3 = keras.layers.Dropout(rate=0.2)(hidden2) hidden3 = keras.layers.Dense(25, activation="elu", kernel_initializer="he_normal",

		kernel_regularizer=keras.regularizers.l1())(droppout3) droppout4 = keras.layers.Dropout(rate=0.2)(hidden3) hidden4 = keras.layers.Dense(25, activation="elu", kernel_initializer="he_normal", kernel_regularizer=keras.regularizers.l1())(droppout4) concat = keras.layers.Concatenate()([input_, hidden4]) output = keras.layers.Dense(1, activation="sigmoid")(concat) wdnn = keras.Model(inputs=[input_], outputs=[output])
5	SRNN	srnn = keras.models.Sequential([keras.layers.SimpleRNN(1, input_shape=[None, 1])])
6	DRNN	drnn = keras.models.Sequential([keras.layers.SimpleRNN(20, return_sequences=True, input_shape=[None, 1]), keras.layers.SimpleRNN(20, return_sequences=True), keras.layers.SimpleRNN(20, return_sequences=True), keras.layers.SimpleRNN(20), keras.layers.Dense(1)])
7	LSTMNN	lstmnn = keras.models.Sequential([keras.layers.LSTM(25, return_sequences=True, input_shape=[None, 1]), keras.layers.Dropout(rate=0.2), keras.layers.LSTM(25, return_sequences=True), keras.layers.Dropout(rate=0.2), keras.layers.LSTM(25, return_sequences=True), keras.layers.Dropout(rate=0.2), keras.layers.LSTM(25), keras.layers.Dropout(rate=0.2), keras.layers.Dense(1)#, activation='sigmoid')])

We add 3 additional models derive from the 7 primary ones using the principles of ensemble methods. Specifically, we take the predictions $p_{m,t} \in \{0,1\}, m = \overline{1..7}$ and implement there different hard voting strategies to define new models. The first ensemble model, VOTE4, predicts a 1 if at least 4 of the 7 primary models also predict a 1 and zero otherwise; this is a decision by simple majority. The second ensemble model, VOTE5, predicts a 1 if at least 5 of the 7 primary models also predict a 1 and zero otherwise; this is a decision by qualified majority. The third ensemble model, VOTE6, predicts a 1 if at least 6 of the 7 primary models also predict a 1 and zero otherwise; this is a decision by super majority.

2.3. Testing Methodology

We construct the TARGET variables that ML models are trained to forecast. We frame the problem as a binary classification task over a predefined prediction interval. Specifically, we define the targets as $TARGET_{t,j} = \mathbb{1}_{\{L_{t+j}>L_t\}}$, where $\mathbb{1}_{\{*\}}$ is the indicator function taking 1 if the condition is true (price is expected to increase) and 0 otherwise (price is expected to decrease), while j is the prediction interval that we set to either 15 minutes, 30 minutes, 60 minutes (1 hour), 180 minutes (3 hours), 1440 minutes (1 day), 10,080 minutes (1 week), or 43,200 minutes (1 month). In the case of the RNN models, which process the data differently, the target is simply the stock price recorded j-minutes in the future.

Next, we use the raw series of returns to calculate several variables–denoted as "features" in ML optimization tasks–that characterize the properties of price movements and can be used by the forecasting models in order to make predictions. Specifically, we calculate the average return over several lookback windows of n observations, $\bar{R}_t^n = \frac{1}{n}\sum_{i=1}^{n} R_{t-i+1}, n = 2^m, m \in \{0,..,9\}$, which can be considered as proxies for price momentum over different intervals. Also, we calculate similar averages for the squared returns, $\bar{V}_t^n = \frac{1}{n}\sum_{i=1}^{n} R_{t-i+1}^2$, the cubed returns, $\bar{S}_t^n = \frac{1}{n}\sum_{i=1}^{n} R_{t-i+1}^3$, and the bi-squared returns, $K_t^n = \frac{1}{n}\sum_{i=1}^{n} R_{t-i+1}^4$, which can be considered proxies for the volatility, skewness, and kurtosis of returns over the same lookback intervals. To this we add two dummy variables: OPEN takes 1 for the first interval of the day and zero otherwise; CLOSE takes 1 for the last interval of the day and zero otherwise. These are intended to handle possible beginning/end of the day effects in market returns (see, e.g., (Anghel *et al.*, 2020). In total, 42 features are defined and used to forecast future stock prices. In the case of RNN models, we use a 120-observation data series of past prices as the input, which is equivalent to a lookback window of 30 trading hours (approximately 1 calendar week).

