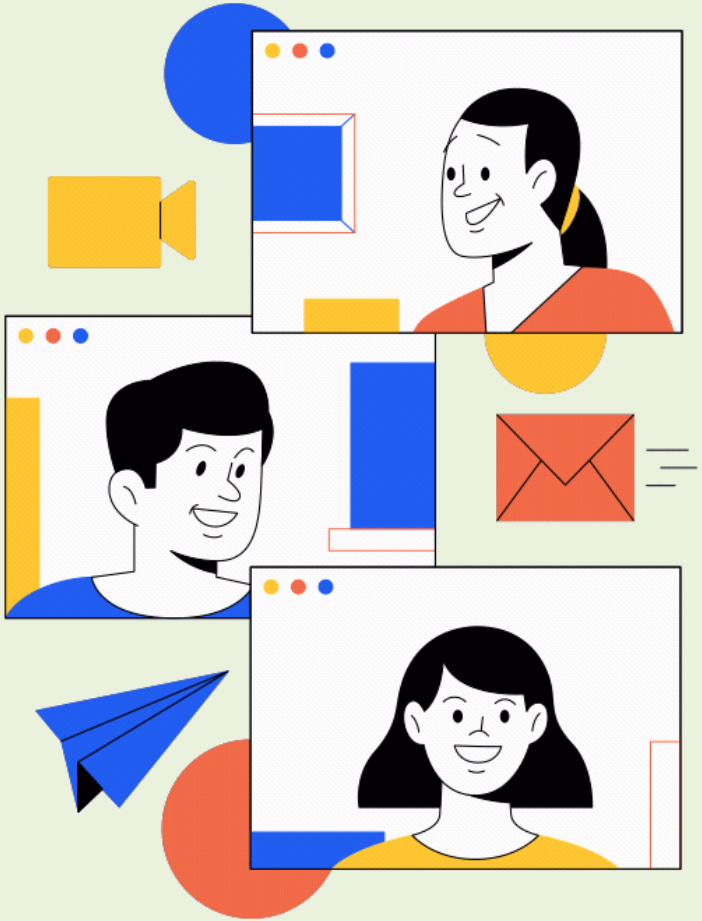
Profitability Prediction through ESG & Sustainability Metrics

Presented By:
Pavithra Lakshmi Venugopal
FD0003363



INTRODUCTION

Problem Statement

Main Problem :

✓ Can ESG and sustainability factors reliably predict company profitability independent of financial metrics?

Sub - Problems Addressed:

✓ How to build models that avoid financial data leakage and remain robust after removing direct profitability features?

✓ How to ensure fair, calibrated, and interpretable predictions while identifying which ESG dimensions drive profitability most?

Target Variable: Profitability

Dataset

01 1,000 companies, 9 industries and 7 regions

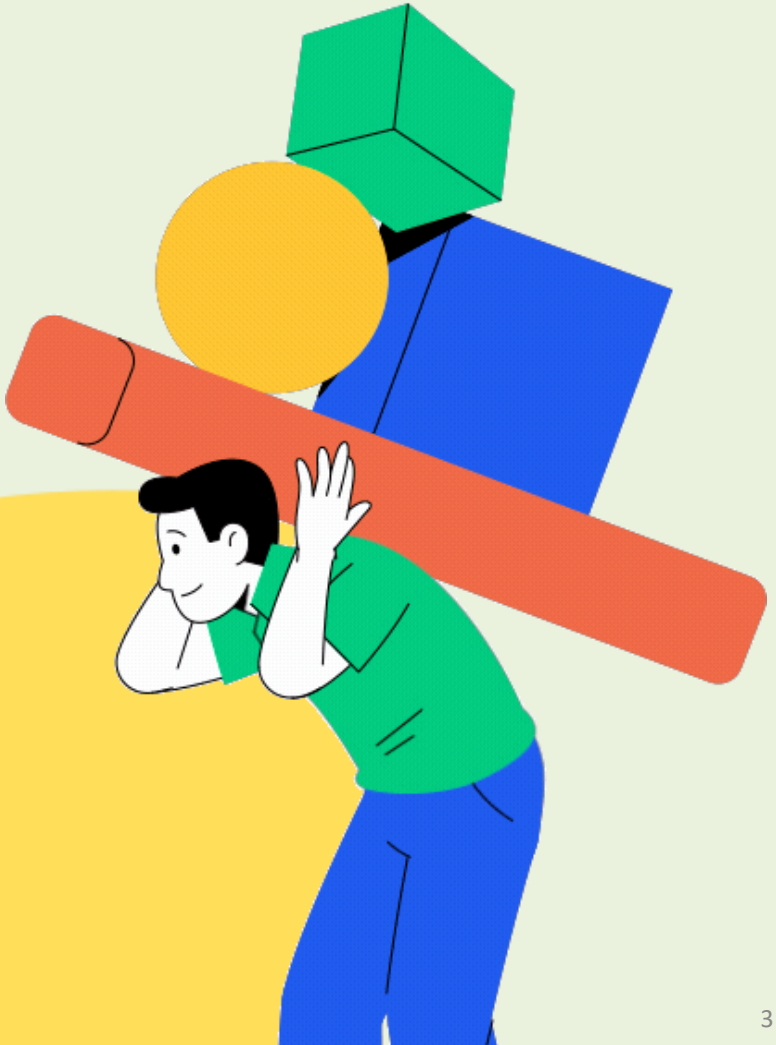
02 ESG scores (Environmental, Social, Governance)

