Decision-Making Simulator for Buying and Selling Stock Market Shares Based on Twitter Indicators and Technical Analysis

Deborah S. A. Fernandes*, Marcio G. C. Fernandes†‡, Geovany A. Borges† and Fabrizzio A.A.M.N. Soares§*
*Universidade Federal de Goiás, Goiânia - GO, Brazil, Email: {deborah, fabrizzio}@inf.ufg.br
†Universidade de Brasilia, Brasilia - DF, Brazil, Email: marcio.giovane@aluno.unb.br,gaborges@unb.br
‡Universidade Estadual de Goiás, Anápolis - GO, Brazil, Email: marcio.giovane@ueg.br
§Departament of Computer Science, Southern Oregon University, Ashland-OR, USA, Email: soaresf@sou.edu

***Abstract*—Microblogs have increasingly been used by the crowd to post their thoughts and speeches about everything. Thus, one of the themes is the stock market that is exploited by many researchers. Although obtaining indicators of stock market dynamics through online social networking has been gaining the attention of academia and the business world, there are many questions to be analyzed about their effectiveness. This work presents the development of a simulator for buying and selling stocks based on microblog data. Therefore, we collected tweets about the Brazilian stock exchange market, produced indicators using sentiment analysis and performed a set of heuristics for decision making. The first technique is the composition of the decision-making strategy for buying and selling stocks composed of pure logic, tweets volume thresholds, profit objective and technical analysis in the stock exchange. The contribution of this work is a decision-making architecture using Twitter's data as an index of future expectation about the social mood that may change the market behavior. As a result, it has pointed to attractive profits for Brazilian market actions and many issues that can be analyzed and improved. Our study showed that it is possible to obtain market dynamics information on the Twitter social network and this information could be used to compose stock buying and selling strategies.**

***Index Terms*—Microblog, Twitter, decision making, stock market, sentiment analysis, Technical Analysis.**

I. Introduction

The financial market suffers at every turn from influences and changes that can hardly be followed in real time without the use of technology. In this way, the price of a stock can be influenced not only by price history or company performance. One piece of news may trigger an avalanche of comments from people who are physically distant but can be perceived through the digital medium. Computational tools that automate the gathering of indicators taken from online data can be important allies of investors for asset decision making.

Many techniques can be employed to track the movement of the financial market for buying and selling stocks. The most widely known are the technical and fundamental analysis. The technical analysis is based on the hypothesis that price movement patterns tend to recur in the future. The fundamental is based on macro indicators and balance sheets of companies to aid in the decision making on assets but also allows to evaluate the monetary value of these [1].

Data generated by online social networking communities have gradually gained credibility as a valid source for stock market analysis [2]. The study results of [3] pointed out that most of the professionals in this area adopt these media in their professional activities. Although many authors as [2], [4], [5] have recognized that trend, they identify questions about the accuracy and significance of models and report that there is much to be done to achieve effective predictions.

Behavioral economics states that emotions can profoundly affect investor behaviors and their decision making [6], [4]. If the individual investor's emotion may affect how he/she responds to a new information, it is likely that the collective sentiment of investors may influence stock market dynamics [7]. As a consequence of this thought, measuring social mood has become a key issue in financial forecasting research [8].

In this work, we investigate the use of Twitter microblogging messages in Portuguese to estimate the dynamics of the social mood that can influence the stock market behavior and the strategy of decision to buy or sell stocks. The novelty of this work is to use Twitter data as a source of some composing indicators on a making-decision simulator system to buy/sell stocks. Thus, our major contributions are listed as follows: (1) an architecture of decision-making system based on indicators obtained from online social networks data for the stock market; (2) strategies for buying and selling stocks based on tweets and technical analysis; (3) a set of heuristics for decision making.

This paper extends our work as follows. First, we had proposed a dataset with tweets and daily (and historical) stock market price in Brazil, section III-A. Second, We created indicators used in a decision-making system, described in sections III-B and III-C. Third, we propose a rule-based simulator for buying and selling of the stock market, section III-D. Finally, the results have presented in section IV. Conclusions and future work in section V.

II. Related Work

Microblogs, online forums, are increasingly being used by the masses to post their thoughts and speeches about everything. One of the themes is the stock market that is exploited by many researches.

Many authors have explored the study of the effectiveness of the data obtained from news, microblogs and online social networks in the predictability of the stock market financial market. Using mining techniques in news articles Schumaker and Chen (2009) [9] researched the prediction of current stock prices listed on S & P 500. Other authors such as Geva and Chen (2014)[10] and Xiaodong et. al. (2014) [11] studied the impact of news on financial market returns. As results, they have shown that the analysis of these sources can help predict market movement.

