발표 대본

1.

안녕하십니까? 죽은 기업의 회생, 기사회생 팀의 발표를 하게 된 박준배라고 합니다.

지금부터 경제충격을 고려한 머신러닝 기반 외감 기업 회생 예측 및 정책 제언에 대하여 발표를 시작하도록 하겠습니다.

2.

발표 순서는 다음과 같습니다.

3.

먼저 본 연구의 개요입니다.

4.

해당 페이지는 본 연구에 대한 전반적인 내용에 대한 요약입니다. 이후 자세하게 설명될 내용이기에 해당 페이지는 빠르게 넘기도록 하겠습니다.

5.

본격적인 발표에 앞서 저희 팀 구성과 역할에 대해 말씀 드리겠습니다.

6.

저희 기사회생팀은 여규동 리더를 필두로 김윤지 팀원, 김민지 팀원, 그리고 저 박준배 팀원으로 구성되어 있고 개별 인원들이 다음 역할을 주도적으로 맡아주었고 모든 팀원들이 모든 작업에 참여를 하면서 본 연구에 기여를 하였습니다.

7.

다음은 저희 프로젝트의 수행 절차 및 방법입니다.

8.

저희 프로젝트는 6월 7일 논문 리뷰를 시작하였고 해당 기간 동안 선행 연구를 찾아보며 프로젝트의 밑그림을 그리는 작업을 시작했습니다. 이후 14일부터 본격적인 기획안을 작성하기 시작하고 20일부터 데이터 수집 및 전처리를 시작했고 중간 평가 이후 모델링 및 성능평가를 시작했습니다. 사실 달력에는 구체적인 날짜가 적혀있지만 20일 이후 최종 발표까지 시행착오를 거치면서 전처리, 모델링,성능평가 등을 끝없이 반복하면서 한 달이 넘게 열심히 달려왔습니다.

9.

이번 프로젝트의 진행 순서에 대해 플로우 차트 형식으로 말씀드리겠습니다.

선행 논문 연구 및 기획 단계에서 관련 논문 탐독 및 선행 연구의 한계점을 분석하고 프로젝트의 밑그림을 그렸습니다.

데이터 수집 및 처리 과정에서 저희는 TS2000을 활용하였고 이 과정에서 EDA를 통해 데이터 분석 및 결측치, 이상치 등을 처리하였습니다.

변수 선정 및 모델링 과정에서 저희는 기간별 유의한 변수 선정에 로짓 분석과 t검정을 활용했고 머신러닝 모델링을 활용하여 파라미터 튜닝 등으로 모델을 최적화 하는 과정을 거쳤습니다.

이후 모델 평가 및 결과 분석 단계에서 F1 score, ROC-AUC를 기간별 모델별 평가를 진행하고 선정 지표를 비교하고 결과를 비교 분석하면서 모델에 대한 강건성 검정을 진행하였습니다.

마지막으로 정책 제언 및 한계점 제시 부분에서 연구 결과의 시사점을 분석하고 이를 금융당국 및 기관에 제언하고 프로젝트의 한계점을 분석한 이후 추가 연구 필요사항을 말씀드리고자합니다.

10.

그럼 프로젝트 진행 순서에 따라 선행논문 연구에 대해 말씀드리겠습니다.

11~12

먼저 부실 기업 예측과 관련하여 가장 대표적인 비버, 알트만, 올슨의 논문을 시작으로

이 외에 다음과 같은 논문을 탐독하였습니다.

(한국의 부실기업 예측 연구 논문 파트 마지막 부분)

1. 해당 부분에서 저희는 전반적인 부실기업 예측에 대한 이해를 높일 수 있었고
2. 부도예측 모형 설계에 대한 참고 자료로 활용하였습니다.
3. 그리고 독립변수 후보군 선정에 대한 참고로 활용하였습니다.

13~14

그리고 한계기업과 관련하여 다음과 같은 논문들을 탐독했습니다.

해당 논문들에서

1. 한계 기업과 회생 기업의 정의에 대해 참고할 수 있었고
2. 본 연구 과정에서 더미변수의 활용 가능성을 확인할 수 있었으며
3. 한계기업 회생을 통한 정책 제언이 가능하다는 사실을 참고했습니다.

15.

또한 경제 충격 기간, 정상기간으로 구분한 본 연구의 목적에 참고 가능하다고 생각한 다음 논문을 읽었고 경기 순환 국면에 따른 예측 모형의 유용성에 대해 참고할 수 있었습니다.

