Directional classification of cryptocurrencies using stock indices: the case of Bitcoin

Chasiotou Valentina

SCHOOL OF ECONOMICS, BUSINESS ADMINISTRATION & LEGAL STUDIES

A thesis submitted for the degree of
Master of Science (MSc) in Banking and Finance

December 2019
Thessaloniki – Greece

Student Name:	Valentina Chasiotou
SID:	1103170001
Supervisor:	Prof. Periklis Gogas

I hereby declare that the work submitted is mine and that where I have made use of another's work, I have attributed the source(s) according to the Regulations set in the Student's Handbook.

December 2019
Thessaloniki - Greece

Abstract

This dissertation was written as part of the MSc in Banking and Finance at the International Hellenic University.

The objective of the current thesis is an effort to examine the relationship between Bitcoin prices and stock market indices, as well as with stock prices of big capitalization companies.
More specifically, it is an attempt to build a forecasting model that predicts Bitcoin prices to the highest accuracy that could be achieved. Since Nakamoto (2008) introduced the Bitcoin, the financial sectors have been struggling to comprehend this financial instrument, as its complexity seems to draw the attention of both investors and researchers. Due to its extreme volatility and compound nature, the engaging in such a trading, requires a more detailed evaluation and knowledge before stepping into the transaction environment.
As the transaction frequency arises, the need to observe and study its behavior, attracts more and more researchers, with ultimate goal to build a forecasting model, with decent forecasting accuracy, that predicts this cryptocurrency's evolution.

In order for the model to achieve its best potential, we have to provide multiple regressors. To support our research, we collected data from stock indices as well as stock prices from companies with big market capitalization.
More specific, T. Panagiotidis et al. (2018) , after examining all potential suggested drivers: stock market indices, oil and gold, interest rates and internet trends, concludes that NIKKEI index is one of Bitcoin's determinants. Regarding other stock market indices, positive correlation is obtained by DJ, SSEC, and Nasdaq , setting the green light that more investigation on stock market indices had to be done.

In addition, Salisu, Isah and Akanni, (2019) examines the effect of bitcoin prices in predicting the stock returns of the G7 countries. The conclusion drawn from this research is that the stock returns of the G7 are better forecasted when using as inputs the bitcoin prices rather than the respective macroeconomic factors. The main reason is that the G7 stock markets are integrated into the world of economy and thus are more sensitive to external shocks than internal shocks. The interesting aspect of this research is that Japan appears to be an exception.

Our study is driven by conclusions mentioned in literature review but also by results that did not appear to have a significant impact on that particular research, but show some level of correction, intriguing us to investigate further on them.

A significant contribution to the present thesis has been made by Professor Gogas, overseeing the whole research and providing accurate and concrete information and guidance on the machine learning process.

Chasiotou Valentina
01/12/2019

Preface

I would like to thank my professor Mr.Gogas and his fellow Phd student Mr.Sofianos for their helpful guidance and commitment in order to finalize this present thesis.

Contents

Introduction

As we are moving to the next generation, where all transactions can be made electronically, through a virtual space, Nakamoto (2008) in his official release paper provides a simple explanation of what Bitocoin is and how it actually operates.

It mainly operates on an electronic payment system based on algorithms that use cryptography, allowing any two parties to perform transactions without any intermediates. These transactions are mathematically impossible to reverse, protecting both parties from fraud. It is considered to be a sequence of digital signatures. Each owner may transfer his coin to the next, simply by providing a document with his digital signature of the previous transaction along with the public key of the next owner and add these to his coin. In order to ensure the payee that the owners did not double-spend the coin, a trusted authority has been introduced to control the transactions. This authority, or else called the mint, checks every transaction thoroughly and only coins that are issued directly from the mint are allowed to be used.

Stepping into the technology that is behind all these transactions, we are elaborating a bit more on the so called blockchain technology. A blockchain is simply a distributed database of records, containing all the digital actions and transactions that have been executed and exchanged between the participants. Once information enters the system, it can never be erased. Its advantage lies to the fact that it contains an undoubtful and valid record of every single transaction and it can be verified by the unanimity of the participants in the system. An analogy that Das, Goel and Misra, (2015) present in one of their lectures is the following: “It is easier to steal a cookie from a cookie-jar kept in a secure place, that stealing the cookie from a cookie-jar in a market place while being observed by thousands of people”.

Although it facilitates the implementation of a multibillion-dollar global market of anonymous transactions without any control from the authorities, blockchain technology has been working perfectly over the past years and its applications both financial and non-financial institutions. The advantages of this technology overcome the regulatory issues making them more and more popular within the finance community. One of the frequently used cases involves the so-called “smart contracts”. Smart contracts are basically computer programs that can automatically execute the terms of the contract. Once the terms are configured from the program, the agreed payments can be automatically made in a transparent way.

Lastly, “Smart property” is another approach regarding the efficient way to control the ownership of a property as defined in the smart contracts. The property can be physical such as a house, or non-physical such as the ownership of shares of a company. In the next chapter the financial areas in which bitcoin technology finds applications, will be further explained.

Financial applications of technology compelled use cases

Blockchain technology no longer appears as a threat to financial institutions and banks. In fact the traditional models are being obsolete and replaced by new, innovative blockchain applications.

1. Private securities: It is widely known that the cost that needs to be undertaken in order for a company to become public is very large. A special syndicate of banks must be formed and decide to underwrite the deal in order to attract any potential investors. The way this usually works is that the stock exchange lists the company shares in the secondary market so that it can function in a secure way with their trades and in due time. Now, it is possible for companies to surpass this process and issue directly those shares via a blockchain network. These shares can be purchased and sold in a secondary market that is based upon the blockchain. Below some examples are indicated:

- *NASDAQ Private Equity*: NASDAQ launched its Private Equity Exchange back in 2014. The current process of stock trading in this particular exchange is inefficient and slow due to high involvement of third parties. So, NASDAQ collaborated with a San Francisco startup called *chain.com* in order to base the private equity exchange upon the blockchain. *Chain.com* is implementing blockchain technology based on smart contracts as discussed before. The outcome is really amazing since it appears to be "fast, traceable and efficient".
- *Medici* is a securities exchange that has as a primary scope to create an innovative stock market by using the counterparty implementations of bitcoin 2.0. Counterparty is a protocol that deals with traditional financial instruments such as the smart contracts. These smart contracts ease, verify or impose the negotiation eliminating at the same time the need for physical evidence. Subsequently, this excludes the demand of an intermediary, such as a broker, bank or exchange.

- *Coinsetter* is a New York based bitcoin exchange. It is working on a project that will settle and clear financial transactions in T+10 minutes rather than in T+2 or T+3 days.
- *Bitshares* are digital symbols that rely on the blockchain technology and can be related to specific assets such as currencies or commodities. The holders of these tokens are leveraged since they earn interest on commodities such as oil, gold as well as dollars, euros and currency instruments.

2. Insurance: Assets with a unique form that can only be identified by specific identifiers are difficult to impair or replicate, can be registered in the blockchain platform. This will help in verifying the ownership and tracing the transaction history. Through this way any property registered, physical like an asset or digital like a laptop, can be verified at any time and by anyone.

- *Everledger* is another company that uses blockchain technology in order to create an official permanent registration of a diamond certification and of the transaction history. Within this ledger the unique characteristics of each diamond will be stated, such as the height, depth, color etc. This way it is ensured that the verification of each diamond can be safely performed by insurance companies, law firms, owners and anyone that is interested in. Through a simple web application called API, one may look for a diamond, create, and be informed about any updates or claims by insurance companies or even about police reports on diamonds.

Economic Analysis of the Bitcoin Payment system

An electronic payment system operates as a tracker of a series of accounts. Each account is uniquely linked to a user. It also allows the user to check his balances as well as debit his balance and credit the debited amount to another account. The right to debit the account is strictly restricted only to the owner of the account. Balances cannot change without legal transfer.

The bitcoin system is designed to operate without any trusted authority. Therefore, its ledger is being updated and maintained by a series of computer servers, called miners. Even though those are not trusted, the system as a whole is secure, i.e. the system that processes all legal actions performs only this action and no other transactions.

Bitcoin's ledger is a public database called blockchain, which can be verified by third parties though cryptography. The system compensates the miners for their services and each of them maximizes his profit ensuring that all the others will act in the same way.

At first all balances amount to zero. As time passes by, the protocol mints new coins which are automatically added to the balances of the miners. All balance changes are recorded by the system and the proof of the transaction is a message which is sent to the miners containing the following information: the sending account, receiving account, amount transferred, transaction fee and a cryptographic signature by the sending account. This cryptographic signature is used in the verification of the transaction by any third party. It can be proven that the transaction was authorized by the holder of the sending account and that he indeed held a balance sufficient for the transfer.

Huberman, Leshno and Moallemi (2017) find that the Bitcoin Payment System can eradicate inefficiencies that occur due to market power, despite the fact that some other costs may arise. In this paper a model of the BPS (Bitcoin Payment System) has been developed as a two-sided market that captures the economic structure implied by the blockchain design. This model allowed analyzing the new market structure and deriving prices and costs. More specifically, it is built on the fact that the blockchain design acts as a two-sided platform and it is constituted by: (i) miners (independent profit oriented parties) who provide the system's infrastructure in exchange for payment; (ii) users who perform transactions and pay the respective fees.

In conclusion, the paper shows that transaction fees have bifold and essential roles in the Bitcoin system and also to support the results their model predicts that the miner's profits are zero and that fees are positively correlated with transaction waiting times.

Another difference that separates the BPS from the traditional payment systems is the support of transactions that are performed only in the system's coin, the bitcoin. The bitcoin has gained its value because the payment parties are prompted to exchange a credit in it for other commodities, services or even traditional currencies.

The blockchain technology is an innovative notion that has already been driving the economists' attention and scrutiny. Presently the BPS handles daily transactions that account for a huge amount of dollars and this surely is a strong motivation for scientists to study its structure and future applications.

Literature Review

As research is an endless ongoing process, many economists are making their own attempts to find out the real properties of bitcoin and the determinants that drive this cryptocurrency. Evidence prove that so far there is not a unanimous result on the real properties and as the discussion continues authors argue that the market plays a significant role in driving the prices while others suggest that a broader view will be more profitable and research should focus on the Asian market.

Poyser, O. (2017), after exploring the affiliation between Bitcoin's price and a set of internal and external factors, concludes that is positively related to stock market indeces and specifically to S&P500. Elaborating a bit more on the internal factors, Poyser,O. (2017) tried to find the supply and demand drive factors that are directly connected to bitcoin prices. Given that supply is deterministic, the internal values chosen where: number of bitcoins in circulation, transaction volume, hash rate and mining difficulty.

On the external factor side, of course there are other forces that might influence bitcoin’s price such as gold, silver, the Financial Stress Index, the S&P500 index.

Results on the internal factors suggest that they can be ruled out from the subclass of determinants. The author confirms that no correlation actually exists between these factors and the bitcoin price.

On the other hand, when examining the internal factors, it can be proven that bitcoin is a speculative asset, thus its behavior is related to other assets or market indices. Looking at the relationship to the S&P500 index it is positively related concluding that bitcoin has mixed properties between a currency and a commodity, thus further investigation is to be done.

