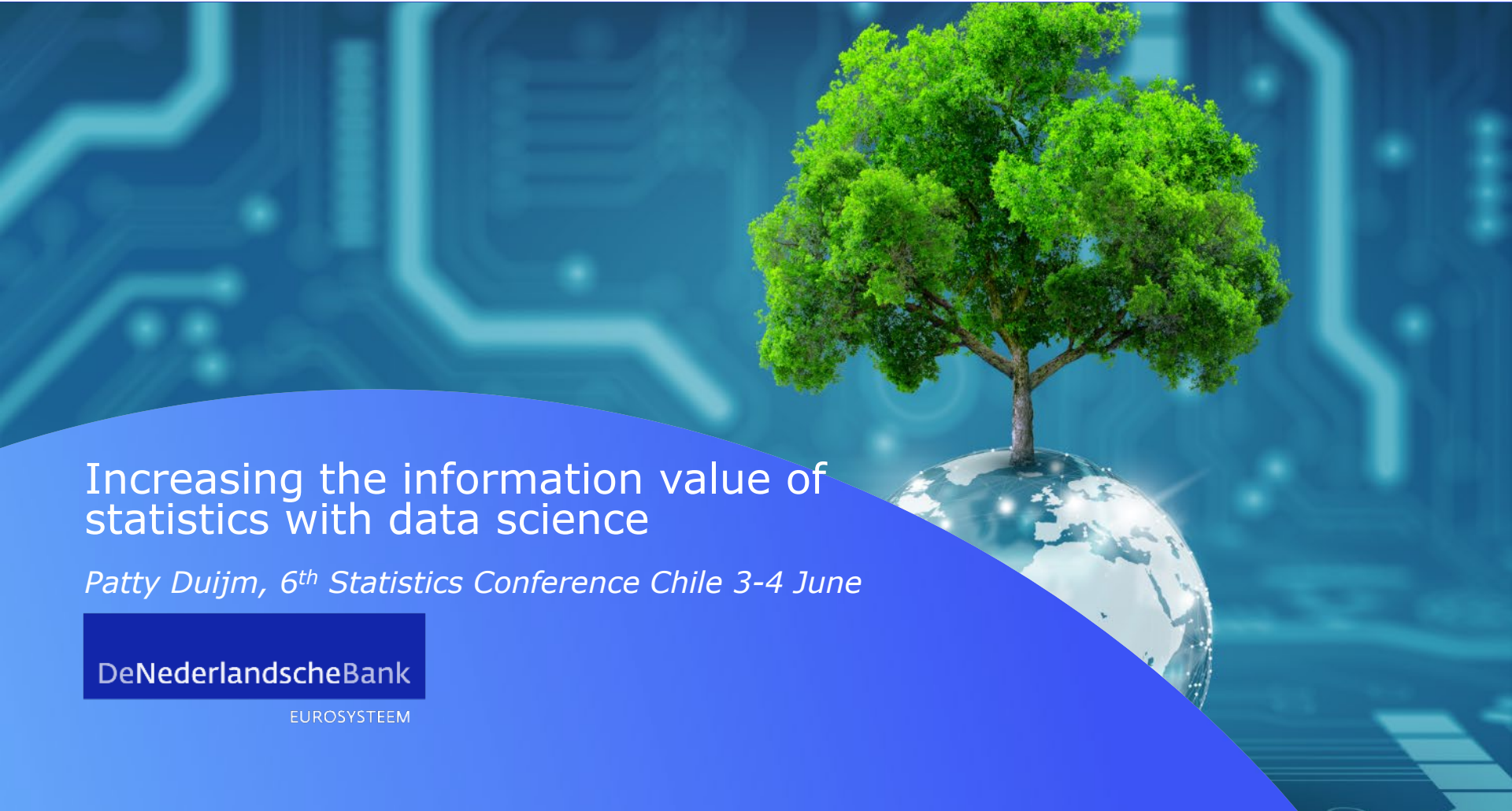


Increasing the information value of statistics with data science

Patty Duijm, 6th Statistics Conference Chile 3-4 June

DeNederlandscheBank

EUROSYSTEEM

Patty Duijm

Background in Econometrics/Risk Management

- 2013 — Started at DNB's Financial Stability Division
- 2015 — Start of part-time PhD
- 2017 — Policy advisor Supervision Policy
- 2019 — PhD defence "On the cyclical nature of finance"
- 2020 — Data Science Hub (involved in setting up the new department)
- 2023 — Head of Department Securities, Sustainability & Payments Statistics