16~17

분석 모델과 관련하여 다음 논문을 탐독했고 여기서 분석 활용 모델링과 평가 지표에 대해 참고했습니다.

18.

이상 저희가 참고한 논문에 대한 소개를 마치고 본 연구의 차별점을 말씀 드리겠습니다.

첫째, 기업의 부실화와 회생에 외부 경제 충격을 고려했다는 점 입니다.

둘째, 그 동안 선행 연구의 주요 분석 대상인 상장 기업 대신 저희는 외감 기업만 집중한 점 입니다.

셋째, 최근 금융 당국이 주목하고 있는 ‘한계기업’에 대해 새롭게 정의를 내려서 분석했다는 점 입니다.

마지막은 후보 독립변수로 기업의 특성을 구분한 더미 변수를 몇가지 추가했다는 것 입니다.

19.

본격적인 프로젝트의 기획에 대해 설명 드리겠습니다.

20.

먼저 한계기업에 대한 정의를 말씀 드리겠습니다. 기존의 많은 기관들은 다음과 같이 한계기업을 정의하고 있습니다.

21.

이에 대해 본 연구에서는 한국은행의 한계기업 정의를 활용한’ ‘2년 연속 이자 보상 배율 1 미만인 기업’, 그리고 신용보증기금과 기술보증기금의 정의를 활용해 ‘2년 연속 영업손실 상태인 기업’을 한계기업으로 정의하였습니다.

22.

다음은 저희의 주요 연구 대상인 회생 기업에 대해 말씀 드리겠습니다. 저희는 한계기업으로 분류된 기업들이 이듬해 해당 기준을 벗어난 경우 회생기업으로 정의하였습니다. 가령 2년 연속 이자보상배율이 1미만으로 한계기업으로 분류된 기업이 이듬해 이자보상배율이 1 이상이 된 경우 회생기업으로 정의하였고 마찬가지로 2년 연속 영업손실 상태로 한계기업으로 분류된 기업이 이듬해 영업 이익을 실현했다면 회생기업으로 정의하였습니다.

23.

다음으로 본 연구에서 외감기업을 연구 대상으로 선정한 이유에 대해 말씀드리겠습니다.

그 이유는 첫째, 취약한 재무 구조입니다.

외감기업들은 상장기업 대비 규모와 특성상 재무적으로 취약하기 때문에 부실화 가능성이 큽니다.

두번째는 주요 정책 제언 대상이 될 수 있기 때문입니다.

위의 이유로 금융기관 및 당국의 주요 관심 대상이 되기 때문입니다.

마지막 세번째는 상대적으로 분석에 용이할 정도로 많은 데이터가 존재했기 때문입니다.

24.

이제 저희가 설정한 기간에 대해 말씀 드리겠습니다.

저희는 충격기간은 통계청이 발표하는 경기 순환 국면의 외부 충격이 발생한 시기의 최저점을 기준으로 -2년의 데이터로 선정하여 닷컴버블, 글로벌 금융위기, 유럽 재정위기의 3개 시점으로 구분하였습니다.

25.

한편, 정상기간은 경기 순환 국면상 외부 충격이 없고 변동이 크지 않은 기간으로 설정하였습니다. 그래서 경기 순환 국면의 11순환의 다음 3개 시점으로 정상기간을 구분하였습니다.

26.

본 연구의 최종적으로 선정된 후보 독립변수는 3개 구분으로 나눠 총 28개를 선정했으며 해당 내용에 대해 자세히 설명 드리겠습니다.

27.

후보 변수군 1은 기업 재무상태를 나타내는 일반적인 재무비율변수로 세부적으로 성장성 지표 3개, 수익성 지표 5개, 활동성 지표 4개, 안정성 지표 8개를 선정하여 해당 부문에서 20개의 후보 변수가 선정되었습니다.

28.

후보 변수군 2,3은 추가비율 변수군 및 더미변수군으로 나뉘는데

먼저 추가비율 변수군은 한국은행의 기업 취약성 지수 선정에 활용되는 변수로 레버리지 비율 3개, 유동성 비율 2개를 선정하여 해당 부문에서 5개의 변수가 선정되었습니다.

더미변수의 경우 저희 팀에서 만든 변수로 좀 더 설명드릴 부분이 있습니다. 업종에 대한 부분은 제조업을 0, 비제조업을 1로 구분하였습니다.

매출과 자산은 명확히 말하면 더미변수가 아니라 범주화 변수로 전체 분포와 이상치를 고려하여 1분위 미만은 0, 1~3분위 사이는 1, 3분위 이상은 2를 부여하였습니다.

