Black Swans, Financial Markets and Machine Learning: when the unthinkable happens

Niccolo' Gentile

Abstract

In recent years, the application of Artificial Intelligence to finance has been steadily growing, in particular regarding one of its most popular subsets, Machine Learning. Aim of this research proposal is to review the applications and subsequent limitations in using these technologies on portfolio management, in particular in the case of sudden, unpredictable negative events as the Covid-19 crisis, and proposing possible solutions to deal with them.

1 Introduction

While become mainstream in the news only recently, *Artificial Intelligence* is a field of study with a long story, made up of alternating periods of fortune and discredit. The main reason why only recently it has again - and now more steadily - gained popularity is thanks to both the exponentially grown computational power of the machines (see *Moore's Law*) and the availability of so called *Big Data*, namely datasets that are both very rich in terms of observations and features, as well as very dynamic in their compositions.

As previously said, while having a long history, Artificial Intelligence as a field of research was founded in 1956, during the Dartmouth Summer Research Project on Artificial Intelligence. John McCarty, the computer and cognitive scientist who coined the term, defined it as *"the science and engineering of making intelligent machines"*. Another definition for Artificial Intelligence defines it as *"the branch of computer science dealing with the simulation of intelligent behavior in computers"*. More precisely, with intelligence, here we refer to "the computational part of the ability to achieve goals in the world". It is therefore immediate, as such, to notice how also *Learning* falls under this definition. *Learning*, loosely speaking, is "the process of gaining knowledge and expertise" and as such it represents *one of the many possible expressions of intelligence.* Hence, it is immediate to define *Machine Learning* as a branch, a subset of *Artificial Intelligence.* The term Machine Learning has been coined by Arthur Samuel in 1959, defining it as is *"the field of study that gives computers the ability to learn without being explicitly programmed"*. A currently widely used definition for Learning applied to machines is the one provided by another pioneer of the field, Tom Mitchell. According to him, *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E"*. In the context of *Statistical Learning*, with *experience* we refer to the *data*, the *task* is that of making a *prediction* and the *performance measure* is one of the many possible *loss functions* or *accuracy metrics* employed in estimating the distance between a predicted value and the corresponding observed one. It is then clear what are *Artificial Intelligent Systems that are not Learning*: basically, they are all those algorithms that are based on $If - Then$ statements, hard coded in the computer, and that *require human intervention to be changed.* In other words, they're *static.* Conversely, Machine Learning algorithms are not explicitly programmed, and adjust themselves - without human interaction - based on the experience they gain over time (as previously said, data).

This distinction is often blurred in the media, where the two terms are so often wrongly used as synonyms or interchanged. It is clear that the application of Artificial Intelligent systems endowed with Learning

capabilities or not responds to very different necessities. Given the aim of this research proposal, we will anyhow indeed focus more on Machine Learning algorithms, since these are the ones that can more flexibly produce predictions starting from a subset of data.

In *Finance*, these algorithms have an enormous appeal. The higher the degree of accuracy by which it is possible to predict the future of a single company, an index, an entire industry or the economy as a whole, the better a portfolio manager can define his or her investment strategy. The rest of this paper is structured as follows: in *Section* 2, we provide a formalized, brief description of what Machine Learning, useful to then better grasp *Section* 3, where we describe its limitations with respect to unpredictable events, going through a review of what has happened to investors who heavily rely on these algorithms during the Covid-19 "Black Swan". In *Section* 4, we finally review where the literature stands in taming these limitations, plus some proposals.

2 Machine Learning: objectives and key algorithms

Machine Learning algorithms can, in turn, be divided in multiple categories. Of these, *three* are commonly considered: *Supervised Learning*, *Unsupervised Learning* and *Reinforcement Learning*.

- *Supervised Learning*: in this case, the *task* is to *learn a function* mapping a set of features onto a target. The *experience* are the *labeled data* contained in a subset of individuals, the *training set*, used to learn the map, where with "labeled" we mean that every individual is defined both in terms of his or her input features and the target one. The *performance measure* is the capability of the learned map (learned on the training set) to *predict accurately* on new individuals, the *test set*. We further talk of *Regression* problems when the the target variable is continuous, of *Classification* when it is discrete. It is called "supervised" since indeed the learning process is supervised by us, in the sense that we defined what the objective is. Some key Supervised Learning algorithms include *Linear and Logistic Regressions, Regression Trees and Random Forests, k-nearest-neighbors, Support Vector Machine, Neural Networks*.