From another perspective, Vo, N. N., & Xu, G. (2017) extend their research and investigate the correlation with other financial market indicators such as the USA market and the Australian market.

As bitcoin is not a legal trading currency, its exchange rate appeared to be a extremely high-risk portfolio with high volatility, requiring very detailed evaluation before proceeding to any decisions. One of the most discussed topics within the bitcoin community is the extremely volatile market prices or we may also encounter them as exchange rates against other strong currencies such as EUR,AUD,USD.

The present paper examines the hourly trade prices against financial indices such as: Dow Jones Industrial Average, Australian Securities Exchange. The correlation analysis performed suggests that bitcoin returns and the financial indices, even more specifically DJIA and ASX, are uncorrelated.

Being unaffected by traditional markets should urge us to look in another market, like the Asian market. Due to limited research, it is only a speculation at the moment that bitcoin may be correlated to the Asian market, but more empirical research should be performed in order to gain a deeper and broader view.

In summary, bitcoin returns has various interesting behavior aspects and it still continues to be unpredictable, thus deeper analytics is definitely required.

Apart from the file of economics, bitcoin tends to attract plentiful attention also in the sectors of cryptography and computer science due to the fact that it combines technology and monetary units. It is well know that certain aspects of the blockchain information are highly involved in bitcoin's supply and demand and are used for training their models and improving their performance.

H. Jang, J. Lee (2017) present an empirical study of forecasting bitcoin price movements with Bayesian Neural Networks, a machine learning algorithm. Considering as independent variables the prices and log prices of Bitcoin, it is confirmed that this model is suitable for forecasting Bitcoin prices. While relatively few studies have been conducted so far on estimation or prediction of bitcoin prices using machine learning algorithms, the present paper presents a practical approach of modeling and predicting the bitcoin prices, by using a Bayesian neural network (BNN). A BNN includes a regularization term into the objective function to overcome the problem of overfitting. This is an attempt to overcome the limitations of typical macroeconomic methods, since it can investigate non-linear influences of each relevant feature of the input variables and the blockchain information. The BNN model is firstly trained using given relevant features of the process and then evaluated in terms training. Errors are tested by using the representative non-linear methodologies and the linear regression model as the benchmark method.

The training of the BNN model is performed according to 10-fold cross-validation technique. The results indicate that the BNN model describes effectively the Bitcoin log price and log volatility. The present model is expected to perform well also on more recent data.

In conclusion, investigating the non-linear relationships between input functions based on network analysis can explain bitcoin price time series. Also, the variability of bitcoin should be treated appropriately. This milestone can be achieved by employing other machine learning methods.

Another interesting application of machine learning in Finance is described by Culkin and R.Das (2017) in their respective paper where deep learning is used for option pricing. There are many conditions that make deep learning useful for the field of finance. Firstly, the availability of large data, the so-called big data renders the application of these techniques a perquisite for understanding these massive amounts of data. Second, several finance applications depend on speed, and these algorithms could achieve high speed levels that can potentially be the key influences to trading processes. Third, many transactions and finance applications involve pattern recognition using data, where multiple factors should be imported as inputs in order to model the predicted outputs. For example, stock market predictions are based on many regressors such as historical prices, streaming data on the stock prices, interest rates, volatilies etc.

As most econometric models today are linear functions or simple transformations of linear functions and due to the fact that the relationship between inputs and outputs in the real world is non-linear, the predicted outcome is not accurate. This exactly what deep learning is adapting to. It transforms the old dataset into a new one and the layers of non-linearity get transformed and adjusted. This implies that a deep learning algorithm can “memorize and learn” almost any function to a high degree of accuracy.

In this paper, the researcher made at attempt to exercise the well-known Black-Scholes option pricing model from simulated data. Using a range of parameters and 300,000 call prices they created an algorithm that learned to price options with very low computing error. This can be further developed to price options in the real world and it is another proof of how capable machine learning is, leading to the conclusion that further research is to be done.

Methodology and Data Analysis

Since the literature review provided us with some insights on the determinants of bitcoin, the present research is an attempt to use these inputs in order to create a forecasting model that actually predicts bitcoin prices. Also, it is an attempt to extend the present literature review by obtaining prices of large capitalization companies and examining the correlation between them and bitcoin.

Data Description

The sample used for the present thesis was collected from the source : https://coinmarketcap.com/. Bitcoin daily prices were obtained ranging from 1/5/13 to 31/8/2019. There are 2314 observations of open, high, low and closing prices of bitcoin. The indendent variables are stock market indices such as S&P 500, Nasdaq, NYSE, DWI, Nikkei,Jasdaq and closing prices of large companies operating in the technology sector such as: Microsoft, Apple, Cisco, Mitsubishi, Fujitsu. In total the sample had 71,734 observations. All data that was used is daily and obtained from the source: https://www.investing.com/ and for the stock markets that do not operate 7 times a week as oposed to bitcoin tradings, the data was adjusted accordingly.

Methodology

Firstly all data was aligned appropriately and the natural logarithm has been calculated for all variables. The closing prices of bitcoin were used to create a binary variable, containing only values of 0 and 1 (0 when the price decreased, and 1 when the price increased). Also, in the initial data file 10 lags of each variable has been created except for the bitcoin closing prices.

Autoregressive model

After this step we have to find the optimum AR model. An autoregressive (AR) model predicts future performance based on previous information. It is widely used in forecasting when there is a high correlation between the values in a time series and the values that come before and after them. It actually works as a linear regression of the data in the current series against one or more values in the same series.

Mathematically, an AR(p) process of a series y_t can be represented by the following equation:

$$y_t=a_1y_{t-1}+a_2y_{t-2}+...+a_py_{t-p}+e_t$$

Where:

- $a_1,a_2,...,a_p$ are autoregressive parameters
- e_t are normally distributed random error terms with zero mean and a finite variance σ^2.

The reason that we needed tha lags from the previous step is that these lagged values work as predictor variables in our model. The value "p" of an AR(p) model is called the order. For example the AR(1) model is called a "first order autoregressive process" where the outcome variable at some point in time *t* is related only to time periods that are one periond behind *(t-1).*

In order to find the optimum AR model, we run all AR regressions in the software *e-views* up to the order of 25 and came up with the following results:

Autoregressive model in order p	Schwarz criterion
AR1	1.449109
AR2	1.451859
AR3	1.454785
AR4	1.457906
AR5	1.461138
AR25	1.518493

Table 1

For better insight it would be better if we provided a detailed explanation of the Schwarz criterion and how we should select the appropriate variable.

Scwarz Information Criterion

In statistical modeling one of the researcher's duties is to choose the most appropriate model among a wide range of potential contestants. In order to reach to a conclusion, the use of a selection criterion is essential. The criterion assigns a score value to every model based on a statistical principle. The fitted model is the one that has assigned the maximum or minimum score, always depending on the respective model and methodology used.

The Schwarz information criterion (SIC) was introduced by Schwarz on 1978 and it is one of the most popular and effective criteria for model selection. The model basically serves as an ssymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. In large-samples, the fitted model favored by SIC ideally corresponds to the candidate model with the smallest information criterion value. The computation of SIC is based on the empirical log-likelihood and does not require the specification of priors. The equation below describes the SIC:

$SIC = k \times \ln n - 2 \times \ln(L)$

Where:

- k is the number of model parameters
- ln(L) is the log-likelihood function for the statistical model.

Moving forward, as we may observe from table 1 the autoregressive model with order 1, the AR(1) has the smallest value of SIC, thus we need to proceed with this one.

Our next step, was our attempt to remove autocorrelation.

Autocorrelation is a mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals. Technically it measures the relationship between a variable's current value and its past values. We may encounter autocorrelation between the range of -1 up to 1, with an autocorrelation of +1 representing a perfect positive correlation, while an autocorrelation of -1 representing a perfect negative autocorrelation. In order not to “cheat”, we have to remove autocorrelation since it has a huge impact on our model due to the fact that pas prices should not have such an impact on future prices.

Correlogram function

In order to test this, we performed the appropriate test in e-views called “correlogram” and the output is presented below:

Sample: 5/02/2013 8/31/2019
Included observations: 2313

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0015	0.969
		2	-0.022	-0.022	1.0800	0.583
		3	0.024	0.024	2.4210	0.490
		4	0.017	0.017	3.0935	0.542
		5	0.009	0.010	3.2825	0.657
		6	0.041	0.041	7.1866	0.304
		7	0.058	0.058	15.047	0.035
		8	0.045	0.047	19.809	0.011
		9	0.003	0.004	19.827	0.019
		10	0.005	0.003	19.890	0.030
		11	0.007	0.003	20.019	0.045
		12	0.011	0.007	20.323	0.061
		13	0.015	0.009	20.849	0.076
		14	0.046	0.040	25.851	0.027
		15	0.001	-0.004	25.853	0.040
		16	-0.017	-0.019	26.560	0.047
		17	-0.002	-0.006	26.571	0.065
		18	0.021	0.017	27.609	0.068
		19	0.017	0.014	28.247	0.079
		20	0.012	0.009	28.609	0.096
		21	0.001	-0.005	28.611	0.124
		22	0.020	0.017	29.566	0.129
		23	0.007	0.008	29.678	0.159
		24	0.008	0.009	29.834	0.190
		25	0.011	0.007	30.126	0.220
		26	0.010	0.005	30.370	0.253
		27	0.035	0.032	33.294	0.188
		28	-0.018	-0.022	34.018	0.200
		29	0.016	0.015	34.602	0.218
		30	-0.022	-0.027	35.742	0.217
		31	0.014	0.012	36.219	0.238
		32	-0.007	-0.014	36.341	0.273
		33	0.016	0.012	36.975	0.290
		34	0.018	0.014	37.726	0.303
		35	-0.002	-0.002	37.734	0.345
		36	-0.006	-0.006	37.818	0.386

Figure 1-

Source: E-views

All variables of this graph will be properly analyzed and their functionality will be properly discussed within this chapter.

- AC stands for Autocorrelation. The autocorrelation of a series *Y* at lag *k* is estimated by the following equation:

$$T_k = \frac{\sum_{t=k+1}^{T}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y})^2}$$

equation 1

Where $\overline{Y}$ is the mean of *Y*. This is the correlation coefficient for values of the series *k* periods apart.

- PAC stands for partial autocorrelation. It mainly produces the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags. It can be estimated by the following equation:

For regression of *y* on x_1, x_2, x_3, x_4, the partial correlation between *y* and x_1 is:

$$\frac{cov(y, x_1 | x_2, x_3, x_4)}{\sqrt{var(y|x_2, x_3, x_4) * var(x_1|x_2, x_3, x_4)}}$$

equation 2

This can be calculated as the correlation between the residuals of the regression of *y* on x_2, x_3, x_4 with the residuals of x_1 on x_2, x_3, x_4.

- The last two columns reported in the correlogram are the Ljung-Box Q-statistics and their respective *p*-values.

The Ljung Box test is a way to test for the absence of serial autocorrelation up to a specific lag *k*. The default set in e-views is 36 lags. The null and alternative hypotheses can be described as:

H_0: The data are independently distributed meaning that there is no autocorrelation.