DeNederlandscheBank
EUROSYSTEEM

This presentation

- How to employ data science techniques to increase the information value of data and statistics.
- Based on real use cases from DNB.
- A strong focus on the process of producing *sustainable finance statistics.*

Background – organization

In this presentation all use cases that are shown are stemming from:

DataScience Hub

Data Science Hub projects

- DNB's Data Science Hub officially set up in 2020
- Main goal: support colleagues throughout the organization to implement data science in their work.
- Work is done on a project basis. Currently ~70 finalized data science projects, ranging from HR, pension funds to integrity supervision and statistics.

Sustainable Finance Statistics

- In line with Governing Council action plan and new roadmap, ECB, DNB and ESCB (all other Eurosystem NCBs) committed to produce euro area statistical indicators.
- No regular data reporting; input data is sourced from public sources and existing ESCB reports.
- Statistics are still in a developing phase; joint effort with ECB and other ESCB countries to improve and enhance the indicators. DNB co-chair of the Expert Group and strongly involved.
- Statistics on green bonds, carbon indicators, and physical risk.

DeNederlandscheBank
EUROSYSTEEM

Background – Sustainable Finance Indicators

Carbon Indicators

- Ability to measure **relative carbon footprints** of financial institutions via their investments.
- To monitor **transition risk** in the financial institutions' portfolios.
- **Combination** of data on corporate securities and loan portfolios with underlying firm-level data and climate data.

Various indicators, of which the WACI (Weighted Average Carbon Intensity):

Note: in our work scope 3 is currently not included

Physical Risk Indicators

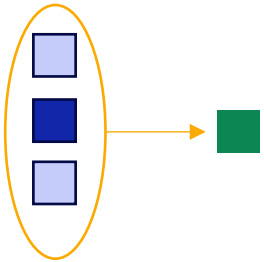
- Capture financial system exposures to companies located in areas susceptible to **natural disasters** (such as flooding, windstorms, wildfires or droughts) and chronic physical risks (heat and water stress).
- The **combination** of location data with climate risk scenarios.

Various indicators, of which the PEAR (Potential Exposure At Risk):

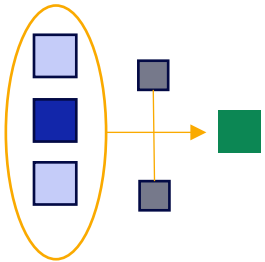
$$PEAR = \frac{\sum_{i=1}^{N}(EXPOSURE_i | RS_{i,j}(j>0))}{\sum_{i=1}^{N}(EXPOSURE_i)}$$

Elements of data analysis and statistical process

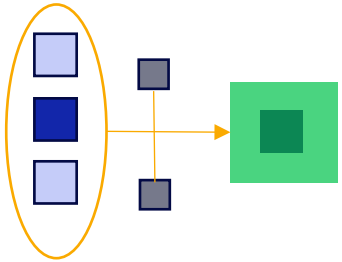
- In data analysis, but also for statistical processes, we often combine, enrich or expand data.
- For each of these three elements we showcase examples where data science helped us.



Combining
Effectively using granular data



Enriching
Use of external sources



Expanding
Increasing the timeliness of the data

Combining

*"The **combination of - sometimes novel - granular data** offers new insights useful for both supervisors and policymakers."*

Examples are SHS, Anacredit, EMIR, MMSR.