Then, we train and test the ML forecasting models as follows. First, we retain the earliest 1/10 of the data (1,701 observations) to initialize training. Note that this makes April 17, 2018, 15:15 the first observation for which we obtain predictions. This initial sample is split into a 70% training subsample and a 30% validation subsample. We train the models until they converge or, in the case of Artificial Neural Networks, for a maximum of 20 epochs, additionally employing early stopping if the prediction accuracy does not improve for 5 consecutive epochs. The combination of cross-sample validation and early stopping is intended to minimize model overfitting. The resulting models are then used to make out-of-sample predictions for the next 28 calendar days. After this period, the models are retrained by adding the new data into the training-validation data pool. Another set of predictions are then made for the next 28 days and the procedure continues in a similar fashion, until the end of the sample is reached.

After we obtain the predictions for all models, we estimate their accuracy and some characteristics of their distribution of returns. We do the same and compare the results with two benchmark prediction models, namely the buy-and-hold model, which buys the asset on the first observation and keeps it until the end (this is the typical passive trading strategy employed in the literature) and a naïve forecasting model that simply takes that last price change and uses it as a prediction. We compare the different models by testing the statistical significance of their predicting performance: the prediction accuracy and the average return obtained. This is done using a standard t-test, which is defined as:

$$t = \frac{\hat{\mu} - \mu}{\widehat{se}(\hat{\mu})}$$

where $\hat{\mu}$ is the estimated performance, $\mu = 0.5$ for prediction accuracy and $\mu = 0$ for average return, and $\widehat{se}$ is the estimated standard error of $\hat{\mu}$. We also test the difference between the means obtained by the different models and the passive benchmark assuming unequal variances of their performance distribution:

$$t_{dif} = \frac{\hat{\mu} - \hat{\mu}_b}{\sqrt{\widehat{se}(\hat{\mu})^2 + \widehat{se}(\hat{\mu}_b)^2}}$$

where N is the size of the sample and the subscript b denotes the benchmark model is used. Moreover, we test for statistically significant excess returns, alphas over the benchmark by estimating the linear regression:

$$R_{m,t} = \alpha + \beta R_{b,t} + \varepsilon_t$$

where $R_{m,t}$ denotes the series of returns obtained by applying model m, $R_{b,t}$ the series of returns of the benchmark model, and $\varepsilon_t \sim N(0, \sigma^2)$ are independent and identically distributed errors. Statistically significant alphas imply that economic returns can be attained by investors in a cost-free trading environment. Because trading costs are important in real markets, we compensate for this limitation by additionally computing and reporting break-even transaction cost for each model, i.e. the total return obtained divided by the number of trades (trading frequency). The higher the break-even cost, the better the model. Note that real trading costs in the Romanian market range between 0.1% and 1%, depending on the selected broker and the type of investor (trading frequency and the total value of the portfolio). Thus, a break-even cost in excess of 0.1% implies that the model can be used by some investors to earn "abnormal" profits, while a break-even cost in excess of 1% implies that economic returns can be obtained by all investors in the market.

3. Results

Table 3 reports the average prediction accuracy recorded by each model over each prediction interval, this being accompanied by the appropriate statistical tests. We find that the various models are consistently capable to predict future stock prices at above-random rates, irrespective of the prediction interval. Also, the prediction rates are significantly higher compared to the ones obtained by the naïve model and to those generated by the buy and hold strategy after accounting for model stability (the standard deviation of the recorded predictions). Moreover, we find that the prediction accuracy increases with the prediction interval, with daily, weekly and even monthly future price movements being predicted at a higher rate compared to shorter, intraday movements. This results implies that significant dependencies exist in intraday stock prices, especially over longer time horizons. Even though short-term price movements seem to be more random, this may be due to microstructural noise. In general, the findings are consistent in rejecting the classical take on weak-form EMH (Fama, 1970) for intraday price movements in the Romanian market.

Table-3. Test results–average prediction accuracy (all trades)

Prediction model	Prediction interval (minutes)						
	15	30	60	180	1440	10080	43200
BH	50.18% [0.46]	50.90% [2.22]**	50.80% [1.97]**	51.46% [3.58]***	55.43% [13.34]***	61.18% [27.98]***	72.05% [59.89]***
NAÏVE	46.63% [-8.22]***	47.40% [-6.33]***	47.84% [-5.26]***	48.93% [-2.59]***	51.14% [2.79]***	50.60% [1.48]	50.98% [2.40]**
SVM	51.21% [2.95]* {1.76}*	50.73% [1.79]*** {5.74}***	50.66% [1.62]*** {4.86}***	50.32% [0.80]** {2.40}**	52.71% [6.64]*** {2.72}***	53.53% [8.64]*** {5.05}***	57.61% [18.78]*** {11.50}***
LOGISTIC	51.38% [3.38]** {2.07}**	50.80% [1.95]*** {5.85}***	50.20% [0.51]*** {4.07}***	49.32% [-1.66] {0.66}	51.25% [3.07] {0.20}	53.70% [9.07]*** {5.36}***	53.39% [8.29]*** {4.16}***
FOREST	52.56% [6.26]*** {4.10}***	51.33% [3.27]*** {6.78}***	50.45% [1.12]*** {4.50}***	48.39% [-3.91] {-0.93}	49.74% [-0.62]** {-2.41}**	55.87% [14.42]*** {9.12}***	58.38% [20.72]*** {12.84}***
WDNN	49.60% [-0.95] {-1.00}	50.91% [2.23]*** {6.05}***	51.09% [2.66]*** {5.60}***	53.41% [8.34]*** {7.73}***	54.45% [10.91]*** {5.72}***	60.83% [27.06]*** {17.85}***	54.75% [11.64]*** {6.51}***
SRNN	50.40% [0.98] {0.37}	50.63% [1.56]*** {5.57}***	50.17% [0.43]*** {4.02}***	49.59% [-0.98] {1.14}	49.55% [-1.08]*** {-2.74}***	58.81% [21.82]*** {14.26}***	30.35% [-52.09]*** {-37.02}***
DRNN	49.86% [-0.33] {-0.56}	50.23% [0.57]*** {4.88}***	51.07% [2.63]*** {5.57}***	53.49% [8.54]*** {7.87}***	54.39% [10.76]*** {5.62}***	58.01% [19.80]*** {12.86}***	63.39% [33.89]*** {21.79}***