H_1: The data are not independently distributed meaning that they exhibit autocorrelation.

As we may observe our *p*-value for lag 36 is 0.386>0.05, thus we cannot reject the null hypothesis that there is no autocorrelation up to lag 36.

During our next step we had to decide which regressors should keep in order to proceed with our model. In the chapter below the stepwise least squares algorithm is being described, as it was a part of our feature selection technique.

Stepwise least squares

Often only theory is not enough to provide a specific direction as to which variables should be included in the final regression model. The actual set of predictor variables that will be finally used must be determined by data analysis. Determining this subset is the so-called feature selection problem.
Finding this subset of potential regressors or independent variables involves two main objectives. First, we need the regression model to be as concrete and realist as possible in order to reflect the reality. For this reason, we would need every regresssor that is even slightly related to the dependent variable to be included. Second, we want to include only a few variables, and this applies mostly to big dataset with many factors, because irrelevant regressors will decrease the precision of the estimated coefficients and predicted values. Also, the presence of redundant variables increases significantly the complexity of data collection and model maintenance. The ultimate goal of this technique is to obtain a balance between simplicity and fit.

There are many methods to approach, and *e-views* uses the forward step-up selection method. This method provides an initial screening of the candidate variables, when a large group of variables exists. The algorithm begins with no candidate variables in the model. Then, it selects the variable that has the highest *R-Squared.* At each step, the candidate variable that increases the *R-Squared* the most, is picked up. It stops adding variables when none of the remaining variables are significant. Once a variable enters the model, it cannot be deleted.

The stepwise forwards method was performed in *e-views* and the results are the following:

Dependent Variable: BINARY_BITCOIN_CLOSE
Method: Stepwise Regression
Date: 11/21/19 Time: 19:54
Sample (adjusted): 5/11/2013 8/31/2019
Included observations: 2304 after adjustments
Number of always included regressors: 2
Number of search regressors: 319
Selection method: Stepwise forwards
Stopping criterion: p-value forwards/backwards = 0.05/0.05

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	1.244982	0.556045	2.238997	0.0253
BINARY_1	-0.020834	0.020834	-0.999997	0.3174
LN_HITACHI_PRICE__10	-0.822575	0.250096	-3.289044	0.0010
LN_MITSUBISHI_PRICE__2	1.564205	0.621491	2.516859	0.0119
LN_MITSUBISHI_PRICE__1	-2.767243	0.868327	-3.186866	0.0015
LN_HITACHI_PRICE__3	0.593016	0.254477	2.330331	0.0199
LN_MITSUBISHI_PRICE	1.367206	0.626089	2.183724	0.0291

R-squared	0.014430	Mean dependent var	0.547743
Adjusted R-squared	0.011855	S.D. dependent var	0.497823
S.E. of regression	0.494864	Akaike info criterion	1.433965
Sum squared resid	562.5125	Schwarz criterion	1.451411
Log likelihood	-1644.927	Hannan-Quinn criter.	1.440325
F-statistic	5.605070	Durbin-Watson stat	2.001436
Prob(F-statistic)	0.000009		

Figure 2-

Source: E-views

As we may notice, the variables that we used as dependent ones are the binary and the first lag of the binary as indicated per our previous AR(1) model. The regression points out 5 variables that seem to be more correlated to our dependent variable. The most interest part is that, all those variables trade in the Tokyo Stock Exchange and this is compliant with our initial question and the purpose of this research. Since there was no literature review on this stock exchange it can play a significant role in future research.

All p-values are statistically significant as they are <0.05, thus the feature selection process gave us the best predictors for our model.

Machine learning

Machine learning is a concept that computer systems use to perform a specific task without human interference. The use of scientific algorithms and statistical models render the system independent to perform any actions by relying on patterns only. It may also be encountered in literature review as artificial intelligence. More specific, machine learning algorithms build a mathematical model based on simple data, that is also called training dataset, in order to perform predictions or decisions without being programmed to do so.

There are three common characteristics that all machine learning algorithms share, regardless of the learning style or function:

1. Representation: how to represent knowledge. A machine learning model can't directly see, hear or sense the inputs that we are providing with. Instead, we must create a representation of the data to provide the model with useful information on the data's specifications and key qualities. In order to train a model, we must choose the set of features that best represent the data uploaded. Examples include decision trees, graphical models, neural networks, support vector machines, models ensembles.
2. Evaluation: the way to evaluate the performance of a machine learning model, which is a mandatory step of any data science model. Model evaluation aims to estimate the generalization accuracy of the model on a future dataset, that the current model has not been exposed to yet. Methods for evaluating a model's performance are divided into two main categories: holdout and cross-validation. In our research the cross-validation technique has been used and will be properly explained in the next chapter. As a general rule, both methods use a test set (dataset that has not been processed by the model) in order to evaluate its performance. Of course it is not recommended to use the data used to build the model for evaluation, since it will simply "remember" the whole training dataset and provide an accurate prediction. This is also known as overfitting.

3. Optimization: often the highest-scoring classifier. It is the most essential ingredient since the choice of an optimization algorithm can make a significant difference between getting a good accuracy in hours or days. It starts with defining some kind of loss/cost function and ends with minimizing it or using another optimization routine. There are three basic elements of any optimization problem:

- Variables: Free parameters which the algorithm can tune
- Constraints: Boundaries within which the parameters must fall
- Objective function: The set of goals towards which the algorithm drives it's solution. Often this is the amount that minimizes some error measure or maximizes some utility function.

Types of learning

There are 3 main types of machine learning.

1. *Supervised learning:* All data is labeled and the algorithm learns to predict the output given the input dataset. More specific, these algorithms are designed to learn by example. The name "supervised" originates from the idea this algorithm is compared to the situation where a teacher in school supervises the whole process of a test in the class. When training a supervised learning algorithm, the training data will consist of inputs paired already with the corresponding correct outputs. After the training, the algorithm will take in new inputs that it has never processed before and will determine which label the new inputs should take, classifiyin them based on prior training data. The objective is to predict the correct class or label of the newly inputed data. At its most basic form, a supervised model can be explained by the basic mapping function used in mathematics:

$$Y= f(x)$$

Where Y is the predicted output that is directly determined by a mapping function that assigns a class to an input value x. The function used to connect the input data with the output data is automatically created by the machine learning model during training.

Supervised learning can be split into two main subcategories: *Classification* and *Regression* that will be approprietly elaborated into the next chapters.

2. *Unsupervised learning:* The algorithm infers patterns from a dataset without reference to any known or labeled outcomes. The goal is to model the underlying structure in the data in order to gain further insight of the data. This technique is called unsupervised because unlike supervised methods, there is no correct answer and there is no teacher in this case. Algorithms are left to make their own decisions in order to discover the structure of the data. Basically they perform three actions:
    - Explore the structure of the information and datect distinct patterns.
    - Extract valuable insights on the dataset structure
    - Implement those in their operations in order to increase the efficiency of the decision-making process.

    Unsupervised models can be used into clustering and association problems.
    Clustering: A clustering problem is where you want to detect the fundamental groups formed in the data, for example grouping customers by their purchasing behavior.
    Association: An association problem is where you want to detect specific rules that describe large parts of your data, for example people that buy *X* also tend to buy *Y.*

3. *Semi-supervised:* These algorithms are to be categorised somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data inputs for their training; typically a small amount of labeled data and a large amount of unlabled data. The systems that use this method are able to improve their learning accuracy. Basically, labeled data is used to help identify that there are specific groups present in the data and what those might be. So the discovery of new patterns and the clustering into several categories based on the dataset's features are the key objectives of this method. Popular examples are the K-means and Latent Dirichlet Allocation algorithms.

Machine learning in Finance

Nowadays different sectors of the economy are processing huge amounts of data available in all kinds of formats and collected from various sources. This huge amount of data, also called big data, has become easily accessible due to the continiously progressive use of technology. Machine learning has found a plethora of applications in finance due to the increased volume of financial historical data available.

The are 3 main reasons why the finance industry is benefited from the use of these techniques.

- Reduced operational costs due to process automation.
- Increased revenues due to better productivity and enhanced user experiences.
- Better compliance and reinforced security.

Below a brief presentation of the main current applications of machine learning in the finance industry follows:

1. *Portfolio Management:* The so-called Robo-advisors are nowadays recognized as one of the most widely used applications of machine learning in finance. These online applications can provide several financial services and guidance through an automated procedure. With the use of statistics these algorithms are able to provide portfolio management services. Of course, their services are more accurate and faster than a human's financial decisions and much cheaper. The process is quite simple for the users. In order to set up a new account with a robo-advisor, one must fill in a questionnaire regarding his investment goals and current financial situation. Then, the robo-advisor automatically starts allocating assets within a range of investment options based on the goals and investor profile that created while analyzing the data from the questionnaire. It uses its algorithms to monitor the portfolio and regularly modifying the initial portfolio. *Betterment* and *Wealthfront* are two online investment companies that provide on-line portfolio management services through the robot-advisors. The software that they use is programmed to perform verified investment

strategies, to automatically check for any improvement in these, while maintaining the optimal investment goal making sure that no deviations occur.

2. *Algorithmic trading:* The computers use a set of predefined instructions in order to carry out programs that trade on behalf of a human trader. This set of instructions may contain for example several parameters, like prices, time, quantity or any other independent variables. Hedge fund managers use these automated trading systems more frequently since they can achieve the execution of large orders, only by sending small increments of the order to all markets at intervals. Also, since the system can perform multitasking by operating in multiple markets simultaneously, the trading opportunities increase significantly. Moreover, these algorithms manage to learn and adapt to real-time changes, another advantage over the human trader. What is more interesting is the fact that since they are not able to experience any sentiment or emotion, the sabotage of any opportunity is minimized reaching zero levels of mistakes. Aidyia is in control of an artificial intelligence mechanism that performs all stock trades without any human interference. Although this system was built by humans, it runs completely autonomously. At the launch of this automated hedge fund, the CEO Dr. Ben Goertzel remarked: “If we all die, it would keep trading”.
3. *High frequency trading (HFT):* High frequency trading is a subset of algorithmic trading, executing hundreds of trades per day. Complex algorithms are working on analyzing multiple markets simultaneously based on current market conditions. Investment banks and hedge funds benefit from these automated trading platforms platforms that are able to perfom multitasking, trading and researching at the same time. The most fascinating thing about HTF is that these algorithms could spot price differences that occur only for a small fraction of a second, giving the big players the chance to make huge profits. Some of the biggest players include companies like Tokyo-based Nomura Securities, Tower Research Capital and DRW. On the other hand, criticism against the practices of HTF states that it can cause inexplicable and unforeseen market movements. The example here is the dramatic drop on the Dow Jones Industrial Average on May 6, 2010, that amounted to 10% in just 20

minutes. After investigation, it was blamed on a massive order that triggered a sell-off and caused the crash.