03 Sustainability metrics (Energy, Water, Carbon)

04 Financials (Revenue, MarketCap, ProfitMargin)

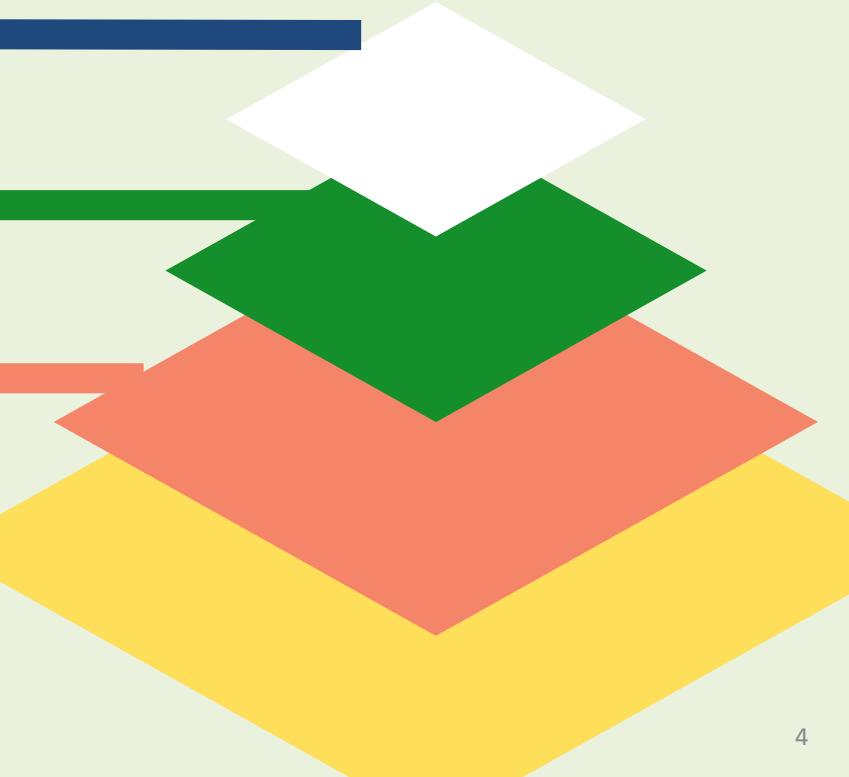
05 GrowthRate: about 9% missing values were detected

Newly added Column: Profitability derived from ProfitMargin.



Methodology

1. Models: Decision Tree, Random Forest, Logistic Regression.
   Train/Test – 80/20
2. Balanced data with SMOTE + class weighting
3. Applied feature reduction to remove leakage.
4. Calibration for reliable probabilities..



Overfitting & Leakage Challenge

01

Initial results: 100% accuracy → too good to be true.

02

Root cause: ProfitMargin, Revenue, MarketCap = leakage.

