
통계적 모델링을 통한 신용평가

- 재무모형과 계량 비재무모형 중심으로



Index

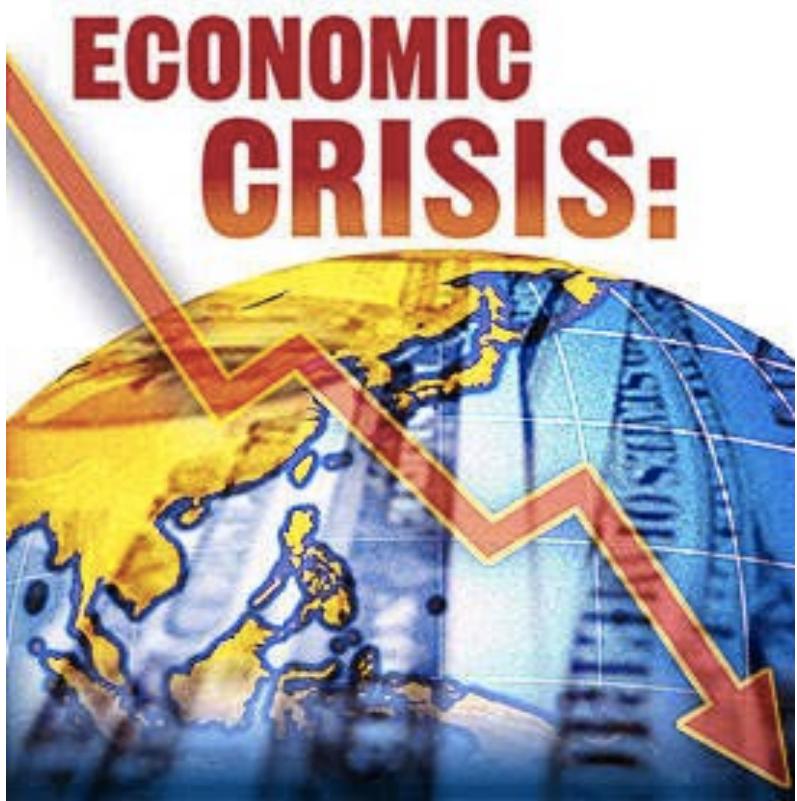
0. Introduction
1. Environment Setting
2. Feature Engineering
 - Variable Select
 - Algorithms
3. Explore Data Analysis (EDA)
4. Model Result
5. Expected Effects
6. REFERENCE



0.

Introduction

이름	담당 업무
강준영	Text Mining, Modeling
박형권	Feature Engineering Modeling
허성욱	Bigdata Platform, Domain Knowledge



✓ 막연한 기대감

- 주택 가격이 계속 오를거라는 기대감으로 인해, 대출 위험 대상인 서브프라임(SubPrime) 등급에게도 돈을 빌려주기 시작
- 부실채권 비율이 높아지려 하자, 그 위험성을 떠넘기려고 또 하나의 증권화를 시켜서 유통 (MBS)
- 서로 위험성이 연결 돼 있다보니, 한 회사가 망한 후 연쇄 파산이 일어남

➤ 신용등급을 객관적으로 평가해 대출을 했으면 서브프라임사태가 나지 않았을 것. 결국, 숫자가 아닌 ‘감’에 의존해서 2008년 세계금융위기 촉발

