

# 재무 비율을 활용한 기업 부도 예측

데이터사이언스(가반)

망고 조 - 이강룡, 이상민, 이정호, 조성원

최종 발표자 : 이상민, 이정호

# 목차

## 1. 배경 및 목표

## 2. 기대 효과

## 3. 분석 및 실험 결과와 해석

### 3-1. 데이터 수집

### 3-2. EDA 및 전처리

- 기초통계
- 논문 **feature selection**
- **t-test**로 **1차 feature selection**
- 극단치 조절

### 3-3. 모델링

- **Linear Regression**
- **Data imbalance**
- 차원 축소 (**2차 feature selection**)
- 결과

### 3-4. 해석 및 시사점

## 4. 일정 및 팀원 별 기여 내용

## 5. 프로젝트 회고

## 1. 배경 및 목표

**부도란?**

어음교환소에 제시된 **어음**에 대해 해당 기업이 약속된 금전을  
결제하지 못하게 될 경우

**어음이란?**

발행인이 소지인에게 일정한 금전의 지급을 약속하거나 제  
3자에게 그 지급을 위탁하는 유가증권

# 1. 배경 및 목표

홈 > 증권 > 국내증시

## 대유위니아 그룹 줄도산 위기에 주가 널뛰기... '투자 주의보'

입력 2023-10-07 09:00:20 수정 2023.10.07 09:00:20 양지혜 기자

뉴스듣기 가

위니아, 36억 원 규모 만기어음 부도 발생  
위니아전자·대유플러스 이어 기업 회생 신청  
신용등급 강등으로 자금 조달 여건도 악화



위니아의 2024년형 덤채 신제품. 사진 제공=위니아

### 오늘의 핫토픽

- # 초일류 외진 이재용
- # ECB 금리인하
- # 서울 아파트값 상승
- # 김호중 음주행소니 구속
- # "중부세 합헌"

### 오늘의 이슈

#커지는 중부세  
[만파식적] 쫓

#재상변 특검법

## '법정관리 신청' 위니아 협력사 450곳 줄도산 위기

송고시간 | 2023-10-10 15:12

정아름 기자

| 협력사 "1천억원 납품대금 못받아" 금융지원 촉구



'법정관리 신청' 위니아 대금 지급 촉구하는 협력사들  
[위니아 채권단 협의회 제공. 재판매 및 DB 금지]

# 1. 배경 및 목표

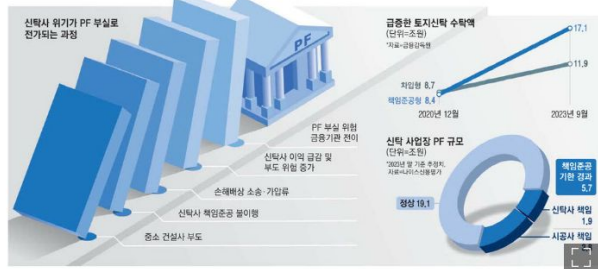
## [단독] 중소건설사 부도에 ... 준공 기한 약속한 신탁사까지 벼랑끝

손동우 기자 aing@mk.co.kr  
이희수 기자 lee.heesoo@mk.co.kr

입력 : 2024-05-08 17:46:20 수정 : 2024-05-08 19:59:06



PF사업·금융 도미노 위기  
국내 건설 불황 장기화되며  
건설사 완공 기한 못 지키자  
'책임준공형 PF' 23%가 위태  
안성·평택 물류센터 채권단  
신원자산신탁에 손해 소송  
업계최초 예금 가압류까지



## 빌려준 돈 40조 넘는데...건설사 부도에 떠는 은행

N 구독

이건업 기자 |  
2024.04.16 06:20:00



### [PF 부실 현실화]③

지난해 은행권 건설업 대출 42조4755억  
기업은행, 7조9864억...전체 18.8% 차지  
고금리+증기 지원 역할 맞물린 결과  
부동산 경기 침체에 부실 가능성 확대

[이데일리 마켓in 이견업 기자] 중소 건설사들을 중심으로 부도 사례가 속출하면서 불안감이 고조되고 있는 가운데 국내 은행권에 대한 건설업 종 대출 규모가 40조원을 돌파했다. 지난해 대내외 불확실성으로 시장금리가 꾸준히 오르면서 건설사들이 회사채 발행보다는 은행 대출 등 단기 자금 시장 문을 적극 두드린 결과라는 분석이다.

