

빅데이터와 AI를 활용한 신용평가의 변화 시도

2018.06.28

박인천 기업/금융평가본부장 iseapark@scri.co.kr 02-6966-2473

박현희 금융평가실장 hyunhee.park@scri.co.kr 02-6966-2466

SUMMARY

서울신용평가(주)(이하 “SCR”)는 빅데이터/AI와 관련된 최근 기술적 진보에 주목하여, 자회사 (주)SAI¹를 활용하여 신용평가 기법과 이들 기술을 접목함으로써 기존 신용평가 방식이 지니는 한계를 극복하기 위한 프로젝트를 진행하고 있다.

SCR이 인식하고 있는 기존 신용평가 방식의 한계는 다음과 같다. 첫째, 신용등급이 업데이트되는 속도가 느려 신용위험과 관련된 최신 정보를 제대로 반영하지 못한다. 둘째, 평가방법론상 정성적 요소에 대한 높은 평가비중에도 불구하고 이에 대한 객관적인 판단 근거가 부족하다. 셋째, 신용등급 커버리지가 일부 우량 대기업 중심으로 제한되어 있어 자본시장의 정보비대칭 해소라는 본연의 역할 수행에 미흡하다.

기존 신용평가의 한계를 극복하기 위하여 SCR이 제시하는 해법은 빅데이터 등 신용평가를 위한 정보 원천의 확대, 그리고 AI 분석방법론 도입이다.

우선 SCR은 회계정보에 대한 분석 의존도를 축소하고, 다각적이고 입체적인 신용분석을 위하여 빅데이터를 구축하고 있다. SAI의 주주사가 보유하고 있는 시장 데이터 베이스(이하 “DB”)와 지적재산(IP) DB 등을 기반으로 하여 빅데이터 구축에 착수하였다. 빅데이터 기술의 활용과 관련하여 SCR은 재무 DB와 시장가격(주가) DB를 결합한 두 가지 종류의 부실확률 예측 모형을 개발 완료한 상태이다. SCR은 지적재산(IP) DB가 기업의 장기 실적 및 생존 여부에 대한 의미 있는 설명변수로 활용될 수 있다는 점을 확인하였고, 현재는 이를 평가모형에 반영하는 작업을 진행하고 있다.

한편, SCR은 AI 분석방법론을 기반으로 한 신용평가 모형들을 개발하였으며, 기존 시장 유효등급을 대상으로 이들 모형에 대한 테스트 작업을 성공적으로 완료하였다. 각각의 모형에서 산출된 신용등급은 97% 수준의 예측 정확도(신용등급 Notch 기준)를 시현하였다. SCR은 향후 신용평가 대상이 중소기업을 중심으로 확대될 것에 대비하여 지속적인 학습 과정을 통하여 모형을 진화시켜 나가고 있다. 또한, SCR은 AI 기반 평가 모형에 빅데이터를 순차적으로 탑재할 예정인데, 이로써 신용등급의 정확도는 점차 제고될 것으로 전망된다.

¹ (주)SAI(Seoul Artificial Intelligence Inc.)는 2017년 11월 설립된 SCR 자회사로서 빅데이터와 AI 기반의 솔루션 개발업체임.

SCR이 빅데이터와 AI 기반의 신용평가 모형 도입에 거는 기대효과는 다음과 같다. 첫째, 신용평가 과정에 있어서 회계 데이터 이외 다양한 데이터(자본시장 DB, 지적재산 DB 등)를 반영함에 따라 무엇보다도 구조적인 시차문제(Time lag)를 극복해가며, 신용등급의 적시성이 크게 개선될 것으로 기대된다. 둘째, 빅데이터, 특히 지적재산 DB를 반영함으로써 상대적으로 낮은 신용등급(High Yield 등급) 구간에 속하는 기업에 대하여 신용등급 변별력을 강화할 수 있을 것으로 예상된다. 셋째, 빅데이터와 AI 기반의 진화된 신용평가 모형은 신용등급 정확도의 훼손 없이 신용등급 커버리지를 획기적으로 확대함으로써 국내 자본시장의 정보비대칭 해소에 기여할 수 있을 것이다.

