Contact
Andreas Posavac
Executive Director
ESG, M&A and Governance Advisory
andreas.posavac@ihsmarkit.com
Telephone +43 (0)664 444 11 30

Author
Christine Chow, PhD
Executive Director
**Strategic Governance Advisory and
ESG Integration**
christine.chow1@ihsmarkit.com

How Data and AI are Changing Investment Analysis

Key Takeaways

Investment professionals use principal component analysis (PCA) to identify key risk factors that impact portfolio performance. This is one of the traditional statistical methods used, along with other techniques, such as multi-factor risk analysis or a hybrid combination of them. The toolbox for statistical analysis has expanded significantly over the last 30 years. Now, we have a catchier option that uses machine learning. We often call it Artificial Intelligence (AI).

For AI to be useful in investment analysis, we need three key elements:

- **Why do we use AI? Know your problem -** Users must understand what problem they are trying to solve and set a clear objective for each application. This is because AI at its core is an optimisation exercise.
- **What do we have and what is missing? Know your data –** Define what data is relevant to meeting the objective. Find out what data is available and what needs to be created or compiled. If it involves unstructured data, such as news feeds or social media posts and videos, there must be a clear methodology that turns unstructured data into usable formats. Some tasks such as peer benchmarking require comparable data. This is what investors are crying out for now. *This article highlights the current debates on common reporting standards with an example of how industry collaboration can make a difference.*
- **How to do it? Know your algorithm** – Once quality and relevant data become available, suitable AI techniques can be applied in portfolio management. Objectives could be maximising key performance metrics like Sharpe ratio or Information ratio, or minimising risk metrics, like Value-at-Risk (VaR), standard deviation or tracking error. AI can be used to construct enhanced exchange treaded funds (ETFs) with fewer holdings that mimic the performance characteristics of a full index. It can also be used for back-testing and scenario analysis. The details of such applications may warrant a separate article on "AI in portfolio management".

Introduction

Not all investment analysis can make use of AI, but when it does, we must first understand what problem we are trying to solve and set a clear objective of application.

For example, "growing the business" is not a well-defined objective, but "optimising sales" is. "Assessing board effectiveness" is not a well-defined objective, but "minimising interlocking board directorship" is. This is because fundamentally, AI is about maximising an objective function using one or a combination of four common techniques as described in the book ***Mathematics for Machine Learning*** by Deisenroth, Faisal and Ong (2020)[1] (Figure 1):

Figure 1: Key Optimisation Techniques in AI

1) **Regression** is fitting a curve to data. It refers to a process of estimating the relationships between a dependent variable (often called the 'outcome variable', 'y') and one or more independent variables (often called 'predictors', 'covariates', or 'features', $x_1 \dots x_n$). For example, in attribution analysis, a multi-factor risk model uses regression.
2) **Dimensionality reduction** is reducing the number of variables dealing with whilst retaining meaningful properties of the original data. For example, in image recognition, it is used to improve clarity of images for machine learning.
3) **Density estimation** is the identification of areas that data clusters together. For example, it is used in unsupervised learning to group data and find concentration with specific characteristics.
4) **Classification** is sorting data into categories. For example, in sentimental analysis, keywords are assigned positive, negative or neutral sentiment.

AI can find an investment fund most likely to be purchased by a customer based on propensity scores that maximise the chance of purchase by presenting relevant offers at the right time and at the right price. This combines behavioural science and customer segmentation. Even when personal data is anonymised, each of us is likely to have been categorised as a particular customer type by each business we interact with that uses AI for targeting. The above has already been put into practice in other industries for many years by companies such as Amazon or Netflix when they recommend the next purchase items or next favourite movie based on our personal profiles, past purchases and viewing history.

