Group Assignment: Supplier Sustainability Analytics — Prediction & Insights (Philips SSP)

In this assignment, you will replicate the predictive analytics core of the CARE framework implemented at Philips, focusing on supplier sustainability assessment and insight generation. You will NOT build the prescriptive optimization model.

1 Case Description

In 2016, Koninklijke (Royal) Philips N.V. (hereafter referred to as Philips) implemented and started a collaborative approach that they refer to as the "Supplier Sustainability Performance" (SSP) program [3] to enhance the sustainability levels of selected suppliers. The program consists of collecting information about suppliers' sustainability levels directly from suppliers. Such an approach is highly resource-intensive, considering the efforts and costs associated with data collection and on-site visits. Many Multi-National Companies (MNCs) rely on structured sustainability audits to monitor their suppliers, followed by incentives and penalties to direct suppliers toward sustainability compliance [4], but these audits have several limitations. Firstly, the auditing activities are often outsourced to third parties, and suppliers finance them themselves, creating a conflict of interest that encourages biased evaluations by the auditors to mediate negative findings and exaggerate the positives [2]. Secondly, audits are based on a pass-or-fail mentality to check supplier compliance, instead of addressing the root problems and focusing on improvement. Empirical evidence suggests that traditional auditing is ineffective in eliminating sustainability violations [4]. Jiang [1] suggests that supplier sustainability compliance can only be achieved by combining supplier assessments with continuous improvement and with consultancy support from buyers. Finally, suppliers are subject to repeated audits, possibly by multiple buyers, resulting in "audit fatigue". Leading MNCs try to alleviate this problem by constructing alliances to jointly conduct audits and/or share audit results [5].

Philips has had similar experiences with traditional auditing. However, Philips had about 15,000 first-tier suppliers when they launched their SSP program, a large portion of whom are located in China. Monitoring and auditing such a large global supply chain without resorting to analytics is challenging and costly. Therefore, the objective is to estimate the level of supplier sustainability without resorting to costly supplier auditing activities. How could this be possible?

Philips started collecting data concerning the sustainability of its suppliers by sending an online Self-Assessment Questionnaire (SAQ) to its selected suppliers. Currently, 1262 self assessments had taken place for a total of 463 suppliers. Of these suppliers' headquarters, around 93% are located in China and the rest are located in other countries such as India, Germany, Indonesia, Italy, Taiwan, Thailand, Türkiye and Yemen. The SAQ consists of a multitude of questions, some of which have a hierarchical structure, and there are sub-questions contingent on answers. The SAQ is divided into five categories: i) General Information (GI) of supplier, ii) Environment (E), iii) Health & Safety (HS), iv) Business Ethics (BE) and v) Labor and Human Rights (LHR). The latter four are referred to as the "sustainability topics", which contribute to the sustainability level (score) of the supplier.

The SAQ contains information requests, such as the location of the supplier, the number of employees, the number of management and executive staff, the gender distribution among workers, ownership of the company, performed activities, obtained certificates and the existence of facilities concerning sustainability, such as water treatment facilities, hospital and kitchen. Thus, GI is used to keep a record of all possible variables that reflect a supplier's characteristics, which can impact or relate to the supplier's sustainability. Naturally, suppliers can make deliberate or unintentional mistakes while filling out the SAQs. To ensure truthful and careful declaration, suppliers are asked to provide physical or digital Evidence Pieces (EPs), depending on their answers to the SAQ. EPs can be in the form of a proof document, certificate or even pictures from the supplier site. Philips then evaluates these EPs to confirm the responses of the suppliers. Furthermore, a limited number of SSP suppliers also receive on-site visits by the assessors from Philips approximately once a year or less frequently to follow critical ones closely. The on-site visits are different from third-party audits. First, the assessors are Philips' employees to avoid conflict of interest. Second, the main purpose is close collaboration with suppliers for supply chain sustainability improvement, rather then mere monitoring. Discussing sustainability issues on-site provides the opportunity to make first-hand improvement suggestions, while also helping with SAQ validation and further data collection, if necessary. But how supplier sustainability scores are calculated?