LSTMNN	50.81% [1.99] {1.08}	50.61% [1.49]*** {5.53}***	51.09% [2.68]*** {5.61}***	52.84% [6.94]*** {6.74}***	52.21% [5.40]* {1.85}*	59.03% [22.39]*** {14.65}***	58.60% [21.29]*** {13.23}***
VOTE4	51.07% [2.63] {1.53}	51.16% [2.84]*** {6.48}***	51.23% [3.02]*** {5.85}***	52.03% [4.98]*** {5.35}***	52.90% [7.08]*** {3.03}***	61.15% [27.89]*** {18.41}***	55.91% [14.52]*** {8.54}***
VOTE5	50.35% [0.87] {0.29}	50.82% [2.00]*** {5.89}***	51.09% [2.66]*** {5.60}***	50.76% [1.87]*** {3.16}***	50.63% [1.54] {-0.88}	54.98% [12.22]*** {7.57}***	49.38% [-1.49]*** {-2.75}***
VOTE6	49.97% [-0.07] {-0.37}	50.25% [0.62]*** {4.91}***	49.51% [-1.18]*** {2.88}***	49.44% [-1.36] {0.87}	48.73% [-3.09]*** {-4.15}***	51.21% [2.95] {1.04}	41.82% [-20.19]*** {-15.89}***

Note: t reported in squared parenthesis; t_{dif} reported in brackets. ***, **, and * denote statistical significance at the 99%, 95%, and 90% levels, respectively.

Table-4. Test results–average prediction accuracy (long trades)

Prediction model	Prediction interval (minutes)						
	15	30	60	180	1440	10080	43200
SVM	46.48% [-6.08]***	58.49% [14.99]***	46.27% [-6.49]***	62.58% [22.73]***	68.74% [36.70]***	64.43% [28.74]***	68.77% [41.92]***
LOGISTIC	48.08% [-3.31]***	56.16% [10.81]***	58.99% [15.89]***	62.30% [22.20]***	64.42% [27.34]***	71.51% [45.45]***	64.73% [31.90]***
FOREST	49.61% [-0.67]	53.36% [5.87]***	53.49% [6.09]***	51.75% [3.07]***	54.95% [9.04]***	70.56% [43.01]***	73.00% [53.60]***
WDNN	11.58% [-103.62]***	53.29% [5.75]***	71.01% [40.23]***	74.74% [49.80]***	70.71% [41.31]***	92.34% [151.82]***	62.85% [27.52]***
SRNN	38.96% [-19.54]***	42.15% [-13.82]***	34.45% [-28.41]***	25.13% [-50.11]***	39.49% [-19.50]***	60.79% [21.07]***	3.37% [-267.22]***
DRNN	27.41% [-43.72]***	39.44% [-18.79]***	38.76% [-20.03]***	64.47% [26.45]***	58.85% [16.32]***	58.23% [15.93]***	63.20% [28.32]***
LSTMNN	35.38% [-26.40]***	43.69% [-11.05]***	44.93% [-8.84]***	57.03% [12.43]***	60.98% [20.44]***	59.10% [17.66]***	52.78% [5.77]***
VOTE4	24.67% [-50.71]***	51.73% [3.03]***	52.17% [3.78]***	63.03% [23.62]***	65.80% [30.24]***	76.61% [59.94]***	62.36% [26.41]***
VOTE5	7.69% [-137.02]***	28.20% [-42.12]***	28.69% [-40.91]***	41.45% [-15.16]***	46.61% [-6.16]***	54.20% [8.04]***	41.11% [-18.68]***
VOTE6	1.15% [-394.99]***	11.93% [-102.07]***	10.02% [-115.60]***	19.73% [-66.49]***	23.99% [-55.27]***	39.40% [-20.67]***	28.12% [-50.32]***

Note: t reported in squared parenthesis. ***, **, and * denote statistical significance at the 99%, 95%, and 90% levels, respectively.

