DATA ANALYST PORTFOLIO

Searching for Value through Data Analysis
데이터 분석을 통해 가치를 찾아냅니다

PARK HYUNWOO (박현우)

Data Analyst Candidate

Profile & History

Education

삼성SDS - 삼성국민취업아카데미
Samsung Academy
AIoT를 이용한 빅데이터 분석 산업솔루션 과정

동양대학교
Dongyang University
철도경영학과 / 경영·물류

Experience

대한민국 육군 장교 (8 Years)
ROK Army Officer (Captain)조직 관리,
리더십, 책임감 배양
Leadership & Responsibility

Core Competencies

Local Insight
Incheon Local Resident
영종도 거주 (지역 이해도 高)

Global Comm.
JLPT N1 (Japanese)
글로벌 고객 데이터 분석 역량

Technical Skills

Data Analysis & ML

Data Preprocessing **Pandas, NumPy**
데이터 전처리 능숙

Machine Learning **Scikit-learn**
머신러닝 모델링 (Regression)

Management & Viz

Database Management **MariaDB, SQL**
복잡한 Join 쿼리 및 DB 구조 이해

Visualization **Matplotlib, Seaborn**
통계적 인사이트 시각화 구현

PORTFOLIO PROJECT

재무 정보를 활용한 기업 ESG 등급 예측 모델 제작

Predicting Corporate ESG Ratings Using Financial Information

목차 (Index)

01. 개요 (Overview)

핵심 목표

한국 ESG 기준원의 평가 등급을 바탕으로 재무 지표와 연계한 **등급 예측 모델 제작**.

* 비 재무적 요소인 ESG를 재무 지표를 통해 역으로 추론

주제 선정 이유

- **기존 연구의 한계점 보완:** 선행연구의 낮은 설명력 (R^2 0.225)을 개선하여 발전된 모델 구축 * 이미지 참조
- **데이터 기반 입증:** 다양한 파생변수를 통해 재무 지표가 기업 가치(ESG)에 미치는 영향을 정량적으로 입증하고자 함.

<Table 3> Model A Multi - Regression Result

	model_A	model_A(E)	model_A(S)	model_A(G)	model_A(ESG)
Intercept	2.361***	2.176***	2.322***	2.087***	2.168***
ROA	0.263***	0.277***	0.244***	0.291***	0.268***
MASR	-0.002**	-0.002***	-0.003***	-0.003***	-0.003***
RD	1.865	1.884	1.820	1.845	1.834
SIZE	-0.101***	-0.085***	-0.102***	-0.067***	-0.084***
ADV	-0.001	-0.002	-0.002	-0.003	-0.003
SGRW	0.000	0.000	0.000	0.000	0.000
LEV	0.000	0.000	0.000	0.000	0.000
FOR	0.008***	0.008***	0.008***	0.008***	0.008***
E	0.013	0.03**			
S	0.064		0.066***		
G	-0.049*			-0.022	
ESG	0.017				0.034*
F-statistic	35.47***	37.83***	38.19***	37.69***	37.74***
MAE	0.469	0.469	0.469	0.471	0.470
MSE	0.759	0.762	0.760	0.763	0.762
RMSE	0.871	0.873	0.872	0.873	0.873
R^2	0.228	0.225	0.227	0.225	0.225

Signif. codes: p<.10, p<.05, p<.01

< 이재영, 차우창(2024) “머신러닝 모델을 활용한 ESG 활동과 기업 가치 분석”, 한국산업경영시스템학회지 47(4), 76-86.>

02. 데이터 수집 및 전처리 (EDA)

기존 연구 대비 평가 대상 기업 **635**개에서 **731**개로 확장하여 연구 진행

1

■ ESG 등급 체계

- ESG 평가 등급은 환경, 사회, 일반상장사 지배구조, 금융사 지배구조 영역별 등급과 ESG 통합 등급이 부여됩니다.
- 등급은 S등급부터 D등급까지 총 7개 등급으로 분류되며, 절대평가로 등급별 점수 기준에 따라 등급이 분류됩니다.