29.

다음은 종속변수에 대한 부분입니다.

본 연구의 관심은 회생기업이기 때문에 이자보상배율 기준에서는 2년 연숙 이자보상배율 1미만인 한계기업 중 이듬해 이자보상배율 1 이상인 기업을 회생기업으로 1을 부여하고 그 외에 0을 부여했고

영업손실 기준에서도 마찬가지로 1과 0을 부여하였습니다.

30.

다음은 데이터 수집입니다.

31.

최초에 저희 팀은 TS2000에서 재무비율을 수집하였습니다. 그러나 해당 자료에서 큰 문제점이 있었습니다.

먼저 보시다시피 전혀 데이터가 없는 부분과 실제 재무데이터를 기반으로 계산 했을 때와 다른 값이 나오는 것을 확인할 수 있었습니다.

또한 마이너스 비율을 확인할 수 있는데 이는 데이터 교란 요인이 되었습니다.

32.

이러한 이유로 저희는 재무 데이터를 직접 수집하여 계산을 진행했습니다.

33.

그 결과 데이터 전처리 진행 이전 최초 데이터의 개수인 row의 개수가 기간별로 다음과 같이 구할 수 있었습니다.

34.

이제 데이터 전처리에 대해 말씀드리겠습니다.

35.

먼저 결측치와 inf 표기 데이터에 대한 처리입니다.

해당 데이터에 대해서 해당 값이 포함된 row 자체를 제거하는 방식으로 전처리를 진행했고 그 결과 기간별로 다음과 같이 데이터가 남았습니다.

36.

이 과정에서 최초 데이터를 가져올 때, 제조업 분류로 한정하여 기업 데이터를 가져왔음에도 특수목적 회사와 금융사 데이터가 다음과 같이 일부 포함되어 있다는 것을 알게 되었습니다. 특수목적 회사의 경우 일반적으로 투자 목적으로 설립되어 이익의 90% 이상을 배당하여 재무적 특성이 일반적인 기업과 다르고 금융사의 경우 회계 기준 자체가 다르기 때문에 해당 데이터를 삭제하였습니다.

37.

이상치에 대한 처리입니다.

저희는 최대한 강건한 모형을 만들기 위하여 이상치에 대해서는 제거하는 방식보다는 윈저라이징을 하는 방식으로 데이터를 유지하고자 했습니다. 또한 최대한 원 데이터를 유지하기 위해서 양측 0.5%, 총 1%에 대해서 윈저라이징을 진행하였습니다.

38.

최종적으로 본 연구에서 정상기간 3개 기간, 충격기간 3개 기간 총 6개 기간에 대해서 이자보상배율, 영업손실 기준 2가지 한계기업 기준으로 구분한 12개의 데이터 세트를 구성했고 해당 데이터를 기반으로 연구를 진행하였습니다.

해당 표가 눈에 잘 들어오지 않을 것으로 생각되어 시각화 자료로 자세히 설명드리겠습니다.

39.

먼저 이자보상배율 기준 한계기업 데이터 세트입니다.

Ppt에 나와있는 그래프는 한계기업 중 회생 된 기업과 회생하지 못한 기업의 숫자를 분류한 것입니다. 00~01 데이터의 경우 2000~2001년 데이터를 기준으로 한계기업을 분류하고 2002년에 회생 여부를 표현한 것입니다.

40.

해당 데이터에 대해서 비율로 표현하면 다음과 같습니다. 충격기간에는 한계기업 중 회생 기업의 비율이 약 30~36% 사이에 있고 정상기간의 경우 약 25~39% 사이에 있습니다.

41.

마찬가지로 영업손실을 기준의 회생기업과 미회생기업의 숫자는 다음과 같습니다.

42.

해당 데이터에 대해서 비율로 표현하면 충격기간에 한계기업 중 회생기업의 비율이 약 35~44%에 있고 정상기간의 경우 약 30~42% 사이에 있습니다.

43)

이제 저희 연구의 수행 결과에 대해 말씀드리겠습니다.

44)

먼저 최종 독립변수의 선정입니다.

45)

최종 독립변수 선정은 이전에 설명드린 내용처럼 로짓 모형을 사용하고 t검정을 통해 유의한 변수를 추출했습니다. 이때, 기본적으로 후진 제거법을 활용하고 p-value 기준으로 1개씩 제거하는 방식을 활용했습니다.

이때, 상관계수가 0.8 이상인 독립변수에 대해서는 p값을 기준으로 더욱 유의한 변수를 선택하고 그 이하인 경우 유의한 변수는 제거하지 않았습니다.

