REPORT - GROUP WORK

PROJECT SUBMISSION 3

Students:

• David Kwasi Nyonyo Mensah-Gbekor (mensah-gbekor@hotmail.com)

• David Wonder Doe-Dekpey (wonderdoe85@yahoo.com)

• Alexander Victor Okhuese (alexandervictor16@yahoo.com)

• Alexandre Botica (alexbotica@yahoo.com)

INTRODUCTION

The use of data mining and time series has been a research topic of great relevance to the market as a whole (Last, Kandel and Bunke, 2004). Companies are making use of data mining to track consumer profile, for example. Competitive intelligence is being used to perform informational monitoring allowing companies to act in advance and reduce operating risks. According to (Bach et al., 2019), in the specific case of forecasting time series related to the stock market, it is important to predict more accurately the trend of movement (rise or fall) of the series, since such characteristics can aid in the strategy of buying and selling them. Prices of financial assets generally depend on other market prices. Volatility may also depend on other events such as government announcements of unemployment rates, inflation, quarterly GDP, interest rate, among others (for International Settlements, 2016).

Considering the foregoing literature, it was notable the concern with the data chosen for analysis, but it is also important, according to (Rauchman, no date), to know that there is external interference, especially when it comes to financial market actions, where they are very volatile to government actions, or investments of third parties. Under these conditions, an opportunity arises in this task 3 to investigate possible variables considered relevant to a specific stock movement.

LITERATURE REVIEW

Performance Metrics

Evaluation metric plays a critical role in achieving the optimal classifier during the classification training (Hossin and Sulaiman, 2015). Thus, a selection of suitable evaluation metric is an important key for discriminating and obtaining the optimal classifier. In Machine learning, evaluation metrics are used to compare and select the best algorithms and modeling strategies to the problem in hand.

ROC

Sensitivity and specificity are often difficult to reconcile, that is, it is complicated to increase the sensitivity and specificity of a test at the same time. Receiver operator characteristic (ROC) curves are a way of representing the normally antagonistic relationship between the sensitivity and specificity of a quantitative diagnostic test over a continuum of cutoff point values (Junge and Dettori, 2018). To construct a ROC curve, a diagram representing the sensitivity as a function of the proportion of false positives (1-Specificity) is plotted for a set of cutoff points.

The ROC curve method was originally developed to evaluate the ability of radar operators to decide whether a spot on the screen represented an enemy target (an airplane or a ship) or an allied spacecraft, or whether it was a noise. In fact, ROC is the acronym for "Receiver Operating Characteristic", which can be freely translated as "signal receiving operator efficiency". It is, therefore, a measure of an observer's ability to correctly classify a data within a dichotomous key.

CONFUSION MATRIX

It is a table that shows the classification frequencies for each class of the model.

• True positive (TP): Occurs when in the actual set, the class we are looking for was correctly predicted. For example, when the woman is pregnant and the model predicted correctly that she is pregnant.

• False positive (FP): occurs when in the actual set, the class we are looking to predict has been incorrectly predicted. Example: The woman is not pregnant, but the model said she is.

• True false (TN): occurs when in the real set, the class we are not looking to predict has been predicted correctly. Example: The woman was not pregnant, and the model predicted correctly that she is not pregnant.

• False negative (FN): occurs when in the real set, the class we are not looking to predict has been incorrectly predicted. For example, when the woman is pregnant and the model incorrectly predicted she is not pregnant.

PRECISION

It tells me how my model got the predictions right. it is the ratio of the sum of the correct predictions (true positives with true negatives) to the sum of the predictions.

RECALL

What is the proportion of positives that were correctly identified? This metric gives us a measurement of how good our model is to predict positives, being positive understood as the class one wants to predict. It is defined as the ratio of true positives to the sum of true positives and false negatives.

F1-SCORE

This metric combines precision and recall to bring a unique number that indicates the overall quality of our model and works well even with data sets that have disproportional classes.

ACCURACY

What is the proportion of positive identifications was really correct? The accuracy gives us an idea of how well our model worked.

AUC

The AUC and ROC curves are among the most used metrics for the evaluation of a Machine Learning model. The AUC is derived from the ROC curve. The AUC emerged as an attempt to simplify the ROC analysis. Hence, the AUC (area under the ROC curve) is nothing more than a way to summarize the ROC curve into a single value by aggregating all ROC thresholds, calculating the "area under the curve ".

The AUC value ranges from 0.0 to 1.0 and the threshold between the class is 0.5.

That is, above this limit, the algorithm classifies in one class and below in the other class.

