

A ESTRATÉGIA INTRADAY VALE A PENA? UM ESTUDO DE CASO USANDO APRENDIZADO DE MÁQUINA

DOES INTRADAY STRATEGY WORTH? A CASE STUDY USING MACHINE LEARNING

ABUABARA, Leila
Universidade Federal de São Paulo (UNIFESP)
Instituto Tecnológico Aeronáutico (ITA)
leila.abuabara@unifesp.br

VERRI, Filipe Alves Neto
Instituto Tecnológico de Aeronáutica (ITA)
verri@ita.br

Recebido em 09/2021 – Aprovado em 12/2021

Resumo

Cada vez mais pessoas têm contado com a renda variável para diversificar seus investimentos e aumentar seus ganhos. Em geral, esse tipo de investimento começa como uma fonte secundária de recursos. O processo de compreensão do funcionamento desse tipo mercado conta com diversas iniciativas educacionais independentes. Com o objetivo de vivenciar a dinâmica e principalmente os riscos associados a tais investimentos, propomos aqui a aplicação de um modelo de predição utilizando árvore de regressão como algoritmo de aprendizado para prever se existe um período ideal em um dia de negociação para realizar uma transação conveniente, conforme os interesses do investidor pela venda ou compra de ativos. Selecionamos uma ação brasileira que faz parte do Índice IBovespa (IBOV) para acompanhar durante um período de dois meses e meio. Nossas conclusões mostram que a negociação de curto prazo é muito arriscada devido à volatilidade dos preços do mercado de ações ao longo de um mesmo dia. Ainda, sugerimos que investidores inexperientes devem contar com técnicas combinadas e informações adicionais para uma negociação bem-sucedida, ou mesmo considerar produtos financeiros mais seguros. Além disso, uma visão holística e realista, que vai além da busca por ganhos imediatos, é necessária entre o crescente público que investe o qual está excessivamente exposto às propagandas que argumentam que este risco compensa.

Palavras-chave: Mercado de ações, Árvore de regressão, Brasil.

Abstract

More and more individuals are counting on variable income to diversify their investments and boost earnings. In general, this sort of investment starts as a secondary source of gains. The process of understanding how this sort of market works counts on many independent instructional initiatives. Aiming to experience the dynamics and mainly risks associated with such investments, we proposed here the application of a prediction model using regression tree as the learning algorithm to predict if there is an ideal period of a business day to perform a convenient trading, according to the investor objective in assets sale or purchase. We selected a Brazilian stock which is part of the IBovespa Index (IBOV) to follow during a two and half month period. Our conclusions show that short-term trading is very risky due to the volatility of the stock market prices over a same day. Yet, we suggest that inexperienced investors should count on combined techniques and additional information for succeeded trading, or even consider safer financial products. Furthermore, a holistic and realistic perspective that goes beyond the search for immediate gains is needed among the growing investing public who is overly exposed to advertisements that argue that this risk pays off.

Keywords: Stock market, Regression tree, Brazil

INTRODUÇÃO

Variable income has been becoming an increasingly investment alternative for ordinary people. We refer to the term 'variable income' (in opposite to 'fixed income') as the investment where interest rate is unknown at the time of application, since they will depend on future factors generally related to the global and local economics and politics. Thus, they are destined to people more prone to risk. However, the high risk can be compensated with high returns if compared to fixed income investments, or even losses, which shall be considered indeed. Some examples of variable income are stock market,

foreign exchange, derivatives, equity funds, among others. In 2019, the Brazilian stock exchange reached 1.5 million individual investors (COTIA, 2020). An increasing figure considering that in the previous year this amount was around 800 thousand, practically it doubled. The first quarter of 2020 has finished summing up more than 2.2 million Brazilian individuals in the stock market (B3[1], 2020). This happens contemporaneously to the COVID-19 pandemic when many Brazilians are losing their jobs (MANFRINI, 2020). However entering in the stock market does not mean to investing a large amount of money (B3[3], 2020) since we can start with a few tens of Brazilian Reais (BRL). The demand (or expectation) for higher profitability goes along with a certain inexperience of such investors that somehow counts on expert tips and hints.

The purpose of this work is to use a machine learning technique to better understand how this market works and support the decision process. We collected intraday data of a stock of a large Brazilian holding company, considered as stable to a certain extent, in order to predict if it is a good time of the day to perform good deals according to the individual objective of each investor, usually purchase at a low stock price and sell at a high stock price. We concentrated our study on the particular investor who has other professional activities and does not follow minute-by-minute variations searching for the right moment for a trade, especially in short-term such as the intraday level trading, where stocks are purchased and sold during the same trading day promoting hourly earnings, but not limited to them.

We argue that ordinary people shall be familiar to investments which they want to risk, not just letting themselves to be attracted to *potential* easy gains. Investments critical appraisal and system thinking are fundamental to novices and experienced investors. That means to look to the boundaries and be aware of *what are the win and loss limits?* In addition, to have a certain self-knowledge (OSMAN, 2020) which goes beyond the usual risk profile investor test and be

especially prepared for losses or knowing how to make choices that suit us best.

The article includes four sections beyond this introduction. In section 2, we give an overview of the Brazilian stock market and investors, and some previous similar studies in this area. In section 3, we describe our dataset and present our learning task and the adopted methodology. In section 4, we present and discuss the experimental results analyzing them. Conclusions and directions for further work are given in section 5.