Towards a Common Reporting Standard

Making effective prediction require quality data for AI analysis. Data is where the biggest challenge lies for the investment industry now. Pressing issues raised by keynote speakers[2] at the bi-annual Climate and Green Finance Regulatory Forum[3] pointed to a '**misguided reality**' where company information, especially sustainability

---

[1] Deisenroth, Faisal and Ong (2020) *Mathematics for Machine Learning* Cambridge University Press April 2020.
[2] Keynotes by Ashley Alder, Chair of the Board of IOSCO; Richard Monks, Director of Strategy at the UK regulator, the FCA; and Eric Usher, Head of UNEP FI.
[3] New tab (cityandfinancialconferences.com)

information, tends to be widely inconsistent. For financial information, the International Financial Reporting Standards (IFRS) Foundation and the International Accounting Standards Board (IASB) have already established a standardised way of describing a company's financial performance and position so that company financial statements are understandable and comparable across international boundaries. For sustainability information however, companies could cherry pick reporting standards and ratings, leading to concerns over greenwashing and/or SDG[4] washing.

The investment industry has called for an alignment of data definitions and reporting standards. To this end, the World Economic Forum (WEF) and Impact Management Project (IMP) have been working with practitioners towards a comprehensive reporting system. It published a conceptual framework that brings the key reporting standards together, including how they relate to each other, using enterprise value and climate-related financial disclosure as examples (Figure 2) [5].

Figure 2: Comprehensive corporate reporting
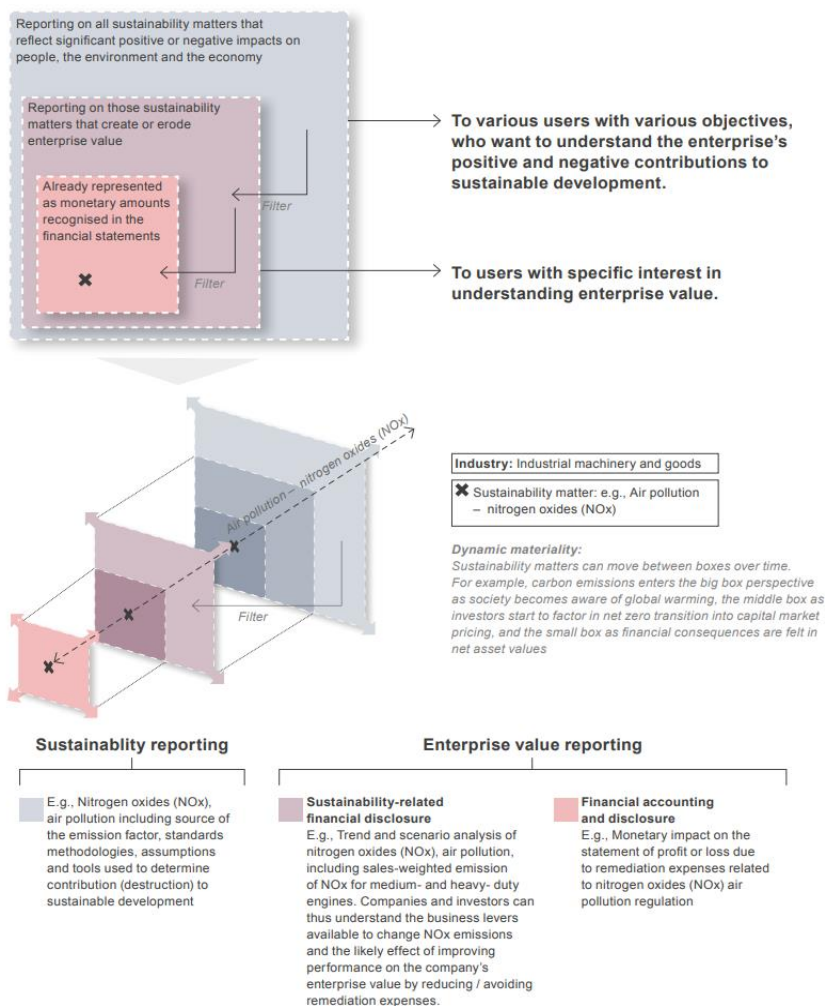
Challenges in Improving the Quality of Company Data

Having an aligned reporting system is work in progress. However, it addresses only some but not all challenges.

