상당 기업부실예측을 통한 보 기업 여신 까동 심사 솔루션 개발



]. 프로젝트 개요 **2.** HOIEL 4 TO

3. 데이터 西州리 **4.** 변수 선정

5. Modeling

6. 3148 **기**。 기업 여긴 심사 시스템

8. 기대효과 및 한계

프로젝트 팀 구갱 및 역할



@이**현주** #자료 수집 및 문서 정리 #ppt 제작



@조은서(조장) # 데이터 수집 및 전처리 # 성능 평가

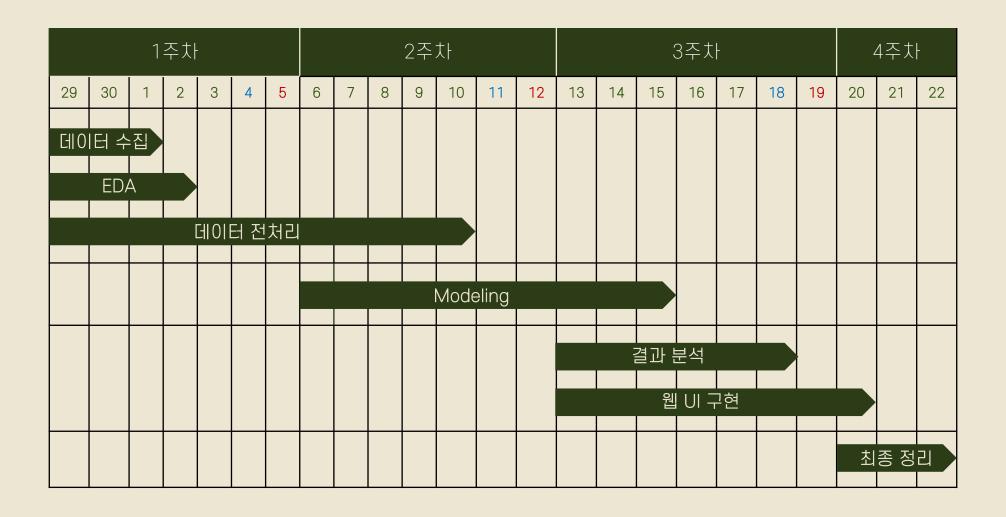


@정동혁 # 자료 수집 # 웹 제작



@하규진(부조장)
데이터 전처리 및 모델링
기업 등급화

프로젝트 구행 갤亦



데이터 수집

데이터 전처리 & Feature Selection

Modeling

검증

TS 2000

상장기업 중 금융업 제외 기업의 2007~2016년 재무비율 및 재무상태

폐지사 포함 총 기업 수 <u>2,207사</u> 총 데이터 19,989 1년 당 약 2,000개 데이터

부도 라벨링

부도 정의에 해당되는 기업 → 부도 기업

1.1

1.1

11

1.1

1.1

1 1

1.1

 $\mathbf{I} \cdot \mathbf{I}$

 $\frac{1}{1}$

1.1

1.1

1.1

1.1

1.1

피흡수 합병, 회사 합병의 경우 → 정상 기업

정상 기업: 2,208 부도 기업: 189

Data 정제

파생 변수 생성

결측치 제거

Robust Scaling

파생 변수: 7개 (부도 여부 포함) 총 50개 변수

변수 선정

T-Test

Wrapper Method

VIF test

등분산 → Student's t-test 이분산 → Welch's t-test 전진 선택/후진 제거/단계별 모두 같은 변수 추출 VIF 지수가 높은 변수 제거 **Oversampling**

Models

Logistic Regression

Decision Tree

Random Forest

Lightgbm

Xgboost Classifier

성능 평가

Recall

H

1.1

+1

+1

H

1.1

1.1

 \mathbf{I}^{-1}

 $\frac{1}{1}$

į I

F1- Score

변별력

KS Statistics

Accuracy Ratio

데이터 무낍

TS2000을 활용한 상장기업(KOSPI+KOSDAQ) 중 <u>금융업을 제외한</u> 제조업의 재무비율 및 자산, 자본, 매출액 등의 재무상태 변수 데이터



폐지사 포함 총 기업 수 2,207사 총 데이터: 19,989 1년당 약 2,000개 데이터

2007~2016년 총 10년 데이터 사용





금융위기 기간 2007~2008년 데이터 추가하여 모델 강건성을 높이고자 함 파생 변수 계산을 위해 2006년 추가 수집 후 삭제

• • •

부도 기업 정의 (DART 공시 자료 활용)

