



[나이스 디앤비 챌린지] 골목대장팀_최종 보고서

1. 개요

- 1-1) 분석 목적
- 1-2) 배경 및 필요성
- 1-3) 대안 신용 평가 모델로서의 차별성 및 독창성

2. 외부 데이터 활용 파생변수

- 2-1. 미시경제 지표-기업 재무 지표
 - 2-1-1) Altman K-score
- 2-2. 거시경제 지표
 - 2-2-1) 환율 (Exchange)
 - 2-2-2) 발탁 운임 지수 (BDI, The Baltic Exchange)
 - 2-2-3) 곡물식량가격지수 (Cereals Price Index) & 식량가격지수 (Food Price Index)
 - 2-2-4) 국제 유가 (WTI, Brent, Dubai)
 - 2-2-5) 원자재 국내 선물가격 (Soft : 솜, 설탕 / Hard : 구리, 금, 은)
- 2-3. 인적 자원 지표
 - 2-3-1) 퇴사율 / 동종업계 내 퇴사율 백분위
 - 2-3-2) 예상평균연봉 / 동종업계 내 예상평균연봉 백분위
 - 2-3-3) 총 재직 인원
 - 2-3-4) 최근 1년 인원증가율 / 동종업계 내 인원증가율 백분위

3. 데이터셋 소개 및 전처리

- 3-1) 재무 데이터
 - 3-1-1) 결측치 처리
 - 3-2-2) 하나의 결산년월 선정
- 3-2) Baseline
- 3-3) 크롤링 데이터 전처리
 - 3-3-1) Altman K - score
 - 3-3-2) 환율
 - 3-3-3) 금리지표
 - 3-3-4) 국제유가
 - 3-3-5) 발탁운임지수(BDI)
 - 3-3-6) 곡물수입단가
 - 3-3-7) Food Price index
 - 3-3-8) Sugar, Cotton 추가
 - 3-3-9) 금, 구리, 은
 - 3-3-10) 크레딧잡 크롤링 데이터

4. EDA 탐색적 자료분석

- 4-1) 퇴사율 분포
- 4-2) 퇴사율과 인원증가율 사이의 연관성
- 4-3) 본점과 지점의 폐업 차이 (HDOF_BR_GB)
- 4-4) 국외투자법인여부 (FR_IVST_CORP_YN)
- 4-5) 벤처기업여부 (VENT_YN)
- 4-6) 홈페이지 url 보유 여부 (HOMEPAGE_URL)
- 4-7) 평균연봉 (avg_wage)

5. 분석 모형

- 5-1) 분석 모형 및 방법론
- 5-2) 분석 (모델링) 결과
- 5-3) 결과 해석
 - 변수 중요도(feature importance)
- 5-4) SHAP(Shapley Additive Exaplations) Value를 이용한 모델 해석

6. Reference

1. 개요

1-1) 분석 목적



미시경제, 거시경제, 인적자원 관점의 다양한 파생변수를 기반으로 중소기업의 휴폐업을 예측하고 유의미한 지표를 발굴하며 최적화된 휴폐업 예측 모델을 개발한다.

- 1 첫째, NICE DNB 제공 데이터, 논리적 근거에 기반한 다양한 파생변수로부터 휴폐업 결정에 유의미하다고 예상되는 변수를 조합 & 선별한다.
- 2 둘째, 휴폐업에 유의미한 영향을 미치는 요인들에 대한 요인분석을 진행한다.
- 3 셋째, 유의미한 변수를 활용하고 성능을 향상시킨 예측 모델 및 인사이트를 도출한다.

1-2) 배경 및 필요성

우리나라는 1997년 IMF 사태 이후 금융시장을 전면적으로 개방하고 사회 및 경제 전반에 걸친 구조조정을 실시하였으나 대기업 뿐만 아니라 벤처 및 중소기업의 휴폐업도 크게 증가하였다.

10년이 지난 2007년부터는 서브프라임 사태의 영향으로 세계적으로 경기 하락 및 침체가 이어져 상대적으로 기반이 취약한 국내 중소*벤처 기업의 휴폐업에 대한 위험이 가중되고 있다. 기업의 휴폐업은 경영자를 비롯하여 투자자, 채권자, 근로자, 거래처 등 직접적인 이해관계자에 대한 손실 뿐 아니라 사회, 경제 전반에 미치는 피해가 크다고 할 수 있다.

따라서, 기업 휴폐업 예측의 중요성은 더욱 부각되고 있다.

비재무 정보란?

재무제표에 나타나지 않는 기업 내부의 계량정보와 외부의 시장정보, 산업정보, 창업자에 대한 정보를 포함한 기업과 관련된 모든 정성적인 정보를 포함하는 개념으로 사용된다.

그런데 휴폐업 기업 중 재무적 요인은 좋지만 비재무적 요인이 좋지 않아 휴폐업된 기업, 재무적 요인은 좋지 않지만 비재무적 요인이 양호하여 지속되고 있는 기업의 수가 일정 부분 보이고 있다. [1]

즉, 단순히 재무적 요인만으로 휴폐업을 예측해서는 안 된다는 결론을 이끌어낼 수 있다.

특히, 중소*벤처 기업에서는 아래와 같은 이유로 재무적 요인과 더불어 비재무적 요인의 중요성을 인지해야 한다.

- 재무자료 수집이 어려운 중소, 창업기업의 부실예측을 위해서는 비재무정보에 관한 검토가 선행되어야 하나, 비재무정보는 그 범위가 매우 다양하며 이를 수집하기가 어렵다는 문제점을 가지고 있다.
- 중소기업은 대기업에 비해 회계정보의 신뢰성이 낮을 뿐만 아니라 재무구조가 취약하고 경영환경 변화에 민감하기 때문에 비재무정보를 평가에 반영해야 한다.
- 회계정보의 신뢰성이 낮은 중소기업에 대해서는 재무정보만을 이용한 모형보다는 비재무정보를 함께 분석한 모형이 더 신뢰성이 높은 것으로 나타났다. [2]

1-3) 대안 신용 평가 모델로서의 차별성 및 독창성

기존의 대안 신용 평가 모델은 주로 대기업이나 중견기업들과 관련된 재무적 요인 중심으로 이루어져, 기업 환경이 열악하고 재무적 요인이 상대적으로 미약한 중소*벤처 기업에 대해서는 정확성이 확보되지 못했다. [2]

- 따라서, 다양한 선행연구에 기반해 논리성을 갖춘 독창성 있는 파생변수를 통해 차별성을 확보했다.
 - 미시경제 관점 : 장부가치를 활용해 비상장기업이 많은 중소기업의 재무적 요인을 보다 정확하게 반영한다.
 - 거시경제 관점 : 핵심 키워드 “시장지향성” 을 바탕으로, 이에 영향을 주는 세부 지표 각각에 대한 선정
 - 인적자원 관점 : 측정이 어려운 중소기업의 ESG 성과를 내포하고 있는 인적자원 관점
- 결측치 처리 과정에서 KNN-Imputer와 K-Means Clustering 기법을 사용해 강건성(Robust)을 최대한 확보하고자 했다.
- 모델링 과정에서 boosting model 3가지 + bagging model 1가지 + 인공신경망 모델과의 비교를 통한 최종 모델을 선정했다. 또한, 성능 측정을 위한 다양한 지표를 사용했다.