Checkley, Higón and Alles [12] focused intraday market data. To predict the stock market they sampled-every-two-minutes data on price direction, price volatility and trading volume and match then with sentiment metrics extracted from micro-blogging sites. They find evidence of sentiments causing market behavior, especially over time horizons of minutes, rather than hours or days. For them, sentiment has some short-term predictive value, but forecasts are better in the cases of predicting volatility and trading volume, than price direction.

Stock market predictability was explored by Oliveira, Cortez and Areal [13]. They found that Twitter sentiment and posting volume were relevant for the forecasting of returns of the S&P 500 index and others. Their results confirm the usefulness of microblogging data for financial expert systems, allowing to predict stock market behavior and providing a valuable alternative for existing survey measures [13].

Zhang, Xu and Xue [14] presented an investigation on the correlation of the collective feeling coming from large scale networks feeds to the stock transaction data over time. They proposed a model that divides the information space into two, one of the beliefs and the other the realistic transactions. Their results point out that the predictability of financial data can be improved with the model. In [4], Bollen, Mao and Zeng correlated moods of collective humor through Twitter data with the value of the DJIA index (Dow Jones Industrial Average). They showed that the accuracy of DJIA predictions can be improved by including public mood measures.

Daniel, Neves and Horta [15] studied the development of a system to detect and discover the importance of special events in the financial area through the Twitter network. They had focused only on events that could change the evolution of specific stocks. Their results have proven the potential influence of the finance community on Twitter in regards to the publication of tweets about companies. The system developed was able to detect financial events.

III. DATASET AND DEVELOPED ARCHITECTURE

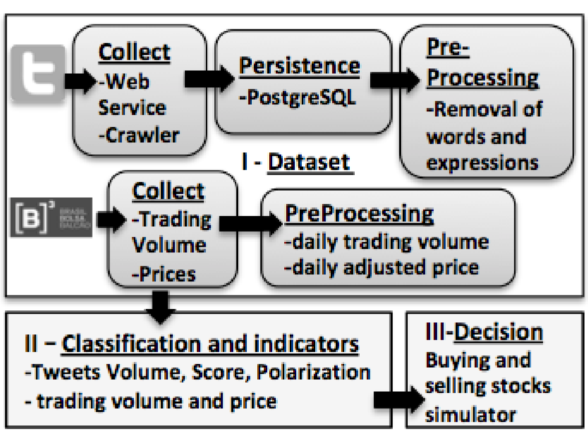


Fig. 1: Architecture model.

This section describes the architecture model and data set used for the study. We proposed a framework in three stages, Figure 1. The first deals with dataset (data from Twitter [1] and data from Brazilian Stock Exchange - B3 [2]). Both of them are pre-processed. The second one obtains indicators based on tweets and the stock market. The third represents the decision-making process which simulates the purchase and sale of stocks.

A. Dataset and Pre-processing

The dataset formed for the experiment is composed of: (1) tweets collected through API Twitter and stored in a PostgreSQL database; (2) tables of price data and trading volume obtained from the stock exchange B3. The data domain for the collection is composed of important Brazilian companies that have shares in the stock exchange shown in the Table I.

TABLE I: The selected Brazilian companies.

	Company	Operation	Stock exchange shares
(1)	Vale S.A.	mining	VALE3 and VALE5
(2)	Companhia Siderúrgica Nacional	mining	CSNA3
(3)	Usiminas	steel segment	USIM5
(4)	Guerdau	steel segment	GGBR3 and GGBR4
(5)	Itaú Unibanco	financial	ITUB3 and ITUB4
(6)	Banco do Brasil	financial	BBAS3
(7)	Petrobrás	oil exploration	PETR4 and PETR5

The dataset was collected from August 13, 2013, to May 4, 2015. The total of tweets for all the companies previously described was 8,144,657. For the experiment, we chose to make use of data obtained on trading days. That is, trading days on the Brazilian stock exchange (from Monday to Friday, except holidays), discounting those presented technical problems to be collected, we gathered 426 days.

After a visual analysis of the database, it was noticed that there were no tweets for the companies (2), (3), (4), (5) and (6) arranged in table I. However, to perform the experiments it was necessary to have daily tweets for each company. Understanding that the Brazilian user was not sharing about those companies at that time, we chose to continue the work with which we had the data. Consequently, to continue the work, we chose to use only those that had a daily collection of messages. They were the stocks VALE5 from *Vale S.A.* and PETR4 from *Petrobrás*.

