The Rise of Machines in ESG Investing

In a report entitled 'Asset & Wealth Management Revolution: Embracing Exponential Change', PwC predicts rapid growth for the asset & wealth management industry and anticipates that global assets under management will grow from approximately US$111 trillion in 2020 to about US$145 trillion by 2025. The report also warns that firms needs to take action now, if they are to survive an exponential level of change. And one of those big changes is the adoption of Artificial Intelligence and Machine Learning whose use is expected to rise dramatically in the investment industry.

The other growth trend will be the adoption of ESG. According to a report by Deloitte, ESG mandated assets will make up more than half of all professionally managed assets globally by 2024.

Authors

Sumali Sanyal, CFA
Managing Director, Senior Portfolio Manager, Systematic Global Equities

Cameron McLennan,
Director, Senior Portfolio Manager, Systematic Global Equities

Chart 1

ESG-mandated Assets Are Projected To Make Up Half Of All Professionally Managed Assets Globally By 2024
Global assets under professional management ($T)

Year	ESG-mandated	Non-ESG mandated
2014	$19	$45
2016	$19	$50
2018	$27	$53
2020	$39	$69
2021E	$46	$71
2022F	$55	$72
2023F	$67	$72
2024F	$80	$71
2025F	$96	$69

Note: All amounts are in US dollars.
Source: Proportion of ESG-mandated data through 2020 from Global Sustainable Investment Alliance; DCFS analysis through 2025. Deloitte Insights | deloitte.com/insights

■ ESG-mandated ■ Non-ESG mandated

This post explores the reasons behind the growing popularity of Artificial Intelligence (AI) and Machine Learning (ML), the opportunities that are available in using these techniques for ESG investing and the challenges encountered in using them.

The Popularity of Machine Learning

Machine learning is an approach to develop data-driven computer applications that train themselves to complete complex tasks without explicit instruction. AI and ML concepts and methods are not new and attempts to apply them to investing date back to the 1990s, when numerous academic articles were published on the use of neural networks in forecasting financial market events. Today, the intense focus on ML reflects the confluence of several trends. First, many ML methods are computationally intensive, and computing costs have dropped dramatically. Commoditization of cloud computing has made

thousands of processors available on-demand. Second, ML algorithms are now freely available in robust and well-documented open-source packages, dramatically lowering cost, time, and knowledge barriers to implementation. Third, the explosion of big data facilitates ML approaches that involve algorithms that require enormous quantities of data to train. Finally, there have been algorithmic breakthroughs in language processing and in the development of better behaved and more computationally efficient neural networks.

In the context of quant investing, interest in ML reflects the desire for greater modeling flexibility in two broad respects. The first is non-linear prediction. ML algorithms are designed to infer the relationship between data attributes and a variable of interest. In contrast, linear regressions impose an assumption that the form of the relationship is a straight line. Although linear models tend to be simple, transparent, robust, and have modest data requirements, in many investing contexts the assumption of linearity is unfounded. With that in mind, investment managers are now using ML algorithms such as Generalized Additive Models (GAM), Decision Tree Models, and Deep Learning to capture non-linear relationships between factors to improve the predictive ability of their models.