SCR은 기존 신용평가 방식의 한계를 극복하는 데에서 한 걸음 더 나아가, 빅데이터와 AI 기반의 진화된 평가모형을 활용하여 신용평가 시장의 새로운 영역, 즉 기술 중소기업을 위한 특화된 신용평가 시장을 개척하고자 한다. 이를 위해 SCR은 지적재산(IP) 관련 DB가 특히 중소기업 신용평가의 정확도를 제고하는 데에 유용한 정보가치를 지니고 있는 사실에 주목하여 AI 기반의 신용평가 모형에 설명변수로 반영하는 작업을 진행하고 있다. 향후 기술 중소기업이 자본시장을 통하여 기술의 사업화 자금을 원활하게 조달할 수 있기 위해서는 이들 기업에 대한 보다 정확한 신용평가가 전제되어야 하는 바, SCR은 기술 중소기업의 신용위험에 대한 차별화된 정보의 제공을 통하여 특화된 신용평가 부문의 토대를 구축하고자 한다.

1. 들어가며

국내 자본시장의 발전과 더불어 투자 의사결정의 주요 기준으로서 **신용등급 정보에 대한 수요가 양적으로 크게 확대된 가운데, 신용평가의 질적 제고를 바라는 투자자의 요구는 여전히 큰 상황이다.** 이에 각 신용평가사는 방법론 개정을 통하여 평가 논리를 보다 정교하게 만들고, 크레딧 이슈 리포트 발간 및 세미나 개최 등을 통하여 투자자와의 소통을 강화하는 등 신용평가의 질적 수준을 향상시키기 위한 노력을 지속하고 있다.

그러나, 기존 신용평가사가 추진하고 있는 신용평가의 질적 제고 노력은 분명한 한계에 직면할 것으로 보인다. 기본적으로 신용평가를 위한 정보 수집 및 애널리스트의 분석 역량(Capacity) 측면에서 분명한 제약요인이 존재하기 때문이다.

이에 **서울신용평가(주)**(이하 “SCR”)은 빅데이터/AI와 관련된 최근 기술적 진보에 주목하여, **신용평가 기법과 이들 기술을 접목함으로써 기존 신용평가 방식이 지니는 한계를 극복하고자 한다.** 본 보고서에서는 먼저 기존 신용평가 방식이 지니는 문제점을 점검해 보고, 이후 신용평가 분야에 응용 가능한 빅데이터 및 AI 기술 분야를 찾아보고자 한다. 마지막으로 현재 SCR이 추진하고 있는 빅데이터와 AI 기술에 기반한 신용평가 Up-grade 프로젝트의 진행경과 및 향후 계획을 설명하고자 한다.

2. 기존 신용평가 방식의 한계

SCR이 인식하고 있는 기존 신용평가 방식의 한계는 다음과 같으며, 신용평가 시장의 외연 확대와 신용평가의 질적 개선을 바라는 투자자의 요구에 부응하기 위해서는 반드시 해결되어야 할 문제들인 것으로 판단하고 있다.

(1) 신용등급 정보의 적시성 저하

기존 신용평가 방식에서는 신용등급이 업데이트되는 속도가 느려 **평가대상의 신용위험에 대하여 최신 정보를 제공하는 데에 구조적인 한계를 지니고 있다.** 이는 기본적으로 신용평가 과정 자체가 회계정보(감사보고서)에 대하여 과도하게 의존하기 때문에 초래되는 문제점이다. 통상 12월 결산법인의 경우 주주총회에서 확정된 재무제표를 3월말까지 공시한다. 이를 기초로 4월초에 신용평가를 개시한다고 가정하더라도, 신용평가사는 재무제표 기준일(12월말)로부터 3개월 이상 경과한 과거(Out-of-date) 회계정보를 기준으로 등급을 결정할 수 밖에 없는 상황이 된다.