LEHMAN BROTHERS

Goldman
Sachs

JPMorganChase



Merrill Lynch

BEAR
STEARNS

파산

LEHMAN BROTHERS



Goldman
Sachs

JPMorganChase



Bank of America

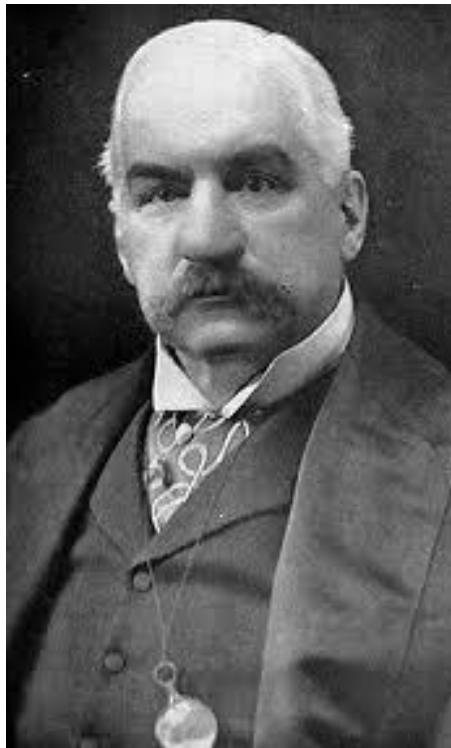


Merrill Lynch

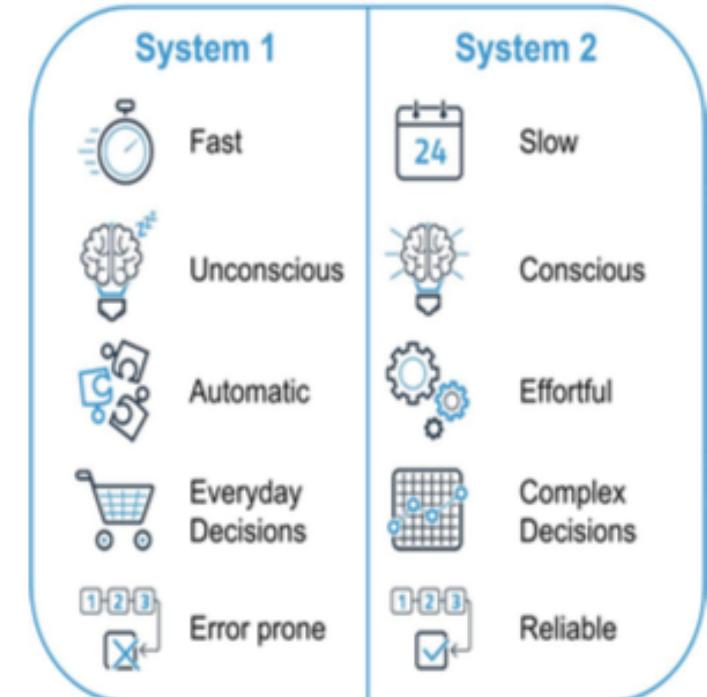
BEAR
STEARNS

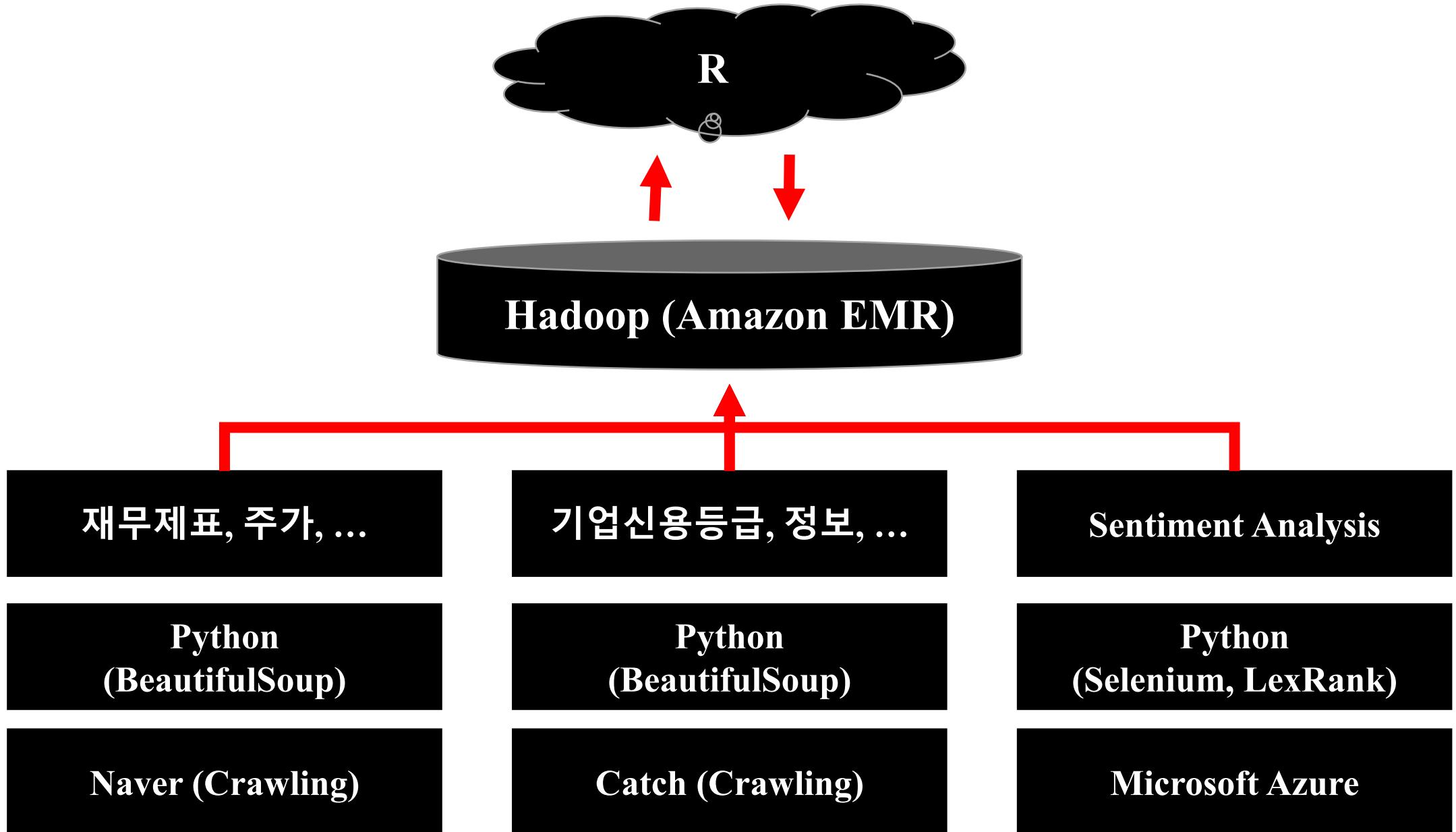
신용이 없는 사람과는 그가 비록
예수님 나라에서 발행한 채권을
내놓아도 거래하지 않았다.

