Advanced Analytics & Artificial Intelligence
Business Unit @Value Digital Services
www.value.com.tn

Unsupervised diversification applied on the tunisian stock market before and during the covid-19 crisis

Ahmed Rebai[1], Louay Boukhris[1], Lotfi Ncib[2], Mohamed Anis Ben Lasmer[2]

[1] Value Digital Services, Tunis Tunisia, Advanced Analytics and Artificial Intelligence Business Unit [2] Esprit Tunis Tunisia, Esprit School of Engineering, Grafics Team

1. Introduction

When investing money in the stock market, rational investors try to avoid the risk of concentration and therefore diversify their investment over several securities. This study focuses on the use of machine learning to achieve better diversification. Financial data, related to companies listed on the Tunisian stock exchange (BVMT), were collected and analyzed according to the methodology applied in machine learning on over the periods before and during the Covid-19 crisis. This study shows, for instance, that the ML algorithms can perform a good segmentation according to several criteria and can discover the aberrant behavior of certain companies with an abnormal financial situation. These results were confirmed by other outlier detection algorithms.

[1] Ahmed Rebai, Louay Boukhris, Lotfi Ncib and Mohamed Anis Ben Lasmer (2021). Unsupervised Learning Diversification Applied on the Tunisian Stock Market Before and During the Covid-19 Crisis. International Journal of Management Research and Economics. 1(4), 24-47. doi: 10.51483/ IJMRE.1.4.2021.24-47.

[2] Ahmed Rebai and Louay Boukhris and Fayad Ali Banna, Microsegmentation for better diversification in the Tunisian Stock Exchange Market (BVMT) (September 20, 2021).

2. Context

The Tunisian financial market is a new market, hampered by strict banking regulations that put several obstacles to investors, such as the law n° 1994-25 of 7 February 1994 modifying the law n°1967-51 of 7 December 1967 regulating the banking profession. On the other hand, this market is located in a developing country ravaged by a deep structural economic crisis accentuated by a political crisis for 10 years and by the current Covid-19 crisis since March 2020. Therefore, the astute reader of this study is advised to understand the difficulties encountered during the analysis of the data and the interpretation of the results.

3. Diversification

Diversification is a fascinating concept in financial risk management that is used by various stakeholders such as investment managers, traders, quantitative analysts, and risk managers in monitoring financial market dynamics and building portfolios. First of all, after introducing diversification and explaining its important role, we will talk about four approaches used to achieve optimal diversification: the scientific approach (Markowitz, 1952; Fama and French, 1993 and 1996), the empirical approach or naive approach (DeMiguel et al., 2007), the new diversification strategies (Cont and Tankov, 2007) appearing after 2008 crisis, and the machine learning diversification strategies (Lopez de Prado, Marcos, 2019). A comparison between these four approaches will be given at the end of this section. Diversification consists in finding the best assets with their best contributions to build an efficient portfolio that allows reaping the highest premium corresponding to the risks incurred. In this respect, diversification is not an objective in itself, in other words, a rational investor should not be happy to hold a diversified portfolio because his ultimate goal is to reap the rewards.

Indeed, the reward per unit of risk is defined in terms of the net ratio. The high ratio of a portfolio is the average return of the portfolio minus the risk-free. This ratio is called the Sharpe ratio:

$$Sharpe\ ratio = \frac{r_p - r_f}{\sigma_p}$$

where: r_p is the portfolio return, r_f is the risk free rate and σ_p is the standard deviation of portfolio's excess return.

This study has enabled us to make some important observations:

- The Tunisian market is characterized by a yield less than 31% and a Sharpe ratio varying between 3.17 and 4.86. These extreme values characterize a very risky market whose data require a fine analysis.
- The execution of the CAPM model requires a high computation time despite the fact that the number of companies is limited to 85. If we add the fact that this execution requires the calculation of a covariance matrix and its inversion during a quadratic optimization problem, this calculation becomes very costly in the case of the NYSE stock exchange with an S&P 500 index containing more than 500 companies or in the case of decentralized finance for crypto-currencies.

