

METHODOLOGY DOCUMENTATION

Likelihood to Buy (LTB)

A machine-learning score describing the probability of companies engaging with the voluntary carbon market

VERSION	DATE	TYPE	PREPARED BY
v1.0	September 2026	Machine Learning Model	AlliedOffsets Product Team

Overview

AlliedOffsets' Likelihood to Buy (LTB) describes the probability of companies engaging with the voluntary carbon market (VCM). It combines the 10,000 largest companies by market cap (per companiesmarketcap.com), each company's previous engagement with the VCM, its corporate climate commitments and obligations (SBTi, net-zero targets, CORSIA, B Corp status, and similar), and a predictive model of buyer activity. Each scored company receives a probability and is grouped into one of five bands: Very Low, Low, Medium, High or Very High. As of the latest run, 29,293 companies have been scored (this total excludes newly added companies not yet assessed).

Data Inputs

Field / Parameter	Description	Source / Notes
Company universe	The 10,000 largest companies globally by market capitalisation.	companiesmarketcap.com
Previous VCM engagement	Whether and when the company last had a tracked carbon credit transaction.	AlliedOffsets transaction data
Abatement potential	Whether the company's sector is classified as 'Easy to Abate' or otherwise, reflecting how readily the company can cut emissions through direct action versus needing offsets.	AlliedOffsets sector classification
Headquarter country / Buyer sector	The company's country of headquarters and industry sector, each of which is one-hot encoded into individual category features (e.g. 'Headquarter country_United States', 'Buyer sector_Airlines') before being fed to the model.	AlliedOffsets company profile data
Net-Zero Target (NZZ) interim and end targets, and removals/offset use	Whether the company has published interim and/or end net-zero targets, whether it has a removals target, and whether it has declared it will use carbon credit offsets as part of its strategy.	Company sustainability disclosures
SBTi near-term target status	Whether the company has set or committed to a Science Based Targets initiative near-term target.	Science Based Targets initiative (SBTi) database
Energy use	Reported energy consumption profile, including whether the company is a data-centre user.	Company sustainability disclosures
Internal Carbon Price	Whether the company has adopted an internal carbon price.	Company sustainability disclosures
Annual profit/loss and annual revenue (USD)	The company's most recently reported financial performance, used as an indicator of capacity to fund carbon offsetting.	Company financial filings



Field / Parameter	Description	Source / Notes
Employees	Total headcount, a general indicator of company size.	Company profile / third-party data providers
B Corp / GHGP membership	Whether the company holds B Corp certification or is a member of the Greenhouse Gas Protocol.	Company sustainability disclosures
Emissions (Scope 1 & 2, Scope 3, total)	The company's reported greenhouse gas emissions across each scope.	Company sustainability disclosures / emissions registries

Logical Flow

1 Build the company universe

Start from the 10,000 largest companies globally by market cap, and link this list to AlliedOffsets' buyer supplementary dataset, which records whether and when each company has previously engaged with the VCM.

2 Assemble company features

For every company, gather firmographic data (sector, headquarter country, size, financials), decarbonisation commitments (net-zero targets, SBTi status, CORSIA obligations, B Corp/GHGP membership) and abatement/energy-use characteristics.

3 Encode and balance the training data

Categorical fields such as buyer sector and headquarter country are one-hot encoded into individual binary features (e.g. 'Buyer sector_Airlines'). To ensure fairness, the training data is balanced by matching the number of known buyers and non-buyers before training.

4 Select informative features

A feature-selection process identifies the most informative predictors and removes variables that contribute little to predictive performance.

5 Train the tree-based model

A gradient-boosted decision tree model (XGBoost) is trained on companies with known VCM engagement history. Tree-based modelling was chosen specifically to capture non-linear relationships between company characteristics and buying behaviour.

6 Score and band every company

The trained model produces a probability score for each company in the universe, which is then grouped into one of five likelihood bands.

7 Apply manual overrides where justified

In at least one case (Airlines), the raw model output is manually adjusted to reflect known real-world context — see 'A note on Airlines' below.

8 Generate SHAP explanations

SHAP (SHapley Additive exPlanations) values are calculated to show how much each feature pushes an individual company's score up or down relative to the model's average prediction, powering the transparency views on the Likelihood to Buy Methodology page.