John Pierpont Morgan (J.P.Morgan)



사람들은 인간의 판단을 지나치게 신뢰한다.
Daniel Kahneman





1. Environment Setting

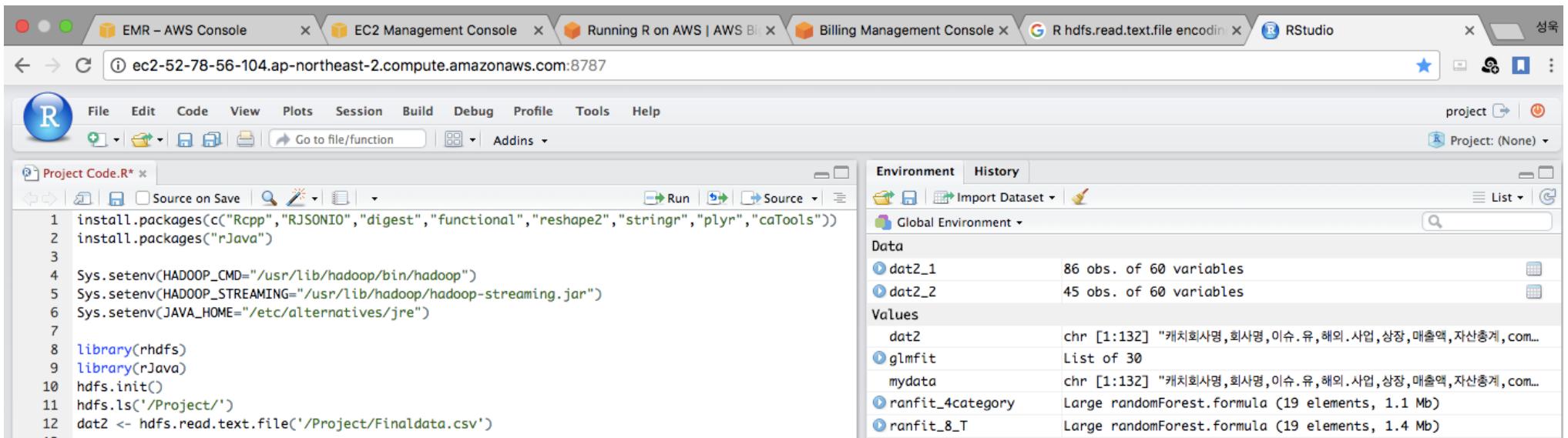
The screenshot shows the AWS EC2 Instances page. The left sidebar has a tree view with 'EC2 대시보드' at the top, followed by '이벤트', '태그', '보고서', '제한', '인스턴스' (selected), '인스턴스' (under '인스턴스'), '스팟 요청', '예약 인스턴스', and '전용 호스트'. The main area has tabs '인스턴스 시작' (selected), '연결', and '작업'. A search bar says '태그 및 속성별 필터 또는 키워드별 검색'. Below is a table with columns: Name, 인스턴스 ID, 인스턴스 유형, 가용 영역, 인스턴스 상태, 상태 검사, 경보 상태, 퍼블릭 DNS(IPv4), and IPv4 퍼블릭 IP. The table contains three rows:

