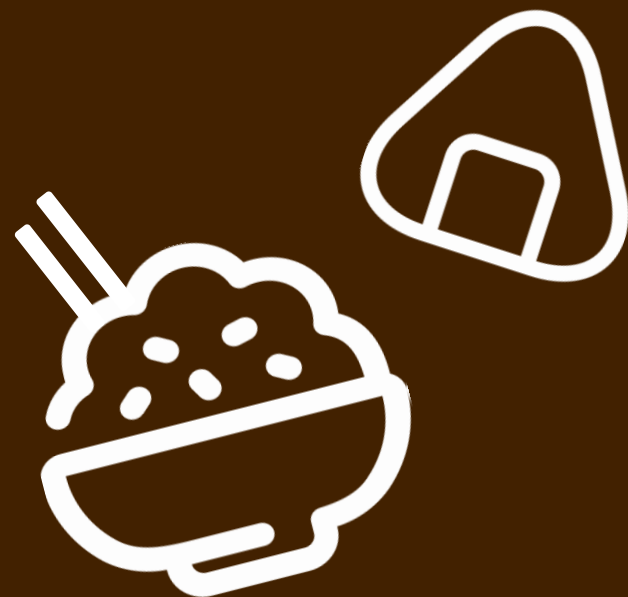


『 딥러닝 기반 부실기업 예측모형에 관한 연구 』

조재혁
안은주
김성수

2021.02.17

오니기리와
여규동



Index

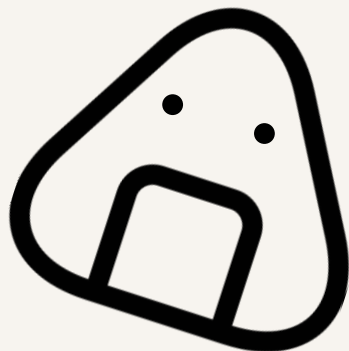
1. 팀 소개
2. 논문 선정 이유
3. 서론 - 선행연구
4. 본론 - 데이터, 변수 설정, 결과
5. 연구 방법 및 검증
6. 결론 - 연구 요약, 기여효과 및 한계점

