머신러닝 기반 KOSDAQ 시장의 관리종목 지정 예측 연구

저자: 윤양현, 김태경[[1]](#footnote-1), 김수영, 박용균

날짜: 2021-10-08

초록

생략(추후 적을 예정)

키워드: KOSDAQ, 관리종목, 로지스틱 회귀분석, 랜덤 포레스트, SVM, 앙상블

1. 서론

관리종목은 주권상장법인이 상장 후 영업실적의 지속적 악화, 상장회사가 갖추어야 할 최소한의 유동성 부족, 기업지배구조 미구축 등의 사유로 상장폐지기준에 해당된 종목 가운데 특별히 지정된 종목들을 의미한다. KRX(Korea Exchange, 한국거래소)에서는 상장폐지기준에 해당하는 상장기업들을 관리종목으로 지정하고, 기업이 경영을 지속 혹은 존속하기 힘든 기업들을 관리종목 지정제도를 통하여 투자자들에게 공시한다. 이러한 정보의 전달로 투자자는 상장폐지 위험이 있는 기업들을 조기에 파악함으로써 투자판단에 있어 주의를 환기하고, 당해 법인에게는 일정기간 경과기간을 부여하여 관리종목 지정사유를 해소함으로써 기업의 정상화를 촉진하는 효과를 지닌다*(김민철, 관리종목지정사유별 주가수익률의 변화, 2004)*.

세포치료제 연구사업 및 줄기세포 보관사업 등을 주 사업으로 영위하는 차바이오텍 기업은 KOSDAQ(Korea Securities Dealers Automated Quotation)에 상장되어 있다. 차바이오텍은 관리종목 지정 전에 시가총액 2조 원의 가치를 지니고 있었으나, 2018년 3월 22일 관리종목 지정이 되고 난 후 주가가 하락하여 3일 만에 관리종목 지정 전 주가 대비 절반이 된 사례가 존재하였다. 이 사례의 경우 시가총액 1조 원의 가치가 단기간 내에 사라지게 되어 기업과 투자자의 입장에서 많은 손실을 가져오게 되었다*(신동인&곽기영, KOSDAQ 시장의 관리종목 지정 탐지 모형 개발, 2018)*.

관리종목으로 지정이 되면 해당 기업은 신용거래 대상에서 제외되고 증권시장에서 일정기간 동안 매매거래가 정지된다. 또한 대용증권으로도 활용할 수 없어 해당 기업의 레버리지가 줄어들게 되는 불이익을 받게 된다. 그 외에도 매매방법에 있어 30분 간격으로 거래가 체결되는 별도의 제한을 받게 되어 관리종목 지정 자체가 기업의 입장에서 상당히 불리한 조치임을 알 수 있다*(박종성, 관리종목 지정을 회피하기 위한 코스닥 기업의 이익조정, 2012).* 과거의 선행연구에서 관리종목 공시 이전부터 주가가 하락함을 알 수 있었으며*(김태혁&엄철준, 관리대상종목의 수익률과 위험 속성에 관한 연구,1997),* 관리종목 지정 자체가 시장에서 부정적인 정보로 인식되어 지정 후에 비기대이익에 대한 누적초과수익률이 감소하는 것으로 나타났다. 이는 관리종목으로 지정 받았다는 사실 자체가 나쁜 정보로 작용하여 상장폐지를 유발하는 결과가 초래될 수 있다*(손성규&오명전, 관리종목 기업의 회계정보 효과, 2008).*

앞선 사례와 관리종목 지정으로 인한 기업과 투자자에게 부정적인 효과를 내비친 선행연구를 통하여 본 연구자에게 다음과 같은 연구 질문을 제기하게끔 하는 계기가 되었다. 기업의 과거 재무비율을 기반으로 하여 머신러닝 기술을 통해 해당 기업의 관리종목 지정 여부를 예측할 수 있을까? 만약 예측이 가능하다면, 관리종목을 가장 잘 예측할 수 있는 머신러닝 알고리즘은 무엇일까?