## 1. 배경 및 목표

### 목표 변경

기업의 **신용등급**  
기준으로 부도 예측



재무 비율 데이터 추적  
**직접 부도 여부를 판단**  
예측 모델 제작

## 2. 기대 효과

조기 경고 시스템 구축

투자자 및 투자기관 보호

사회적 충격 예방



재무 비율 분석



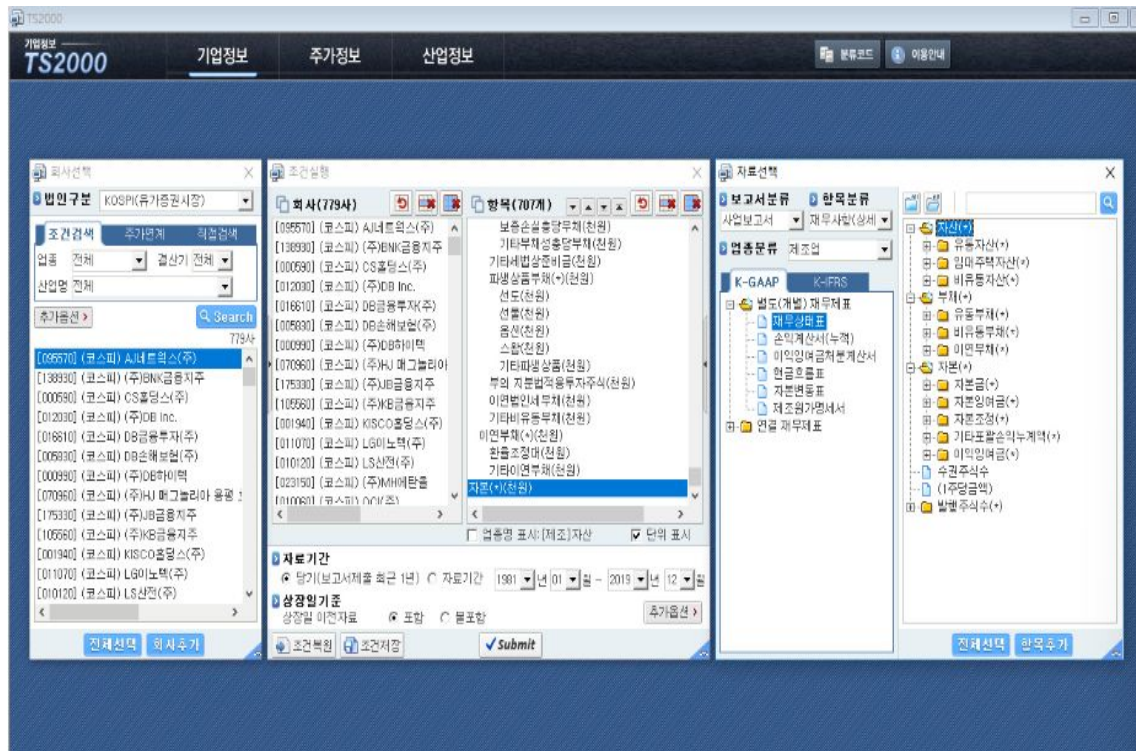
사업 구조조정



흑자 전환 성공

## 3-1. 데이터 수집

# 2007년~2020년 상장기업 분기별 데이터





### 3-1. 데이터 수집\_재무비율

#### 재무비율을 사용한 이유

재무비율: 재무제표 간의 상대적인 비율

- 데이터의 정규화
- 대다수의 부도예측모형(Altman Z score 등)이 재무비율 사용
- 비전문가가 활용하기 쉬움

### 3-1. 데이터 수집\_재무비율

#### 안정성 비율

채무상환능력 판단

ex) 유동비율, 당좌비율

#### 수익성 비율

경영성과 판단

ex) 매출총이익율, ROA, ROE

166

#### 성장성 비율

성장성 판단

ex) 매출액증가율, 영업이익증가율

#### 활동성 비율

경영의 효율성 판단

ex) 총자산회전율, 총부채회전율

### 3-1. 데이터 수집\_부도 기업 정의

#### 소속코드

03 주권상장법인 관리종목

06 코스닥상장법인 관리종목

04 상장폐지사

#### 감사의견코드

DS - 의견거절

DU - 부정적의견

DI - 한정의견

위 코드에 해당 기업을  
**재무 위험 기업**으로 가설 설정

### 3-1. 데이터 수집\_거시경제지표

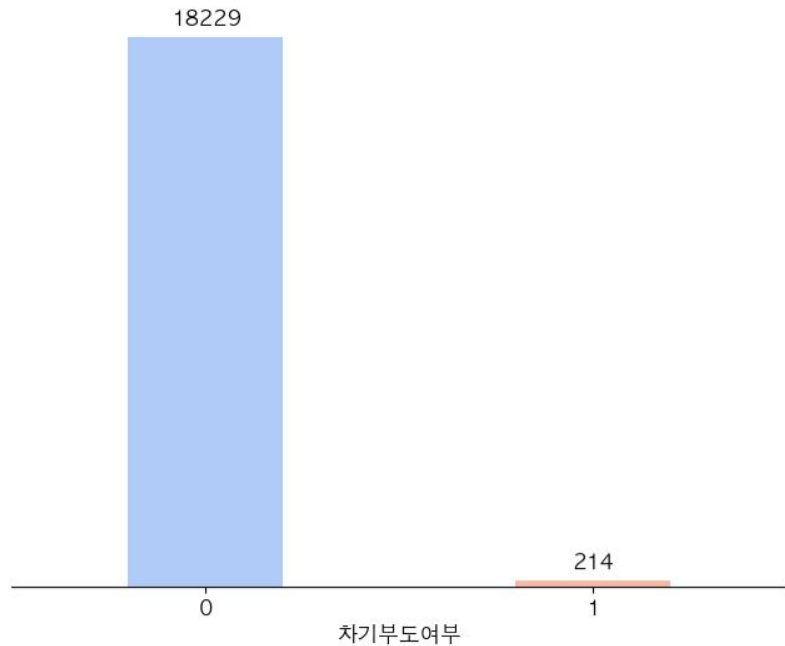
회사명	거래소코드	차기부도여부	회계년도
(주)CMG제약	58820	0	2007.0
(주)CMG제약	58820	0	2008.0
(주)CMG제약	58820	0	2009.0
(주)CMG제약	58820	0	2010.0
(주)CMG제약	58820	0	2011.0

**한국은행 경제통계시스템**의 거시경제현황 데이터 활용  
회계년도 당시 거시경제지표를 추가

ex) 통화량, 기준금리, 국제수지, GDP

## 3-2. EDA 및 전처리

### EDA



-데이터에 존재하는 전체 칼럼은 총 **166**개