1. 상장 폐지된 기업 중 감사의견코드가 DS or DU 2. 자본잠식률이 50% 이상 (2년 이상) 3. 자진상장폐지

[표1]		
감사의견코드	AG	GAAP 위반
감사의견코드	DI	독립성
감사의견코드	DS	감사 범위 제한
감사의견코드	DU	불확실성
감사의견코드	NS	보고서 미제출
감사의견코드	QA	계속 기업 전제
감사의견코드	QC	GAAP 위반. 계속성 변경
감사의견코드	QG	GAAP 위반
감사의견코드	QQ	한정의견
감사의견코드	QS	감사 범위 제한
감사의견코드	QU	불확실성

피흡수 합병되었거나 회사 합병 된 경우 정상 기업으로 라벨링 (DART 공시 자료 활용)

회사명	거래서코드	회계 연도	상장폐지일	부도여부	폐지사유
부산방직공업(주)	025270	2009 /03	2016/08/18	0	
부산방직공업(주)	025270	2010 /03	2016/08/18	0	
부산방직공업(주)	025270	2011 /03	2016/08/18	0	
부산방직공업(주)	025270	2012 /03	2016/ 08 /18	0	
부산방직공업(주)	025270	2013 /12	2016/08/18	0	
부산방직공업(주)	025270	2014 /12	2016/ 08 /18	0	
부산방직공업(주)	025270	2015 /12	2016/ 08 /18	0	피흡수합병

회계연도와 상장폐지 연도가 같은 경우

직전 연도를 부도로 라벨링

[±3]

회사명	거래서코드	회계 연도	상장폐지일	부도여부
(주) 네프로아이티	950030	2009 /03	2011/11/05	0
(주) 네프로아이티	950030	2010 /03	2011/11/05	1
(주) 네프로아이티	950030	2011 /03	2011/11/05	1

회계연도와 상장폐지 연도가 2년 이상 차이 난 경우

상장폐지 연도에만 부도로 라벨링

[丑4]

회사명	거래서코드	회계 연도	상장폐지일	부도여부
(주) 로케트전기	000420	2006 /12	2015/ 02/24	0
(주) 로케트전기	000420	2007 /12	2015/ 02/24	0
(주) 로케트전기	000420	2008 /12	2015/ 02/24	0
(주) 로케트전기	000420	2009 /12	2015/ 02/24	0
(주) 로케트전기	000420	2010 /12	2015/ 02/24	0
(주) 로케트전기	000420	2011 /12	2015/ 02/24	0
(주) 로케트전기	000420	2012 /12	2015/ 02/24	0
(주) 로케트전기	000420	2013 /12	2015/ 02/24	1

