

ESG Big Data Mining and Analytics using AI/ML

Budha Bhattacharya

October 2022



ESG in the news...

*Sustainability is ultimately a complex societal challenge, and it is human nature to try and understand complex challenges by boiling them down to simple and understandable concepts, scores and ratings. If we can measure it and we can see it, then we can analyse it and manage it.**

Deliveroo: The worst IPO in history, with a side-order of ESG investor boycott

The Gig Economy is under scrutiny for legal and governance sustainability risks

Goldman Sachs Group ✓ Added

SEC investigating Goldman Sachs for ESG claims

US regulator has been putting more focus on potential 'greenwashing' in finance

DWS Group ✓ Added

German police raid DWS and Deutsche Bank over greenwashing allegations

Asset manager accused of misleading investors about ESG factors in its financial products

[The ESG Mirage - Bloomberg](#)



Elon Musk
@elonmusk · Follow



Exxon is rated top ten best in world for environment, social & governance (ESG) by S&P 500, while Tesla didn't make the list!

ESG is a scam. It has been weaponized by phony social justice warriors.

5:09 PM · May 18, 2022



MSCI, the largest ESG rating company, doesn't even try to measure the impact of a corporation on the world. It's all about whether the world might mess with the bottom line.

By [Cam Simpson](#), [Akshat Rathi](#), and [Sajjel Kishan](#)

[ESG Ratings Are Not What They Seem](#)

...stressed that E.S.G. was crucial for BlackRock, after writing in 2020 that climate change would be a "defining factor" in its investment assessments. Failing key E.S.G. tests could lead BlackRock to sell its holdings in a company: "Access to capital is not a right," Fink wrote

Leaders | Sustainable investing

ESG should be boiled down to one simple measure: emissions

Three letters that won't save the planet

Anti-ESG ETF gets off to a roaring start

Launch coincides with moves by some US states against ESG investing and contrasts with slower progress of similar vehicles in the past

The Big Read ESG investing ✓ Added

The factory by a Tuscan beach and the future of ESG investing

Some investors who want to assess the environmental record of companies worry about how ratings are compiled

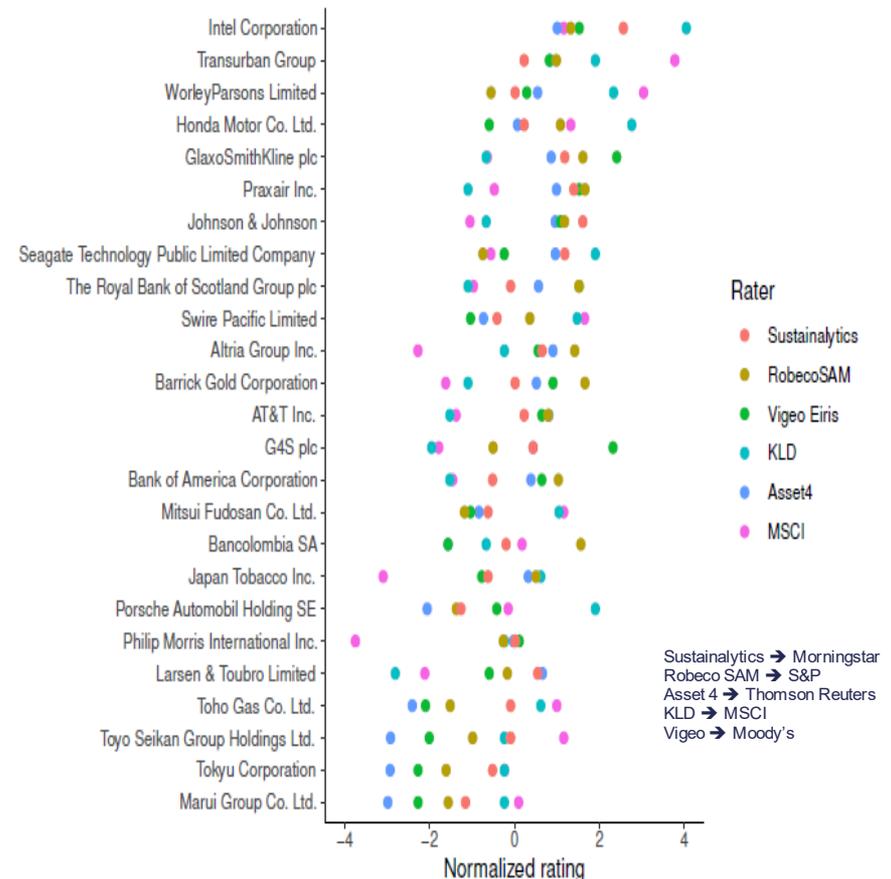
[This White Italian Beach Has a Dark Secret](#)