- *Unsupervised Learning*: in this case, the *task* is to learn useful properties of the structure of the dataset, in order to provide an easier representation of the data without loss of information, only *experiencing* each individual's features. *Performance measures* are used to balance the loss of information due to

the representation and the its accuracy. We talk about *Dimentionality Reduction* algorithms when

the aim is that of reducing the number of input features (without loss of relevant information), while

we talk about *Clustering* algorithms when the aim is finding clusters of individuals in the dataset

based on similarity measures in their features. Some key Unsupervised Learning algorithms include

(Kernel) Principal Component Analysis, Compressed Sensing, t-SNE, $k-means-clustering$.

- *Reinforcement Learning*: in this case, the *task* of the machine is that of maximizing a cumulative re-
  ward. Here, the machine interacts with an environment, namely a world in which the agents operate
  in his or her state, hence choosing an action based on a policy function defining the probabilities of
  taking one of many possible actions. Once the action is performed, the *performance* is defined via
  a reward associated with a value function, and the *experience* is as such defined via an interaction
  between the data and the environment, hence it is not just a fixed dataset as in supervised and un-
  supervised learning. Some key Reinforcement Learning algorithms include *Monte Carlo, Q-Learning,*
  State–action–reward–state–action (SARSA), Deep Q Networks.

In Finance, Supervised Learning algorithms can intuitively be useful to predict stock prices, the future states

of an industry or the economy as a whole, as well as binary outcomes. Unsupervised Learning can be useful

to empirically identify, for instance, cluster of stocks or industries we didn't think of - simply out of theory

- by ourselves, as well as identify a subgroup of stocks or industries that already suffice in subsequently

build predictive models for all the others. Reinforcement Learning algorithms find applications in Portfolio

optimization, Optimized trade execution and Market-making. While all of the three assume a key role in

Finance and Trading, an analysis of the limitations of all of them is beyond the scope of this work. Here, we

focus mostly on *Supervised Learning*: its limitations are, anyhow, comparable to those of the other two,

especially *Reinforcement Learning*.

Suppose we observe k features encoded in vectors $\boldsymbol{x}_i \in \mathcal{X} \subseteq \mathbb{R}^k$ for each individual i in the sample, and a

target of interest to be predicted $y_i \in \mathcal{Y} \subseteq \mathbb{R}$. Let $\mathcal{F}$ be the set of functions we are considering, mapping the

features $\mathcal{X}$ onto $\mathcal{Y}$, $Z = \mathcal{X} \times \mathcal{Y}$ and $l : \mathcal{F} \times Z \to \mathbb{R}_+$ a *loss function* used to assess the performance of the

a mapping function $f \in \mathcal{F}$. The task of the algorithm is finding a mapping function $f \in \mathcal{F}$ minimizing the

risk function, namely the *Expected Value* of the loss function l:

$$Err_D(f) = E_{(\boldsymbol{x},y)}\left[l(f,(\boldsymbol{x},y))\right] = \int_{\mathcal{X}\times\mathcal{Y}} l(f,(\boldsymbol{x},y))g(\boldsymbol{x},y)d\boldsymbol{x}dy \tag{1}$$

with $g(\boldsymbol{x}, y)$ following a distribution $\mathcal{D}$ over Z being the *data generating process*. Clearly we don't know

$g(\cdot)$, and as such, upon extracting a subset of individuals (the *training set*) from $\mathcal{D}$, the algorithm minimizes the *Emprical Risk Function*, namely:

$$Err_{train}(f) = \frac{1}{n} \sum_{i=1}^{n} l(f, (\boldsymbol{x}_i, y_i)) \tag{2}$$