2012년	2013년~현재	등급 기준
A+(매우우수)	S(탁월)	탁월한 지속가능경영 체제를 구축하고 있어 타 기업과 지속가능경영 전반에 모범이 되는 상태
	A+(매우우수)	매우 우수한 지속가능경영 체제를 구축하고 있으며 지속적으로 우수한 성과를 보이고 있는 상태
A(우수)	A(우수)	비교적 우수한 지속가능경영 체제를 구축하고 있으며 체제 고도화를 위한 노력이 필요한 상태
B+(양호)	B+(양호)	양호한 지속가능경영 체제를 구축하고 있으며 체제 개선을 위한 지속적 노력이 필요한 상태
B(보통)	B(보통)	다소 취약한 지속가능경영 체제를 구축하고 있는 상태로 체제 개선을 위한 지속적 노력이 필요한 상태
C(취약)	C(취약)	취약한 지속가능경영 체제를 구축하고 있으며 체제 개선을 위한 상당한 노력이 필요한 상태
	D(매우취약)	매우 취약한 지속가능경영 체제를 구축하고 있으며 체제 개선을 위한 상당한 노력이 필요한 상태

Target Data (Y)

평가 기업 등급별 정리
한국 ESG 기준원 등급 데이터

2

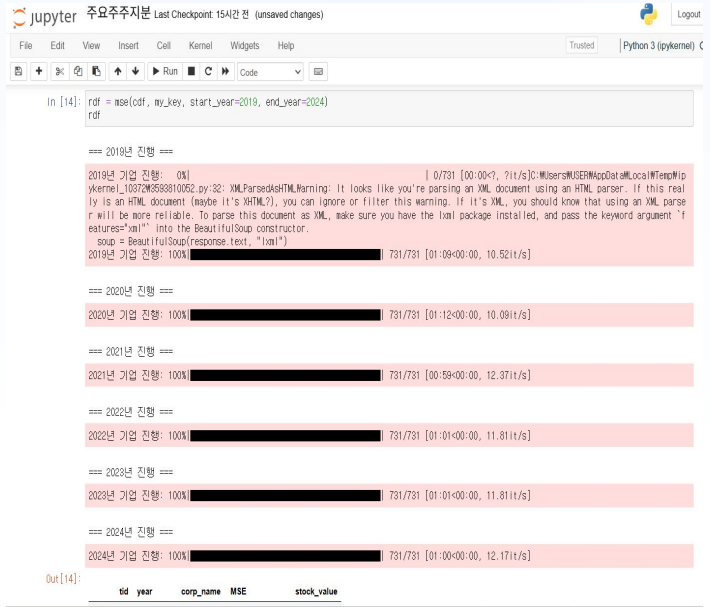
- X_features

1. A_SIZE : log(total_asset), log(기업의 총자산)
2. LEV : Leverage, 총 부채 / 총 자산
3. TQ : Tobins'Q, (시가총액 + 총 부채) / 총 자산
4. FOR : Foreign Ownership Ratio, 외국인 지분율
5. MSE : Major Shareholder Equity, 주요주주지분
6. ROA : Return On Assets, 총 자산 수익률
7. ADV : Advertising Intensity, 광고비
8. SGR : Sales Growth Rate, 매출성장률
9. R&D : Research & Development, 연구개발비

Feature Selection (X)

주요 재무지표 선정
(기존 연구 답습)