- **Label**의 분포는 정상기업이 **18, 229** 개

부도기업이 **214**개로 불균형 데이터임을

확인할 수 있다

## 3-2. EDA 및 전처리

### 정상기업군과 차기 부도 기업군에 대한 기초통계 비교

- 부채 비율에 대한 지표는 차기부도 기업군이 정상 기업군보다 높다
- 이익을 같은 경우 정상기업군이 더 높았다.
- 이자보상배율의 경우 이상치가 존재함을 확인, 이후

	mean	std	min	max
부채비율	393.626	1153.595	0.000	12464.470
비유동부채비율	78.914	306.874	0.000	2767.390
유동부채비율	314.713	1043.827	0.000	12309.500
차입금의존도	37.899	23.255	0.000	163.400
자기자본구성비율	41.459	23.785	0.000	93.930
자기자본배율	24305.988	350159.144	0.000	5122701.680
이자보상배율(이자비용)	-60.026	635.625	-9195.000	168.770
총자본영업이익률	-11.772	17.306	-128.530	28.110

부도기업군의 기초통계량

	mean	std	min	max
부채비율	98.375	256.776	0.000	13769.960
비유동부채비율	24.084	103.211	-0.220	9605.620
유동부채비율	74.291	197.483	0.000	13602.960
차입금의존도	20.197	17.800	0.000	257.880
자기자본구성비율	61.593	20.309	0.000	99.950
자기자본배율	1640.254	46472.680	0.000	6266409.710
이자보상배율(이자비용)	79545590.797	270592150.446	-4493653.000	999999999.000
총자본영업이익률	3.304	9.244	-169.750	76.970

정상기업군의 기초통계량

## 3-2. EDA 및 전처리

### 1차 feature selection

전체 칼럼이 총 **166**개로 분석에 사용하기에는 **feature**의 수가 너무 많아 **t-test**를 통해 집단간 유의한 차이를 갖는 **feature**만 분석에 사용

최종적으로 **63개의 feature**만 남는다.