In practice, though a regularization term is added in (2), controlling the complexity or roughness of the function:

$$\hat{f}_{min} \in \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} l(f, (\boldsymbol{x}_i, y_i)) + \lambda R(f) \tag{3}$$

with $\hat{f}_{min}$ the learned map and λ a *tuning hyperparameter* trading-off *bias* and *variance* of $\hat{f}_{min}$.
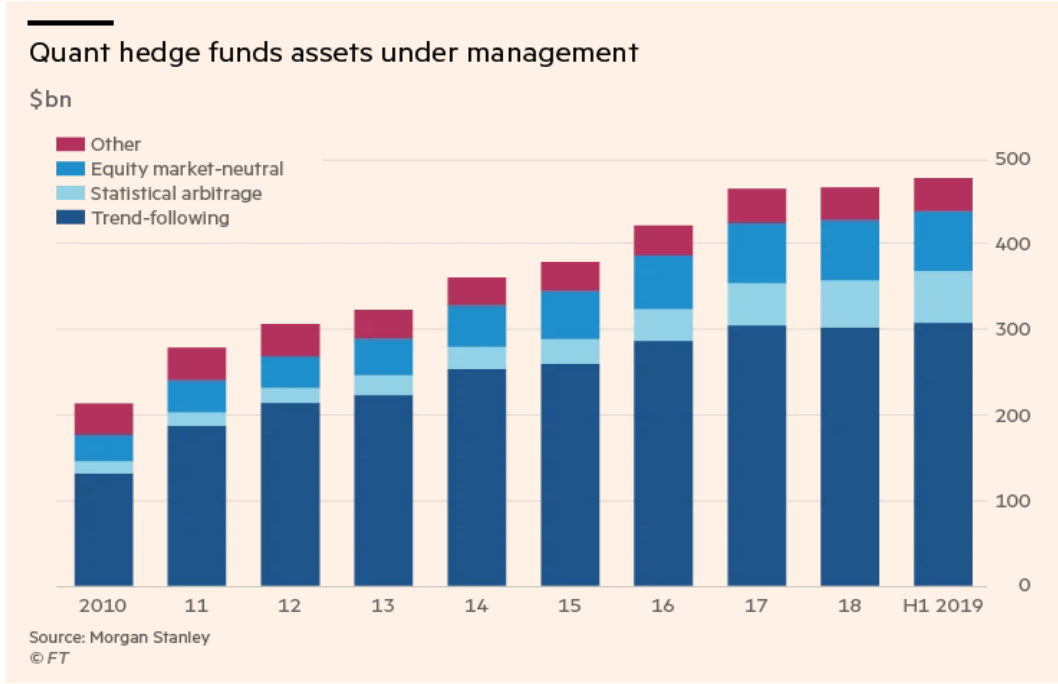
It is immediate to infer how this framework can be applied identically to *Time Series* data, considering T periods of observations for a single unit instead of one period for n different of them ($Cross - Sectional$ data).

3 Machine Learning and Finance: applications, limitations and Covid-19 crisis

As we said in the Introduction, Machine Learning algorithms are increasingly applied in Finance. In their work, Berat Sezer et al. (2019) propose a comprehensive review, focusing especially on the applications of *Deep Learning* algorithms to forecasting. Deep Learning is a subset of Machine Learning in which the algorithms not only learn a map from the features to the target, but also the best possible *representation* of the data for the predictive work. *Artificial Neural Networks* are the quintessential Deep Learning algorithm. In a 2015 bulletin of the Bank of England, David Bholat, Stephen Hansen, Pedro Santos and Cheryl Schonhardt-Bailey explored the potential applications of Machine Learning algorithms to perform *Natural Language Processing* (NLP) and text mining over Central Bankers' statements, well-known to be of particular interest for financial markets, since even using one word instead of another can strongly halter trades and economic perspectives ("close but below to 2%"). In a recent paper, *"How to talk when a machine is listening : corporatedisclosureintheageofAI"*, Cao et al. (2020) investigate how NLP algorithms are now also applied to CEOs' disclosures. As of today, top Quantitative Hedge funds include *Two Sigma, Renaissance Technologies, DE Shaw, AQR, and Capula,* all actively employed quantitative strategies, and Ray

