1. **과제(꼭 하세요!!!!!)**
2. **팀명: 오니기리와 여규동**
3. **프로젝트 목표**

* 참고 논문 재확인
* 주제 구체화

1. **진행 사항**
2. 참고 논문 재확인
   1. 논문의 절차적 측면에서 활용 가능성이 클 것으로 생각된다.
   2. 더미 변수를 활용해 데이터의 특성을 구분하는 것은 우리 프로젝트에 활용 가능 할 것으로 생각된다.
   3. 전반적인 내용의 전개가 우리가 하고자 하는 주제와 약간 떨어진 느낌이 든다.
   4. 금융위기, 닷컴 버블, 코로나 등에 대해 더미 변수를 활용해 적용하는 것도 활용 가능성이 있을 것으로 생각된다.
3. 주제 구체화
   1. 한계기업의 정상화 예측
   2. 한계기업 정의 관련
      1. 굳이 한국은행 정의 보다는 기타 정의에 레퍼런스를 해서 자체 한계기업을 정의할 수 있을 것 같다.
      2. 이상호 자본시장 연구원 자료: 우리나라 실정에 맞는 새로운 한계기업 정의가 필요할 것으로 생각된다.
      3. 만성적인 한계기업에 대한 정의도 하면 좋을 것 같다.

BOK 연구 보고서에서 만성적 한계기업의 정의를 관측 기간 내에서 1번이라도 한계기업으로 분류된 기업으로 정의한다.

* + 1. 한계기업을 선정한 이후 여기서 별도로 만성적 한계기업을 따로 뽑는다.(전체 데이터 세트에서든 기존에 정한 한계기업 데이터 세트에서든.)
  1. 기업 분류에 관한 정의
     1. 중소기업의 경우 데이터의 신뢰성 문제가 발생할 것으로 생각된다.
     2. 상장 기업을 활용할 경우 기존 연구와의 차별성이 힘들 것 같다.
     3. 스타트업의 경우 관련 재무 정보를 벤처협회에서 구할 수는 있을 것 같다.
  2. 한계기업의 정의 후 회생을 예측하고 중요 피처값을 기반으로 향후 해당 기업이 정상화 이후

Ex: 이자보상배율 2년 1 이하 -> 40개 선정, 15개 회생 가능성 있는 것으로 판단 -> 앞으로 좋을지 예측하는 데에 있어 다른 지표를 선정하여 향후 1~2년 혹은 그 외 기간동안 정상상태 유지를 할 경우 해당 기업을 정상상태로 정의, 그렇지 않는다면 회복 후 다시 한계기업으로 돌아갔다고 판단. -> 한국 은행의 취약성 지수 활용 가능성 제시

1. 한계 기업 정의
   1. 한국은행:
      1. ‘3년 연속 이자보상배율 1 미만’
   2. 블룸버그:
      1. 레버리지 비율(부채/자본) 100% 이상 및 ROE 5% 이하인 기업
   3. 채권은행
      1. 최근 3년간 연속 이자보상배율 1미만
      2. 최근 자본총계(-)인 기업
   4. 중진공
      1. 2년 연속 적자기업 중 자기자본 전액 잠식 기업(= 완전잠식)
      2. 2년 연속 매출액이 50% 이상 감소한 기업
   5. 기술보증기금:
      1. 당기적자 시현기업 중 자기자본 전액 잠식
      2. 최근 2개년 연속 당기순손실 실현
   6. 신용보증기금:
      1. 최근 2개년 연속 총차입금이 매출액 초과 당기적자 시현기업 중 자기자본 전액 잠식
      2. 최근 2개년 연속 당기순손실 실현
      3. 최근 2개년 연속 당기순손실 실현
   7. 기업은행
      1. 매출액 30% 이상 감소
      2. 최근 1년 당행 당좌교환결제금액이 최근 1년 매출액의 70% 초과
      3. 당기 차입금의존도 70% 이상

의견 정리

중진공 정의를 써 보자

2년 연속 적자기업 중 자기자본 전액 잠식 기업(= 완전잠식) or

2년 연속 매출액이 50% 이상 감소한 기업 or

최근 2개년 연속 총차입금이 매출액 초과 당기적자 시현기업 중 자기자본 전액 잠식 or

2년 연속 이자보상배율 1 미만 or

등등으로 정의

1. 조사 기간 설정
   1. 코로나 시기를 생각하면 2020~2021 데이터 포함 필수적
   2. 타겟: 코로나 시기에 한계기업으로 분류된 애들을 볼 것인지, 17~19년도를 기준으로 해도 되고…
   3. 만약 코로나 시기를 고려한다면 과거 위기 시기의 데이터를 기반으로 이번 코로나 시기를 예측하면 과거 데이터도 필요하고
   4. 코로나 이전의 데이터를 기반으로 한다면 2017~2018년 이전 자료를 사용할 것으로 생각됨.
   5. 한계기업으로 분류되는 재무 데이터의 영향이 당해가 아니라 시차가 존재할 수 있기 때문에 정의한 한계기업 기간에 추가적으로 기간을 덧붙여서 조사 가능.

기간: 금융위기 2008~2009 2개년, 닷컴버블 기간: 2001~2002, imf: 1997~1998

1. 대상 기업
   1. 상장 및 외감기업
   2. 규모에 따른 데이터 세트 분리
   3. 상장, 비상장 외감 기업 분리
   4. 업종 분리

외감기업이 6만개 가량 나오는데 코스피, 코스닥, 코넥스 포함하면 더 많을 것으로 생각된다.

1. 목표
   1. 1차: 기업 데이터를 기반으로 한계기업 상태를 벗어나는지 확인한다.
   2. 2차: 한계기업 상태를 벗어난 기업들을 대상으로 얘네들이 정상상태를 유지할 수 있는지 확인한다.
2. 분석 방법
   1. 피처 선정: 각종 재무 데이터
   2. 타겟: 각 데이터별 3년차 이자보상배율(‘가정’) 기반 한계기업 여부
   3. 데이터 전처리
      1. 결측치

제거

* + 1. 이상치

윈저라이징, quantiledistributon, 양쪽 0.5% 삭제

* + 1. 스케일링

Quantiledistributon

Minmax

Log

Standardscale

* + 1. 데이터 검정

정규성 검정

다중 공선성

T 검정

변수 선택

* 1. 모델
     1. 로지스틱 회귀
     2. 회귀 모형
     3. 분류 모형
        1. SVM
        2. KNN
        3. Random forest
        4. ANN, DNN
        5. RNN: LSTM, GRU
        6. CNN
        7. XGBoostClassifier, catboost, lightgbm
  2. 모델 평가
     1. 혼동행렬
     2. 재현율, 정밀도, 정확도, f1-score
     3. 모델별 선정 변수와 재무비율을 기반으로 코로나 기간에 가져와서 예측을 진행한다. -> 회생 가능성이 있는 애들을 분리한다.

예측 결과가 나와요.

1차: 과거 데이터를 기반으로 한계기업 중 회생 가능성이 있는 기업들을 분리하는 예측 모형을 만들어 내고 이것을 현재 시점에 적용해 결과물을 보여준다.

이때, 회생 가능성이 있는 기업들이어떤 재무비율의 특징을 갖는지 설명하면 좋을 것 같다.