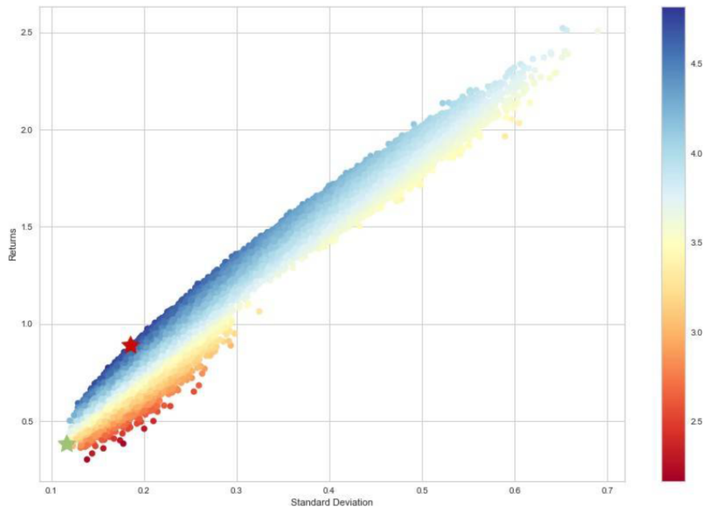

Figure 1: The Figure Represents the Efficient Frontier in the Plane Annual Return Vs Annual Volatility. The Red Star Represents the Maximum Sharp Ratio and the Green One Represents the Global Minimum Variance Portfolio During the Second Period

The Hierarchical Risk Parity (HRP) This diversification technique was first introduced in 2015 by Professor Marcos Lopez de Prado. He believes that machine learning models and algorithms can find patterns in financial data that only business experts are capable of finding. This algorithm solves 3 problems encountered in portfolios constructed based on Markowitz theory, which are instability, concentration, and underperformance.

Segmentation UMAP (Uniform Manifold Approximation and Projection) is a new dimension reduction technique based on manifold learning. The theoretical underpinning for UMAP is based on Riemannian geometry and algebraic topology. UMAP adopts the idea of tSNE in general, but adds a number of enhancements such as a second cost function and the lack of normalization of high- and low-dimensional probabilities.

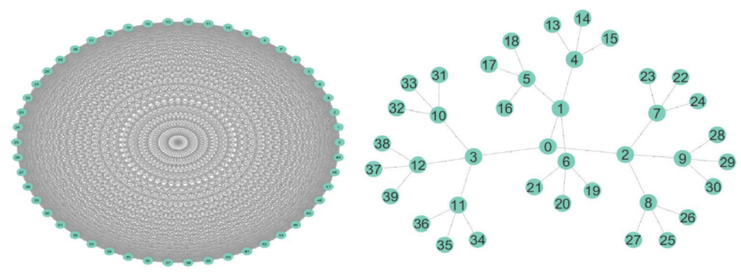

Figure 2: Intuitively it would be desirable to drop unnecessary edges. Adding a Hierarchical Structure can solve the numerically-ill conditioned covariance matrix.