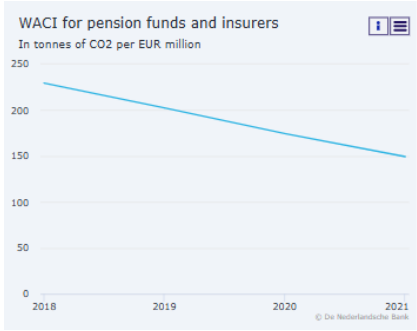
Three challenges:

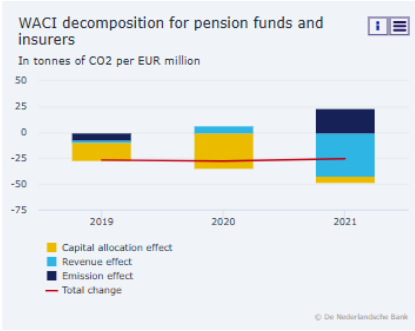
- Counterparty names as free-format text
- No single unique identifier
- One needs to consider different sources of granular data

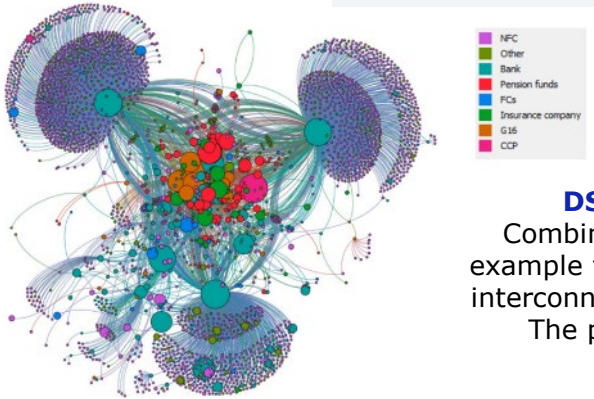
DeNederlandscheBank
EUROSYSTEEM

Sustainable finance statistics

For the compilation of these statistics we combine various (internal and external) data sources, e.g. financial statistics, company information, climate data.







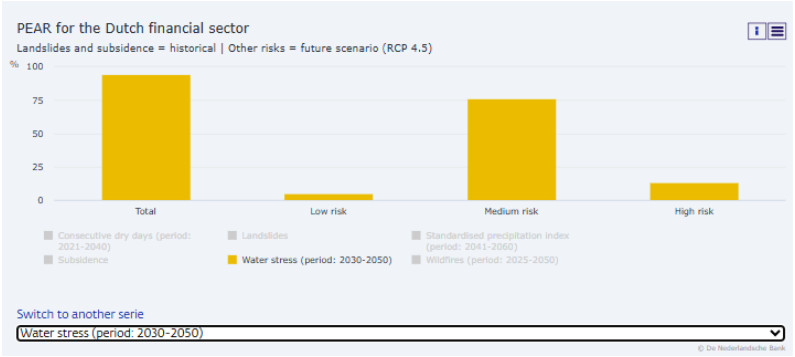
DSH project: Network analysis

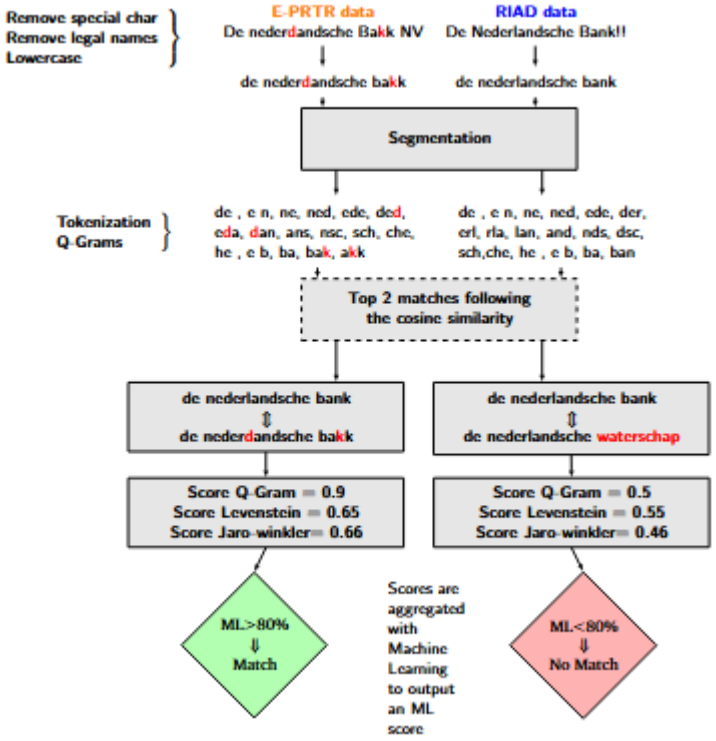
Combining granular sources is valuable, for example for network analyses and for detecting interconnectedness within the financial system. The picture shows interconnections via derivative trades solely.

