

## 설명 가능한 인공지능 기반 기업부도 예측에 영향을 미치는 요인에 관한 연구

박 중 현<sup>1</sup> · 김 경 영<sup>2</sup> · 주 진 혁<sup>3</sup> · 이 현<sup>4</sup> · 최 희 정<sup>5\*</sup>

<sup>1</sup>선문대학교 컴퓨터융합전자공학과 박사과정 <sup>2</sup>선문대학교 컴퓨터공학부 학사과정 <sup>3</sup>선문대학교 SW융합학부 연구교수

<sup>4</sup>선문대학교 컴퓨터공학과 부교수 <sup>5\*</sup>선문대학교 경영학과 조교수

## Factors Affecting Corporate Insolvency Prediction Based on Explainable Artificial Intelligence

Joong-Hyun Park<sup>1</sup> · Gyeong-Yeong Kim<sup>2</sup> · Jin-Hyeok Ju<sup>3</sup> · Hyun Lee<sup>4</sup> · Hee-Jung Choi<sup>5\*</sup>

<sup>1</sup>Doctor's Course, Department of Computer and Electronics Convergence Engineering, Sun Moon University, Asan-si 31460, Korea

<sup>2</sup>Undergraduate Program, Division of Computer Science and Engineering, Sun Moon University, Asan-si 31460, Korea

<sup>3</sup>Research Professor, Division of Software and Convergence, Sun Moon University, Asan-si 31460, Korea

<sup>4</sup>Professor, Division of Computer Science and Engineering, Sun Moon University, Asan-si 31460, Korea

<sup>5\*</sup>Professor, Department of Business Administration, Sun Moon University, Asan-si 31460, Korea

### [요 약]

최근에는 기업 신용상태를 평가하고 부실기업을 예측하는 데에는 전통적인 방법과 더불어 인공지능 기술이 널리 활용되고 있다. 그러나 머신러닝 및 딥러닝 모델은 고차원의 복잡한 데이터와 파라미터들 간에 복잡한 상관관계를 학습하고 이를 통해 예측이나 결론을 도출하기 때문에 모델들이 학습한 패턴과 결론을 설명하기 어렵다. 또한 복잡한 형태의 블랙박스이기 때문에 그 결론을 설명하거나 이유를 명확히 파악하기가 더욱더 어렵다. 본 논문에서는 이러한 불투명성을 고려하여 설명가능한 인공지능의 적분 경사를 활용하여 기업부도예측 모델에서 어떤 특성들이 기업부도예측과 관련이 있고 이러한 특성들이 어떻게 영향을 미치는지를 시각화하여 모델의 결론에 영향을 주는 요인을 분석함으로써 모델의 예측 결과에 대한 신뢰도 및 투명성을 향상시키고자 한다.

### [Abstract]

Nowadays, artificial intelligence (AI) technology is widely being used alongside traditional methods to evaluate corporate credit status and predict potential bankruptcies. However, machine learning and deep learning models learn complex correlations among high-dimensional and intricate data and parameters to derive predictions or conclusions, making it difficult to explain the patterns and decisions learned by these models. Moreover, owing to their complex black-box nature, proving clear explanations or understanding the reasons behind their conclusions becomes challenging. Considering this opacity, this study aims to enhance the reliability and transparency of the corporate bankruptcy prediction model by utilizing interpretable AI's integrated gradients. We analyze the features that are relevant to corporate bankruptcy prediction and visualize how these influence the model's conclusions, thereby exploring the factors that affect the prediction results, ultimately improving the credibility and transparency of the predictions.

**색인어** : 설명 가능한 인공지능, 적분 경사, 기업 부도, 예측, 위험 요인

**Keyword** : Explainable Artificial Intelligence, Integrated Gradients, Corporate Insolvency, Prediction, Risk Factors

<http://dx.doi.org/10.9728/dcs.2023.24.9.2093>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Received** 29 July 2023; **Revised** 08 August 2023

**Accepted** 28 August 2023

**\*Corresponding Author; Hee-Jung Choi**

**Tel:** +82-41-530-2539

**E-mail:** choihj0105@gmail.com

## I. 서 론

최근에는 기업 신용상태를 평가하고 부실기업을 예측하는 데에는 전통적인 방법과 더불어 인공지능 기술이 널리 활용되고 있다. 전통적인 Machine Learning(머신러닝) 및 Deep Learning(딥러닝) 모델은 고차원의 복잡한 데이터와 파라미터들 간에 복잡한 상관관계를 학습하고 이를 통해 예측이나 결론을 도출한다. 그러나 이러한 모델들은 학습된 패턴과 결론을 설명하기가 어려운 경우가 있으며 특히 딥러닝의 경우에 복잡한 블랙박스(Black-Box) 형태이기에 더욱 뚜렷하게 나타난다. 따라서 모델이 어떻게 결과를 도출하는지 이해하는 것이 중요한 응용 분야에서는 Explainable AI(설명가능한 인공지능)의 필요성이 대두되었다.

Explainable AI란 복잡한 인공지능 모델이 도출한 결과를 사람이 이해 가능하게 설명하고 분석하는 방법론을 말한다. 일반적으로 Explainable AI 방법은 심층신경망(Deep Neural Network, DNN) 기반의 복잡한 구조를 가지는 인공지능 모델의 결론 도출 과정을 분석하려는 기술이다.

과거 금융위기 등의 주요 발생 원인으로 신용 위험이 지목되면서 기업부도를 선제적으로 예측하는 방법론은 그간 꾸준히 발전되어 왔으며 최근에는 부도사건 자체와 재무정보 등의 관계를 머신러닝을 통해 파악하고 이를 토대로 부도를 예측하는 연구가 활발하게 이루어지고 있지만, 부도사건이 정상 기업에 비해 희소하다는 점에서 오버샘플링이나 언더샘플링이 요구되고 있고 이는 정보 왜곡을 발생시키는 요인으로 작용할 가능성이 높은 문제점을 고려하여 머신러닝과 딥러닝 기법을 활용한 인공지능 기반 기업부도 예측 모델에 시간적 특성을 반영하고 인공지능 모델 평가지표 F1 Score를 통해 신뢰도 및 정확도가 높은 기업부도 예측 모델 연구가 진행되었다[1]. 이러한 연구 결과를 바탕으로 Explainable AI를 활용하면 기업부도 예측 모델이 어떻게 결론에 도달하는지 이해할 수 있으며 모델의 어떤 특성(Feature)이 기업부도와 관련이 있는지, 그리고 이러한 특성이 어떻게 영향을 미치는지를 시각적으로 표현할 수 있다. 이를 통해 기업의 재무 상태를 해석하고, 모델의 결론에 영향을 주는 주요 요인을 식별할 수 있으며, 모델의 결론이나 예측 결과를 설명할 수 있기 때문에 의사결정자나 신용 평가 기관은 기업부도 예측 모델을 기반으로 결정의 타당성을 높일 수 있다. 이는 금융 기관이나 투자자들이 부도 위험이 있는 기업을 식별하고 적절한 선제적 조치를 취하는 데 도움이 될 수 있게 될 것이다.

이에 본 논문에서는 Explainable AI의 Integrated Gradients(적분경사)에 기반하여 기업부도 예측 모델에서 기여도를 분석하고 어떤 특성들이 기업부도와 관련이 있는지, 그리고 이러한 특성들이 어떻게 영향을 미치는지를 시각적으로 표현하여 기업부도요인을 해석하고 모델의 결론에 영향을 주는 주요 요인을 분석하여 모델의 결론이나 예측 결과를 설명하여 기업부도예측 모델의 결과에 대한 신뢰도 및 투명성을 향상시키고자 한다.

또한 추가적으로 분석된 결과를 바탕으로 특성 기여도가 높은 상위 10개의 특성을 가지고 인공지능 모델인 DNN, Convolutional Neural Network(CNN), Long Short-Term Memory(LSTM)의 3가지 기업부도 예측 모델에 적용하고 성능 평가를 실시하여 분석된 Feature Attribution(특성기여도)를 검증하고 그 차이에 대하여 비교 분석하고자 한다.

## II. 관련 연구

### 2-1 전통적 기법을 활용한 기업부도 예측 선행연구

기업의 신용평가와 부실 정도를 예측하는 방법으로 과거부터 다양한 모형들을 활용한 연구들이 활발하게 진행되어 왔다. Beaver[2]는 기업의 재무정보를 이용하여 현금흐름 비율, 순이익비율, 총자산 대비 부채비율, 총자산 대비 유동자산비율, 부채비율 대비 유동자산비율, 매출액비율의 6개 그룹에 속하는 30개의 변수를 선정하고 단일변량분석을 이용한 기업부도 예측 모형으로 연구를 진행하였으며 연구결과로 제시한 예측의 정확도는 78%로 다소 낮은 예측 정확도를 나타내었다.