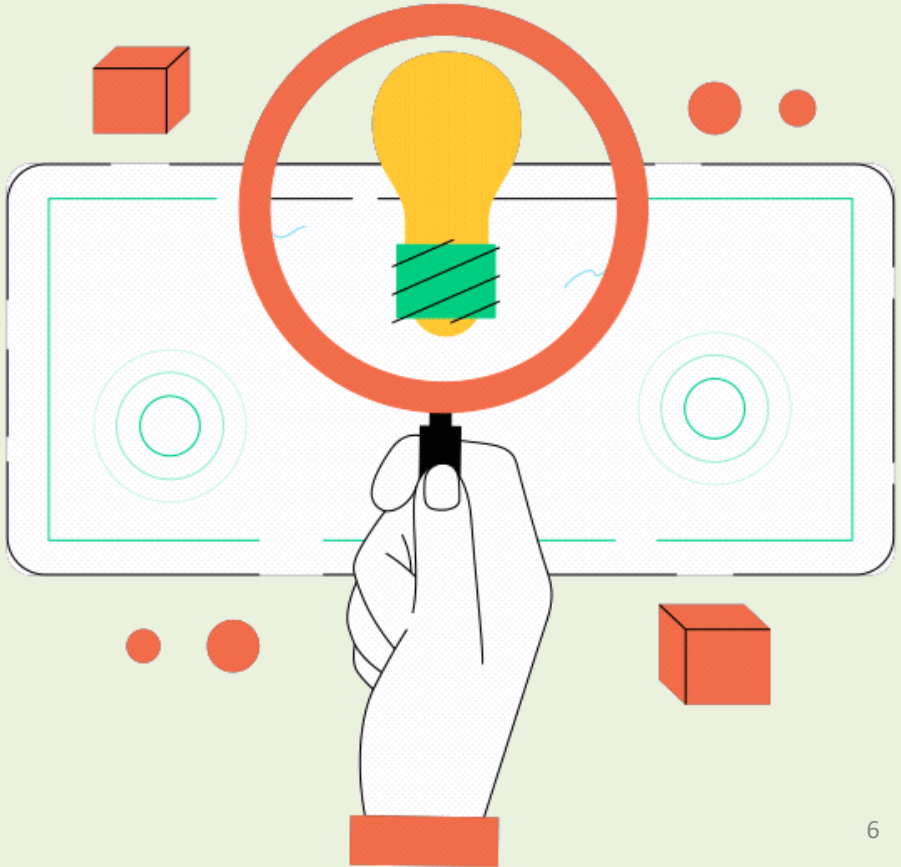
03

Solution: removed them → performance dropped, but became realistic.

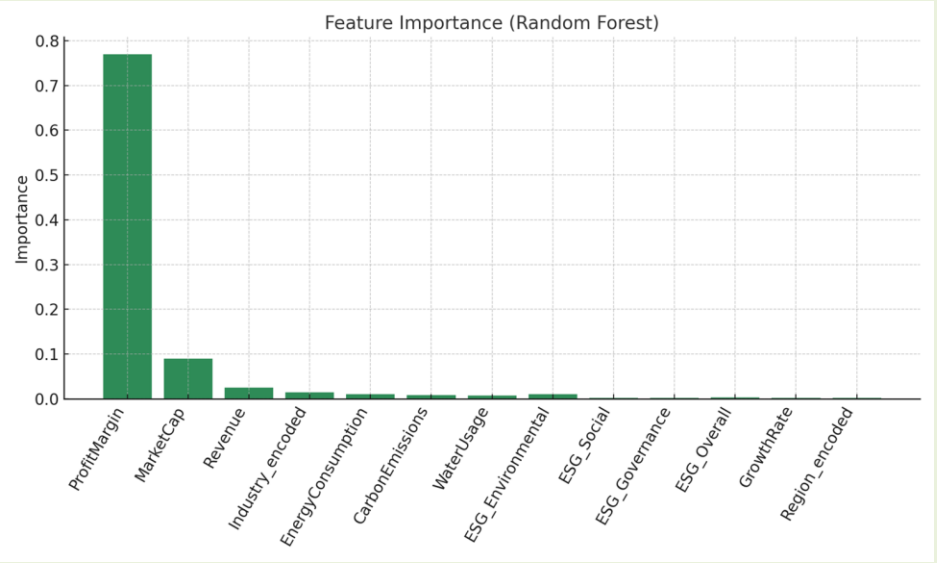
.

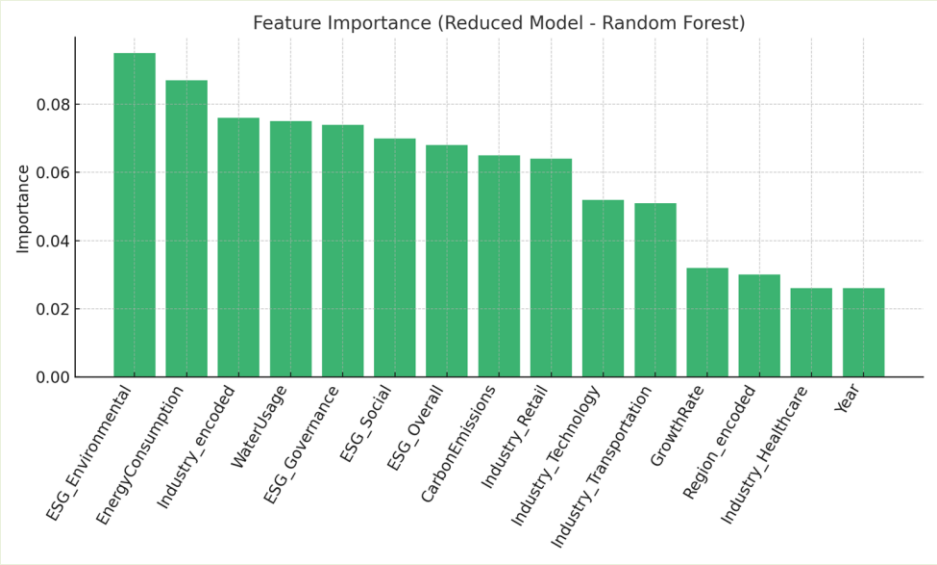
Baseline vs Reduced Models

- ✓ Financial leakage removed: ProfitMargin, Revenue, MarketCap.
- ✓ Decision Tree: ~57% | Logistic Regression: ~55% | Random Forest: ~65%.
- ✓ ESG-only features remain predictive.
- ✓ More realistic and scientifically sound results.