그래서 95% 신뢰수준에서 유의한 변수를 최종적으로 선정하였습니다.

46)

먼저 충격기간 이자보상 배율 기준으로 한계기업을 분류한 데이터에서 기간별로 다음과 같이 변수가 선정되었고

47)

정상기간 이자보상 배율 기준 데이터에서는 기간별로 다음과 같이,

48)

이어서 충격기간 영업손실 기준 데이터

49)

정상기간 영업손실 기준 데이터에서 다음과 같이 변수가 선정되었고

50)

이것을 종합하여 정상기간, 충격기간 분류에 따라 표로 표현하면 다음과 같습니다.

해당 표를 충격기간, 이자보상배율 행을 예시로 들면서 설명하겠습니다.

충격기간 이자보상배율 행에 대해서 적혀져 있는 변수는 충격기간으로 정의된 3개의 기간에 대해 1번이라도 유의 하다고 분류된 변수들 입니다. 여기에서 파란색 변수의 경우 3개의 기간 중 2개의 기간에 유의 하다고 분류된 변수이고 빨간색 변수는 3개의 기간 모두 유의 하다고 분류된 변수들입니다.

전반적으로 수익성 지표가 정상, 충격 기간에 관계 없이 유의한 것을 확인 할 수 있고 그 외에 자산, 매출 더미 변수, 부채/매출액 변수와 차입금 평균 금리가 눈에 띄는 것을 알 수 잇습니다.

51.

이것을 한계기업에 대한 분류 기준에 따라 모아서 확인하면 분류 기준에 따른 유의한 변수가 차이가 잘 보이는 것을 알 수 있습니다.

52.

이제 머신러닝 모델링 결과에 대해 말씀 드리겠습니다.

53.

본 연구에서 train, test set의 분류는 8:2의 비율로 분리했으며

사용 모델은 금융권에서 가장 일반적으로 사용하는 전통적인 로지스틱 모형, 머신러닝 기법으로 의사결정나무, knn, 인공신경망 모형, 앙상블 모형으로 랜덤포레스트와 xgboost, 딥러닝 모형을 활용하였습니다.

54.

그리고 데이터셋 별로 최상의 모델을 구현하기 위해 다음 하이퍼 파라미터에 대해 gridsearch cv를 통해 최적 파라미터를 찾고자 했으며 해당 파라미터 튜닝 값의 세부 설정은 다음과 같습니다.

55~56

57~58.

그래서 한계기업 기준에 따른 데이터 세트에 대한 개별 모델의 성능을 기간별로 평균한 값은 다음과 같습니다. 여기서 빨간색은 제일 좋은 성능을 주황색은 2번째, 노란색은 세번째로 좋은 성능을 보여주는 값을 표기했습니다.

59.

해당 표들의 내용을 요약하여 최고의 성능을 보여주는 모델을 기간별, 한계기업 분류 기준별로 3위까지 모아서 보여드리면 다음과 같습니다.

전반적으로 로지스틱, xgb, randomforest 모형이 우수한 성능을 보여주었고 일부 사례의 경우 decisiontree와 딥러닝, ann 모형이 우수한 성능을 보여주었습니다.

60~61.

해당 모델들의 성능을 차트로 보여드리면 다음과 같습니다.

이자보상배율 기준으로 상위 3개의 모델의 성능은 다음과 같고

영업손실 기준 상위 3개의 모델의 성능은 다음과 같습니다.

62.

다음은 강건성 검증입니다. 강건성 검증은 서로 다른 기간의 데이터를 교차로 넣어 강건성 검증을 진행했고 모델별 강건성 검증 결과치의 평균값과 기존 모델의 결과치와 비교한 표는 다음과 같습니다.

63.

이자보상배율 기준으로 모델의 강건성을 검증한 결과

전반적으로 딥러닝 모형이 강건성 검증에 있어 다소의 편차를 보였고 randomforest와 xgb 모형은 편차가 작거나 오히려 더 좋게 나오는 것을 확인할 수 있었습니다.

64.

영업손실을 기준으로 모델의 강건성을 검증한 결과

로지스틱 분류의 성능이 유독 낮게 나온 것을 확인할 수 있었습니다.

65.

본 연구의 시사점 및 정책적 제언입니다.

66.

시사점에 대해서 일종의 키워드로 설명 드리겠습니다.

첫번째는 매출 to the 차입금입니다.

이전에 보신 것 처럼 매출과 자산 그리고 차입금과 관련된 지표는 대부분의 데이터셋에서 유의한 변수로 선정 된 것을 확인할 수 있었습니다.