Problems with ESG Data & Ratings

- ✓ **ESG Ratings are disparate** often 4 to 5 standard deviation apart.
- ✓ **ESG data** is often **inconsistent**; often **scanty** and **imputation models** are varied.
- ✓ Underlying **ESG related data** is often self reported therefore can often be **greenwashed**.
- ✓ Underlying **subjectivities** used in in **ESG Rating methodologies** are **opaque**.
- ✓ Often when buying 3rd party ESG Ratings firms end up buying into **ESG ratings provider's judgement**
- ✓ When it comes to ESG related unstructured data, it's **difficult to separate ESG news from ESG related noise**.
- ✓ Core functions and support functions have **limited capabilities** to comply with the evolving requirements.
- ✓ There are **multiple reporting frameworks** at play, its difficult for clients to decide which is the right one.
- ✓ Most importantly, clients find it **difficult to configure their own subjectivity** into ESG related decisions

Firms with high disagreement

Normalized ratings for the 25 firms with the highest mean absolute distance to the average rating (MAD) within the normalized common sample (n=924). Firms are sorted by their average rating. Each rating agency is plotted in a different color.



Berg, F., Koelbel, J.F. and Rigobon, R., 2019. Aggregate confusion: The divergence of ESG ratings. Forthcoming Review of Finance.

Kotsantonis, S. and Serafeim, G., 2019. Four things no one will tell you about ESG data. Journal of Applied Corporate Finance, 31(2), pp.50-58.

Mining for ESG Big Data



Unstructured Data - Company Specific
 Management Announcements,
 Shareholder results, Company policies,
 Regulatory filings/disclosures

Unstructured Data – Publicly available
 News articles (including controversies),
 NGO Campaigns, social media, web
 pages, industry forums

Legal Agreements
 Trade confirms, loan agreements, IMAs
 and other such legal or marketing
 documents, prospectuses, insurance
 policy documents etc.