LITERATURE REVIEW

Brazilian Investors

In Brazil, financial education is introduced already in the early years of elementary school. A recent law decree of 2010 established the National Financial Education Strategy (Estratégia Nacional de Educação Financeira – ENEF 2010) aiming to contribute to the strengthening of citizenship, the efficiency and solidity of the national financial system and conscious decision-making by consumers. For those already grown up (and that did not have a similar opportunity at school), they enjoy the scattered and independent educational initiatives that put efforts to train and inform them through courses, workshops, YouTube videos, websites, books, a true ‘specialized media’ which also includes the Brazilian Stock Exchange itself (known as “[B]3”, the acronym of Bolsa, Brazil, Balcão) (BONALDI, 2018). The popularization of stock investment in Europe took place during the 19th century (PREDA, 2006). In Brazil, the expansion process has been quite recent, in the end of the 20th century. Beyond the legitimizing discourse as mentioned above, we come to count on socio-technical apparatus, the home-broker system which was introduced in 1999 and operationally linked the individual investor to the stockbroker (BONALDI, 2018) without which it is impossible to access this sort of investment.

Investors Profile

The most recent information released by [B]3 (B3, 2020) in relation to the Brazilian individual investor profile dates of December 2020. In general words, we can say that they are mostly (74%) men. They are on average 32 years old. And they mostly (51%) live in the southeastern States of Brazil (São Paulo, Rio de Janeiro, Espírito Santo e Minas Gerais). Literature (PREDA, 2006; TSAI; WANGng, 2009) points us two important common indicators for forecasting stock price: (i) fundamental analysis, which uses company’s financial (e.g. balance sheets) and market (e.g. prospects and opportunities) information; and, (ii) technical analysis, which includes a graphical analysis based on trends acquired from past prices and trading volumes. Brazilian studies suggest (ROGERS; SECURATO; RIBEIRO, 2007) a third one: (iii) subjective analysis, as part of the behavioral finance which includes the analysis of emotions and cognition that influence the decision of individuals. We add to this a dash of rationality and self-knowledge since many investors especially men are overconfident in their investment decisions (BARBER; ODEAN, 2001). However, the reality is that they know less than they consider knowing.

A Wispy of Similar Studies

The desire of making predictions in the stock market using machine learning techniques is not a novelty. Tsai and Wang (2009) combined Artificial Neural Networks (ANN) and Decision Tree (DT) to enhance the rate of prediction accuracy in the stock price forecasting of an electronic industry in the Taiwan market. Another hybrid technique was used by Choudhry and Garg (2008) through the combination of Genetic Algorithm (GA) and Support Vector Machines (SVMs). ANNs were also used by Olson and Mossman (2003) and compared to ordinary least squares (OLS) and logistic regression (logit) techniques in the context of Canadian stock returns forecasts. While Swales and Yoon (1992) compared stock performance using ANN and linear multiple discriminant analysis (MDA) approaches with satisfied results to the first. However, in relation

to day trading, it doesn't seem to make any sense for any profile of investor (CHAGUE; GIOVANNETTI, 2020). The dropout rate is as high as the financial loss witnessed in this type of investment, 92% and 91%, respectively. However, experts say that by getting experience, earnings can improve.

A System Approach

Behind each stock of the stock exchange, there is a business in a certain sector of activity, which is inserted in a market with competitors and cooperators, customers and suppliers, providing a service or within a production chain. This means a complete “ecosystem” (WATSON, 2018) that should be widely understood and followed by investors beyond the economic and political scenario. After all, “behaviors and relationships are constantly adapting in response to the other elements of the environment” (p.2). This system view has also been widely offered for investors and discussed in accessible media by consultants, influential videos from expert speakers (M1 Finance [1], 2020) and so on. In practice, five relevant lessons shall be kept in mind, committed and continually updated by investors (M1 Finance [2], 2020):

- i. Be aware that financial goals shall not be decoupled from the investor's other life goals;
- ii. A perfectible self-knowledge about investor risk tolerance is fundamental;
- iii. Rely on digital tools (not just any tool, but useful) that can assist investor to make better financial decisions and save time, that means information automation & consolidation (e.g. tradingview.com);
- iv. Study and follow the strategies of great investors. *Have good references!* And;
- v. Revisit the strategy frequently over time. *We and the world are always changing!*

MATERIAL AND METHODOLOGY

Dataset

The original data has a single source: a digital newspaper about economics (<https://economia.uol.com.br/>). The data collection was performed on a daily basis from March 27, 2020, to June 15, 2020 (53 business days). Each end of day, the minute-by-minute quotation of that day was extracted. The data set of each day was just available up to the following day at 10 AM when a new set then started to be registered since that was the opening time of the stock market operation. Among many stocks available for analysis, we selected Itaúsa (Code: ITSA4.SA; Description: ITAUSA PN), a Brazilian holding company which controls several solid companies including financial, real estate sectors and many other industries (wood panels, health, chemicals, fashion and so on). Moreover, Itaúsa is part of the Bovespa Index (known as IBovespa) which is the main performance indicator of the stocks negotiated at [B]3, and, is composed by the most important companies of the Brazilian stock market (B3[2], 2020). To be part of such an index, the company shall comply with certain criteria since it is a consolidated reference by investors around the world. The current data set is composed by information of 53 business days meaning an amount of 20,431 initial objects (n) which includes initially the following seven attributes (d):

- **Date** (day/month/year) and **Time** (hour: minute).
- **Stock quota**: final price (BRL) of the stock for exchange at each specific minute during the open market (10AM to 5PM).
- **Variation (BRL)**: quota variation (BRL) in relation to the opening value (or the last minute of the previous day in the case of the first minute of each day).
- **Variation (%)**: quota variation (%) in relation to the opening value (or the last minute of the previous day in the case of the first minute of each day).
- **Maximum**: the maximum quota value (BRL) at each specific minute.