## 3-2. EDA 및 전처리

### 극단치 조절

t-검정을 하고나서도, feature 별 min, max 정도가 심함

=> min-max 정규화 추가 적용

$$x_{scaled} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$



## 3-3. 모델링

### Logistic Regression 사용

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lr	Logistic Regression	0.8985	0.8806	0.6667	0.0801	0.1412	0.1231	0.2048	1.6160
svm	SVM - Linear Kernel	0.7923	0.8666	0.7733	0.0554	0.1015	0.0817	0.1690	0.9030
knn	K Neighbors Classifier	0.8713	0.8144	0.6133	0.0548	0.1002	0.0807	0.1550	1.4950
nb	Naive Bayes	0.7840	0.8316	0.6800	0.0472	0.0859	0.0657	0.1365	0.8660
qda	Quadratic Discriminant Analysis	0.8885	0.6377	0.1200	0.0165	0.0243	0.0067	0.0087	0.9170
ridge	Ridge Classifier	0.9591	0.8035	0.0200	0.0082	0.0116	-0.0051	-0.0059	0.8330
dt	Decision Tree Classifier	0.9796	0.4955	0.0000	0.0000	0.0000	-0.0099	-0.0101	0.9680
rf	Random Forest Classifier	0.9768	0.6079	0.0000	0.0000	0.0000	-0.0116	-0.0116	1.4480
ada	Ada Boost Classifier	0.9753	0.6781	0.0000	0.0000	0.0000	-0.0121	-0.0123	1.4340
gbc	Gradient Boosting Classifier	0.9734	0.6194	0.0000	0.0000	0.0000	-0.0131	-0.0133	3.3160
lda	Linear Discriminant Analysis	0.9597	0.5947	0.0000	0.0000	0.0000	-0.0167	-0.0185	0.8560
et	Extra Trees Classifier	0.9772	0.6530	0.0000	0.0000	0.0000	-0.0113	-0.0114	0.9830
lightgbm	Light Gradient Boosting Machine	0.9801	0.5128	0.0000	0.0000	0.0000	-0.0096	-0.0098	1.0360
dummy	Dummy Classifier	0.9884	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8540