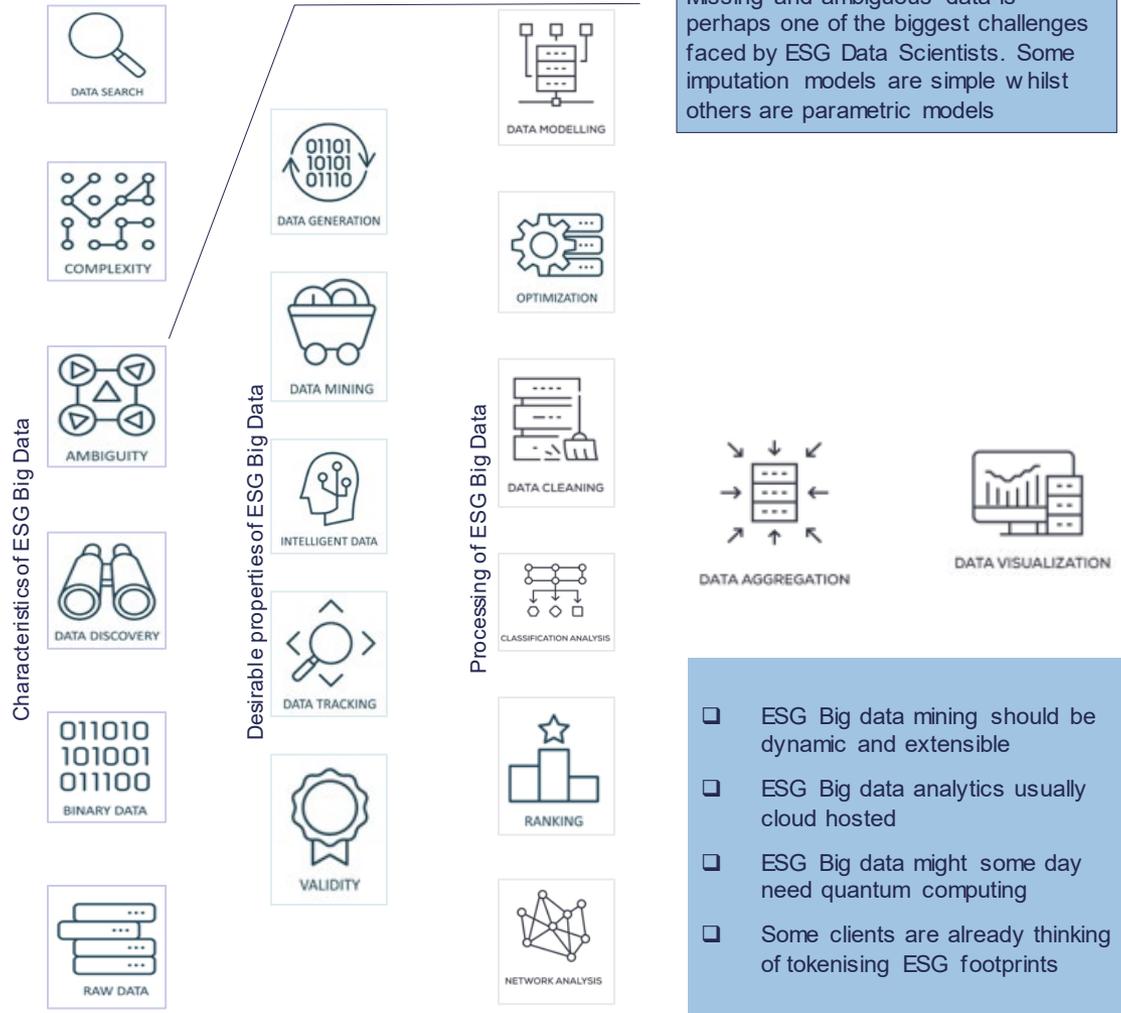
**Satellite, Geospatial, LIDAR, GPS,
 Weather related data such as Climate
 Engine.**

Structured Data
 BBG/ S&P Trucost/ Truvalue Labs/
 Refinitiv, MSCI etc.
 Extended public data sources UN,
 OECD, World Bank, NASA et al.

Client Outreach
 Digital questionnaire - web based UI to
 collect data from clients

Internal Data
 Internally available ESG data such as
 network data etc. (could also be
 unstructured data)