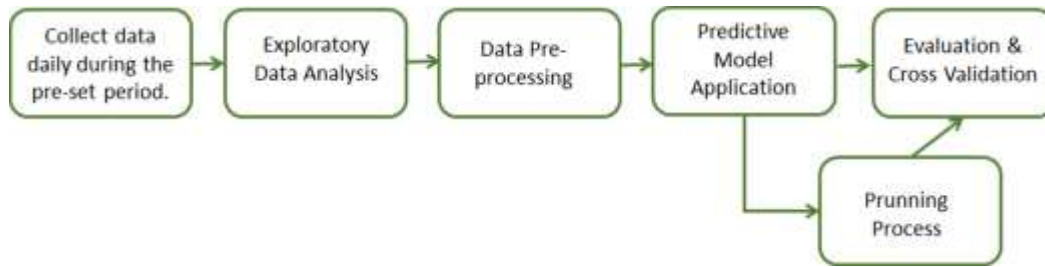


Figure 1. Methodological framework.

- **Minimum:** the minimum quota value (BRL) at each specific minute.
- **Volume:** trading volume (BRL) up to a specific minute (accumulated data).

The target attribute is related to the quota value which is the main interest information about which we want to make predictions. For such a reason, we analyzed especially this subset of date. The unrepresentative variables (attributes) were eliminated during the pre-processing phase.

Methodology

Before presenting the adopted methodology, we defined here our learning task:

In the role of an investor, we want to define if it is a good moment of the day to make a trade and, thus, decide to perform it (or not). By trading we mean the operation of purchasing new stocks or selling the ones we already have aiming to make profit. Considering a specific moment of the day, we will be able to predict if the trade will be performed in an increasing or decreasing variation period. That means it will be a good moment for a sale order or a purchase order respectively.

This is a predictive and supervised learning problem. In relation to the training and testing data, we allocated 80% of them to train our model. The remaining 20% was used for testing the model.

Figure 1 illustrates the sequential methodological framework.

We follow to next section where we put in practice the methodology, presenting and discussing the outcomes step by step.

RESULTS

In this section, we present and discuss the main results of the applied methodology.

Exploratory Data Analysis

As the Brazilian operating day which is based in São Paulo city lasts 7 hours (from 10AM to 5PM), the maximum daily data set will be composed by 420 samples (corresponding to 420 minutes of such period). However, in some moments of the day, there may have been no trading, so that in such minutes, nothing is registered taking to a different number of registers each day. The data that support the findings of this study are available from the corresponding author upon request.

The behavior of a stock on a given day is independent of what happened in the previous days. We can even use the following expression: The new day holds new surprises! Additionally, it is sensitive to political and economic factors that are occurring on that specific day in the company, in the sector to which the business belongs, in the country, and in the world geopolitics as well. This volatility can be clearly seen through the undulation in Figure 2 where each boxplot represents a day analyzed.

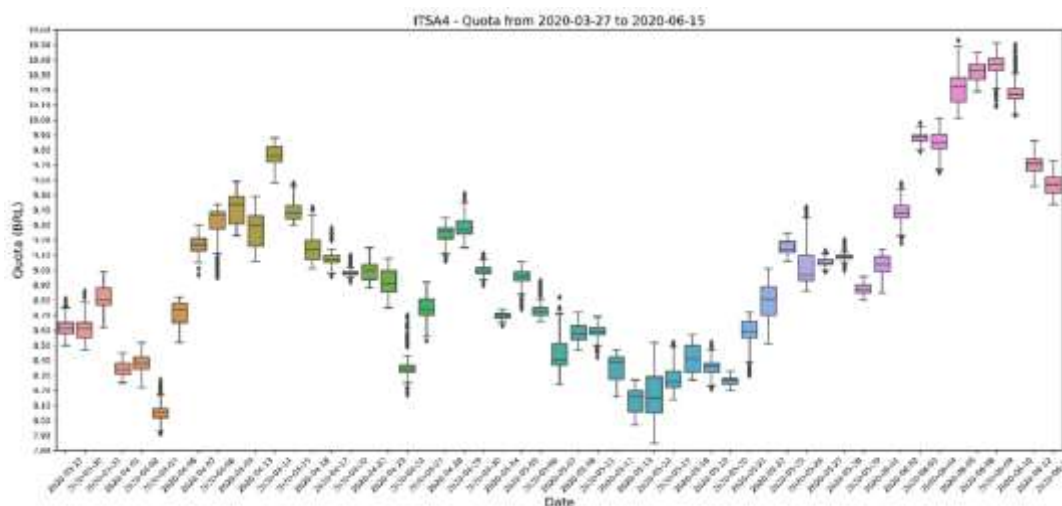


Figure 2: Daily Mean boxplots.