Altman[3]은 도산기업과 건전기업을 대상으로 총자산 대비 운전자본, 총자산 대비 이익잉여금, 총자산 대비 세전이익, 총부채 대비 자본의 시장가치, 총자산 대비 매출액을 변수로 선정하고 다변량 판별분석을 이용한 기업부도 예측 모형으로 연구를 진행한 결과 도산 1년 전의 재무비율을 이용한 예측 정확도가 95%로 나타났다.

Ohlson[4]의 연구에서는 정상기업과 도산기업을 표본으로 구성하여 기업부도에 영향을 미치는 요인에 대한 연구와 함께 기업부도 예측 정확도를 로짓분석 모형을 이용한 분석하였으며 연구결과로 기업규모, 재무구조, 경영성과, 활동성 등이 기업부도에 영향을 미치는 요인이며 기업부도 예측의 정확도가 96%라는 연구결과를 제시하였다.

전통적인 기업부도 예측 모형을 이용한 대표적인 선행연구에서는 모두 기업의 재무정보를 활용한 재무비율을 토대로 실증분석을 실시하였으며 분석결과에 따른 예측정확도를 연구결과로 제시하였으나 시대가 변화하면서 1990년 이후 단순 재무정보를 이용한 기업부도 예측에 대한 연구가 아닌 기업특성이나 거시경제와 같은 비재무적인 시장정보를 이용한 기업부도 예측에 관한 연구들로 발전하게 되었다.

Mcquown[5]의 연구에서는 자본시장에서 시장 가격을 이용한 옵션가격 평가모형으로 기업부도 위험 수준인 EDF(Expected default frequency)를 측정하는 KMV 모형을 제시하였다. 연구결과 채무불이행 발생 확률에 대한 추정치 도출은 단순 기대손실과는 차이가 있으며 채무불이행 예측에 있어서 효율적이라는 것을 EDF를 이용하여 검증하였다.

오세경[6]의 연구에서는 로짓모형을 이용하여 다변량 관별분석과 함께 옵션가격 평가모형을 이용하여 EDF의 시간별 변화 추이를 분석하였으며 부실 기업의 EDF가 부도가 발생하기 수개월 또는 1년 이상 전부터 급격히 올라가는 것을 확인하였으며 시장정보에 의한 기업부도 예측의 유용함을 연구 결과로 제시하였다.

기업부도에 또 다른 요인인 거시 경제 변수의 영향을 반영한 통합 모형 연구도 수행되었다. Nam, C., T. Kim, N. Park, and H. Lee[7] 시간 가변적인(Time-varying) 헤저드 모형을 사용하여 거시경제 변동이 기업의 부도(헤저드) 확률을 상승시킬 수 있음을 실증 분석하였다. Tinoco and Wilson[8]은 재무 지표, 시장 정보와 함께 거시경제 변동 수준을 설명변수로 포괄하는 Panel Logit 기반의 다중회귀분석을 활용한 부도예측 모형을 연구하였다. 연구결과 거시경제 변동은 부도에 매우 결정적인(conclusive) 영향을 미치는지는 않았지만, 시장 정보 등 타 요인의 한계적(marginal)으로 영향을 미칠 수 있음을 연구하였다.

## 2-2 인공지능을 활용한 기업부도 예측 선행연구

인공지능을 활용한 기법은 머신러닝과 딥러닝 모델이 대표적이라고 할 수 있는데 오늘날 4차산업혁명 기술의 발전과 더불어 우리의 다양한 산업에 적용되고 있다.

인공지능을 활용한 기업부도 예측에 대한 선행연구로는 기존 재무정보를 활용한 기업부도 예측의 한계를 보완하기 위해 비재무적인 정보를 활용한 인공신경망 기반의 부도예측 모형을 제시하고 재무정보가 불투명한 중소기업의 경우 예측 모형의 효과성을 연구결과로 제시하였다[9]. 또한 딥러닝 개념의 인공신경망 기법 중 하나인 Deep Belief Networks (DBN)이 기존 Support Vector Machine(SVM) 보다 기업 부도의 예측 성능이 우수하다는 연구결과를 제시하였으며[10], 인공신경망(Artificial Neural Network)과 K-군집분석 등의 방법론을 이용한 기업부도 예측 모형의 우수한 예측력을 실증분석을 통해 연구결과로 제시하였다[11].

기업신용등급의 예측에 있어서 인공지능 기법 중 랜덤 포레스트(Random Forests) 기법을 적용하여 다중관별분석과 인공신경망, 다분류 SVM 등 전통적 기법을 활용한 기업부도 예측 모델의 방법론과 비교를 통해 예측 성능이 우수하다는 것을 실증분석을 통해 연구결과로 제시하였다[12].

또한 부도위험 예측에 있어서 사례 선택을 활용한 배깅(Bagging) 모형이 기존 SVM보다 예측력이 뛰어난 것으로 나타났으며[13], 두 개의 인공신경망 모델을 이용하는 하이브리드 모형을 토대로 부도 예측을 제안한 연구[14],[15],와 딥러닝 시계열 알고리즘인 RNN과 LSTM 기반의 부도예측모형이 다른 알고리즘에 비해 성능이 우수함을 확인한 연구가 제시되었다[16].

박중현 외[1] 연구에서는 인공지능 기반 기업부도 예측 모델을 시간적 특성을 반영하여 실증분석을 실시하였다. 머신러

닝 기반 SVM, RF 모델과 딥러닝 기반 DNN, CNN, LSTM 모델을 도입하여 기업부도 예측을 진행하였으며 Cross Validation(교차 검증)을 통해 머신러닝 및 딥러닝 모델의 성능을 향상시켰다.

선행연구와 같이 전통적인 기법을 활용한 기업부도 예측뿐만 아니라 금융이나 재무 분야에서도 인공지능을 활용한 기업부도 예측이 다양하게 연구되고 있다. 그러나 전통적 기법을 활용한 기업부도 예측 모델에 대한 연구에 비해서 그 양과 질은 부족한 상황이지만 전통적인 기법을 활용한 기업부도 예측과 인공지능을 활용할 기업부도 예측과의 비교를 통해 인공지능을 활용한 기업부도 예측 모형이 우수하다는 연구결과를 제시하고 있다.

## III. 연구방법

### 3-1 분석 대상 기업

본 연구에서는 어떠한 특징들이 기업부도에 영향을 미치는지 분석하기 위해 기업 기준은 2001년부터 2021년까지 KOSPI, Kosdaq 시장에서 상장된 기업과 그에 따른 부도에 관련된 공시가 발생한 기업을 대상으로 연구를 실시하였으며, 보다 정확한 공시를 확인하기 위해 Web crawling(웹크롤링) 방법을 이용해 전자공시 시스템인 DART, KIND를 이용하여 상장폐지 관련 정확한 공시를 확인하였다. 상장폐지 사건은 부도와 반드시 연결된다고 볼 수는 없으나 거래 정지 및 추가 하락이 발생하여 투자자와 채권자가 큰 손실을 볼 수 있는 사건이므로 상장폐지를 부도로 인식하는 것은 보다 보수적인 기준에서 부도를 적절하게 평가하는 방법이라고 할 수 있으며[17]~[19], DART 및 KIND에 공시된 상장폐지 사유 중 부도와 직접적인 관련이 없다고 판단되는 “신규/이전 상장”, “특수 목적에 의한 상장폐지”, “폐흡수합병”, “완전 자회사화” 등의 사유는 분석 대상 기업에서 제외 시켰다. 또한 상장폐지 사유 중 부도와 직접적인 관련이 있는 “부도발생”, “파산선고”, “감사의견 거절”, “회사정리절차개시”, 및 “은행거래정지” 등의 사유는 부도기업으로 정의 하였으며 KOSPI 시장의 정상기업 수는 711곳, 부도기업 수는 118곳이며, Kosdaq의 경우 정상기업 수는 1,300곳, 부도기업 수는 411곳으로 분석되었다(표 1).