Baseline vs Reduced Models





Feature Importance before reduction

vs

Feature Importance after reduction

Calibration Impact

Calibration = process of adjusting a model so predicted probabilities reflect true outcome frequencies..

Pre - calibration: ROC - AUC ~0.27, Gini negative → unreliable.

01

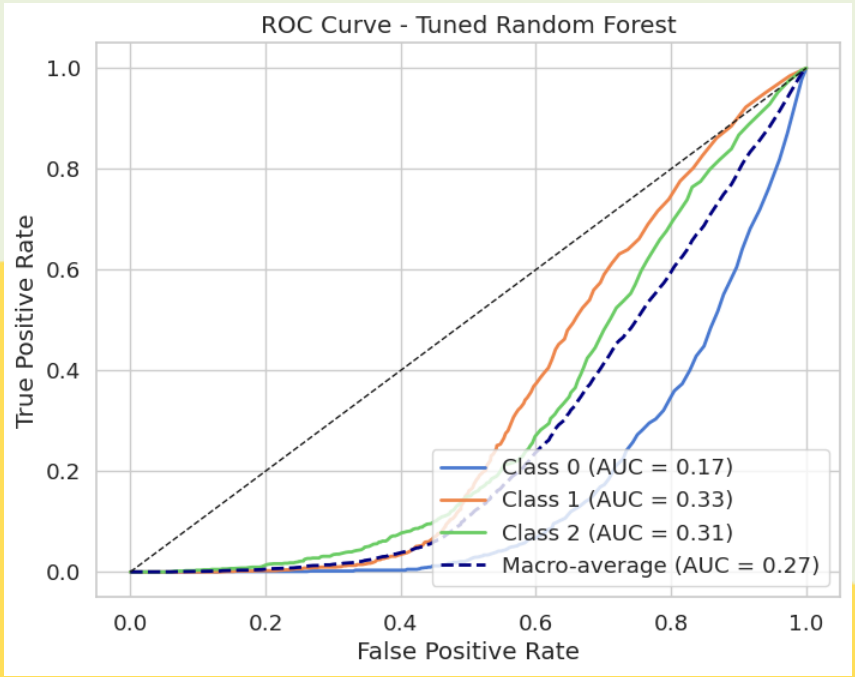
Post - calibration: ROC - AUC ~0.96, Gini ~0.92 → excellent.

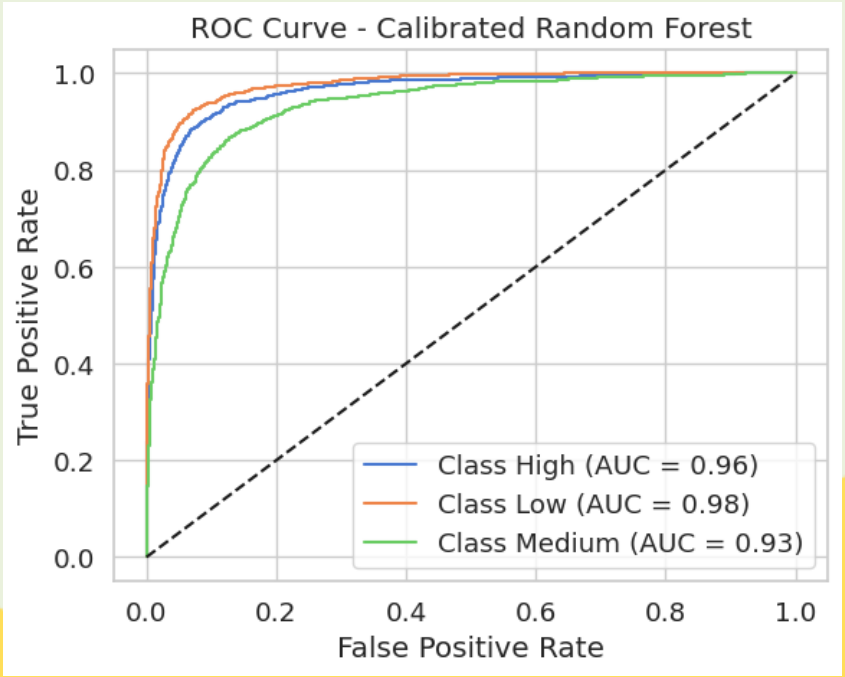
02

Calibration made probabilities trustworthy.