3



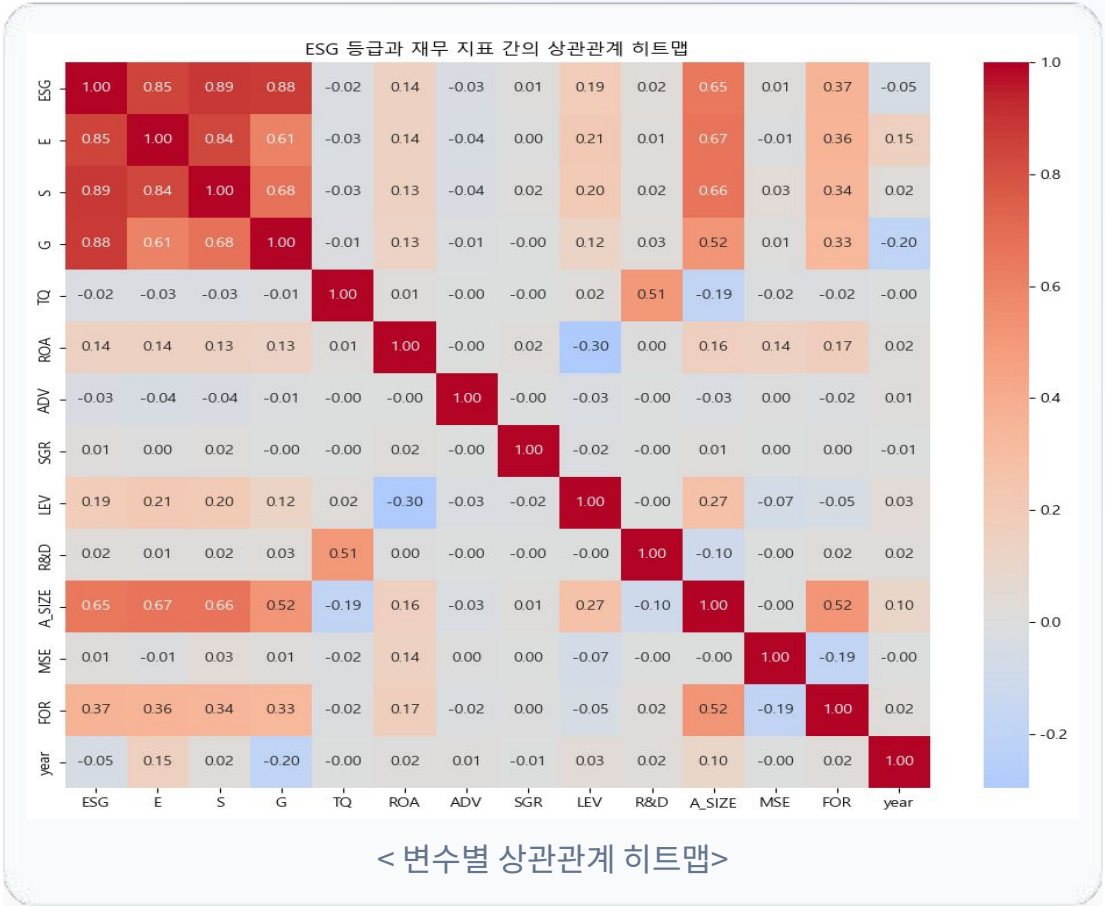
Preprocessing

지표 수집 및 전처리
결측치 제거 및 정규화

03. 중간 모델링 (Base Model)

가) 변수간 히트맵 분석

- **자산(0.65), 외국인 지분율(0.38)**이 타겟과의 가장 높은 양(+)의 상관관계 확인.
- 타겟과의 상관관계 : 자산 외에는 명확한 관계 파악 어려움.
  다중공선성 우려 : 전 **변수별 상관계수 0.55 이하**



< 변수별 상관관계 히트맵>

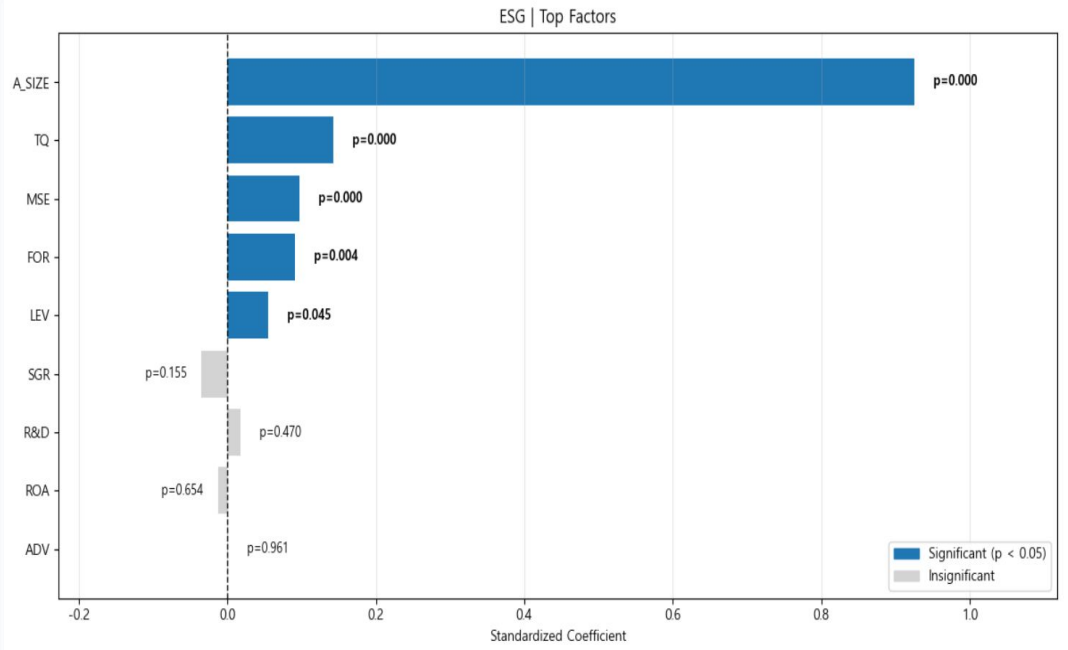
03. 중간 모델링 (Base Model)