팀 소개



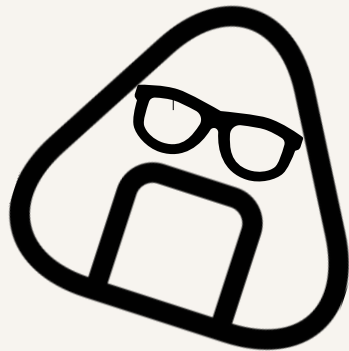
이름 : 여 규동

- 온화한 리더
- 팀 이름 부끄러워함
- 코딩 & 금융



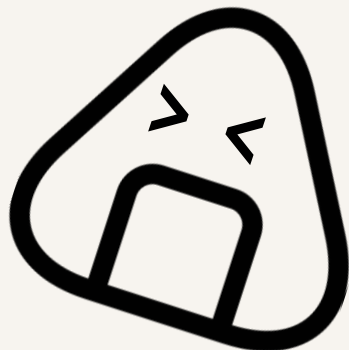
이름 : 김 윤지

- 오니기리즈 첫째
- 세상에 찌뚱
- 아이디어 & 통계



이름 : 박 준배

- 오니기리즈 둘째
- '제가 할 게요!' 담당
- 코딩 & 정리



이름 : 김 민지

- 오니기리즈 막내
- 해맑은 간식 담당
- 회계 도메인 담당

논문 선정 이유

딥러닝 기반 부실기업 예측모형에 관한 연구

조재혁 *

안은주 **

김성수***

< 요약 >

부실기업 예측은 회계와 재무 분야에서 중요하게 다루어져 온 연구 주제이다. 특히, 급변하는 기업 환경과 최근 COVID-19 대유행으로 국내의 많은 기업이 기업 부실화로 재무적 어려움에 직면하고 있어 그 연구의 필요성이 더욱 강조되고 있다. 대표적인 관련 연구로는 기업부도 예측이 있지만, 부도기업은 사실상 사업활동을 중단한 기업으로 계속기업 간에 어떠한 기업이 부실 징후를 보이는지를 판단하는 기준으로 부적합하다는 한계점이 존재한다. 본 연구는 부실기업의 범주 중 하나인 한계기업을 그 예측대상으로 선정하였다. 한계기업은 3개년도 연속 이자보상비율이 1 미만인 기업으로, 사업활동을 영위하고 있으나 적정 수준의 이익을 지속적으로 확보하지 못하고 있는 부실기업을 의미한다. 본 연구에서는 한계기업 예측을 위한 방법으로 딥러닝 기법을 활용하였다. 딥러닝은 다양한 분야에서 그 우수성을 인정받아 최근 주목받고 있는 머신러닝 기법의 하나이지만 한계기업 예측을 위한 연구에서는 적용된 바가 없다. 본 연구는 여러 재무비율 변수를 독립변수로 하여 딥러닝 기법 중 RNN과 CNN를 적용하고, 선행연구에서 예측력이 뛰어나다고 보고된 머신러닝 앙상블 모형들과 그 성과를 비교하였다. 2017~2019년의 기업 데이터를 학습용 및 테스트용 데이터로 설정하여 분석한 결과, RNN-LSTM, RNN-GRU, CNN이 재현율(Recall)의 관점에서 우수한 성과를 보이는 것으로 나타났다. 그러므로 딥러닝 모형이 한계기업 예측에서도 널리 사용될 수 있는 분석방법이 될 것으로 기대된다.

1. 최초 팀 회의에서 딥러닝을 활용하여 한계기업의 회생 가능성을 탐구하는 이야기가 나옴.

2. 펜데믹이 완화됨에 따라 한계기업 중 정상기업으로 회생 가능성이 있는 기업들이 생길 것으로 예상함.

논문 서론

1. 기업 부실에 대한 예측 시도

- Altman의 5가지 재무 변수를 활용한 다중 판별 분석 기반 부도 예측
- Ohlson의 로지스틱 회귀 모형 기반 부도 예측
- Zmijewski의 프로빗 모형 기반 부도 예측

3. 해당 한계점을 보완하기 위해

계속 기업 중 부실기업으로 볼 수 있는 새로운 정의(한계기업)를 내리고,
통계에 대한 엄격한 가정을 필요로 하지 않는 방법을 활용해 종속변수를 설명하고자 한다.

2. 기존 연구의 한계점

- 실무적 한계점 : 계속기업 중에서 관리 대상인 기업들을 판단하고, 부실기업을 예측하는 데에는 부적절하다.
- 방법론적 한계점 : 엄격한 통계적 가정을 만족해야 한다.

4. 앙상블 모형과 딥러닝 기법을 적용하여

기존 한계기업 연구에서 많이 사용된 머신러닝 기법과는 다른 방법으로 진행하고자 한다.

논문 서론 - 한계기업에 대한 정의

1. 오영택, 정성필(2011, 152)

정의 : 경영이 부실하고 재정적 위기에 빠진 기업

2. 송양호, 정성필(2011, 182)

정의 : 영업이익으로 이자를 감당하지 못하거나,
시장 점유율 저하 등으로 경쟁력을 상실한 기업



3년 연속 이자보상
배율이 1 미만인 기업

Defined by. 한국은행



3. 이기영, 우석진(2015, 76)

정의 : 1) 최종 3개년 연속 총 차입금 > 매출액
2) 최종 2개년 연속자기자본 전액 잠식 기업
3) 최종 3개년 연속 이자보상배율 < 1

논문 서론 - 선행연구

머신러닝

- 윤호중. (2019). 머신러닝을 이용한 한국 스포츠산업 기업의 부실 예측에 관한 연구
- Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction.
- Shin, K. S., & Lee, T. S. (2005). An application of support vector machines in bankruptcy prediction model.