Twitter is a quite noisy [1] online social network [1] . To perform automatic sentiment analysis of data, it is essential for employing pre-processing techniques. Some usual techniques such as removal of stop-words, punctuation, links, and retweets [16] had evaluated for use. We use punctuation filter. After the visual analysis of the database, we found that many messages containing links also contained useful information about the market. We also assume that the retweets, in the

[1] Twitter - An online microblogging platform was chosen for limiting the size of messages by 280 characters, has mostly public profiles and the provision of an API that allows real-time collection HTTP://www.twitter.com.

[2] B3 -*Brazil Bolsa Balcao* - It is the Brazilian Stock Exchange available in HTTP://www.b3.com.br/en_us/.

analyzed context, expressed a reinforcement of the opinion that was being shared. That way, if a tweet is positive and a person performs a retweet, the second person is reinforcing the thought of the first and should be kept in the bank.

So, as pre-processing data, we analyzed data samples to capture irrelevant expressions and words as notification of sports events and musical shows sponsored by the companies to remove them from data. This blacklist contains 412 items. The database contained 51,849 and 23,430 tweets for PETR4-*Petrobrás* and VALE5-*Vale S.A.*, respectively. After pre-processing, they had 28,459 and 13,366.

The data used from the stock exchange were historical and closing stock prices. We applied logarithmic scale normalization for daily returns r_t. Return is the percentage change in asset value. There is scant evidence of the predictability of a return [7]. However, it provides useful information about the probability distribution of the asset price. It is calculated on the adjusted closing price (for dividends, interest rate on capital, splits and groupings) of the day p_t subtracted from the previous day price p_{t-1}, Eq. (1) [7], [16].

$$r_t = ln\,(p_t) - ln\,(p_{t-1}) \tag{1}$$

B. Twitter Indicators Modeling

We formed four indicators from the data collected from Twitter. Three based on a naive count and one on sentiment analysis. For the naive counters, we had used the raw database. That's because we were interested in the number of tweets containing mentions to some term. Although the data is collected uninterrupted on the Twitter platform, the analysis done is per day. That is, the indicators express the day t, they are:

- B_t: Related to Twitter buzz, it represents the tweets volume posted on day t, adopted by several authors [8], [2], [7].
- M_t: It represents the day mood, optimistic to buy and pessimistic to sell stocks. It is the volume of tweets containing the words radicals in Portuguese to buy and sell. $H_t = BUY_t - SELL_t$, positive when is optimistic to buy, pessimistic when negative and neutral if zero.
- E_t: Points out how many tweets contain words and trend-indicating expressions on day t to bullish and bearish. Such words and phrases were obtained from a financial market specialist. $E_t = HIGH_t - LOW_t$ it is positive to bullish, negative to bearish.
- S_t: It represents day sentiment obtained through sentiment analysis (details below). $S_t = Positive_t - Negative_t$.

The sentiment analysis (SA) has several levels of classification, one of them is the textual identification in positive, negative and neutral [17]. Considering the granularity, the SA acts on document levels, sentence and aspects. It adopts linguistic resources to annotate documents with sentiment labels. It is a "multi-faceted problem including many sub-problems, but not a single task" [18]. From the perspective of the methods, it is considered the supervised, the unsupervised and the hybrid (or semi-supervised) [17].

In the proposed architecture, document-level and supervised methods were used. Each tweet is considered a document because it is atomic in its semantic context, enclosing in it all necessary elements and objectives for the identification of the global sentiment.

The 280 character limitation, the informal text in linguistic aspects, the abbreviations, the sarcasm, the emoticons, the curse words and jargon practiced by the financial market players strongly restrict the generality of the SA, leading to a supervised approach to the classification of polarity. The approach of the other methods will be considered in specific future works.

The purpose of SA is an indicator of a measure of collective expectation for the respect of a subject or object of investment. In addition to technical analysis, it is public and notorious as a random reaction of the stock market, facing a prospective prospect perspective, may benefit or undermine an investment strategy.

We did the sentiment analysis using the free and multilingual Lingpipe [3] tool.