표 1. 분석 대상 기업 수

Table 1. Number of analyzed corporations

Corporations	KOSPI	KOSDAQ
Non-Defaulted	711	1,300
Defaulted	118	411

표 2. 분석 데이터 구성

Table 2. Composition of training data

Index	Description
TL/TA	Debt to Assets
TL/MV	Debt to Market Vale
FL/TA	Financial Liability to Assets
FL/TL	Financial Liability to Total Liability
CA/CL	Current Ratio
NCA/TA	Non-Current Assets to Assets
OI/TA	Operating Income to Assets
NI/MV	Net Income to Market value
ROE	ROE(Return on Equity)
RE/TA	Retained Earning to Assets
AGR	Assets Growth Rate
SGR	Sales Growth Rate
NIGR	Net Income Growth Rate
Cash/TA	Cash to Assets
Cash/MV	Cash to Market Value
OCF/TA	Operating Cash-Flow to Assets
Sales/TA	Assets Turnover
Sales/AR	Account Receivables, Turnover
ln(Sales)	The natural logarithm of Sales
ln(TA)	The natural logarithm of Total Assets
△ Employee	Employee growth rate
△ Salary	Salary growth rate
MajorOwn(%)	Equity ratio of majority shareholders
△ MajorOwn	MajorOwn growth rate
CD	91-day Certificate of Deposit yield
GDP	GDP in billion won
FX	Foreign exchange rate
KOSPI	KOSPI closing price
KOSDAQ	KOSDAQ closing price
△ CPI	Consumer Price Index growth rate
△ PPI	Producer Price Index growth rate
Bond rate(%)	3-year government bond yield

### 3-2 분석 데이터 구성

분석 대상이 되는 기업의 총 수는 KOSPI 및 KOSDAQ 포함 2,540개 기업이며, 각 기업의 건전성, 수익성, 성장성, 유동성, 활동성, 규모, 기업특성, 거시경제지표를 활용하여 분석 데이터를 구성하였다. 건전성에는 부채비율(TL/TA), 시장부

채비율(TL/MV), 금융부채비율(FL/TA), 금융부채비율2(FL/TL), 유동비율(CA/CL), 고정자산비율(NCA/TA), 수익성에는 총자산영업이익율(OI/TA), 시장자산영업이익율(NI/MV), 자기자본순이익율(ROE), 총자산이익잉여금비율(RE/TA), 성장성에는 총자산증가율(AGR), 매출액증가율(SGR), 당기순이익증가율(NIGR), 유동성에는 현금자산비율(Cash/TA), 시장현금자산비율(Cash/MV), 자산대비용업현금흐름(OCF/TA), 활동성에는 자산회전율(Sales/TA), 매출채권회전율(Sales/AR), 규모에는 총매출액규모(ln(Sales)), 총자산규모(ln(TA)), 기업특성에는 직원수 증감(△ Employee), 직원평균임금 증감(△ Salary), 최대주주지분율(MajorOwn(%)), 최대주주지분율 증감(△ MajorOwn), 거시경제지표에는 CD 유통수익율(CD), GDP, 원/달러 환율(USD)(FX), KOSPI 종가지수(KOSPI), KOSDAQ 종가지수(KOSDAQ), 소비자물가지수 증감(△ CPI), 생산자물가지수 증감(△ PPI), 국고채3년(Bond rate(%))으로 구성되어 있으며(표 2), 총 32개의 특성으로 구분하여 Explainable AI의 Integrated Gradients 방법론으로 Feature Attribution을 분석하고 시각화에 적용하였다. 또한 분석된 Feature Attribution을 바탕으로 상위 10개의 특성을 가지고 F1 Score, Accuracy를 측정하여 본 논문에서 제시한 전체 특성을 가지고 기업부도를 예측한 결과와 비교 분석하였다.

### 3-3 훈련 데이터셋(Training Dataset) 구성

훈련 데이터셋 구성은 각 총 3개의 데이터셋으로 구분하였으며, Set 1은 기업의 재무적 특성인 건전성과 수익성, 성장성, 유동성, 활동성, 규모를 반영한 재무비율로 구성하였고 Set 2는 Set 1에 기업특성으로 직원수 증감, 직원평균임금 증감, 최대주주지분율, 최대주주지분율 증감을 추가하였다. Set 3는 Set 2에 거시경제지표로 CD유통수익율, GDP, 원/달러 환율(USD), KOSPI 종가지수, KOSDAQ 종가지수, 소비자물가지수 증감, 생산자물가지수 증감, 국고채3년을 추가하여 훈련 데이터셋을 구성하였다.

표 3. 훈련 데이터셋

Table 3. Training dataset

Set 1	Set 2	Set 3
financial ratios	financial ratios	financial ratios
-	corporate characteristics	corporate characteristics
-	-	macroeconomic indicators

### 3-4 딥러닝(Deep Learning) 모델

선행연구에서는 인공지능 기반 기업부도 예측 모델 중 통계 기반인 머신러닝에서는 SVM, RF 모델이 우수한 예측력을 보였고 딥러닝 모델에서는 DNN, CNN, LSTM 모델이 높



은 예측력을 보였다[1]. 하지만 딥러닝 기반의 경우 고차원의 복잡한 데이터와 파라미터들 간에 복잡한 상관관계를 학습하고 이를 통해 예측이나 결론을 도출하기 때문에 모델들이 학습한 패턴과 결론을 설명하기가 어렵고 사람들이 이해하기 어려운 블랙박스 형태이기 때문에 그 결론을 설명하거나 이유를 명확히 파악하기가 더욱더 어렵다.

이에 본 논문에서는 딥러닝 기반의 DNN, CNN, LSTM 모델을 활용하고 Integrated Gradients 방법론으로 각 기업의 어떤 특성들이 기업부도 예측에 기여도를 높였는지 분석하고 시각화하여 블랙박스 형태의 딥러닝 학습 방법의 신뢰성과 투명성을 제공하고자 한다. 또한 Integrated Gradients로 분석된, 기여도가 높은 상위 10개 특성을 가지고 다시 기업부도 예측을 실시하여 본 연구에서 제한한 최대 32개의 특성을 가지고 예측한 결과와 비교하고자 한다.

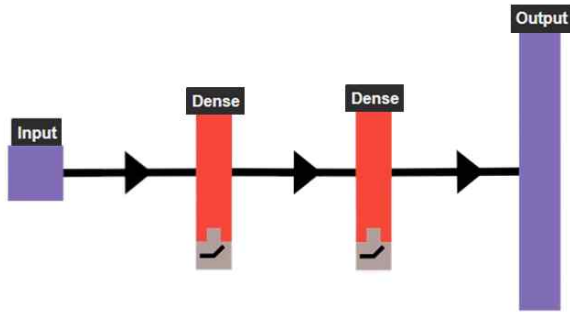


그림 1. DNN의 구조  
Fig. 1. The structure of a Deep Neural Network

DNN은 Input Layer와 Output Layer 사이에 여러 개의 Hidden Layer 구조를 가지고 있으며 본 연구에서는 2개의 Dense Layer를 사용하였다(그림 1). DNN의 경우 다른 신경망에 비해 변수 간의 비선형 조합이 가능하고, 범주형 변수에 상관없이 분석할 수 있지만, 신경망 가중치의 의미를 정확히 해석하기가 어렵기 때문에 해석이 어려운 단점이 존재한다.

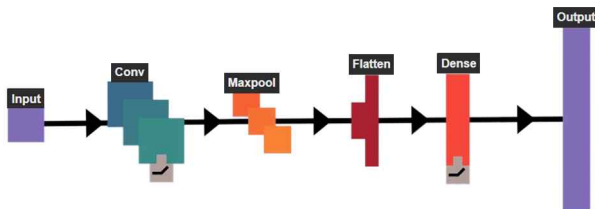


그림 2. CNN의 구조  
Fig. 2. The structure of a Convolutional Neural Network

CNN은 이미지 패턴을 분석에 유용한 알고리즘으로 Convolutional Layer와 DNN 모델이 결합된 형태로 본 연구에서는 기업데이터를 이미지화하여 Input(입력)으로 사용하였으며 Convolutional Layer를 통해 이미지의 공간적인 정보

를 추출하고 이를 DNN의 입력데이터로 사용함으로써 이미지를 분류하였다(그림 2).