SVM, 인공신경망이 다중 판별 분석이나 귀납 추론 등 전통적인 통계 기법에 비해 우수하다.

머신러닝 앙상블

- 민성환. (2014). 개선된 배깅 앙상블을 활용한 기업부도예측.
- 김승혁, & 김종우. (2007). Modified Bagging Predictors를 이용한 SOHO 부도예측.
- 김명종. (2009). 기업부실 예측 데이터의 불균형 문제해결을 위한 앙상블 학습.

단일 인공신경망 모형에 비해 앙상블 모형이 더 우수하다.

딥러닝 CNN

- Yeh, S., Wang, C., & Tsai, M. (2015). Deep belief networks for predicting corporate defaults.
- 권혁건, 이동규, & 신민수. (2017). RNN을 이용한 기업부도예측모형에서 회계정보의 동적 변화 연구.
- 차성재, & 강정석. (2018). 딥러닝 시계열 알고리즘 적용한 기업 부도예측모형 유용성 검증.

기존의 머신러닝 기법보다 RNN과 기업의 재무 데이터를 이미지화한 CNN이 한계기업 예측에 더 우수한 성과를 보인다.


논문 본론 - 데이터, 전처리

☑ 분석 대상

- 2017년~2019년 코스피, 코스닥, 코넥스, 외감기업
33916개 중 결산월이 12월인 비(非)금융기업
- 분석 기간 내에 재무제표의 적절성 판단 지표인
감사의견이 '적정' 이외의 기업을 제외
- 연차 재무제표만 활용해 최종적으로
16813개의 기업을 분석 대상으로 삼음

☑ 전처리

- 결측치 : 중앙값으로 대체
- 영업이익 이자보상배율에서 결측치가 존재하는 경우
해당 기업정보 전체를 제거
- 이상치 처리 방법으로 윈저라이징 사용
> 변수 별 누적확률 0.01, 0.99에 해당하는
값으로 변환

 최종적으로 정상기업과 한계기업의 수를 각각 13818개, 2995개로 집계

논문 본론 - 독립변수 설정

구분	분류	명칭	산식
1	성장성	총자산증가율	$(\text{기말총자산} - \text{기초총자산}) / \text{기초총자산} \times 100$
2		유동자산증가율	$(\text{기말유동자산} - \text{기초유동자산}) / \text{기초유동자산} \times 100$
3		매출액증가율	$(\text{당기매출액} - \text{전기매출액}) / \text{전기매출액} \times 100$
4		순이익증가율	$(\text{당기순이익} - \text{전기순이익}) / \text{전기순이익} \times 100$
5		영업이익증가율	$(\text{당기영업이익} - \text{전기영업이익}) / \text{전기영업이익} \times 100$
6	수익성	매출액순이익률	$\text{순이익} / \text{매출액} \times 100$
7		매출총이익률	$\text{매출총이익} / \text{매출액} \times 100$
8		자기자본순이익률	$\text{순이익} / \text{자기자본} \times 100$
9	활동성	매출채권회전율	$\text{매출액} / \text{매출채권}$
10		재고자산회전율	$\text{매출원가} / \text{재고자산}$
11		총자본회전율	$\text{매출액} / \text{총자본}$
12		유형자산회전율	$\text{매출액} / \text{총자산}$
13		매출액대매출원가	$\text{매출원가} / \text{매출액} \times 100$
14		매출액대판매관리비	$\text{판매관리비} / \text{매출액} \times 100$
15	안정성	부채비율	$\text{부채} / \text{자기자본} \times 100$
16		유동비율	$\text{유동자산} / \text{유동부채} \times 100$
17		자기자본비율	$\text{자기자본} / \text{총자산} \times 100$
18		당좌비율	$\text{당좌자산} / \text{유동부채} \times 100$
19		고정비율	$\text{고정자산} / \text{총자본} \times 100$
20		순운전자본비율	$\text{순운전자본} / \text{총자본} \times 100$
21		차입금의존도	$(\text{장기 및 단기차입금} + \text{사채}) / \text{총자본} \times 100$
22		현금비율	$\text{현금예금} / \text{유동부채} \times 100$