Dalio - founder of *Bridgewater Associates* - in his book, *Principles*, stresses in multiple occasions the role that quantitative methods have assumed and continue to assume in his hedge fund's decision processes. Reinassence Technologies is defined as "one of the most secretive and successful" hedge funds in the world, and on Two Sigma's website it is possible to find multiple publications and papers also on Machine Learning. Some of the leading researchers in the field include Dr. Marcos Lopez de Prado - whose most famous textbooks on the topics are *"Machine Learning for Asset managers"* and *"Advance in Financial Machine Learning"* - and Dr. Matthew Dixon, whose most famous textbook on the topic is *"Machine Learning in Finance: From Theory to Practice"* (clearly, for both, along with a long list of papers on the topic).

As a final remark, a distinction between $Factor-based\ Quantitative$ strategies and $Machine-Learning-based$ strategies is due. In the former strategy, an hypothesis is made, and tested against past data. If confirmed, a new test is run on live data, and then it is constantly checked if the factor, or signal, remains. Conversely, in the latter, no hypothesis is made, no underlying theory is formulated, and the whole process is merely data-driven, with ample margin left to the algorithms to learn and develop "their theories". This approach, can lead to better performances since allows to uncover patterns and relationships otherwise unthinkable. Nonetheless, its often "black box" nature can make them hard to explain, and as Mr. Kharitonov, CEO of Voleon spcified in a Financial Times' article of the 17th of October 2019 *Why hedge fund managers are happy to let the machines take over*, quoting him, the majority of the patterns discovered "have no simple economic rationale".

Quant hedge funds assets under management

$bn

Legend:
- Other
- Equity market-neutral
- Statistical arbitrage
- Trend-following

Source: Morgan Stanley
© FT

One of the key issues of Machine Learning is that they are prone to *overfitting*, a situation in which the algorithms *model the noise rather than the signal*, meaning that the learned model is too specific to the training data and poorly generalizable to new test, unobserved individuals. This implies that the models may be poorly capable of dealing with extreme values in the testing phase. When referring to *Time Series* data, we refer to testing as predicting on unobserved periods. More precisely, the idea is that of learning a model capable of predicting one period ahead. That is, for instance, using data in $t-3$ we learn a function $\hat{f}$ capable of predicting in $t-2$, and then we test it using $t-1$ to predict t. Considering $train\ set = (\boldsymbol{x}_{t-3}, \boldsymbol{x}_{t-2})$, $test\ set = (\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)$, we learn the mapping $\hat{f}_{t-3}$ s.t. $\hat{x}_{t-2} = \hat{f}_{t-3}(x_{t-3})$. Later, we then try to use this same function to predict from $t-1$ to t, using this test loss to assess the validity of the learned map: $\hat{x}_t = \hat{f}_{t-3}(x_{t-1})$. If in period t, however, the *data generating process* - that is, the underlying distribution - changes sharply and becomes significantly different from those of the previous periods, inevitably our model will fail to represent the new data distribution, hence leading to problems with forecasting. This limitations was well-described by Robert Pozen in an article of the 2nd of October 2019 on the Financial Times, *Will bots replace humans in active equity investment?*. Quoting it: "ML derives its conclusions from existing data points fed into trained algorithms. It cannot predict future to the extent that the future patterns are not rooted in the past, such as major discontinuities in the 2008 financial crisis". A premonition of what would have happened less than one year after.

This issue is also related with the research for *correlations* that Machine Learning algorithms perform when learning from the data. Such correlations - found by algorithms without a specific understanding of the underlying rationale - can be the result of an overfitting behavior, hence a modeling of noises rather than signals.