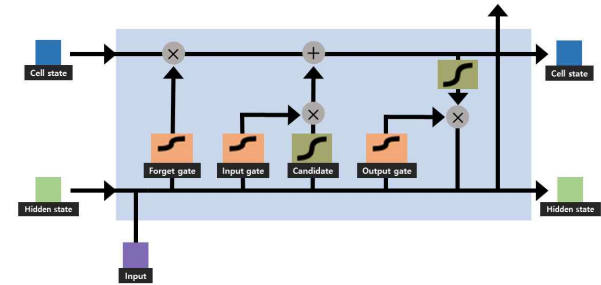


그림 3. LSTM의 구조  
Fig. 3. The structure of Long Short-Term Memory

LSTM은 RNN의 hidden State에 cell state를 추가한 구조를 통해 RNN의 Long-term dependencies 문제를 해결한 방법이다. LSTM은 4단계를 통해 Output을 출력하는데 Forget gate layer에서는 Cell state로부터 어떤 정보를 버릴 것인지 정하고, Input gate layer는 새로운 정보 중 어떤 것을 Cell state에 저장할 것인지를 정한다. 그다음 Cell state에서는 업데이트를 진행한 후에 Output gate layer를 통해 최종 출력값을 출력한다. 본 연구에서는 LSTM 모델의 이러한 특징을 바탕으로 기업부도 예측에 높은 성능을 발휘하는 것을 기대하였다.

### 3-5 Integrated Gradients

Explainable AI 분야는 Input feature(입력 특성)가 모델의 Output(결과)에 얼마나 기여를 했는지 나타내주는 Attribution(기여도)을 파악하는 방법론으로 다양한 연구가 진행되고 있다. 그 중 Integrated Gradients 방법론은 Attribution의 중요한 Axioms(공리)에 대해 Sensitivity(감도)와 Implementation invariance(구현 불변)를 만족하는 Feature Attribution을 구하는 방식으로 이미지처리, 자연어처리, 정형 데이터 등 다양한 분야에 적용이 가능하다는 특징이 있다[20].

Feature Attribution란 그 특성이 모델의 결과에 기여한 정도 혹은 영향도를 정량화 하는 기술이다. Feature Attribution을 통해 Input feature들 중 인공지능 모델의 판단에 영향을 끼친 특성을 찾아 인공지능의 판단 근거를 설명할 수 있다. 만약 Feature Attribution의 절대값이 크다면 해당 특성이 결과에 기여한 바가 크다는 의미이고, 기여도가 0에 가깝다면 그 특성이 결과에 미친 영향이 미미하다는 의미이다. 또한 Feature Attribution이 양수라면 해당 특성이 모델의 결과에 긍정적인 영향을 행사한 것이고, 기여도가 음수라면 부정적인 방향으로 영향을 미쳤다는 것을 의미한다.

Integrated Gradients는 Attribution을 파악하기 위해 Baseline을 사용하고 Input feature간의 차이를 비교하여 계산하며 Sensitivity와 Implementation invariance를 만족하

여야 한다. Sensitivity는 Baseline과 Input의 차이가 오직 하나의 Feature이고 그 둘의 예측 결과가 다르다면, 차이나는 Feature가 모델의 예측에 영향을 끼쳤다고 말할 수 있으며, Feature의 영향이 0이 아니라면 Sensitivity 조건을 만족한다. Implementation invariance는 서로 다른 Network라도 Input 과 Output의 관계가 형성된다면 두 Network는 동일한 Attribution을 가져야 한다. Integrated Gradients는 이러한 조건을 만족시키는 방법론이며, Baseline에서 Input까지의 모든 Gradient를 고려하는 방법으로 특정 지점에서 Gradient값이 0이 되는 Sensitivity를 해결할 수 있으면서 Gradient를 활용한다. Integrated Gradients의 수식은 다음과 같이 정의 할 수 있다.

$$IG_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha \quad (1)$$

본 연구에서는 기업부도 예측에 사용된 재무비율, 기업특성, 거시경제 등 총 32개의 특성을 이용하여 Integrated Gradients 방법론을 적용하고 어떠한 특성들이 기업부도에 예측에 영향을 주었는지 분석 및 시각화 하고자 한다. 또한 기업부도예측에 기여도가 높은 상위 특성 10개를 가지고 F1-Score, Accuracy를 다시 측정하여 비교하고자 한다.

## IV. 모델 검증 방법 및 분석결과

### 4-1 평가지표

본 연구에서는 F1 Score와 Accuracy(정확도)를 사용하여 기업부도예측의 평가 지표로 사용한다. 먼저 F1 Score는 Precision(정밀도)과 Recall(재현율)의 조화 평균이며 Precision(정밀도)과 Recall(재현율)로 F1 Score가 구성되어 불균형한 데이터에서 잘 동작하는 평가 지표이다.

이는 Goutte, Gaussier(2005)에서 F-Score의 신뢰도를 실험적으로 검증하였다[21]. 실험 결과, F-Score는 거대 데이터 셋에서 무작위로 분할시킨 다른 데이터셋에도 적용될 수 있었고 F1 Score는 F Score의 베타 값을 1로 해서 쓴 식이다.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

Accuracy는 인공지능 모델을 평가하는데 가장 직관적으로 모델의 성능을 보여주는 평가 지표이며, True를 True라고 예측한 경우, False를 False라고 옳게 예측한 경우 모두 고려하며 일반적으로 Classification Metrics에서 성능 평가 척도로 사용되고 있다.

표 4. 분류 기준

Table 4. Classification metrics

Division	True	False
True	True Positive(TP)	False Positive(FP)
False	False Negative(FN)	True Negative(TN)

가로 행은 모델이 예측한 결과를 나타내며 세로 열은 실제 결과를 의미한다(표 4). Classification Metrics(분류 기준)에서 Accuracy를 나타내면 수식은 다음과 같다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

Accuracy는 인공지능 모델의 정확도를 측정할 때 주로 사용되었던 평가지표로 본 연구에서도 F1 Score와 함께 평가 지표로 사용하고자 한다.

### 4-2 데이터 전처리 <- 설명하기

본 연구에서는 K-fold Cross Validation을 사용하고자 한다. 일반적으로 선행 연구에서는 전체 데이터를 8대 2 비율로 나누어 8을 Training Set(훈련세트), 2를 Test Set(테스트 세트)로 사용하고, 다시 Training Set을 8대 2로 나누어 2를 Validation Set(검증세트)로 사용하였다.

이는 전체 데이터의 64%만 Training Data(훈련데이터)로 사용하기에 Training Data로 사용하는 Sample(표본)의 개수가 부족하여 인공지능 모델이 다양한 특성을 학습하기 어려워 높은 성능의 인공지능 모델을 만들기 어렵다. 이에 데이터를 충분하게 확보하기 어려운 상황에서는 데이터의 특성을 유지하고 양을 늘리는데 사용하는 기법이 Cross Validation(교차 검증) 방법이다. K-fold는 Training Data Set을 K 등분하여 그 중 1/K는 Validation Set으로 사용하고 (K-1)/K는 Training Set로 사용하여 각 등분된 세트에 대해 총 K번 반복한다. 이를 통해 K 만큼의 모델을 만들고 각 모델의 MSE값을 평균 내어 해당 모델의 MSE값을 결정하게 한다. 본 연구에서는 5-fold를 사용한다.

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (4)$$

Data Set은 부도 발생 1년 전의 재무비율, 재무비율 + 기업특성, 재무비율 + 기업특성 + 거시경제로 구성하여 총 3개의 Data Set로 구성하였고 인공지능 모델은 DNN, CNN, LSTM을 사용하여 F1 Score 및 Accuracy를 측정하였다.

표 5. 예측 결과

Table 5. Result of prediction

Division	Set 1		Set 2		Set 3	
	F1	Acc	F1	Acc	F1	Acc
DNN	92.90	92.57	90.99	90.47	99.56	99.56
CNN	90.05	89.48	91.34	90.87	99.64	99.64
LSTM	93.12	92.97	93.09	92.85	99.72	99.72

#### 4-3 기업부도 예측 결과

Set 1은 기업의 재무비율 특성을 가지고 예측한 결과이다. F1 Score와 Accuracy를 살펴보면 **DNN, CNN, LSTM 모두 준수한 예측력을** 보였으며, **LSTM 모델의 경우 좀 더 우수한 성능을** 보였다. Set 2는 재무비율 특성에 기업 특성을 더하여 데이터셋을 구성하였으며 기업적 특성이 기업부도에 영향을 줄 것으로 가정하였으나 실험 결과 Set 1과 비슷하거나 F1 Score 및 Accuracy가 조금 내려가는 경향이 있었다. Set 3의 경우는 거시경제가 기업 부도에 영향을 줄 것이라고 가정하고 **Set 2에 거시경제를 추가**하였다. 실제 IMP 이후 기간에 기업 부도 추이가 높았고 2008년 글로벌 경제위기 이후 기간인 2009년에서 2011년 사이에 집중적으로 부도 기업이 발생하였다. 거시경제지표가 추가된 Set 3에서는 F1 Score 및 Accuracy가 높은 성능을 보였다. 이는 **기업부도에** **측에서 거시경제지표가 미치는 영향이 크다고** 해석할 수 있다. 이처럼 기업부도예측에 우수한 성능을 보이는 딥러닝 모델을 제안했지만 실제 딥러닝 모델이 어떤 특성을 기반으로 어떻게 학습이 이루어져 결론에 도달하였는지 알 수 없다.