나) 다중회귀분석 (Linear Regression)

- **설명력: R^2 0.440 달성** (논문 대비 약 2배 상승).
- 구조적 데이터일수록 유의미.
  R&D 및 광고비(ADV)는 기업 미공시로 인해 유의성 저조

```
=== ESG 등급 예측 모델 분석 결과 ===
[다중회귀 ESG모델 성능 보고서]
MAE:  1.0489
MSE:  1.5217
RMSE: 1.2336
R2:   0.4484
```

< 다중회귀분석 결과 (1) >



< 다중회귀분석 결과 (2) >

03. 중간 모델링 (Base Model)

다) E, S, G 분야별 설명력 분석 비교

* G(지배구조) 분야는 정성적 요소가 강해 외형적 재무 지표만으로는 설명하기 어렵다는 한계 확인.

03. 중간 모델링 (Base Model)

문제점 도출

현재의 X_feature 만으로는 정확한 예측 모델 제작에 한계가 있음.

개선 전략 (Strategy)

- **변수 통합:** 유의성이 낮은 R&D와 광고선전비(ADV)를 통합하여 '기타 판매비와 관리비'로 반영.
- **X_feature 추가:** 설명력이 낮은 G 분야 보완을 위해 재무지표 중 정성적 요소를 대변할 수 있는 변수 추가.

- 추가 X_features

1. SGAE_R : 기타 판매비와 관리 비율
2. Fe_R : 남성 대비 여성 직원의 비율
3. Re_R : 총 직원 대비 정규직 비율
4. SA : log(1인당 평균 임금)
5. Pay_Gap : 남녀 임금 평균 차이 비율(남성 - 여성) / 남성
6. W_YEAR : 기업별 평균 근속연수
7. DIV : 주가배당율
8. DIR_FE : 임원의 여성비율
9. DIR_OUT : 임원의 사외이사 비율
10. DIV_enco : 주가배당이 있으면 1, 없으면 0
11. DIR_FE_enco : 여성임원이 있으면 1, 없으면 0
12. IND : 산업군

< 추가 X_feature 도출 >

04. 추가 데이터 수집 및 전처리 (Plus Data)