As such, one of the arguments for critics of this approach, is that in periods of downturn (in particular) the automated selling can easily exacerbate the situation (a similar critique is moved from active investors against ETFs and passive investors in general). This is in particular an issue of $Volatility\ Targeting\ Funds$, as well as $trend-following\ hedge\ funds$ and $risk-parity$ funds. In August 2019, they represented 1 trillion dollars in assets, as reported in a Financial Times' article of the 8th of August 2019, *"Automated selling has exacerbated US market swings, say analysts"*. In the same article, it is quoted Michael Lewis, head of US equity cash trading at Barclays, who said that "Systemic flows were a large part of the market activity on Monday", in which a 3% drop on the SP 500 index had occurred. During November 2019, also, trend-following hedge funds were accused of exacerbating the sell-off of Japanese government bond, reverberating the effect on other debt markets (*"Fingers point at hedge funds after Japan bond sell-off"*). As of January 2020, in another article on the Financial Times, *"'Quant winter' raises tricky questions for a hot industry*, it is reported that Joseph Mezrich, head of quantitative strategy at Nomura Instinet, argued that the industry was going through a 2 years "quant winter", with just 15% of quant mutual funds having beaten the US stock market in the preceding year, a performance trailing even traditional stockpiling funds (already suffering their own bad performance). In the article, the main reference is to those quantitative funds that use factors. Possible explanations ranged from the increasing popularity of these strategies eroding their capability to generate alpha as well as the current season of seesawing interest rates, sending mixed signals. Another occasion in which quant-driven investment strategies have been accused of exacerbating the downturn, incurring in particularly marked losses, was in August 2007. The $Quantitative\ Investment\ Strategies$ department at Goldman Sachs had under management 165 billions of dollars as of August 2007, as reported in Financial Times' article *"Goldman Sachs' lessons from the quant quake"*, published on March the 8th, 2017. In another article - *"Quant quake shakes hedge-fund giants"* - published on $Market\ Watch$ on the 13th of August 2007, it is reported that "its [of Goldman Sachs] 9 billion Global Alpha quant fund is down 27 so far this year, with more than half of those losses coming last week", along with another 30% decrease in another of its quant funds, $Global\ Equity\ Opportunities$. The widely employed leverage surely did not help, in this occasion. The similarity of positions in the field triggered a chain reaction, with $Renaissance\ Institutional$ $Equities\ Fund$ (RIEF), losing 9% in the same month, while $Medallion$, the Reinassance fund that manages

the firm's employees' money, remained overall positive up to the end of the year. These results have to be compared with the historical results of these two funds: RIEF, founded in 2005, had returned about 20% the previous year, while Medallion had had an average yearly annual return of about 36% from 1988. Jim Simons himself referred to the *overcrowding issue* in a letter sent after the aforementioned downturns of his funds as possible explanation for the particularly severe downturn. A similar argument was given by Mr. Asness, of *AQR Capital Management* ,another quantitative investor caught up in the August's 2017 downturn. Going back to the 2017's Financial Times article, it sounds close to prophetical when in it there's written that "some analysts fear that another 2007-style meltdown would be more severe due to the proliferation of quant strategies". This prophecy manifested itself fully during the Covid-19-induced financial markets' crash of March and April 2020. In another Financial Times' article of the 24th of March 2020, *"Glitchy coronavirus markets cause quant funds to misfire"*, it is reported that according to a Credit Suisse's, report the average quant fund had lost 14% during the year. Reinassence's aforementioned RIEF dropped 18% during March, leading to a total 24% fall on year-to-date performance. Such a fall has again to be compared with, instead, for instance, the Medallion fund, that has we said has lead to an about 40% average yearly return since 1988. The figure amounted at a loss of 20% as of June the 12th. Similarly, *Two Sigma*'s flagship fund *Spectrum* lost 2%, while its global macro *Compass* was down 13%. Another illustrious victim of March's meltdown was the fund *DE Shaw*, which fell down 2.6% in the month, the second time it fell in negative territory in two years. In a more recent Financial Times' article, *AI hedge fund Voleon suffers in choppy markets*, of the 7th of September, the bad performance of the previously mentioned *Voleon* fund is reported, who had suffered a loss of 9% on a yearly basis on its flagship fund *Investors*, a surprising fall for a fund that in the 2018, while the SP 500 4.4% after a sell-off, still was managing to make double-digits gains. This case is particularly interesting for our discussion, since this fund is well-known to strongly rely on Machine Learning algorithms to implement its statistical-arbitrage strategy, consisting of betting that short-term discrepancies from the mean will inevitably revert. According to Andrew Beer, managing member at US investment firm Dynamic Beta Investments, the problem with said algorithms is that in the 2015 - 2019 period they have learned to "buy in the dip", from which, as such, quoting him as in the article, "they caught the proverbial falling knife in the first quarter". Anyhow, in the end, AI hedge funds have overall gained 14.5% on a yearly basis, thanks to a 12.2% in an unusually strong August.