부도예측에 중요한 성능 지표인

**AUC, F1-score**가 높은 모델을 고르기 위해,  
autoML library인 **pycaret**을 사용해 모델 비교

모델 비교 결과

**Logistic Regression**의 성능이 제일 좋게 나옴

### 3-3. 모델링

#### Data Imbalance(불균형) 문제

부도 기업의 비율이 **1%** 정도로 굉장히 불균형한 데이터셋

문제 해결

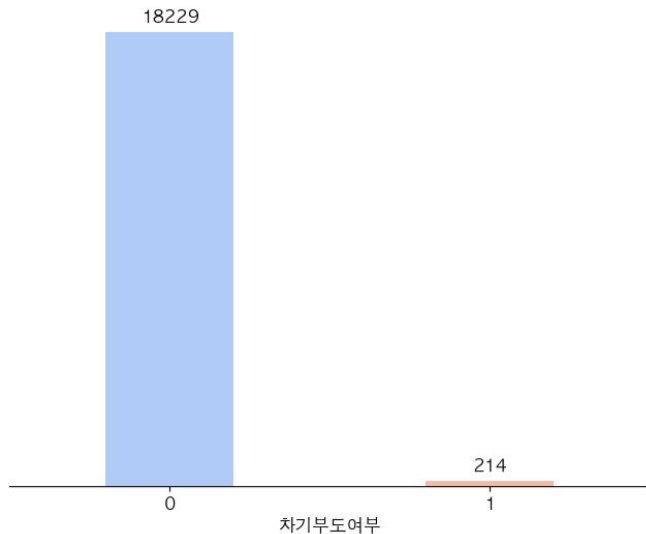
- **SMOTE** 기법

**Oversampling**을 통해 부도 기업 데이터의 수를 늘림

- **class\_weight='balanced'**

클래스가 적을수록 가중치를 높게,

클래스가 많을수록 가중치를 낮게 설정하여 균형을 맞추는 방식



### 3-3. 모델링

#### 차원 축소 - 2차 feature selection

t-검정을 통해 선택한 63개의 feature 中 다시 selection 수행

#### 2차 feature selection 방법

try 1 : 외부 도메인 지식(논문 등)을 통해 feature를 N개로 줄이기

try 2 : Logistic Regression 모델을 만들어 feature 중요도를 기반으로 2차 feature selection

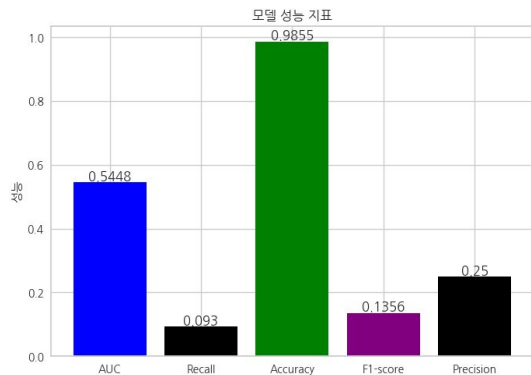
두 방법을 사용해본 뒤, 성능이 좋은 model 선택

### 3-3. 모델링

## 결과

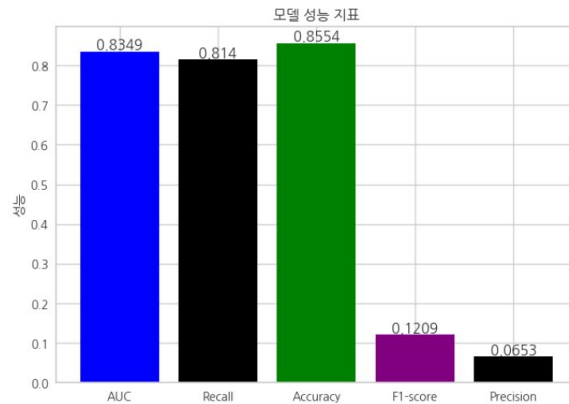
#### 1) baseline

63개의 feature 모두 학습



#### 2) data imbalance 문제 해결

SMOTE기법을 사용 부도기업군 데이터를  
증강

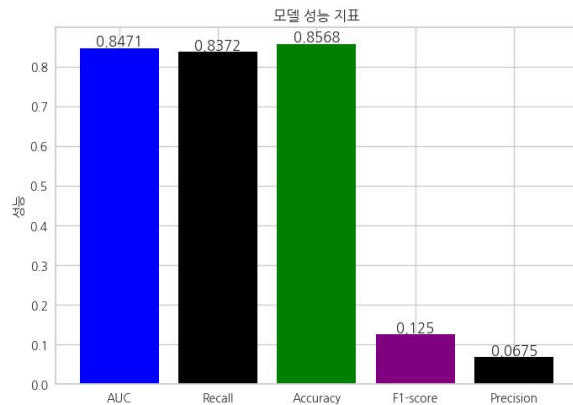


### 3-3. 모델링

## 결과

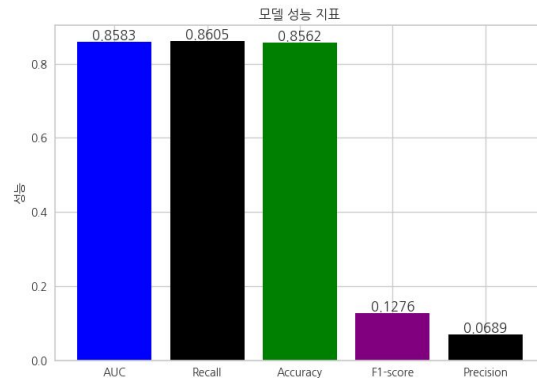
#### 3) 논문 feature 사용

기준에 알려진 부도 예측시 유의미한 변수를 활용



#### 4) 2차 feature selection 후

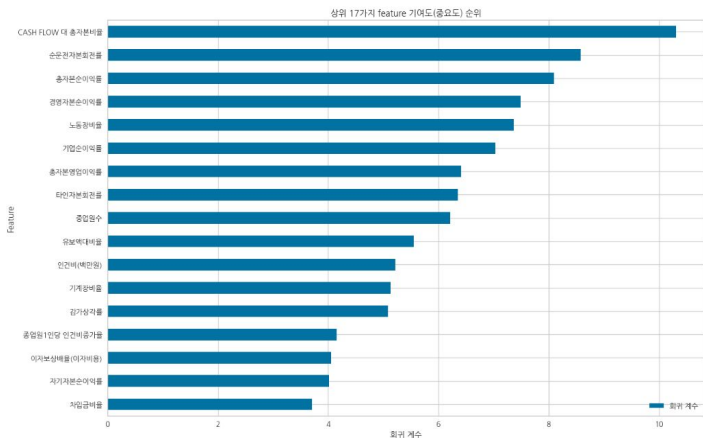
feature 중요도 기반 변수 선정



### 3-3. 모델링

## 결과

data imbalance 해결한 모델에서  
추출한 중요한 feature 상위 17가지



#### 4) 2차 feature selection 후

ROC AUC: 0.8583

Precision: 0.0689

Recall: 0.8605

F1 Score: 0.1276

Accuracy: 0.8562

### 3-4. 해석 및 시사점

## 주요 feature 분석

### 1. 수익성 관련 지표

일정 기간 동안

**순이익을 얼마나 벌었는지와**

**관련이 깊었음**

총자본순이익률

경영자본순이익률

기업순이익률

총자본영업이익률

자기자본순이익률

### 3-4. 해석 및 시사점

## 주요 feature 분석

### 2. 활동성 관련 지표

일정 기간 동안 얼마나

**효율적으로 매출을 만드는지와**

관련이 깊었음

CASH FLOW 대 총자본비율

타인자본회전률

순운전자본회전률