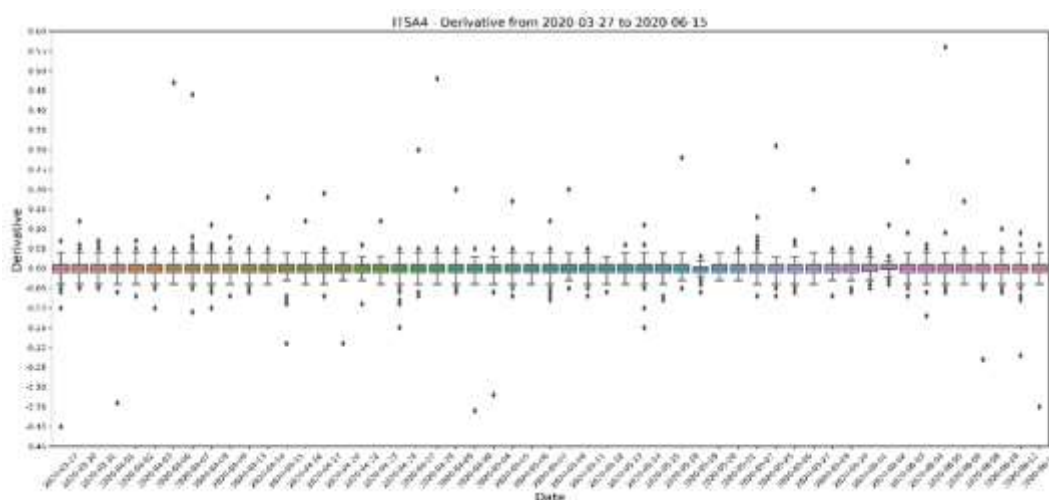


Figure 3: Daily Derivative boxplots.

The time series of intraday stock prices is sensitive to unexpected information and has a very particular behavior, which makes it difficult to be dealt with (PATELL; WOLFSON, 1984). Considering that, we can conclude that in the decision process of our context, the direction of price is more relevant than the price itself. So that, we analyzed the price direction, that is, the derivative of each day, and we found a more uniform data set to work with. Figure 3 shows, also through boxplots, the variations of the stock

price at each day which has a linear pattern over time, but with very particular variations each day, including outliers. This data-processing was better detailed in the pre-processing subsection that follows.

Additionally, we present in Figure 4 the individual histograms of the stock price unified in the same plan. Each day is represented by a different color in this graphic. This shows another perspective for the volatility profile of this market.

decision. They provide short-, medium- and long-

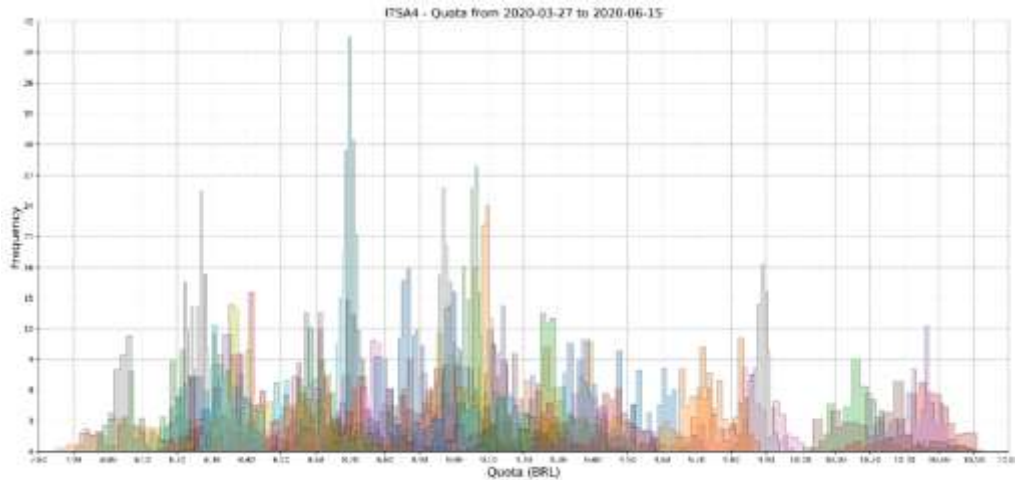


Figure 4: Unified means histograms.

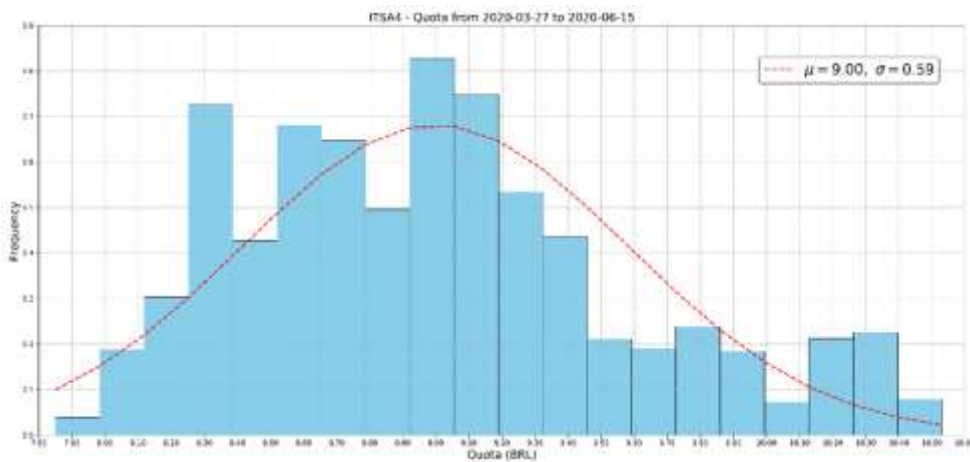


Figure 5: Histogram of Means the entire period.

We added to our data set analysis two other histograms. First, Figure 5 considers the entire original data set joined in a single histogram and which tends to a widely distributed normal curve.

On the other hand, when plotting the corresponding derivatives (direction of the price), we have a more concentrated density distribution normal curve as seen in Figure 6. The reason was previously discussed and is due to the behavior of the stock prices versus the behavior of price direction. We selected to work with the second.

The visual information provided by charts are important tools for analyst sensitivity and

term historical information and shall be followed by the investor. As the data accumulates, and investors can then consider them in other investment strategies such as 'buy & hold', a more conservative (if the stock market can be considered that way) approach and reasonable way to save for retirement for instance (SANDERSON & LUMPKIN-SOWERS, 2018).