The higher the AUC, the better.

Theoretical Considerations for Fund Facts

In this section we will present some concepts related to the creation of the fund facts. Namely: Maximum Drawdown, Annualized Returns, Sharpe Ratio e Plot the Equity Curve.

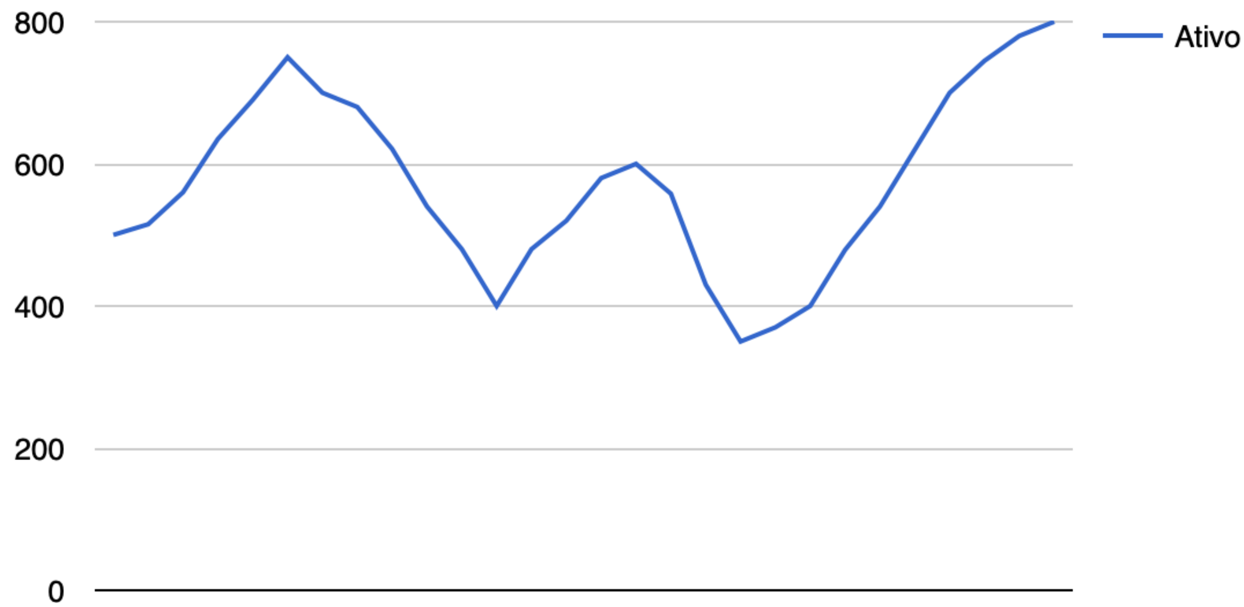
1. Maximum Drawdown

Maximum Drawdown (MDD), or maximum loss, is a risk indicator that shows the highest loss occurring from a point of high (peak) to a minimum point in a historical series (Sancetta and Satchell, 2004). That is, it measures the largest decrease in the value of an asset, in percentage, between a peak that occurred in the past and a later date. It can be used individually as a risk measure, but can also be used to compose an analysis.

An example: suppose an investent portfolio has an initial value of $ 500.00.

During the analyzed period, the amount increases to $ 750.00 and then begins to fall, reaching $ 400.00. Then it goes up to $ 600.00 and falls back, reaching $ 350.00.

Finally, the price more than doubles and reaches $ 800.00. Looking at the chart below, we can have an idea of the behavior of the asset of this example.



The MDD in this case is = ($ 350 - 750) / $ 750 = - 53.33%

2. Annualized Return

According to (Chen, no date), an annualized total return is the geometric average amount of money earned by an investment each year over a given time period. It is calculated as a geometric average to show what an investor would earn over a period of time if the annual return was compounded. An annualized total return provides only a snapshot of an investment's performance and does not give investors any indication of its volatility.

3. Sharpe Ratio

The Sharpe Ration - also known as the Sharpe Ratio - is an indicator used to analyze the statistical performance of funds and portfolios. Its main function is to draw a parallel between the return and the volatility of the analyzed portfolio. The sharpe ratio can help us bring the answer to the following question: What is the investment that offers the highest possible return for the lowest possible risk? Sharpe ratio was created to solve a fairly common problem in the financial market: the tendency to look only at the profitability of assets.

According to (Carter, Macdonald and Cheng, 1997), analyzing just the profitability of an investment is a very simplistic (and wrong) way of assessing whether or not it is good. Past return is no guarantee of future return.

