ESG Score Prediction

16.10.2023 - Team SustainBrAIn
Marius Bosch, Nikita Wilms, Selchuk Hadzhaahmed

Team Slide

Selchuk Hadzhaahmed,
Mathematical Engineer

Nikita Wilms,
Data Scientist

Marius Bosch,
Power Systems Engineer

Agenda

1. Introduction to ESG
2. How are ESG Rating created
3. Machine Learning approach & data product
4. Model & Results
5. Outlook

There are three pillars of sustainability in today's organisations

Environmental

How a company impacts nature and the climate

Social

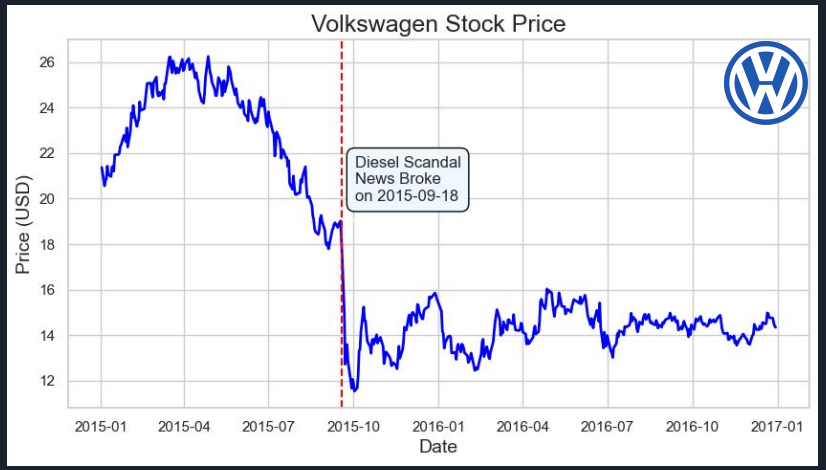
How a company treats its employees, suppliers, customers, and communities

Governance

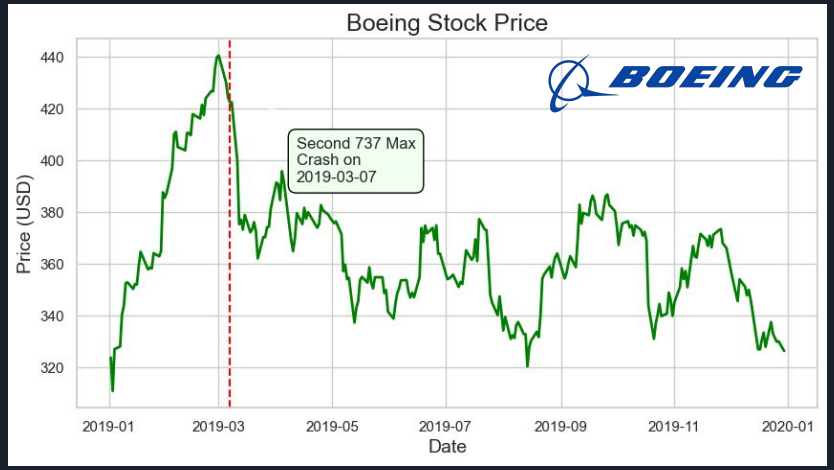
How a company is run, executive pay, internal controls, shareholder rights

Investors and businesses are putting more emphasis on sustainability

Volkswagen stock performance after 2015 **Dieselgate**



Boeing stock performance after **737 MAX crashes** in 2018/19



ESG performance is crucial for **investment decisions, stakeholder trust**, and **long-term sustainability**

Quantifying sustainability performance is difficult and ESG scoring methodologies are mostly intransparent

ESG Rating life cycle (yearly)

Company X creates annual sustainability report

Submission to third party ESG Rating providers

Providers **quantify ESG performance** based on given input

Providers award & **publish ESG Rating**

Reports with ~20-250 pages of text, images, tabular/numerical data

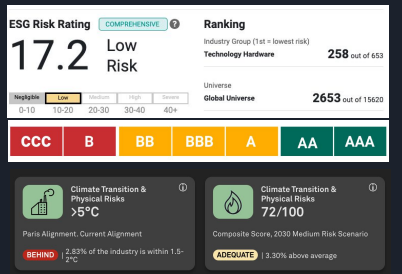
MSCI

SUSTAINALYTICS

Bloomberg

~10 private providers dominate market for ESG scoring

ESG report as starting point, enhanced with publicly available data on:
Industry performance/outlook, financial statements, desk research



All providers with distinct rating methodologies

We use Machine Learning to predict ESG Scores based on textual data from sustainability reports

ESG Rating life cycle (yearly)