국내, 해외 기업의 부실화로 인한 부도예측에 관한 연구가 대부분이지만, 상장기업 내 부실화 가능성이 높은 기업에 대한 경고를 하는 관리종목 지정에 관한 연구는 상대적으로 부족하다. 즉, 한국에서는 관리종목 지정 제도가 기업의 부실화를 예방하기 위한 중요한 제도적 역할을 함에도 불구하고 관련된 선행연구가 부족하다는 것이다. 이는 관리종목 지정 예측이 학문적으로 희소성이 있음을 시사한다. 또한 기존 선행연구와는 다르게 2011년 이후부터 국제회계기준이 적용된 K-IFRS(Korea-International Financial Reporting Standards, 한국채택국제회계기준)을 통해 데이터를 수집하였으므로 더 정확한 관리종목 지정 예측 모형이 될 수 있음을 시사한다.

본 연구는 이러한 논의를 바탕으로 코스닥 기업의 재무 데이터를 활용하여 관리종목 지정 예측 모델을 설계하고자 한다. 다양한 머신러닝 기술을 이용하여 관리종목과 비관리종목을 분류하고 이를 예측할 수 있는 분류 모형을 제안하고, 이를 통하여 기업과 투자자들은 기업 부실화에 가까운 종목을 조기에 인지할 수 있도록 기여하고자 한다.

본 논문의 구성은 1 서론에 이어 2 에서는 연구 배경과 선행연구를, 3 에서는 연구방법을, 4 에서는 실증 분석 결과를, 5 에서는 결론을 내린다.

2. 연구 배경/선행연구

2-1 관리종목 지정

상장폐지란 상장된 유가증권이 한국증권거래소가 정한 일정한 기준에 해당되었을 때 증권거래소에서 매매될 수 있는 자격이 박탈당하는 조치를 말한다. 당해 법인이 상장폐지기준에 해당되었을 경우에 증권거래소는 상장된 기업을 폐지할 수 있게 된다. 그러나 해당기업이 일시적으로 상장폐지기준에 해당하는 경우가 존재하기에 즉각 상장폐지조치를 취하면 회생할 기회도 없이 증권시장에서 퇴출 당하는 결과를 초래할 수 있다. 이는 해당기업의 투자자들에게 투자액 회수의 기회도 상실할 수 있다는 문제점 또한 가지게 된다. KRX의 관리종목 지정은 바로 이러한 문제점들을 해결하기 위해 내려지는 조치이다*(표영인&김일, 관리종목지정 시점에 발생하는 산업내 정보전도효과, 2002).* KRX는 기업이 상장폐지기준에 해당하면 해당 기업의 의사와는 관계없이 강제적으로 상장폐지 시킬 수 있다. 그러나 상장폐지기준에 해당하는 기업들을 바로 상장폐지 하기 보다, 기업에게 회생기회를 부여함과 동시에 투자자에게 투자손실을 예고함으로써 주의를 환기시키기 위해 관리종목을 지정한다. KRX는 기업이 공시서류 미제출, 자본잠식, 매출부진, 영업손실, 시장 내 유동성 부족 등과 같이 부실화 가능성이 높은 일정요건에 해당하는 경우 관리종목으로 지정하게 된다. 한국의 유가증권시장(KOSPI Market: Korea Composite Stock Price Index)과 코스닥시장(KOSDAQ)은 각각의 관리종목 제도가 존재한다. 최근에는 기술성장기업이나 이익미실현기업과 같이 추후 성장이 기대되는 기업이지만 현재 재무구조가 좋지 않은 기업들에 대해서는 관리종목 지정에 있어 예외를 두고 있다. KRX에서 정한 코스닥 시장의 관리종목 지정 요건은 다음 <Table 1>과 같다.