2. 외부 데이터 활용 파생변수

파생변수

≡ 관점	Aa 변수	≡ 변수명
미시경제	Altman K-score	K_score K_부실여부판단
거시경제	환율 (Exchange).	환율
거시경제	발틱 운임 지수 (BDI, The Baltic Exchange).	BDI
거시경제	곡물식량가격지수 (Cereals Price Index) & 식량가격지수 (Food Price Index).	Corn_price Food_Price_Index Soybean_price Wheat_price
거시경제	국제 유가 (WTI, Brent, Dubai).	Brent Dubai WTI
거시경제	원자재 국내 선물가격 (Soft : 설탕, 설탕 / Hard : 구리, 금, 은).	Cotton_price Sugar_price gold_price silver_price
인적자원	회사율 / 동종업계 내 회사율 백분위	turnover_rate turnover_rate_ranking
인적자원	예산평균연봉 / 동종업계 내 예산평균연봉 백분위	avg_wage avg_wage_percent
인적자원	총 재직 인원	total_employee
인적자원	최근 1년 인원증가율 / 동종업계 내 인원증가율 백분위	employee_growth_rate employee_growth_rate_ranking

2-1. 미시경제 지표-기업 재무 지표

2-1-1) Altman K-score

• 정의

Altman(1996)은 한국 기업을 대상으로 연구한 도산예측모형을 제시했다.-모형 추정에 사용된 기업은 1989년~1992년의 기간 중 부실화된 34개 기업과 61개의 건전기업이다. 20개의 재무변수들을 검토한 후 최종적으로 기업규모, 총자산회전율, 누적수익성, 재무구조의 4개 변수들이 선정되었다.

$$K1score = -17.682 + 1.472X1 + 3.041X2 + 14.839X3 + 1.516X4$$

$X1$: $Log(\text{총자산}) \rightarrow$ 기업규모

$X2$: $Log(\text{매출액/총자산}) \rightarrow$ 총자산회전율

$X3$: 이익잉여금/총자산 \rightarrow 누적수익성

$X4$: 자기자본 장부가치/총부채 \rightarrow 재무구조

$K1score$	판정
$K1 < -2.00$	부실가능성 심각
$-2.00 \leq K1 \leq 0.75$	판정 유보
$0.75 < K1$	부실가능성 없음

• 배경

기업의 지속가능성 분석시 많이 사용되는 Altman Z-score 모델은 기업의 지속가능성을 예측하기 위한 많은 연구들 중에서 재무비율을 이용한 것으로 폭넓게 사용되고 있다. MDA(Multivariate Discriminant Analysis : 다변량판별분석) 방법 중 이 모델은 1년 내 약 95%, 2년 내 약 72%수준의 높은 예측력과 사용편의성 등으로 인해서 가장 폭넓게 채택되고 있다.

그리고 이를 한국 기업에 특화시켜 재정의한 모델이 Altman K-score이다. 해당 지표의 선정 배경은 다음과 같이 정리해볼 수 있다.

1. 한국 기업에 특화된 기업 도산 예측 모델로서 IMF 당시 높은 적중률을 보여주어 신뢰도가 확보되었다.
2. K1 score 모형은 재무구조(자기자본/총부채)에서 자기자본을 장부가치로 사용할 수 있는 모형으로 비상장기업이 존재하는 중소기업 데이터에 적용될 수 있다.
3. 단일변량분석과는 달리 여러 재무비율들을 동시에 고려할 수 있다.
4. 각 재무비율의 상대적 기여도를 확인할 수 있으며, 예측 결과를 바탕으로 각 재무비율에 부여할 가중치를 도출할 수 있다. [3]
5. 재무제표 데이터만을 기반으로 간단한 공식을 통해 도출할 수 있기 때문에 범용성, 타당성, 간결성 측면에서 장점을 갖는다.

• 데이터 수집 방법 및 출처

NICE DNB 측에서 제공된 "재무데이터.txt" 데이터에서 사업자 번호 별 가장 최근 날짜의 재무데이터를 사용한다.

2-2. 거시경제 지표

다양한 거시경제 지표의 활용배경을 "시장지향성"이라는 단어를 중심으로 전개한다.

기업의 지속가능을 위한 경쟁역량 창출과정에서 시장지향성은 핵심적인 요소로 대두되어왔다.

시장지향성(marketorientation)이란 기업이 지속적인 경쟁우위와 높은 수익을 실현하기 위해서 시장요구에 대해 경쟁자들보다 더 잘 대응해야 하며, 시장의 변화를 잘 예측해야 한다는 것을 의미한다. [4]

시장지향적인 문화는 고객에게 보다 높은 가치와 만족을 제공하여 기업이 경쟁우위를 유지하고 경쟁기업보다 고객에게 더 나은 가치를 제공함으로써 경영성과를 높이는데 중요한 요인이라고 할 수 있다. [5][6]

이러한 시장지향성은 경영 환경이 국내와 다른 해외 시장에서 더 주요한 요인이 된다. 한국 수출업체 중 상당수가 중소기업이라는 점에서 중소기업의 해외 시장에 대한 시장지향성이 중소기업의 지속성에 유의미한 영향을 줄 것으로 기대한다. [7]

중소기업은 대기업에 비해 상대적으로 자원이 부족해 수출 활동에 여러 장벽이 존재하는 한편, 조직 유연성을 바탕으로 환경 변화에 신속하게 대응할 수 있다는 점에서 상대적 우위가 존재한다(Chen & Hambrick, 1995; Fiegenbaum & Karnani, 1991). 이러한 중소기업의 특성은 “수출시장지향성”을 보다 효과적으로 활용할 수 있게 한다.



중소기업에게 “수출시장지향성”은 중요한 경쟁우위 요소이다.

따라서, 해외 시장을 겨냥한 “수출시장지향성”에 영향을 미치는 대표적인 거시경제 변인 환율, 발틱 운임 지수, 곡물식량가격, 국제 유가, 원자재 지표를 선정한다.

2-2-1) 환율 (Exchange)

• 정의

서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인이다.

원-달러 환율 상승	수출 증가 & 수입 감소
원-달러 환율 하락	수출 감소 & 수입 증가

• 배경

환율은 다른 거시경제 변수에 비해 변동 폭과 빈도가 높고 국내정책보다 대외환경이 더 많은 영향을 준다. 이 때문에 외환위기 이후 환율은 우리나라 거시경제적 주요 변수 가운데 경제적 영향이 가장 큰 변수가 되었다. [8]

또한, 기업 규모별 환율의 영향에 관한 선행연구에서 규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응하는 것으로 나타났다. 또한 기업 규모별 개별기업의 요소소득 중에서는 규모가 작은 수출기업의 자본소득이 가장 민감하게 반응하는 것으로 나타났다. [9]

• 데이터 수집 방법 및 출처

한국은행 경제통계시스템(<https://ecos.bok.or.kr/#/>) 의 데이터를 우리 데이터셋과 시기적 일치성 확보를 위해 날짜 기준으로 병합한다.