EPs mentioned above are used to obtain sustainability scores of suppliers. No matter a supplier is visited on-site or evaluated online, EPs provides the basis for the level of sufficiency. The SAQ answer of a supplier is validated by the EPs, and then validated scores constitute the actual scores for each sustainability topic and the final sustainability score. Observe that, this process actually corresponds to labeling the sustainability scores of the suppliers. Once the sustainability scores are available, one can start working on predicting supplier sustainability scores based on the SAQ responses. Briefly, EPs are used to label the training data which requires years of effort to accumulate. Then, predicting sustainability level of a new supplier can be accomplished based on its GI and SAQ responses. Here, one can start using machine learning tools for sustainability predictions. Please refer to the work by Tan et al. [6] for more details and inspiration.

Note on Data Confidentiality: Due to proprietary restrictions, SAQ question text is not provided. Questions are represented as Q-codes (e.g., Q1272) mapped to topics (Policy, Procedures, Implementation, etc.) and chapters (Environment, Health & Safety, Business Ethics, Labor & Human Rights). This is sufficient for feature importance analysis and business interpretation.

2 Learning Objectives

- Build and evaluate ML models that predict supplier sustainability scores from General supplier Information (GI) and Self-Assessment Questionnaires (SAQs) without on-site visiting and time-consuming auditing activities.
- Identify which features/questions matter using model explanation techniques (e.g., SHAP) or regularization techniques to ensure interpretability.
- Calibrate and translate model outputs into risk segments/action thresholds.
- Understand the trade-off between model accuracy, interpretability, and data collection cost.

3 Data

Primary source: SSP Data.xlsx. Use the data dictionary to navigate sheets and columns. Elements include: supplier identifiers, general info (GI), topic scores (E, HS, BE, LHR) and overall score (OS), assessment dates/sequence numbers, and question-level evidence flags.

Important: The dataset provided may differ from the exact dataset used in Tan et al. [6]. Your results may not exactly replicate the paper's performance metrics. Focus on methodology, interpretation, and insights rather than exact numerical replication.

4 Tasks

4.1 Data Preparation & Feature Sets

- Encode categorical variables (one-hot or target encoding), create missingness flags, and standardize numerics as needed.
- Document any data cleaning assumptions (duplicates, outliers, NA handling).
- **Important:** Split data by supplier ID (not by random rows) to prevent data leakage across assessments from the same supplier.

4.2 Model Development

Stage 1: Start with GI-only features (Activities, Facilities, Workforce, Country — 31 features)

- Build baseline and tested models using only observable supplier characteristics
- Analyze feature importance: Which characteristics predict sustainability?
- Compare with findings in Tan et al. [6] Table D.1
- *Learning goal:* Understand that observable characteristics provide modest but interpretable predictions

Stage 2: Add SAQ features (379 additional questions)

- Rebuild models with GI+SAQ features (410 total)
- Measure performance improvement from adding detailed questionnaires
- Discuss trade-off: accuracy vs. data collection burden vs. interpretability
- *Learning goal:* Understand when detailed questionnaires are worth the cost

Task Details:

In this case study, we are interested in predicting the sustainability scores of Philips' suppliers with minimal effort. The target variable definition might change depending on the perspective, as follows:

A1) Regression
Predict Overall Sustainability (OS) score at an assessment based on GI and Self-Assessment Questionnaire (SAQ) features.

A2) Classification
Classify suppliers as Best in class (BIC) or Room for improvement (RFI) based on their sustainability scores (e.g., using a threshold OS $\geq$ threshold).

4.3 Baseline vs Tested Models

Baseline Models: Baseline models need to be simple and self-explanatory. For regression tasks, a trivial baseline can be the average (mean value) of sustainability scores. For classification tasks, majority class might be used.

Tested Models: Logistic/Linear Regression (with and without regularization), and at least one tree-based model (e.g., Gradient Boosting, XGBoost, LightGBM, CatBoost, Random Forests or simple decision tree).