Company X creates annual sustainability report	**Submission to third party** ESG Rating providers	Providers **quantify ESG performance** based on given input	Providers award & **publish ESG Rating**

Data product

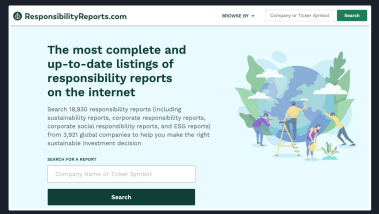
- Allow companies to upload ESG report
- Model will predict potential ESG score based on the contents of the report
- Provide recommendations on areas-of-development
- Create less dependency on expensive consulting services regarding report structure, writing, focus areas

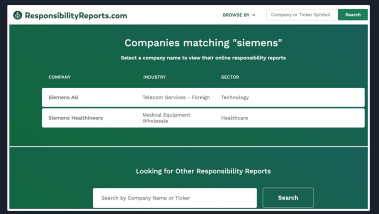
SustainSight

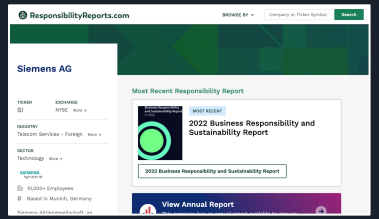
Over 1300 sustainability reports from S&P 500 companies serve as training data for NLP model

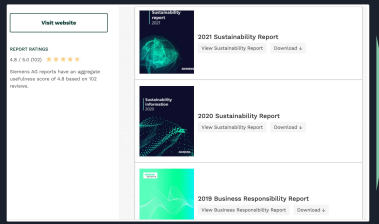
Data collection

- **Developed Web-scraper** automatically select company and relevant year, close pop-ups and downloading relevant sustainability report from ResponsibilityReports.com (runtime ~8 hours)
- Report dates range: **2014-2022**
- Accessed Y**ahooFinance API adding company & industry data** (~400 model features - e.g. revenues, no. employees, emissions)







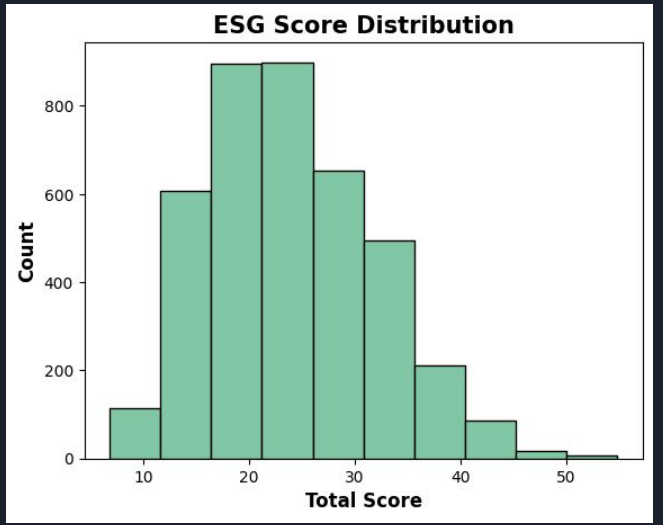


~1300 reports
(15GB)

Target scores range from 0 to 50 and dataset is approximately normally distributed

Target data collection

- Accessed YahooFinance API to access **historic ESG-Ratings provided by Sustainalytics**
- Presented **as Risk-Score** (lower is better)
- **Rating from 0 to ~50** (max value for this dataset at 55)



Classification of exposure to ESG Risks

Negligible	Low	Medium	High	Severe
0 - 10	10 - 20	20 - 30	30 - 40	40+

Extensive text preprocessing on report files was necessary

ESG Score prediction process

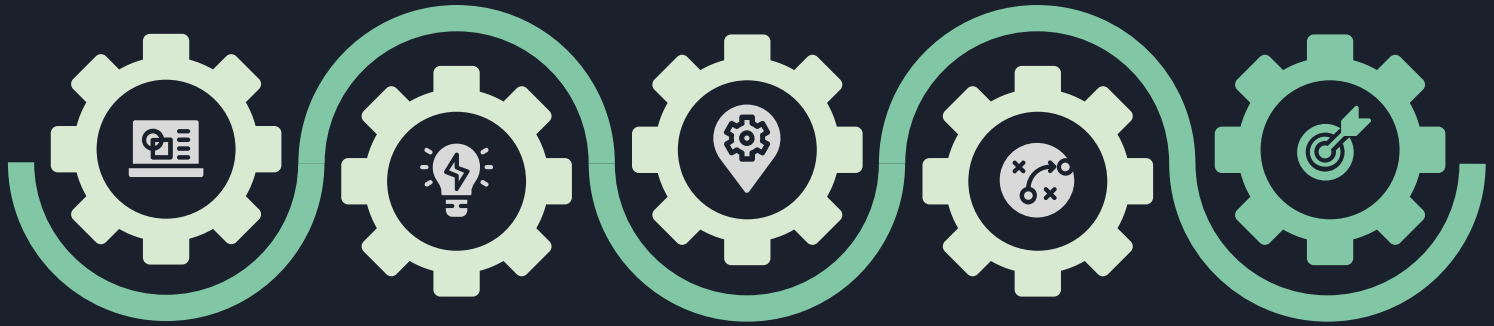
1. Collecting sustainability reports and historical ESG-Ratings