2-2-2) 발틱 운임 지수 (BDI, The Baltic Exchange)

• 정의

발틱 운임 지수란 석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수이다. BDI는 운송비에 대한 가장 중요한 지표 중 하나이자 전세계 무역과 제조 활동량에 대한 중요한 지표로 해운시장과 세계 경기의 대표적인 선행지표라 할 수 있다. [10]

1985년 1월 4일을 기준(=1,000)으로 놓고 그 움직임을 관찰한다.

BDI는 운반하는 품목이 생산의 기초가 되는 원자재이기 때문에 경기선행지표의 역할을 한다. 원자재 수요가 늘어 교역량이 증가하면 선박의 운송비용이 올라 BDI가 높아지는데, 이는 해운업과 조선업의 주가를 상승시킨다. [11]

BDI 상승	경제 호황
BDI 하락	경기 침체

• 배경

해상운임은 세계 경제 활동에 대한 여러 정보를 포함한다. 금융시장은 실물경제의 변화에 비교적 신속하게 반응하는 시장이므로 세계 경기의 변화를 대표하는 지표인 BDI와 주식시장 간에 높은 관련성이 있다고 판단할 수 있다(Bakshiet al., 2012; Endogan et al., 2013; Alizadeh and Muradoglu, 2014; 이상호 · 김진수, 2014).

한국의 수출의존도는 2017년 68.8%로 높으며 기업의 수익 대부분이 해외사업, 무역, 투자에서 나오고 있기에 국제무역에서 해상운임의 변화는 기업의 주가 및 변동성에 유의미한 영향을 미칠 것으로 예상할 수 있다.

• 데이터 수집 방법 및 출처

금융시장 플랫폼 "investing.com"에서 제공하는 BDI 데이터

<https://www.investing.com/indices/baltic-dry-historical-data>

2-2-3) 곡물식량가격지수 (Cereals Price Index) & 식량가격지수 (Food Price Index)

• 정의

곡물식량가격지수란 옥수수, 콩, 밀을 대상으로, 식량가격지수란 곡물, 식물성 기름, 유제품, 육류, 설탕 등 5개 상품군을 대상으로 유엔식량농업기구(FAO)가 국제거래 가격을 종합해 매달 발표하는 지수이다.

• 배경

한국은 밀, 옥수수, 콩, 쌀 등의 곡물 약 1천 7백만 톤을 매년 해외로부터 수입한다.

이는 국내 곡물 생산량의 4배를 상회하는 양으로, 쌀 이외 곡물 소비의 대부분을 해외로부터의 수입에 의존하고 있다. 따라서 국제 곡물 가격변동은 국내 곡물 가공식품 및 배합사료 산업의 경영 성과뿐만 아니라 가공식품 및 사료 물가변동의 주요인으로 작용하게 된다.

수입곡물의 가격변동은 1, 2차 가공 산업을 통해 특히 축산물, 가공식음료품, 외식업 생산원가를 변동시키며 이는 이들 품목의 물가변동을 촉발시키는 요인으로 작용하게 된다. [12]

• 데이터 수집 방법 및 출처

국제연합식량기구 (Food and Agriculture Organization of the United Nations) <https://www.fao.org/worldfoodsituation/wfs-home/foodpricesindex>

2-2-4) 국제 유가 (WTI, Brent, Dubai)

• 정의

WTI는 미국의 서부 텍사스, 브렌트유는 북해의 브렌트 유전, 두바이유는 중동 아랍에미리트에서 생산된다. 이들은 미주, 유럽, 아시아 일대의 원유 가격을 대표한다. [13]

WTI, Brent, Dubai 원유는 국제 원유시장에서 거래시장이 잘 발달해 가격이 투명하게 정해지는 3대 원유이다. 이들 세 유종은 세계 각 지역 원유 가격의 기준이 된다.

WTI는 국제 유가의 상징적 지표로 활용된다. 한국은 원유의 70~80%를 중동에서 수입해 두바이유에 가장 큰 영향을 받으며, 배로 운송하므로 2~3주의 시차를 두고 석유제품 가격에 반영된다.

- 배경

원유에 대한 의존도가 높은 우리나라의 경우 국제유가의 변화는 기업의 생산비용에서의 변화를 유발하고, 인플레이션을 유발하고 이는 가계 구매력 감소로 이어져 소비를 위축시키는 등 경제활동에 영향을 미칠 수 있다. [14]

실제로 원유 가격 변화는 전 산업에 걸쳐 1.96%의 생산비 파급효과를 유발하는 것으로 나타났다. 특히, 제조업은 2.80%로 생산비가 더 민감하게 반응하는 것으로 나타났다. [15]

- 데이터 수집 방법 및 출처

한국석유공사

<https://www.petronet.co.kr/v3/index.jsp>

2-2-5) 원자재 국내 선물가격 (Soft : 숨, 설탕 / Hard : 구리, 금, 은)

- 정의

원자재란 공업 생산의 원료가 되는 자재이며, 선물거래(futures contract)란 현재시점에서 확정된 매매가격으로 거래대상 상품을 미래의 일정시점에 인수·수도하기로 약정한 매매거래이다. 숨, 설탕, 구리, 금, 은 등을 거래하는 선물거래는 “상품선물거래” 유형에 속한다. [16]

- 배경

2021년 기준 주요 원자재 가격 상승폭을 토대로 원자재 가격 상승이 전 산업 생산비용에 미치는 효과를 가격파급모형을 통해 추정 한 결과, 전 산업에서 2.28%, 제조업에서 3.46%의 가격 상승 효과가 나타날 것으로 추정되었다. [15]

원자재 가격이 상승하는 경우 통상 기업들은 이를 제품 가격에 반영할 것으로 기대된다. 그러나 경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화되거나, 가격경쟁력을 상실하여 구조조정을 겪을 수 있다. 또한 원자재 가격 상승은 장기적으로 전반적인 물가상승을 통해 모든 경제 주체에게 영향을 미칠 수 있다.

원자재 가격이 상승하였을 때 생산자가 받는 영향은 해당 원자재를 중간투입물로 사용하는 기업들의 가격 전가 행태에 따라 결정된다.

기업이 생산원가 상승분을 제품 가격에 전가하는가?	파급효과
O	1. 기업의 채산성 ▼ 2. 다른 경제 주체에 영향 X
X	1. 기업의 채산성 영향 X 2. 경제 전체의 비용 ▲ → 소비자 잉여 ▼ → cost-push 인플레이션 압력

원자재 가격의 상승은 기업의 생산비 증가를 초래하여 수출단가를 상승시킴과 동시에 수출단가 상승에 따른 수출물량 감소로 연결된다. 원자재 가격 10% 상승은 수출단가 0.7% 상승, 수출물량 0.25% 감소를 초래하여 수출금액 면에서는 0.45% 증가 효과를 발생시킨다. [17]

- 데이터 수집 방법 및 출처

KREI 해외곡물시장정보

http://www.krei.re.kr:18181/new_sub17

2-3. 인적 자원 지표

인적 자원 (HR, Human Resource) 관점의 리스크를 신중하게 점검하고, 다양성과 인권을 존중하는 HR 베스트 프랙티스를 갖춘 기업의 경쟁력이 더욱 높아졌다.