최종 결과

■ 정상기업: 2208, 부도기업: 189

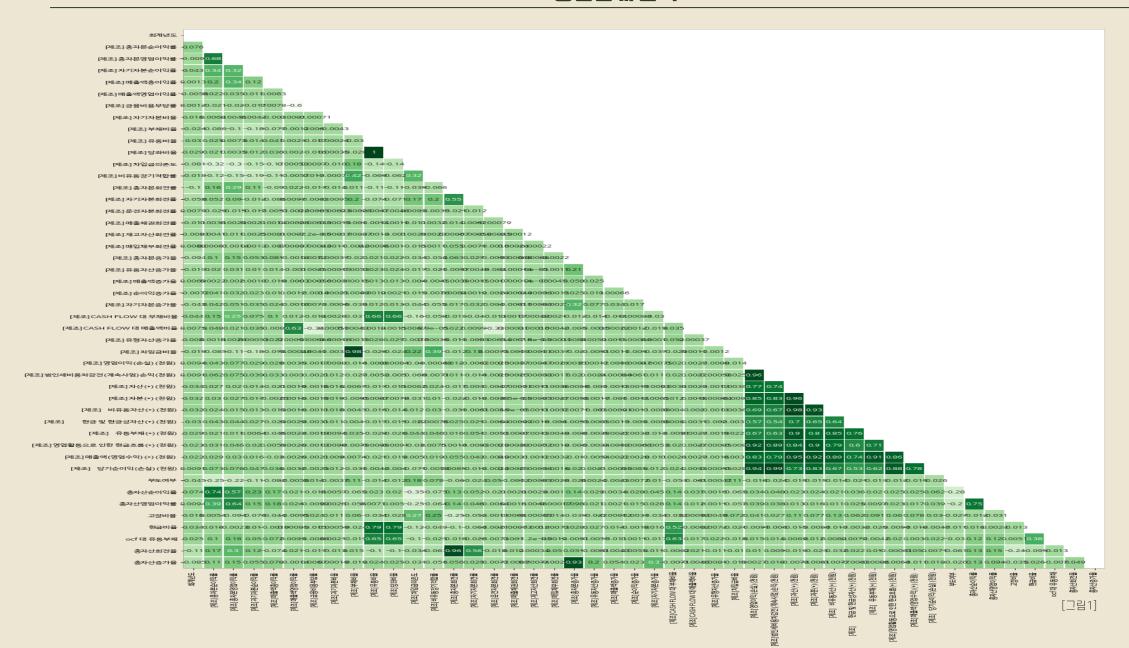
■ 정상 데이터: 19802, 부도 데이터: 194



논문 활용하여 총 46개 변수 선정

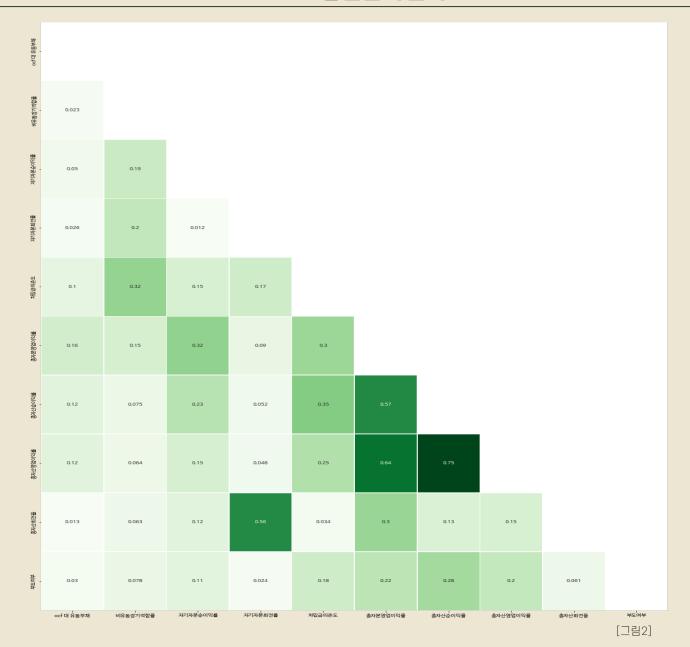
1. 약 500개 결측치 제거
 전체 데이터 약 20,000개 중 1/40에 해당하므로 완전 삭제
 2. 중복 데이터 43개 제거
 3. 파생 변수에 사용한 2006년 데이터 제거
 4. 이자비용 이상치 발견 → 금융기업 리드코프 제거 //

앙관관계분객



- 0.8 - 0.6 - 0.4 - 0.2 - 0.0 - 0.2 - 0.4

앙관관계분객





변구 앤쟁 process

01 T-Test

등분만 검쟁 후 t-test를 통한 변구 앤쟁

등분산 -> Student's T-test 이분산 -> Welch's T-test

<46 TH →19 TH>

02 Wrapper Method

래퍼 방식: 전진 선택법, 후진 제거법, 단계적 선택법 모두 같은 변수 추출

<197H**→9**7H>

03 VIFTest

다풍공산생을 고려하여 회꽁 변수 산택

<97H→97H>

흥 변구 (46개)



[丑5]

총자본순이익률 총자본영업이익률 총자산순이익률 총자산영업이익률

자기자본순이익률

매출액총이익률 매출액영업이익률 금융비용부담률 매출액순이익률 영업이익이자보상비율

수익성 변수(10)

총자산회전율

총자본회전율 자기자본회전율 운전자본회전율 매출채권회전율 재고자산회전율 매입채무회전율

활동성 변수(7)

안정성 변수(14)

성장성 변수(7)

자기자본비율

고정비율 부채비율 유동비율 <mark>현금비율</mark> 당좌비율 차입금의존도

차입금비율

비유동장기적합률 총자산대비유동부채 총자본대비 CASH FLOW 대 부채비율 CASH FLOW 대 매출액비율 OCF 대 유동부채 총자산증가율 총자본증가율

총자본증가율 유동자산증가율 매출액증가율 순이익증가율 자기자본증가율 유형자산증가율 기타 변수(8)