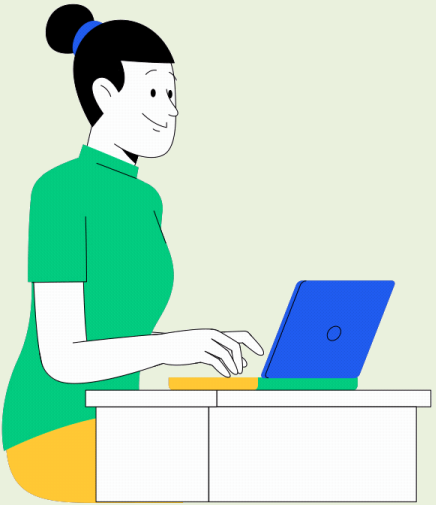
03

Calibration Impact





ROC Curve before calibration (AUC ~0.27) vs ROC Curve after calibration (AUC ~0.96)



Advanced Metrics

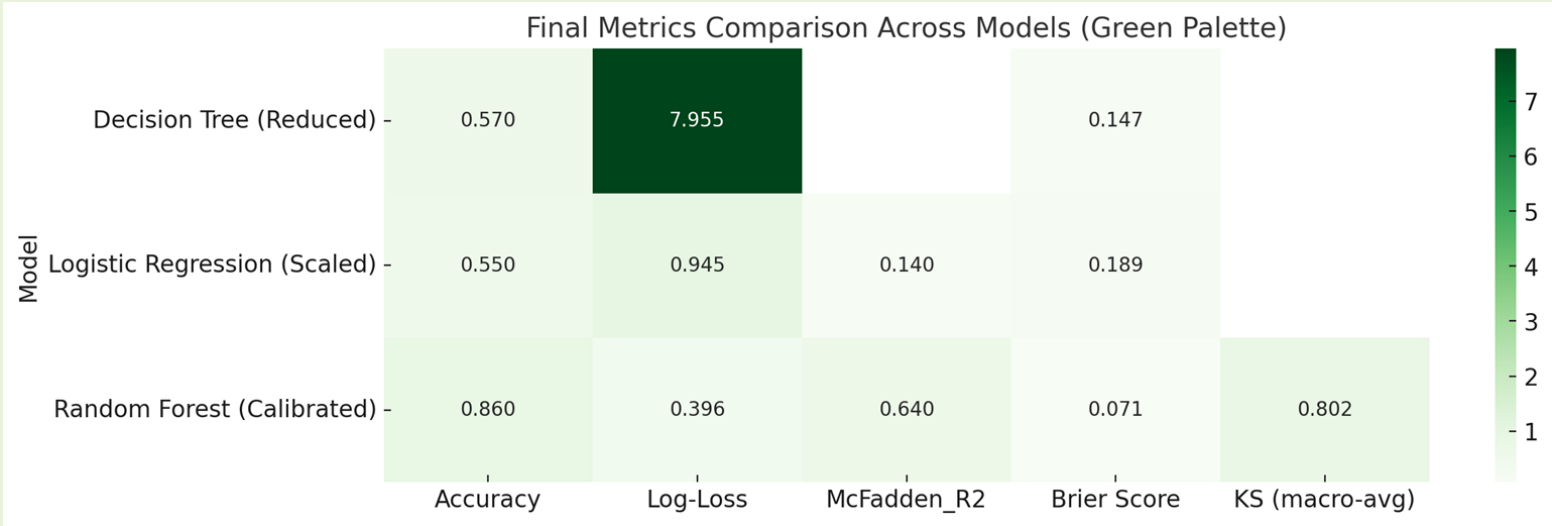
✓Random Forest (Calibrated) consistently dominates across metrics.