Name	인스턴스 ID	인스턴스 유형	가용 영역	인스턴스 상태	상태 검사	경보 상태	퍼블릭 DNS(IPv4)	IPv4 퍼블릭 IP
Master	i-0d9b91090535654...	r3.xlarge	ap-northeast-2c	running	2/2 검사 통과	없음	ec2-52-78-56-104.ap-n...	52.78.56.104
Slave1	i-0441e79b632c85ab0	r3.xlarge	ap-northeast-2c	running	2/2 검사 통과	없음	ec2-13-124-222-56.ap-...	13.124.222.56
Slave2	i-030c5ce15b3fee5e1	r3.xlarge	ap-northeast-2c	running	2/2 검사 통과	없음	ec2-13-124-222-75.ap-...	13.124.222.75

➤ Amazon Web Service에서 서버 3대를 빌려서, RHadoop 구축.

1.

Environment Setting



➤ RStudio 포트(8787)를 열어 같이 작업할 수 있게 환경 설정.

2.

Feature Engineering

2-1. 재무 모형

- ✓ 단순 값이 아니라 비율을 구하는 것이 중요.

	A 기업	B 기업
매출액	46,845억원	46,845억원
영업이익	12,216억원	12,216억원
당기순이익	8,344억원	8,344억원

2.

Feature Engineering

2-1. 재무 모형

- ✓ 단순 값이 아니라 비율을 구하는 것이 중요.

	A 기업	B 기업
매출액	46,845억원	46,845억원
영업이익	12,216억원	12,216억원
당기순이익	8,344억원	8,344억원
사원수	10,000명	1,000명

2.

Feature Engineering

2-1. 재무 모형

- ✓ 단순 값이 아니라 비율을 구하는 것이 중요.

	A 기업	B 기업
이자비용	57억원	143억원
금융기관차입금	42억원	219억원
유동부채	134억원	350억원

2.

Feature Engineering

2-1. 재무 모형

- ✓ 단순 값이 아니라 비율을 구하는 것이 중요.

	A 기업	B 기업
이자비용	57억원	143억원
금융기관차입금	42억원	219억원
유동부채	134억원	350억원
매출원가	200억원	1000억원

지표 구분	비율명	대기업
성장성	매출액증가율	$(당기매출액/전기매출액) \times 100 - 100$
성장성	자기자본증가율	$(당기말자본총계/전기말자본총계) \times 100 - 100$
수익성	기업순이익률	$(당기순이익+이자비용)/(전기자산총계+당기자산총계) \times 200$
수익성	매출액순이익률	$(당기순이익/매출액) \times 100$
수익성	총자본순이익률	당기순이익/(전기자산총계+당기자산총계)×200
수익성	시내유보율	사내유보/처분가능이익잉여금×100
수익성	수지비율	총비용/총수익×100
수익성	유보액/납입자본	$(자본잉여금+이익잉여금-배당금)/자본금 \times 100$
수익성	적립금비율	$(자본잉여금+이익잉여금)/자본총계 \times 100$
안정성	금융비용 대 매출액	이자비용/매출액×100
안정성	금융비용 대 총부채	이자비용/총부채×100
안정성	수정자기자본비율	$((총자산-이연자산-무영자산)-$ $(단기차입금+유동성장기부채+사채+장기차입금+매입채무+부가세예수금+미지급법인세+미지급비용))/(총자산-이연자산-무영자산) \times 100$
안정성	순금융비용 대 매출액	$(이자비용-이자수익)/매출액 \times 100$
안정성	영업이익 대 금융비용	영업이익/이자비용×100
안정성	유보액 대 총자산	$(자본잉여금+이익잉여금-배당금)/자산총계 \times 100$
안정성	고정비율	비유동자산/자기자본×100
안정성	무채총계 대 매출액	무채총계/매출액×100
안정성	비유동부채 대 총자산	비유동부채/총자산×100
안정성	유동부채 대 총자산	유동부채/총자산×100
안정성	유동부채비율	유동부채/자본총계×100
안정성	자기자본비율	자본총계/자산총계×100
안정성	재고자산 대 유동자산	재고자산/유동자산×100
안정성	차입금의존도(%)	총차입금/자산총계×100
이상치 모형변수	금융비용부담율 증가분	당기(금융비용/매출액)-전기(금융비용/매출액)
활동성	총자산회전율	매출액/총자산
활동성	상(제)품회전율(회)	매출액/(전기상품+전기제품+전기반제품+당기상품+당기제품+당기반제품) ×2
활동성	영업자산회전율	매출액/(전기영업자산+당기영업자산) ×2
활동성	운전자금회전율	$1/(1/\text{매출채권회전율} + 1/\text{재고자산회전율} - 1/\text{매입채무회전율})$
활동성	자본금회전율(회)	매출액/(전기자본금+당기자본금) ×2
활동성	재고자산회전율	매출액/(전기재고자산+당기재고자산) ×2
CashFlow 관련지표	손익활동 CF/매출액	손익활동현금흐름/매출액×100
CashFlow 관련지표	중장기차입금상환계수	$(당기순이익+감가상각비+이자비용)/(전기유동성장기부채+이자비용)$