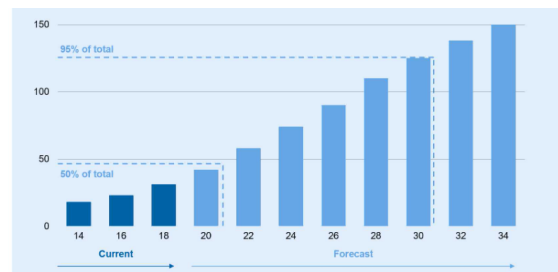
대표적인 ESG 평가지표인 다우존스지속가능지수 (DJSI, Dow Jones Sustainability Indices)의 경우, 노동지표(3), 인권(5), 인적 자원개발(3), 인재 확보와 유지(4), 기업시민정신(2), 보건/안전(9) 등 HR분야가 100점 만점에 36점으로 상당한 비중을 차지하고 있다. HR 관련 정책을 잘 펼치느냐가 기업의 투자매력도를 결정하는 데 1/3 정도의 높은 영향력을 가지고 있는 것이다. [2]

중소기업은 중견기업이나 대기업에 비해 ESG에 대한 인식이 부족하고 대비가 되어있지 않기 때문에, ESG 확산에 따른 인센티브 요인보단 손실리스크가 보다 실제적이고 직접적인 상황이다. [1]

글로벌 자산 운용사들을 중심으로 부정적인 이슈의 기업을 배제하는 소극적 지속가능투자에서, 보다 나은 ESG 성과를 추구하는 적극적 지속가능투자자로 전환해나가고 있다. [1]

[그림 1] 전 세계 ESG 투자 총 운용자산(AUM) 규모 및 추이 예측

(단위: 1조 달러)



자료: Deutsche Bank, Global Sustainable Alliance, zeb research.

저명한 경영학자 마이클 포터의 연구에 따르면 지속 가능성 프로그램(ESG)이 강력한 회사의 사기가 열악한 회사에 비해 55% 더 높았으며, 직원 충성도는 38% 더 높았다. 또한 생산성이 16% 증가하고, 연간 퇴직률도 3% 이상 하락하여 이직으로 인한 교체 비용도 절감할 수 있다고 한다. [18]

따라서, 근로환경, 인권, 급여 수준 등의 인적 자원 지표를 통해 측정이 상대적으로 어려운 중소기업의 전반적인 ESG 실천 정도 및 기업 경쟁력, 지속가능성을 역으로 유추할 수 있을 것으로 기대한다.

• 데이터 수집 방법 및 출처

‘크레딧잡’은 아시아 전역 230만 유저를 보유한 채용 플랫폼 ‘원티드’가 운영하는 기업정보 제공 사이트이다. 국민연금에서 제공받은 데이터를 기반으로 전국 42만개의 기업에 대한 업력/사원수 /올해 입사율/올해 퇴사율/연봉상위율/월별 평균 급여에 대한 정보를 제공하고 있으며, 타 플랫폼보다 정량적 지표의 정확도가 높은 것으로 알려져 있다.

2-3-1) 퇴사율 / 동종업계 내 퇴사율 백분위

• 정의

이직자 수를 전기 말 근로자 수로 나누어 계산하며, 이직자는 업체가 근로자들을 정당한 사유로 면직하는 해고, 근로자 자신의 희망 사직·퇴직, 동일 기업 내의 다른 사업체 배치전환에 따른 전출 등으로 이루어진다. [19]

• 배경

몇몇의 휴폐업 기업 검색 결과 100% 이상의 퇴사율, 즉, 평균적인 기업 근로자 수 만큼 퇴사했다는 것을 볼 수 있었다. 따라서, 휴폐업한 기업의 퇴사율은 액티브 기업보다 높은 곳에서 분포할 것으로 기대한다.

2-3-2) 예상평균연봉 / 동종업계 내 예상평균연봉 백분위

• 정의

평균임금이란 근로기준법 제19조 1항에서 "이법에서 평균임금이라 함은 이를 산정하여야 할 사유가 발생한 날 이전 3월간에 그 근로자에 대하여 지급된 임금의 총액을 그 기간의 총일수로 나눈 금액을 말한다"고 명시하고 있다.

• 배경

선행연구를 통해 기업의 경상이익과 경영자의 현금보상 즉, 연봉 사이에 아래와 같은 관계가 성립함을 확인할 수 있었다. [20]

기업의 경상이익 ▲	경영자의 현금보상 ▲
기업의 경상이익 ▼	경영자의 현금보상 ▼

따라서, 휴폐업 기업의 평균연봉이 액티브 기업의 평균연봉보다 적을 것으로 기대한다.

2-3-3) 총 재직 인원

• 배경

생산성이 높고 노동자 수와 자본집약도 측면에서 규모가 크며 수입비중이 높은 수출기업일수록 보다 많은 수출을 하는 경향이 있다는 선행 연구 결과[1]를 통해 총 재직 인원이 기업의 매출 및 존폐에 유의미한 영향을 줄 것으로 예상할 수 있다.

2-3-4) 최근 1년 인원증가율 / 동종업계 내 인원증가율 백분위

• 정의

국민연금의 데이터를 바탕으로 최근 1년 사이의 재직 인원 증가율을 계산한다.

동종업계 내 인원증가율 백분위 : 인원증가율을 기준으로 상위 ~%로 표현된다.

• 배경

재직 인원 증가는 다양한 부분에서 기업의 성장을 의미한다고 볼 수 있다.

3. 데이터셋 소개 및 전처리

3-1) 재무 데이터

재무 데이터의 경우 부도 예측을 하는데 중요한 지표로 활용할 수 있지만 제공된 데이터의 경우 결측치가 매우 많았다.

재무 데이터의 경우 결측치 처리 기법으로 채우는 것이 민감한 사안이지만, 주요 지표라고 판단하여 최대한 bias가 생기지 않도록 결측치 처리를 진행해보았다.

3-1-1) 결측치 처리

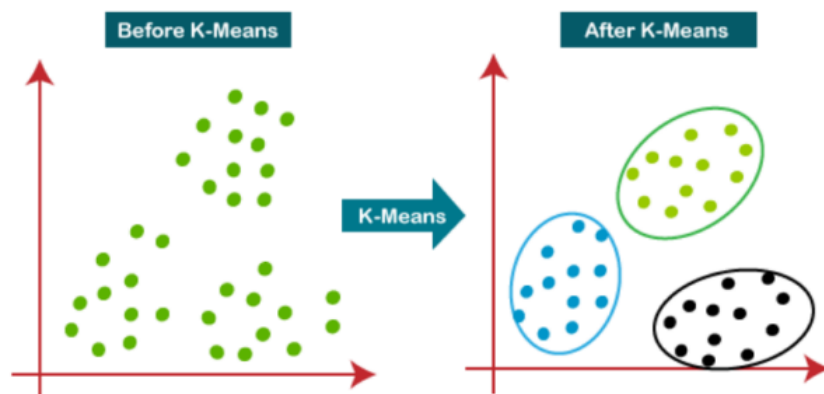
- 1차 결측치 처리 : Knn-Imputer



→ 같은 사업자번호일 경우 재무지표를 모아서 결산년월 이웃평균으로 채웠다.