Pre-Processing

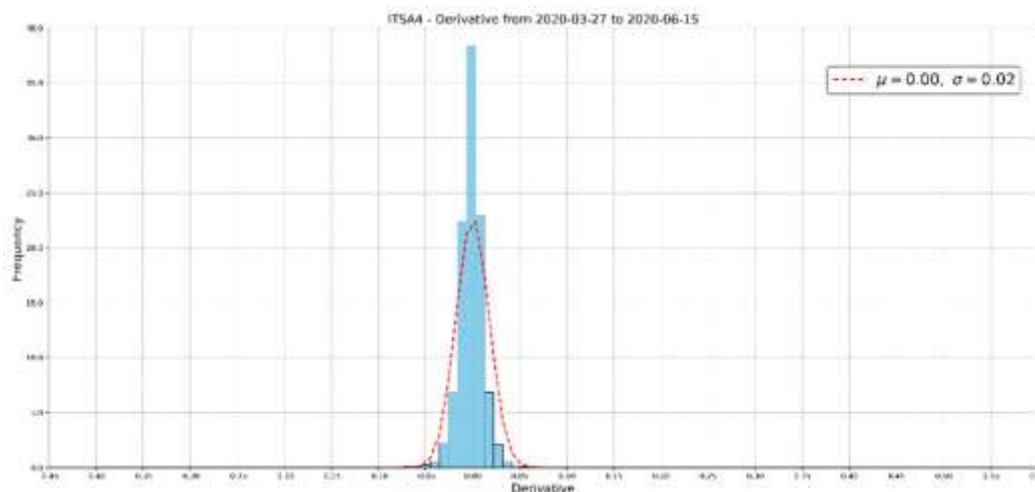


Figure 6: Histogram of Derivatives of the entire period analyzed.

It was not necessary to make so much effort to tidy our data set as it was quite consistent. We are assuming the reliability of our data set firstly because it came from a single source, secondly because it usually feeds stock market specialist analysts. However, we did not escape to perform some fundamental changes to bring sense to our analysis. Before any pre-processing action we manually eliminated unnecessary attributes such as variation (values and percentage), maximum and minimum values, and negotiated volume, keeping only date, time, and stock quota. Additionally, for any investor, looking at the clock and think whether it is a good 'minute' to invest does not seem practical or realistic. It took us for our first pre-processing activity:

- (i) Join the minutes into periods of hours. So, each day of our analysis would be represented by seven points. Each point meaning a period as follows:

- 10.00-10:59 AM: opening [1]
- 11.00-11:59 AM: morning [2]
- 12.00-12:59 AM: lunch [3]
- 1.00-1:59 PM: early afternoon [4]
- 2.00-2:59 PM: middle afternoon [5]
- 3.00-3:59 PM: late afternoon [6]
- 4.00-5:00 PM: closing [7]

The number between brackets is the numerical representation used in the pre-processed data set. As our data set was originally composed by data from 53 business days, this first activity reduced the initial collected 20,431 objects into 371 (n'). The second pre-processing activity was to define the quota value that better represented each period. A measure that is usually used by analysts of the stock market is the most updated value meaning the stock price of the last minute of each period. However, we considered the mean. We chose to use the mean as it better represents a central tendency in the period.

- (ii) Calculate the mean in each hour period: from minute 00 to minute 59.

The third activity was to generate the direction of price of each period instead of working with the value of the stock price. Studies have shown that predicting direction as compared to value can generate higher profits (CHOUDHRY; GARG, 2008) beyond making sense when comparing data from different days.

- (iii) Transform the mean quota price previously calculated into derivative values hour by hour (current in relation to the previous hour). Followed this transformation, we still have numerical values attribute:

$$price\ direction = \frac{\Delta\ quota}{\Delta\ hour} = \frac{\Delta\ quota}{1} = \Delta\ quota.$$

In this study we are dealing with a simple time series with a single variable that iterates over in sequence. To address the dynamic of a series over time, we used a sliding window approach in which we used k-periods worth of data to train the model and then validated the model's performance on data from the succeeding period. We used four different modelling setups (k-values). This process was needed to make data independent and identically distributed (i.i.d.) variables. The sliding window technique determined the target-attributes free of any trend or seasonal component of our data set:

- (iv) Implement *sliding window technique* to arrange the data set in attribute & target-attribute format. We tested four fixed k-sized windows (subset), and chose the one with the best performance:

k = 2; k = 7; and,
k = 5; k = 10.

This process took to a change in the data set size. Table 1 includes the new data set sizes for each attributed k-value, as well the amount used for training (80%) and for testing (20%). The mean and standard deviation of the new data sets were also presented.

Table 1: Data set information after sliding window technique.

	Number of periods (n)			Data sets	
	Total	For training	For testing	Mean	Standard Deviation
original	371	-	-	0.001725	0.104295
k=2	369	296	73	0.002710	0.103199
k=5	366	294	72	0.002705	0.103473
k=7	364	293	71	0.002857	0.103656
k=10	361	291	70	0.002687	0.104042

Prediction Model

Among the regression models available to support our case, we selected Regression Tree techniques since they are quite simple to understand and to interpret, and easily to be

visualized (KOTSIANTIS, 2013). Its logic combines a sequence of simple Boolean tests by comparing attributes. CART (Classification Algorithm and Regression Tree) is a nonparametric approach which finds the best division through analysis of variance criteria (BRIEMAN et al., 1984).