2.

Feature Engineering

2-2. 변수 설정

- ✓ 반응변수 설정 : (AAA → D 까지 총 22등급이 존재)
 - 본 프로젝트에서는 전문신용평가사들의 기준, 관련 논문들의 기준에 따라 4개의 범주, 혹은 8개의 범주로 변환
- ✓ 비재무 모형 : 계량모형
 - 사원수, 복지 후생, 감정 분석, 해외 사업 유무 ...
- ✓ 비재무 모형 : 비계량 모형 – 애널리스트들의 정성적 평가 모형
 - 본 프로젝트에서는, 재무모형과 계량 비재무모형에 초점을 맞추어 분석을 진행
- ❖ 파생변수 포함 총 58개의 변수로 모델링

2.

Feature Engineering

2-3. Algorithms I : Logistic Regression

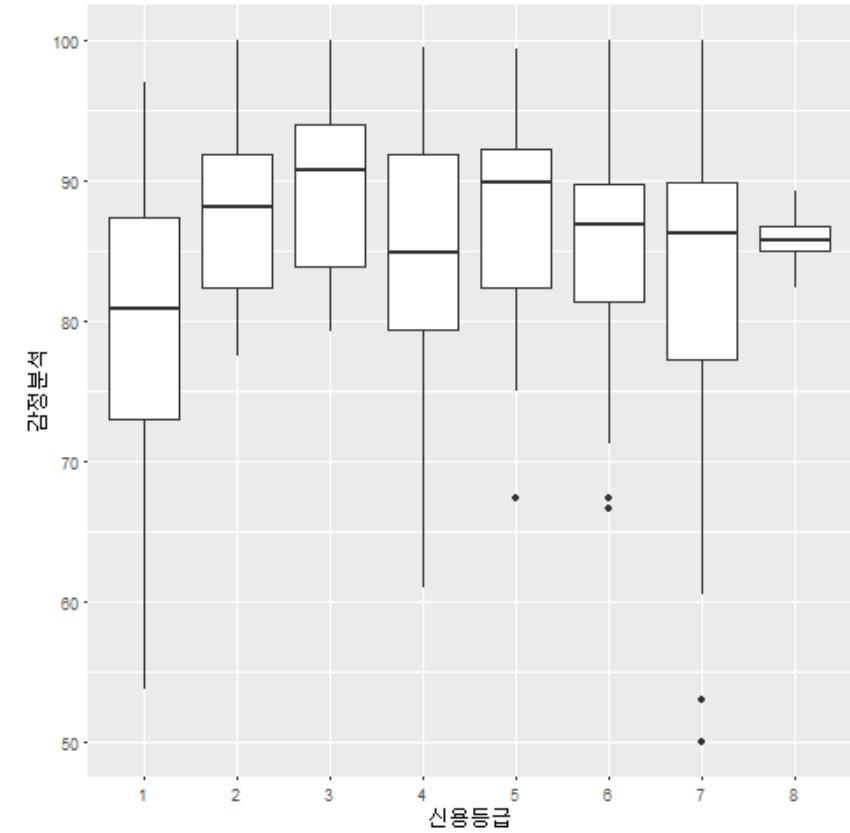
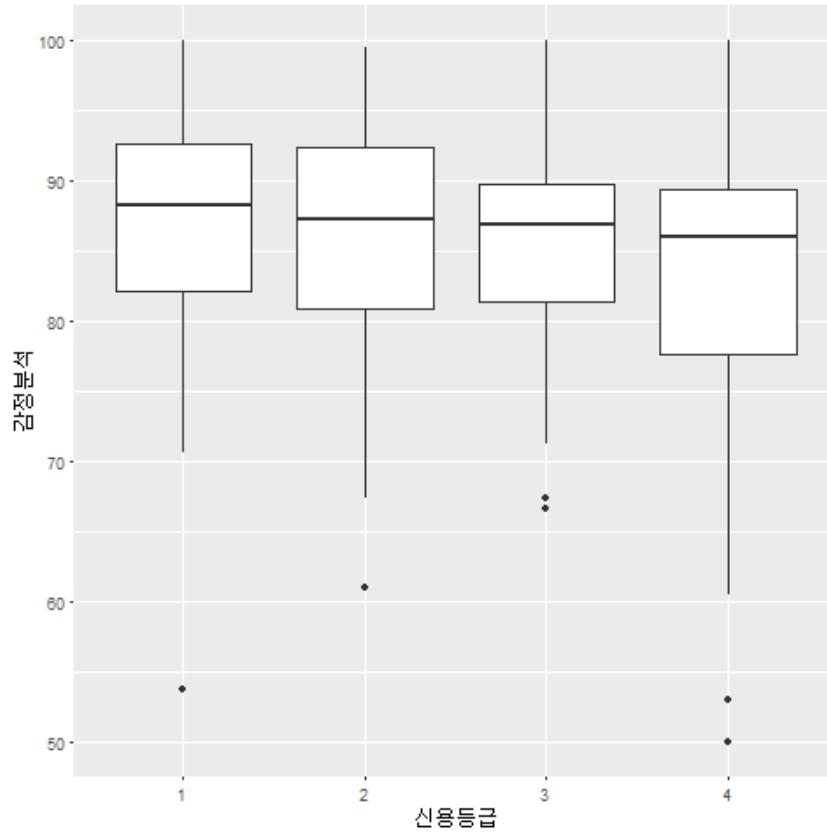
- ✓ 문제점 : 부도 기업에 대한 정보가 없다. (부도 : 1 , 생존 : 0)
 - 다항 분포에 대한 모델링으로 접근을 시도
 - 부도 확률을 반응변수로 하다보니, 특정 등급 사이에서의 차이가 미미해 구별 되지 않을 가능성이 존재.
 - 가장 낮은 등급 (B-)을 부도(1)로 놓고 모델링을 해도, 반응변수가 확률로 나오다 보니 모델평가시 주관성이 들어갈 수 있음.