자산 (천원) 자본 (천원) 비유동자산 (천원) 매출액 (천원) 영업이익 (천원) 당기순이익 (손실) (천원) 현금 및 현금성자산 (천원) 법인세비용차감전 (계속사업)손익 (천원)

T-Test (197H)

[丑6]

총자본순이익률 총자본영업이익률 <mark>총자산순이익률</mark> <mark>총자산영업이익률</mark> 자기자본순이익률 매출액총이익률

총자본회전율 자기자본회전율 총자산회전율

수익성 변수(6)

활동성 변수(3)

안정성 변수(4)

차입금의존도 비유동장기적합률 CASH FLOW 대 부채비율 OCF 대 유동부채 기타 변수(6)

자산 (천원) 매출액 (천원) 영업이익 (천원) 당기순이익 (손실) (천원) 현금 및 현금성자산 (천원) 법인세비용차감전(계속사업)손익 (천원)

깐낀 샌택법, 후낀 제거법, 단계꺽 샌택법 (9개)

구익샘 변구



총자본영업이익률 총자산순이익률 총자산영업이익률 자기자본순이익률 안껑갱 변구



차입금의존도 비유동장기적합률 OCF 대 유동부채 활동생 변구



자기자본회전율 총자산회전율

다풍공앤앵

보통 VIF가 10이 넘으면 다중 공선성이 있다고 판단

[丑7]

	VIF Factor	features
0	3.171445	총자산회전율
1	2.781158	총자산영업이익률
2	2.536515	총자산순이익률
3	2.483940	차입금의존도
4	2.291957	총자본영업이익률
5	2.271203	비유동장기적합률
6	2.260316	자기자본회전률
7	1.173913	자기자본순이익률
8	1.040347	OCF 대 유동부채



모두 통과

데이터 스케일링

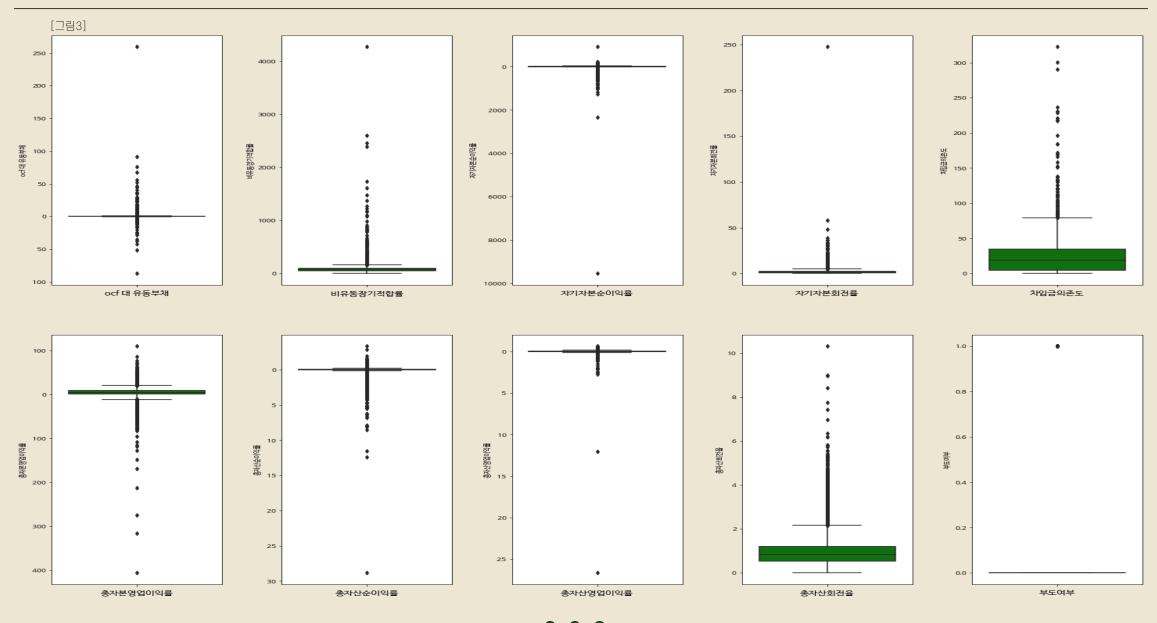
정상기업, 부도기업 비율 맞취 층화 무작위 추출

[⊞8]

y_ train							
0.0 12810							
1.0	1.0 183						
	y_ test						
0.0 5221							
0.0	5221						