4. *Fraud Detection*: One of the most feared problem for financial institutions that is ideally overcomed with machine learning. Due to the fact that these systems are able to scan huge data sets, detect any suspicious activities and label them instantly, such unfortunate occurences may be eliminated. What is more efficient, is the fact that the algorithms are able to detect the so-called "false positives". This usually happens when vendors or financial institutions wrongly reject legitimate financial transaction requests. False-positive card declines are a huge problem that financial institutions have to face, since the customer's loyalty is compromised when a company incorrectly declines customer's cards. IdentityMind Global is a company which helps mechants, financial institutions and payment service provides to indentify fraudsters. The company has patented a machine learning software called electronic DNA that is able to cross-check an individual's identity by using more than 50 data points. Through their services, many other companies perform identity checks, risk-based authentication and regulatory identification. What is most important is the fact that this incorporated monitoring process also allows anti-money laundering and counter-terrorism financing.
5. *Loan and Insurance Underwriting:* Algorithms that check whether an applicant is entitled to receive a loan or insurance by performing automated tasks like matching data records and looking for exceptions. ZestFinance is a company in LA that helps other companies with the assessment of their loan applicants with limited or no credit history. This automated platform utilizes a big amount of data points in order to correctly assess applicants that no other institutions would normally consider assessing. Their services also include the assessments of client's creditworthiness for mortgage, financing and refinancing of student loans, small business loans and house reformation loans.
6. *Risk Management*: The increased use of machine learning helps large corporations and financial institutions to quickly process current data in order to spot trends and predict potential risks. Risk management is improved this way as these institutions rely on accurate market forecasts for the prosperity of

their businesses. *Dataminr* is such an example, where the employment of these algorithms has helped them to manage risk. It claims that it can reveal high-impact events and critical information before it is actually spread through the news. The company uses real-time social media in order to obtain this information and possibly reveal any high-impact event or a breaking news in the market.

7. *Chatbots:* A chatbot is an artificial intelligence software that it is created with the intention of mimicking the spoken or written speech patterns. They are also known as conversational agents as they are designed to simulate and sustain a conversation with a real person. Thanks to natural processing algorithms and the ability to learn from previous communication, chatbots are able to provide an immediate answer to any customer query. They can also adapt to the changes of customer behavior. A company that uses chatbots to supervise its p finance actions is *Kasisto.* Users can download Kasisto's AI application on their mobiles and internet platforms. It uses train statistical models based on collected data in order to be able to provide accurate information.
8. *Document Analysis:* The ability of machine learning systems to scan and analyze documents at high speed levels assists banks to detect fraud and meet their compliance requirements. At *JP Morgan* a program called COIN completed 360,000 hours of work in a matter of seconds. This job entailed analysis of 12,000 commercial credit agreements. COIN stands for Contract Intelligence. *JP Morgan* is investing in technology on a budget of 9,6 billion dollars.
9. *Trade Settlements:* Trade settlement is the process of transferring securities into the account of a buyer and cash into the seller's account by following real time trading stocks. Despite the fact that most of the trades are settled automatically nowadays, some 30% of trades fall through the system and need to be settled manually. The use of machine learning here is quite simple: it identifies the failed trades, then it analyzes whey the trades were rejected, and lastly it provides a solution along with a prediction which trades are more likely to fail in the future. *BNY Mellon* has implemented robotic process automation software which allow them to perform research on failed trades, spot the actual cause and fix the problem.

10. *Money-Laundering Prevention:* According to the United Nations report, 800 billion-2 trillion dollars are laundered globally in the period of one year. The softwares used collect internal, public and transactional data from the client's extented network trying to identify money laundering "behaviors". *Commerzbank* is applying machine learning technology to automatically prepare compliance reports. It intends to automate about 80% of all compliance-based checks related to trade finance process by 2021. This technology uses character recognition and machine learning in order to extract data from physical documents, recognize patterns and mark any deviations.

k-fold Cross-Validation

As previously mentioned, cross-validation is a statistical method for evaluating the model's performance. It is basically a resampling procedure used to detect overfitting. All cross-validation methods follow the same basic procedure:

1. Divide the dataset into 2 parts: The training and the testing.
2. Train the model on the training set.
3. Evaluate the model on the testing set.

This is one of the most important steps as we can check whether our model performs well on data seen and not. Without this technique, we could not use the model for further applications.

In *k-fold cross validation*, the data is divided into *k* subsets such that each time one of the *k* subsets is used as the test/set and the other *k-1* are put toghether to form the training set. Below there is an illustration of this technique:

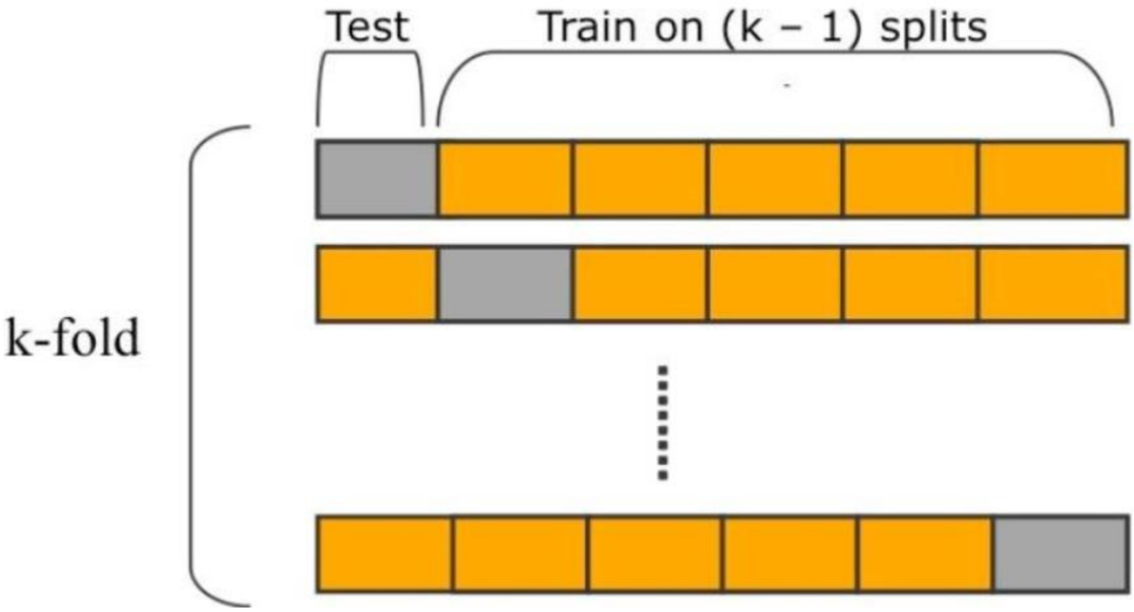

Figure 3-

Source: Researchgate.net

The error estimation is averaged over all *k* trials to achieve the total effectiveness of our model. Every point gets to be validated exactly once, and gets to be trained exactly *k-1* times. This reduces bias and variance significantly as the most of the data is also being used in the validation step. Interchaning the training and test sets also adds effectiveness to this method. As a general rule gained from empirical evidence, *k=5* and *k=10* is generally preferred.

The disadvantage of this method is that some data points may never be selected to be in the test subset at all while some data points might be selected multiple times. This is a direct result of randomisation. Yet with *k-fold* there is a guarantee that all points will at some time be tested.

Below you may find the breakdown of the simplest case of *k-fold* technique:

The goal is to estimate a tuing parameter λ (such as the subset size).

First the data is divided into *K* roughly equal parts.

1	2	3	4	5
Validation	Train	Train	Train	Train

Figure 4-

Source: mc.ai

Second, for each *k=1,2,… K* fit the model with the parameter λ to the other *K-1* parts, giving $\hat{\beta}^{-k}(\lambda)$ and compute its error in predicting the *k*th part:

$$E_k(\lambda) = \sum i \in kth\ part^{(y_i - x_i\hat{\beta}^{-k}(\lambda))^2}$$

equation 3

This gives the cross-validation error:

$$CV(\lambda) = \frac{1}{K}\sum_{k=1}^{K} E_k(\lambda)$$

equation 4

Typically we use *K*=5 or 10.

Results

Since all methodologies used have been explained in the previous chapters, we will now present the actual machine learning models that Matlab selected as the optimum and elaborate on the findings.

Our approach contains the use of both *5-fold* and *10-fold* cross-validation across our models in order to gain further insight on the differences in the outcome.

Quadratic Discriminant

Our first model presented is the Quadratic Discriminant. A short introduction on the model follows for better understanding of the algorithm that Matlab used to produce the results.

Discriminant analysis finds a set of prediction equations based on independent variables that are used to clasify individuals into groups. There are two objectives in the discriminant analysis: First, to find a predicitive equation for classifying new individuals and second to interpret the predictive equation to better understand the relationships that may exist among the variables. In we may parallel this method, it would be with multiple regression analysis.

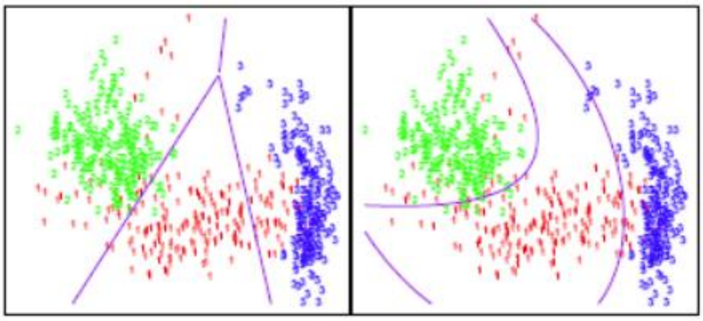

Figure 5-

Source: scikit-learn.org

As it can be observed in the figure above, the correct classification is the right one, where the quadratic discriminant was able to capture the differing covariances and provide more accurate non-linear classification results.

Firstly, matlab resulted that the model with the best forecasting accuracy with 5-fold cross-validation would be the quadratic disciminant. The forecasting accuracy was 56%. From the scatter plot shown in the figure below we may notice that the relationship between the two variables is non linear.