Chart 2 | **Linear Regression vs. GAM: Fitting a Straight Line vs. a More Flexible Function**
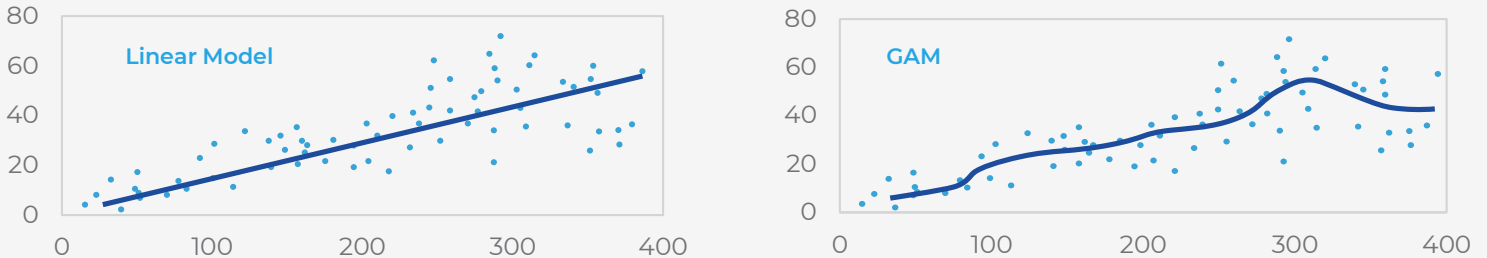
Source: Interpretable Machine Learning, Christoph Molnar

A second motivation for ML is to reveal hidden structure in complex, large data sets. For example, in a company's peer group, ML algorithms may help to illuminate the economic relationships among hundreds of firms, simultaneously accounting for numerous attributes. ML-based textual analysis can derive quantitative metrics from qualitative information - sentiment analysis being a popular example. For quantitative investment managers, it opens new avenues of research that have exciting potential. However, their successful exploitation will demand considerable effort and firm commitment to a disciplined research process.

Use of Machine Learning in ESG

Investment managers are coming under increasing pressure to measure ESG criteria in their portfolios. There is much less structure in ESG data, with some fields purely based on the data mining of unstructured text documents or records. This increases subjectivity and the intuition required by data researchers to convert the ESG factors into quantifiable metrics. There can be varying amounts of noise introduced into the measurement of factors owing to parameters used in the methodology of mapping text or records to a quantified metric. Not surprisingly, many factors have short histories, as providers have only begun sourcing or recording underlying data. Data can be missing for some fields or not updated, particularly for specific time periods. Noise, short histories, and gaps pose challenges in the use of ESG data. AI & ML can help in the integration of ESG data into more stable, comprehensive databases. Advances in natural language processing (NLP), deep learning, and downstream machine learning ensemble techniques can make it possible to integrate similar fields from different datasets, reducing noise while retaining most of the information and value.

Chart 3 | **The Convergence of Three Megatrends**
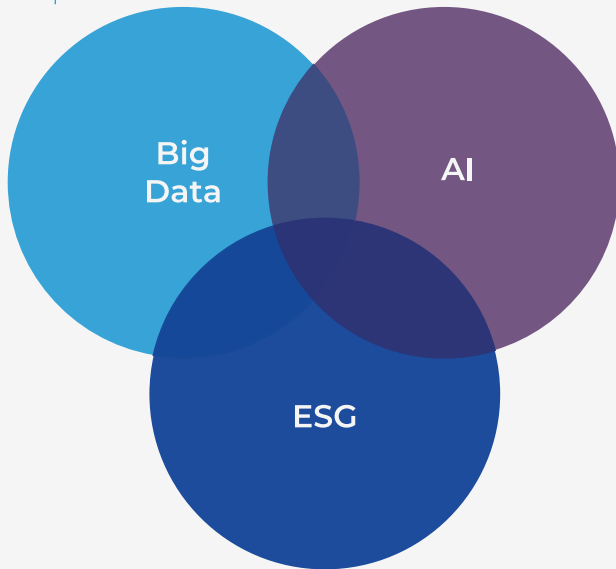
- ESG investing strategies are in demand
- Investors are focused on long-term value creation and sustainability
- "New world" unstructured data is rapidly growing
- NLP and artificial intelligence technologies are advancing

Source: TrueValue Labs

Data augmentation can increase the diversity of data in training models without collecting new data. It can fill in gaps in time or cross sections, thereby taking another step toward standardization. Tensor-completion techniques currently at the forefront of machine learning show promise to address the problem of ESG data gaps. Interpolation can fill missing data in time series, and tensor-completion techniques can be used to extrapolate fields with short histories based on related fields with longer histories. Tensor completion goes above and beyond traditional interpolation techniques by combining cross-sectional and temporal information in a way that carries over many characteristics of the original data to fill in the missing data.