Background: "Estimating Initial Margins – The COVID-19 market stress" (Van den Boom et al., 2021)

Combining: Name Matching (1/2)

- For **sustainable finance statistics**, name matching is currently applied to combine data on:
  - E-PRTR (European Pollutant Release and Transfer Register)
  - RIAD (Register of Institutions and Affiliates Database)





DeNederlandscheBank
EUROSYSTEEM

Combining: Name Matching (2/2)

DSH Project: Name Matching package

At DNB, the Data Science Hub developed a fuzzy name matching package (available on GitHub!)

https://github.com/DeNederlandscheBank/name_matching

AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CRDITO COOPER
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO C
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO CCOPERATIVO CASSE RURALI ED ARTIGIANES.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPE
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO - CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE - S
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RUTALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO E CASSE RURALI E ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO/CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO=CASSE RURALI ED ARTIGIANE - S
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO=CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO=CASSE RURALI ED ARTIGIANE - S.P.A. DE
EFFETTO "BANCA AGR
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO-CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO-CASSE RURALI ED ARTIGIANE- S.P.A.
AGRILEASING - BANCA PER IL LEASING DELLE BANCHE DI CREDITO CPERATIVO CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING - BANCA PER LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE - S.P.A.
AGRILEASING BANCA LEASING BANCHE CREDITO COOPERATIVO SPA RM
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CF
ARTIGIANE S. P. A.
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE S.P.A
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE S.P.A.
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE SPA
AGRILEASING BANCA PER IL LEASING DELLE BANCHE DICREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE SPA
AGRILEASING BANCA PER IL LEASINGDELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE S.P.A.
AGRILEASING BANCA S.P.A.
AGRILEASING-BANCA PER IL LEASING DELLE B.C.C.- SPA
AGRILEASING-BANCA PER IL LEASING DELLE B.C.C.-C.R.A.
AGRLEASING BANCA PER IL LEASING DELLE BANCHE DI CREDITO COOPERATIVO CASSE RURALI ED ARTIGIANE S.P.A.
BANCA AGRILEASING S.P.A.

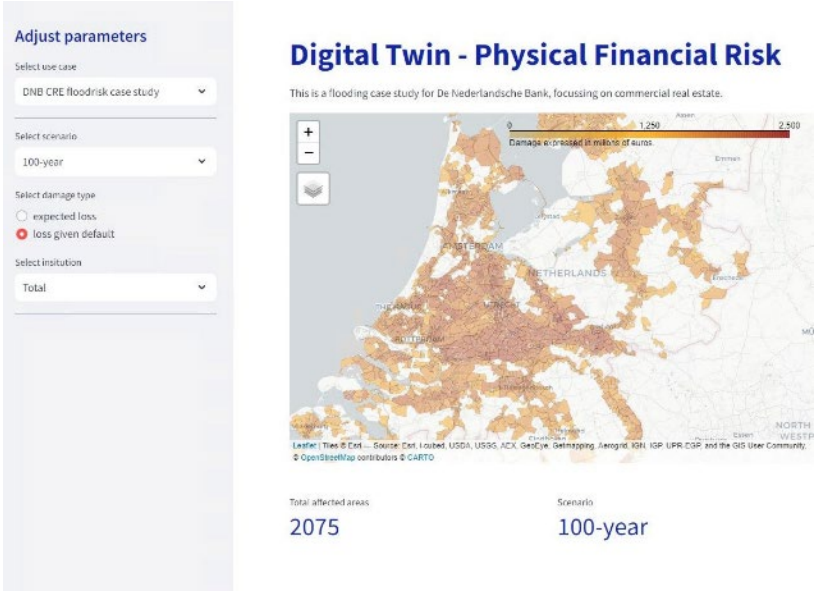
Nijhuis (2022), Company name matching

Enriching: external sources

"Enriching internal data with external resources increases the information value of the data, allowing supervisors and policymakers to improve the incorporation of external factors and risks."