- 2차 결측치 처리 : 군집별 평균으로 채우기

→ 하나의 결산년월만 갖는 사업자번호가 있을 경우, Knn-Imputer로 채워지지 않는다. 이웃하는 결산년월이 없기 때문이다. 이에 따라 두번째 단계로 널값을 갖지 않는 주요재무지표를 Z 표준화(변수간 유사도 비교를 위해) 진행 후 클러스터링을 진행했다. 그리고 군집별 평균으로 결측치를 처리해주었다.



→ 클러스터 개수는 Silhouette Coefficient, Davies bouldin score를 기준으로 선정하게 되었다.

→ 클러스터로도 안채워지는 법인은 분석대상에서 제외시켰다. (drop)

3-2-2) 하나의 결산년월 선정

- 법인마다 결산년월도 다르고 재무제표가 주어진 개수도 제각각이다.
 - 가장 최근 결산년월을 기준으로 한다.
 - 최근 결산년월을 선택한 이유는 액티브 기업, 휴폐업 기업의 가장 최근 재무정보를 반영할 수 있기 때문이다.

3-2) Baseline

- 휴폐업_여부 컬럼 생성(이진분류)
 - 휴폐업한 경우 : 1로 라벨링
 - 액티브 기업인 경우 : 0으로 라벨링
- (주)여부 컬럼 생성
기업명에 (주)가 있는지 없는지 여부만 판단하는 컬럼으로 CMP_PFIX_NM, CMP_SFIX_NM, CMP_ENM 처리.
(주)가 있으면 1로 없으면 0으로 라벨링
- 상장코드
상장여부를 알려주기 위해서 새롭게 라벨링
상장코드 있으면 1, 없으면 0으로 라벨링
- 본점기업코드
본점기업코드는 value_counts를 통해 개수를 대신 넣어줬다.
숫자가 높을수록 체인점을 많이 보유한 것임을 판단할 수 있다.
- 홈페이지 url
홈페이지 url이 있는지 없는지 여부에 따라 1, 0으로 라벨링한다.
- 직원수
널값은 일단 0으로 채운다.

3-3) 크롤링 데이터 전처리

3-3-1) Altman K - score

부실지표 K - score를 산출하기 위해서 재무제표의 자산총계, 총자본회전율, 이익잉여금(결손금), 자본총계를 사용했다.

수식을 적용하고 np.inf나 -np.inf가 나오는 경우가 있다. 이 경우 기업을 드롭한다.

→ 주식 중 자산총계를 분모로 나누는 경우가 있는데 자산총계가 0일때 이런 현상이 나타났다.

3-3-2) 환율

2018년 1월 1일부터 미국달러 / 한국 환율을 뽑아왔다.

연월별로 평균값을 대입하기 위해 연월 컬럼을 새롭게 구성했다.

→ 재무제표의 결산년월을 기준으로 조인 진행을 위해

3-3-3) 금리지표

CD91일, 콜금리, 국고채3년, 회사채3년 금리지표를 가져와서 하나의 데이터프레임을 형성한다. Date변수를 연월컬럼으로 재구성하고 baseline 데이터프레임의 결산년월을 기준으로 조인 작업을 진행한다.

3-3-4) 국제유가

WTI, Brent, Dubai 세계의 국제유가 컬럼을 baseline에 추가한다.

마찬가지로 연월별 평균값을 넣는다.

3-3-5) 발탁운임지수(BDI)

증가를 기준으로 월별 평균값을 넣는다.

3-3-6) 곡물수입단가

Corn, Soybean, Wheat Price가격을 연월별 평균값으로 추가한다.

3-3-7) Food Price index

기존의 Date와 다른 형태이기 때문에 연월을 만들어주는 개별 함수를 생성했다.

연월별 Food_Price_Index 값을 추가했다.

3-3-8) Sugar, Cotton 추가

연월별 선물가격을 변수로 추가했다.

3-3-9) 금, 구리, 은

퀀틀에서 API KEY를 받고 금, 은, 구리 가격 크롤링한다.

→ 금, 은의 경우 우리 데이터셋에 널값이 없었는데 구리의 경우 가격이 기재되지 않은 일자가 많아 결측치가 많이 생성됐다. 따라서, 금, 은 데이터만 추가했다.

3-3-10) 크레딧잡 크롤링 데이터

크레딧잡에서 평균임금, 평균 임금 상위 %, 총임직원수, 퇴사율, 퇴사율 상위%, 임직원수 성장률, 임직원수 성장률 상위%를 크롤링했다.

데이터가 문자열과 같이 섞여있어 정규표현식을 적용하여 수치만 반환할 수 있도록 전처리했다.

중소기업의 특성 상 불투명한 정보공시로 인해 결측치가 다수 나오게 되어 컬럼별 평균값으로 대체했다.