Robust scaler



Robust scaler

[⊞9]

	총자산 순이익률	부채비율	차입금 의존도	총자본 영업이익률	고정비율	Cash flow 대 매출액 비율	자기자본 회전율	총자산 영업이익률	운전자본 회전율	Ocf 대 유동부채	총자본 증가율	자기자본 순이익률	부도여부
count	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00	17591.00
mean	-0.023681	126.415145	21.741482	3.717373	0.614479	-30.366664	1.990701	0.028559	0.381371	0.303268	12.914385	-3.266259	0.010801
std	0.425431	0.425431	19.815872	11.843383	0.368341	2364.589694	2.880039	0.252665	6.890418	2.858432	47.187345	88.517143	0.103368
Min	-28.844252	-28.844252	0.00000	-406.4600	-10.057269	-243519.2200	0.00000	-26.682140	0.00000	-87.348397	-95.43000	-9535.6400	0.00000
25%	-0.010763	-0.010763	4.46000	0.4700	0.418255	-0.1200	0.82000	0.004611	0.00000	-0.025346	162000	-1.34000	0.00000
50%	0.028889	0.028889	18.8000	3.9400	0.583949	5.150000	1.44000	0.037800	0.02000	0.156276	5.94000	5.28000	0.00000
75%	0.069514	0.069514	34.3500	8.4900	0.772152	11.16500	2.44000	0.080275	0.13000	0.444592	18.54000	12.3000	0.00000
max	33.364062	3.364062	322.7100	109.7100	17.781073	8064.33000	247.81000	0.678386	563.6900	259.726413	2512.3500	917.14000	1.00000



Robust scaler 사용한 이유

min max값이 q1, q2, q3에 비해 매우 극단적이므로 이상치의 영향을 최소화하는 robust scaler 사용

은행의 관점에서 정상 기업을 부도기업으로 오분류 하는 것보다 부도 기업을 정상 기업으로 오분류 하는 것이 치명적이기 때문에 recall 값 고려 불균형 데이터임을 감안하여 f1-score도 함께 높이는 방향으로 진행

[표11] 모델 결과 비교표

임 상위 모델 기 차상위 모델

ROS: Random Over Sampling RUS: Random Under Sampling

LR	SMOTE	30%	60%	10%	BLSMOTE	ROS	RUS
recall	0.8421	0.6316	0.6842	0.6053	0.8421	0.7895	0.8684
F1-score	0.132	0.1627	0.1667	0.466	0.1557	0.1207	0.117
ROC AUC	0.8726	0.7823	0.8065	0.7966	0.8721	0.8331	0.8634

DT (max_depth=5)	SMOTE	30%	60%	10%	BLSMOTE	ROS	RUS
recall	0.8684	0.6316	0.6842	0.6053	0.8421	0.7895	0.8684
F1-score	0.132	0.1627	0.1667	0.4466	0.1557	0.1207	0.117
ROC AUC	0.8726	0.7823	0.8065	0.7955	0.8721	0.8331	0.8634

Modeling

RF	SMOTE	30%	60%	10%	BLSMOTE	ROS	RUS
recall	0.5526	0.4211	0.4474	0.3158	0.4211	0.1053	0.9474
F1-score	0.2545	0.3232	0.3063	0.3158	0.3232	0.1569	0.1261
ROC AUC	0.7611	0.7041	0.7156	0.6542	0.7041	0.5513	0.9023

Xgb	SMOTE	30%	60%	10%	BLSMOTE	ROS	RUS
recall	0.8684	0.6053	0.7632	0.5263	0.7105	0.7895	0.8684
F1-score	0.1626	0.21	0.1908	0.3604	0.2298	0.2214	0.1218
ROC AUC	0.8861	0.7799	0.8475	0.7555	0.8308	0.8658	0.8666

Lightgbm	SMOTE	30%	60%	10%	BLSMOTE	ROS	RUS
recall	0.6316	0.1939	0.1493	0.3947	0.5	0.3421	0.8421
F1-score	0.2201	0.5	0.2326	0.33529	0.3042	0.325	0.1262
ROC AUC	0.7934	0.7387	0.7469	0.6928	0.7402	0.6669	0.8583

학습 데이터에만 오버샘플링 SMOTE 적용



[丑10]