As we are working with a time series, the attributes are all given by the behavior of the direction of price in the previous k periods that we assigned during sliding the window process. Partitions of the trees are achieved by successive binary splits of independent variables based on the predictors from the different k periods. For instance, in the case of a value 2 assigned to k, we have two attributes, given by x1 and x2 to evaluate our target-attribute y, where x2 is the direction price in the moment t immediately before the one to be predicted, and, x1 refers to two moments (t-1) before the predicted period.

We applied the learning algorithm to all proposed k-values. However, a further analysis was made just for the sub-set which had the best performance.

Analysis

We presented in this section, in Figure 7, only the 5-k decision tree which was further explained and explored in the following section. At each node, the number n of cases satisfying the criteria specified is shown. The other trees are available with the authors under request.

Evaluation

We used root mean square error (RMSE) value to compare the performance amongst the different subsets. We observed a different morphology of each tree. It is not a usual item of evaluation but the minor root mean square error (RMSE) value came by chance from the most balanced tree. Finally, Table 2 summarizes train and test data information for the different k-values beyond the RMSE of each subset tested.

Even considering 5-k tree as the model with the best performance in our study, it is far to provide

an advisable result to support the investment decision. The RMSE values are quite high, but realistic when compared to the standard deviations.

Cross-Validation & Friedman Test

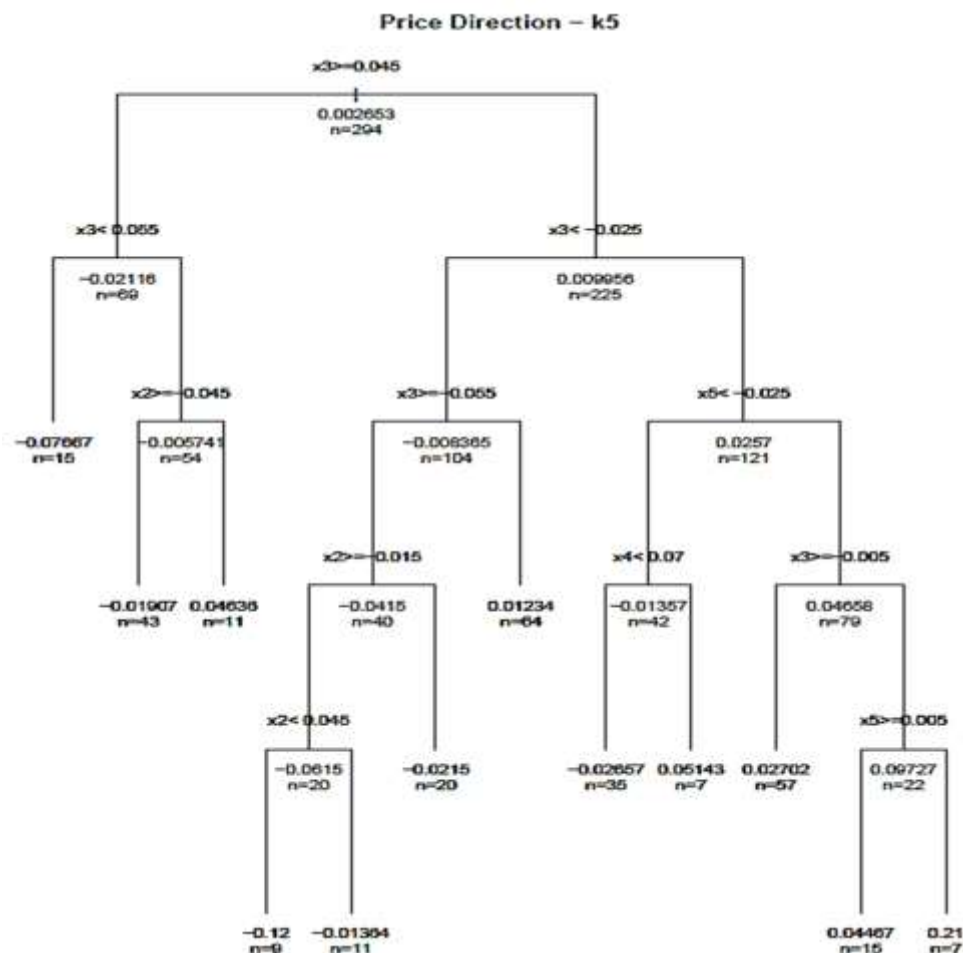


Figure 7: Decision Tree (k=5).

Table 2: Train/test data sets information and RMSE for each k-value.

	k-sized window			
	k=2	k=5	k=7	k=10
Train set				
- Mean	0.00598	0.002653	0.005017	0.002337
- Min.	-0.32000	-0.36000	-0.32000	-0.32000
- Max.	0.63000	0.63000	0.63000	0.63000
Test set				
- Mean	-0.01055	0.002917	-0.006056	0.004143
- Min.	-0.36000	-0.22000	-0.36000	-0.36000
- Max.	0.17000	0.35000	0.17000	0.40000
RMSE	0.1043017	0.1008763	0.1052439	0.1178306