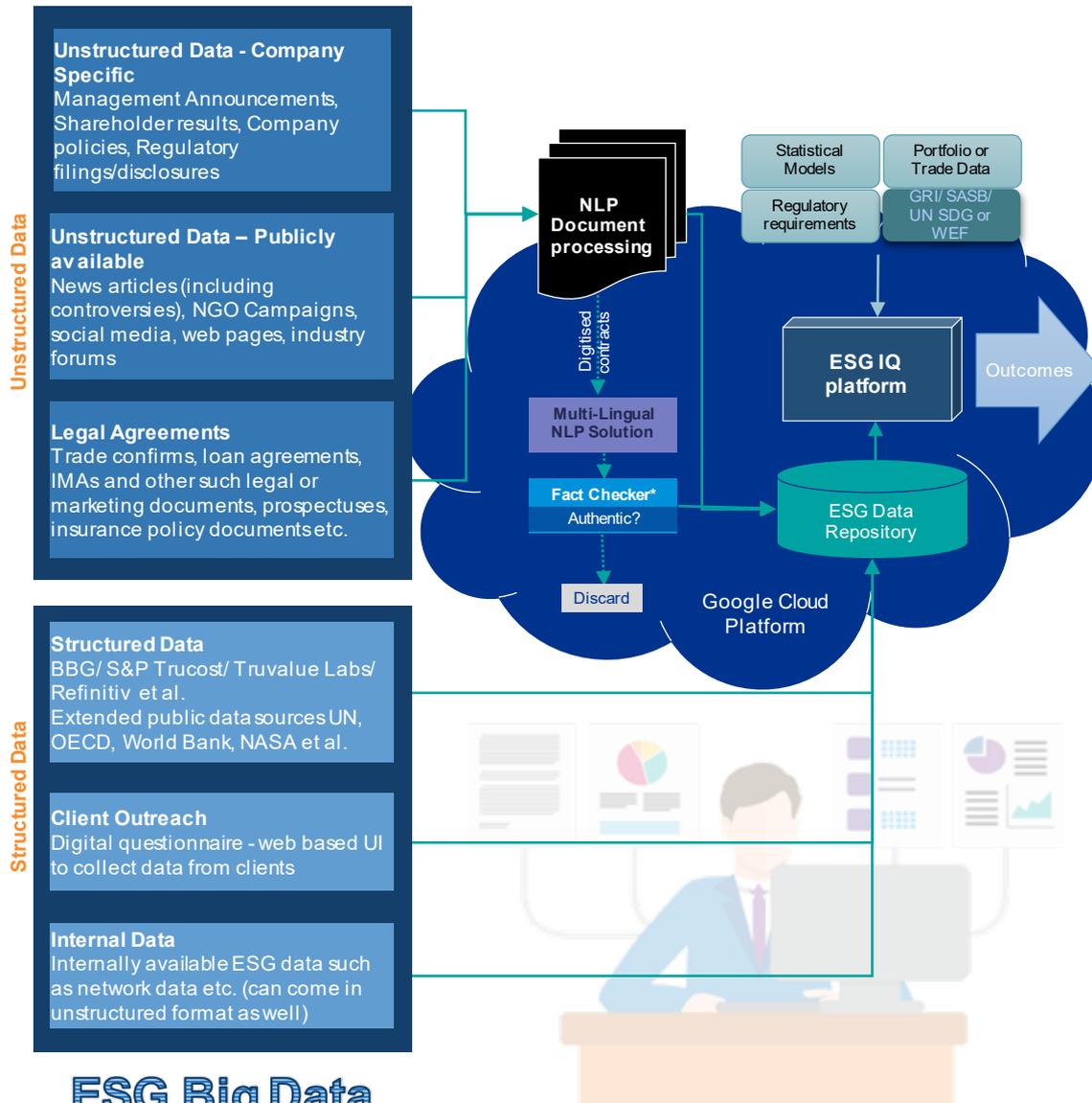
ESG Big Data



Missing and ambiguous data is perhaps one of the biggest challenges faced by ESG Data Scientists. Some imputation models are simple whilst others are parametric models

- ❑ ESG Big data mining should be dynamic and extensible
- ❑ ESG Big data analytics usually cloud hosted
- ❑ ESG Big data might some day need quantum computing
- ❑ Some clients are already thinking of tokenising ESG footprints

Big Data Mining & Analytics application



ESG IQ - modules

Analytics and Bespoke Scoring

- ❑ Customisable data driven analytics platform with multiple data providers and data formats & sources
- ❑ Robust 'Glassbox' ESG methodology - ability to define investment criteria and materiality and toggle across multiple frameworks
- ❑ Client defined bespoke ESG Score per asset/ book/ portfolio / supply chain
- ❑ Dynamic AI/ML driven ESG analytics dashboard with detailed and dynamic data driven insights on ESG standing.
- ❑ Providing clients with the ability to proactively analyse greenwashing risk