The data you own is much more valuable to you if it is augmented with data owned by others
(Mewald, 2023)

Especially these "alternative" sources of data may generate unique insights.



DSH Project: Digital Twin of Climate Risks

In cooperation with the BIS Innovation Network, the Data Science Hub has developed a digital twin pilot of climate risks. The digital twin was developed to measure the effects of climate events on the financial system via real estate exposures of financial institutions.

→ At the moment, the sustainability statistics team and the DSH are investigating whether the visualization can also be applied to the physical risk indicators produced by the statistics team

DeNederlandscheBank
EUROSYSTEEM

Enriching: DataFetcher

"Automating data processes results in more efficient, accurate and highly frequent economic (prediction) models and allows policymakers to focus on modeling rather than preparing the data."

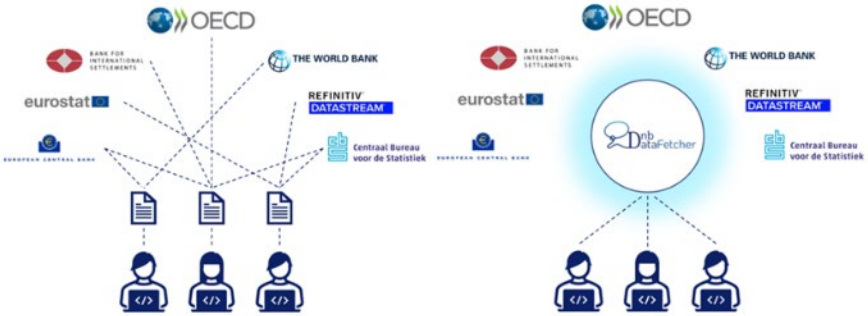
Manual data wrangling is a labor intensive job, prone to human errors.

Automating data collection processes results in more efficiency and accuracy.

DeNederlandscheBank
EUROSYSTEEM

Sustainable finance statistics

For the production of these statistics we also rely on public external sources (e.g. inflation rates and emission data from Eurostat). The team uses the DataFetcher to effectively "fetch" the data!



DSH Project: the DataFetcher

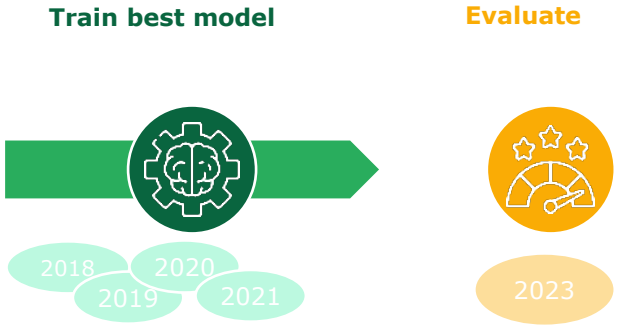
The DataFetcher is a package developed by the DSH and the figure illustrates its purpose. Multiple processes within the central bank use external data sources, resulting in colleagues collecting (the same) data manually or via ad hoc scripts. This may also result in cases where different (or even outdated) versions of the same data set are used within DNB. This is the left panel. In an ideal situation, i.e., the right panel, colleagues within the same institution have immediate access to the same data while restrictions dictated by privacy and confidentiality should be respected.

Expanding: Nowcasting emission data (1/2)

- For the production of our **sustainable finance statistics** we have to rely on firm-level data on emissions and financial performance, sourced from annual statements and offered by commercial providers. Some sources are only available with a substantial time lag.
- Can we forecast the 2024 revenue, company value, scope 1 and scope 2 emissions for the compilation of the statistics early 2025?