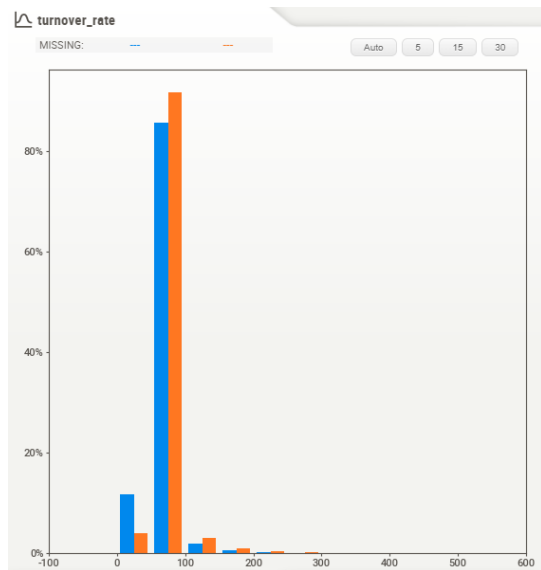
4. EDA 탐색적 자료분석

※ '휴폐업_여부' 변수 기준

◆ : 액티브 기업

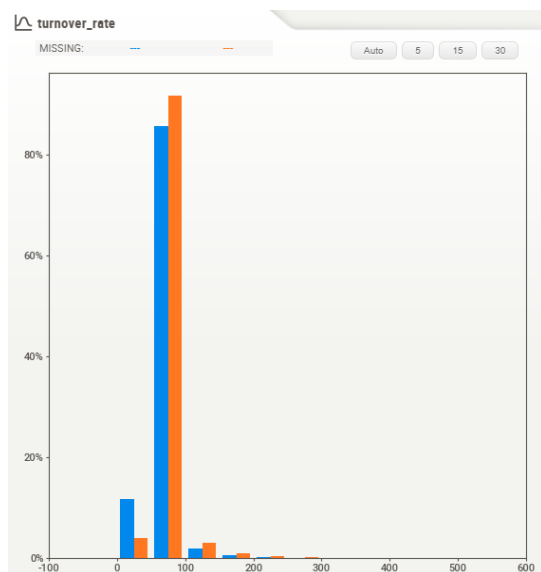
◆ : 휴폐업 기업

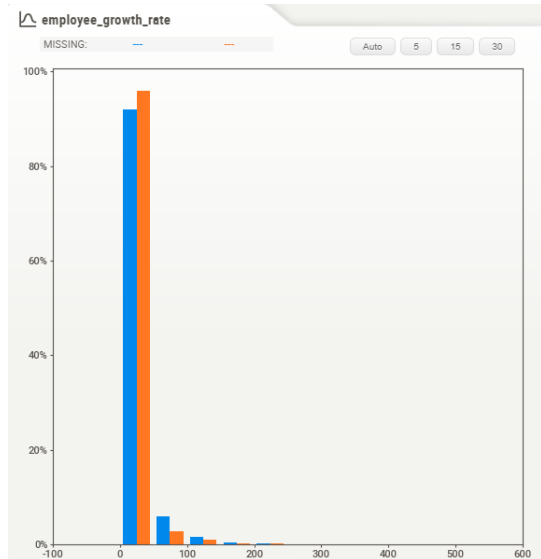
4-1) 퇴사율 분포



해당 그래프는 액티브 기업과 휴폐업 기업 각각의 퇴사율 분포를 보여준다. 실제 휴폐업 기업의 퇴사율이 액티브 기업의 퇴사율보다 높은 곳에 더 많이 분포한 것을 통해, 앞선 [인적 자원 관점]의 선행연구 시 기대했던 바와 동일한 결과가 나온 것을 확인할 수 있다.

4-2) 퇴사율과 인원증가율 사이의 연관성

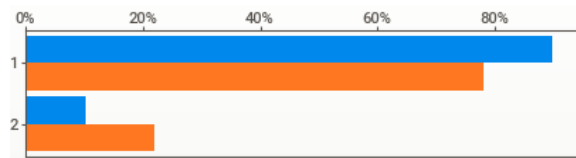




앞선 선행연구를 통해 인원증가와 기업의 성장세가 동일한 방향으로 작용할 것으로 기대했으나, 액티브 기업과 휴폐업 기업 각각을 비교한 시각화를 통해 실제 인원증가율 시각화 자료는 다음과 같았다.

이는 휴폐업 기업의 퇴사율이 더 높기 때문에 고용인 총당 과정에서 인원증가율이 높게 나타난 것으로 해석할 수 있다.

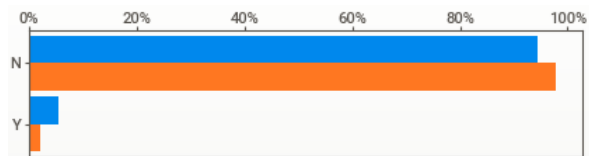
4-3) 본점과 지점의 폐업 차이 (HDOF_BR_GB)



1: 본점 & 2: 지점

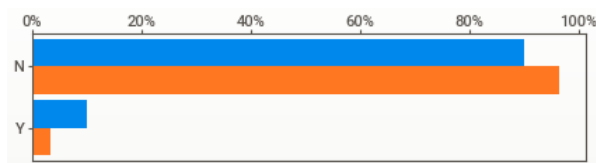
기업이 지점일 경우 휴폐업 가능성이 높다는 것을 알 수 있다.

4-4) 국외투자법인여부 (FR_IVST_CORP_YN)



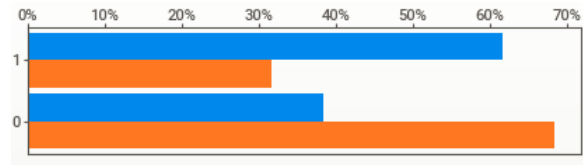
국외투자를 받은 기업이 액티브 기업일 가능성이 높다는 것을 알 수 있다.

4-5) 벤처기업여부 (VENT_YN)



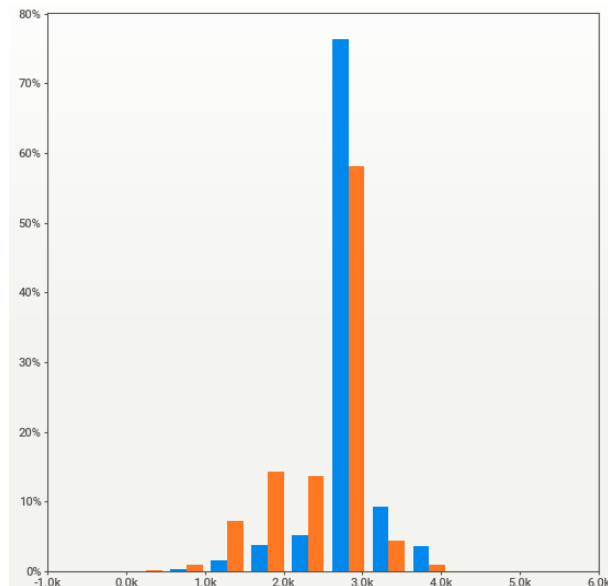
벤처기업이란 타 기업에 비해 기술성이나 성장성이 상대적으로 높은 중소기업을 의미한다. 따라서, 벤처기업인 기업이 액티브 기업일 가능성이 높다는 것을 알 수 있다.

4-6) 홈페이지 url 보유 여부 (HOMEPAGE_URL)



자사 홈페이지를 보유할수록 휴폐업 가능성이 낮다는 것을 알 수 있다.

4-7) 평균연봉 (avg_wage)



3.0k 를 기준으로 연봉이 낮을수록 휴폐업 기업의 분포가 많고, 연봉이 높을수록 액티브 기업이 많이 분포하는 것을 볼 수 있으며, 선행연구에서의 가정이 성립함을 알 수 있다.

5. 분석 모형

5-1) 분석 모형 및 방법론

- 타겟 데이터셋 비율 : 5%

휴폐업 여부	데이터 개수
0	28211
1	6672

→ 심각한 데이터 불균형은 아니지만 5%정도 imbalance가 존재한다.

5-2) 분석 (모델링) 결과

- 타겟 데이터(휴,폐업 여부)의 불균형 존재 → StratifiedKfold이용하여 10번 반복 검증 수행
- 분석 모델은 분류 모형에 대해 성능이 좋은 3가지 boosting model과 1가지 bagging 모델을 사용 추가적으로 인공신경망 모델과의 성능 비교를 진행하고 최종 모델을 선정
- 지표(Metric) 선정 - 이진분류(Binary Classification)을 진행
- 단순, Accuracy(정확도) 지표만으로는 모델의 성능을 정확히 측정하기가 어려우므로 대표적인 metric 측정지표인 recall, precision, f1_score를 이용하여 성능을 비교함

Model	RandomForest	XGBoost	LightGBM	CatBoost	ANN(Artificial Neural Network)
Accuracy	0.966	0.966	0.965	0.964	0.940
Recall	0.974	0.974	0.978	0.974	0.961
Precision	0.961	0.961	0.956	0.958	0.929
AUPRC	0.948	0.948	0.945	0.946	0.911
F1_SCORE	0.966	0.966	0.965	0.964	0.940
AUROC	0.998	0.998	0.965	0.998	0.992

- 최종 모델 선정

Model	Accuracy	Recall	Precision	AUPRC	F1_Score	AUROC
최종 모델	0.966	0.974	0.961	0.948	0.966	0.998