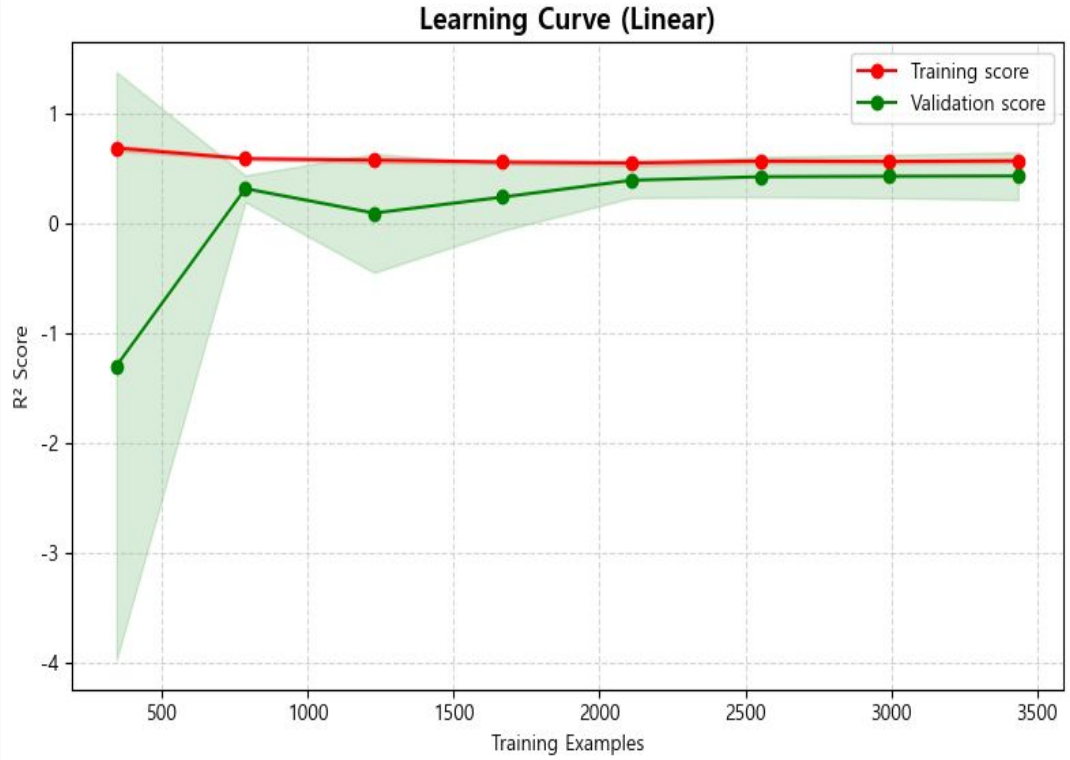
모델 성능 대폭 향상

0.5717

Updated Model R^2

< 다중회귀분석 모델 최종 학습 결과 >

- **추가 X_feature 확정:** 정성적 요소를 반영한 새로운 변수 수집 및 적용.
- **성능 개선:** 중간 모델 대비 R^2값이 **0.1367 상승**.
- 회귀 분석 모델로서 유의미한 설명력을 확보하였으나, Learning Curve 분석 결과 Base Model 학습 한계점 도달



< Learning Curve 분석 >

05. 최종 모델링 (Final Model)

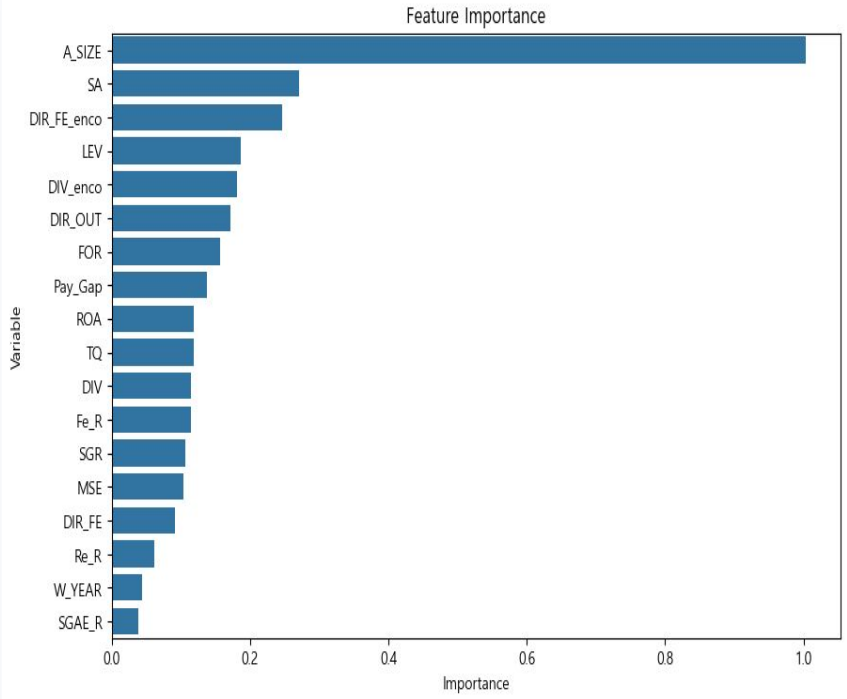
회귀 모델의 한계를 확인하고, 보다 발전된 실무 활용을 위해 **분류 모델(Classification)**로 전환.