4.4 Evaluation

- **Regression:** RMSE and/or MAE and compare against baseline.

- **Classification:** ROC-AUC and accuracy against baseline.

4.5 Interpretability

Global explanations: Identify the most important features (coef. magnitudes, permutation feature importance or SHAP values) — list top 10 drivers and discuss plausibility with managerial interpretation.

For GI+SAQ model: Use SHAP analysis to reveal which features matter most. Interpret Q-codes using the topic mapping provided (e.g., Q1272 → Environment - Corrective Action)

Local explanations: Pick 2 suppliers (one low-score, one high-score) and show local explanation (e.g., apply LIME technique or SHAP to produce waterfall graphs) for their predicted scores.

4.6 Decision Memo — Thresholding

- Propose a risk segmentation (e.g., 5 tiers: Critical, High, Medium, Moderate, Low Risk) or thresholding rule to flag suppliers for further engagement (no optimization step).
- Include a short fairness/guardrail check (e.g., FNR across regions) (e.g., false negative rate, false positive rate, performance across regions/countries). Document any potential biases and mitigation strategies.

4.7 (Optional) Improvement Prediction

- If multiple assessments exist per supplier, model $\Delta\mathrm{OS} = \mathrm{OS}_{(n+1)} - \mathrm{OS}_{(n)}$ using Elastic Net; report RMSE and key features.

4.8 (Optional - Advanced) Hyperparameter Tuning

- Conduct grid search or random search on your best model
- Report performance improvement from tuning
- Discuss: Is tuning worth the computational cost?

5 Deliverables

1. 4–5 page brief with: approach, two-stage results (GI-only vs GI+SAQ), top drivers, calibrated decision rule, and risks/mitigations and business interpretation with executive summary.
2. Appendix with at least two artefacts to support the analysis or EDA section (one for GI-only, one for GI+SAQ if using for comparison).
3. A Jupyter Notebook (or Python script) with clear sections: data prep, GI-only modeling, GI+SAQ modeling, comparison, evaluation, interpretability, decision memo.

6 Evaluation

Your work will be assessed on:

- Data preparation & documentation
- Baselines & modeling correctness
- Metrics & validation
- Interpretability & insights
- Decision memo & overall clarity

7 Starter Guidance

- **Suggested target:** OS score column (Val_Score in the dataset).
- **Train/test split by supplier** (avoid leakage across assessments). Use sklearn's `GroupShuffleSplit` or manual splitting.
- Use `sklearn` pipelines; for SHAP use `shap` (TreeExplainer) or coefficient-based interpretation for linear models.

8 Academic Integrity & Ethics

Use the dataset for course purposes only. Be cautious with sensitive attributes. Document features that could proxy protected characteristics and how you mitigated risks (e.g., monotonic/feature constraints, review by human).

References

[1] Jiang, B. (2009). The effects of interorganizational governance on supplier's compliance with SCC: An empirical examination of compliant and non-compliant suppliers. *Journal of Operations Management*, 27(4), 267–280.

[2] Locke, R. M., Qin, F., & Brause, A. (2007). Does monitoring improve labor standards? Lessons from Nike. *Industrial and Labor Relations Review*, 61(1), 3–31.

[3] Philips. (2022). Supplier sustainability performance. Retrieved November 29, 2022, from https://www.philips.com/a-w/about/environmental-social-governance/environmental/supplier-sustainability/supplier-sustainability-performance.html

[4] Porteous, A. H., Rammohan, S. V., & Lee, H. L. (2015). Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties. *Production and Operations Management*, 24(9), 1402–1413.

[5] Stauffer, B. (2018). "Soon there won't be much to hide": Transparency in the apparel industry. *Human Rights Watch*. Retrieved August 20, 2019, from https://www.hrw.org/world-report/2018/essay/transparency-in-apparel-industry

[6] Tan, T., Akyüz, M. H., Urlu, B., & Ruiz, S. (2024). Stop auditing and start to CARE: Paradigm shift in assessing and improving supplier sustainability. *INFORMS Journal on Applied Analytics*, 54(3), 241–263.