2. Text extraction and preprocessing using NLP techniques

3. Extensive topic modelling using trained and untrained NLP models

4. Creating and tuning ensemble model to predict ESG-Rating

5. Evaluating model predictions with new reports



Before vs. After preprocessing

Example report page

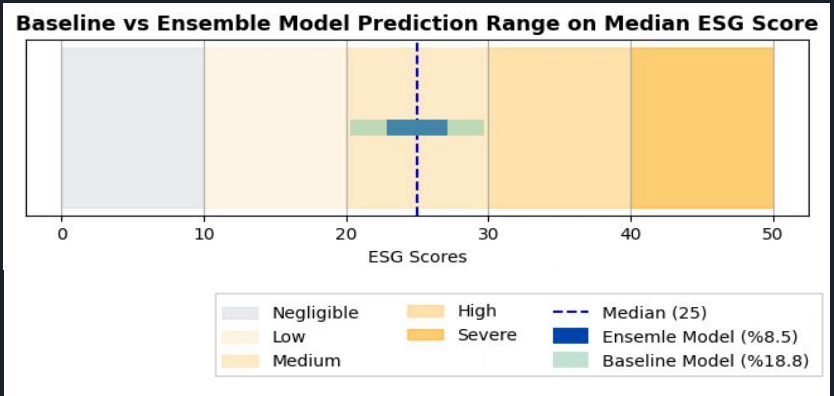
[…low_cost estimate initiative help save sheet_paper lead reduce waste carbon_emission convert waste renewable_resource turn food_scrap compost local farm compost estimate pound_food waste_diversion program read climate_change measure initiative impact health understand impact thing climate_change center support healthy environment future company carbon_impact limited climate_risk_assessment process focus impact climate_change member issue second include discussion governance_structure strategy risk opportunity metric relate manage climate_change utilize risk_management…]

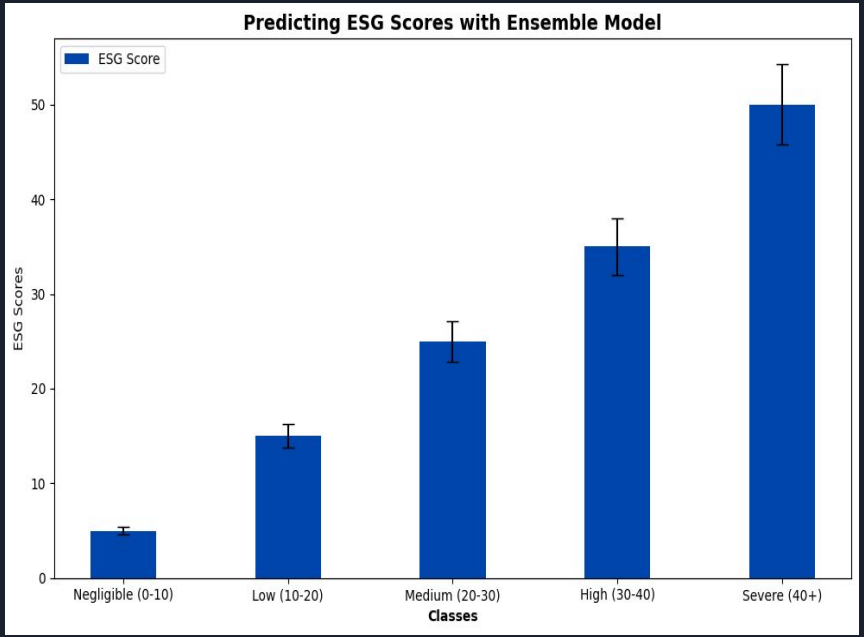
The final ensemble model predicts within an average range of 8.5% from actual ESG-ratings

Model evaluation

- **Baseline** model (Linear Regression) is within **18.8%**
- **Final** model (ensemble) is within **8.5%**

of actual score[1]





1. Root Mean Squared Error (RMSE) served as evaluation metric for all models

Outlook

- Further increase model quality (increase no. of sustainability reports, incorporate tabular data from reports)
- Improve training runtime
- Deploy simple web-application to upload and predict ESG-Rating

Thank you for your attention!