	SMOTE 10%			
Before	counts of label '1'	152		
	counts of label '0'	13920		

After	counts of label '1'	1392	
	counts of label '0'	13920	

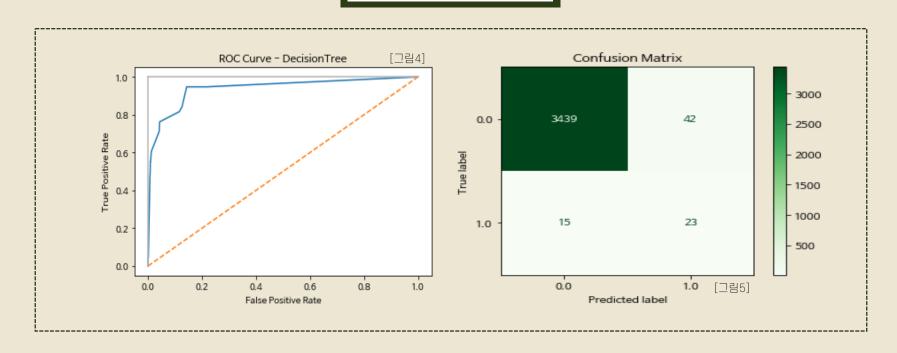
	SMOTE 30%		
	counts of label '1'	152	
Before	counts of label '0'	13920	

After	counts of label '1'	4176	
Aitei	counts of label '0'	13920	

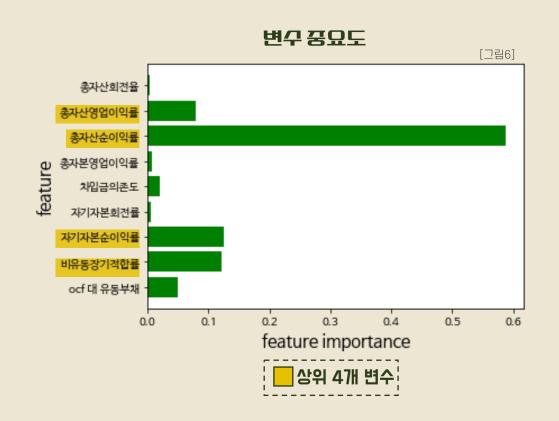
Decision Tree Classifier recall 값: 0.6053, f1-score 값: 0.4466

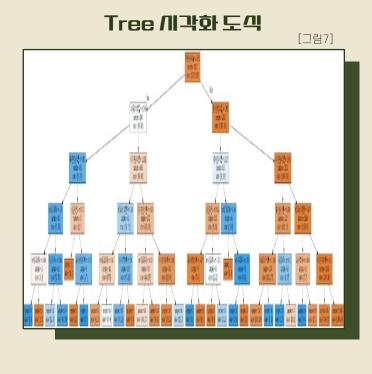
Accuracy: 0.9838022165387894 Precision: 0.35384615384615387 Recall: 0.6052631578947368 F1-Score: 0.44660194174757284 ROC AUC: 0.7965988297373712 Confusion Matrix: [[3439 42]

15 23]]



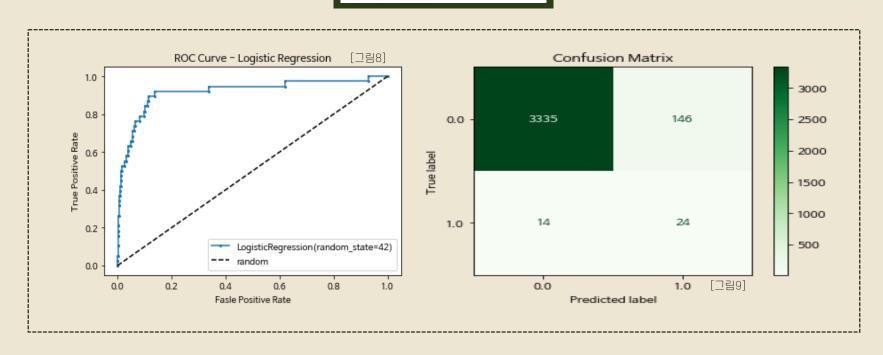
오버샘플링 SMOTE 10%





Logistic Regression recall 과 : 0.6316, f1-score 과: 0.2308

Accuracy: 0.9545325376527423 Precision: 0.1411764705882353 Recall: 0.631578947368421 F1-Score: 0.23076923076923078 ROC AUC: 0.7948184883351729 Confusion Matrix: [[3335 146] [14 24]]