Key Calculations



Likelihood band thresholds	The model's output probability (p) is converted into a band using fixed thresholds: $p \geq 0.80 \rightarrow$ 'Very High'; $p \geq 0.60 \rightarrow$ 'High'; $p \geq 0.40 \rightarrow$ 'Medium'; $p \geq 0.20 \rightarrow$ 'Low'; $p < 0.20 \rightarrow$ 'Very Low'.
SHAP feature contribution	Each prediction can be reproduced as: model prediction = base value + sum of SHAP values, where the base value is the average prediction across the training data and each SHAP value shows how much one feature pushed an individual company's score away from that average. Feature importance is examined at two levels: 'base feature' level, which aggregates a variable's contribution across all its categories (e.g. all buyer sectors combined), and 'specific feature' level, which examines individual categories on their own (e.g. just 'Buyer sector_Airlines').
Global feature importance ranking	Averaged across all scored companies, Abatement potential and Headquarter country are the two strongest overall drivers of the score. Abatement potential is closely correlated with Buyer sector, whose aggregated importance looks smaller only because individual sector categories push the score in different directions and partly cancel out when combined. NZT end target, SBTi near-term target status and NZT removals target are the next most influential group, followed by governance/financial-capacity indicators such as internal carbon price, profit/loss, GHGP membership and B Corp status.

Model Insights

Companies by Likelihood to Buy Band (n = 29,293)

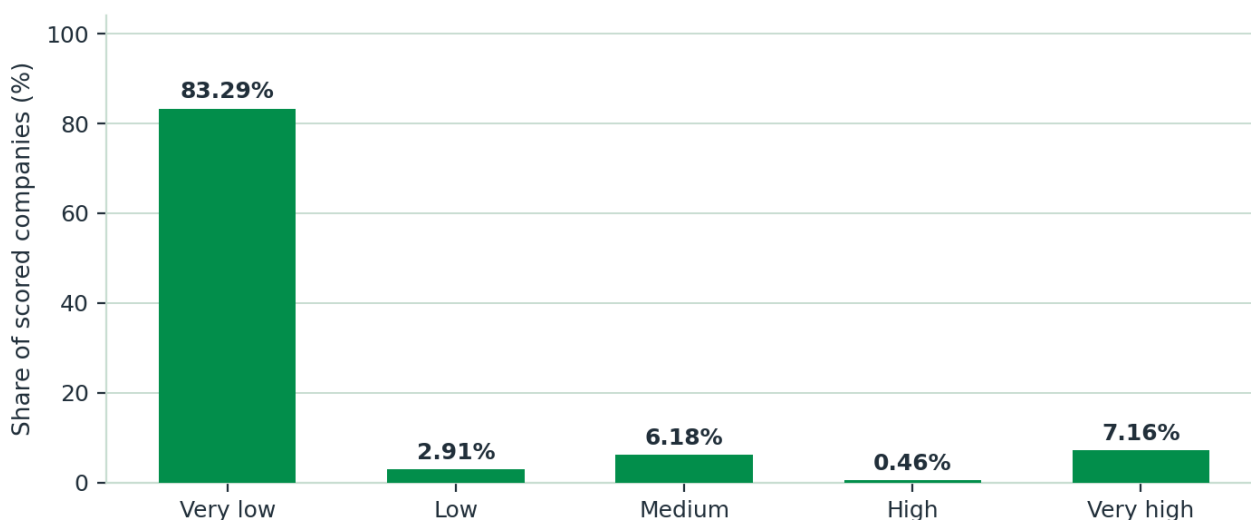


Figure 1. Official distribution of the 29,293 currently scored companies across the five Likelihood to Buy bands.