Equity Curve

An equity curve is a graphical representation of the change in the value of a trading account over a time period (Chen, 2018). An equity curve with a consistently positive slope typically indicates that the trading strategies of the account are profitable, while a negative slope shows that they are generating a negative return.

METHODOLOGY

The methodology used to perform the tasks required to submission two consisted of a thorough literature review to understand the concepts and the nature of the metrics we were supposed to implement.

For the modeling part we followed an order that started with acquiring the data, preprocessing, selecting, cleaning, featuring and modeling. The strategy we chose was to download the stock in S&P 500 index. We downloaded 504 stock prices plus the index prices itself. Then we performed some tasks to prepare the data to the next steps. Our final goal was to perform a modeling to predict whether the stock prices were going up or down, hence, a classification problem. From the 504 stocks within our index, we have chosen 5 of them using PCA. These 5 components explained around 88% of our system behavior. After that we ran some data transformation tasks, including outlier handling, log transformation and normalization. Appling the knowledge we gained from Submission 2. Then we split the data into training and test sets to start the modeling step.

For the modeling tasks, we selected 4 algorithms for classification tasks: logistic regression, OneVsRestClassifier, Gaussian Nayve Bayes and Neural Network. We implemented all 4 using scikit learn from Python. The results are discussed in the next session.

RESULTS

The graphical and detailed numeric results are available in the notebook file and HTML file. Here we will briefly discuss our results.

From our results we can see that our models do not explain very well the market behavior. With all the models with accuracy no bigger than 0.53, which is a poor result. The weakest results were the ones from the Gaussian Nayve Bayes model, which is expected because this is a very simple classification algorithm. On the other hand, neural networks, that tend to have very good results, did not outperformed all other models. What we had was what we judged as a tie between neural net and SVC, because one performed better than the other depending on the metric analyzed. For example, for accuracy, confusion matrix and f1-score results, SVC outperformed neural nets by a small margin. For precision and area under the curve, neural net was the winner.

CONCLUSION

We had a tough time working though this submission. However, the learning was worthy and rewarding. The experience in applying machine learning to financial data was challenging and fun. Regarding the content, we worked on financial data modeling and validation. On this submission we focused on modeling and applying formal metrics to a classification problem using S&P index data. After some analysis, we concluded that, to our case, SVC and neural networks had very similar performance. However, no model had and overall good performance, even though the 5 most important components explained more than 88% of the variation.

REFERENCES

Bach, M. P. et al. (2019) 'Text mining for big data analysis in financial sector: A literature review', Sustainability (Switzerland), 11(5). doi: 10.3390/su11051277.

Carter, S., Macdonald, N. J. and Cheng, D. C. B. (1997) Basic finance for marketers

(Marketing and agribusiness texts-1) by the FAO Regional Office for Africa. Available at: http://www.fao.org/docrep/W4343E/W4343E00.htm (Accessed: 20 May 2019).

Chen, J. (2018) Equity Curve, Investopedia. Available at: https://www.investopedia.com/terms/e/equity-curve.asp (Accessed: 20 May 2019).

Chen, J. (no date) Annualized Total Return, Investopedia. Available at: https://www.investopedia.com/terms/a/annualized-total-return.asp (Accessed: 20 May 2019).

International Settlements, B. (2016) 'Inflation mechanisms, expectations and monetary policy', in. Available at: www.bis.org (Accessed: 20 May 2019).

Hossin, M. and Sulaiman (2015) 'A REVIEW ON EVALUATION METRICS FOR DATA CLASSIFICATION EVALUATIONS', International Journal of Data Mining & Knowledge Management Process (IJDKP), 5(2). doi: 10.5121/ijdkp.2015.5201.

Junge, M. R. J. and Dettori, J. R. (2018) 'ROC Solid: Receiver Operator Characteristic (ROC) Curves as a Foundation for Better Diagnostic Tests.', Global spine journal. SAGE Publications, 8(4), pp. 424–429. doi: 10.1177/2192568218778294.

Last, M., Kandel, A. and Bunke, H. (2004) 'Data Mining in Time Series Databases', 57(January 2011). doi: 10.1142/5210.

Rauchman, M. (no date) Big Data in Capital Markets. Available at: https://www.sigmod.org/2013/keynote1_slides.pdf (Accessed: 20 May 2019).

Sancetta, A. and Satchell, S. E. (2004) 'Calculating hedge fund risk: the draw down and the maximum draw down', Applied Mathematical Finance, 11(3), pp. 259–282. doi: 10.1080/1350486042000220553.