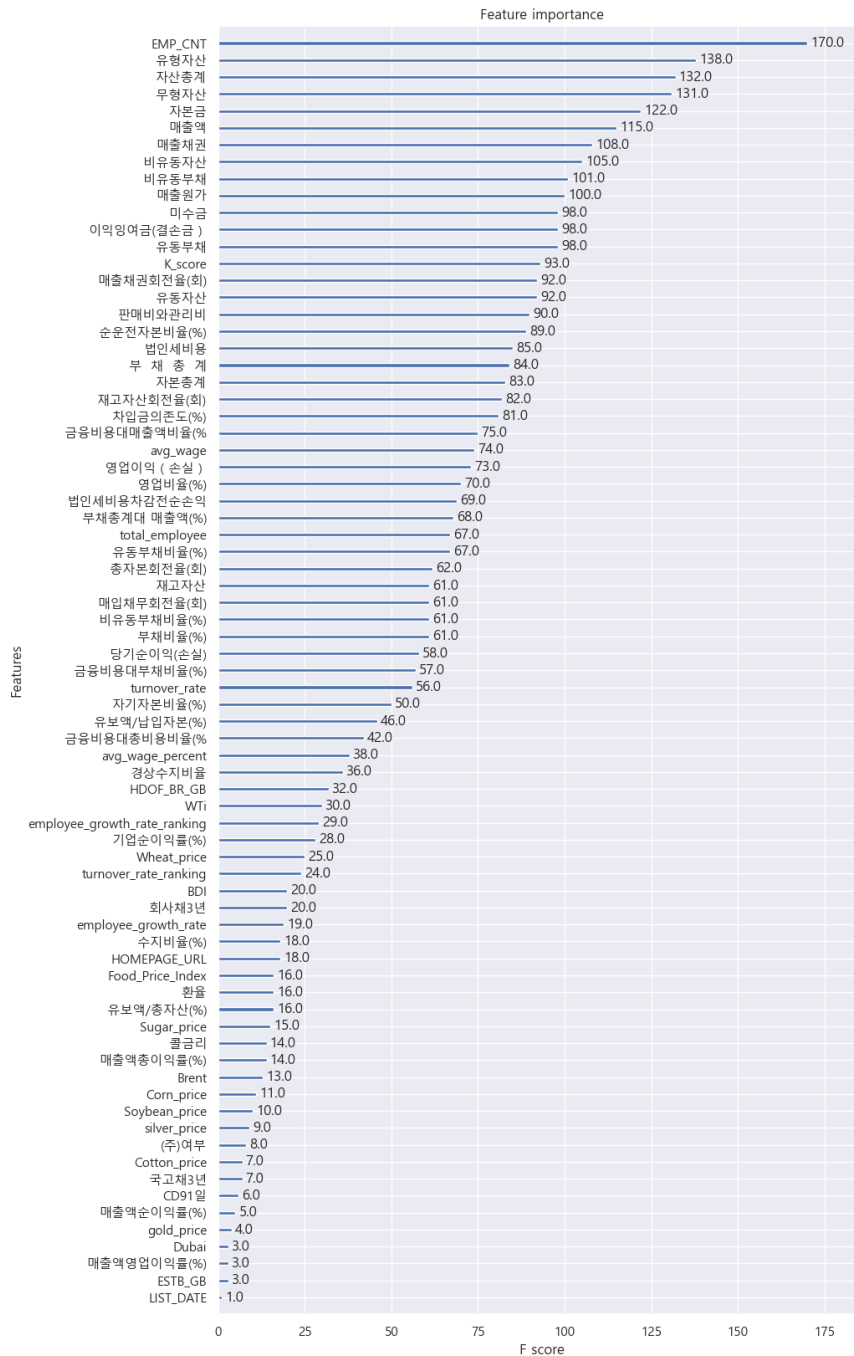
- 최종 모델은 거의 모든 모델이 비슷한 성능을 내었지만 부스팅 계열 모델 중 가장 범용성이 뛰어난 XGBoost를 선정하였다.

5-3) 결과 해석

변수 중요도(feature importance)

- 변수의 중요도만 알 수 있고, 얼마나 어떻게 영향을 미치는지(양수/음수, 크기) 알 수 없다.
- 해석이 불가능, 변수 중요도 측정 기준별로 중요 변수가 달라진다. (SHAP과는 다르다.)
- 주의 : “feature importance가 높다고 중요한 지표다” 라고 말할 수 없다. (inconsistency)
- 단지 어떤 변수가 모델의 예측에 어떻게 영향을 끼치는지 만을 보여준다.

최종 모델 변수 중요도 Plot



- 결과표를 보면 대부분 재무 데이터가 높은 수치로 반영되었음을 알 수 있음

5-4) SHAP(Shapley Additive Exaplations) Value를 이용한 모델 해석

- 목적 : 예측에 대한 각 형상의 기여도를 계산하여 인스턴스(데이터) X의 예측을 설명하는 것
- 장점
 1. SHAP Value의 합리적인 성질 덕분에 개별 예측값에 대한 각 변수의 중요도를 상식적으로 제공한다.
 2. Model Agnostic한 방법이다. 즉 SHAP은 예측 모형 클래스의 관계없이 적용할 수 있어 유연하다.
 3. 특정 모형 클래스에 대해서는 계산이 빠르다. 예측 모형의 종류(선형, Tree)에 따라 빠른 계산 속도를 지원한다.
 4. 개별 예측값(Local)에 대한 변수 중요도 뿐만 아니라 예측에 대한 전반적인 변수 중요도를 계산할 수 있다.

- SHapley Additive exPlanation의 수학적 정의는 아래와 같다.

$$\phi_i = \frac{1}{|N|!} \sum_R [v(S_i^R \cup \{i\}) - v(S_i^R)]$$

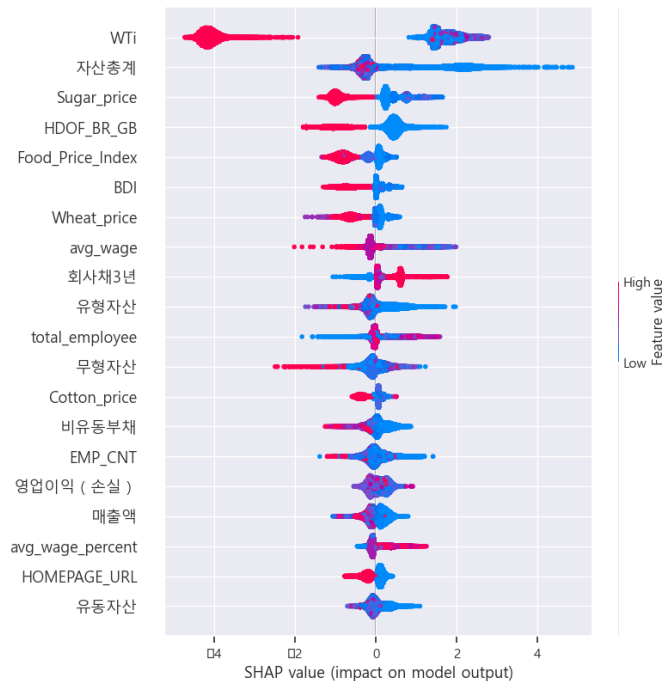
- waterfall plot, force plot - 개별화된 변수 중요도