이에 따라 본 논문에서 Integrated Gradients 방법론을 활용하여 어떠한 특성들이 기업부도예측에 기여를 하였는지 분석하고 시각화 하여 모델의 높은 예측율에 따른 신뢰성을 확보하고자 한다.

#### 4-4 Feature Attribution 분석 피쳐 기여도

본 연구에서는 Set 1, Set 2, Set 3으로 구성된 훈련 데이터셋을 DNN, CNN, LSTM 딥러닝 모델로 학습시켜 우수한 기업부도 예측력을 보였으며, 그에 따라 인공지능 모델이 부도라고 예측한 결과를 토대로 Integrated Gradients를 측정하였다. 그림 4, 그림 5, 그림 6에서는 Feature Attribution 값이 양수일수록 Input Feature가 부도라고 예측한 결과에 긍정적인 영향을 미친 것이고, 음수일수록 예측 결과에 부정적인 방향인 비부도로 예측하게 영향을 미쳤다는 것을 나타내며 Deep Learning 모델이 데이터를 학습하는 과정에서 5-fold Cross Validation 사용하였다. 그에 따라 Feature Attribution 분석 및 시각화에서도 fold 별로 특성의 분포를 나타냈으며, 각 분석된 특성을 절대값으로 환산하고 5-fold 평균을 구하여 어떤 특성들이 기업부도 예측에 기여를 했는지 분석하였다.

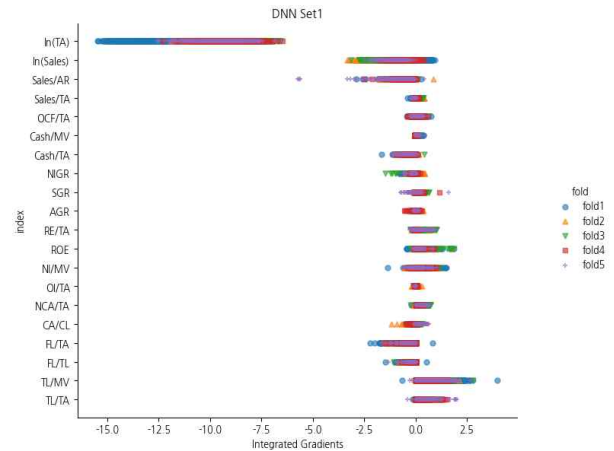


그림 4. Set 1에서 DNN의 Integrated Gradients  
Fig. 4. Integrated Gradients in Set 1 based on DNN

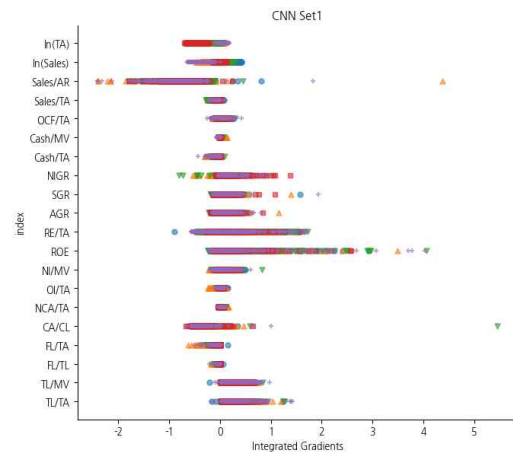


그림 5. Set 1에서 CNN의 Integrated Gradients  
Fig. 5. Integrated Gradients in Set 1 based on CNN

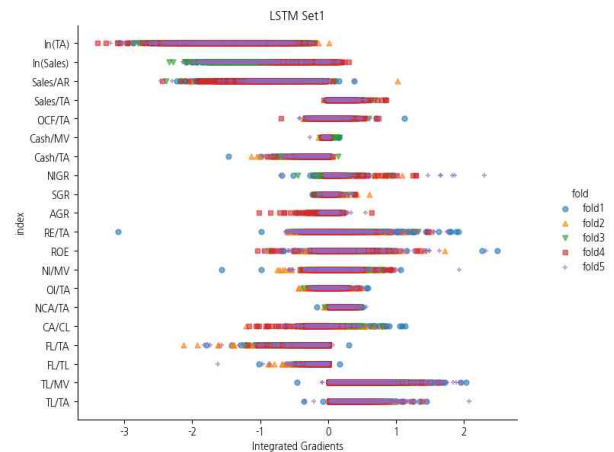


그림 6. Set 1에서 LSTM의 Integrated Gradients  
Fig. 6. Integrated Gradients in Set 1 based on LSTM

Set 1의 경우 CNN 모델의 비해서 상대적으로 예측력이 높았던 DNN, LSTM 모델의 특성 기여도는 총자산규모(In(TA)), 총매출액규모(In(Sales)), 시장부채비율(TL/MV), 매출채권회전율(Sales/AR), 부채비율(TL/TA), 고정자산비율(NCA/TA) 순으로 비슷하게 측정되었으며 Set 1의 Integrated Gradients 비교 분석은 표 6과 같다. 또한 표 6, 표 7, 표 8은 Integrated Gradients 값이 평균적으로 높은 순으로 나열하였다.

표 6. Set 1에서 Integrated Gradients 비교

Table 6. Comparison of Integrated Gradients in Set 1

DNN	CNN	LSTM
In(TA)	Sales/AR	In(TA)
In(Sales)	Current Ratio	In(Sales)
TL/MV	TL/MV	Sales/AR
Sales/AR	TL/TA	TL/MV
TL/TA	In(TA)	TL/TA
NCA/TA	ROE	NCA/TA
Cash/TA	Sales/TA	FL/TL
FL/TL	Cash/TA	Cash/TA
FL/TA	In(Sales)	Sales/TA
ROE	FL/TL	ROE
NI/MV	NIGR	Current Ratio
NIGR	RE/TA	FL/TA
RE/TA	NI/MV	NI/MV
Cash/MV	NCA/TA	NIGR
Current Ratio	FL/TA	AGR
SGR	SGR	Cash/MV
OCF/TA	AGR	OCF/TA
Sales/TA	Cash/MV	OI/TA
AGR	OI/TA	SGR
OI/TA	OCF/TA	RE/TA

Set 2는 Set 1에 직원평균임금 증감( $\Delta$  Salary), 직원수 증감( $\Delta$  Employee), 최대주주지분율(MajorOwn(%)), 최대주주지분율 증감( $\Delta$  MajorOwn)의 기업특성이 추가된 데이터셋이며, DNN, CNN, LSTM의 Integrated Gradients 측정 결과를 시각화하면 그림 7, 그림 8, 그림 9와 같다.

추가된 기업특성 중 최대주주지분율(MajorOwn(%))의 경우 Feature Attribution이 높게 측정되었으며, 직원평균임금 증감( $\Delta$  Salary), 직원수 증감( $\Delta$  Employee), 최대주주지분율 증감( $\Delta$  MajorOwn)의 Feature Attribution은 낮게 측정되었다.

Set 2의 Integrated Gradients 비교 분석은 표 7과 같으며, 기업특성이 추가된 Set 2의 경우 Set 1에 비해 기업부도 예측률이 표 5와 같이 다소 떨어지는 경향이 있었다. 이는 기업특

성 자체가 기업부도예측에 있어서 Feature Attribution이 높지 않다는 점을 시사한다.