C. Stock Exchange Indicators

The stock purchase and sale indicators had obtained through the application of technical analysis methodologies. There are several trend and oscillators indicators. The first help to detect market slopes. The second one aids in the discernment of levels of support and resistance. We used trend indicators that were adopted by traders:

- Exponential Moving Averages (EMA_t) - It is the weighted average of past observations. It promotes the smoothing of the number of buys and sells signals present in the average stock price chart. It allows identifying the trends price of the asset, whether high or low. It also serves as a support when the price is above average, and resistance when the price is below the average [19]. Defined by $EMA_t = (1-k) * EMA_{t-1} + k * P_1$ on day t, considering n as the number of days evaluated and $k = 2/(n+1)$ and P_1 the current asset price. This indicator emphasizes the most recent values because they receive greater weight. The older ones have weights that fall exponentially. Thus, the indicator will react more quickly to a change in current prices. We adopted the popular short average EMA values for the periods of 5 and 20 periods for the system [19].
- Moving Average Convergence / Divergence ($MACD_t$)- An indicator used to observe a trend reversal point. It helps in the making decision to the right moment to enter or exit the market. It is composed of MACD and signal lines [20]. The MACD is formed by the difference between two exponential averages, one of short and one of longer-term (usually 12 and 26 periods [20]): $MACD =$

[3]Library containing a set of word processing tools. Available in http://alias-i.com/lingpipe/

$MME[12] - MME[26]$. The signal is computed by the exponential moving average of nine periods of the values obtained for the MACD line: $Signal = MME[9]_{MACD}$. The crossing of the signal (slower) and MACD (faster) lines reflect changes in the balance of forces between buyers and sellers. When the MACD line crosses the signal upwards then the market is conducive to buy, the opposite happens if the MACD crosses the signal down.

D. Decision Simulator

The simulator architecture is presented in Fig. 2.

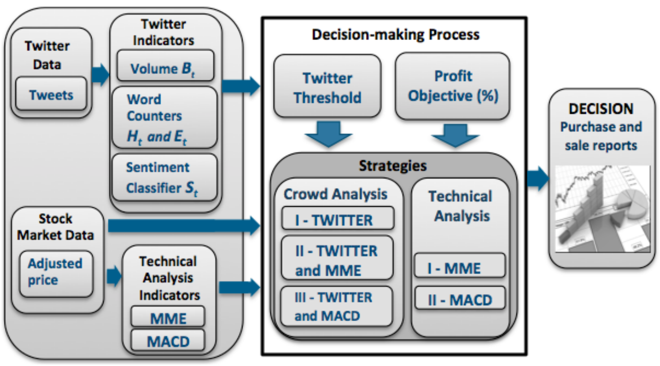


Fig. 2: Architecture Simulator.

1) Inputs: As input it receives the data:

- *Stock_market* - the adjusted closing prices p_t, returns r_t as in Eq. (1) and calculates the $EMA[5]$, $EMA[20]$, and $(MACD, Signal)_t$ for each day t of the data window chosen;
- *Twitter_Indicators* B_t, E_t, Mt and S_t;
- *Threshold* - It is the minimum amount of tweets to be considered per day to perform an operation. We randomly choose the values 0 and $\frac{1}{n}\sum_{n}^{t=1} B_t$ with n samples. B_t is the indicator which reflects the total amount of tweets collected on day t. This variable sets a minimum security threshold for the number of tweets. Ensures that the architecture discards days with a null or tiny amount of tweets.
- *Profit_Goal-PG* - We randomly choose the values $0\%, 3\%, 5\%, 8\%$ or 10%. The profit objective is an important item in the decision to buy and sell stocks, as it establishes the level of risk required of the investor.
- *Initial_Date* and *Final_Date* - The considered period of time.

2) Strategies: In the decision-making process, there are five strategies for buying and selling stocks. They had divided into two categories. One uses tweets and is called crowd analysis, the other is technical analysis - no tweets. Once a strategy is chosen, the simulator starts its activity. They are:

- Crowd Analysis: (I)Twitter - Tw - the system only uses tweets; (II) Tw + EMA; (III) Tw + MACD;
- Technical Analysis: (I) EMA; (II) MACD.

The system allows the user to choose crowd analysis with or without a tweet threshold for the evaluated day. It is the *Twitter Threshold* on Fig. 2. Another possibility of choice is the *Profit_Goal*.

3) Module Simulator: The simulator module had presented on ALG. 1. The *Status* variable reflects the state of the simulator initially NULL. It informs the system that there is no purchase operation in progress. The iteration structure is executed while the Data (initially *Start_Date*) does not reach the *End_Date* selected by the user. The structure of the simulator allows the purchase of a single stock at a time. The *Status* variable is checked at each iteration. Being its content NULL or SOLD, the Decision module (ALG. 2) verifies the possibility of acquiring a stock. In this case, if the decision module returns BOUGHT, *Status* becomes PURCHASED consequently, purchase data is stored. If the *Status* variable is setted as PURCHASED, then the possibility of selling is verified through the Decision module. Being possible, *Status* is changed to SOLD and the sales data is stored. At the end of the repeat structure, the Data variable for SOLD and BOUGHT are updated.