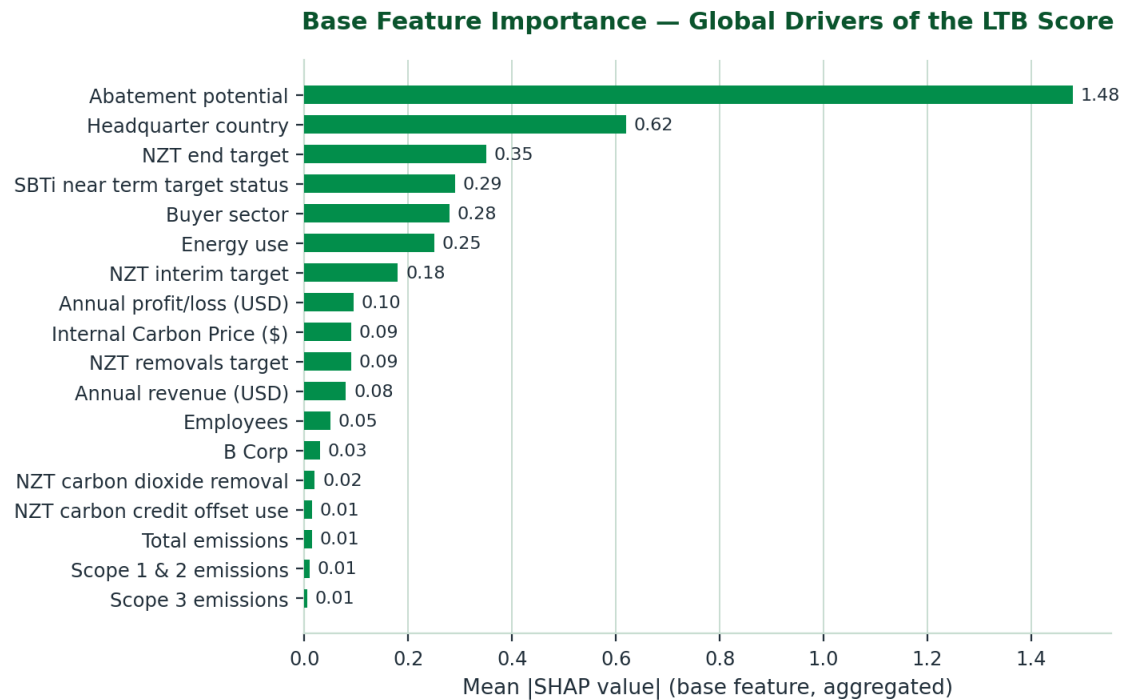


Figure 2. Base feature importance — mean absolute SHAP value for each variable, aggregated across all its categories. Abatement potential and Headquarter country are the strongest overall drivers.

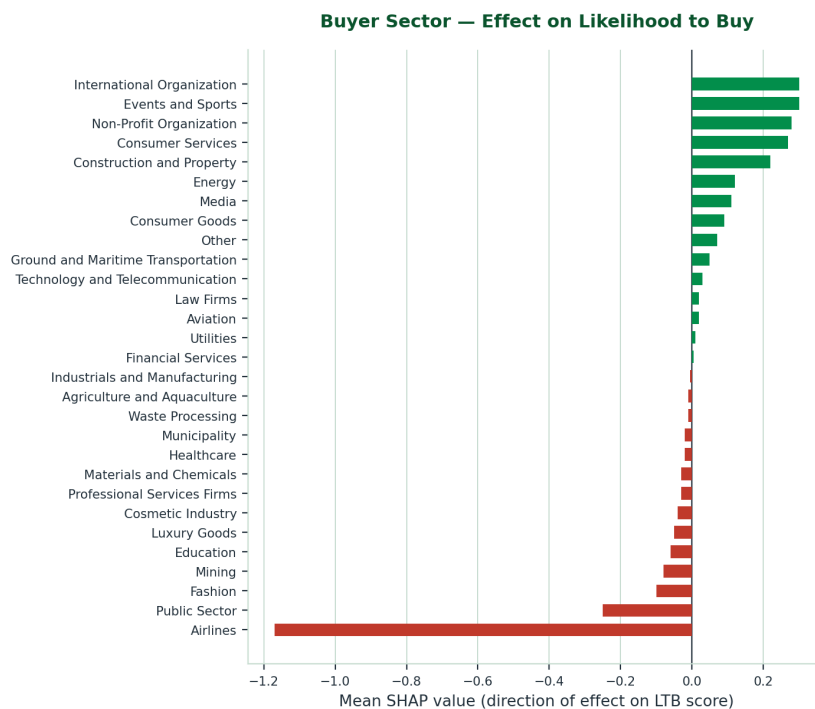


Figure 3. Buyer sector effect on the LTB score. International/Non-Profit Organizations, Events and Sports, and Consumer Services push scores up; Public Sector and, most strongly, Airlines push scores down — see the note on Airlines below for why this is manually overridden in practice.