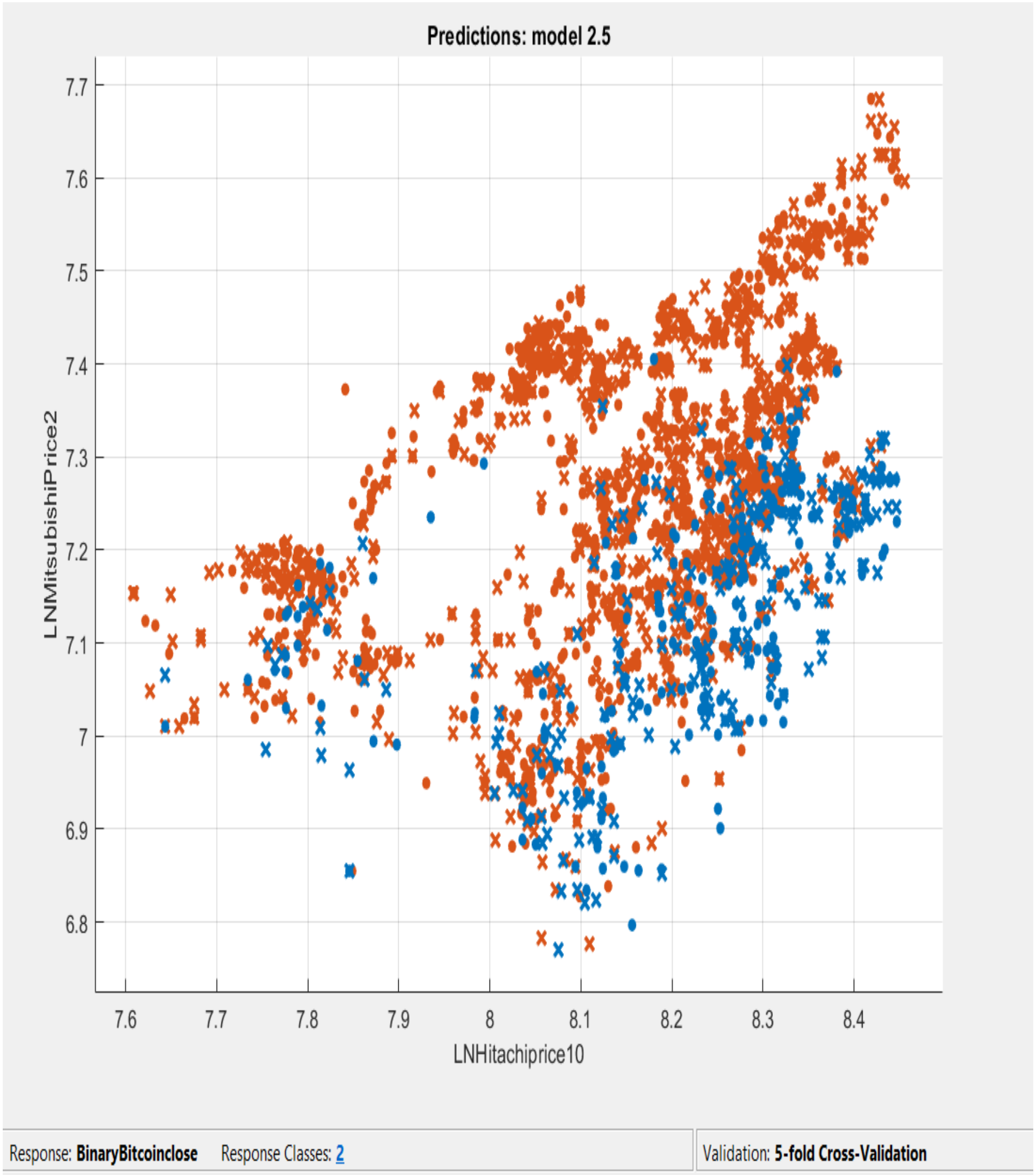

Figure 6-

Source: Matlab

Moving on to the confustion matrix, we will also elaborate a bit on what exactly this matrix is and how to read it.

So, a confusion matrix is often used to describe the performance of a classification model on a set of test data for which though the true values are known. There are two possible predicted classes.

True positives (TP): These are cases in which we predicted yes and actually it was no

True Negatives (TN): We predicted no and actually it was a no

False Positives (FP): We predicted yes, but actually it was no (Type I error)

False Negatives(FN): We predicted no, but actually it was yes. (Type II error).

What is more, regarding the different types of errors, Type I error occurs when one rejects the null hypothesis when it is actually true. In other words, this is the error of accepting an alternate hypothesis where the results can be attributed to chance.

Type II occurs when one does not reject the null hypothesis when the alternate hypothesis is the true one. In other words, one is failing to accept an alternate hypothesis when he doesn't have adequate power.

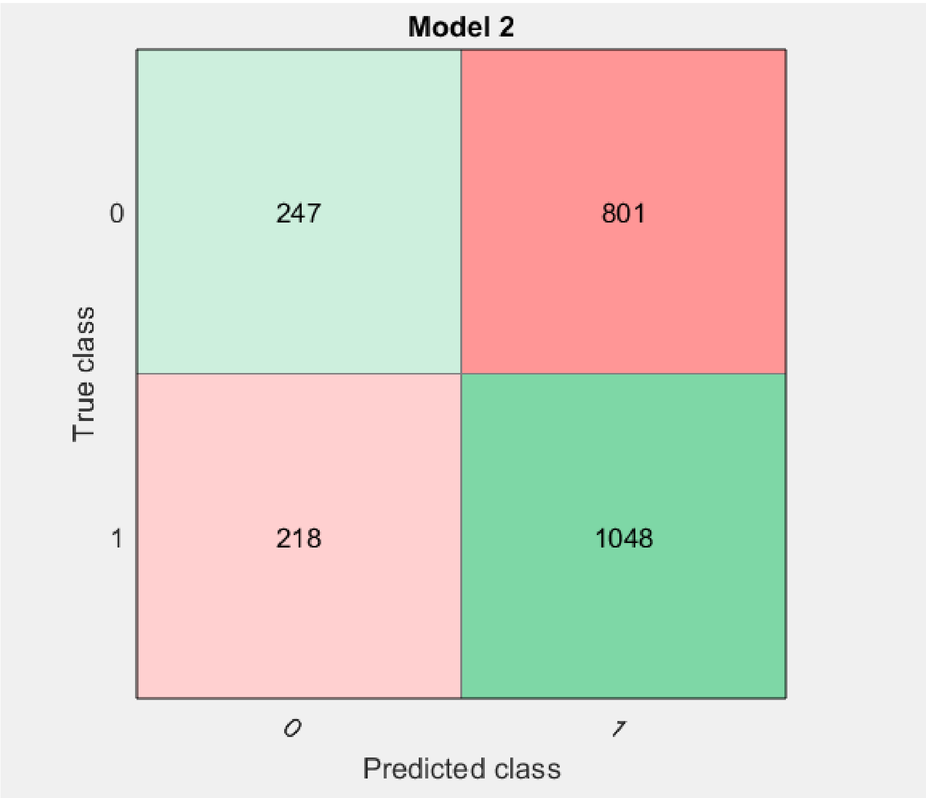

Figure 7-

Source: Matlab

From the above figure it seems that 218 observations that were predicted as 0 (meaning that a decrease occurred) and what was truly happened was an increase in price. 247 observations were predicted to have no change or decreased and this actually was the case.

On the other hand, 1048 observations were found to have an increase in price and this is what actually happened. Lastly, for 801 observations the model predicted that the price will increase, but the opposite happened. The conclusion here is that out of 2314 observations, the 1295 were predicted correctly. The most unfortunate event could happen for the 801 observations, where we predicted that the price will go up, so as an impact an investor might have invested in the bitcoin based on this, but actually the price decreased.

Last but not least, the ROC curve is shown in the below figure. A receiver operating characteristic curve or ROC curve, is a performance measurement for a classification problem. Basically it is a probability curve and the Area Under Curve (AUC) represents the degree or measure of separability. It actuatlly assesses how much the model is capable of distinguishing between classes Higher the AUC, better the model is at predicting.

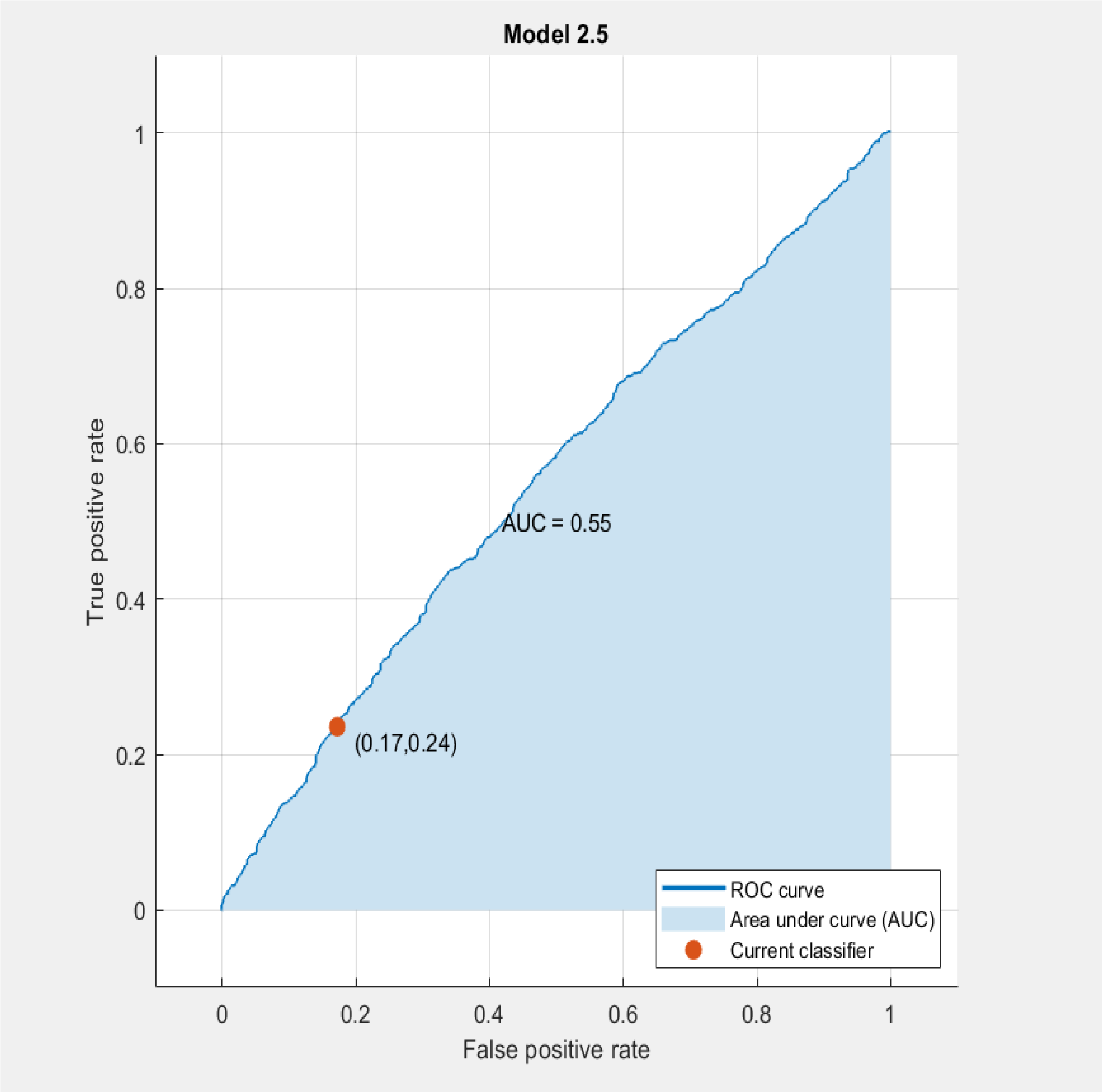

Figure 8-

Source: Matlab

The current classifier has a low forecasting accuracy. The higher we move on the ROC curve and more to the right, the highest accuracy we obtain.

Our second attempt was to use a 10-fold cross validation. The results are almost the same with matlab choosing again the quadratic discriminant model with forecasting accuracy 55,8% this time.

