

Augvest Event Recap: Alt Data & ML/AI for Allocators

Through our network we have seen that while there has been a great deal of buzz around how data science and Machine Learning / Artificial Intelligence can be used in investing, only a few dozen firms can be considered to be truly competitive in this space.

In spite of the difficulty of deriving edge from data, more and more funds have been marketing their data science capabilities. These firms have a growing number of conferences and vendors to take talking points from - but there is no similar ecosystem for allocators.

To bridge that knowledge gap, on January 17th 2019 we co-hosted a panel discussion to explore how allocators can best assess funds' data science capabilities. The panel featured:

- David Loaiza - formerly Chief Data Scientist at Point72
- Matei Zatreanu - formerly Head of Data at King Street
- Brian Lichtenberger - CEO & Co-Founder at 7Park

General background:

- The use of alternative data has in the past mostly been limited to consumer names but we have seen interesting datasets and use-cases emerge in healthcare and industrials
- Data sourcing is becoming increasingly commoditised and so edge is shifting from away from data discovery and towards analytics
- A lot of market participants are singularly focused on using data to track company revenues but there is a limit to how much alpha you can generate trading quarterly announcement surprises - especially when your competitors are likely buying the same data
- The next wave of development in alt data will come from learning to look at more granular performance indicators (e.g. customer retention) and thereby developing or testing long-term investment theses.
- Doing this requires close collaboration between traditional investors and data science, which has proven difficult to implement due to cultural differences as well as conflicts over P&L attribution.
- Deeper analysis is also essential to making use of data at credit funds and in private equity.
- Private equity also has the benefit of being able to take ownership of data, instead of just getting an annual license
- While alternative data tends to be too expensive for small funds, even start-up GPs can get involved in this space by entering into partnerships with vendors where they get access in exchange for helping develop the vendors' products.

A few key red flags to notice when GPs are pitching their data science capabilities:

- **Claiming to be using AI on alt data**

The whole premise of alternative data is that it allows you to directly track the fundamental performance of a business, be it by estimating revenues or more granular KPIs. Even for publicly traded companies those metrics are only published quarterly, meaning that there are four data points per year per firm for models to be trained. That's simply not enough for AI to be relevant.

There are valid uses for AI in data processing - e.g. in tagging transaction record data or analysing imagery - but those tend to be quite niche.

- **Lack of adequate resources**

If a fund claims to derive edge from data but does not include engineers in their hiring plans then they are unlikely to be competitive.

Funds attempting to integrate data science into traditional investing should also be able to talk about the investments they have made into bridging those two areas. Just hiring a few data scientists is not enough; lack of alignment between data scientists and investment analysts is the single most common reason for the failure of alt data efforts.

- **Lightning speeds**

It is not unusual for investment firms to apply the 'data science' label to work they have been doing for many years. If you see a fund claiming to have developed a data science team of dozens of people in a few months, chances are they simply re-branded an existing team.

The successful efforts we have seen grow to meaningful scale have taken years to get there. A few large funds tried to speed things up by investing heavily into data and the end result was rapid churn within their data teams.

Allocators looking to make use of data science themselves should first and foremost look to it as a means of automating recurring internal processes. Doing so will allow you to start gathering your data in a way that would then make it possible to start applying machine learning to streamline and otherwise optimise your operations.