신용등급 정보의 적시성을 개선하려면 평가대상 기업에 대한 다양한 정보를 신용평가에 반영함으로써 **회계정보에 대한 의존도를 축소하는 것이 필요하다고 판단된다.** 예를 들어 상장 기업의 경우에는 주가 데이터에 최신 기업정보가 신속하게 반영되는 점을 감안하여 신용평가에 주가 정보를 반영하는 방안을 고려할 수 있을 것이다. 또한 우리나라 산업구조가 고도화됨에 따라 기업 활동에 있어서 지적재산권, 상표권 등 무형자산의 중요성이 높아지고 있는

점을 감안하여 이를 신용평가에 반영하는 방안도 검토할 필요가 있다고 본다.

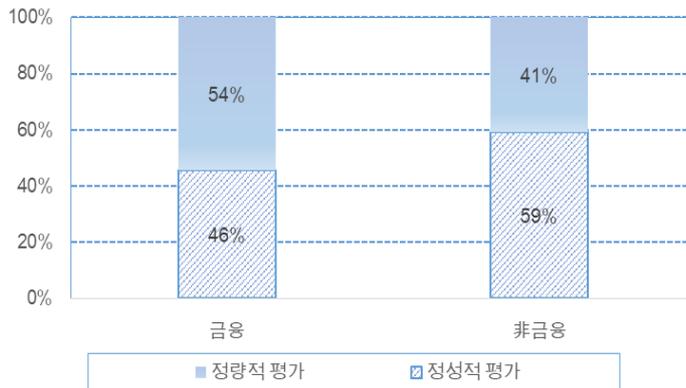
특정 신용평가 시점부터 그 다음 평가시점 사이의 기간 중 신용등급 사후관리(모니터링)에 있어 실질적인 공백이 있는 점도 신용등급 조정의 적시성을 저하시키는 요인이다. 특히 비상장 기업의 경우 공시 의무가 없어 중요 크레딧 이벤트가 발생한다고 하더라도 평가사가 이에 대한 정보를 적시에 획득하고 신용등급에 반영하는데 있어 상당한 애로 요인이 존재한다. 또한, 평가사 애널리스트의 평가대상 커버리지가 확대됨에 따라 시간적, 물리적 제약으로 개별 담당 업체에 대한 모니터링 강도는 약화될 수 밖에 없다.

(2) 정성적 평가에 있어 객관적 근거 제시 미흡

평가방법론상 높은 평가비중(가중치)을 두고 있음에도 불구하고 현실적으로 정성적 요소에 대한 객관적 평가 근거를 확보하기가 어려운 상황이다. 사업위험 평가는 그 특성상 대부분 정성적 평가요소로 구성되며, 재무위험 평가조차 정성적 평가요소(재무정책과 유통성 등)를 포함한다. 정성적 평가요소 중 일부는 시장점유율, 매출규모(또는 자산규모) 등 비교적 단순한 수치를 활용하여 평가할 수 있지만, 거래처와의 관계, 브랜드 인지도, 기술력 수준, 금융시장 내 평판 등 순전히 질적인 요소의 경우에는 명확한 평가 근거를 제시하기가 특히 어렵다.

[그림 1] SCR 평가방법론상 정성적 평가 비중

(단위: %)



주: 정성적 평가 비중은 금융 및 비금융 각각의 범주에 속하는 여러 업종별 방법론상 비중들의 평균치임.

객관적 근거가 부족한 상황에서는 경우에 따라서 애널리스트의 자의적인 해석이 가능하여 평가의 일관성 유지가 어렵다. 또한, 평가 기초정보의 부족으로 등급이 실제 신용위험 대비 보수적으로 산정될 가능성이 존재한다. 주식 애널리스트와 비교할 때 크레딧 애널리스트는 잠재적 위험(Down-Side Risk) 점검에 보다 주안을 두고 있기 때문이다.