Table-5. Test results–average prediction accuracy (short trades)

Prediction model	Prediction interval (minutes)						
	15	30	60	180	1440	10080	43200
SVM	55.97% [10.35]***	42.68% [-12.63]***	55.19% [8.94]***	37.33% [-22.23]***	32.77% [-29.85]***	36.35% [-21.54]***	28.82% [-30.11]***
LOGISTIC	54.71% [8.15]***	45.23% [-8.17]***	41.13% [-15.40]***	35.54% [-25.64]***	34.87% [-25.81]***	25.63% [-42.38]***	24.15% [-38.90]***
FOREST	55.54% [9.59]***	49.23% [-1.30]	47.31% [-4.59]***	44.84% [-8.81]***	43.26% [-11.06]***	32.72% [-27.95]***	20.68% [-46.62]***
WDNN	87.91% [100.08]***	48.44% [-2.66]***	30.51% [-36.17]***	30.79% [-35.33]***	34.22% [-27.04]***	11.16% [-93.60]***	33.86% [-21.97]***
SRNN	61.93% [21.14]***	59.44% [16.42]***	66.40% [29.69]***	75.53% [50.43]***	62.06% [20.24]***	55.68% [8.69]***	99.90% [1036.37]***
DRNN	72.48% [43.31]***	61.42% [20.05]***	63.79% [24.53]***	41.84% [-14.04]***	48.85% [-1.87]*	57.66% [11.78]***	63.89% [18.64]***
LSTMNN	66.36% [29.79]***	57.78% [13.45]***	57.45% [12.89]***	48.39% [-2.73]***	41.29% [-14.38]***	58.91% [13.76]***	73.60% [34.50]***
VOTE4	77.67% [57.17]***	50.56% [0.97]	50.27% [0.47]	40.37% [-16.66]***	36.84% [-22.19]***	36.78% [-20.81]***	39.28% [-14.14]***
VOTE5	93.33% [149.49]***	74.27% [47.43]***	74.22% [47.33]***	60.63% [18.49]***	55.63% [9.23]***	56.22% [9.53]***	70.71% [29.32]***
VOTE6	99.16% [463.94]***	89.99% [113.78]***	90.29% [116.38]***	80.94% [66.89]***	79.51% [59.50]***	69.82% [32.79]***	77.14% [41.65]***

Note: t reported in squared parenthesis. ***, **, and * denote statistical significance at the 99%, 95%, and 90% levels, respectively.

Table 4 and 5 break down the overall prediction accuracy into one specific for long trades and another specific for short trades. We find a significant heterogeneity in results, with some models being better at predicting price increases and others at predicting price decreases. Also, the accuracy of the various models significantly varies depending on the prediction interval, which is a sign of important heterogeneities in price dependencies and in model capabilities to detect them. For example, the Random Forest model is not able to predict short-term price increases very well (average accuracy is between 49% and 53% for short-term predictions up to 180 minutes) but is better able to predict long-term price increases (accuracies between 55% and 73% for the same intervals). However, it behaves exactly the opposite for short trades, starting with an above-average prediction rate of 55% over an interval of 15 minutes, this continuingly decreasing towards 20% as we increase the prediction horizon. Conversely, the SRNN is very bad at prediction price increases (accuracy generally below 40%), but is very good at predicting price decreases (accuracy generally above 60%). Interestingly, all ensemble methods show the best prediction accuracies for price declines, but the worst ones for price increases, irrespective of prediction interval.

On the one hand, given the observed heterogeneity in performance, the results show that investors face a difficult task when ex-ante choosing a prediction model that can met their needs. On the other hand, they show an asymmetry between bull and bear markets, at least when intraday prices are examined: price increases are slow but more instable, inconsistent, and more difficult to predict; while price decreases are sharp but more stable, consistent, and can be consistently predicted, especially by RNN's and ensemble methods. This implies that bear markets are less efficient than bull markets, hinting that investors asymmetrically respond to information, a result which echoes previous findings in the behavioral finance literature (e.g., (Hirshleifer *et al.*, 2016).