2-4. Algorithms II : Random Forest & Binomial Logistic Regression

- ✓ Random Forest로 모델링을 한 후 예측력 조사.
- ✓ Binomial Logistic Regression 과 결과 비교

3.

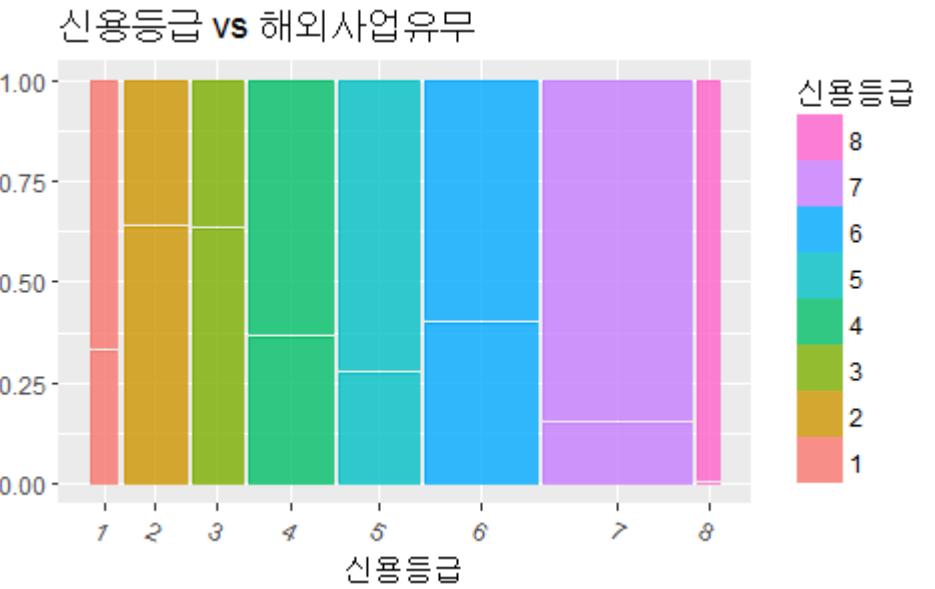
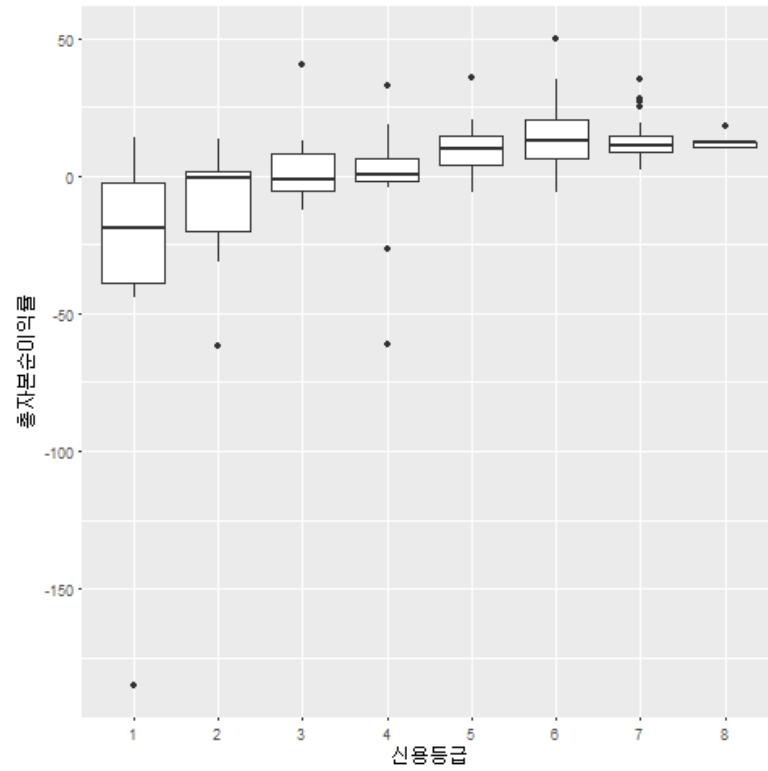
Explore Data Analysis (EDA)