(3) 신용도와 비선형적인 관계를 가지는 평가요소에 대한 부정확한 판단 가능성

평가방법론은 각 평가요소와 신용도 사이에 선형적인 상관 관계를 전제로 하고 설계되어 있어 실제 비선형적인 관계에 있는 경우 평가결과의 정확성이 다소 떨어질 수 있다. 예를 들어 금융기관의 주요 수익성 평가지표 중 하나인 ROA(총자산이익률)와 관련하여, 평가방법론에

서는 선형적인 상관관계를 가정하여 동 지표 수준이 높을수록 이익창출력이 우수하고 결과적으로 신용도에 긍정적인 것으로 평가한다. 하지만, 실제 평가과정에서 애널리스트는 평가대상 금융기관의 ROA가 적정 수준을 크게 상회하면 오히려 직관적으로 신용도에 부정적 요소로 간주하고 초과 수익성이 가능했던 이유를 찾게 된다. 대부분의 경우 평가대상 금융기관이 수익성 증대를 위하여 무리하게 위험을 감수한 결과로 밝혀지게 된다.

(4) 제한적인 신용등급 커버리지에 따른 자본시장 인프라로서의 역할 미흡

현재 신용등급은 일부 공기업 및 우량 대기업을 중심으로 생성되고 있으며, 광범위한 기업을 대상으로 한 신용등급 정보는 자본시장에 충분히 공급되지 못하는 실정이다. 유가증권 및 코스닥 상장기업조차 신용등급을 보유한 기업의 비중이 각각 30%와 3% 수준에 불과하다. 이는 우리나라 신용평가 시장이 법적 규제 하에서 기업어음, 회사채 등 특정 금융상품에 대한 평가로 제한되어 있기 때문이다.

결국 투자자 입장에서는 투자대상 업체 탐색(신용등급 정보 확보)을 위하여 자체적인 비용과 추가적인 시간을 투자해야 하는 상황이며, 이는 자본시장에서의 거래를 위축시키는 요인이 되고 있다. 신용평가의 본질적인 기능 중 하나가 자본시장에 공정한 신용정보를 제공함으로써 기업(평가대상)과 투자자 사이의 정보 비대칭을 해소하고, 자본시장의 효율성을 제고하는데에 있음을 되짚어 볼 일이다.

[표 1] 신용등급을 보유한 상장기업 비중

	상장기업 수	신용등급 보유업체	신용등급 보유비중
유가증권 시장	777 개	231 개	29.7%
코스닥 시장	1,268 개	43 개	3.4%
(소 계)	2,045 개	274 개	13.4%

자료: KIS Pricing, 한국거래소, SCR

자본시장을 통한 기업의 자금조달을 활성화시키기 위해서는 신용등급 커버리지가 획기적으로 확대되어야 하지만, 기존 신용평가 방식을 통한 등급 생성 역량(Capacity)으로는 이를 충족시키지 못할 것으로 판단된다. 자본시장을 통한 자금 공급이 소수의 우량 대기업을 넘어 보다 많은 기업(중소기업 포함)으로 확산되려면, 이들 기업들에 대한 신용등급 정보가 사전적으로 충분히 생성되는 것이 핵심적인 필요조건일 것이다. 하지만, 기존 채무증권 발행업체 의뢰에 기반한 신용평가 제도², 그리고 사실상 모든 신용분석 과정을 애널리스트가 커버하는 현재의 평가방식 하에서는 관련 소요시간 및 비용 등을 고려할 때 자본시장 활성화를 위한 조건이 충족되기 힘들 것으로 보인다.

² 2018년 2월부터 투자자(제3자) 요청 신용평가 제도가 도입되었으나, 신용평가 소요비용 수준 및 그 부담주체와 관련된 이슈 등으로 평가실적이 극히 미미한 상태임.