Table-6. Test results–average return (annualized)

Prediction model	**Prediction interval (minutes)**						
	15	**30**	**60**	**180**	**1440**	**10080**	**43200**
BH	0.15% [0.01]	0.15% [0.01]	0.15% [0.01]	0.15% [0.01]	0.15% [0.01]	0.15% [0.01]	0.15% [0.01]
NAÏVE	2.62% [0.37]	2.62% [0.37]	2.62% [0.37]	2.62% [0.37]	2.62% [0.37]	2.62% [0.37]	2.62% [0.37]
SVM	7.14% [0.85] {0.49}	8.20% [0.90] {0.50}	-2.78% [-0.29] {-0.46}	4.33% [0.44] {0.14}	0.00% [0.00] {-0.23}	3.03% [0.30] {0.03}	5.25% [0.51] {0.21}
LOGISTIC	9.07% [0.96] {0.60}	17.05% [1.92]* {1.32}	-0.30% [-0.03] {-0.24}	0.17% [0.02] {-0.20}	-0.39% [-0.04] {-0.25}	0.58% [0.05] {-0.16}	2.01% [0.19] {-0.05}
FOREST	27.04% [4.00]*** {2.15}**	6.80% [0.73] {0.37}	7.17% [0.78] {0.40}	-3.30% [-0.37] {-0.52}	-2.34% [-0.31] {-0.47}	4.25% [0.40] {0.13}	5.32% [0.58] {0.24}
WDNN	-9.73% [-1.21] {-0.70}	7.71% [0.75] {0.42}	5.31% [0.49] {0.21}	9.17% [0.91] {0.55}	5.54% [0.54] {0.24}	6.47% [0.58] {0.30}	-4.66% [-0.44] {-0.57}
SRNN	9.94% [2.20]** {0.79}	16.70% [3.94]*** {1.82}*	6.22% [1.44] {0.44}	5.10% [1.41] {0.32}	6.59% [1.44] {0.48}	14.68% [2.85]*** {1.45}	1.19% [1.01] {-0.20}
DRNN	-5.59% [-0.78] {-0.42}	10.17% [1.80]* {0.86}	13.53% [2.98]*** {1.36}	13.86% [2.52]** {1.31}	19.86% [4.05]*** {2.15}**	14.88% [3.02]*** {1.50}	15.97% [3.10]*** {1.61}
LSTMNN	11.42% [2.76]*** {0.93}	9.48% [1.94]* {0.83}	12.83% [2.70]*** {1.26}	15.66% [3.04]*** {1.57}	-11.68% [-1.31] {-1.27}	16.61% [3.43]*** {1.73}*	11.86% [2.56]** {1.14}
VOTE4	13.70% [2.06]** {1.04}	21.85% [3.03]*** {2.02}**	9.62% [1.17] {0.66}	14.44% [1.92]* {1.19}	0.64% [0.08] {-0.18}	18.28% [2.09]** {1.45}	8.72% [1.02] {0.56}
VOTE5	4.15% [0.98] {0.32}	14.95% [3.07]*** {1.51}	13.68% [3.15]*** {1.40}	10.30% [2.20]** {0.94}	4.43% [0.73] {0.20}	12.83% [2.51]** {1.22}	9.39% [2.01]** {0.83}
VOTE6	3.37% [2.44]** {0.27}	10.37% [3.36]*** {1.05}	4.09% [1.60] {0.20}	1.53% [0.42] {-0.14}	2.99% [0.82] {0.05}	13.24% [3.15]*** {1.36}	6.86% [1.69]* {0.53}

Note: t reported in squared parenthesis; t_{dif} reported in brackets. ***, **, and * denote statistical significance at the 99%, 95%, and 90% levels, respectively. Statistically significant positive returns are highlighted in green. Statistically significant negative returns are highlighted in red.

Table-7. Test results–average excess return (alpha) vs. buy and hold benchmark

Prediction model	Prediction interval (minutes)						
	15	30	60	180	1440	10080	43200
SVM	7.09% [1.21]	8.12% [1.44]	-2.87% [-0.50]	4.23% [0.77]	-0.09% [-0.02]	2.92% [0.55]	5.14% [1.04]
LOGISTIC	9.01% [1.65]*	16.95% [3.21]***	-0.40% [-0.07]	0.07% [0.01]	-0.49% [-0.08]	0.45% [0.11]	1.89% [0.38]
FOREST	27.00% [5.22]***	6.71% [1.19]	7.09% [1.25]	-3.39% [-0.58]	-2.40% [-0.41]	4.14% [0.89]	5.23% [0.91]
WDNN	-9.82% [-1.71]*	7.61% [1.57]	5.19% [1.20]	9.07% [1.85]*	5.44% [1.10]	6.34% [1.82]*	-4.78% [-1.05]
SRNN	9.95% [2.39]**	16.71% [4.26]***	6.20% [1.54]	5.09% [1.48]	6.57% [1.56]	14.65% [3.21]***	1.19% [1.02]
DRNN	-5.65% [-0.99]	10.13% [2.07]**	13.51% [3.25]***	13.81% [2.88]***	19.82% [4.53]***	14.88% [3.36]***	15.95% [3.49]***
LSTMNN	11.37% [2.96]***	9.45% [2.14]**	12.78% [2.97]***	15.65% [3.42]***	-11.74% [-2.13]**	16.59% [3.81]***	11.80% [2.80]***
VOTE4	13.63% [2.54]**	21.79% [4.07]***	9.57% [1.66]*	14.35% [2.58]***	0.57% [0.09]	18.22% [3.45]***	8.63% [1.50]
VOTE5	4.13% [1.05]	14.94% [3.40]***	13.69% [3.41]***	10.26% [2.41]**	4.39% [0.84]	12.78% [2.81]***	9.38% [2.19]**
VOTE6	3.37% [2.45]**	10.38% [3.49]***	4.09% [1.64]	1.52% [0.44]	2.98% [0.86]	13.21% [3.39]***	6.83% [1.79]*