**<Table 1> 코스닥 시장 관리종목 지정 요건(2021년 7월 21일 기준)**

|  |  |
| --- | --- |
| 구분 | 관리종목 |
| 매출액 | 최근년 30억원 미만(지주회사는 연결기준)  -기술성장기업, 이익미실현기업은 각각 상장후 5년간 미적용 |
| 1) 법인세비용차감전계속사업손실 | 자기자본50%이상(&10억원이상)의 법인세비용차감전계속사업손실이 최근3년간 2회 이상(&최근연도계속사업손실) -기술성장기업 상장후 3년간 미적용, 이익미실현 기업 상장후 5년 미적용 |
| 장기영업손실 | 4)최근 4사업연도 영업손실(지주회사는 연결기준)  -기술성장기업(기술성장기업부)은 미적용 |
| 2) 자본잠식/자기자본 | 사업연도(반기)말 자본잠식률1) 50%이상  사업연도(반기)말 자기자본 10억원미만  반기보고서 제출기한 경과후 10일내 반기검토(감사)보고서 미제출 or 검토(감사)의견 부적정·의견거절·범위제한한정 자본잠식율 = (자본금 - 자기자본) / 자본금 X 100 |
| 3) 감사의견 | 반기보고서 부적정, 의견거절, 감사범위 제한으로 인한 한정 |
| 시가총액 | 보통주시가총액 40억원미만 30일간 지속 |
| 거래량 | 분기 월평균거래량이 유동주식수의 1%에 미달 -월간거래량 1만주, 소액주주 300인이상이 20%이상 지분 보유 등은 적용배제 |
| 지분분산 | 5)소액주주200인미만or소액주주지분20%미만  -300인이상의 소액주주가 유동주식수의 10%이상으로서 100만주이상을 소유하는 경우는 적용배제 |
| 불성실공시 | - |
| 공시서류 | 분기, 반기, 사업보고서 법정제출기한 내 미제출 |
| 사외이사등 | 사외이사/감사위원회 요건 미충족 |
| 회생절차/파산신청 | 회생절차 개시 신청  파산신청 |
| 기타(즉시퇴출) | 기타 상장폐지 사유 발생 |

**주 1)**연결재무제표 작성대상법인의 경우, 연결재무제표상 법인세비용차감전계속사업손실 및 자기자본 기준

**주 2)**연결재무제표 작성대상법인의 경우, 연결재무제표를 기준으로 하되 자기자본에서 비지배지분을 제외

**주 3)**연결재무제표 작성대상법인의 경우, 연결재무제표에 대한 감사의견을 포함

**주 4)**기술성이 있고 연구개발 투자가 많은 연구개발기업에 대해 장기영업손실로 인한 관리종목 지정을 한시적으로 면제

**주 5)**자진상장폐지를 위한 공개매수시 분산기준 미달로 인한 관리종목 지정 등 유예

관리종목 지정과 같은 시장규제 제도는 해외의 각 국가마다 차이가 존재한다. 증권시장에서 가장 대표적인 미국의 나스닥(NASDAQ: National Association of Securities Dealers Automated Quotations)은 별도의 상장폐지 요건이 존재하지 않고 상장된 기업이 지속적으로 상장을 유지할 수 있도록 상장유지 조건을 운영하고 있다. 나스닥의 상장유지조건은 순 유형자산, 시가총액, 유동주식수, 유동주식의 시가총액, 주주수 등의 기준을 포함하고 있다. 미국 뉴욕증권시장(NYSE: the New York Stock Exchange)은 거래소 자체가 상장폐지권한을 지니고 있지 않지만 뉴욕거래소에서 상장폐지요구서를 제출하면 미국 증권거래위원회(SEC: Securities and Exchange Commission)에서 상장폐지 심의를 거쳐 상장폐지를 결정한다. 해당 기업은 이러한 상장폐지 심의에 대해 이의를 제기할 수 있으며, 상장폐지의 결정은 법원에서 결정한다. NYSE의 상장폐지 조건은 일정 주가 30일 거래일 지속, 시가총액의 일정 조건 미충족, 상장유지 비용 미납부, 재무상황 악화, 거래소와의 계약 위반 등의 기준을 포함하고 있다. 일본의 자스닥(JASDAQ)은 주주수, 상장시가총액, 채무초과, 치부율 기타 구분(사업활동의 정지, 부당한 합병 등)으로 상장폐지기준에 해당할 경우 상장폐지를 진행하고 있다*(김승열, 코스닥시장의 상장폐지실질심사에 관한 연구, 2010)*. 이 외에도 중국의 상해증권교역소(SSE: Shanghai Stock Exchange), 유럽의 런던증권거래소(LSE: London Stock Exchange), 홍콩증권거래소(Hong Kong Stock Exchange) 등도 각 나라의 증권소 별 상장폐지기준에 대한 차이는 있지만, 큰 틀은 NYSE와 비슷하게 상장폐지기준을 포함하고 있다*(Martinez and Serve, Reasons for delisting and consequences: a literature review ad research agenda,2016).*