There are multiple challenges in reading reports to understand how an organization pursues and achieves ESG activities. First, the types of activities and impacts are not described by any standard taxonomy. The use of language changes with the times, and in different sectors and industries, different language is used to describe similar activities. Second, as with any human activity, issues of repeatability and reproducibility need to be considered to maintain consistent results across organizations and topics over time. Lastly, it is challenging to clearly differentiate what an organization aspires to do versus what they achieve. Technology — in particular, artificial intelligence technology — can be used to address these challenges.

Much of the potential for artificial intelligence in ESG investing comes from sentiment analysis algorithms. Sentiment analysis programs can be trained to read a certain type of conversation and analyze the tone by comparing the words used to a reference set of existing information. For example, a program trained to read the transcripts of a company's quarterly earnings calls could determine the tone of the words when the CEO speaks, use natural language processing to easily identify in which parts of the conversation the CEO talks about ESG-related topics, and then infer from those words how committed a company appears to be about mitigating ESG risks. Machine-learning algorithms can identify patterns from the data they receive and learn from their own results. When a company releases an update, these algorithms enable analysis of the reaction from its stakeholders and the public. Investors are using machine learning to mine all kinds of information, from the minutiae of earnings disclosures to the content of LinkedIn posts. Using this sentiment data, they can sift through the hype around environmental, social and governance (ESG) issues and get an accurate picture of a company's credentials.
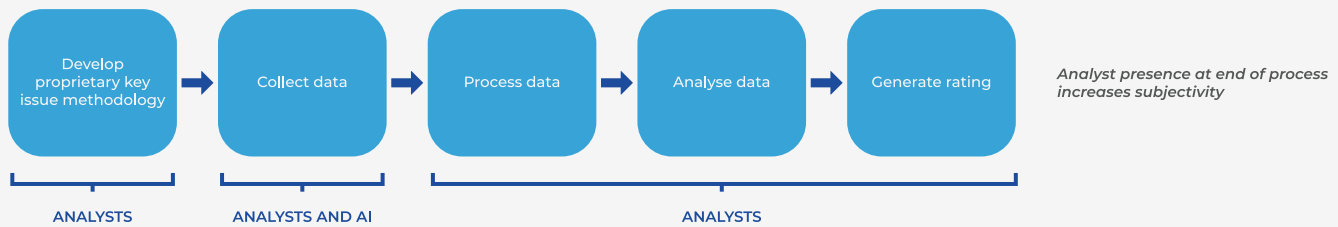
Using NLP, researchers can go beyond traditional market reports to analyze both written and verbal communication to understand how ESG commitments are both presented and received. Analysts mostly use NLP to analyze the language companies themselves use; whether their declarations are concrete or vague, whether they use the first person or take refuge in the third. By using machine learning applied to the news, an investment manager can effectively highlight the exact ESG actions a company is taking to promote positive impact. One focus area is how the presenters handle impromptu questions from analysts, as the answers are often much less positive than the pre-prepared statement at the start. This can give a truer picture and can help in aggregating the sentiment across markets and regions. Other firms focus on how external parties react to these statements. They comb thousands of news stories to get instant reactions and often combine this with comments scraped from social media. By understanding what people are saying about these companies, a true picture of their perceived brand value can be obtained. ML models give investors the ability to detect whether managers are 'greenwashing' when they talk about their firm's ESG policy.