	2022	2023	2024
Investment	Observed	Observed	Observed
Revenue	Observed	Observed	**Nowcast**
Company value	Observed	Observed	**Nowcast**
Emissions	Observed	Observed	**Nowcast**

Model evaluation phase

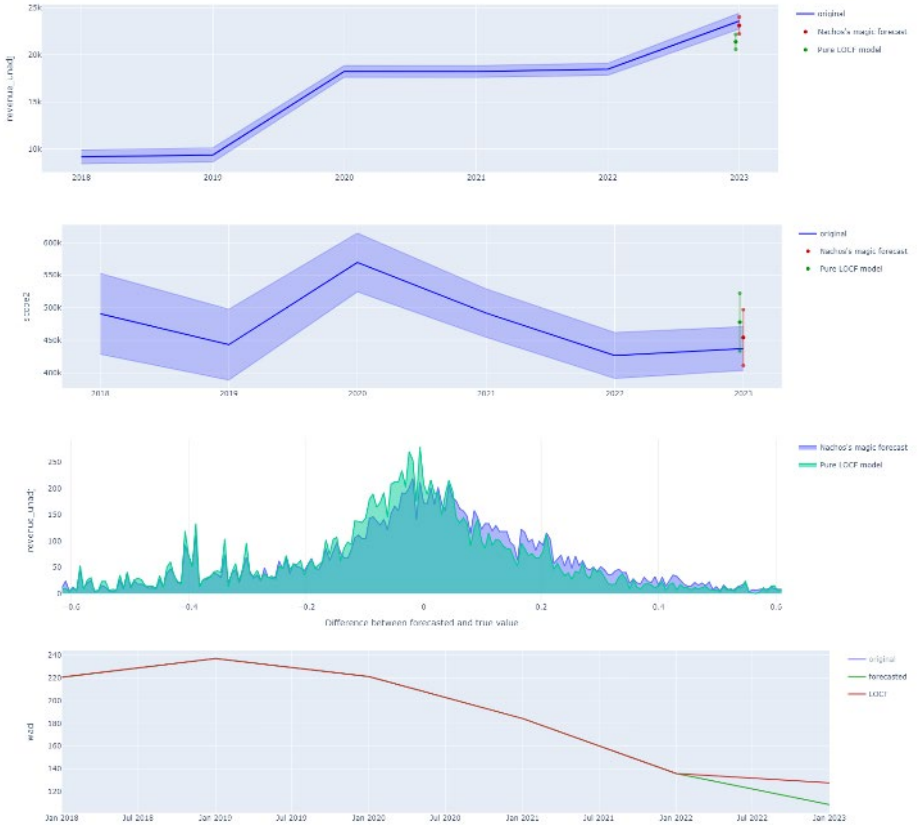


DeNederlandscheBank
EUROSYSTEEM

Expanding: Nowcasting emission data (2/2)

Experiences so far:

- Nowcasting elements (revenue, emissions) of the indicator WACI gives plausible results.
- LOCF (Last Observation Carried Forward) used as benchmark: typically, high accuracy but potentially biased
- Adding credit rating data to nowcast revenue improves the model by reducing bias
- Figures based on OLS (scope 2 emissions) and LGBM/XGBoost (revenue). Other models that implemented in the package: simple heuristics such as WMA, WEX; machine learning models such as Random Forest and kNN.



Conclusion

- Advanced data techniques and data science have great potential for statistical processes.
  - The **cooperation** between colleagues within the sustainability statistics team and the data science hub is really valuable.
- Central banks (and statistical agencies) are in a unique position to combine traditional and unconventional data to create new and insightful data products.
  - Furthering the work on **unique identifiers** is needed to facilitate the opportunities.
- Techniques can be applied to **other statistics**
  - For example, nowcasting can be used to increase the timeliness of the published statistics by adding forecasted statistics.
- But **much more** is possible…
  - Sometimes a single new data point will add more value to your organization than ten policy notes combined.

Want more?

Sustainable finance data *Want to know more about our statistics? Have a look!* Sustainability Dutch financial sector | Dashboard | De Nederlandsche Bank

Interested in more data science use cases, and applications beyond statistics? Have a look at the **Data Science Hub** *Annual Reports via: Data Science Hub - De Nederlandsche Bank | De Nederlandsche Bank*

DeNederlandscheBank
EUROSYSTEEM