

DATA ANALYST

MIN SUN-AH

PORTFOLIO

회계 실무 5.5년의 도메인 지식과 데이터 과학의 고도화된 만남.
수치 너머의 맥락을 읽어 비즈니스 가치를 증명합니다.

Profile & Core Competencies

Accounting Domain Specialist

총 경력 5년 6개월: 회계/자금 실무 전문가에서 데이터 분석가로의 도약

(주)지지푸드 (관리부

사원)

ERP WEHAGO 활용 자체 기장, 부가세 신고, 월별 손익보고서 작성.

Notion/ChatGPT 기반 출고 일정 및 지출결의 자동화 구현.

(주)인성저축은행 (총무부

사원)

저축은행 전용 ERP 운용, 법인카드 및 비품 재고 관리. 2년간 반기별 외부

감사 대응 및 임원 비서 실무 수행.

Skills &

Education

영광홍익대 경영학 학사(Minor 일본어/JLPT N2). ADsP, AICE ASSOCIATE,

재경관리사, 펀드투자상담사 등 10여개 자격 보유.



Hybrid Value Proposal

- ✓ **수치 무결성:** 5.5년 결산 경험 기반의 데이터 검증
- ✓ **연락 사고:** 비즈니스 로직에 기반한 통계적 가설 수립
- ✓ **실무 자동화:** 생성형 AI 및 협업 툴을 통한 업무 혁신

Technical Competency & Tech Stack



Data Engineering

Python(Pandas, NumPy) 기반 정밀 전처리 및 EDA. Scikit-learn, XGBoost 머신러닝 파이프라인 구축.



AI Innovation

ChatGPT 프롬프트 엔지니어링 및 Notion AI 활용 실무 자동화. 새로운 도구에 대한 빠른 적응력.



Finance Domain

기업 재무제표 및 ESG 공시 체계 분석. 재경관리사 기반의 고도화된 비즈니스 데이터 해석.

PROJECT CASE STUDY

기업 **ESG** 통합 등급 예측 모델

머신러닝을 활용한 기업의 비재무적 가치 정량 추론 프로젝트

01. Research Design & Data Collection

Objectives

연구 배경: 선행 연구(2024)의 ESG 예측 성능($R^2=0.225$)은 실무 활용 임계치에 미달.

연구 목표: 정량적 재무 지표를 활용하여 한국 ESG 기준원 등급을 80% 이상의 정확도로 예측하는 분류 모델 구축.

Dataset Expansion: 기존 연구 대비 평가 기업을 635개에서 **731개**로 확장하여 연구의 객관성 및 일반화 성능 확보.

Feature Selection (X_features)

Group	Primary Variables
Financial	A_SIZE (Asset), LEV (Leverage), ROA (Return)
Market	TQ (Tobin's Q), SGR (Growth Rate)
Structure	FOR (Foreign Ownership), MSE (Major Equity)
Intangible	R&D Intensity, Advertising Intensity (ADV)