재무상태나 경영성과가 전기 대비
얼마나 증가했는지

전기 대비 이익이나 손실이
얼마나 발생했는지

기업 보유 자산을 당기에 얼마나
효율적으로 운용하고 있는지

기업의 재무상환 능력과
현금 창출 능력

논문 본론 - 독립변수 설정

- ☒ T 검정을 진행하여 P-value가 0.05보다 큰 재무비율 제외
- ☒ 해당 재무 비율을 기반으로 로지스틱 회귀분석 실시
- ☒ 후진 선택법 활용
 - > 우수한 설명력을 갖는 재무 변수 선정
- ☒ Train Data : Test Data = 70 : 30
 - > 데이터 정규화(Normalization)

〈표 3〉 최종 선정된 연도별 재무비율

구분	2017년	2018년	2019년
1	총자산증가율	총자산증가율	총자산증가율
2	● 유동자산증가율*	유동자산증가율	● 유동자산증가율*
3	매출액증가율	● 매출액증가율*	● 매출액증가율*
4	순이익증가율*	● 순이익증가율*	● 순이익증가율*
5	● 영업이익증가율*	● 영업이익증가율*	영업이익증가율
6	매출액순이익률	● 매출액순이익률*	매출액순이익률
7	● 매출총이익률*	매출총이익률	매출총이익률
8	● 자기자본순이익률*	● 자기자본순이익률*	● 자기자본순이익률*
9	매출채권회전율	매출채권회전율	매출채권회전율
10	재고자산회전율	재고자산회전율	재고자산회전율
11	총자본회전율*	총자본회전율	● 총자본회전율*
12	유형자산회전율	유형자산회전율	● 유형자산회전율*
13	매출액대매출원가	● 매출액대매출원가*	매출액대매출원가
14	● 매출액대판매관리비*	● 매출액대판매관리비*	매출액대판매관리비
15	부채비율	● 부채비율*	부채비율
16	유동비율	● 유동비율*	● 유동비율*
17	자기자본비율	● 자기자본비율*	● 자기자본비율*
18	● 당좌비율*	● 당좌비율*	당좌비율
19	● 고정비율*	● 고정비율*	● 고정비율*
20	순운전자본비율	순운전자본비율	● 순운전자본비율*
21	● 차입금의존도*	차입금의존도	● 차입금의존도*
22	● 현금비율*	현금비율	현금비율

