**Business Challenge Description :**

When considering different resources and comparing different points of views on the use of Machine Learning in Finance and more specifically, Machine Learning for active portfolio management and financial time series prediction, the opinions are very mixed. People like Yajiao Tang focus their ML models on predicting financial time series while others like Giuliano De Rossi consider a broader use of ML integrating portfolio optimization using risk measures for signal generation for instance.

The problem here is that ML models for portfolio management optimization are often very innovative and sophisticated which can also be a limit in real-world applications because of the interpretability difficulties that it creates for the population and especially the investors.

On the other hand, ML for FTS predictions are usually more understandable and thus more applicable because it takes into account the difficulty of FTS known for being short in length and noisy which leads to overfitting risks.

Thus, my goal here will be to focus on FTS predictions and will do so by implementing traditional prediction techniques mentioned above and my intent will be to use more and more sophisticated models during the project to see how it impacts the transparency and the risks of my models considering the complexity of FTS forecasting.

**Data Description and Sources :**

Here, I will start by focusing on the historical data of the S&P 500 index FTS and more specifically, my first idea is to take the data 1 year from today and assume that the daily close price is the mid price for the day which is a strong assumption.

Here, my purpose is to simplify as much as possible the analysis and thus the models, and making it more complex and trying different models over the course of time.

I will take the data from the Nasdaq website with the link below:

*https://www.nasdaq.com/market-activity/index/spx/historical?page=1&rows\_per\_page=25&timeline=y1*

**Scope of the Project :**

When considering a project involving FTS prediction, the complexity resides in the FTS themselves because they can be very disturbing with noisy data but the purpose here is to understand well their issues and limits in order to be able to expand the methods used in FTS forecasting and have a more integrated approach to be more relevant.

The problem here is that when trying to expand the models used in FTS prediction, many risks arise such as data requirements: some models like deep neural networks require a huge amount of data. As I mentioned before overfitting risks and interpretability also come whenever we try to implement more sophisticated models but also computational complexity can also become a real issue with processing power and time.

The possible and easiest statistical models that can be implemented could go from Moving Average, Exponential Smoothing and Auto-Regressive models to ARIMA and GARCH models.

The more difficult and advanced models used in Machine Learning could be in the spectrum of LSTM networks, SVMs, Random Forest, XGBoost or even Transformers models.

**Work Plan :**

Regarding the timetable, I will be working on this project 1 hour a day on Mondays, Tuesdays, Fridays, Sundays and 2 hours on Saturdays.

As I will be working on myself on this project, I will contribute to all the aspects of it but you can always check my progression on my GitHub : [*https://github.com/kyle75gto/ML-Project.git*](https://github.com/kyle75gto/ML-Project.git).