ALG. 1: Module Simulator.

```
Data: Stock_Market,Twitter,Threshold, Profit_Goal,
      Initial_Date,Final_Date
Result: Report(Bought_Data,Sold_Data)
Date=Initial_Date
Status = NULL //Simulator status: BOUGHT/Sold.
Strategy_option = STRATEGY_CHOICE // Twitter, MME, MACD,
  Twitter with MME or MACD.
while Date ≠ Final_date do
    if Status ≠ BOUGHT then
        if DecisionToBuy(Date, Strategy_Option, Stock_market,
          Twitter, Threshold, Profit_Goal) = BOUGHT then
            Status = BOUGHT;
            Store(BOUGHT_data);
    if Status ≠ SOLD then
        if DecisionToSell(Date, Strategy_Option, Stock_market,
          Twitter, Threshold, Profit_Goal) = SOLD then
            Status = SOLD;
            Store(SOLD_data);
```

ALG. 2: Module Decision.

```
Data: Date, Strategy_option, Stock_Market, Twitter, Threshold,
      Profit_Goal
Result: BOUGHT, SOLD or NULL
if (Status ≠ BOUGHT) then
    //For Strategy_Option = Twitter its necessary to //check the
     threshold. If technical analysis, //Twitter data is null.
     if Strategy(Date, Stock_Market, Twitter) = BUY) then
        return BOUGHT
else
    if Profit_Goal > 0 then
        if Strategy (Date, Stock_Market, Twitter) = SELL) AND
          Probable_Profit(Date, stock_market)>=Profit_Goal then
            return SOLD
    else
        if Strategy (Date, Stock_Market, Twitter) = SELL) then
            return SOLD
return NULL
```

The decision module receives as input the date, strategy, stock and twitter data, the threshold and the profit objective.

The invocation of the module returns BOUGHT, SOLD, or NULL. If the invocation of the decision module was to check the possibility of purchasing (ALG. 2) and the strategy is Twitter then the variable *Threshold* is queried. Be its value greater than zero, a purchase can only be made if the total of tweets collected for the day t is at least equal the number of tweets pointed out by *Threshold*. Being zero any amount of tweets can be used to make a decision to buy and sell a stock. If the decision module invocation was to check the possibility of a sale and the strategy is Twitter, then the variable *Profit_Goal* is queried. A sale will only be performed if the strategy indicates it and if the *Profit_Goal* is in agreement with the one desired by the user. If the Strategy module return (ALG. 2) is different from "BUY" or "SELL", then the Decision module will return "NULL".

TABLE II: Decision Making Rules.

Indicators			**Rules**	**Action**
E	**M**	**S**		
	B	POS	If (H and Pos) then	BUY
	S	POS	If (H and Pos) then	BUY
	N	POS	If (H and Pos) then	BUY
	B	NEG	If (H and B) then	BUY
H	S	NEG	If (S and Neg) then	SELL
	N	NEG	-	NIL
	B	N	If (H and B) then	BUY
	S	N	-	NIL
	N	N	-	NIL
	B	POS	If (B and Pos) then	BUY
	S	POS	If (L and S) then	SELL
	N	POS	-	NIL
	B	NEG	If (L and Neg) then	SELL
L	S	NEG	If (L and Neg) then	SELL
	N	NEG	If (L and Neg) then	SELL
	B	N	-	NIL
	S	N	If (L and S) then	SELL
	N	N	-	NIL
	B	POS	If (B and Pos) then	BUY
	S	POS	-	NIL
	N	POS	-	NIL
	B	NEG	-	NIL
N	S	NEG	If (S and Neg) then	SELL
	N	NEG	-	NIL
	B	N	-	NIL
	S	N	-	NIL
	N	N	-	NIL

The technical analysis strategies EMA and MACD can be performed by the decision module. In any of these options, the module receives the stock exchange data and calculates the averages. Indications for purchase and sale are configured at the crossing of averages. If the user chooses the option TWITTER, the system applies the heuristics developed in this work and available in TABLE II for decision making. These heuristics take into account the agreement of at least two indicators. if all indicators disagree, no action is taken. The sentiment indicators are the ones defined in the Subsection III-B . **E** for the price trend (it can be H-high, L-Low or N-Neutral), **M** for the mood (it can be B-Buy, S-Sell or N-Neutral) and **S** for the sentiment (it can be POS-positive, NEG-Negative or N-Neutral). There is also the option of adopting the crowd and technical analysis strategies together. In this case, if the option selected is Twitter with EMA or Twitter with MACD a purchase or sale will only be performed if both strategies return the same indication.