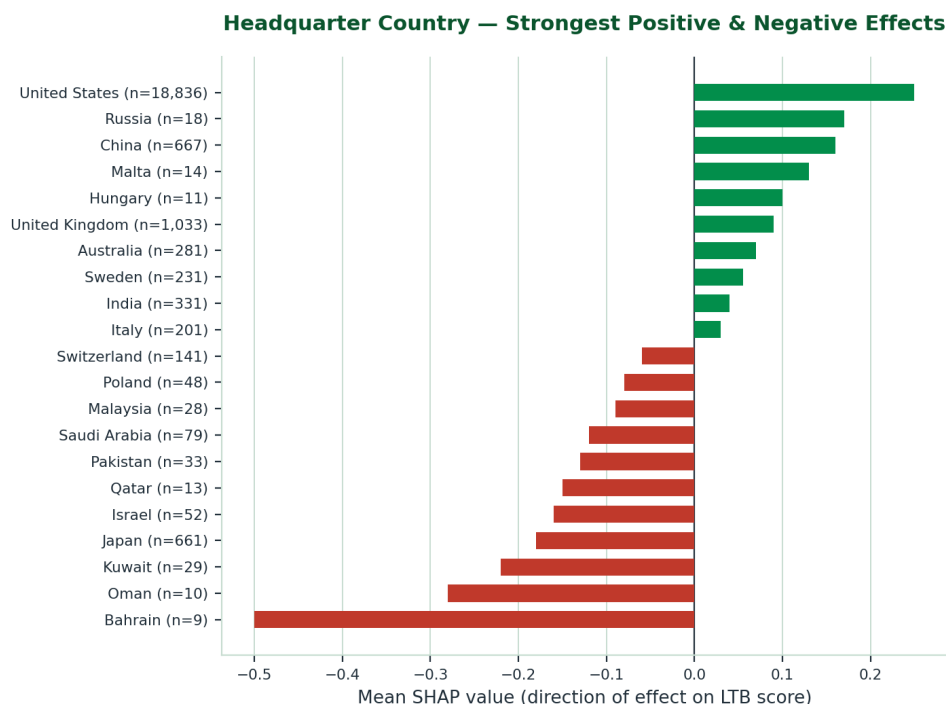


Figure 4. Strongest positive and negative headquarter-country effects on the LTB score. US-headquartered companies see the largest positive lift; companies headquartered in Bahrain and Oman see the largest negative effect. Sample size (n) is shown per country.

Assumptions

- The model is trained only on the 10,000 largest companies by market cap (companiesmarketcap.com); companies outside that universe are not scored.
- Training data is rebalanced to match the number of known buyers and non-buyers before training, which affects how strongly the model weights each feature relative to a naturally imbalanced dataset.
- The reported total of 29,293 scored companies is lower than the full company universe because newly added companies that have not yet been assessed are excluded.
- Categorical variables (buyer sector, headquarter country, etc.) are one-hot encoded, so a company's score reflects the specific category it falls into rather than a continuous scale.
- Airlines are a known, deliberate exception to the raw model output: the model naturally scores the sector low due to limited historical buyer data, but airlines are manually reclassified into the High likelihood category in the company search table to reflect their CORSIA offsetting obligations. Any client-facing use of Airline-sector scores should account for this override.
- Dashboard content (feature importance figures, banding, country/sector effects) reflects the version live as of February 2026 (per the 'Latest update' note on the Headquarter Country chart); the model is periodically retrained and these figures may shift over time.

Limitations

- The model estimates likelihood from observable company characteristics and disclosed commitments; it cannot detect buying intent that has not yet been publicly disclosed.



- Because the model is trained on historical buyer/non-buyer patterns, sectors or regions with limited past VCM activity (such as Airlines, prior to the manual override) can be under-scored relative to their real-world potential — this is why manual overrides exist and should be checked for other similarly under-represented segments.
- SHAP-based explanations describe correlation with the model's prediction, not a proven causal relationship between a feature and a company's actual intent to buy.
- Global (base) feature importance reflects average effects across the whole company universe; the drivers of any individual company's score can differ substantially from this aggregate ranking.
- Scores are refreshed periodically rather than in real time, so a company's band may lag its most recent disclosures until the next model run.

Output

For each company, the model returns an adjusted probability (0-100%) of VCM engagement and a Likelihood Band (Very Low / Low / Medium / High / Very High) derived from that probability using the fixed thresholds above.