Note: t reported in squared parenthesis. ***, **, and * denote statistical significance at the 99%, 95%, and 90% levels, respectively. Statistically significant positive alphas are highlighted in green. Statistically significant negative alphas are highlighted in red.

Up to now, the results show that historical intraday stock prices can be successfully used to predict future price movements at intervals of up to one month. This is a clear violation of the Efficient Market Hypothesis in its classical specification (Fama, 1970). But can investors use this information to earn surplus returns compared to a passive buy and hold strategy? To answer this question, Table 6 reports the average return associated to each model and each prediction interval, while Table 7 reports the alpha obtained versus the buy and hold benchmark. The results show that only the prediction models based on Recurrent Neural Networks are able to systematically earn statistically significant excess returns over the buy and hold benchmark, while the others either are not capable of such a task (NAÏVE, SVM, LOGISTIC, FOREST, WDNN), or are influenced by the bad predictions of some alternatives and have reduced capacities (VOTE4 VOTE5, VOTE6). Interestingly, the DRNN and the LSTMNN are better than the simple SRNN, consistently earning average returns and alphas of about 10%-20% per year, irrespective of the prediction interval. On the one hand, this result shows that stock prices display long memory even at the intraday level, this making the RNN-type models more profitable. However, these long-term dependencies are of a more complex, nonlinear nature and can only using more models with a more sophisticated architecture. On the other hand, the ability by at least some models (RNN's) to earn excess returns is yet another piece of evidence that contradicts the classic EMH.

Table-8. Test results–break-even transaction costs

Prediction model	Prediction interval (minutes)						
	15	30	60	180	1440	10080	43200
SVM	0.1517%*	0.1517%*	0.1517%*	0.1517%*	0.1517%*	0.1517%*	0.1517%*
LOGISTIC	0.0003%	0.0003%	0.0003%	0.0003%	0.0003%	0.0003%	0.0003%
FOREST	0.0026%	0.0039%	-0.0014%	0.0024%	0.0000%	0.0023%	0.0097%
WDNN	0.0024%	0.0055%	-0.0001%	0.0001%	-0.0004%	0.0011%	0.0093%
SRNN	0.0049%	0.0016%	0.0023%	-0.0019%	-0.0021%	0.0133%	0.0253%
DRNN	-0.0089%	0.0019%	0.0025%	0.0047%	0.0039%	0.0226%	-0.0144%
LSTMNN	0.0564%	0.0300%	0.0030%	0.0018%	0.0036%	0.0175%	0.0017%
VOTE4	-0.0069%	0.0089%	0.0215%	0.0270%	0.0423%	0.0617%	0.0902%
VOTE5	0.0081%	0.0108%	0.0132%	0.0149%	-0.0080%	0.0404%	0.2757%*
VOTE6	0.0047%	0.0066%	0.0037%	0.0078%	0.0004%	0.0347%	0.0229%

Note: * denotes weak economically-significant break-even transaction costs above the 0.1% threshold.