2-2 통계적 모형을 활용한 기업부도예측 선행연구

관리종목 지정 제도는 한국에만 존재하는 특수한 제도이기에, 예측 모델에 관한 연구는 부족하지만 이와 유사하게 기업의 부실 예측 모형에 관한 선행연구는 많으며, 이에 관한 연구들은 현재도 활발히 연구 중에 있다. 부실 예측 모형의 초기 연구는 Beaver*(Financial Ratios as Predictors of Failure, 1966)*의 단일변량 분석으로 재무 변수를 통하여 기업 부도를 예측하였다. 하지만 현실에서는 복합적인 요인으로 인하여 기업이 파산하기 때문에 단일변수로는 기업의 부실화예측을 설명하기에는 명백히 한계점이 존재하였다. 그 후 Altman*(Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, 1968)*이 단일변량 분석의 한계점을 보완하기 위하여 다변량 판별분석을 적용한 Z-Score 모형을 개발하였다. Z-Score 모형은 기업의 도산 1년 전 재무 데이터를 활용하여 기업의 도산 여부를 설명하였다. Beaver의 단일변량 분석에 비해 기업의 도산 여부를 설명하기에 많은 변수들을 고려하였지만, 한계점으로는 파산기업과 정상기업 모두 독립변수의 분산과 공분산 행렬이 동일하다고 가정한 후 연구를 수행하였다. 실제로는 독립변수의 분산과 공분산이 일정하지 않으며, 데이터의 분포도 정규성을 띄지 않는 경우가 존재하였기에 이 연구에서도 한계점이 존재하였다. Ohlson*(Financial Ratios and the Probabilistic Prediction of Bankruptcy ,1980)*은 1970~1976년에 파산한 105개 기업과 건전 기업인 2058개를 표본으로 하여 도산 1년 전의 재무 데이터를 이용하여 로지스틱 회귀분석을 통한 파산예측모형을 제시하였다. Alaka et al*(Methodological approach of construction business failure prediction studies: a review, 2016)*은 부실예측모형에 관련된 70개의 논문을 분석하였다. 검토를 통하여 기존 연구 논문들의 한계를 제시하며, 새로운 패러다임과 이에 맞는 연구 설계 프레임 워크를 제시하였다.