Currently each ESG data provider has its own rigorous set of metrics, but with no universal standard for what constitutes "good ESG", their methodology, and thus their ratings, differ markedly. Some providers score companies on an absolute basis, so everyone is judged by the same criteria, but others score relatively, which can reward the least bad companies in less progressive industries. The data these providers rely on is also heavily influenced by periodic corporate disclosures. This data isn't just prone to bias, as companies omit the factors that paint them in a bad light, it is also backward looking.
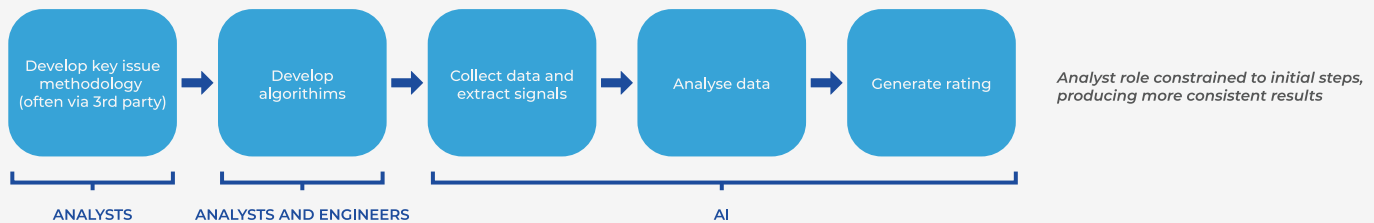
Traditional ESG ratings are composed of only three actors: rated corporations, rating agencies and end-users. The few actors involved and the opacity of the dialogue between them risks encouraging biases in the ratings. For example, higher scores have been found to have been assigned to larger companies who have more resources to fill out the questionnaires that ratings providers send them. Conflicts of interest may also exist whereby higher ratings are sometimes given to the holdings of asset managers who are also heavily invested in the ratings providers themselves.

Chart 4 | **The Contrasting Role of Human Research Analysts In the Processes of Traditional (top) and Alternative (bottom) ESG Ratings Providers.**

Analyst-driven ESG research: derives ratings in a structured data model

| Develop proprietary key issue methodology | Collect data | Process data | Analyse data | Generate rating |

Analyst presence at end of process increases subjectivity

ANALYSTS | ANALYSTS AND AI | ANALYSTS

AI-driven ESG research: derives signals from unstructured data

| Develop key issue methodology (often via 3rd party) | Develop algorithims | Collect data and extract signals | Analyse data | Generate rating |

Analyst role constrained to initial steps, producing more consistent results

ANALYSTS | ANALYSTS AND ENGINEERS | AI

Source: Author, adapted from information from TVL.

Technology, however, potentially offers a solution to these problems by limiting the subjectivity and cognitive bias that often stems from human-led analysis. These innovations have begun to be applied to ESG sustainability insights in the form of "Alternative Data" sources that supplement the core financial information, using data which has been scraped from the Internet. In this way, AI is arguably being utilized for social good. Alternative ESG ratings based on AI provide a more objective, outside-in perspective of a companies' sustainability performance. They use natural language processing (NLP) to synthesize vast amounts of unstructured data from online media and the Internet to extract the public sentiment on a company through automatic summarization, relationship extraction and sentiment analysis that effectively judges what the world thinks about the company. Machine learning is being applied to sustainable investment to give structure to unstructured datasets.

Fundamentally, these AI techniques have started to redistribute the control of sustainability information away from just a handful of powerful actors in financial centers. Firstly, there are many more stakeholders than just the corporations themselves feeding into the discussion on relevant ESG issues. The Internet acts as a quasi-objective collection of third-party public information on companies, coming from NGOs, national and international media sources, academic journals, trade blogs. Secondly, sustainability reporting standards initiatives such as SASB play an important role in defining the relevant ESG issues in these networks. This is in contrast to the traditional model, where the ratings agencies are responsible for identifying the key issues themselves. Instead of the 'inside-out' perspective that this creates, Alternative ratings follow an 'outside-in' perspective, which is based on a more democratic system that creates ratings using no company disclosure and analyzing only publicly available data sources based on public perception. Alternative ratings methods also have safeguards in place to deal with fake news and use updated watch lists that help to avoid unreliable sources, particularly on social media.