Sentiment Analysis (정성적 분석) 결과 신용등급간 **유의미한 차이**가 없음
➤ 인간의 심리가 '**모델**', 상으로는 영향이 없다는 가설을 반증

3.

Explore Data Analysis (EDA)



4.

Modeling Result (4개 범주화)

	1	2	3	4
1	13	16	1	1
2	10	18	6	3
3	1	5	13	6
4	0	1	4	33

Accuracy **0.58778626**



4.

Modeling Result (8개 범주화)

	1	2	3	4	5	6	7	8
1	0	2	2	1	0	0	1	0
2	1	4	2	6	1	0	0	0
3	1	1	1	3	4	1	0	0
4	0	4	1	5	6	3	0	0
5	0	1	1	3	7	5	1	0
6	0	1	0	1	3	12	8	0
7	0	0	0	0	1	4	28	0
8	0	0	0	0	0	0	1	4

Accuracy 0.465648855

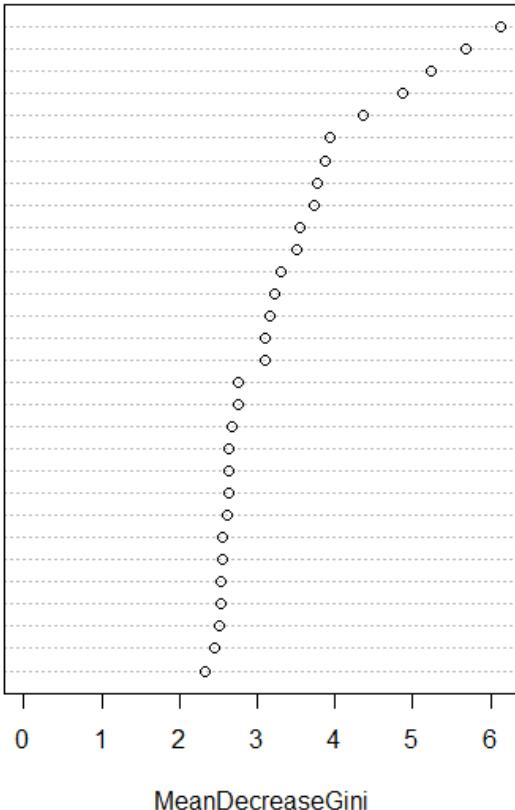
Effective Accuracy : 0.763358779



4.