02. Base Modeling: Regression Results

Regression Metrics

MAE

1.0489

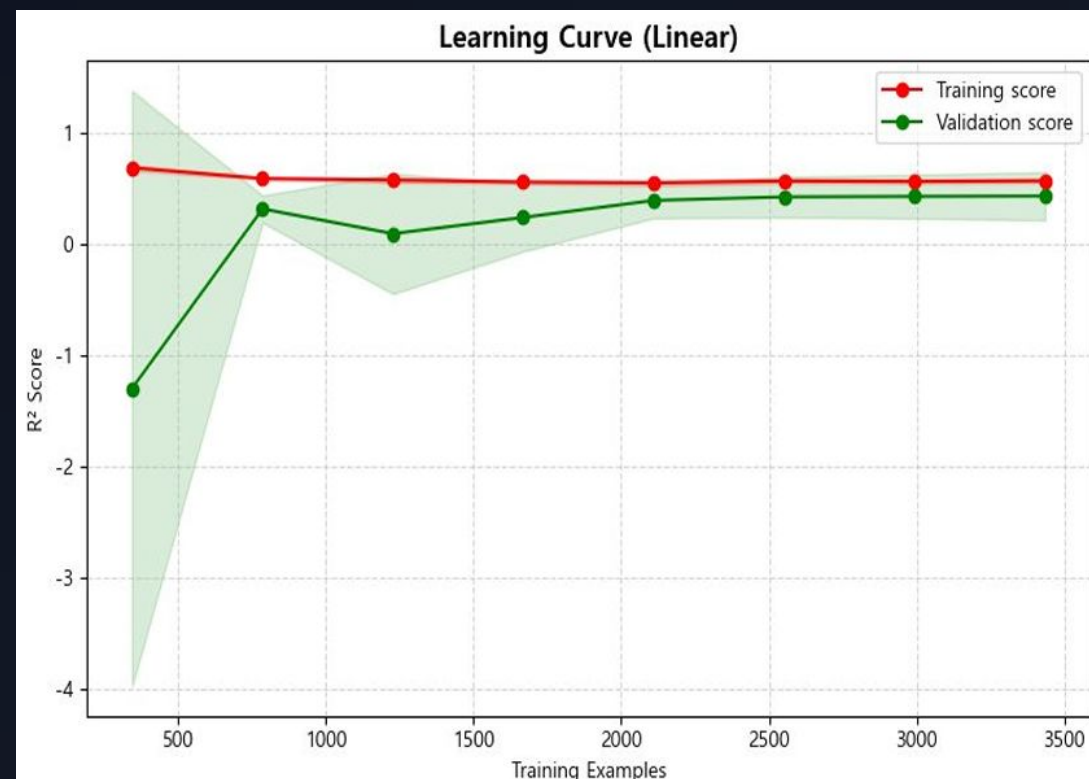
RMSE

1.2336

Explanatory Power (R^2)

0.4484

성과: 선행 연구 대비 설명력을 약 2배 향상시켰으나, ESG 등급의 비연속적 특성과 정성적 요소를 선형 회귀로 설명하는 데 **구조적 한계** 발견.



03. Paradigm Shift: Transition to Classification

Bottleneck Found

회귀 모델은 연속적인 등급 변화를 포착하기에 부적합하며, 실무적인 등급 판별(Classification) 능력이 부족하다는 결론에 도달.

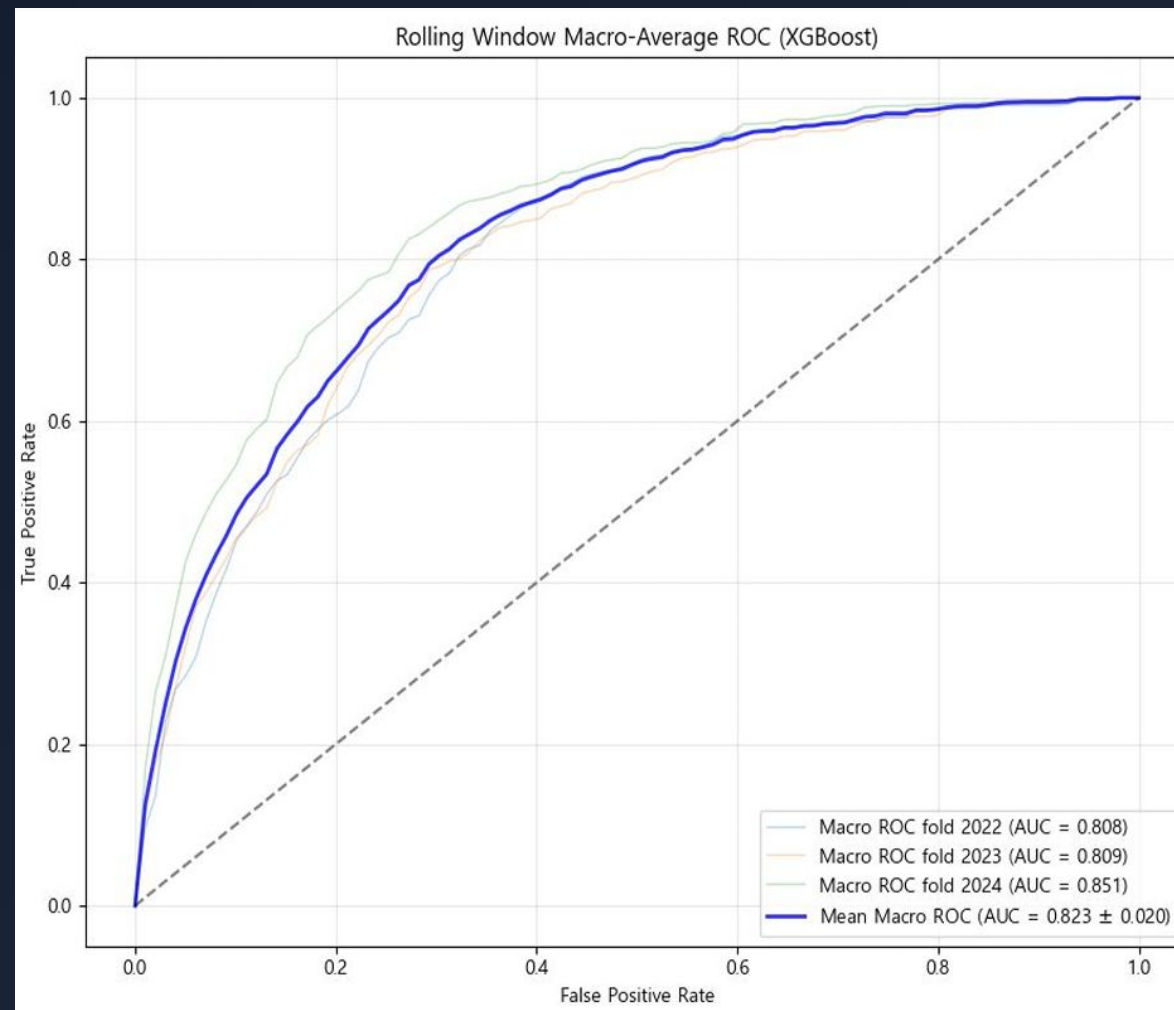
Classification Strategy

- 7개 등급 체계의 **다중 분류(Multiclass)** 모델로 전환
- 알고리즘별 성능 비교를 통한 최적 Base 모델 선정
- 시계열 특성을 반영한 **Window Rolling** 검증 수행