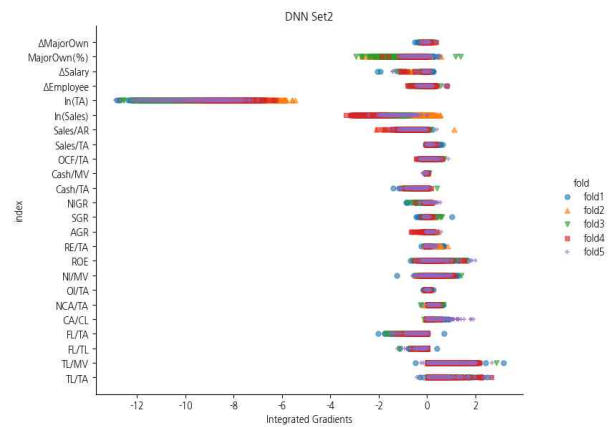


그림 7. Set 2에서 DNN의 Integrated Gradients  
Fig. 7. Integrated Gradients in Set 2 based on DNN

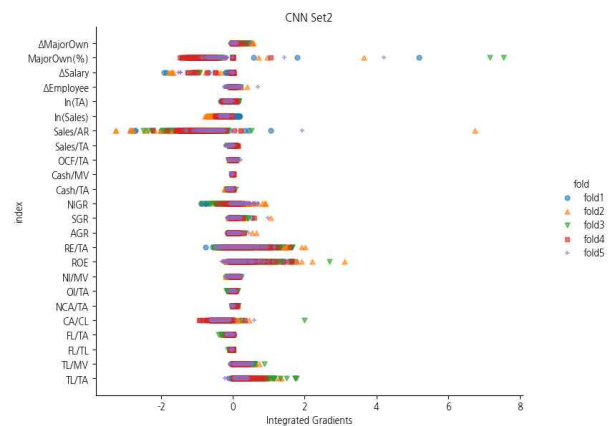


그림 8. Set 2에서 CNN의 Integrated Gradients  
Fig. 8. Integrated Gradients in Set 2 based on CNN

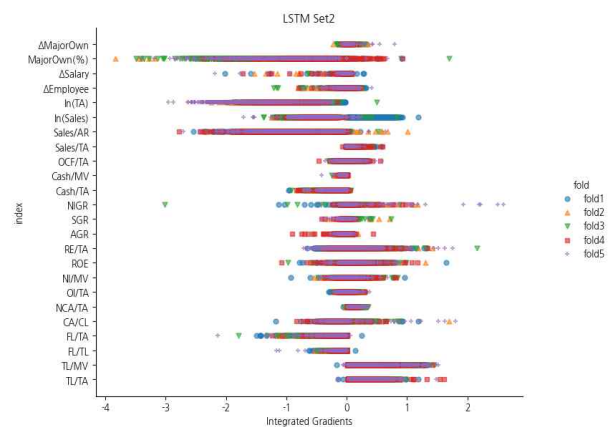


그림 9. Set 2에서 LSTM의 Integrated Gradients  
Fig. 9. Integrated Gradients in Set 2 based on LSTM



표 7. Set 2에서 Integrated Gradients 비교

Table 7. Comparison of Integrated Gradients in Set 2

DNN	CNN	LSTM
In(TA)	Sales/AR	In(TA)
In(Sales)	MajorOwn(%)	MajorOwn(%)
MajorOwn(%)	Current Ratio	Sales/AR
TL/MV	TL/TA	TL/MV
Sales/AR	TL/MV	In(Sales)
TL/TA	In(Sales)	TL/TA
NCA/TA	ROE	Cash/TA
Cash/TA	In(TA)	FL/TL
FL/TL	Sales/TA	NCA/TA
Current Ratio	NCA/TA	Current Ratio
Sales/TA	Cash/TA	Sales/TA
FL/TA	FL/TL	FL/TA
ROE	△ Salary	ROE
NI/MV	△ MajorOwn	Cash/MV
△ Salary	RE/TA	NIGR
RE/TA	AGR	△ Salary
Cash/MV	NI/MV	AGR
AGR	FL/TA	NI/MV
SGR	△ Employee	RE/TA
OCF/TA	SGR	△ Employee
△ MajorOwn	Cash/MV	△ MajorOwn
OI/TA	OCF/TA	OCF/TA
NIGR	OI/TA	OI/TA
△ Employee	NIGR	SGR

Set 3은 Set 2에 CD 유통수익률(CD), GDP, 원/달러 환율(USD)(FX), KOSPI 종가지수(KOSPI), KOSDAQ 종가지수(KOSDAQ), 소비자물가지수 증감(△CPI), 생산자물가지수 증감(△PPI), 국고채3년(Bond rate(%))의 거시경제지표가 추가된 데이터셋이며, DNN, CNN, LSTM의 Integrated Gradients 측정 결과를 시각화하면 그림 10, 그림 11, 그림 12와 같다.

거시경제지표가 추가된 Set 3에서 DNN, CNN, LSTM 모델 모두 표 6과 같이 가장 우수한 기업부도 예측률을 보였으며, 3개의 Deep Learning 모델에서 Integrated Gradients 측정값은 Set 2와 그 결과가 비슷한 양상이었다. 거시경제지표 특성들의 Integrated Gradients 측정값이 높지 않기 때문에 그 특성들이 기업부도 예측에 기여한 정도가 높다고 할 수 없으나, 거시경제지표가 추가된 Set 3에서 기업부도 예측률을 높였다는 것은 거시경제지표가 기업부도예측에 있어서 유의미한 특성으로 사용되어질 수 있다는 것을 시사한다.

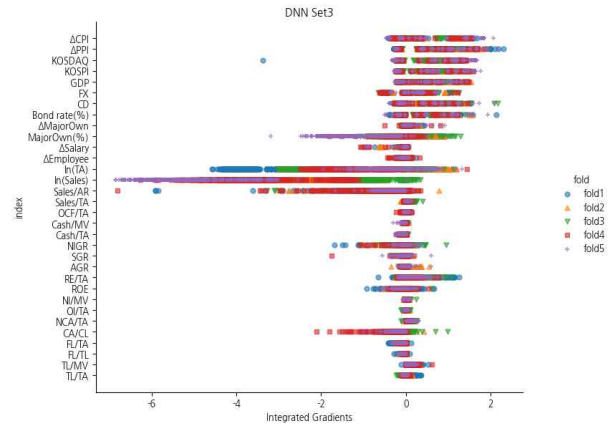


그림 10. Set 3에서 DNN의 Integrated Gradients  
Fig. 10. Integrated Gradients in Set 3 based on DNN

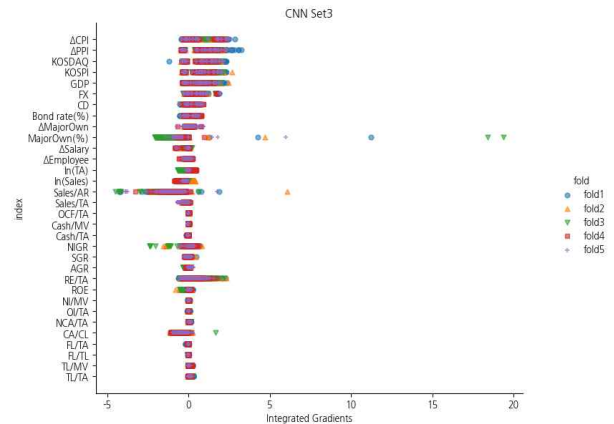


그림 11. Set 3에서 CNN의 Integrated Gradients  
Fig. 11. Integrated Gradients in Set 3 based on CNN

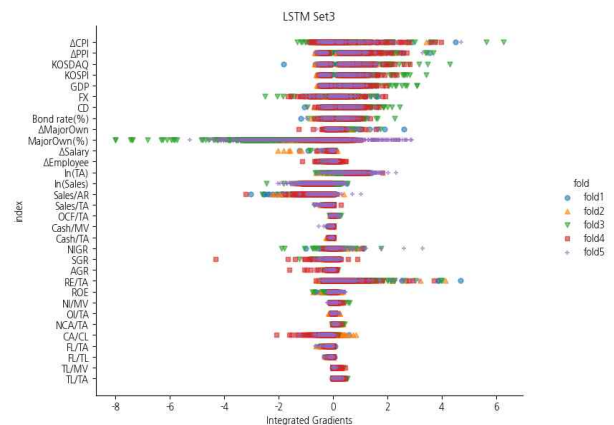


그림 12. Set 3에서 LSTM의 Integrated Gradients  
Fig. 12. Integrated Gradients in Set 3 based on LSTM

표 8도 Set 3에서 각 특성들의 Integrated Gradients 평균값이 높은 순으로 정리하였다. 위에서 언급한 바와 같이 Set 3에서 가장 높은 기업부도 예측률을 보였으며, 그에 따라

거시경제지표가 추가된 데이터 시각화에서 각 특성의 분포 정도에 따라 각 특성들이 기업부도 예측에 기여한 정도를 직관적으로 알 수 있었다.