Deep rooted issues that impact the credibility of reporting lies in the quality of **company level data**. This could be related to constraints in legacy IT systems; inconsistency in policy applications across global operations; and difficulties in categorising certain business activities into definitive and exclusive buckets. Examples can be found in reports such as ***European Sustainable Finance Survey[6]*** by Adelphi and ISS ESG and ***Testing the Application of the EU taxonomy to Core Banking Products***[7] by the European Banking Foundation (EBF) and UNEP FI.

There are operational challenges in tagging data for specific purposes. For example, banks corporate lending systems tend to be set up by sector and industry to suit relationship managers team structure and expertise. The systems do not necessarily separate use of proceeds for general purpose loans or revolving credit. These issues are recognised in the EBF and UNEP FI report as some of the most significant

challenges (See page 32 of the report and forestry case on page 36 and 39). These challenges are amplified when loans are approved at group level and funds are disbursed to multiple, local legal entities.

Unstructured Data from Machine reading and Satellite Imagery

Besides reporting and data systems issues described above, investors are bombarded by more unstructured data. According to Wikipedia, unstructured data is information that either does not have a pre-defined data model or is not organised in a pre-defined manner. It is typically text-heavy, but may contain data such as dates, numbers, and images as well.

According to the NBER paper ***How to talk when a machine is listening: corporate disclose in the Age of AI***[8], annual reports are increasingly read by machines for content extraction and sentiment analysis. It uses a common AI technique called natural language processing (NLP) or natural language understanding (NLU). In the US, the number of machine downloads from the US SEC EDGAR database as a fraction of all downloads increased from 39% in 2003 to 78% in 2016. It is likely that this number had significantly increased by 2021, given the explosion of AI-powered web scraping activities in recent years.

Data and analytics firms such as IHS Markit has developed an **ESG reporting repository** powered by its AI engine Goldfire[9] to organise unstructured data, help companies to improve reporting and build sustainability profiles[10]. Companies can upload their own reports and user-defined materials. The platform uses NLP to align relevant information against reporting standards, such as GRI (Global Reporting Initiative), SASB (Sustainability Accounting Standards Board), TCFD (Taskforce for Climate-related Financial Disclosure) and UN SDG. Its DigitalGlobe Satellite

---

[6] European Sustainable Finance Survey 2020

[77] Testing-the-application-of-the-EU-Taxonomy-to-core-banking-products-EBF-UNEPFI-report-January-2021.pdf

[8] How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI | NBER

[9] (1) New Message! (ihsmarkit.com)

[10] News Release (theesgreporting.com)

Imagery service provides geospatial data to track land uses, detect deforestation and monitor energy pipeline conditions and transmission[11].

Investors are increasingly using geospatial data to monitor business performance of investee companies real time or near real time, and to improve supply chain transparency. Insurance companies use geospatial data for underwriting policies and pricing as well as claims settlement. Ping An Insurance has established a 'greenwashing' indicator[12] using NLP techniques to understand the breadth and depth of companies' climate disclosure, and link them to firm characteristics and financial performance. TruevalesLab has created tools and products to help investors discover hidden insights[13]. The sophistication in using data and AI for investment analysis is only set to increase. Using anchor or official information sources, running weekly averages to screen out news spikes are some of the many ways to combat misinformation risk.

Deciding and Creating Useful Data requires Industry Level Collaboration

In the era of data explosion, deciding what matters and how to use the information is of vital importance. How should investors tackle this? Taking an industry-based approach seems most sensible.

The Poseidon Principles launched in New York on 18 June 2019, is an agreement between the finance sector and the shipping industry to integrate the policies of the International Maritime Organisation (IMO) on climate change into shipping finance decision making processes[14]. Companies must follow the Technical Guidance for all business activities, down to the coding system of vessel types, the calculation of specific carbon intensity metric known as the Annual Efficiency Ratio (AER) for the shipping industry, and the choice of three permissible methods of calculating decarbonisation pathways. Signatories, mainly shipping finance entities within banks, rely exclusively on data types, data sources and service providers identified by the Technical Guidance to conduct assessment that determines lending decisions. The '**value chain'** of data and analysis are therefore aligned between companies and financiers.