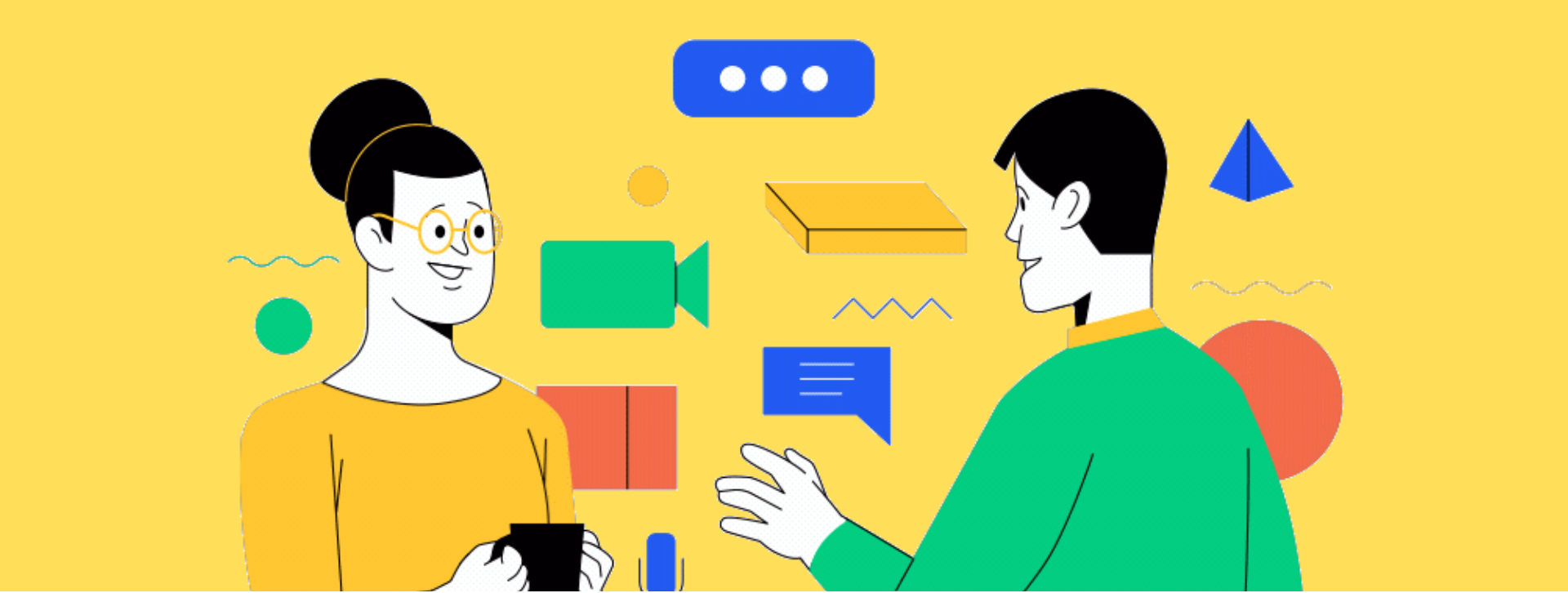
✓Log-loss → how wrong the probability predictions are (lower is better).

✓McFadden's R^2 → model fit quality (higher is better).

✓Brier Score → accuracy of predicted probabilities (lower is better).

✓KS statistic → how well the model separates classes (higher is better).





Logistic Regression Equation: logit(P(Profitability)) = 0.0000 + (0.0126 * ESG_Environmental) + (0.0037 * ESG_Overall) + (0.0017 * GrowthRate) + (0.0006 * Year) - (0.0012 * ESG_Governance) - (0.0004 * ESG_Social) + (Industry/Region Effects)