Company disclosure standards also vary dramatically across different markets, with weak regulation in developing regions. Although initiatives like the TCFD, SASB and the Global Reporting Initiative are helping to standardize company disclosures relating to ESG matters, at the moment globally recognized standards are not as developed as in financial accounting. Consequently, the outside-in perspective of Alternative ratings is a refreshing development in the proliferation of unbiased and transparent ESG ratings.

Chart 5 | **Company Disclosed Information Is Only Part of the Story**
To truly evaluate a company and eliminate ESG "blind spots" requires a comprehensive approach.

Company Disclosed Information

External Stakeholder Information
News, NGOs, Watchdogs, Trade Journals, Indepedent Reports
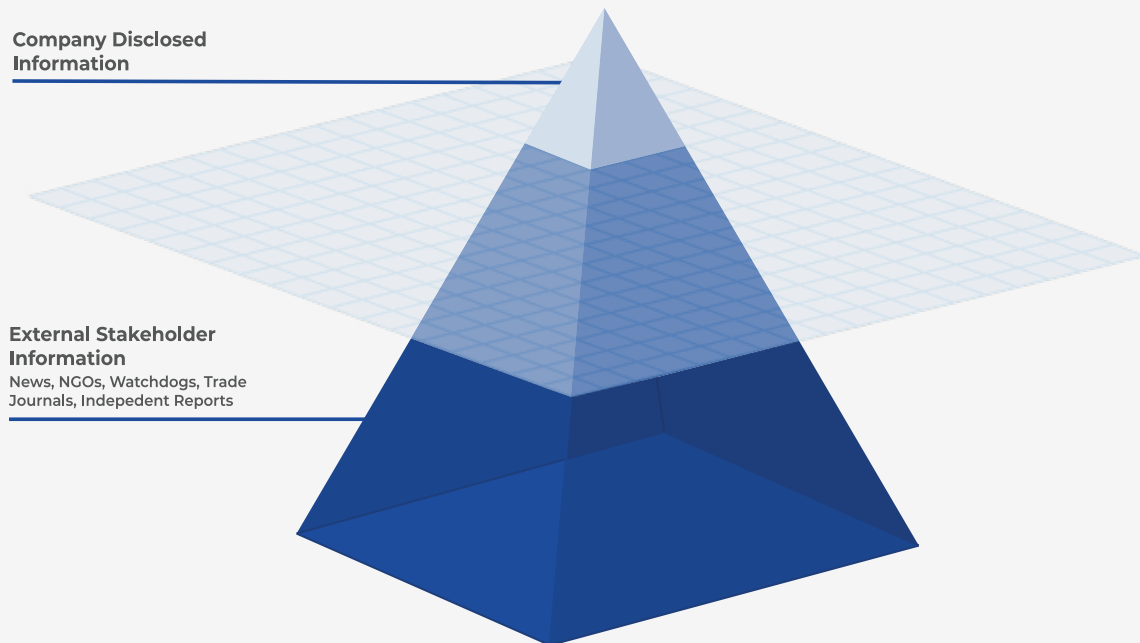
Source: TrueValue Labs

For Alternative ratings, the weight setting process is very different. The weightings are determined by the impact of the issue which is measured by the datapoints analyzed by the algorithms, on that key issue. Weightings also change in real time instead of per year, considering smaller ESG events, rather than just shifting every so often in light of major events. The kind of data sources that AI/ML monitors on the Internet are based primarily on the sustainability controversies that generate the most 'noise' by commentators, which fluctuate based on what seem to be important during the time period. A key difference between the spaces of AI and Traditional ESG ratings is that while traditional ESG matters are defined by internal debate and corporate decision making within ratings agencies, AI allows sustainability to become more of a 'public sentiment' issue. This increases public participation and a greater element of democracy in the ratings process.