Regulatory compliance

Taxonomy eligibility and alignment of portfolio / supply chain or trading book across multiple jurisdictions for EU as well as US, UK, HK, Singapore and more

ESG IQ – ESG DaaS

- ❑ ESG data collation, normalisation, curation, maintenance, lineage and discovery
- ❑ ESG Data Aggregation and repository creation

Risk Modelling – ESG as a principal risk class

Trade level or book level or portfolio level ESG VaR exposure calculation with Risk modelling

- ❑ ESG stress testing / scenario analysis

Reinforcement learning based ESG asset recommendations for portfolio construction

ESG Big Data



AI & ML - Applications

Extracting ESG data

- Most organisations report their ESG standings through their CSR reports, Annual reports and company Websites. We use ML techniques to extract this data and curate this data whilst creating and automating data pipelines.
- Often extensive SOTA models are being used using Google toolkits such as Custom Workbench and more.

Context analysis

- Context analysis can be best executed by contrasting self reported datasets (that is available via 3rd party vendors) against regulatory filings such as EDGAR database. We use advanced scraping and NLP methodologies to achieve this.

Risk Management

- Integrating ESG factors in risk models has now become mainstream for most Financial Services institutions. One of the independent variable in most PD equations now is an ESG score

Recommendation Models

- As companies define their respective ESG appetites we can use products such as ESG IQ to use reinforcement learning based recommendation models to pick assets (equities, bonds, funds and more) to build portfolios, for trading or even to optimise their supply chains.

Asset scoring

- NLP is not only used to extract ESG data but also to score financial assets such as bonds, equities, funds and also illiquid assets such as ABS, RMBS, CMBS and structured bonds

Questions



Budha Bhattacharya
Chief Technology & Product Officer, ESG IQ
budha.bhattacharya@kpmg.co.uk

Thank you

Appendix



What are some of the ESG factors?



Environmental

- Climate change
- Greenhouse gas (GHG) emissions
- Resource depletion, including water
- Waste and pollution
- Deforestation
- Hazardous materials
- Biodiversity



Social

- Working conditions, including slavery and child labour
- Impact on local communities
- Conflict regions
- Health and safety
- Employee relations and diversity
- Product mis-selling
- Data protection



Governance

- Executive pay
- Bribery and corruption
- Political lobbying and donations
- Board diversity and structure
- Tax strategy
- Data breaches

Below are some examples of E, S and G risks for financial services and the wider corporate world:

- Banks' loans, Asset Managers' investments or advice to companies that are carbon intensive or polluters e.g. Goldman advising Aramco, BlackRock investing in ExxonMobil, Royal Dutch Shell, BP
- Insurance companies' lack of consideration for climate change risks
- Outside FS:
 - Companies that use unsustainable sources of resources e.g. palm oil
- The PPI mis-selling for the banks in the 2010s
- The closure of bank branches in more remote areas
- Pay day loans companies' perceived impact on the poorer members of society
- On the positive side, LBG's social housing £500m fund and Barclays Life Skills initiative are examples of attempts at positive social impact
- Outside FS
 - Big tech companies use of consumer data to sell advertising
 - Soft drink companies and high sugar content
- Fines for banks such as HSBC on lack of controls on money laundering
- E.g. Equifax data breach and insider trading accusations 2017
- Outside FS
 - Coca-Cola has a low score due to its high spend on lobbying
 - Competition concerns over big tech companies' dominance
 - Tax avoidance accusations for big tech companies
 - The #MeToo scandals that have impacted retailers such as Ted Baker
 - Accounting scandals e.g. Tesco