3. 신용평가의 진화를 위한 빅데이터와 AI 활용 방향

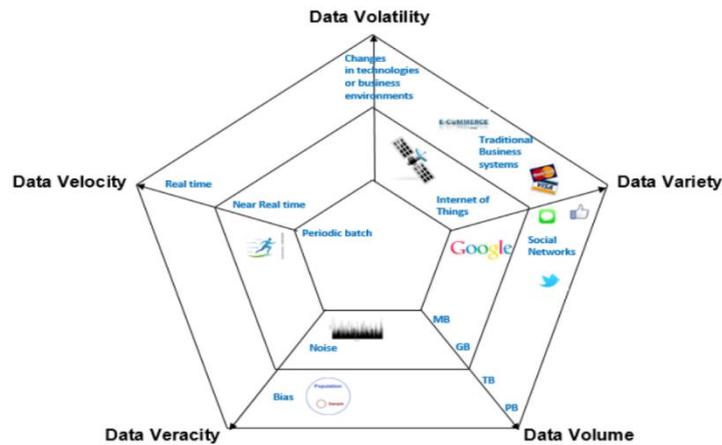
(1) 빅데이터(Big Data) 개념

빅데이터는 특정목적을 위해 만들어진 통계적 데이터와는 달리 경제·경영 관리시스템, 소셜 네트워크, 인터넷 상에 축적되어 발견된(found) 부산물이다. 디지털 경제의 확산과 소셜네트워크의 활성화, 사물인터넷 보급 등으로 다양한 분야에서 빅데이터가 빠르게 축적되고 있다.

아직까지 빅데이터에 대한 공통된 정의는 없으나, 통상 대량의 정보(high-volume), 빠른 처리 속도(high-velocity), 종류를 불문한 다양성(high-variety)이라는 3V로 특징지어진다.³ “대량의 정보”는 기계, 네트워크, 대인간의 상호작용을 통해 기하급수적으로 생성되는 디지털 처리 용량(Exabyte)의 데이터를 지칭한다. “빠른 속도”는 데이터의 수집, 분석, 활용 속도를 지칭하며, “다양성”은 데이터의 종류와 원천에 관련되어 있다.

최근 들어 빅데이터를 기존의 비즈니스 영역에 결합시키는 노력을 거치며 3-V에 새로운 2-V가 추가되었다. 정확성(Veracity)과 변동성(Volatility)이 바로 그것이다. 정확성은 빅데이터를 분석하는 데 있어 기업이나 기관에 수집된 데이터가 정확한 것인지, 분석할 만한 가치가 있는지 등을 살펴보는 것이고, 변동성은 빅데이터가 생성된 사업환경이나 기술이 변화하여 분석의 실효성이 떨어지는 것은 아닌지 등을 살펴보는 것을 지칭한다.

[그림 2] 빅데이터의 5V - Volatility, Variety, Velocity, Veracity, and Volume



Based on Doug Laney, 2001

자료: IMF, Big Data: Potential, Challenges, and Statistical Implications, September 2017

³ Gartner사의 애널리스트인 Doug Laney가 2001년에 3-V개념을 고안하였다.

[표 2] 국제유럽경제위원회 (UNECE)의 빅데이터 분류

개인정보 등	사회적 관계서비스: 페이스북, 트위터, 링크드인
	블로그, 논평
	인터넷 등의 검색엔진 (구글)
	모바일 정보 콘텐츠: 문자메시지, 통화녹음기록, 상세자료, 위치 정보
비즈니스 정보	공공기관의 생성 자료: 행정적으로 모은 정보
	기업의 생성자료: 거래정보, 금융정보, 전자(온라인) 상거래정보, 신용카드 정보
센서 정보	고정형 감지장치로부터 생성: 주택 자동화, 날씨/오염정보 감지장치, 교통 감지장치, 교통카메라, 과학 감지장치
	모바일 감지장치로부터 생성: 모바일 위치 (GPS), 자동차, 위성사진

자료: IMF, Big Data: Potential, Challenges, and Statistical Implications, September 2017