1. 회귀 모델 고도화

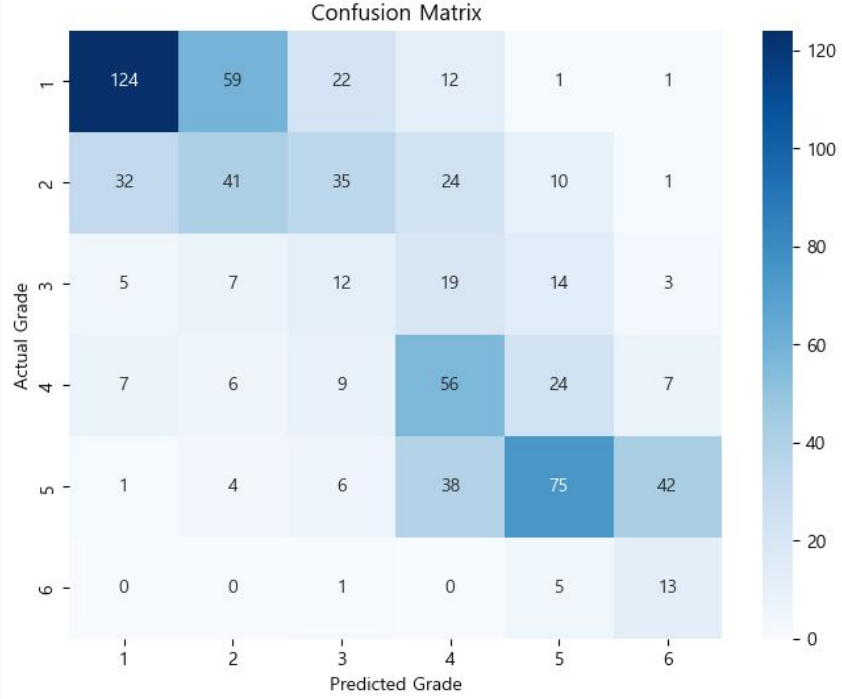
- XGBoost, RF, LGBM 적용
- **XGBoost 우수 (R^2 0.649)**
- Train-Test 격차 0.1236
- 회귀모델 한계 도달
- 실무 활용 위해 분류 모델 전환

2. 모델 전환 (분류)

- 다중분류모델로 전환
- Logistic Base Model 구축
- 다중분류분석 시작
- Rolling Window 기법 적용



< 다중분류 분석 (1) >



< 다중분류 분석(2) >

05. 최종 모델링 (Final Model)

회귀 모델의 한계를 확인하고, 보다 발전된 실무 활용을 위해 **분류 모델(Classification)**로 전환.