연구 방법 및 검증 - 머신러닝

Random Forest

n_estimator : 50

max_depth : 5

min_sample_split : 10

SVM 앙상블

n_estimator : 50

Kernel : Linear Kernel

벌칙 모수 C : 2

KNN 앙상블

n_estimator : 50

거리 추정 방식 : 유클리디언 거리

k : 30

연구 방법 및 검증 - 딥러닝

RNN-LSTM

은닉층 : 4개, 128개 셀

epoch : 200

Activation Function: Relu

EarlyStopping

RNN-GRU

은닉층 : 4개, 128개 셀

epoch : 200

Activation Function: Relu

EarlyStopping

CNN

은닉층 : 4개, Filter 크기 : 128

MaxPooling

Dropout : 0.25

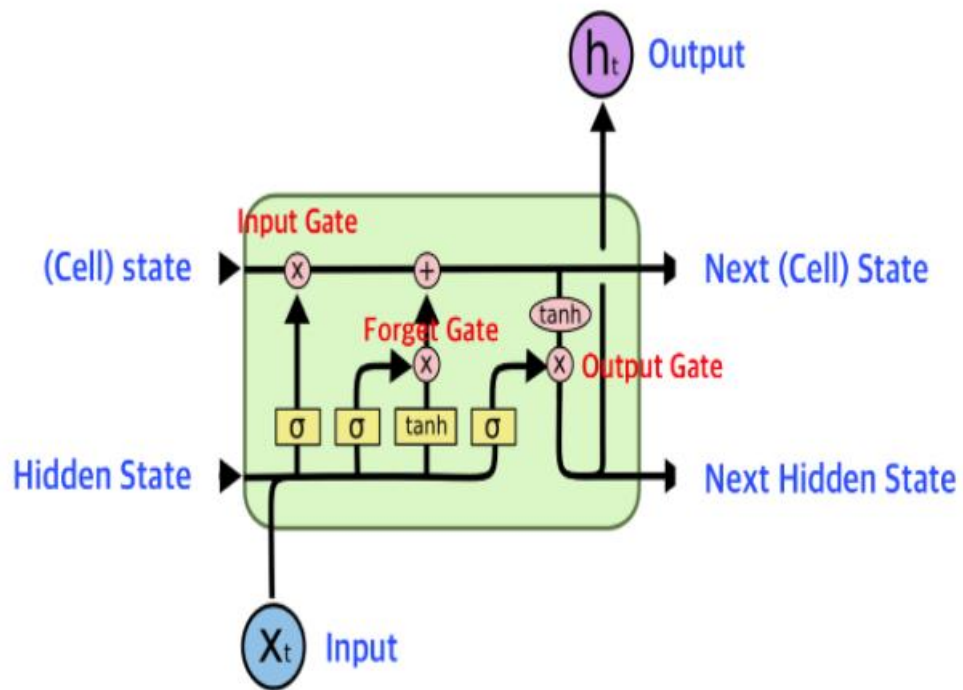
Loss Function: Binary

Crossentropy

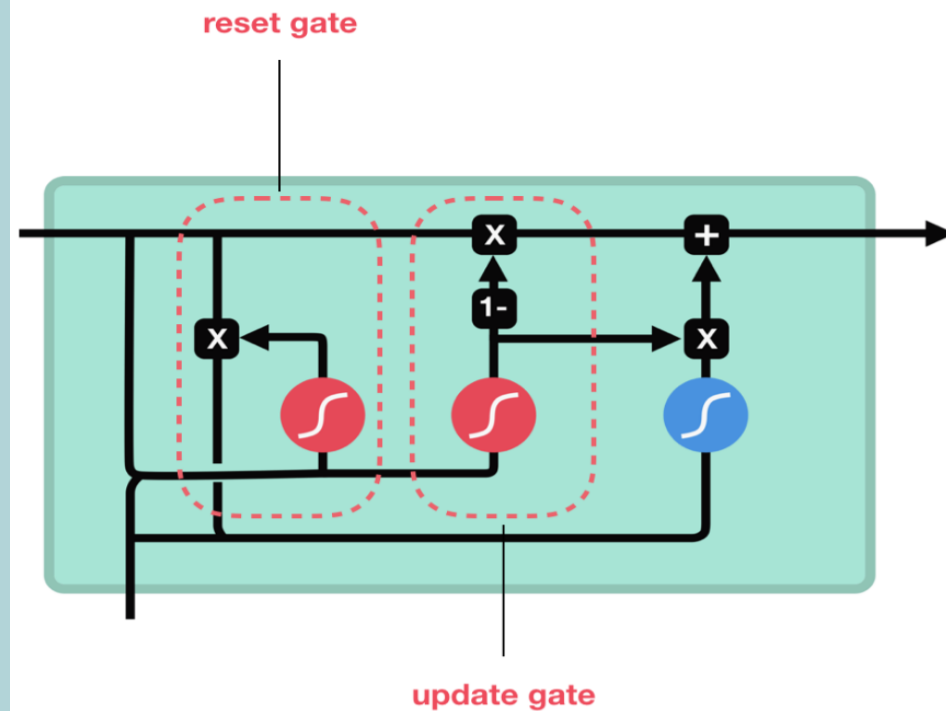
Optimizer : Adam

연구 방법 및 검증 - 딥러닝

RNN-LSTM



RNN-GRU