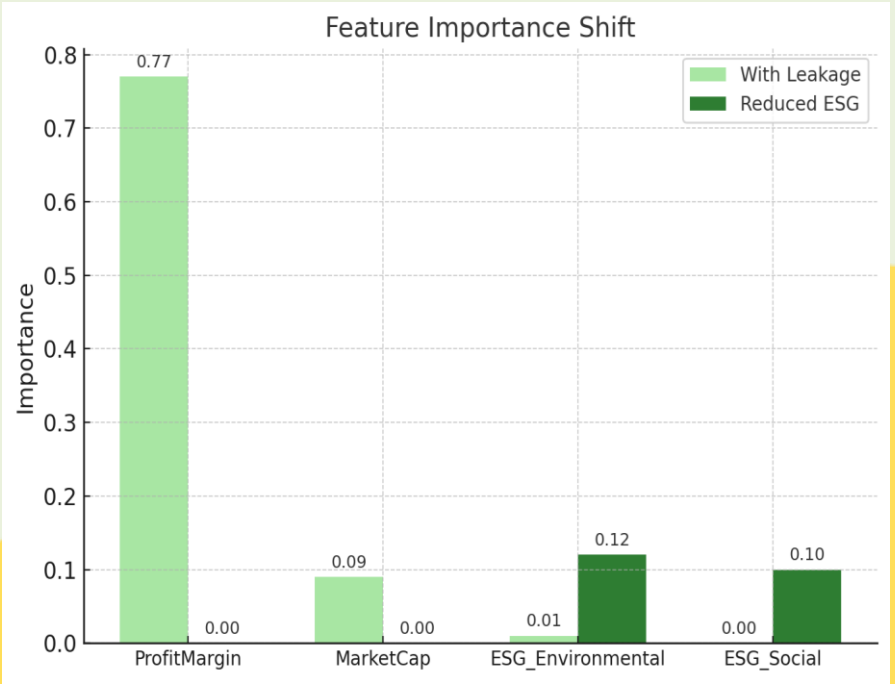
Key Findings

01 ESG-only models achieved 73% accuracy strong predictive power without financials.

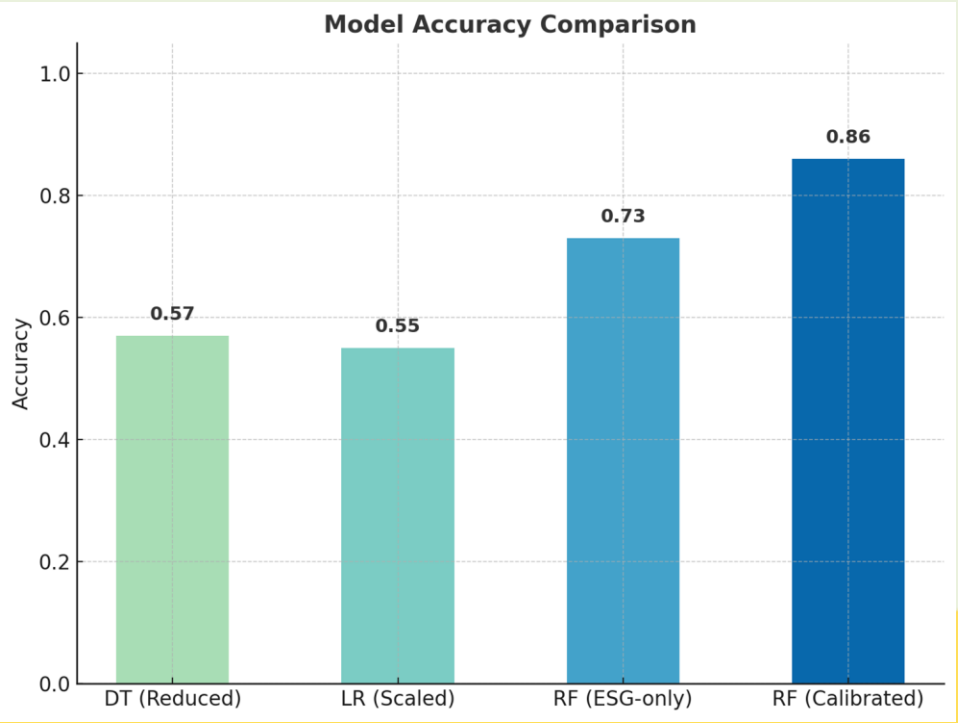
02 Calibrated Random Forest delivered 86% accuracy, ROC-AUC 0.96, KS ~0.80.

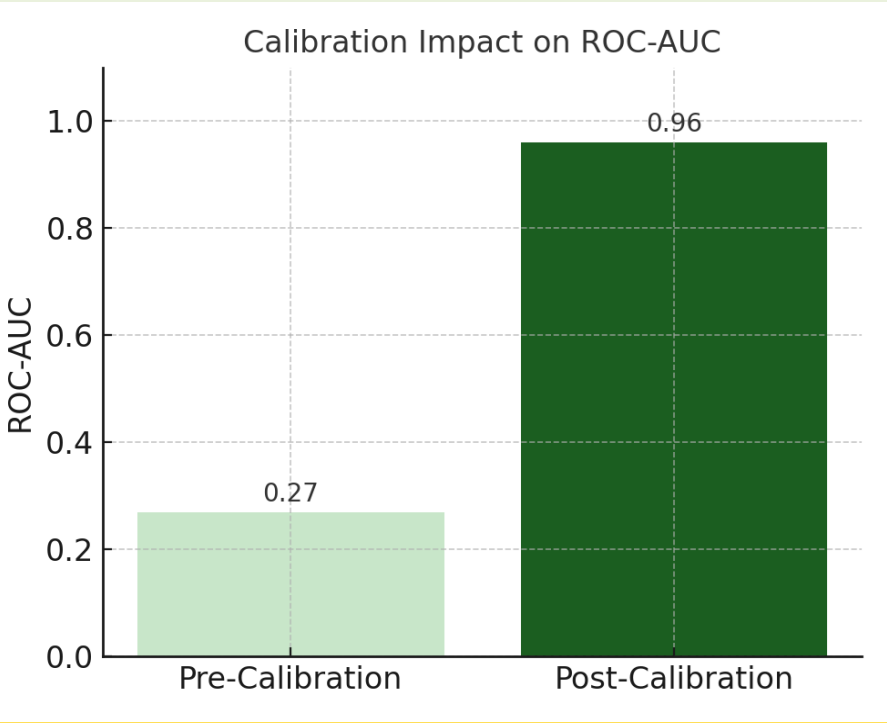
03 Environmental and Social factors were the most influential ESG drivers.