ESG factor integration (old school)



idBbGlobalCompanyNam	cntryOfIncorporati	cntryOfDomic	latestPeriodEndC	eqyConsolidati	directCo2Emissio	indirectCo2Emissio	totalCo2Emissio	totalGhgEmissio	scp3BusinessTravelEmissio	gasFlar	emissionReducti	climateChgPro	climateChgPoli	percentOfDisclo
Mattioli Woods PLC	GB	GB	01/05/2021	N				0.235	0.003		TRUE	FALSE	FALSE	
Victoria PLC	GB	GB	01/04/2021	N							TRUE	FALSE	TRUE	
TECHNO ASSO Co Ltd	JP	JP	01/03/2021	Y							TRUE	FALSE	TRUE	
Hinokiya Group Co Ltd	JP	JP	01/12/2021	Y							FALSE	FALSE	FALSE	
Aptiv PLC	JE	IE	01/12/2021	N										
Empresas Cop SA	CL	CL	01/12/2021	N										
Oricon Enterprises Ltd	IN	IN	01/03/2021	Y							FALSE	FALSE	FALSE	
Huvis Corp	KR	KR	01/12/2021	Y										
Askari Bank Ltd	PK	PK	01/12/2021	N										
Leeds Group Ltd	GB	GB	01/05/2021	N							FALSE	FALSE	FALSE	
NorthWestern Corp	US	US	01/12/2021	N										
Luceco PLC	GB	GB	01/12/2021	N										
Biffa PLC	GB	GB	01/03/2021	N				533.187			TRUE	FALSE	TRUE	
Kuraudia Holdings Co Ltd	JP	JP	01/08/2021	Y							FALSE	FALSE	FALSE	
PI Industries Ltd	IN	IN	01/03/2021	Y				143.008			TRUE	FALSE	TRUE	
Oblong Inc	US	US	01/12/2021	N										
Toukei Computer Co Ltd	JP	JP	01/12/2021	Y										
Ellington Financial Inc	US	US	01/12/2021	N										
a2 Milk Co Ltd	NZ	NZ	01/06/2021	N				1.97			TRUE	FALSE	TRUE	
Yotai Refractories Co Ltd	JP	JP	01/03/2021	Y							FALSE	FALSE	FALSE	
Palomar Holdings Inc	US	US	01/12/2021	N										
Teikoku Electric Manufact	JP	JP	01/03/2021	Y							TRUE	FALSE	TRUE	
Young & Co's Brewery PLC	GB	GB	01/03/2021	N				8.43			TRUE	FALSE	FALSE	
Spirit Airlines Inc	US	US	01/12/2021	N										
BW LPG Ltd	BM	SG	01/12/2021	N				1335.44			TRUE	FALSE	TRUE	
Altech Corp	JP	JP	01/12/2021	Y							FALSE	FALSE	TRUE	
Pennar Industries Ltd	IN	IN	01/03/2021	Y							TRUE	FALSE	TRUE	
Grammer AG	DE	DE	01/12/2021	N							TRUE	FALSE	FALSE	

600 factors

20,000 companies

Fund Construction

Utilise traditional portfolio construction & optimisation strategies

- Equity long only
- Equity Long/ Short
- Global macro strategy
- Event driven
- Relative value
- CTA/Systemic/Quant

External ESG Ratings

Use MSCI or Sustainalytics or another or a combination of other ESG rating agencies

Create 'personalised' weightings on the scores – often difficult to explain investors and regulators alike