TABLE III: Results for VALE5.

Strategy	**PG**	**Threshold = 0**				**Threshold = 15**			
		Min (%)	**Mean** (%)	**Max** (%)	**SD** (%)	**Min** (%)	**Mean** (%)	**Max** (%)	**SD** (%)
	0	-74	29	196	35	-69	18	139	46
	3	-77	36	224	37	-74	31	195	41
Tw	5	-69	20	147	41	-63	15	119	42
	8	-71	25	171	40	-63	15	121	42
	10	-71	25	171	40	-63	15	121	42
	0	-47	9	77	38	-42	4	64	69
	3	-41	1	58	72	-37	1	50	58
Tw +	5	-41	1	58	71	-37	1	50	59
EMA	8	-41	0	56	71	-37	0	50	59
	10	-42	1	61	71	-37	0	48	62
	0	-45	-1	47	61	-48	0	52	74
	3	-52	4	69	34	-55	6	75	20
Tw +	5	-50	2	62	53	-55	6	75	20
MACD	8	-49	14	58	62	-54	4	70	29
	10	-46	9	49	53	-49	24	56	62
EMA	Any	-45	-1	47	62	-	-	-	-
MACD	Any	-43	3	65	65	-	-	-	-

Any - Same value to any Profit Goal-PG.

4) Outputs: The simulator output consists of a table with data of all the buying and selling operations performed by the system during the requested time interval. The investor can act in the market following the tendency of indicators or counter-trend. In both can be positioned as (1) Bought; (2) Sold, or (3) Bought/sold. (1) In this, the investor acquires an asset in the low of the price with the expectation to sell if there is a rise in prices. (2) It expects to profit from lower prices. In this type of operation, the asset is sold when the price is high because it is expected to buy it again when the price falls. (3) The gain is through the loan or stock rental. If the investor does not have the share to sell he borrows it by paying a rental fee to the owner, then he sells it. Being sure that the stock price will fall the investor buys it back for a lower price and returns it to the borrower. The investor earns with the fall of an action that at first he did not have, initially assuming the position of SOLD and later of BOUGHT when acquiring the action again.

The system will issue a set of data for each simulation with a period determined by the start date and end date. The simulation results are presented for all types of actuation described in the last paragraph, such data will be described in the next section.

IV. Results and Discussion

The results refer to the data collecting interval between August 13, 2013, and May 4, 2015. We considered for collecting only exchange trading days (from Monday to Friday, except holidays), discounting those presented technical problem to be collected, we gathered 426 days, in about of 50,000 tweets for PETR4 and 13,218 for VALE5.

For the variables *Threshold* and *Profit_Goal* we use the values as described in section III-D1. In the technical analysis, we used the returns specified on Eq. (1) to obtain the indicators

TABLE IV: Results for PETR4.

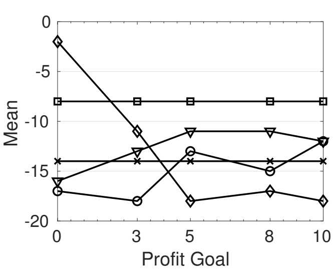
Strategy	PG	Threshold = 0 Min (%)	Mean (%)	Max (%)	SD (%)	Threshold = 40 Min (%)	Mean (%)	Max (%)	SD (%)
Tw	0	-49	-16	18	27	-85	31	222	42
	3	-60	-13	31	38	-86	34	235	39
	5	-64	-11	38	44	-88	43	278	33
	8	-64	-11	38	44	-88	41	267	35
	10	-62	-12	35	41	-88	43	277	34
Tw + EMA	0	-79	-17	35	45	-80	-15	38	78
	3	-81	-18	37	52	-82	-5	85	46
	5	-83	-13	58	61	-82	3	78	24
	8	-80	-15	44	54	-82	-8	73	44
	10	-81	-12	58	60	-83	-4	91	45
Tw + MACD	0	-58	-2	88	51	-70	-2	162	44
	3	-43	-11	18	25	-45	-16	0	68
	5	-32	-18	-3	12	-32	4	45	71
	8	-32	-17	-2	13	-41	-16	-3	12
	10	-34	-18	-2	13	-47	-15	4	12
EMA	Any	-80	-14	44	48	-	-	-	-
MACD	Any	-49	-8	57	35	-	-	-	-

Any - Same value to any Profit Goal.