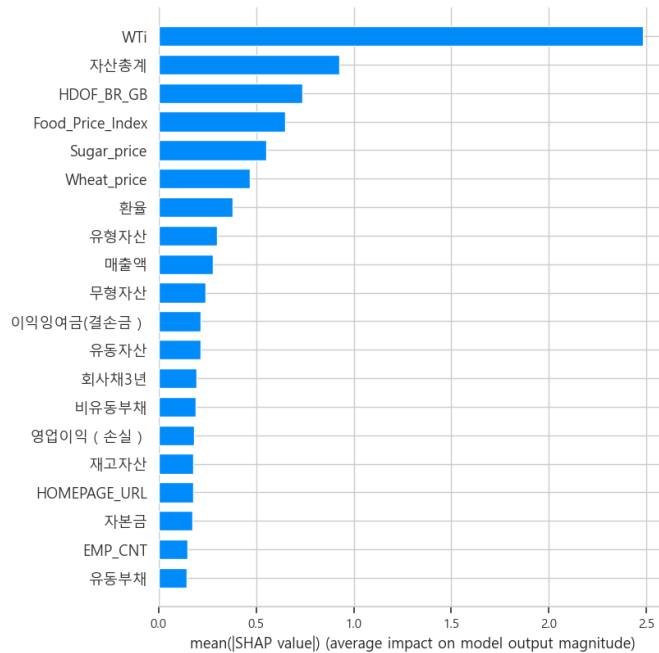
아래 예시는 특정 기업이 예측될 시의 중요도



- Summary_plot 혹은 bar_plot을 통한 전체적인 변수 중요도 시각화 (다중 클래스 결과에 대해 누적 막대 생성)

특정 기업이 아닌 전체적인 기업들의 SHap value 특징을 보여준다.





- 결과를 보면 유가 중 하나인 WTI가 가장 높게 나온 것을 확인 할 수 있고, 해당 모델은 1 즉 휴폐업을 예측하는 모델이므로 유가가 높을수록 오히려 마이너스(-) 적인 영향을 줄을 확인할 수 있다.
- 해당 변수 외에도 외부 파생 비재무 데이터들의 다수가 상위권에 ranking됨을 결과로 확인할 수 있다.
- 따라서 이러한 SHAP Value를 통해 우리의 모델에서 비재무 데이터의 중요성을 다시금 확인 할 수 있었고, 더 나아가 대안 신용평가 모델에서의 중요성 역시 마찬가지로 점검했다.

[23]

6. Reference

- [1] 남기정, 이동명 and 진로. (2019). 비재무정보를 이용한 창업기업의 부실요인에 관한 실증연구. 벤처창업연구, 14(1), 139-149.
- [2] 조옥래(1998), 『중소기업 도산예측과정에서 비재무제표 항목의 유용성검증과 기업 분석시스템 개발에 관한 연구』, 경희대학교, 석사학위논문.
- [3] 또 다른 발자국, 2017년 02월 08일, <https://nicksstory.tistory.com/182>
- [4] 정대용, 유봉호, "기업가 지향성과 시장지향성이 조직유효성과 사업성과에 미치는 영향", 벤처경영연구, 제10권, 제2호, pp.69-93, 2007.
- [5] A.Kohli and B.Jaworski, "Market Orientation: The Construct, Research Proposition, and Managerial Implication," Journal of Marketing, Vol.54, No.2, pp.1-18, 1990
- [6] J.C.Narver and S.F.Slater, "The Effect of a Market Orientation on Business Profitability," Journal of Marketing, Vol.54, No.4, pp.20-35, 1990
- [7] Chung, H. (2012), "Export Market Orientation, Managerial Ties, and Performance," International Marketing Review, 29(4), 403-423.
- [8] 윤덕홍 · 김효상. (2017). 환율변화가 한국기업에 미치는 영향분석과 정책적 시사점: 기업데이터 분석을 중심으로. 대외경제정책연구원.
- [9] 한민수 and 편주현. (2021). 환율 변화가 기업 규모별 우리나라 수출기업에 미치는 영향. 시장경제연구, 50(1), 99-127.
- [10] 최기홍, 김동윤. (2019). 발틱인덱스가 한국 주가 변동성에 미치는 영향. 한국항만경제학회지, 35(2), 61-75.
- [11] 서지영. 서현원/KDI 경제정보센터, 2010.08.02, https://eiec.kdi.re.kr/material/clickView.do?click_yymm=201512&cid=1267

- [12] 김종진, 박성진 and 박지원. (2020). 수입곡물 가격변동이 국내 소비자물가에 미치는 영향 분석. 농촌경제, 43(2), 29-58.
- [13] 임현우, "WTI·브렌트유·두바이유...'3대 原油'는 고향이 모두 달라요", 한국경제, 2021.03.01, <https://sgsg.hankyung.com/article/2021022654681>
- [14] 김상배. (2021). 유가 변화가 우리나라 주식시장 유동성에 미치는 영향. 에너지경제연구, 20(1), 79-103.
- [15] 김바우, 김정현 and 강성우, 최근 원자재 가격 상승의 배경과 국내 제조업에 미치는 영향, KIET 산업경제, 2021.11.09
- [16] 금융위원회 증권제도과, 1999-07-06, [해설] 선물거래의 기본구조, <https://www.fsc.go.kr/po010101/63344?srchCtgr=&curPage=1160&srchKey=&srchText=&srchBeginDt=&srchEndDt=>
- [17] 강내영 and 강성은. (2021). 국제원자재 가격의 변동요인 및 우리 수출에의 영향 분석. 한국무역협회, 17.
- [18] 천성현. "왜 HR이 ESG에 주목해야 하는가", HR insight, 2020.12.31, https://www.hrinsight.co.kr/view/view.asp?in_cate=112&bi_pidx=31723
- [19] 통계청, 통계용어및지표, https://kostat.go.kr/understand/info/info_lge/1/detail_lang.action?bmode=detail_lang&pageNo=7&keyWord=0&cd=SL4175&sTt=
- [20] 지성권 and 김태수. (2001). 경영자 보상과 회계이익특성의 관계에 관한 실증연구. 회계학연구, 26(1), 25-54.
- [21] 한민수, 편주현. (2021). 환율 변화가 기업 규모별 우리나라 수출기업에 미치는 영향. 시장경제연구, 50(1), 99-127. Minsoo Han, Ju Hyun Pyun. (2021). Effects of exchange rate changes on Korean export firms by firm size. SOGANG ECONOMIC PAPERS, 50(1), 99-127.
- [22] 오낙교, 윤성수 and 박원구. (2014). 알트만 Z-스코어를 이용한 신생 중소기업의 지속가능성 분석: 신재생에너지산업을 중심으로. 기술혁신연구, 22(2), 185-220.
- [23] 분석가 꿈냥이, 2022.08.25, <https://zephyrus1111.tistory.com/271>
- [그림1] KOSI 중소벤처기업연구원. 중소기업 포커스 제 21-14호 2021.08.09 : ESG 확산이 중소기업에 미치는 영향 및 지원 방향. 나수미