04 Calibration transformed probability outputs from unreliable (AUC ~0.27) to business-ready (AUC ~0.96).



Key Findings





Other Projects / Papers

- ✓ Combined ESG + financial variables (Profit, MarketCap, Revenue).
- ✓ Reported accuracy or R^2 but not probability calibration.
- ✓ Often let financials dominate → ESG contribution unclear.
- ✓ Focused more on predicting ESG scores or financial returns.
- ✓ Limited analysis of overfitting or leakage.

My Approach

- ✓ Removed financial leakage → ensured ESG alone predicts profitability.
- ✓ Used advanced metrics (Log-loss, Brier Score, KS) beyond accuracy.
- ✓ Applied calibration → probabilities became realistic & business-ready.
- ✓ Identified key ESG drivers (Environmental, Social most influential).
- ✓ Explained overfitting (100% accuracy issue) and corrected it.
- ✓ Delivered both accuracy (86%) and reliable probabilities (AUC ~0.96).

Sample Predictions

Sample Predictions – Calibrated Random Forest

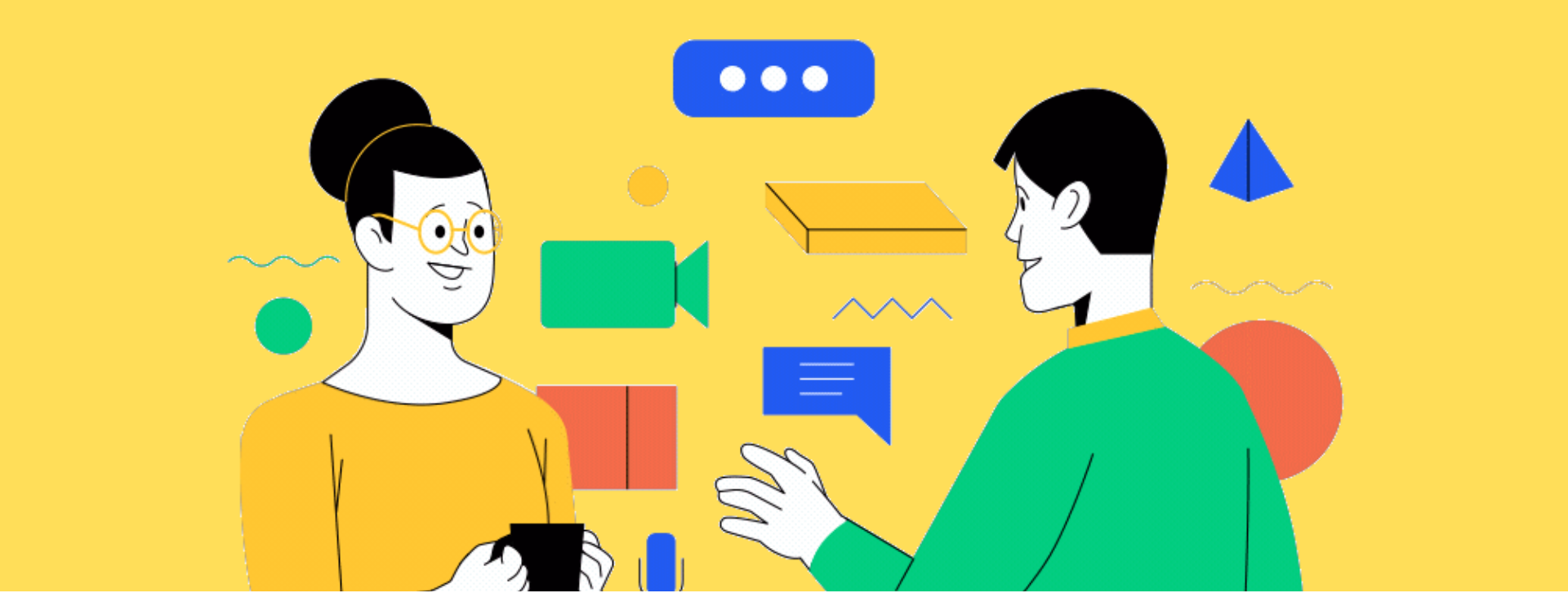
Actual Class	Predicted Class	Prob (High)	Prob (Medium)	Prob (Low)
High	High	0.87	0.09	0.04
Low	Low	0.06	0.08	0.86
Medium	Medium	0.18	0.74	0.08
High	High	0.91	0.06	0.03
Low	Low	0.07	0.12	0.81

✓ This table shows example predictions with their probabilities.

✓ Probabilities are well-calibrated, making the model trustworthy for business use.

Application & Impact

This project transforms ESG factors into a reliable profitability prediction tool, enabling data-driven decisions for investors, companies, and policymakers



Thank you!