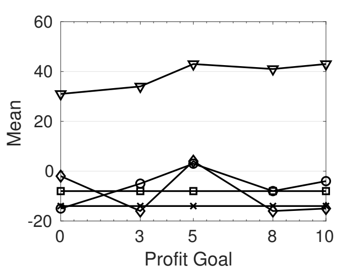
described in section III-C. For technical analysis there is no use of tweets, so there is no data for the Threshould variable.

Simulations were performed for each strategy (Tw, Tw + EMA and Tw + MACD) and produced 60 results each, half/half with and without threshold for tweets. For the technical analysis strategies (EMA and MACD) we obtained 30 outputs for each. We tested all strategies of purchase (Section III-D2) with all the profit goal possibilities in the interval of 426 days. The same simulation conditions had performed for PETR4 and VALE5. As outputs we had obtained the measurements for the investor acting such as BOUGHT, SOLD AND BOUGHT/SOLD following or not the trend (Section III-D4).

TABLES III and IV present the minimum, average, maximum and standard deviation (SD) accumulated profit values obtained from all simulations carried out under the described conditions. PG is the profit goal. The sets of accumulated profit values had obtained for each PG were submitted to the Shapiro-Wilk normality test, and all fit to normal distribution.



(a) Threshold= 0



(b) Threshold=40

Fig. 3: Profit Averages for PETR4. $\bigtriangledown$ - Tw, $\bigcirc$ - Tw + EMA, $\diamondsuit$ - Tw + MACD, $\times$ - EMA, and $\square$ - MACD.

In Fig. 3(a) and Fig. 4(a) it is possible to observe the average of the accumulated profits for each *Profit_Goal* without the use of tweets threshold. In the first case, PETR4, there is no profitable strategy. The opposite occurs with VALE5 which presented most of the strategies as profitable, even varying the *Profit_Goal*. In this case, the Twitter strategy had the best results. An interesting factor, even with PETR4 having more tweets, in case of threshold $= 0$, VALE5 did better in the simulation.

The best results for PETR4 were obtained using the threshold variable, 3% of *Profit_Goal* and with Twitter strategy. The crowd analysis with technical analysis options was profitable with only 5% profit-goal.

In Fig. 4 (b) are shown the averages of the accumulated profit of VALE5 with threshold. Although the Twitter strategy remains profitable, using the threshold the profits have declined.

In general, the proposed architecture demonstrates the superiority in relation to traditional methods by maximizing profits by considering only the perceived expectation of social humor through Twitter. It is important to note that the results obtained in the tables do not take into account the costs of financial operations. A next step in the development of architecture is to consider them.

Another important issue is the effect on the results observed because of the relationship between the threshold setting and the volume of tweets collected. In the case of the PETR4 action, where the volume of the tweets is high, the application of the threshold was determinant for the good average result of the profit obtained in the simulation. In the VALE5 action, where the volume of tweets was reduced, the use of the threshold worsened the average results of the profits obtained. Thus, it was clear the importance of this item for the composition of the decision-making architecture, pointing to the need for a study that proposes a better model for the calculation of the threshold.

Regarding the profit objective, a careful study should be carried out to elaborate a model that defines a possible causal relationship between it, the algorithms of the architecture and the profit achieved.

Our study had a limitation that it has taken into account a small sample of stock market: PETR4 and VALE5. Although this limitation, it points good insights into the variables‘s analysis because of distincts presented features in the tweets sets related to the couple of sample.

V. CONCLUSIONS

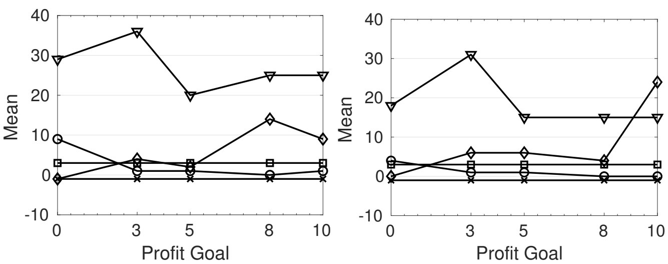
The objective of this work is to investigate indicators from Twitter in the Portuguese language that could help in the decision-making process of a computational system, this had reached.