Often scores are aggregated to create composite scores

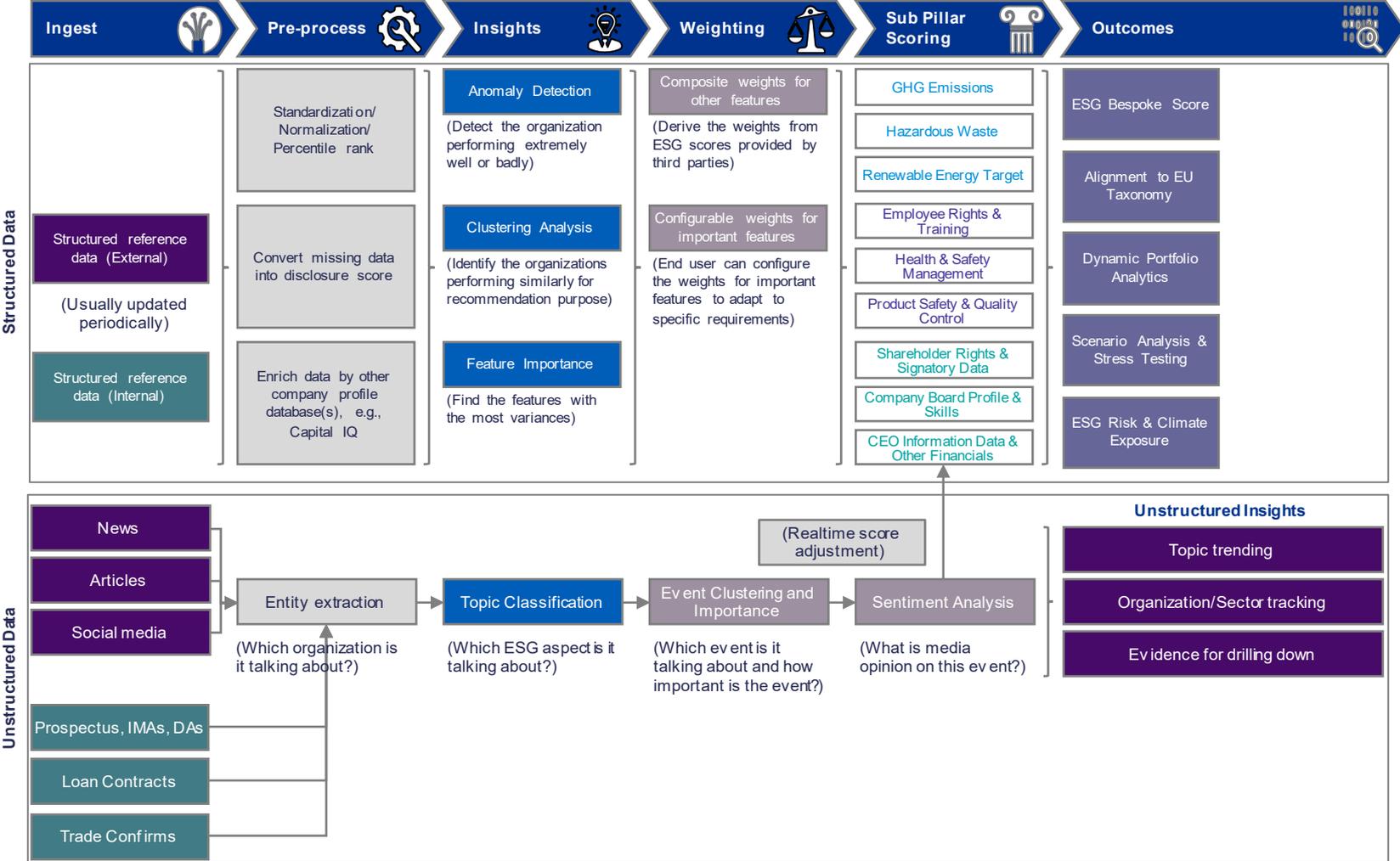
Market funds as Green/Responsible

Exposed to investor protection issues or mis-selling thus regulatory risk

Often lose RfPs due to lack of customisation

Trying to define internal weightings and materiality, marrying 'new standards' to ESG strategy