In another recent Financial Times' article, emblematically titled *"A terrible, horrible, no-good year for quants"* of the 3rd of November of the current year, a more detailed analysis of the fallout of these investors is introduced. In it, it is additionally referred a report titled *"Why I am no longer a quant"* by Mr. Inigo

Freser-Jenkins, head of quantitative strategy at Bernstein. In it, Mr. Freser-Jenkins echoed the same issue we (formally) described, regarding a possible change in the underlying joint data distribution of features and targets moving from one period to another, rendering a learned map on past data useless (if not harmful). In it, he argues indeed that one of the key issues in the development of quant strategies is that "quants mine historical data looking for clues of what can happen in the long run, glossing over the fact that market regimes come and go", quoting the Financial Times' article. In an October's survey by Refinitiv, it is coherently reported that 75% of the surveyed quants stated that their models had been hurt by the Covid-19 crisis, witha small yet considerable 12% of them stating that their models had become "obsolete". To put some end-of-years numbers, in the year to-end September, the average quant US mutual fund is up just 3.3% against the stockpicker's 8.3% gain and the Russel 1000's 6.4%, according to Bank of America. Weighting by assets, as at the end of August the average quant hedge fund *lost* 5.7% as the end of August, against the average hedge fund's *gain* 5.2%, according to Aurum Fund Management.

Source: Bank of America, Aurum
© FT

4 Lessons from the Covid-19 crisis and quantitative investing and possible proposal

The growth and the over-the-market returns quantitative strategies have produced historically - especially in the last 2 to 3 decades, in which data availability and computational powers of the machines have grown exponentially - have been remarkable. Nonetheless we have, so far, described how this year's turmoil has disrupted this sector of investors in an unprecedented manner. While not new to this kind of turmoils - see the aforementioned "quant quake" of 2007 - the disruption of this year's Black Swan, the Covid-19 pandemic, has been unprecedented. Given how recent the crash has been, a comprehensive literature about "what went wrong" has already to be developed. Nonetheless, in a paper published on the 30th of April, called *"Three Quant lessons from Covid-19"*, Alex Lipton and Marcos Lopez De Prado outlined three main lessons and patterns for future development for the field. The first is *"More Nowcasting, Less Forecasting"*. This critique is in line with the problem we have described previously: the key difference between $forecasting$ and $nowcasting$ is that under the former we focus on using structured data to predict long-term outcomes, while in the latter we rely on unstructured data to make short-term predictions, if not contemporary ones. Examples of nowcasting could be predicting the current level of inflation via continuous web-scraping of prices across Internet's marketplaces, or using satellite images for parking slots of malls. Before the sudden market sell-off started, there was a plethora of signals that Chinese supply chains were being disrupted: the sell-off would have not been a "Black Swan" for a careful nowcaster. The second critique is titled *"Develop Theories, Not Trading Rules.* The critique in this case is related with a sort of inversion that has happened in Finance with respect to the Scientific Method, regarding the use of data backtesting for validation. In the Scientific Method, backtesting on data assumes a central role in refusing an $already\ defined$ theory, summarized under the $Null\ Hypothesis$, to be assumed being the state of nature, hence to be true unless the data do not provide strong empirical evidences against it. In Finance, however, the opposite is now happening, since quants basically *use data directly to create theories rather than testing them.* In other words, backtesting has become part of the research process rather than of the validation one. The proposed solution is that of restarting to develop theories *without backtesting*, but rather using explanatory tools based on randomization of features that are resistant to overfitting, hence establishing a $cause-effect$ dynamic. Backtesting, in this framework, would get back to be a validation tool for trading strategies to be employed *under the correctly specified and already tested theoretical framework*. Their third and final lesson/critique is titled *"Avoid All-Weather Strategies"*. Their critique starts from the assumption, done by

practitioners and academics, that there exist a strategy capable of guaranteeing alpha under many different market regimes. According to them, empirical observations have invalidated this assumption, given the bad performance that multiple funds were already experiencing under the zero-interest regimes, and then the disastrous one during the Covid-19 storm. Structurally, the critique revolves once again around the problem of the different underlying joint distribution of target and features - the *data generating process*, DGP - relying under each different market regime. Hence, the practical solution would once again be based on nowcasting. With a proper nowcasting strategy, it could be possible to understand under which market regime the markets are more likely to currently be and will be in the narrow future, from which the possibility to continuously adjust the portfolio.