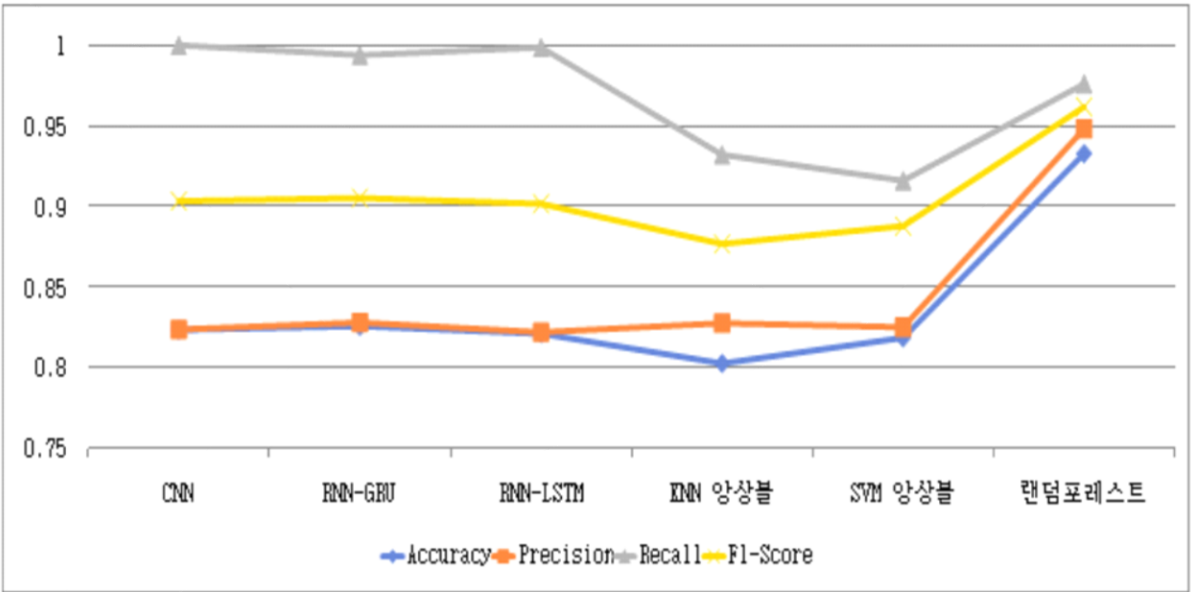


연구 방법 및 검증 - 결과

- ☑ 모델 평가 지표 :
정확도, 정밀도, 재현율, f1점수
- ☑ 랜덤포레스트 : 정확도, 정밀도, f1점수에서
가장 우수한 성적
- ☑ 재현율 : 딥러닝 모델이 머신러닝 앙상블 모델에
비교해 우수한 결과
- ☑ F1 score : 랜덤포레스트를 제외하고 딥러닝
모델이 머신러닝 앙상블 모델에 비교해 우수한
결과

〈표 6〉 모형별 정확도, 정밀도, 재현율, F1-점수

모형 \ 평가척도	정확도 (Accuracy)	정밀도 (Precision)	재현율 (Recall)	F1-점수 (F1-Score)
Random Forest	0.9328	0.9481	0.9760	0.9618
SVM Ensemble	0.8186	0.8252	0.9610	0.8879
KNIN Ensemble	0.8027	0.8277	0.9321	0.8768
RNN-LSTM	0.8212	0.8221	0.9986	0.9017
RNN-GRU	0.8259	0.8281	0.9938	0.9053
CNN	0.8236	0.8238	1.0000	0.9034



〈그림 2〉 모형별 정확도, 정밀도, 재현율, F1-점수

연구 방법 및 검증 - 재현율

☑ 한계기업 예측에서는 재현율이 중요한 지표로 다뤄짐.