Modeling Result (Random Forest vs Logistic Regression)

EBITDA
 총자본경상이익률
 매출액증가율
 자본금순이익률
 이의영여금.유동자산
 순금융비용.대.매출액
 자본금회전율
 자산총계
 총자본순이익률
 매출액
 X.부채총계.이익영여금..유형자산
 기업순이익률
 매출액순이익률
 부채총계.대.매출액
 유보액.납입자본
 고정비율
 총자산회전율
 자기자본비율
 적립금비율
 금융비용.대.매출액
 금융기관자입금.부채
 KED현금비율
 유보액.대.총자산
 현금성당좌자산.자산
 영업자산회전율
 금융비용.대.총부채
 수지비율
 비유동부채.대.총자산
 부채.유형자산
 현금성당좌자산.유동자산



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.277e-01	9.411e-01	0.348	0.728746
매출액	2.068e+00	9.100e-01	2.273	0.026153 *
자산총계	-2.949e-01	1.235e-01	-2.388	0.019706 *
매출액증가율	1.566e-03	9.744e-04	1.607	0.112572
기업순이익률	-1.809e-02	4.527e-03	-3.995	0.000159 ***
수지비율	-3.764e-05	1.656e-05	-2.273	0.026126 *
고정비율	8.055e-03	2.343e-03	3.438	0.000996 ***
비유동부채.대.총자산	-2.572e-02	1.030e-02	-2.497	0.014932 *
유동부채비율	-8.553e-03	2.362e-03	-3.621	0.000557 ***
자기자본비율	-2.335e-02	8.074e-03	-2.892	0.005109 **
차입금의존도...	-2.461e-02	1.264e-02	-1.946	0.055725 .
총자산회전율	-6.359e-01	2.829e-01	-2.248	0.027756 *
자본금회전율	8.769e-04	6.241e-04	1.405	0.164494
KED현금비율	5.572e-04	3.775e-04	1.476	0.144495
현금성당좌자산.유동자산	-8.274e-03	4.176e-03	-1.981	0.051538 .
현금성당좌자산.자산	1.239e-02	6.476e-03	1.913	0.059844 .
금융기관자입금.부채	9.088e-03	3.706e-03	2.452	0.016735 *

❖ 유동부채비율, 차입금의존도 제외하고 일치

➤ 나이스디앤비 기준 - 규모: 2 수익성: 2 안정성: 7 활동성: 1

4. Modeling Result (다른 논문들의 모델링 결과)

	<i>Sample</i>	<i>Method</i>	<i>Accuracy</i>
<i>Kim and Ahn, 2012</i>	1295	SVM	67.98%
<i>Ogut et al., 2012</i>	18	MDA, SVM	65.11%
<i>Lee et al., 2004</i>	555	Discriminant Analysis	68.32%
<i>이정우 and 서용무, 2015</i>	1158	Decision Tree, Random Forest	Tree – 64.98%, RF – 71.52%



5.

Expected Effects

기업신용에 중요한
평가 요소를 탐색하여
**기업 경영 방향 설정 및
마케팅
에 도움**

전문적이고 일관된 판단 능력
을 기초로 하여 신용위험에 대
한 적절한 정보를 투자자에게
제공해
정보의 비대칭성 완화

저성장, 저금리,
새로운 회계기준 도입으로
인한
리스크관리
어려움 일부 해소

6.

REFERENCE

- 안현철, 김경재. “다양한 다분류 SVM을 적용한 기업채권평가”, 2009
- “신용등급의 핵심 요소에 대한 설명서” – Standard & Poors (S&P), 2013
- “KED 신용평가 방법론” – CONTACT 평가정책팀
- 이성우, 서용무. “주요 재무변수들과 그 변화율을 이용한 기업신 평가 예측모델”, 2015
- “전자산업평가방법론” – 한국신용평가, 2013
- “기업신용평가체제” – 이크레더블, 2016
- 윤우영. “신용평가의 필요성, 그 패러다임의 변화를 중심으로”
- 김상봉, 조경준. “부도예측모형을 이용한 기업부실화의 원인분석”
- 이영섭, 박주완. “로지스틱 회귀모형을 이용한 기업 인적자원 관련 변수와 기업 신용점수와의 관계 연구: HCCP 1차년도 자료를 중심으로”, 『패널리뷰』, 2008
- NICE, D&B “나이스디앤비 기업신용평가 방법론”
- 이건창, 한인구, 김명종. “통계적 모형과 인공지능 모형을 결합한 기업신용평가 모형에 관한 연구”, 1996
- 이승민, 조근희. “2014년 고성장 ICT기업 현황 및 특징” 『한국통신학회 학계종합학술발표』, 2016

Q & A



감사합니다.