### 3-4. 해석 및 시사점

## 주요 feature 분석

### 3. 부가가치 관련 지표

설비, 고용 투자를 했을 때  
투자 대비 가치와  
관련이 깊었음

기계장비율

감가상각률

인건비(백만원)

종업원1인당 인건비증가율

종업원수

노동장비율

3-4. 해석 및 시사점

주요 feature 분석

4. 민스키 분석

기업의 영업이익과  
부채 상환 능력에 따라서  
관련성이 깊었음

유보액대비율

차입금비율

이자보상배율(이자비용)

민스키분석

기업의 영업이익과 원리금 상환과의 관계를 통해 기업의 안정성을 분석합니다.

[민스키분석으로 분석한 기업의 현 상태]

상태	내용
헛지	재무제표상 회사의 영업이익이 이자비용과 단기성차입금보다 많은 상태입니다.
투기	재무제표상 단기성차입금이 회사의 영업이익보다 많은 상태입니다.
폰지	재무제표상 이자비용이 회사의 영업이익보다 많은 상태입니다.

### 3-4. 해석 및 시사점

#### 시사점

주요 features 활용  
모델링 구축



유료 데이터를 이용하지 않고도

**기업 재무 안정성**

**투자 위험 판단**

**기업 부도 예측**

### 3-4. 해석 및 시사점

#### 보완점

1. 테스트 데이터 확보 및 수행
2. 부도 예측 결과 시각화

## 4. 일정 및 팀원 별 기여 내용

### 일정

팀 이름 및 조장 설정  
프로젝트 주제 설정 및 주제 관련 정보 조사  
제안 발표서 작성 및 발표자 선정

1주 차

2주 차

3주 차

분류 모델 선정 및 F1 score 개선  
분석 결과 도출 및 시각화  
최종 보고서 작성

데이터 수집, EDA, 전처리 과정

## 4. 일정 및 팀원 별 기여 내용

### 팀원 별 기여 내용

**이강룡 (조장) :** 제안 발표, **PPT** 제작, **feature selection**, **Modeling** 및 성능 개선, 결과 시각화

**이상민 :** 최종 발표 및 **PPT** 제작, **EDA** 및 **feature selection**, **t-test**, 거시경제지표 수집 및 분석

**이정호 :** 최종 발표 및 **PPT** 제작, **EDA** 및 분석 결과 시각화

**조성원 :** 제안 발표, **PPT** 제작, 데이터 수집 및 전처리, **feature** 분석 및 결론 도출

## 5. 프로젝트 회고

이강룡 (조장) : 금융지식에 대해서도 많은 것을 알아갔고, 데이터 마이닝 및 모델링 과정 전부를 직접 몸으로 부딪혀볼 수 있었던 뜻깊은 경험이었다.

이상민 : 금융데이터를 활용하여 데이터 사이언스 수업때 배운 기법들을 활용할 수 있는 좋은 기회였다  
특히 금융도메인 프로젝트는 처음해보는 프로젝트이기에 **feature**를 해석하는 것이 어려웠지만  
결과적으로 많은 것을 얻어 갈 수 있었다.

이정호 : 실제 사회의 데이터가 불균형하고 분석하기 어렵다는 것을 알았습니다. 모델링 이전에 데이터 전처리의 중요성을 깨달았습니다.

조성원 : 금융 데이터를 활용해 데이터 분석을 해보고 싶은 생각으로 데이터사이언스를 수강했고,  
실제 프로젝트에서 팀원들과 함께 수행할 수 있어서 보람을 느꼈습니다.

향후 관련 내용을 더 공부하고, 테스트를 수행해서 직접 결과를 시각화 해보고 싶습니다.

감사합니다