2-3 기계학습을 활용한 기업부도예측 선행연구

2-4 관리종목 관련 선행연구

관리종목에 관한 선행 연구는 다음과 같다. 신동인&곽기영*(KOSDAQ 시장의 관리종목 지정탐지 모형 개발, 2018)*은 2008년부터 2018년까지의 전체 코스닥 기업 재무 데이터를 수집하고 관리종목과 무작위로 선택한 비관리종목을 쌍대표본으로 구성한 후 관리종목 지정 예측에 대한 연구를 하였다. 로지스틱 회귀분석과 의사결정나무 분석을 이용하여 관리종목 지정을 예측하였으며, 전체 평균 예측 정확도는 검증용 데이터셋에 대해 86%, 87%의 예측 정확도를 보여주었다. 손성규&오명전*(관리종목 기업의 회계정보 효과, 2008)*은 관리종목 지정 전후를 비교 분석하여 관리종목 지정이 회계정보에 어떠한 역할을 수행하는지 검토하였다. 연구 결과 관리종목 지정 자체가 시장에서는 부정적인 정보로 인식되어 회계정보효과를 감소시킬 가능성이 있으며, 반대로 관리종목에서 해제된 후에 비기대이익에 대한 누적초과수익률이 증가하는 방향성이 나타났음을 확인하였다. 이를 통하여 관리종목으로 지정 자체가 시장에서 부정적인 정보로 받아들여지고 있음을 제시하고 있다. 김태혁&엄철준*(관리대상종목의 수익률과 위험 속성에 관한연구, 1997)*은 1984년 이후로 관리대상종목으로 지정된 68개 기업을 대상으로 하여 각 종목의 수익률과 위험 특성에 대하여 연구하였다. 연구 결과 관리종목 지정 이후 대부분 주가는 음(-, negative)의 수익률을 기록하며, 특히 관리종목 공시일에 가까울수록 주가의 변동성이 크며, 공시 이전 관리대상종목을 사전적으로 발견할 수 있음을 확인하였다. 김민철*(관리종목 지정사유별 주가수익률의 변화****,*** *2004)*은 상장기업들이 관리종목으로 지정되는 경우 관리종목지정사유와 주가수익률의 관계를 조사하였다. 2000~2003년 관리종목을 대상으로 공시 30일 이전부터 공시 40일 이후까지 주가수익률을 관찰하였다. 연구 당시 관리종목지정사유를 구분하여 회사정리절차관련, 자본잠식, 부도발생, 화의관련, 감사의견한정, 감사의견부적정 6개 요인을 재무적 요인으로 구분하였고, 영업활동정지, 주식분포미달, 거래량요건미달, 공시의무위반, 회계처리위반 5개 요인을 비재무적 요인으로 구분하여 조사하였다. 연구 결과 재무적 요인으로 관리종목지정 10일전부터 유의적인 차이가 발생하였음을 확인하였으며, 비재무적요인에 의한 관리종목의 경우에는 시장평균수익률과 큰 차이가 없음을 확인하였다.따라서 도산의 정의로 관리종목을 선정하는 경우 비재무적 요인에 의하여 지정된 기업은 배제되어야 함을 제시하였다. 박종성*(관리종목 지정을 회피하기 위한 코스닥 기업의 이익조정, 2012)*은 2011년부터 K-IFRS가 도입됨에 따라 영업이익에 산출기준이 없어지고 항목분류 조정을 통한 이익조정이 가능해졌음을 확인하였다. 또한 관리종목 지정 위기에 처한 기업들이 관리종목 지정을 회피하기 위해 영업이익을 조정하였는지 분석하였다. 분석 결과에 따르면 기타 손익의 분류를 달리하면 영업 손실을 영업이익으로 전환하여 관리종목 지정을 회피할 수 있었으며, 이는 관리종목 지정 위기에 처한 기업들이 영업이익에 도움이 되는 항목은 영업손익에 포함시키되 영업손실을 초래하는 항목은 영업손익에서 제외시키고 있음을 확인하였다. 이러한 경우 예외적으로 관리종목에 해당하는 기업임에도 불구하고 회계장부 조작을 통하여 관리종목 지정에서 벗어날 수 있음을 보여주었으며, 한국에 맞는 영업이익에 대한 산출기준을 별도로 정할 것을 제시하였다. 김일*(관리종목을 지정된 기업의 재무적 특성에 관한 연구, 2005)*은 관리종목으로 지정된 기업과 동종산업 내 비관리종목을 비교하여 재무적으로 다른 특성이 있는지 조사하였다. 연구 결과 관리종목 지정 기업은 수익성, 안정성, 활동성에서 동종업계 비관리종목보다 관리종목 지정 전 5년동안 유의한 차이를 보여주지만 성장성을 통하여 두 비교집단 간 유의한 차이를 확인할 수 없었다. 또한 유동성은 관리종목 지정 5년 전부터 2년 전까지는 유의한 차이가 없고, 1년 전부터 큰 차이를 보여줌을 확인하였다. 따라서 수익성, 안정성, 활동성, 유동성을 나타내는 과거 재무적 특성 요인이 관리종목 지정에 영향을 미친다는 결과를 제시하였다.