Figure-1. Average prediction accuracy of DRNN model-intraday intervals

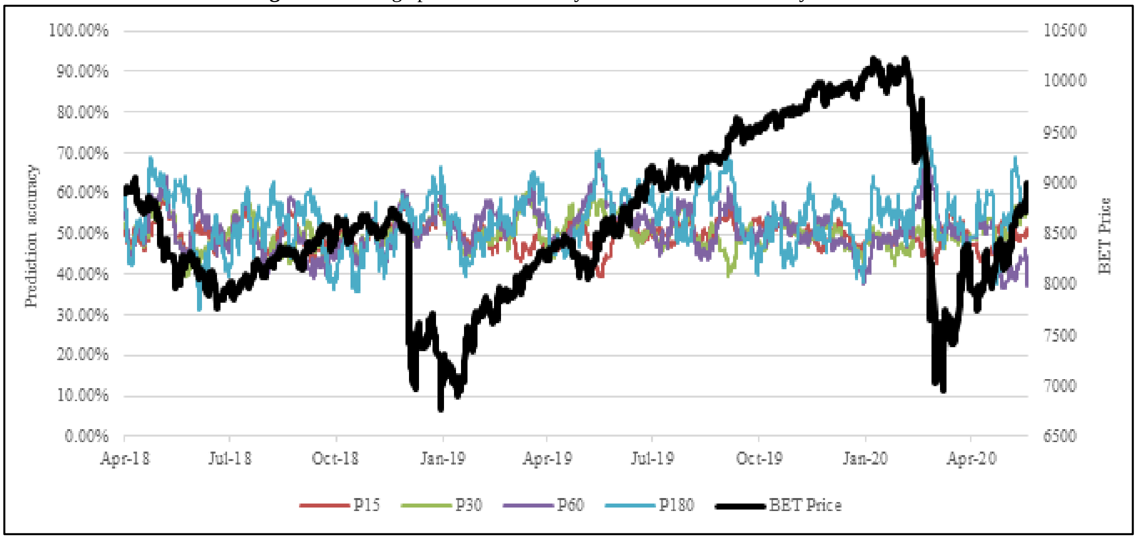
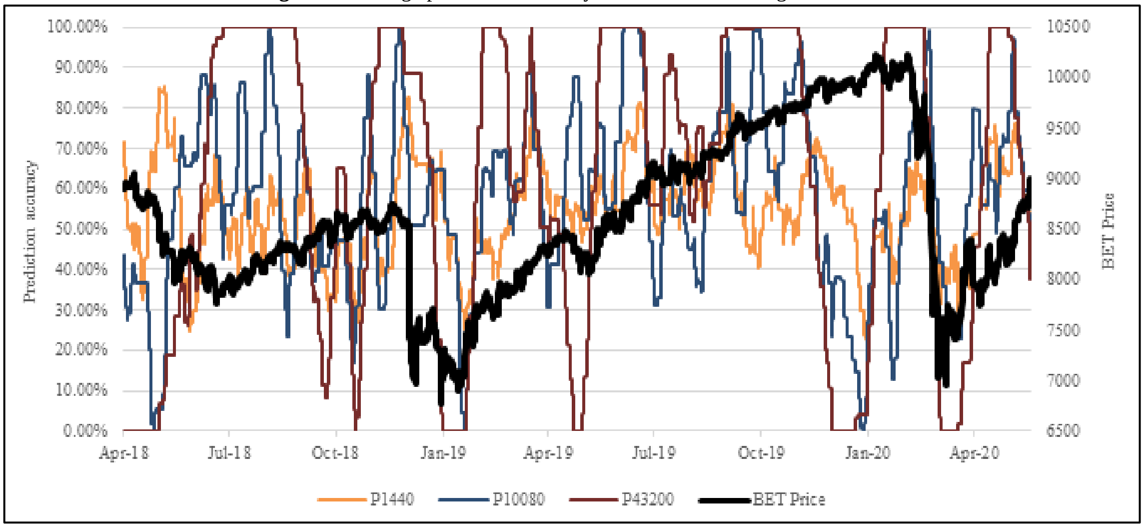


Figure-2. Average prediction accuracy of DRNN model-longer intervals



Even though excess returns can be earned in a frictionless market by some models, this does not necessarily translate into gains for investors in real markets. Thus, the modern take on the EMH is not necessarily rejected. To get a better understanding, we report break-even transaction costs in Table 8. Interestingly, with the minor exception of the VOTE5 model over the 1-month interval, no prediction model is capable of generating excess returns that compensate the trading costs needed to implement them. Moreover, the break-even costs are generally much lower compared with the 0.1% threshold. Thus, unless traders pay no fees, timing intraday prices using ML models would actually lead to financial losses. As a result, we cannon reject (nor are we close to rejecting) the modern take on the weak-form Efficient Market Hypothesis, as specified by Timmermann and Granger (2004), based on state of the art ML prediction models. This is an important results that complements and extends the findings of Anghel (2015;2017), which showed that intraday trading strategies based on technical analysis or statistical patterns in stock returns do not lead to economic profits.

Finally, Figures 1 and 2 show the average prediction accuracy of the DRNN model for a backward-forward window of 301 observations (equivalent to about 2 calendar weeks) centered on the calendar time. Even though the regions of high/low prediction accuracies of the other models are different (the results are available upon request), the general characteristics of the series are similar and point to the same conclusion regarding the time-varying predictability of intraday stock prices. We find that the average prediction accuracy over intraday intervals generally varies between 40% and 60%, with occasional spikes towards the 70% level. Regions of high predictability generally correspond with periods of persistent price movements, especially in the negative direction. The maximum average prediction accuracy is 69.10% for a 60 minute interval and 75.42% for a 180 minute interval; these are both recorded at the beginning of March 2020, during the recent market collapse due to the COVID-19 pandemic. The lowest values that fall below 40% are mainly recorded in October 2018, during a period of generally rising prices. Conversely, when using longer prediction intervals, bull trends seem to be better predicted by the RNN model, even with accuracies of 100% in a limited time interval; while price corrections generally corresponds to regions of low predictability.