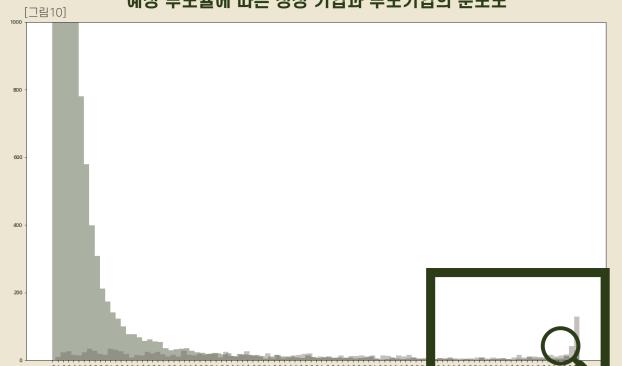


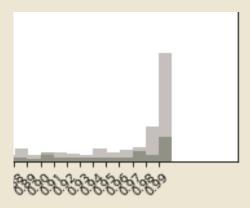
기업 등급화 (9등급)

로짓 분석 기반 예상 부도확률 산출 🗼 9등급 cutoff 선정 🛶 나머지 구간 균등하게 8등분하여 기업을 총 9등급으로 구분

• • •

예상 부도율에 따른 정상 기업과 부도기업의 분포도

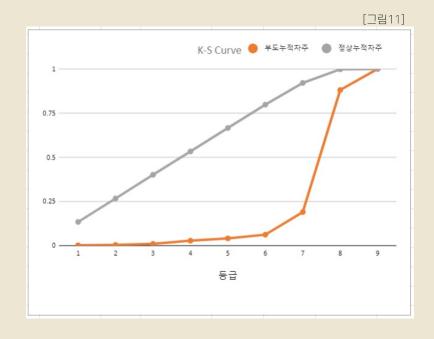




정상 기업과 부도 기업간의 빈도 차이가 크게 나는 구간인 <u>0.98이</u>상

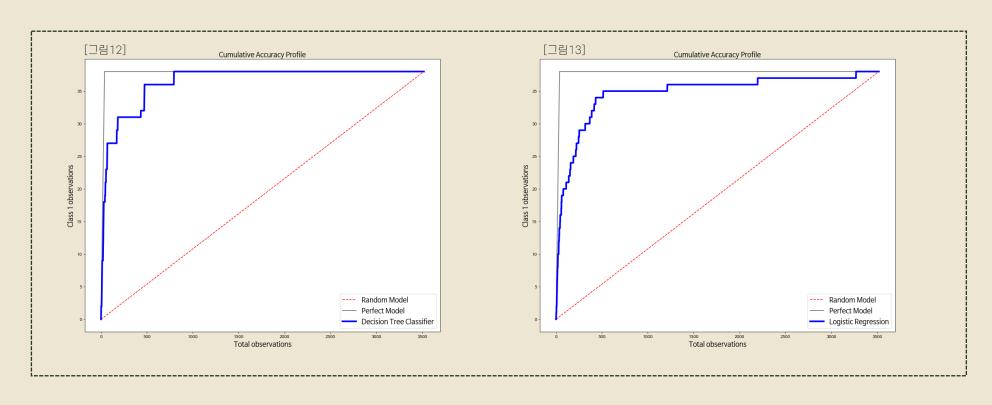
변별력 검증 - KS Statistics

예상부	도확률	=7	시케니시르	нг	T-1 1 1	누적분	포비율	К 0
From	То		실제부실률	부도	정상	부도	정상	K-S
0.000006	0.0083	1	0.002	3	2326	0.002	0.134	0.132
0.0083	0.0142	2	0.003	3	2326	0.004	0.267	0.263
0.0142	0.0203	3	0.004	9	2320	0.01	0.401	0.391
0.0203	0.0276	4	0.003	25	2304	0.028	0.533	0.505
0.0276	0.0376	5	0.009	19	2310	0.041	0.666	0.625
0.0376	0.0564	6	0.016	31	2298	0.062	0.798	0.736
0.0564	0.1278	7	0.079	184	2145	0.19	0.921	0.731
0.1278	0.9799	8	0.379	996	1333	0.881	0.998	0.117
0.9799	1	9	0.227	171	37	1	1	1



KS통계량: 0.736 → 4~7등급의 변별력이 우수함을 확인

변별력 검증 - Accuracy Ratio(AR)



AR for Decision Tree Classifier: <u>0.92888</u> AR for Logistic Regression: <u>0.83882</u>