표 8. Set 3에서 Integrated Gradients 비교

Table 8. Comparison of Integrated Gradients in Set 3

DNN	CNN	LSTM
In(Sales)	Sales/AR	MajorOwn(%)
In(TA)	MajorOwn(%)	Sales/AR
Sales/AR	Current Ratio	In(TA)
MajorOwn(%)	Sales/TA	In(Sales)
TL/MV	In(Sales)	Sales/TA
Current Ratio	TL/TA	TL/TA
NCA/TA	△ PPI	△ PPI
KOSDAQ	TL/MV	NCA/TA
TL/TA	In(TA)	TL/MV
FL/TL	CD	KOSDAQ
Cash/TA	△ CPI	KOSPI
NIGR	NCA/TA	Current Ratio
KOSPI	GDP	GDP
△ PPI	SGR	RE/TA
Sales/TA	Bond rate(%)	CD
CD	NIGR	Bond rate(%)
ROE	Cash/TA	△ CPI
Bond rate(%)	FX	FL/TA
GDP	KOSPI	FX
FL/TA	KOSDAQ	SGR
SGR	RE/TA	Cash/TA
△ Salary	ROE	FL/TA
RE/TA	FL/TL	Cash/MV
△ CPI	△ Employee	AGR
Cash/MV	AGR	△ Salary
△ MajorOwn	△ Salary	△ MajorOwn
NI/MV	FL/TA	NIGR
AGR	△ MajorOwn	ROE
OCF/TA	NI/MV	NI/MV
△ Employee	OCF/TA	△ Employee
FX	Cash/MV	OCF/TA
OI/TA	OI/TA	OI/TA

#### 4-5 Feature Attribution 분석을 통한 기업부도 예측 결과

Explainable AI 기반 Feature Attribution 분석을 위해 Integrated Gradients 방법론으로 분석하였고, 분석 결과 중 기업부도 예측에 기여한 상위 10개의 특성을 가지고 추가적으로 기업부도 예측을 실시하여 본 연구에서 사용한 32개의 특성과 비교 분석하여 그 결과에 대해 비교하고자 하였다.

표 9. 예측 결과

Table 9. Result of prediction

Division	Set 1		Set 2		Set 3	
	F1	Acc	F1	Acc	F1	Acc
DNN	90.94	90.31	90.61	90.04	88.88	88.09
CNN	81.70	79.95	87.89	87.10	99.64	99.64
LSTM	90.27	89.76	91.22	90.79	99.68	99.68

표 9의 예측 결과는 4-3 장 기업부도 예측에서 사용한 동일한 방법으로 데이터 구성, 평가지표, 데이터전처리를 실시하여 DNN, CNN, LSTM 딥러닝 모델의 결과를 예측하였다. Set 1, Set 2, Set 3 데이터셋 결과를 앞서 실시한 기업부도 예측 결과(표 5)와 비교하면 평균적으로 큰 차이를 보이지 않았지만 CNN 모델의 경우 Set 1, Set 2에서 F1 Score 와 Accuracy 예측 값이 다소 떨어지는 경향이 있었고 Set 3에서는 동일한 성능을 보였다. DNN 모델이 경우 Set 3에서 예측 값이 다소 떨어지는 경향도 보였지만 CNN과 LSTM 모델은 Set 3에서 모두 우수한 성능을 보였다.

이는 다양한 특성들을 바탕으로 기업부도 예측을 실시하여 어떠한 특성들이 기업부도 예측에 영향을 주었는지 Feature Attribution을 분석하고, 분석된 기여도를 바탕으로 기업부도 예측 연구를 실시한다면 보다 성능 좋은 기업부도 예측 모델을 만들 수 있을 것이다. 특히 인공지능이 어떠한 특성을 가지고 학습하고 그 결과를 예측하였는지 설명이 필요한 주요 분야에서는 더욱더 깊이 연구할 필요성이 있을 것으로 생각된다.

#### 머신러닝과 ai의 최대 한계점

-> 블랙박스

#### V. 결 론

-> 해결책 : XAI - 설명가능한 ai

본 연구에서는 인공지능 모델 Input Feature의 Feature Attribution을 측정하는 Integrated Gradients 방법론을 적용하였다. Feature Attribution을 통해 Input Feature들 중 인공지능 모델의 판단에 영향을 끼친 특성을 찾아 인공지능의 판단 근거를 설명할 수 있다. 만약 Feature Attribution의 절대값이 크다면 해당 특성이 결과에 기여한 바가 크다는 의미이고, 기여도가 0에 가깝다면 그 특성이 결과에 미친 영향이 미미하다는 의미이다. 또한 Feature Attribution이 양수라면 해당 특성이 모델의 결과에 긍정적인 영향을 행사한 것이고, 기여도가 음수라면 부정적인 방향으로 영향을 미쳤다는 것을 의미한다.

기업부도 예측 모델의 훈련 데이터는 건전성, 수익성, 성장성, 유동성, 활동성, 규모의 총 6가지로 구분한 재무비율과 기업특성, 거시경제지표를 이용하였다. 재무비율은 연간재무제표의 계정과목에서 추출하였으며, 기업특성 데이터는 기업부도에 영향을 미칠 수 있는 직원 평균 임금의 증감, 최대주주 지분율과 같이 공시된 기업 현황을 이용하였다. 거시경제지표

의 경우 실제로 2008년 세계 금융 위기, 2020년 코로나19 팬데믹 상황에 주요한 거시경제지표가 하락세를 띠자 기업의 부도 발생 건이 증가하는 추세였다. 따라서 주요 거시경제지표 중 소비자물가지수(CPI) 증감, 국내총생산, 원/미국달러 매매기준율 등을 훈련 데이터로 채택하였으며 기업특성 데이터와 마찬가지로 기업부도에 영향을 미친다는 가정 하에 이용하였다.

Set 1에서는 인공지능 모델이 부도라고 예측한 결과에 대해서는 건전성이 높은 영향을 미쳤으며 반대로 인공지능 모델이 비부도라고 예측한 결과에 대해서는 기업의 규모, 활동성이 많은 영향을 미쳤다. 즉, Set 1에서는 기업의 건전성에 대한 재무비율이 좋지 않을수록 모델이 부도라고 예측하는 데에 많은 영향을 미쳤으며 기업 규모, 활동성에 대한 지표가 좋을수록 비부도라고 예측한 결과에 높은 기여를 했다고 볼 수 있다.

기업 특성에 대한 정보가 추가된 Set 2에서는 Set 1과 같이 기업의 규모, 활동성에 대한 지표가 좋을수록 기업부도 예측모델이 비부도라고 예측하는 데에 많은 영향을 미쳤고, 최대주주지분율과 건전성이 높을수록 기업부도 예측모델이 부도라고 예측하는 데에 높게 기여한다는 결과를 나타내었다.

거시경제지표가 추가된 Set 3에서도 앞선 Feature Attribution과 그 양상이 대체로 유사했다. 기업의 규모 및 활동성에 대한 재무비율은 인공지능 모델의 예측 결과가 비부도인 방향으로 크게 기여하였으며, 건전성 및 최대주주지분율 같은 특성이 인공지능 모델이 부도라고 예측하는 데에 높은 영향을 미쳤다. 또한 소비자물가지수 증감, 생산자물가지수 증감과 같은 거시경제지표 또한 인공지능 모델이 부도라고 예측하는 데에 영향을 미쳤다고 볼 수 있다. 다만 거시경제지표는 기업부도 예측에 기여하는 정도가 기업의 건전성, 활동성 등에 비해 상대적으로 낮은 결과를 보였다. 그럼에도 거시경제지표를 추가한 인공지능 모델의 예측력이 우수하므로 기업의 재무비율 뿐만 아니라 기업 외적인 변수인 거시경제지표의 영향도 무시할 수 없다고 해석할 수 있다.

본 연구에서 Integrated Gradients 방법론으로 Deep Learning 모델의 블랙박스 형태를 해석한 것에 대하여 의의를 둘 수 있으나 Integrated Gradients는 Deep Learning 모델을 설명하기 위한 방법론으로 Machine Learning에는 적용이 불가능하기에 단순히 도출 결과를 설명하는 방법이지 모델 자체를 해석하고 설명하는 것이 아니라는 것은 한계점이라고 할 수 있겠다. 따라서 Integrated Gradients 방법론 뿐만 아니라 다른 XAI(Explainable Artificial Intelligence) 기법을 적용해 인공지능 모델을 해석한다면 인공지능 모델의 설명력을 향상시킬 수 있을 것이며 Machine Learning, Deep Learning 등의 모델에 상관없이 모든 인공지능 모델에 적용 가능한 XAI(Explainable Artificial Intelligence) 방법론에 대한 연구가 지속적으로 진행 된다면 보다 완전하게 인공지능 모델에 대한 설명이 가능할 것으로 기대한다.