기존의 연구*(신동인&곽기영, KOSDAQ 시장의 관리종목 지정 탐지 모형 개발, 2018)*에서는 2008년부터 2018년의 데이터를 모아서 분석하였다. 하지만 박종성*(관리종목 지정을 회피하기 위한 코스닥 기업의 이익조정, 2012)*의 연구에 따르면, 회계장부가 K-IFRS로 변화되고 난 후 관리종목에 해당하는 기업임에도 불구하고 회계장부 조작을 통하여 관리종목 지정에서 벗어날 수 있음을 시사하였다. 이는 2011년 전후로는 회계장부에 있어 차이점이 분명히 존재하기에 2011년 이후로 작성된 회계장부 기반 데이터 마이닝을 해야 더 정확한 예측 모형을 만들어 낼 수 있다. 또한, 2018년 말 회계 외부감사법이 개정된 이후로 관리종목 지정에 있어 이전보다 더 엄격한 기준이 적용되었다. 본 연구는 K-IFRS 개정과 최근 외부감사 개정을 반영한 2011년~2020년의 관리종목 표본 기업들을 반영하였기에 기존 선행연구보다 더 정확한 결과가 나올 것으로 기대된다. 선행연구와 이러한 점을 반영하여 전통적인 통계 기반 분석인 로지스틱 회귀분석과 의사결정나무 분석, 기계학습을 기반으로 한 서포트 벡터 머신(Support Vector Machine), 앙상블-보팅(Voting), 앙상블-랜덤포레스트(Bagging), 앙상블-부스팅(Boosting) 분석을 하여 이를 비교 분석하고자 한다.

3. 분석방법론

3-1 로지스틱 회귀분석

종속변수가 연속형 변수가 아닌 범주형 변수인 경우 일반적인 선형회귀식으로는 설명하기 어렵다. 이러한 문제는 로지스틱 회귀식을 통해 식을 도출하여 해결할 수 있다. 종속변수가 2개인 이진 변수의 경우 이항형 로지스틱 회귀분석을 통하여 식을 도출한다. 관리종목으로 지정될 경우 종속변수는 1, 관리종목으로 지정되지 않으면 종속변수는 0인 두 집단에 속할 확률을 예측하기 위한 비율을 측정한다.

위 식에서 는 관리종목으로 지정될 확률이며, 는 관리종목에 해당할 확률에 영향을 주는 재무변수들을 의미한다. 이는 승산비(odds ratio)로도 표현하며, 관리종목에 해당될 확률 는 다음과 같이 표현할 수 있다.

관리종목 지정확률 가 일정수준을 넘으면 관리종목 기업으로, 그렇지 않으면 비관리종목 기업으로 판단한다. 일반적으로 정하는 일정수준인 의 값을 0.5를 기준으로 판단하지만, 이에 대해서는 임의로 정할 수 있다.