굅// 과쟁 모형



기업 여신 심사 웹 구현

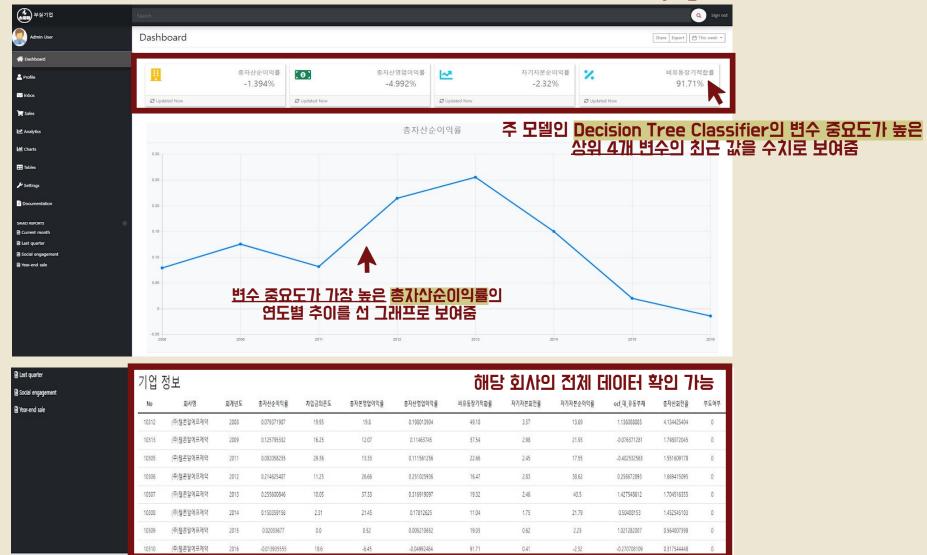
<Main page>



기업 여긴 낌가 웹 구현

사용자가 회사명 🔾 을 입력하면.

<Admin Dashboard page>



기업 여신 심사 웹 구현

<Model page>



사용자가 변수 값을 입력하면 차기 연도의 예상 부실 확률과 기업 등급이 출력됨

기업 여신 심사 웹 구현

- ☑ Decision Tree Classifier 모델을 주 모델로 사용하여 예상부도율을 산출하고,
 Logistic Regression 모델을 보조 모델로 사용하여 부도 모형의 설명력을 보충함 (모형의 회귀계수 사용)
- ⊘ 등급화: 일정 구간별로 기업별 등급 부여 후 대출

[丑13]

변수	회귀 계수
차입금 의존도	0.894026
총자본영업이익률(-)	0.839256
총자산회전율(-)	0.430028
자기자본회전율(-)	0.256373
총자산영업이익률	0.138856
자기자본순이익률(-)	0.094466
비유동장기적합률(-)	0.053321
총자산순이익률(-)	0.037343
OCF 대 유동부채(-)	0.00415

기대 효과

• 부실채권 방지를 통한 은행의 기업 여신 리스크 관리 가능

■ Logistic Regression 모델의 계수를 통해 부도 모형의 변수별 설명력을 보충함으로써 대출이 힘든 기업의 컨설팅 제공 가능



• • •

한계점

- 성능(Recall) 향상을 위한 모델 추후 개선 필요
 - 데이터 순도를 높일 수 있는 데이터 전처리에 대한 연구 필요



Reference

- 1.이상봉(2013) 부실기업의 재무적 특성에 관한 연구
- 2.박종원 안성만(2014) 재무비율을 이용한 부도예측에 대한 연구
- 3.최정원, 오세경, 장재원(2017) 빅데이터와 인공지능 기법을 이용한 기업 부도예측 연구
- 4.권혁건, 이동규, 신민수(2017) RNN(Recurrent Neural Network)을 이용한 기업부도예측모형에서 회계정보의 동적 변화 연구
- 5.차성재, 강정석(2018) 딥러닝 시계열 알고리즘 적용한 기업부도예측모형 유요성검증 6.이현상, 오세환(2020) 시계열 예측을 위한 LSTM 기반 딥러닝: 기업 신용평점 예측 사례
- 7.이경수, 임희석(2018) 기계학습을 이용한 저축은행 부실 예측모형 검증
- 8.남기정, 이동명, 진로(2019) 비재무정보를 이용한 창업기업의 부실요인에 관한 실증연구
- 9.이현상, 오세환(2020) 시계열 예측을 위한 LSTM 기반 딥러닝: 기업 신용평점 예측 사례