(2) 빅데이터 활용 방향

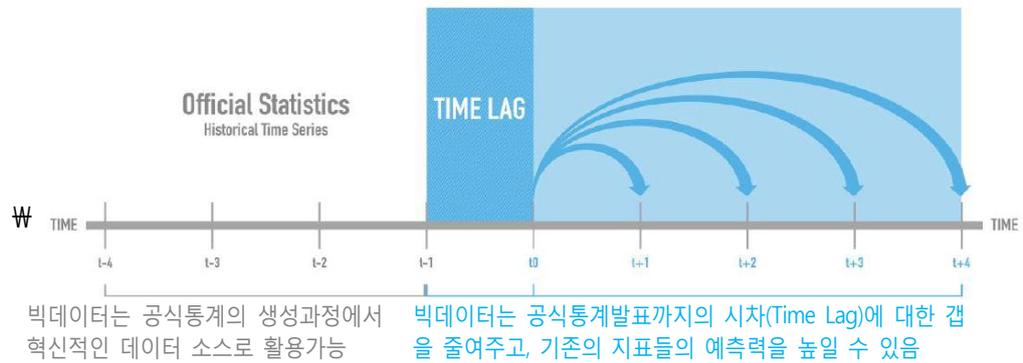
신용평가 분야에서 빅데이터의 활용 방향은 크게 2가지로 볼 수 있다.

첫째, 실시간 빅데이터를 활용하여 광범위한 데이터 사이의 상관관계를 기반으로 새로운 모니터링 지표를 도출해낼 수도 있고, 보다 효과적인 조기경보 시스템(Early Warning System)을 구축할 수 있을 것으로 기대된다. 빅데이터를 활용하더라도 어떤 현상의 발생 원인에 대한 명확한 설명은 어려울 수도 있으나, 적어도 어떤 현상이 발생하고 있다는 경보차원의 상관관계나 패턴을 찾는 데에는 빅데이터가 매우 유용할 수 있다. 전통적인 데이터 활용이 필요한 데이터(거시경제, 산업, 재무정보 등)를 수집하고 생성하여 통계모형을 이용해 인과관계를 설명하는 방식이었다면, 빅데이터의 활용은 기존과는 다른 새로운 데이터 접근방식을 가져올 것이다.

둘째, 빅데이터의 활용으로 공식적인 통계에 대한 분석 의존도를 축소하고 신용등급 정보의 적시성을 개선할 수 있을 것이다. 공식적인 통계(거시경제, 산업동향 등) 및 재무 데이터의 수집과 분석에 필요한 시차(Time Lag)로 인하여 이미 과거 수치(Out-of-date)가 되어 버린 산업 통계와 회계감사 보고서에 대한 신용분석 의존도가 높을수록, 신용등급의 적시성은 구조적으로 저하되는 문제점이 존재한다. 이에 기존 정보 원천(재무 데이터 등)에 새로운 정보 원천(인터넷 데이터 등 실시간 현장의 움직임을 반영한 지표들)을 결합하여 분석함으로써 통계의 시차 문제를 줄이고, 현재 이용하고 있는 평가지표들의 예측력을 높일 수 있을 것이다.

일례로 신용평가 영역에서 빅데이터를 이용하면, 회사의 분기별 재무제표가 공시되는 시점까지 기다리지 않고도 실시간으로 회사의 매출 트렌드를 파악할 수 있다. 예를 들어, 대형 소매 유통업 분석의 경우 입점 고객에 대해서는 인공위성 데이터를 이용해 주차장 입출차 정보를 분석하고, 인터넷 고객에 대해서는 제품에 대한 인터넷 검색 빅데이터를 분석하여 재무제표가 확정되기도 전에 매출 규모를 예측할 수 있다.