5-Model Performance Comparison (Baseline)

Logistic Regression	0.796
Random Forest	0.804
SVM	0.807
LightGBM	0.824
XGBoost (Selected)	0.827

XGBoost Window Rolling 결과: **Macro ROC AUC 0.823** 확보



04. Strategy: Feature Optimization (Contribution)

Baseline 성능(0.823)을 돌파하기 위해 비즈니스 맥락을 반영한 **2가지 핵심 신규 변수**를 직접 설계하여 추가했습니다.

ESG_lag (관성 변수)

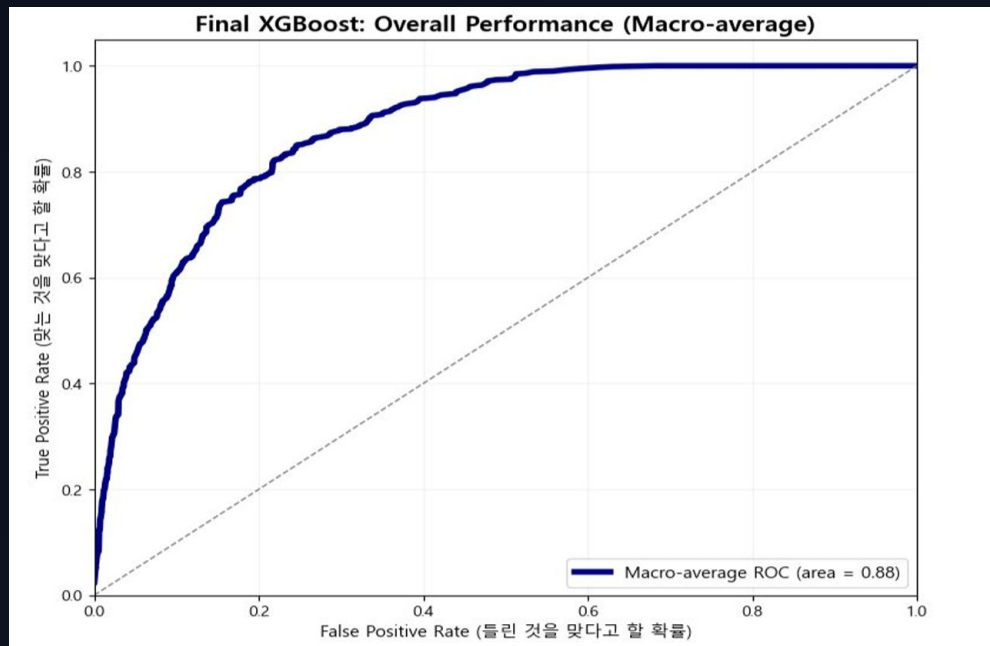
기업 경영의 **일관성과 관성**을 반영. 전년도 ESG 등급 점수를 피처로 활용하여 모델이 시계열적 등급 유지 패턴을 학습하도록 설계.

A_SIZE_FOR_inter (상호작용)

기업 규모와 외국인 지분율의 **상호작용 변수**. 글로벌 자본 유치가 대형 기업의 ESG 공시 체계 강화에 미치는 시너지 효과를 수치화.

"도메인 지식 기반의 피처 엔지니어링이 모델 성능 한계 돌파의 유일한 해법"

05. Final Optimized Model Results



AUC 0.88

신규 변수 도입 후 성능 쿼텀 점프 달성

Final Performance Analysis

최고 성능 달성: Baseline(0.823) 대비 약 7% 성능 향상을 이뤄내며 도달.

SHAP 기여도 확인: 설계한 `A_SIZE_FOR_inter`와 `ESG_lag` 변수가 모델 예측의 최상위 기여 인자로 확인됨.

| 06. Results & Business Application



ESG Value Discovery

재무 지표만으로 기업의 비재무적 등급을 88% 수준으로 정밀 예측. 정보 비대칭 해소 및 진정한 기업 가치 발견.



Investment Support

ESG 등급 공시 전 사전 등급 예측을 통한 투자 포트폴리오 리스크 선제 관리 및 의사결정 지원.



Corporate Consulting

예측 모델을 활용하여 등급 하락 요인을 진단하고, 기업 맞춤형 ESG 개선 전략 및 로드맵 제시 가능.

감사합니다

데이터 분석가 민선아

숫자 너머의 맥락을 읽는 통찰력으로 비즈니스의 성장을 돕겠습니다.



010-7175-7319



sunah_1024@naver.com