두번째는 수익성, 수익성, 수익성입니다.

수익성 지표는 모든 데이터셋에 대해 유의한 결과를 보여주었습니다.

세번째는 기간보다 기준입니다.

저희의 연구는 회생기업의 주요 변수가 기간에 따라 차이가 있을 것으로 생각했으나 그것 보다는 한계기업의 정의인 이자보상배율, 영업손실 기준에 따라 명확한 차이를 보여주었습니다.

마지막은 돌고 돌아 앙상블입니다.

Randomforest와 xgboost 같은 앙상블 모형은 성능적 측면과 강건성 측면에서 모든 데이터 셋에서 좋은 성능을 보여주었습니다.

67.

이제 본 연구를 기반으로 금융 당국 및 기관에 제언을 하면 다음과 같습니다.

첫째, 본 연구에서 제시한 유의한 변수와 관련된 지표에 대한 집중적인 관리가 필요합니다. 모든 데이터 세트에 대해 공통적으로 유의한 지표 외에도 기간이나 기준별 유의 지표에 대해서 세심한 관리가 필요합니다.

둘째, ‘한계기업+’ 관리를 제안합니다. 한계기업 선정 기준을 보다 다양화하여 보수적인 관리를 할 필요가 있습니다.

셋째, 앙상블+딥러닝의 활용을 제안합니다. 전반적인 좋은 성능을 보이는 앙상블 모형과 더욱 예측력이 좋아질 딥러닝의 복합 사용을 제안합니다.

68

저희 팀의 자체 평가 및 의견입니다.

69.

먼저 본 연구의 기여 효과입니다.

본 연구의 의의로

첫번째, 회생 예측에 경제 충격을 도입해 이전 선행연구에서 고려하지 않은 경제 충격 기간과 정상기간에 대한 분석을 진행했습니다.

두번째, 외감 기업만을 분석하여 기존의 주 연구 대상인 상장기업과 규모나 성격이 다른 대상을 연구했습니다.

세번째는 다양한 독립변수의 다양화입니다. 일반적으로 활용되는 재무비율 외에 레버리지나 유동성 변수 및 더미변수를 추가해 연구를 진행하였습니다.

마지막은 다양한 데이터셋을 구성해 기간별, 한계기업 기준별 분할을 진행해 많은 데이터셋에 대해 평가 및 비교를 시도했습니다.

70.

이제 본 연구의 한계점을 말씀 드리겠습니다.

먼저 기간 설정 및 충격 상황에 대해서 2년이라는 기간 설정에 대한 적절성 여부입니다. 실질적인 경제 충격 상황은 2년으로 딱 떨어지지 않고 경제 충격 상황이 모두 동일하지 않은데 이것을 분리하지 않고 일괄적으로 충격상황이라는 분류를 했기 때문에 각기 다른 충격 상황에 대한 고려가 필요할 것으로 생각됩니다.

두번째는 연간 기준 데이터 관련입니다. 저희는 본 연구에서 연말 결산 데이터를 사용했는데 이는 연중 동적 상황에 대한 미시적 분석이 불가능하다는 문제점이 있습니다. 가령 기업의 경영 상황이 연말에 급격하게 나빠질 경우 해당 기업의 현재 상황은 한계기업 상태이지만 연말 결산 데이터 상에서는 정상기업으로 분류하는 문제입니다.

세번째는 시차요소를 고려하지 않았다는 점 입니다.

실제 경제 충격이 기업에 미치는 영향은 수년에 걸쳐 진행되는데 이 부실화에 대한 시차적 요소를 고려하지 못했습니다.

71.

추후 연구과제를 끝으로 본 발표를 마무리 하겠습니다.

먼저 더욱 세밀한 충격기간 설정이 필요합니다. 즉, 연 단위가 아닌 실제 경제 충격에 영향이 있는 기간으로 설정할 필요가 있습니다.

두번째는 시차 변수 도입의 연구입니다. 경제 충격으로부터 부도, 회생까지의 장단기 시차를 고려할 수 있는 변수를 도입하는 것 입니다.

마지막은 한계기업에 대해 더욱 다양하게 정의하여 많은 데이터 셋을 구성하여 비교 분석할 필요가 있습니다.

이것으로 기사회생팀의 경제충격을 고려한 머신러닝 기반 외감 기업 회생 예측 및 정책 제언에 대한 발표를 마무리 하도록 하겠습니다. 긴 시간 발표를 들어주셔서 감사합니다.