With the development of the research, we realized the potential of the information obtained through the Twitter network for decision making. As a result shows many values can be better configured and tested.

This study showed that it is possible to obtain market dynamics information on the Twitter social network and this information could be used to compose stock buying and selling strategies.

This research becomes promising as the use of online social networks, especially Twitter, has strengthened since the beginning of this research. In addition, Brazil has been undergoing changes in political and economic strategies since 2016 leading to a medium-term change in the behavior and interests of Brazilians in these issues. This behavior is observed with the increase in the volume of collections that continue to be performed.

Future works include: (1) improving noise removal on Twitter database; (2) improve sentiment analysis to obtain more accurate sentiment indicator; (3) study and calculate brokerage expenses; (4) add other indicators to the framework (e.g. Google Trends); (5) improve the decision process by adding machine learning and (6) develop day-trade in the framework.



(a) Threshold= 0 (b) Threshold=15

Fig. 4: Profit Averages for VALE5. ∇ - Tw, $\bigcirc$ - Tw + EMA, $\Diamond$ - Tw + MACD, $\times$ - EMA, and $\Box$ - MACD.

References

[1] R. F. H. N. Daniel, M.; Neves, "Company event popularity for financial markets using Twitter and sentiment analysis," *Journal of Expert Systems with Applications*, no. 71, pp. 111–124, 2017.

[2] P. Dondio, "Predicting Stock Market Using Online Communities Raw Web Traffic," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 230–237, Ieee, Dec. 2012.

[3] H. Chen, P. De, Y. J. Hu, and B.-H. Hwang, "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *Review of Financial Studies (RFS), Forthcoming.*, December 2013.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, pp. 1–8, Mar. 2011.

[5] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction," in *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 800–807, Ieee, Dec. 2011.

[6] John R. Nofsinger, "Social Mood and Financial Economics," *Journal of Behavioral Finance*, vol. 6, no. 3, pp. 144–160, 2005.

[7] N. Oliveira, P. Cortez, and N. Areal, "Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from Twitter," in *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics - WIMS '13*, (New York, New York, USA), p. 1, ACM Press, 2013.

[8] H. Mao, S. Counts, and J. Bollen, "Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data," 2011.

[9] R. P. C. H. Schumaker, "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems*, no. 27(2), pp. 1–19, 2009.

[10] J. Geva T.; Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," *Decision Support Systems*, no. 57, pp. 212–223, 2014.

[11] X. H. C. L. W. J. D. X. Li, X., "News impact on stock price return via sentiment analysis," *Knowledge-Based Systems*, no. 69, pp. 14–23, 2014.

[12] H. A. M. S. Checkley, D.Añón Higón, "The hasty wisdom of the mob: How market sentiment predicts stock market behavior," *Expert Systems with applications*, vol. 70, pp. 256–263, 2017.

[13] N. A. Nuno Oliveira, Paulo Cortez, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," *Expert Systems with applications*, vol. 73, pp. 125–144, 2016.

[14] Y. X. Gaowei Zhang, Lingyu Xu, "Model and forecast stock market behavior integrating investor sentiment analysis and transaction data," *Cluster Comput*, vol. 20, pp. 789–803, 2017.

[15] N. H. Mariana Daniel, Rui Ferreira Neves, "Company event popularity for financial markets using Twitter and sentiment analysis," *Expert Systems with Applications*, vol. 20, pp. 111–124, 2016.

[16] M. Arias, A. Arratia, and R. Xuriguera, "Forecasting with twitter data," *ACM Transactions on Intelligent Systems and Technology (TIST) - Special Section on Intelligent Mobile Knowledge Discovery and Management Systems and Special Issue on Social Web Mining*, vol. 5, no. 1, 2013.

[17] B. Liu, *Sentiment Analysis and subjectivity - Handbook of Natural Language Processing*. 2010.

[18] C. W. L. X. Z. W. . Y. M. Yue, L., "A survey of sentiment analysis in social media.," *Journal of Knowledge and Information Systems*, 2018.

[19] StockCharts, "Moving Averages - Simple and Exponential [Online]." Available: https://stockcharts.com/school/doku.phpid=chart_school:technical_ indicators:moving_averages, February, 2019.

[20] StockCharts, "MACD (Moving Average Convergence/ Divergence Oscillator) [Online]." Available: https:// stockcharts.com/school/doku.phpid=chart_school:technical_indicators: moving_average_convergence_divergence_macd, February, 2019.