안녕하십니까? 이번 딥러닝 기반 부실기업 예측모형에 관한연구 논문 발표를 맡게된 오니기리와 여규동의 박준배입니다.

팀소개: 오니기리와 여규동

저희 팀은 규동을 맡고 있는 리더 여규동님을 필두로 오니기리 1,2,3에 김민지, 김윤지, 박준배로 구성되어 있습니다.

팀원 및 역할 소개

저희 팀 역할은 다음과 같이 구성되어 있습니다.

논문 선정 이유

저희가 해당 논문을 선정한 이유는 코로나 기간 한계기업으로 분류된 기업들이 많아 졌을 것이라생각했습니다. 그래서 판데믹이 끝나감에 따라 이런 한계기업 중 정상기업으로 회복할 수 있는 기업들이 많이 있을 것으로 생각했습니다. 저희는 이런 기업들을 예측함에 있어 딥러닝을 활용해 예측을 해보자고 생각했습니다. 그래서 해당 논문을 선정하게 되었습니다.

최초 팀 회의에서 딥러닝을 활용해 한계기업의 회생 가능성을 탐구하자는 이야기가 나왔고 해당 논문을 적용할 수 있을 것으로 생각했기에 이번 발표 논문으로 선정하게 되었습니다..

논문 서론

* 최근 경제 위기로 부실 예측에 대한 시도가 1960년대 이후로 많았다.

대표적인 이론은 다음과 같다.

Altman의 5가지 재무 변수를 활용한 다중 판별 분석 기반 부도 예측

Ohlson의 로지스틱 회귀 모형 기반 부도 예측

Zmijewski의 프로빗 모형 기반 부도 예측

* 기존 연구의 한계점

1. 실무적 한계점: 이들 논문의 한계점 은 <각각의 논문에서 말하는 부도의 기준을 이야기 하면서> 이것들은 계속 기업 중 관리 대상을 고르기에 예측의 판단 기준이 되기 어렵다.
2. 방법론적 한계점: <분석에 대한 어떤 가정을 만족해야하는지>를 이야기하면서 이것을 만족해야만 하는 한계점이 있어

* 해당 한계점을 보완하기 위해서 계속 기업 중 부실기업으로 볼 수 있는 새로운 정의를 내리고 통계에 대한 엄격한 가정을 필요로 하지 않는 데이터를 활용해 종속변수를 설명하고자 한다.
* 기존 연구에서 머신러닝을 사용한 예시는 많지만 딥러닝 기법을 적용하고 예측대상을 한계기업으로 한정한 연구는 전무한 상황이기에 이번에 해당 방식을 진행하고자 한다.

관련 선행 연구

1. 한계기업에 대한 정의:

해당 논문에서 한계기업에 대한 정의를 다음과 같이 다양한 논문에서 인용했습니다. 이들 중에서

본 연구는 한국은행의 한계기업에 대한 정의인 ‘3년 연속 이자보상배율이 1 미만인 기업’’을 사용했습니다.

* 해당 논문에서는 선행연구에 대해서 머신러닝, 앙상블, 딥러닝으로 구분하여 관련 내용을 인용하고 있습니다.

1. 머신러닝을 이용한 부실기업 예측

윤호중 – 머신러닝은 전통적 기법의 성능을 앞선다

Odom – 인공신경망 모형은 다중판별분석에 비해 우수하다

이건창 – 인공신경망이 다중판별분석, 귀납추론, 인공신경망 모형에 비해 예측력이 좋다

Shin – 부도기업 예측에 svm이 인공신경망에 비해 우수하다

Kim,sohn - svm을 이용한 중소기업의 부도 가능성 예측이 인공신경망과 로지스틱 회귀모형보다 우수하다.

Le&viviani – (한계점) 전통적 통계, 인공신경망, svm을 활용해 부도 예측을 진행한 결과 인공신경망과 svm의 성능이 우수함을 확인했다.

오우석,김진화 – 전통적 통계, 인공신경망, Decision tree 모형을 이용한 부도 예측에서 Decision tree가 가장 우수한 성능을 보임을 확인했다.

이찬호 – 부실 징후 기업 중 한계기업을 예측하고 다양한 머신러닝(dnn, rf, knn, boosting, bagging, linear) 기법 중 knn의 성능이 가장 우수한 것을 확인했다.

-위의 논문을 보여준 이유

-> 1990년대 이후 특정 머신러닝 기법이 전통적인 통계기법이나 다른 머신러닝 기법보다 우수함을 보이고자 하는 목적

* 본 연구에서 선행연구의 기조에 따라 딥러닝 모형과 머신러닝 모형을 비교해 한계기업 예측에 딥러닝의 적용가능성을 제시하고자 한다.

1. 앙상블 모형 부실기업 예측

민성환 – 개별 분류기보다 다수 분류기의 성능이 더 향상되었다.

김승혁,김종우 – 기업 부도 예측에 인공신경망, decision tree, 훈련:테스트 데이터 비율8:2를 적용한 rf 모형, 훈련:테스트 데이터 비율 6:2를 적용한 rf 모형을 비교한 결과 배깅 앙상블 모형이 가장 성능이 우수했다.