Climate change is a global issue. It is reassuring that signatories[15] of Poseidon Principles include banks from developed and emerging economies; from Europe and Asia. Having an industry aligned approach allows investors to compare apples with apples, making informed investment decision possible.

---

[11] [DigitalGlobe Satellite Imagery | IHS Markit](#)
[12] [https://group.pingan.com/dam/jcr:3a9a1369-5474-4034-85fd-ffe4ab4c3b31/Climate Disclosures and Financial Performance.pdf](#)
[13] [ESG Data | AI Powered ESG Insights & Analytics | Truvalue Labs](#)
[14] https://www.poseidonprinciples.org/wp-content/uploads/2019/07/Poseidon_Principles.pdf
[15] [Signatories - Poseidon Principles](#)

Conclusion

Quality data is the lifeblood of effective AI applications in investment analysis simply because of the well-documented phenomenon of 'garbage in garbage out'. Once quality and relevant data become available, suitable AI techniques can be applied in portfolio management. Objectives could be maximising key performance metrics like Sharpe ratio or Information ratio, or minimising risk metrics, like Value-at-Risk (VaR), standard deviation or tracking error. AI can be used to construct enhanced exchange treaded funds (ETFs) with fewer holdings that mimic

Key contacts

Andreas Posavac, MBA
Executive Director - Global Head of ESG, M&A and Corporate Governance Advisory

Andreas runs a global team that supports companies and their C-suite and Investor Relations teams as well as their banking and legal advisers with market intelligence, risk analytics and advisory services focused on institutional investors in the lead up to M&A transactions, general meetings or special situations. His team also works with companies to develop a coherent ESG and engagement strategy to follow legal and industry best practice standards and take advantage of the opportunities in sustainable investments and increased ESG integration. Andreas has deep knowledge of the ESG and corporate governance sensitivities of institutional investors, their relationships to external advisors as proxy advisory and ESG rating firms and can help stakeholders better understand how to address governance and ESG in order to have the greatest impact in minimizing risks and maximizing opportunities in the current capital markets landscape

Christine Chow, Ph.D.
Executive Director - Global Head of Strategic Governance Advisory & ESG Integration

Before joining IHS Markit, Christine was Head of Federated Hermes EOS in Asia and global emerging markets. Her PhD thesis on responsible investment was short-listed for a United Nations award in Sweden for industry relevance and academic excellence. She is a Member of Court and Investment Committee of the London School of Economics (LSE), a Board Member of the International Corporate Governance Network (ICGN) and appointed an honorary adviser to the Financial Reporting Council (FRC) Hong Kong. She was named as one of the top 30 Inspirational Women in the City of London. In 2020, she won the Finance Monthly Women in Finance Award as the Investment Management Leader of the Year (Asia). Christine is a graduate of the London School of Economics and the University of Melbourne. She completed an executive education course on financial engineering at Stanford University. She was a member in the Data Governance Task Force of the UK All Party Parliamentary Group (APPG) on Artificial Intelligence (2018 – 2021) and an Adjunct Professor in Finance at the Hong Kong University of Science and Technology (2014-2016).

About IHS Markit

IHS Markit (NYSE: INFO) is a world leader in critical information, analytics and expertise to forge solutions for the major industries and markets that drive economies worldwide. The company delivers next-generation information, analytics and solutions to customers in business, finance and government, improving their operational efficiency and providing deep insights that lead to well-informed, confident decisions. IHS Markit has more than 50,000 business and government customers, including 80 percent of the Fortune Global 500 and the world's leading financial institutions. Headquartered in London, IHS Markit is committed to sustainable, profitable growth.

IHS Markit Customer Care

CustomerCare@ihsmarkit.com
Japan +81 3 6262 1887
APAC   +604 291 3600
EMEA   +44(0) 134 432 8300
AMER   +1 800 447-2273

IHS Markit®