## 감사의 글

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음.

## 참고문헌

- [1] J.-H. Park, G.-Y. Kim, G.-M. Jung, H. Lee, and H. Choi, "An Empirical Study on the Artificial Intelligence-Based Corporate Bankruptcy Prediction Model - Focusing on Improvements Reflecting the Characteristics of Time," *Journal of the Korea Society of Information Technology Policy & Management*, Vol. 15, No. 2, pp. 3179-3188, June 2023.
- [2] W. H. Beaver, "Financial Ratios as Predictors of Failure," *Journal of Accounting Research*, Vol. 4, pp. 71-111, 1966. <https://doi.org/10.2307/2490171>
- [3] E. I. Altman, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, pp. 589-609, September 1968. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- [4] J. A. Ohlson, "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131, 1980. <https://doi.org/10.2307/2490395>
- [5] J. A. McQuown, A Comment on Market vs. Accounting-Based Measures of Default Risk, KMV Corporation Working Paper, 1993.
- [6] S. K. Oh, Prediction of Corporate Bankruptcy Using Multivariate Discriminant Analysis Model and Stock Option Model, Korea Development Bank, Seoul, KDB Monthly Vol. 549, pp. 1-29, 2001.
- [7] C. W. Nam, T. S. Kim, N. J. Park, and H. K. Lee, "Bankruptcy Prediction Using a Discrete-Time Duration Model Incorporating Temporal and Macroeconomic Dependencies," *Journal of Forecasting*, Vol. 27, No. 6, pp. 493-506, September 2008. <http://dx.doi.org/10.1002/for.985>
- [8] M. H. Tinoco and N. Wilson, "Financial Distress and Bankruptcy Prediction among Listed Companies Using Accounting, Market and Macroeconomic Variables," *International Review of Financial Analysis*, Vol. 30, pp. 394-419, December 2013. <https://doi.org/10.1016/j.irfa.2013.02.013>
- [9] J. S. Lee and J. H. Han, "Usability Test of Non-Financial Information in Bankruptcy Prediction Using Artificial Neural Network," *Journal of Intelligence and Information Systems*, Vol. 1, No. 1, pp. 123-134, January 1995.

- [10] S.-H. Yeh, C.-J. Wang, and M.-F. Tsai, "Deep Belief Networks for Predicting Corporate Defaults," in *Proceedings of the 24th Wireless and Optical Communication Conference (WOCC)*, Taipei, Taiwan, pp. 159-163, October 2015. <https://doi.org/10.1109/WOCC.2015.7346197>
- [11] S. Addai, Financial Forecasting Using Machine Learning, African Institute for Mathematical Science(AIMS), Muizenberg, South Africa, May 2016.
- [12] S. Kim and H. Ahn, "Application of Random Forests to Corporate Credit Rating Prediction," *Journal of Industrial Innovation*, Vol. 32, No. 1, pp. 187-211, March 2016. <http://dx.doi.org/10.22793/indinn.2016.32.1.006>
- [13] S.-H. Min, "Bankruptcy Prediction Using an Improved Bagging Ensemble," *Journal of Intelligence and Information Systems*, Vol. 20, No. 4, pp. 121-139, December 2014. <http://dx.doi.org/10.13088/JIIS.2014.20.4.121>
- [14] N. Jo, H. Kim, and K. Shin, "Bankruptcy Type Prediction Using a Hybrid Artificial Neural Networks Model," *Journal of Intelligence and Information Systems*, Vol. 21, No. 3, pp. 79-99, September 2015. <http://dx.doi.org/10.13088/JIIS.2015.21.3.79>
- [15] N. Jo and K. Shin, "Bankruptcy Prediction Modeling Using Qualitative Information Based on Big Data Analytics," *Journal of Intelligence and Information Systems*, Vol. 22, No. 2, pp. 33-56, June 2016. <http://dx.doi.org/10.13088/JIIS.2016.22.2.033>
- [16] S. Cha and J. Kang, "Corporate Default Prediction Model Using Deep Learning Time Series Algorithm, RNN and LSTM," *Journal of Intelligence and Information Systems*, Vol. 24, No. 4, pp. 1-32, December 2018. <http://dx.doi.org/10.13088/JIIS.2018.24.4.001>
- [17] I. Lee and D. Kim, "An Evaluation of Bankruptcy Prediction Models Using Accounting and Market Information in Korea," *Asian Review of Financial Research*, Vol. 28, No. 4, pp. 625-665, November 2015.
- [18] J. W. Choi and S. K. Oh, "The Prediction of Corporate Bankruptcy by Combining Survival Analysis and KMV Model," *The Kon-Kuk Journal of Business and Economic Studies*, Vol. 41, No. 1, pp. 69-104, March 2016.
- [19] J. W. Choi, S. K. Oh, and J. W. Jang, "A Study on the Prediction of Corporate Bankruptcy Using Big Data and Artificial Intelligence Techniques," in *Proceedings of Korean Finance Association Autumn Conference*, Seoul, pp. 396-435, November 2017.
- [20] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in *Proceedings of the*

*34th International Conference on Machine Learning (ICML '17)*, Sydney, Australia, pp. 3319-3328, August 2017. <https://doi.org/10.48550/arXiv.1703.01365>

- [21] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in *Proceedings of the 27th European Conference on IR Research (ECIR 2005)*, Santiago de Compostela, Spain, pp. 345-359, March 2005. [https://doi.org/10.1007/978-3-540-31865-1\\_25](https://doi.org/10.1007/978-3-540-31865-1_25)



### 박중현(Joong-Hyun Park)

2004년 : 선문대학교 컴퓨터정보학부  
컴퓨터정보 (이학사)

2011년 : 선문대학교 대학원 컴퓨터공  
학과 (공학석사)

2023년 : 선문대학교 대학원 컴퓨터용  
합전자공학과 (공학박사수료)

2012년~2014년: (주)국제기술직업전문학교 대표이사

2014년~2019년: 한결직업전문학교 대표

2020년~현 재: 선문대학교 SW융합학부 연구교수

※관심분야 : 인공지능, 사물인터넷, 빅데이터 분석, 컴퓨터  
교육 등



### 김경영(Gyeong-Yeong Kim)

2023년 : 선문대학교 컴퓨터공학부  
빅데이터전공 재학

2019년~현 재: 선문대학교 컴퓨터공학부 빅데이터전공 재학

※관심분야 : 인공지능, 생성모델, 컴퓨터비전, 자연어처리 등



### 주진혁(Jin-Hyeok Ju)

2003년 : 호서대학교 세무회계학과  
(경영학사)

2006년 : 호서대학교 대학원 경영학과  
(경영학석사)

2010년 : 호서대학교 대학원 경영학과  
(경영학박사-재무회계)

2012년~2016년: 한국경영문화연구원 연구위원

2017년~2018년: 충남지역인적자원개발위원회 책임연구위원

2019년~2020년: ㈜에코텍 대표

2021년 9월~현 재: 선문대학교 SW융합학부 연구교수

※관심분야 : 기업재무, 기업도산, 빅데이터 분석 등



### 이현(Hyun Lee)



1998년 : 선문대학교 전자계산학과  
(이학사)

2002년 : 선문대학교 대학원 전자계산  
학과 (이학석사)

2010년 : The CSE Dept. of the Univ.  
of Texas at Arlington (공학  
박사)

2011년 : 대구경북과학기술원(DGIST) 로봇시스템 연구부 선  
임연구원

2021년~현 재: BK21+ 바이오 빅데이터 기반 충남 스마트  
클린 전문인력 양성사업단

2012년~현 재: 선문대학교 컴퓨터공학과 부교수

※관심분야 : 실시간 의사결정시스템, 자율컴퓨팅, 휴먼케어  
시스템, 바이오 빅데이터 분석 등 초거대 데이  
터 기반 인공지능 활용분야

### 최희정(Hee-Jung Choi)



2003년 : 이화여자대학교 경영학과 (경  
제학사, 경영학사)

2011년 : 이화여자대학교 대학원 경영  
학과 (재무석사)

2017년 : 고려대학교 대학원 경영학과  
(재무박사)

2003년~2010년: 포스코 재무실/자금관리실

2015년~2017년: 고려대학교 경영학과 강사

2017년~2020년: 국민연금연구원 기금정책분석실 부연구위원

2020년~현 재: 선문대학교 경영학과 조교수

※관심분야 : 기업재무, 기업지배구조, 기금정책 등