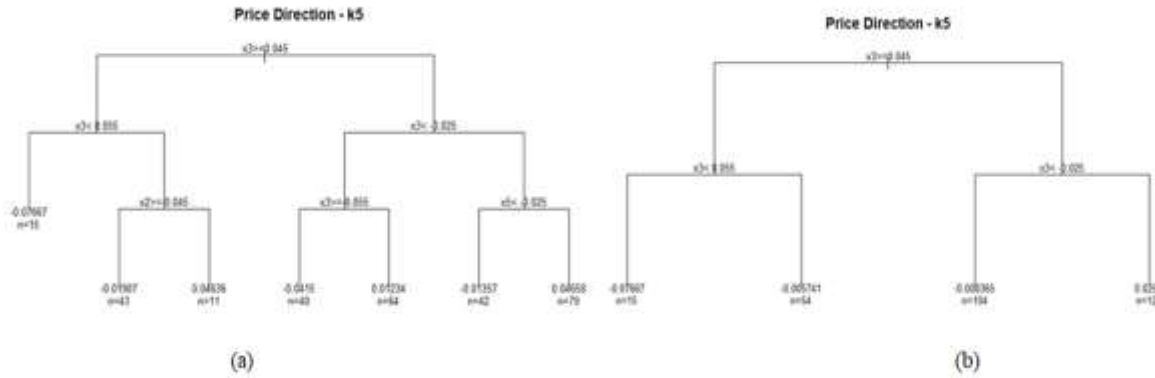


Figure 8: Pruning - (a) level 3 and (b) level 2.

In addition to the previous analysis, we also carried out a comparison for the train / test data set splits through cross validation followed by a statistical test. Cross validation is usually used to compare different machine learning methods. However, in the present work we only used regression tree, so that, cross validation was used to measure the accuracy of the model in different data sets from different k-values. The data set was uniformly at random partitioned into 5 folds of similar size. Table 3 summarizes the scores gotten from a run for each k-sized window based on RMSE.

Table 3: Scores from cross-validation at 5 folds (RMSE).

train/test data sets	k-sized window			
	k=2	k=5	k=7	k=10
1	0.15163758	0.17042078	0.17535483	0.16190222
2	0.15126767	0.1332574	0.16874639	0.17761147
3	0.13611981	0.15468926	0.19072321	0.1724094
4	0.1280169	0.14221824	0.139505	0.12089941
5	0.13649561	0.19523782	0.1264966	0.13450527

We followed by performing Friedman test. The Friedman test tests the null hypothesis that repeated measurements of the same individuals have the same distribution. Back to Table 3, rows r represent the blocks of different train/test data sets, and, the columns k state the distinct conditions and treatments for the data set. The data of test is ranked by block (R_{ij} , $i=1, \dots, r$; $j=1, \dots, K$). The χ^2 distribution with 3 ($=K-1$) degrees of freedom. Pereira et al. (2015) details the test procedure. We aimed to know if there is any difference among the different k-values sets using $\alpha=0.05$ level of significance. So, we have the following hypothesis:

H_0 : there is no difference among the k-sized windows.

H_1 : there is a difference indeed.

Running the test for the 4 distinct k-value set of scores, we had as result the test statistic of 2.520 and a p-value of 0.472. The interpretation is that we shall not reject H_0 meaning that it does not matter k-sized window for a best prediction as all of them has similar performance. In the following sub-section, we continued our analysis just using the 5-k value tree in a pruning process.

4.7. Pruning

We initially allowed the decision tree to grow fully for all k-values and then we calculated the RMSE value for all trees. Considering that the 5-k tree was our best prediction tree model, we performed the pruning process only for such tree. The original 5-k tree has five levels. We simulated two different levels: 3, and then 2 as seen in Figure 7.

Pruning process gives us the possibility to simplify our analysis. In fact, the tree at level 2 which is the simplest possible tree, provided us the best RMSE value as seen in Table 4. We can observe that in such tree only attributes related to x_3 were considered. Remembering that we are working with a time series, and a sliding window at 5-k, that means that two previous period ($t-2$) provides us rich information about our target attribute.

Table 4: RMSE after pruning process at 5-k tree.

	Pre-pruning	Pos-pruning	
	5-k tree	Level 2	Level 3
RMSE	0.100876	0.087280	0.094917

CONCLUSIONS

The objectives of investing shall be clear in particular to the investor him/herself. Obviously, every investor bet with the aim of making money, but for what reason? Before any action, it is necessary to be aware of the future destination of such gains? A not so trivial question, but that will guide the strategy.

Our study focused mainly on the short-term investment which can be both in the intraday and inter-day levels trading. By using a regression tree as our supervised learning method, we intended to assist investors by promoting a validated model to predict the direction of price in a future moment. To learn our model, it was used a 2.5-month data set of a consolidated stock in the Brazilian market.

The adoption of regression tree technique was satisfactory since it is simple to understand and to interpret, and ease to be visualized beyond providing realistic results which reflected how difficult predictions in such markets are indeed.

We tested four different sizes for the time of windows. One of them ($k=5$) showed a slightly better performance. Despite this result, we performed a statistic test to verify it, and we concluded that any k -value would respond in a similar pattern.

We followed by carrying out a pruning process that decreased the complexity of the tree and provided a more reliable outcome. However, a successful model to predict short-term investment behavior remains a challenge. Thus, the expectation for profitable intraday trading keeps being a very risky activity which we do not recommend especially to novice investors.

It is known that breaking news and rumors information quickly impact in the stock prices minute by minute. So, we would suggest that

investors consider different sources of information, for instance textual news, to increment the decision process and support a more conscious trade recommendation (GEVA; ZAHAVI, 2014).

We must assume that prediction for short-term trading recommendation is a quite difficult task by the nature of the business itself. And, in case the long-term investment is not an option (and, in fact, it should be an option, as a growing trend is seen in the Figure 2), we can succumb to the expertise's advice and limit the amount invested to the point of losing its entirety without major damages (HUNGRIA, 2020) or sorrow.