Methodology – Data & Outcomes

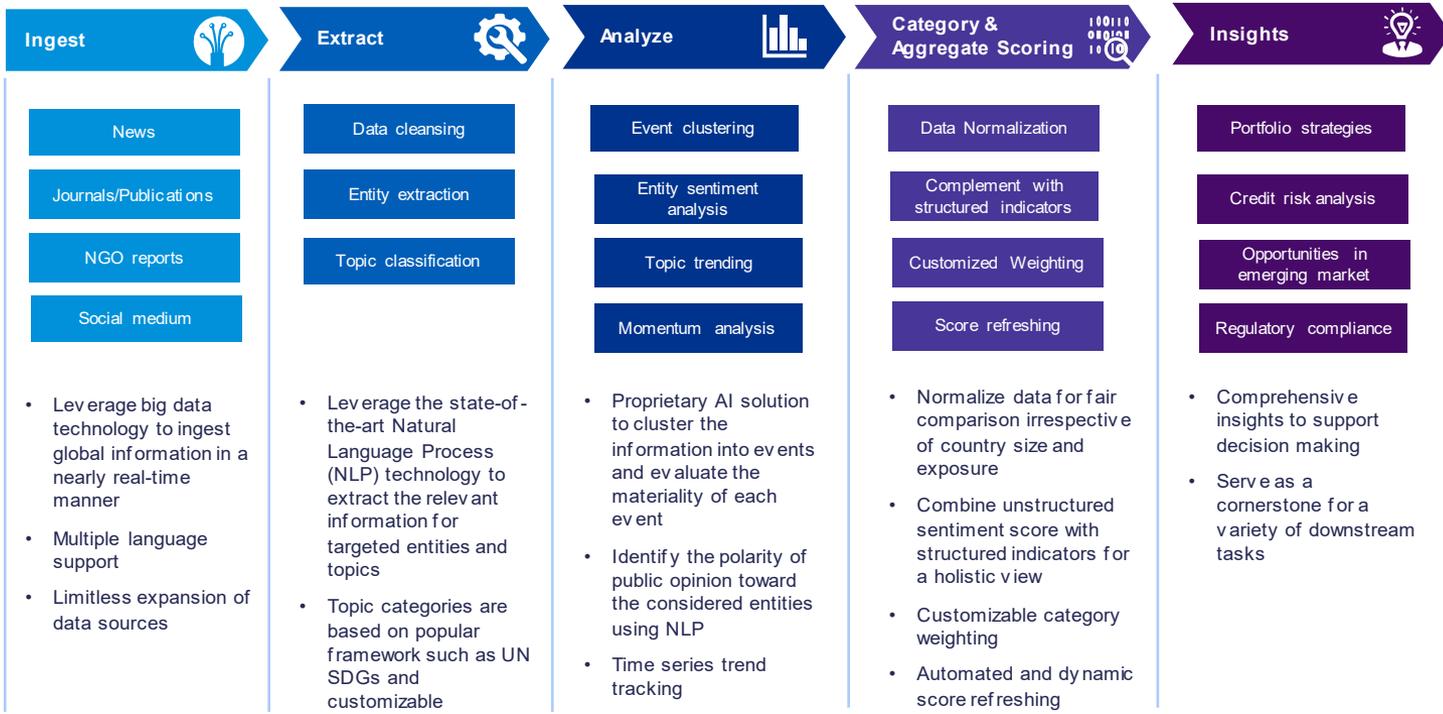


Methodology – Bonds

Challenges

Sovereign ESG performance has been dominantly evaluated using various structured data indicators, e.g. <https://datatopics.worldbank.org/esg/>. However, the current approach faces the following challenges:

- There are many more other ESG aspects which have not been incorporated in the collated datasets or difficult to be quantified
- The data points are usually updated monthly or yearly
- Lack of a gauge to really measure what they are doing rather than saying



Solution

Thanks to the latest innovation of big data and AI technologies.

We are now in a position to innovate ESG investing by being able to exploit the invaluable signals embedded in numerous unstructured data sources, which can provide:

- A limitless scope covering various ESG aspects in various unstructured data sources such as news, journals, reports, and social media posts
- Timely refresh based on the latest information
- Public consensus to gauge the sovereign ESG performance

Enabler: Partnership with Google Cloud Platform Partnership with third-party data vendors