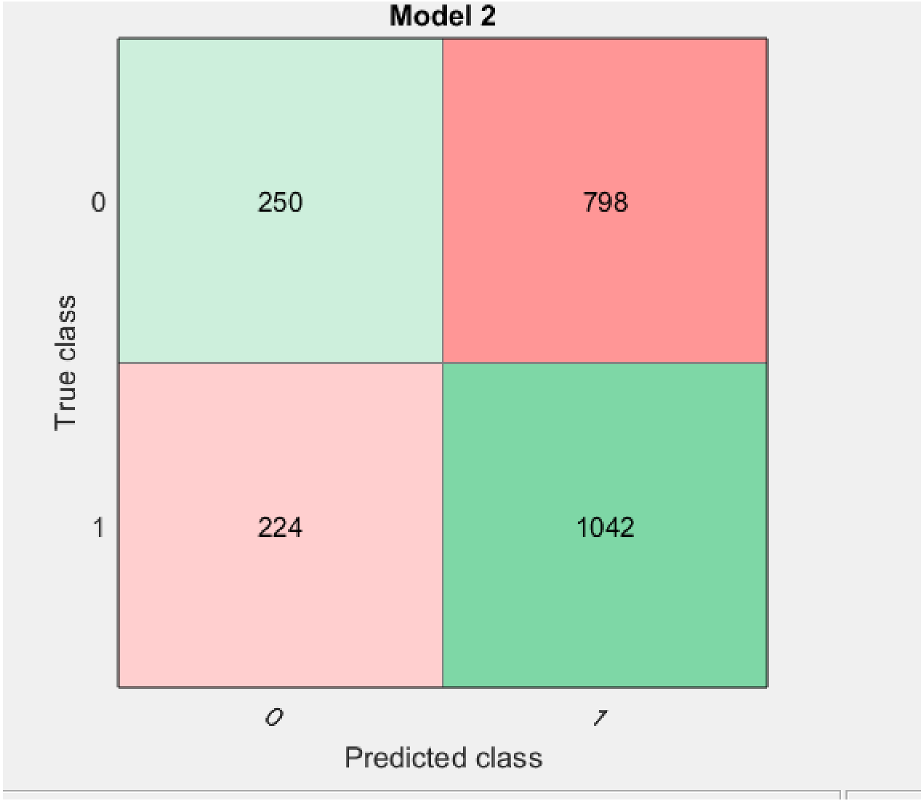

Figure 9-

Source: Matlab

Comparing the confusion matrices, it seems that with this model we increased the wrongly predicted classes from 218 to 224 but decreased the classes where we had predicted that the price will go up but instead it went down, from 801 to 798. What is more, concerning the correct predictions we improved our results from 247 to 250, but on the other hand for classes that we predicted an increase in price and this was what actually happened, we dropped from 1048 to 1042.

Regarding the ROC curve, the results are mostly the same, as the forecasting accuracy did not improve compared to our previous model.

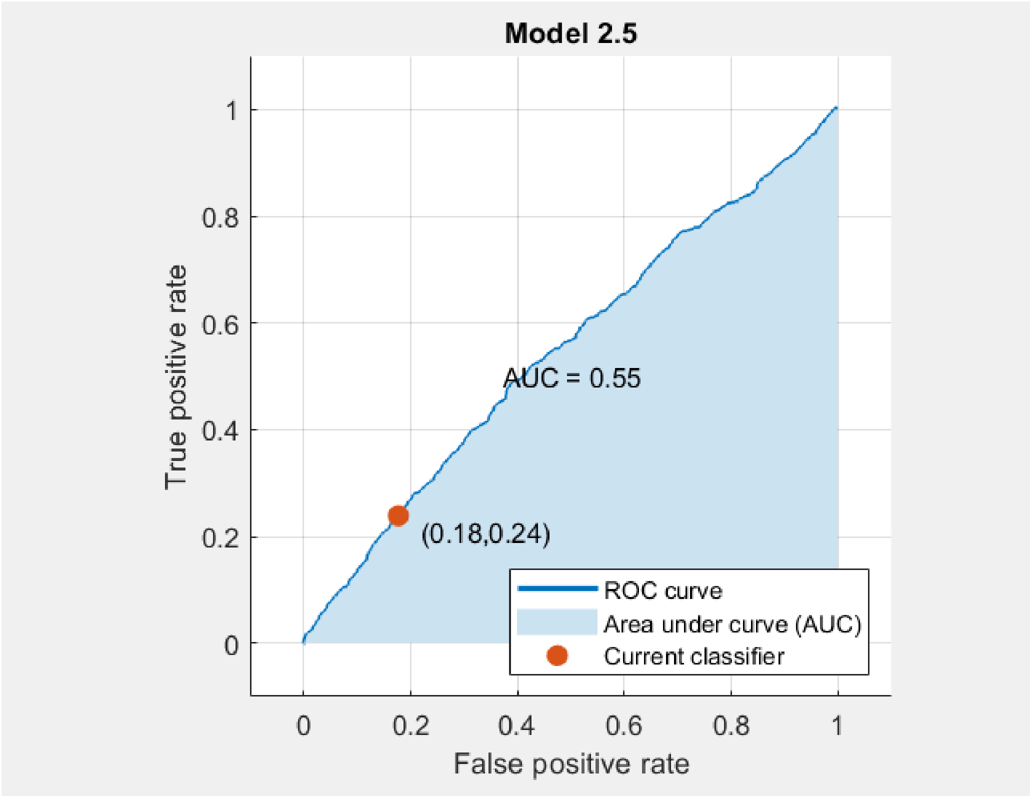

Figure 10-

Source: Matlab

The classifier is still on the left and down side of the curve, and we would need it to move on the right and up in order to achieve better forecasting results.

The third model that was tested had as inputs not only the specific lagged values that our stepwise least squares procedure proposed, but also all their previous lags. For example for the variable LN Hitachi our regression proposed the 10^{th} lag and this time all the previous lags from -1 to -10 were used. The results shown below suggest a different choice of models this time.

First of all, we used the 5-fold cross validation and our software suggested that the model Ensemble, with subspace disciminant and forecasting accuracy of 56.2% percent was the best one.

Ensemble methods are algorithms that combine several machine learning techniques in one predictive model in order to improve the forecasting accuracy. The most popular are bagging and boosting. The bagging algorithms try to reduce the variance of an estimate and the boosting try to convert weak learners to strong learners.

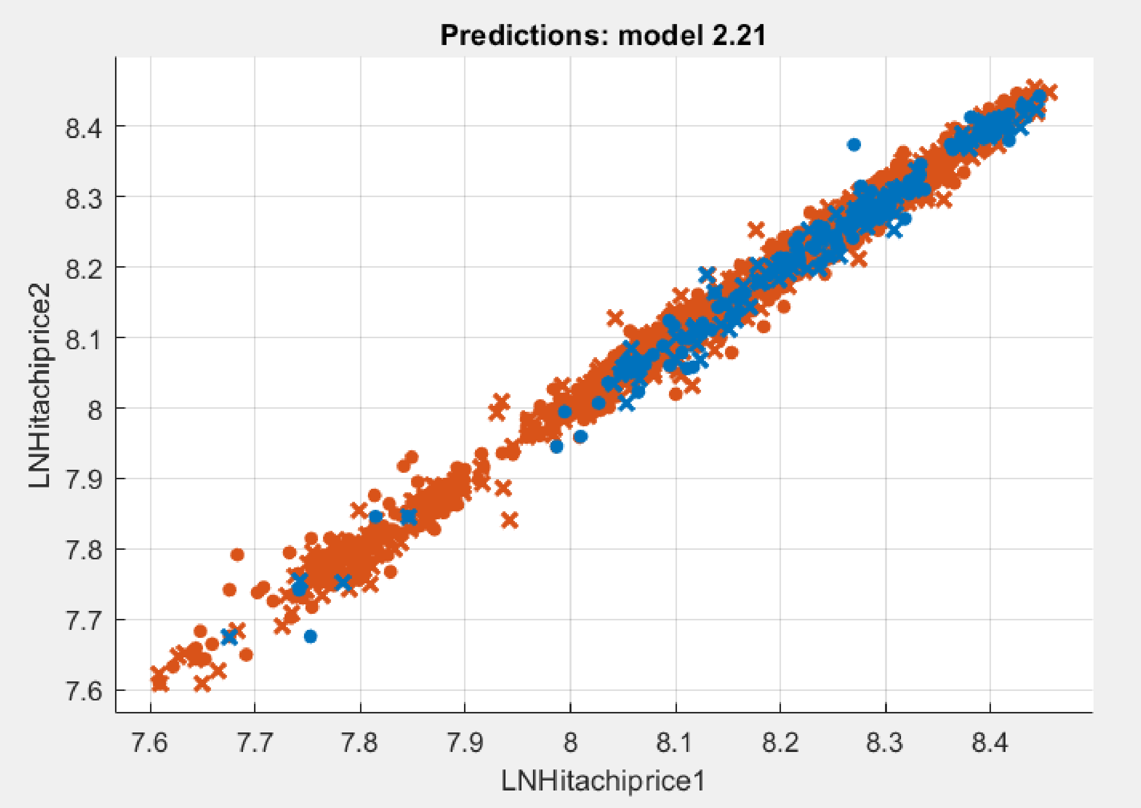

Figure 11-

Source: Matlab

Starting from the scatter plot, where we may observe that the relationship seems linear at the beginning but it gets really non-linear as we move to the up right side of the graph.

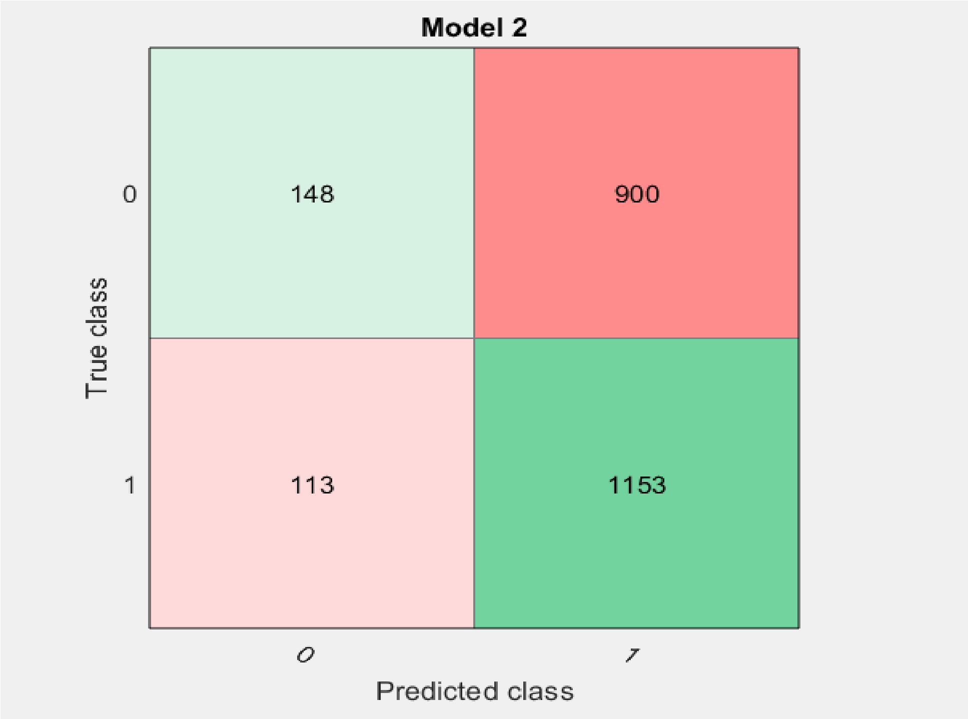

Figure 12-

Source: Matlab

From the confusion matrix, it seems that we have a high amount of observations that were classified correctly, meaning 1153 predictions were accurate. On the other hand, we observe a high amount, 900 to be exact, of observations that were wronlgy classified and this is the most important problem for the model, since in these cases we predicted wronlgy that the price will go up, while it really went down.