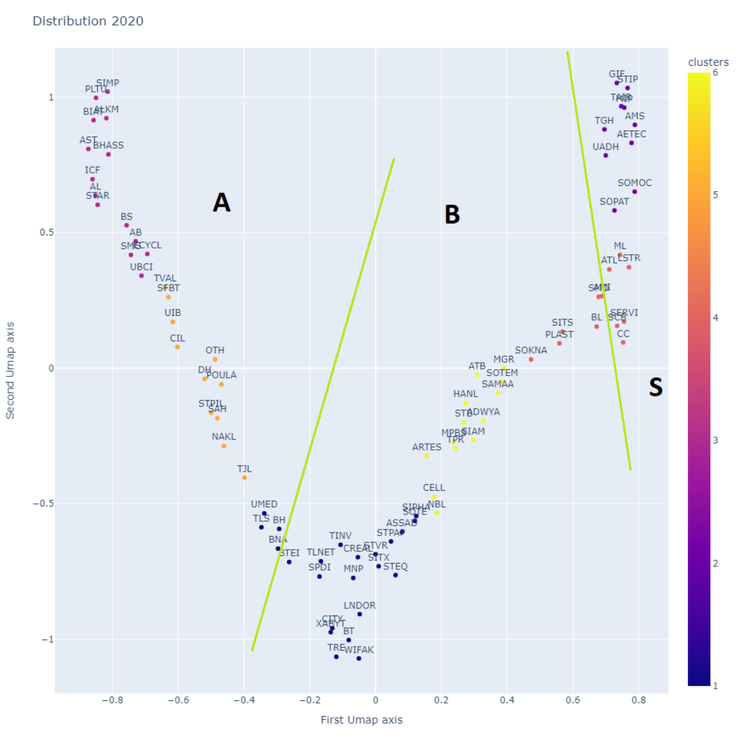

Figure 3: The figure shows the projection of the stocks on the plan obtained by the U-MAP algorithm with a hand drawn borders. It is clear that applying the U-MAP algorithm can reproduce the segmentation given by the BVMT.

4. Financial machine learning between two curses

Machine Learning gives a proven good approximation for high dimensional functions. ML learning processes is based on various types of data (tabular data, image data, sound data, text data...) and is leading to success in various fields. ML learns and generalizes patterns in high-dimensional and highly non-linear spaces without being specifically guided. This opens up new possibilities for attacking problems that suffer from the "curse of dimensionality". A second curse is added in finance which is the Markowitz's curse: Quadratic optimization is likely to fail precisely when there is a greater need for finding a diversified portfolio (De Miguel et al. [2009] show that many of the best known quadratic optimizers underperform the Naïve 1/N allocation, The more correlated the assets, the higher the condition number, and the more unstable is the inverse matrix).

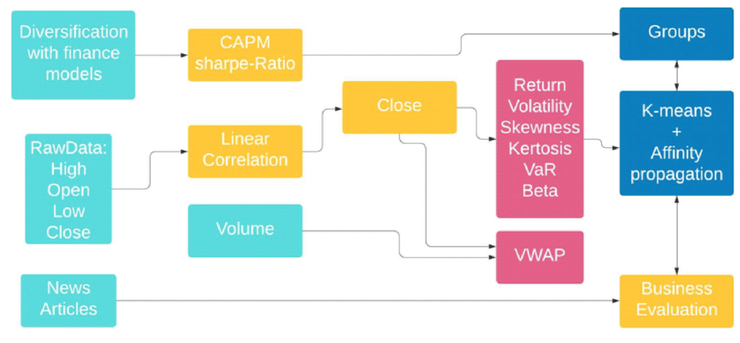

Figure 4: Data analysis pipelines.

5. Conclusions

Finance is concerned with the decision to spend available capital in an uncertain manner over a set time span in order to make a profit. As a result, such a decision could put the allocated funds at risk. As a result, there is a need for diversifiable risk diversification since there is another incompressible factor that contributes to systemic risk (non-diversifiable). Historically, the CAPM (Capital Asset Pricing Model) theory has provided a sound theoretical foundation for achieving such diversification and obtaining asset portfolios that minimize risk while maximizing return through quadratic optimization. Unfortunately, when it comes to risk management, a portfolio that maximizes the Sharpe ratio is not always the best option, especially during extremely rare 4 sigma events like financial or health crises. This was seen during the subprime mortgage crisis of 2008 and the current Covid-19 crisis. Indeed, in a parametered space, the problem of diversifying a portfolio of financial assets can be reformulated as a segmentation problem in the same space. With its unsupervised learning methods, Artificial Intelligence provides algorithms for detecting weak signals in large amounts of data. After working on the Tunisian stock market we were able to detect groups that are valuable to invest in and on the other hand very risky groups that their stock value is unpredictable. We first used isolation forest in order t obtain the companies that are really far from the average fluctuations and to confirm the first iteration we used the clustering techniques explained previously and the results were confirmed as the companies in question were either in the same group alone or each one of them was provided a center of a cluster and occupied the group by itself. All the results we got were confirmed by news articles specializing in the Tunisian stock market.