Limitations and Future Avenues

There are endless possibilities of research in the financial market due to the huge amount of data that organizations are making available each day besides the current sophisticated analytical tools. This study covered only one stock in a short period of two and a half months. Other stocks can be followed and for longer periods and then have the performance compared among them. Further, we can also evaluate different strategies, for instance short and long-term such as intraday and buy and hold.

The local and global political-economic scenarios can bring the uncertainties of this market that moves fast, and this was also the case of this period in which the covid-19 pandemic was also an event. However, the data granularity (minute by minute) was able to reflect any ordinary event that really impacted. Despite these limitations, the study achieved its objective of awakening investors to the reality (including the dangers of losses) that intraday investment presents.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCE

- B3[1]. (2020). Histórico Pessoas Físicas. [Online]. www.b3.com.br
- B3[2]. (2020). Índice Bovespa (IBovespa) [Online]. www.b3.com.br
- B3[3]. (2020). Uma Análise da Evolução dos Investidores Pessoas Físicas na B3 [Online]. http://www.bmf.com.br/portal/pages/newsletter/BMFBOVESPA/Downloads/Estudo_PF-final.pdf
- BARBER, B.M, ODEAN, T. (2001). Boys will be boys: gender overconfidence, and common stock investments. *Quarterly Journal of Economics*. February, 261-292.
- BONALDI, E. V. (2018). O pequeno investidor na bolsa Brasileira: Ascensão e queda de um agente econômico. *Revista Brasileira de Ciências Sociais*, 33(97). <https://doi.org/10.590/339710/2018>
- BRIEMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., STONE, C. J. (1984). *Classification and regression trees*. Wadsworth.
- CHAGUE, F., GIOVANNETTI, (2020). É possível viver de day-trading? *Brazilian Review of Finance*. 18 (3), 1-4.
- CHOUDHRY, R., GARG, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. July, 315–318.
- COTIA, A. (2020, October 29). Bolsa alcança 1,5 milhão de investidores pessoa física. Valor Investe. <https://valorinveste.globo.com/>
- ENEF - Estratégia Nacional de Educação Financeira [National Financial Education Strategy]. Decreto Lei No 7.397 de 22/Dez/2010. [Online], (2010). www.planalto.gov.br
- GEVA, T., ZAHAVI, J. (2014). Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision Support Systems*, 57(1), 212–223. <https://doi.org/10.1016/j.dss.2013.09.013>
- HUNGRIA, R. (2020, June 19). Quem disse que o ruído nos mercados atrapalha seus investimentos? Veja como lucrar com ele. [Online]. Seu Dinheiro. www.seudinheiro.com
- KOTSIANTIS, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. <https://doi.org/10.1007/s10462-011-9272-4>
- MANFRINI, S. (2020). 40% dos brasileiros já tiveram perda total ou parcial da renda durante pandemia, aponta CNI. O Estado de São Paulo. Economia. <https://economia.estadao.com.br/noticias/geral,percepcao-de-inflacao-alta-incentiva-procura-por-investimentos-atrelados-ao-ipca,70003431945>
- M1 Finance [1] (2020, October 22). A smart investor’s guide to systems thinking. M1 Finance. <https://www.m1finance.com/blog/a-smart-investors-guide-to-systems-thinking/>
- M1 Finance [2] (2020, October 20). The 5 steps to smarter financial decisions every investor needs to hear. M1 Finance. <https://www.m1finance.com/blog/5-steps-to-smarter-financial-decisions/>
- OLSON, D., MOSSMAN, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, 19(3), 453–465. [https://doi.org/10.1016/S0169-2070\(02\)00058-4](https://doi.org/10.1016/S0169-2070(02)00058-4)
- OSMAN, O. (2020). Using Systems Thinking to Understand Personal Finance. Medium, 16/Dec. <https://oao84.medium.com/using-systems-thinking-to-understand-personal-finance-db2261eda460>
- PATELL, J. M., WOLFSON, M. A. (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics*, 13(2), 223–252. [https://doi.org/10.1016/0304-405X\(84\)90024-2](https://doi.org/10.1016/0304-405X(84)90024-2)
- PEREIRA, D. G., AFONSO, A., MEDEIROS, F. M. (2015). Overview of Friedmans Test and Post-hoc Analysis. *Communications in Statistics: Simulation and Computation*, 44(10), 2636–2653. <https://doi.org/10.1080/03610918.2014.931971>
- PREDA, A. (2006). Socio-technical agency in financial markets:: The case of the stock ticker.

Social Studies of Science, 36(5), 753–782.
<https://doi.org/10.1177/0306312706059543>

ROGERS, P., SECURATO, J. R., RIBEIRO, K. C. de S. (2007). Finanças comportamentais no Brasil: um estudo comparativo. *Revista de Economia e Administração*, 6(1).
<https://doi.org/10.11132/rea.2002.144>

SANDERSON, R., LUMPKIN-SOWERS, N. (2018). Buy and Hold in the New Age of Stock Market Volatility: A Story about ETFs. *International Journal of Financial Studies*, 6(79);
<https://doi.org/10.3390/ijfs6030079>,

SWALES, G. S., YOON, Y. (1992). Applying Artificial Neural Networks to Investment Analysis. *Financial Analysts Journal*, 48(5), 78–80.
<https://doi.org/10.2469/faj.v48.n5.78>

TSAI, C.-F., WANG, S.-P. (2009). Stock Price Forecasting by Hybrid Machine Learning. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009, I*, 2210.

WATSON, W.T. (2018). System Thinking and Investment: Introducing the ecosystem perspective. *Thinking Ahead Institute*, 2/Feb.
<https://www.thinkingaheadinstitute.org/research-papers/system-thinking-and-investment-introducing-the-ecosystem-perspective/>