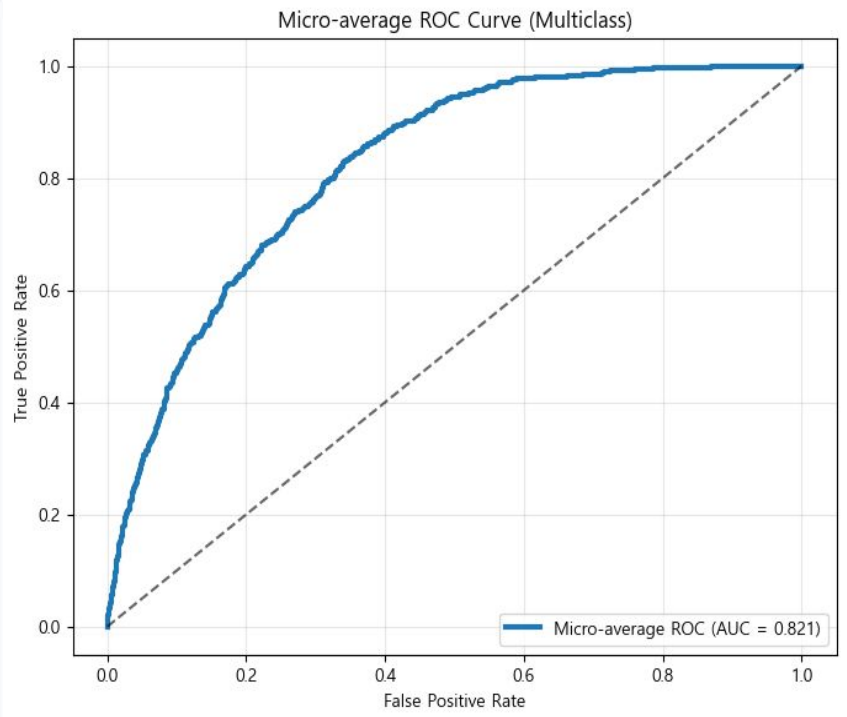
3. 검증 방법

- 클래스 불균형 고려
- 임계값(Threshold) 조정
  ROC AUC 지표 선정

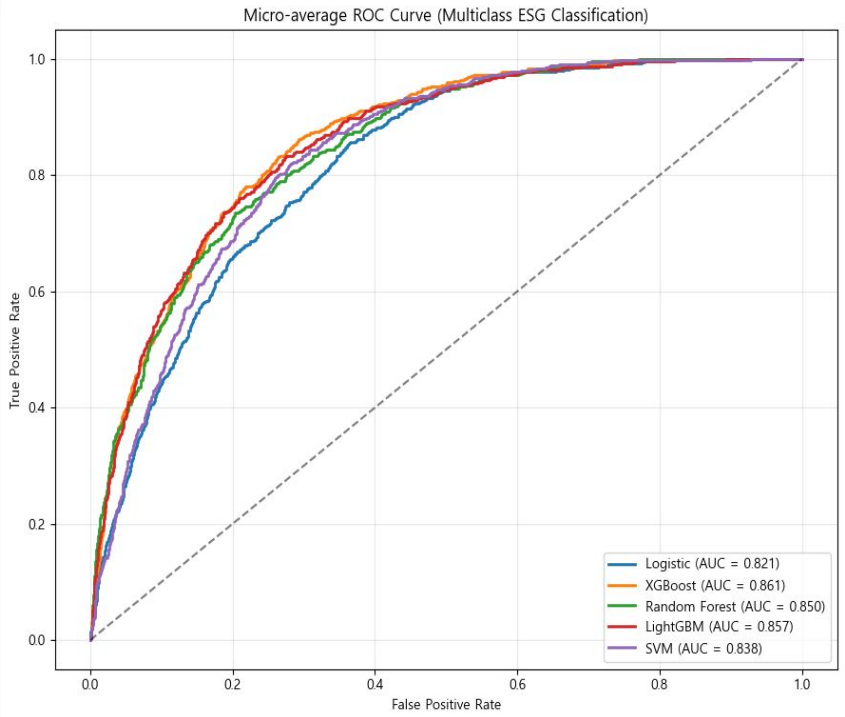
.

4. 발전 모델 적용

- XGBoost, Random Forest
- LGBM, SVM 등 테스트
- 다양한 알고리즘 성능 비교



< 검증지표 선정(ROC_AUC) >



< 우수 모델 선정 >

05. 최종 모델링 (Final Model)

최종 모델 선정: XGBoost

0.861

Updated Model ROC AUC

< Final Model 결과 >

- Base Model 대비 **ROC AUC 0.4 향상**으로 가장 우수한 성능 기록.
- 분류 모델 전환을 통해 등급 예측의 실효성 확보.

결과 및 인사이트

활용 방안

- **투자 전략 활용**: ESG 등급으로 **기업의 투자 가치 평가**
- **기업 컨설팅**: 어떤 방면으로 개선해야 ESG등급이 상승할지 **기업 중장기 전략 수립**
- **리스크 관리**: ESG등급 하락 조기 예측, **기업의 리스크 관리지표로 활용**

Ready to Work

Ready to create the best customer experience through data.

데이터를 통해 최고의 고객 가치를 만들어갈 준비가 되어 있습니다.

010 - 4023 - 9072

hw135@naver.com