On the one hand, this result shows that the optimum prediction interval is not unique but depends on price momentum and prediction model characteristics. Although intraday prices are predictable on average, this signals that return generating processes are not ergodic and add an additional layer of randomness for investors aiming to ex-ante choosing a prediction model capable of "beating" the market. On the other hand, this result reinforces our previous conclusion regarding the asymmetric behavior of investors in bull and bear markets, showing that price predictability is greater in the latter. But because no general improvements in prediction accuracy can be observed though time, this again does not contradict the modern take on the EMH. The latter result also shows that the AMH (Lo, 2004) is not a better theory for explaining intraday price movements in this case.

Overall, the results show that intraday stock prices can be systematically predicted, although with average rates not too much above the 50% random threshold. However, the predictability of intraday price movements rises with the prediction interval, reaching levels of around 60% at the 1 month interval. In specific cases, longer-term price movements can be predicted at average rates of above 70% using historical intraday fluctuation, this even rising to 100% for specific periods in which price movements persistently occur in a single direction. Even though this finding leads to a clear rejection of the EMH in its classical form (Fama, 1970), we cannot reject the more modern take discussed by Timmermann and Granger (2004). Several other factors also contribute to this conclusion. First, there is a very large heterogeneity in the accuracy of the tested models when long and short trades are analyzed separately. Second, the prediction accuracy of the various models widely vary trough time, in a seemingly random fashion. More importantly, given the significant number of trades required to apply ML models in the market, the break-even transaction costs are significantly lower compared to actual trading fees that real investors pay. As a result, no economic profits are attainable by timing the market at an intraday horizon using ML forecasting models.

4. Conclusion

This papers investigates the predictability of intraday stock prices in the frontier market of Romania using forecasting models derived from Artificial Intelligence and Machine Learning. Our main finding is that future price movements can be predicted using historical intraday data, and that this predictability increases as we extend the forecasting interval (look-ahead period). Also, in a frictionless environment, investors can use this information to develop trading strategies that earn statistically significant excess returns compared to a passive buy and hold strategy and a naïve strategy that simply predicts prices will continue to move in the same direction. However, when considering trading fees that real investors actually pay, we find no trading strategy that is able to "beat" the passive benchmark in an economically significant way. Moreover, even though ML models display above-random predictive ability on average, this seems to fluctuate randomly in time and throughout the prediction model cross-section. As a result, we conclude that intraday stock prices in the frontier stock market of Romania move in a way that is consistent with the modern take on the weak-form Efficient Market Hypothesis (Timmermann and Granger, 2004), while that Adaptive Market Hypothesis (Lo, 2004) does not seem to provide a better perspective.

On the one hand, our results imply that investors cannot use intraday trading information to time the market, even when using modern prediction models derived from Artificial Intelligence and Machine Learning. This complements and extends previous findings based on simpler prediction models (Anghel, 2015;2017). On the other hand, even though we find evidence of investor asymmetric reaction to information, which provides support to theoretical concepts in the behavioral finance literature (Hirshleifer *et al.*, 2016), we show that the Romanian stock market is efficient at pricing historical intraday information obtained via ML techniques, this being consistent with similar evidence obtained on more developed, international markets (e.g., (Caporale *et al.*, 2016)).

References

Anghel, D. G. (2015). Market efficiency and technical analysis in Romania. *International Journal of Financial Research*, 6(2): 164-77.

Anghel, D. G. (2017). Intraday market efficiency for a typical central and eastern european stock market: The case of Romania. *Romanian Journal of Economic Forecasting*, 20(3): 88.

Anghel, D. G., Țilică, E. and Dragotă, V. (2020). Intraday patterns in returns on frontier stock markets. *Romanian Journal of Economic Forecasting*, 23(2).

Caporale, G. M., Gil-Alana, L., Plastun, A. and Makarenko, I. (2016). Intraday anomalies and market efficiency: A trading robot analysis. *Computational Economics*, 47(2): 275-95.

Dragotă, V. and Oprea, D. S. (2014). Informational efficiency tests on the Romanian stock market: a review of the literature. *The Review of Finance and Banking*, 6(1): 15-28.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2): 383-417.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3): 393-408.

Hirshleifer, D. A., Huang, C. and Teoh, S. H., 2016. "Information Asymmetry, Market Participation, and Asset Prices." In *6th Miami Behavioral Finance Conference.*

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2/3): 95-101.

Lim, K. P. and Brooks, R. (2011). The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys*, 25(1): 69-108.

Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5): 15-29.

Park, C. H. and Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4): 786-826.

Ruxanda, G. and Badea, L. M. (2014). Configuring artificial neural networks for stock market predictions. *Technological and Economic Development of Economy*, 20(1): 116-32.

Taylor, N. (2014). The rise and fall of technical trading rule success. *Journal of Banking and Finance*, 40: 286-302. Available: https://www.sciencedirect.com/science/article/abs/pii/S0378426613004664

Timmermann, A. and Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1): 15-27.