3-2 의사결정나무

의사결정나무는 의사결정 규칙을 나무 구조로 나타내어 전체 자료를 몇 개의 소집단으로 분류하거나 예측하는 분석 방법이다. 규칙 노드와 리프 노드로 이루어져 있으며, 규칙 노드는 어떤 속성에 대한 조건문을 통해 데이터를 분리한다. 리프 노드는 규칙 노드로 분류된 구조에서 최종적으로 결정된 결정값을 의미한다. 의사결정나무는 정보의 균일도에 따라 규칙을 생성하기 때문에 데이터를 따로 가공할 필요가 없으며, 의사결정이 이루어지는 시점과 성과 파악을 시각화 할 수 있기 때문에 결과를 해석하고 이해하기 쉽다는 장점이 있다. 하지만, 나무의 깊이에 따라 학습 데이터에 대해 과소적합(underfitting) 혹은 과적합(Overfitting)이 이루어지는 위험에 노출되어 있기 때문에 편향된 결과를 초래할 수 있다.

3-3 서포트 벡터 머신

서포트 벡터 머신은 데이터를 기반으로 하여 분류 또는 회귀를 하는 초평면(hyperplane)을 찾아내어, 이를 분류하는 기계학습 알고리즘 중 하나이다. 초평면인 결정경계를 최대화하는 마진(margin)을 찾아내어 구분을 좀 더 확실히 하는 데 목적이 있으며, 특히 분류 문제에 있어서 좋은 성능을 이루어내고, 오래 걸리지만 과적합의 오류가 적은 편이다.

3-4 앙상블:보팅

보팅(Voting) 방식은 데이터셋은 그대로 놔두고 여러 분류기를 학습하여 다른 예측결과를 만들어내는데, 이를 투표(Voting)를 통하여 가장 좋은 결과값을 산출하게 하는 앙상블 방법의 한 방법이다. 투표를 하는 방식은 hard 방식과 soft 방식으로 나뉘어져 있는데, hard 방식을 적용하면

3-5 앙상블:배깅-랜덤포레스트

랜덤 포레스트는 배깅(Bagging) 분석방법의 일종으로, 데이터셋을 일정한 크기로 복원추출하는 부트스트래핑(Bootstrapping)을 하여 복원된 데이터셋을 다수의 의사결정나무 분류기를 통하여 거대한 숲을 만들어 내는 기계학습 방법 중 하나이다. 랜덤 포레스트는 학습 데이터에 대해 과소적합, 과적합이 많고 편향된 분산값을 만들어내어 실제 예측에 대해서는 예측력이 약했던 단점을 보완한 것이 랜덤 포레스트 방법이다. 랜덤 포레스트로 다양한 의사결정나무를 만들어 분산을 낮추어 주며, 일반화 오류가 의사결정나무보다 상대적으로 적은 편이다.

-앙상블:부스팅

4. 실증분석

4-1 데이터 설명

문제: 정확도가 과적합 되어 있을 가능성이 있다. 따라서 3겹 교차검증을 활용한 정확도를 산출한다.

표현: ROC-AUC 커브를 6개를 합쳐놓고 비교하는 형태의 그래프를 그려야한다

결과표

혼동 행렬

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | 3-Fold Acc Mean | Precision | Recall | F1-Score | Kappa | ROC-AUC |
| 로지스틱 회귀분석 | 71.94% | 75.93% | 73.97% | 72.92% | 73.47% | 0.4370 | 0.7187 |
| 의사결정나무 | 70.50% | 74.92% | 73.53% | 68.49% | 70.92% | 0.4107 | 0.7061 |
| 서포트 벡터 머신 | 74.82% | 75.94% | 75.32% | 78.38% | 76.82% | 0.4929 | 0.7457 |
| 앙상블-보팅 | 77.70% | 77.23% | 74.36% | 84.06% | 78.91% | 0.5543 | 0.7774 |
| 앙상블-배깅(랜덤포레스트) | 74.10% | 76.08% | 80.00% | 63.77% | 70.97% | 0.4812 | 0.7403 |
| 앙상블-부스팅(lightGBM) | 78.42% | 75.50% | 81.16% | 76.71% | 78.87% | 0.5685 | 0.8562 |

5. 결론

1. 교신저자(corresponding author), kimtk@kw.ac.kr [↑](#footnote-ref-1)