5 Conclusions

Overall, given the continuous growth of computational power, data availability and their instantaneous growth, quantitative investment strategies are likely to grow in popularity, including high-frequency-trading. Nonetheless, both as a consequence of possible overcrowding, as well as of sudden changes in the economic framework, this kind of strategies can easily backfire, especially those based on Machine Learning algorithms that may exacerbate the financial swings via their reinforced behavior. However, this problem is also for those strategies based on $Symbolic, Rule\ Based\ Artificial\ Intelligence$, in which a nested sequence of $if - else$ behaviors can also lead to overbuying and overselling. The proposed solutions by Lipton and Lopez De Prado represent a first step to tame these issues from the perspective of the agents in the markets, but more research is needed on the topic. Overall, opening up the so called *Black Box* of Machine Learning algorithms is now a recurring theme in the discipline, independently from the field in which these algorithms are employed to analyze data. Understanding the causal relationships between the features and the targets, and as such being able to infer a *causal framework*, as well as a *general theory*, is fundamental to assess and update accordingly the learned models, making them more resilient to any possible change in the environmental conditions.

References

[1] Bank of England - Financial Conduct Autority, *´'Machine Learning in UK Financial Services"*, October 2019.

[2] Barr A., Marketwatch, *"Quant quake shakes hedge-fund giants"*, Marketwatch, 13th of August 2007.

[3] Berat Sezer O., Ugur Gudelek M., Murat Ozbayoglu A., *"Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005 - 2019"*, Preprint submitted to Applied Soft Computing, November 2019.

[4] Bernot K., Gebauer S., Schäfer D., *"Monetary policy implications of financial innovation: In-depth analysis"*, DIW Berlin: Politikberatung kompakt, DIW Berlin, German Institute for Economic Research, volume 120, number pbk120, December 2017.

[5] Bholat D., Hansen S., Santos P., Schonhardt-Bailey C. *"Text mining for central banks "*, Centre for Central Banking Studies, Bank of England, 2015.

[6] Cao S., Jiang W., Yang B., Zhang A.L., *"How to talk when a machine i listening: corporate disclosure in the age of AI"*, NBER Working Paper No. 27950 , October 2020.

[7] Fletcher L., *"AI hedge fund Voleon suffers in choppy markets"*, Financial Times, edition of the 7th of September 2020.

[8] Financial Stability Board, *"Artificial intelligence and machine learning in financial services: market developments and financial stability implications"*, November 2017.

[9] Hastie T., Tibshirani R. Friedman J, *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"*, Springer Series in Statistics, Second Edition, January 2017.

[10] Henderson R., *"Automated selling exacerbated US market swings, say analysts"*, Financial Times, edition of the 8th of August 2019.

[11] Lipton A., Lopez De Prado M., *"Three quant lessons from COVID-19"*, April 2020.

[12] Pozen R., *"Will bots replace humans in active equity investment?"*, Financial Times, edition of the 17th of October 2019.

[13] Wigglesworth R., *"A terrible, horrible, no-good year for quants"*, Financial Times, edition of the 3rd of November 2020.

[14] Wigglesworth R., *"Glitchy coronavirus markets cause quant funds to misfire"*, Financial Times, edition of the 24th of March 2020.

[15] Wigglesworth R., *"Goldman Sachs' lessons from the 'quant quake'"*, Financial Times, edition of the 8yh of March 2017.

[16] Wigglesworth R., *"Quant winter' raises questions for a hot industry"*, Financial Times, edition of the 17th of January 2020.

[17] Wigglesworth R., *"Why hedge fund managers are happy to let the machines take over"*, Financial Times, edition of the 17th of October 2019.