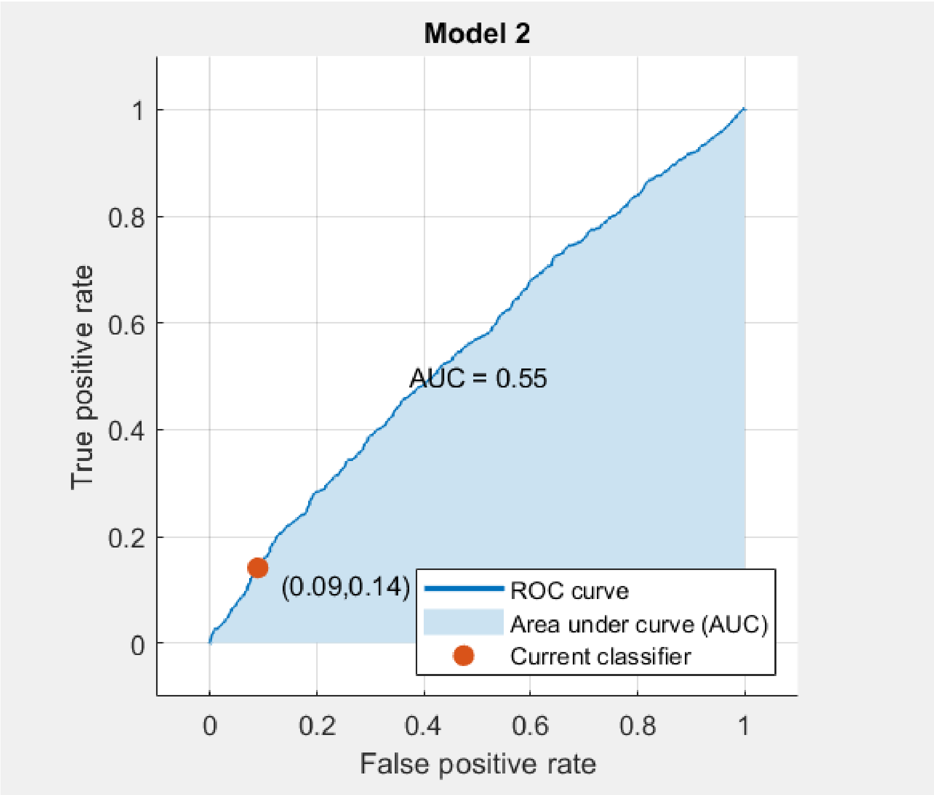

Figure 13-

Source: Matlab

As far as the ROC is concerned, the results are aligned with our initial forecasting accuracy. The classifier stands close to 0, the beginning of the curve meaning that the forecasting accuracy is not high.

Linear Discriminant

Lastly, our last model was performed on a 10-fold cross validation and this time matlab concluded to the linear discriminant with forecasting accuracy of 56.1%.

The linear discriminant is a widely used dimentionality reduction technique that was developed in 1936 by Ronald. A.Fisher. The original linear discriminant applied to only 2-class problems but in 1948 C.R.Rao generalized it to apply to multi-class problems.

Basically this algorithm helps to reduce the high-dimensional dataset into a lower dimensional space. The ultimate goal is to do this while having a proper separation between calsses and reducing resources and costs of computing.

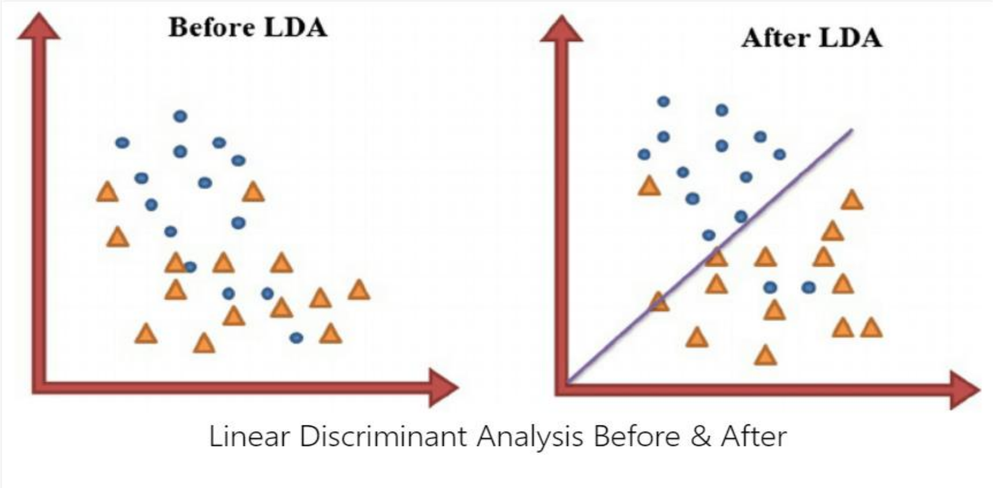

Figure 14-

Source: Researchgate.net

Comparing our scatter plot and the above figure, it seems that the software chose the correct model according to the dispersion of the data. A theoretical line was drawn in order to separate the classes as effectively as possible.

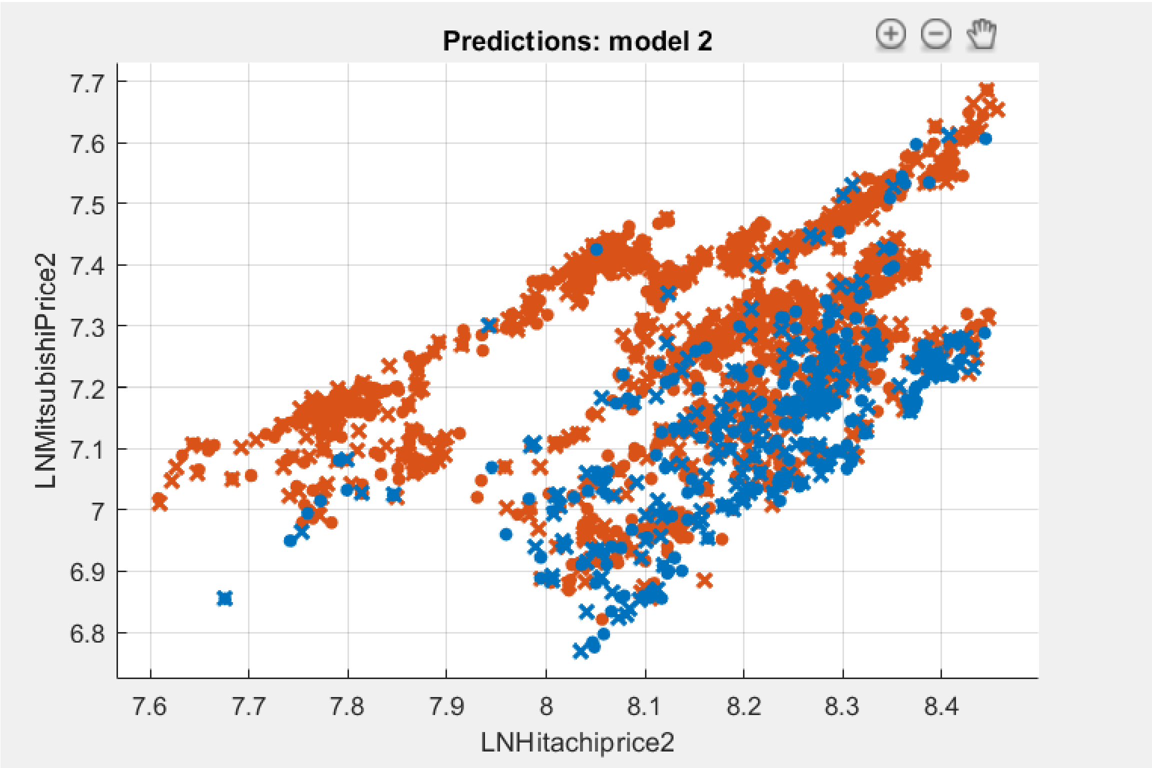

Figure 15-

Source: Matlab

Moving on to the confusion matrix, it seems that we reduced the classes that were corrected classified from 1153 to 1067 but the important thing is that we also reduced the ones that were not correctly classified before from 900 to 816.

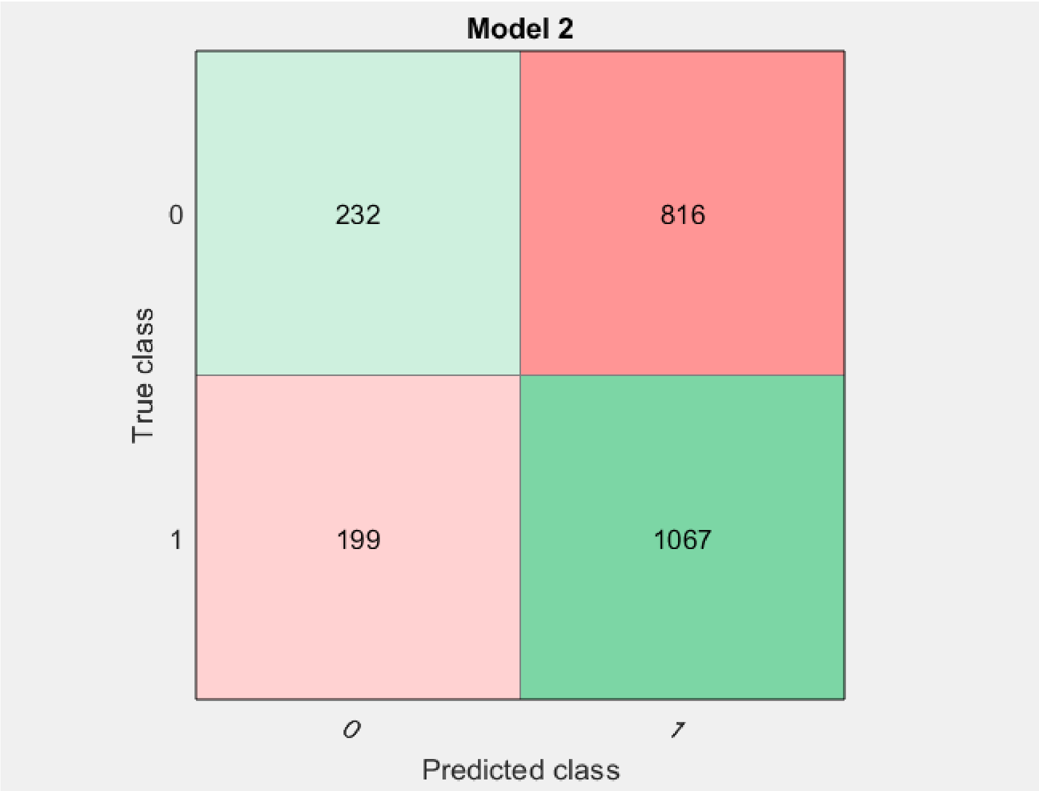

Figure 16-

Source: Matlab

Lastly, from the ROC curve, the results that are drawn are similar to our previous model. The forecasting accuracy remains more or less the same with our current classifier not being able to achieve high accuracy while standing on the left and down side of our curve.