Tsai&Wu – 인공신경망 모형을 활용한 앙상블 모형과 단일 인공신경망 모형을 비교한 겨로가 앙상블 모형을 적용한 인공신경망의 성능이 더 좋게 나왔다.

김명종 – 기업 부도 예측에 불균형 문제 해결을 위해 앙상블 모형(GM-Boosting)을 적용한 결과 불균형에 상관없이 일정한 예측 결과가 나왔다.

배재권 – 전통 통계기법, 머신러닝 기법을 통합한 5가지 모형(다변량 판별분석, 로지스틱 회귀분석, 인공 신경망, 규칙유도기법, 베이지안망)을 제안한 결과 보팅과 ANN을 적용한 모형의 예측 정확성이 가장 우수했다.

최하나,임동훈 – 단일 svm보다 앙상블(보팅 기법을 적용한) svm 기법이 더 좋았다.

민성환 - knn으로 앙상블 기법을 적용한 결과 단일 knn 기법보다 우수했다.

Barboza – 기업 부도 예측에 다중판별 분석, 로지스틱 회귀, svm, 인공신경망, rf 방식 등을 비교한 결과 의사결정나무에 배깅 방식을 적용한(rf) 결과가 가장 우수했다.

엄,김,최 – rf, 인공신경망, cnn을 결합한 스태킹 앙상블 모형을 제안하고 이것이 예측 편향을 완화할 수 있다고 말한다.

* 많은 연구에서 단일 모형보다 앙상블이 더 좋다는 것을 보여주는 목적
* 선행연구에서 머신러닝 앙상블 모형과 딥러닝 기법의 성능 차이를 비교하는 연구가 업었기에 이를 비교해 딥러닝 한계기업 예측 적용의 필요성을 제시하고자 한다.

1. 딥러닝 기반 부실기업 예측

Bengio – 딥러닝 정의

오,최,장 – 딥러닝은 인공신경망 이론에 기반을 둔 것이다.

Mikolov, leCun& bengio - CNN 정의

* 다양한 딥러닝 기법 정의

Yeh – 심층신뢰망을 이용한 기업 도산 예측이 머신러닝보다 좋다

권혁권 – 순환신경망 모형이 타 모형보다 좋았다.

차성재,강정석 – LSTM이 적용된 순환신경망 기법이 전통기법과 머신러닝 기법보다 우수했다.

Mai – 기업 재무 데이터 기반 부도 예측에서 딥러닝 모형이 머신러닝 모형보다 우수했다.

Alexandropoulos – 다층 퍼셉트론 로지스틱 회귀모형 등 모형과 딥러닝 모형을 예측한 결과 딥러닝 모형이 가장 우수했다.

Arratia&Sepulveda – 기업 재무 데이터를 이미지 처리를 하여 CNN을 사용하면 예측 성능이 좋아 질 수 있음을 보여주었다.

Hosaka – 기업 재무 데이터를 이미지 처리한 이후 분석한 결과가 전통 기법 보다 CNN이 더 성능이 좋았다.

Vochozka – LSTM을 적용한 기업 부도 예측을 시도했다.

* 국내 상황에 맞춰 딥러닝 기반 부실기업 예측이 없어 해당 방식을 적용해 앙상블 모형과 비교해 그 우수성을 검증하고자 한다.

\*논문 서론 부분에 대한 한계점

-

데이터 관련

1. 분석 대상

* 해당 논문에서는 2017~2019년의 코스피, 코스닥, 코넥스, 외감기업 33916개 중 비금융업이고 결산월이 12월인 기업을 대상으로 했다.
  + 이유
    - 금융업은 영업활동 및 재무제표 계정 특성이 타 업종과 다르다.
    - 결산월 차이에 따른 데이터 왜곡을 방지하고자 한다.
    - 데이터 객관성 확보를 위해 2016년 이후 설립 기업이면서 분석기간 내에서 ‘적정’ 이외의 감사의견(적정, 한정, 부적정, 의견거절)을 받은 기업은 분석 대상에서 제외하였다.
    - 분기, 반기 재무제표는 회계감사를 거치지 않아 연차 재무제표 정보만 활용해 16813개의 기업만이 분석 대상으로 선정되었다.

1. 데이터 전처리

* 결측치: 데이터 편향 문제 발생 우려로 중앙값 대체를 활용

(단, 종속변수에 활용되는 영업이익이자보상비율에서 결측치가 존재할 경우 해당 기업 정보를 제거했다.)

* 이상치 처리: 윈저라이징 적용
* 분류 기준

‘2017~2019년 3개년 영업이익이자보상배율 1미만인 기업’ = 한계기업

그 외 = 정상기업

* 최종적으로 정상기업 수 13818개, 한계기업 2995개 집계

1. 변수 선정

* 종속변수로 선정된 한계기업을 예측하기 위해 독립변수로 각 기업의 재무비율을 활용했으며 그 분류는 다음과 같이 구분했다.

\*데이터 관련 한계점

상관분석, 정규성 검정 부분은 없었다.