IVEY EMBA CAPSTONE EVENT - A DEEP DIVE INTO AI

Saturday, January 11, 2020

In an industry where almost every player has access to the same core financial and fundamental data, investment firms look for data science and alternative data in order to find differentiated insights into companies not found in filings, earnings calls or fundamental datasets. The data used in earnings forecasts today could include satellite imagery, real-time spending information, meta-data details of every product on every store, and much more besides. This dramatic increase in the availability of alternative data will profoundly change the investment landscape. As more investors adopt a more data driven investment style, the market will start reacting faster and will increasingly anticipate traditional data sources (e.g. quarterly corporate earnings, low frequency macroeconomic data, etc.). This gives an edge to both fundamental and quantitative managers willing to learn about new datasets and methods. Below I discuss how we went about building the effort (e.g. what talent is required to build a data science effort in-house, robust technology setups, and collaboration) to adopt a more quantitative data driven investment style.

Talent

As a logical individual who has spent the past decade working with quantitative investing techniques and risk premia investing, I gained the right mix of soft skills (such as ability to teach data science) and hard skills (such as the ability to design a systematic strategy which produces robust weights) that allowed us to fully internalize and integrate data science and alternative data. Securing talent in quantity and quality was the key to research workflow innovation, and a competitive advantage. To encourage a good flow of talent over the long term, I created education platform across all teams: data engineering, quantamental research and portfolio construction & execution. This ranges from sharing internal materials that stimulate interest in programming, data engineering, statistical learning and computational finance. Luckily, I taught two data science courses[1] for nine years at the University of Western Ontario with a focus on fundamental and quantitative investment styles, consumption of increasing amounts and differentiated types of data, and adoption of new methods of analysis such as those based on machine learning.

---

[1] http://www.rnfc.org/courses/

IVEY EMBA CAPSTONE EVENT - A DEEP DIVE INTO AI

Technology

Any of our investment strategy relies heavily on the availability and quality of data. By working with major data vendors for the past decade, we were able to collect and clean a lot of data, including S&P Global and Bloomberg Professional Services, which was subsequently used for building our data engineering capabilities for storing and pre-processing data. Typical data sources for our research are regularly categorized as "Big Data", which means that they may be very large and complex and often cannot be handled by programming tools traditionally used for storing or handling data (such as the popular R data.table and Python pandas). Our proprietary database solution can handle the distributed nature of storing and analyzing data. It is designed for real time generation of analytical data reports using SQL queries and is blazing fast. What does "blazing fast" mean? It means that our researchers can iterate over billions of rows under a second! Empowering and accelerating the work of our data scientists through more efficient and scalable technology is my ardent interest, and never-ending goal. In fact, alternative datasets are often multi-dimensional (e.g. might be a mix of different formats such as images, text, numbers) and require robust pre-processing solutions. At OneSixtyTwo capturing, storing, calculating and sharing features is coded exclusively in distributed C++ and supported by a data engineering team with educational backgrounds in computer science, statistics, and engineering.
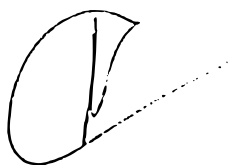
Collaboration

We also learned that alone, we cannot accomplish anything! We do collaborate with our investors and partners who have deep industry knowledge. In 2016 we joined the NVIDIA Inception Program and were able to setup our own high-performance computing solution to handle the distributed nature of storing and analyzing data at scale. Many popular data science tools have traditionally been slow and cumbersome, relying on CPUs to load, filter and manipulate data and train and deploy models. Graphics processing units (GPUs) are commonly used to accelerate computational these tasks. We optimized our technology infrastructure to run a single instruction over multiple segments of data, which allowed our researchers to **easily** dispatch efficient portfolios computations and back test multiple ideas. In addition, by working with Goldman Sachs we were able to design, code and optimize our portfolio construction and execution algorithms,

which would involve an additional layer of system and software engineers linking signals to our investment strategies.

These past three years, we have made good progress in building a multi-disciplinary team of data scientists and engineers obsessed with data and code, searching for very slight probabilistic edges, and eventually building systematic strategies to exploit those edges[2]. We view our data technology stack as enterprise software that can be easily customized in a cost-effective and scalable manner. There is much more to do, and we will continue to work hard on this while building lasting relationships with our stakeholders and partners.

Sincerely,

Rafael Nicolas Fermin Cota (Nico)
Co-Founder & CEO
OneSixtyTwo Group

---

[2] https://www.linkedin.com/pulse/2019-year-end-review-rafael-nicolas-fermin-cota/