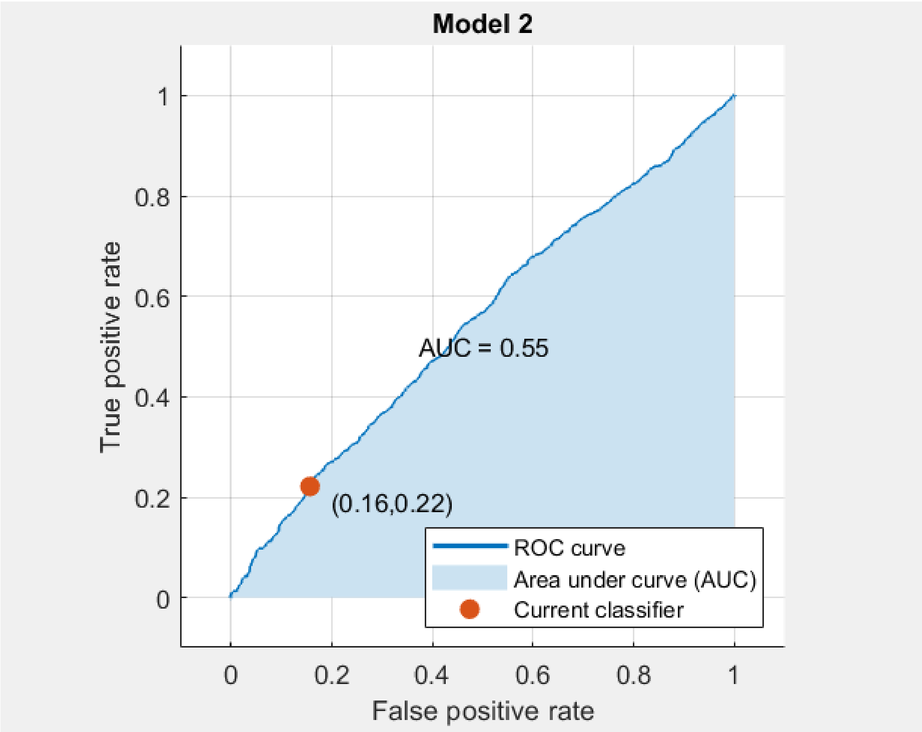

Figure 17-

Source: Matlab

Conclusions and Recommendations

Taking into account the existing literature review, our research does not outperform the already established models. Despite the fact that all the main stock indices and the closing prices of big technology companies were taken into account, the highest forecasting accuracy that we achieved was 56,2%. Additionally, it is worth mentioning that the model that achieved the best forecasting accuracy followed the algorithm ensemble, combining techniques in order to have this result.

We could conclude that the bitcoin follows the efficient market hypothesis or EMH. According to Fama (1970), the efficient market hypothesis is an investment theory where the share prices reflect all information. Theoretically, only inside information can result in big risk-adjusted returns. According to EMH, the stocks always trade at their fair value and investors could not purchase undervalued stocks or sell stocks for inflated prices. Considering this, it would be impossible to outperform the market.

In our case, bitcoin seems to fall under the category of being consistent with the efficient market hypothesis, since our analysis could not predict with high accuracy the next day's prices, using all the information that is available today.

As a recommendation, we conclude that more factors should be taken into account, like indices from global markets combined all together. What is more, since our stepwise least squares regression pointed out only Asian market traded big technology companies, it should be of high importance to perform further investigation on more companies that could affect bitcoin's prices. Companies across sectors could be used and of course we suggest to mix big caps with mid and small caps as well.

In the next part, we find really interesting to discuss about some principle of the prudential treatment of cryptocurrencies as these would definitely affect the future trends and it would be of high importance for next researchers to take those into account.

Prudential Treatment of crypto-currencies

Interesting insights on principles of a prudential treatment for crypto-assets can be found in the recently published paper from Bank of International Settlements. Since Bitcoin plays a significant role in our everday life affecting the financial institutions, some general principles should be followed by anyone how is engaged in transactions including crypto-assets.

- Crypto-assets for intra and inter bank settlements: Crypto-assets that are used exclusively for intra and inter bank settlements should have a different risk profile compared to highly-risky crypto-assets.
- Crypto-assets that use stabilization tools linked to other assets: A different risk profile could be also applied to those that represent a claim on an underlying asset and are verifiably backed by the other tangible asset. The risk of course would depend on the governance in order to ensure valuation stability.
- Disclosure requirements: Banks should be required to disclosure information on any material crypto-asset holdings on a quarterly basis and more specifically the report should include the following information: (i) the exposure amounts of different direct and in-direct crypto-asset exposures, (ii) the capital requirement for these exposures and (iii) the accounting disclosure of such exposures.
- Same risk, activity and treatment: A crypto-asset may not be a “traditional asset” and should be treated equally with the aforementioned. The prudential framework should not be designed in a way to advocate or avert specific technologies linked to them, but instead it should account for any traditional risks resulting from the unique features and all the other factors that drive traditional assets.
- Simplicity: As the use of crypto-assets is growing rapidly, the prudential treatment should be simple and flexible in nature. Since complex approaches may take time to develop and the prudential treatment is to be built on the existing framework, simplicity is of essence.

- Minimum standards: Any prudential treatment should require a minimum standard. Jurisdictions of course are free to apply any additional or more conservative measures.

For future research, it would be recommended that all factors rising from the above prudential treatment should be taken into consideration. Bitcoin is evolving rapidly and only if we are able to outrun it, we will gain further insight on its trends.

Bibliography

Almeida, J., Tata, S., Moser, A., & Smit, V. (2015). Bitcoin prediciton using ANN. Neural networks, 1-12.

BIS (2019). Designing a prudential treatment for crypto-assets. [online] Available at: http://www.bis.org [Accessed 3 Dec. 2019].

B. Trafalis., T. (2003). Data Mining Techniques for Pattern Recognition: Tornado Signatures in Doppler Weather Radar Data. *International Journal of Smart Engineering System Design*.

Cheung, K. (2019). *Home*. [online] Algorithm-X Lab. Available at: https://algorithmxlab.com/ [Accessed 6 Dec. 2019].

Chowdhury, Reaz & Arifur Rahman, M & Sohel Rahman, M & Mahdy, Mahdy Rahman Chowdhury. (2019). Predicting and Forecasting the Price of Constituents and Index of Cryptocurrency Using Machine Learning.

Culkin, R. and R.Das, S. (2017). Machine Learning in Finance: The Case of Deep Learning for Option Pricing. [online] Srdas.github.io. Available at: https://srdas.github.io/Papers/BlackScholesNN.pdf [Accessed 14 Dec. 2019].

Crosby, M., Nachiappan, Pattanayak, P., Verma, S. and Kalyanaraman, V. (2016). Blockchain Technology: Beyond Bitcoin. Applied Innovation Review, (2).

Das, D., Goel, A. and Misra, R. (2015). *Broadcast Algorithm for Wireless Networks*. Saarbrücken: LAP LAMBERT Academic Publishing.

E. Cavanaugh, J. (1999). Generalizing the derivation of the schwarz information criterion", Communications in Statistics - Theory and Methods.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), p.383.

Hastie and Tibishari (2009). Cross-validation and bootstrap. [online] Available at: https://web.stanford.edu/~hastie/Papers/ESLII.pdf [Accessed 6 Dec. 2019].

Heaton, J. and Polson, N. (2016). Deep Learning for Finance: Deep Portfolios. SSRN Electronic Journal.

Huberman, G., Leshno, J. and Moallemi, C. (2017). Monopoly Without a Monopolist: An Economic Analysis of the Bitcoin Payment System. SSRN Electronic Journal.
Dyhrberg, A. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. Finance Research Letters, 16, pp.85-92.

Investopedia. (n.d.). *Investopedia*. [online] Available at: https://www.investopedia.com/ [Accessed 16 Nov. 2019].

Jang, H., & Lee, J. (2017). An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. Ieee Access, 6, 5427-5437.

Klein, T., hien, p. and Walther, T. (2018). Bitcoin Is Not the New Gold: A Comparison of Volatility, Correlation, and Portfolio Performance. SSRN Electronic Journal.

Liew, Venus Khim−Sen, (2004) "Which Lag Length Selection Criteria Should We Employ?." Economics Bulletin, Vol. 3, No. 33 pp. 1−9

Moallemi.com. (2019). [online] Available at: https://moallemi.com/ciamac/papers/seqrisk-2010.pdf [Accessed 6 Dec. 2019].

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. Finance Research Letters, 27, 235-240.

Poyser, O. (2017). Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series. arXiv preprint arXiv:1706.01437.

Salisu, A., Isah, K. and Akanni, L. (2019). Improving the predictability of stock returns with Bitcoin prices. The North American Journal of Economics and Finance, 48, pp.857-867.

Showsiteinf.org. (2018). *svitsmbnr.ac.in - Sree Visvesvaraya Institute of Technology & Science*. [online] Available at: http://showsiteinf.org/sites/svitsmbnr.ac.in [Accessed 3 Dec. 2019].

Srdas.github.io. (2015). *SCU Web Page of Sanjiv Ranjan Das*. [online] Available at: https://srdas.github.io/ [Accessed 14 Dec. 2019].

Vo, N. N., & Xu, G. (2017, October). The volatility of Bitcoin returns and its correlation to financial markets. In 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC) (pp. 1-6). IEEE.

Yang, S. and Kim, J. (2015). Bitcoin Market Return and Volatility Forecasting Using Transaction Network Flow Properties. SSRN Electronic Journal.

Appendix

Dependent Variable	*Source*
Bitcoin Close	www.coinmarketcap.com
Explanatory Variables	***Source***
S&P 500	www.investing.com
Dow Jones Industrial	www.investing.com
Nasdaq	www.investing.com
Nasdaq_100	www.investing.com
Nyse_Composite	www.investing.com
FTSE_100	www.investing.com
FTSE_250	www.investing.com
FTSE_All_share	www.investing.com
FTSE_Tech	www.investing.com
Nikkei_225	www.investing.com
Nikkei_1000	www.investing.com
Jasdaq	www.investing.com
TOPIX	www.investing.com
Topix_1000	www.investing.com
Hitachi	www.investing.com
Tokyo_electro	www.investing.com
Mitsubishi	www.investing.com
Fujitsu	www.investing.com
NTT	www.investing.com
Itochu	www.investing.com
Apple	www.investing.com
Microsoft	www.investing.com
IBM	www.investing.com
Intel	www.investing.com
Cisco	www.investing.com