- Lee et al, 2002, 251

☑ 제2종 오류는 부실기업 예측의 관점에서 '실제 부실기업임에도 불구하고 정상기업으로 예측하는 오류'로 해석.

- Reibiro, Lopes, & Silva, 2010, 5

-> 한계 기업 예측에서 2종 오류 최소화가 중요

☑ 재현율은 실제 분류대상을 분류기가 얼마나 성공적으로 분류하는지를 나타내므로 분류기의 완결성(Classifier's Completeness)으로 정의되는 중요한 평가척도라고 할 수 있음.

- Sun & Vasarhelyi, 2018, 182

논문 결론

연구 요약

딥러닝 모델을 한계기업 예측에
적용하여 그 성능을 측정,
머신러닝 앙상블 모형과 비교
-> 이를 통해 딥러닝 모형들이
한계기업을 예측하는데
그 유용성이 있음을 제시

시사점

- ☑ **학문적 관점** : 부실기업 예측에 딥러닝 모델을 적용하여
그 유용성을 검증
- ☑ **실무적 관점** : 국내 상황에 적합한 부실기업 범주인 한계기업
을 그 예측대상으로 함으로써 우리나라의 경제 실정에 맞는
기업 부실화 연구를 수행
 > 계속기업들 사이에서 어떠한 특성을 가진 기업이 관리대상
 으로 분류되어야 하는지를 보였다는 점에서 기여점이 있다.

논문 결론 - 한계점

- ☑ 한계 기업 예측에 있어서 재무비율만을 이용하여 대외적 변수를 고려하지 않았다.
- ☑ 기업 회계정보만을 활용하여 한계기업을 예측했다는 점.
 - > 역사적 데이터가 갖는 적시성, 미래 지향성의 부족
- ☑ 일정한 규모 이상의 외감기업 및 상장기업만 고려한 점.
 - > 비외감대상 중소기업을 분석대상에 포함하지 않아 다소 편향된 연구결과를 제시할 가능성을 내포하고 있다.

논문의 한계 - 이론적 배경

☑ Le&Viviani의 논문 인용 : '부실기업 예측에 ANN과 SVM이 우수하다.'

-> 실제 해당 논문 : SVM 성능이 좋지 않았음

☑ 차성재, 강정석의 논문을 인용 :

'LSTM이 전통적인 통계기법과 머신러닝 기법을 포함하여 가장 우수하다.'

-> 실제 해당 논문 : 일부 사례에서만 LSTM이 우수한 것으로 확인

논문의 한계 - 데이터

- ☑ 기초 통계량이 주어지지 않았고, 회귀분석의 기본 가정을 만족했는지에 대한 언급이 없다.
- ☑ 기업 규모에 따른 데이터 세트의 분리 없이 하나의 데이터 세트로 분석.
- ☑ 결측치 처리 방식에서 독립변수의 중앙값 대체와 종속변수 제거 방식에 대한 타당성 부족.
- ☑ 변수별 데이터 분포 현황과 t검정에 대한 결과를 제시하지 않았다.
- ☑ 사용되는 재무 비율에 대한 회계 기준을 밝히지 않았다.

논문의 한계 - 분석 결과

- ☑ 모델 별 혼동행렬을 제시하지 않았다.
- ☑ 연도별로 선택된 변수가 다른 상황에서 LSTM과 GRU 모델을 사용.
- ☑ 딥러닝 모델 중 출력층에 사용된 분류 함수에 대해 명확히 밝히지 않았다.
- ☑ CNN 모델의 구성에 대한 설명이 부족하다.

Q & A

-Thank you-