AI allows the incorporation of controversies more frequently into updating scores, offering users more detailed, up-to-date insights. AI/ML is more effective at incorporating big data in real-time. Human analysts simply cannot keep pace with the data coverage capabilities of AI. AI/ML technology used to carry out analysis can also be scaled up. They use multipipeline architecture that can incorporate different frameworks and increase the number of languages analyzed.

A shift away from human-based analysis allows ESG information to flow more quickly through space, enabling the immediate spread of knowledge from companies and markets across the world that are more physically isolated. In asset management, real-time analytics is crucial, giving firms the ability to reallocate resources dynamically in response to unforeseen events and delays can have serious consequences for portfolio values. Additionally, AI allows the analysis of controversies with more objectivity, relying just on patterns in the data rather than on human biases and subjectivities.

Challenges Associated with Machine Learning and Artificial Intelligence

While giving ESG investing the opportunity to grow and expand, AI can itself be an ESG risk for companies that aim to undertake the effort. Adopting AI for any purpose can pose a significant environmental impact. The process for creating and training AI algorithms requires large amounts of computing power, which in turn consumes large amounts of electrical energy.

New algorithms can also replicate existing problems in society if the dataset that teaches programs is itself biased. For example, some facial-recognition systems are reportedly better at recognizing white men than black women, because existing image datasets tend to include more men and white people. Thus, the risk of reinforcing biases is a real concern.

There are also concerns over how transparent AI can be, with analysis of digital geographies documenting the unequal global access to the Internet, the positionalities of the analysts who write the code, and the potential for algorithms to 'go bad'. The virtual spaces that Alternative ESG ratings are drawn from are still facilitated by physical infrastructures, such as data centers, cables, and routers, which can limit uninhibited global access. Proprietary algorithms of Alternative ratings can also be vulnerable to 'Google governance' that allows platforms to control what users see, mediating the mobility of information.

It is not possible to access any of the code that drives AI driven ESG ratings, and it could be argued that having opaque algorithms lacks transparency just as much as Traditional ratings based on opaque methodologies. Research has indeed been carried out across the social sciences into how only a small number of technocratic elites have control over these algorithms, hindering public participation. A final criticism levelled at the democratic nature of Alternative ratings is that the Internet is not entirely open to participation but is a contested and unequal space. Although being an 'ethereal, alternate dimension' that is everywhere, access to the Internet is still dominated by developed nations, and thus most online knowledge in the form of academic journals and newspaper articles is generated by developed regions. Thus, the big data that Alternative sources analyze is likely to "exhibit spatial cores and peripheries of knowledge." Fake news and echo chambers online may also skew results. Alternative ratings may also be influenced by analyst subjectivity in the engineering and design phases.

An important final caveat is the distinction between ESG ratings and research. Traditional ESG ratings are paired with written company, industry, and thematic research reports, as well as analyst calls and discussion. This qualitative element is important to many asset managers. Alternative data providers on the other hand prioritize the data and the numbers and employ far fewer analysts to produce accompanying research reports.

Using Machine Learning Effectively

A good starting point for using machines in ESG investing would be a calibrated approach involving a careful mix of human intervention and capability, backed by AI-based tools which provide the most practical and meaningful results. Best practices in ML's application to ESG investing should be guided by a clear understanding and algorithms must be carefully controlled and validated. While ML methods require new skill sets, domain knowledge from finance and investments will remain crucial to beneficial research. The key components of successful ML applications are robust algorithms, infrastructure, and experienced teams. Asset management firms that seek innovative data, find unique alpha, and incorporate machine learning techniques to enhance their ESG investment processes will be most likely to excel and adapt in an ever-evolving industry.

References:

"How machine learning is helping investors find ESG stocks," Gareth Platt, Feb 14 2021

"Alternative ESG Ratings: How Technological Innovation Is Reshaping Sustainable Investment," Arthur Hughes, Michael A. Urban, and Darius Wojcik