[그림 3] 빅데이터의 잠재력



자료: 국제통화기구, BIG DATA: POTENTIAL, CHALLENGES, AND STATISTICAL IMPLICATIONS, 2017년 6월

(3) AI(Artificial Intelligence) 개념

AI는 머신러닝(Machine learning)과 데이터 발굴(Data mining)을 토대로 하여 반복적인 데이터 작업 수행을 자동화함으로써 많은 양의 데이터로부터 정보를 추출하는 데이터 집약적 기술이다. 그동안 사람이 수행하던 데이터로부터 정보를 추출하는 작업을 기계가 대신 하는 AI 기술은 최근 컴퓨터 연산능력의 강화 및 알고리즘의 진보 등으로 괄목할 만한 발전을 보여왔다.

머신러닝은 컴퓨터에게 데이터를 분석하고 그 결과를 통해 학습을 반복하게 함으로써, 새로운 데이터가 주어졌을 때 결과를 도출해내고 예측하도록 하는 방식이다. 많은 양의 데이터와 알고리즘들을 사용해 기계를 훈련시키고 업무를 수행하는 방법을 배우도록 하는 것이 머신러닝이다. Apple의 Siri나 Facebook의 뉴스 피드 (News Feeds), Netflix의 영화 추천은 머신러닝의 일종이다.

데이터 발굴(Data mining)은 대량의 데이터를 효율적으로 검색함으로써, 언뜻 보아 알 수 없는 새로운 경향이나 패턴을 발견하는 것이다. 각 데이터의 상관관계를 AI 기법을 통해 밝혀주는 것이다.

(4) AI 활용 방향

신용평가에 AI를 활용함으로써 기대할 수 있는 효과는 크게 2가지 관점으로 나누어 볼 수 있다.

첫째, AI는 애널리스트(human analyst)가 해왔던 단순 반복적인 정보 수집 및 모니터링 업무를 대체할 수 있을 것이다. 기계(컴퓨터)는 외부로부터 어떤 방해도 받지 않고, 중간 휴식 없이도 신용평가를 위한 기초 작업(뉴스 분석, 통계 데이터 분석 등)을 동일한 집중도로 수행할 수 있다. 예를 들어 65,000여 개나 되는 수많은 기관이 발행하는 미국 지방채(municipal bond)의 신용위험을 애널리스트가 일일이 모니터링 하기에는 현실적으로 어려움이 많았다.

그러나, 지금은 신용평가사를 비롯, 기관투자자들, 브로커들이 off-the-shelf AI를 이용해 미국 지방채에 조기경보가 될만한 필요한 뉴스를 걸러내고 있다.

둘째, AI는 신용위험 모형 응용(Credit risk modeling applications)에 상당히 기여하였는데, 특히 머신러닝을 응용한 방법은 설명변수와 부도위험 간의 관계가 비선형적인 경우 좋은 설명력을 보여주었다. 머신러닝 모델은 통계모형의 가정에 제한되지 않아 애널리스트가 데이터를 통해 추론하는 것보다 더 나은 통찰력을 제공할 수 있을 뿐만 아니라, 애널리스트와는 대조적으로 편향된 의견에 빠질 위험(Bias risk)을 배제할 수 있다.

(5) 빅데이터 및 AI 기술 활용과 관련된 기존 제약 요인

물론 빅데이터 및 AI 기반의 신용평가 모형을 개발하고 운용함에 있어, 기술적 또는 제도적 요인으로 인하여 다음과 같은 한계점이 존재한다.

우선 AI 기반의 평가모형은 데이터 분석 및 신용도 예측에 있어서 우수한 역량을 보이고 있으나, 예측 결과를 산출한 원인에 대하여 모형 자체적으로 설명할 수 없다는 한계를 아직 지니고 있다. AI 모형의 분석 과정은 마치 블랙박스과 같아 평가 의뢰인에게 신용등급 산출 논리를 직접적으로 설명할 수는 없다. 따라서 현재의 기술 수준에서는 애널리스트들이 AI 평가모형이 생성한 신용등급 및 평가요인에 대하여 사후적인 검증절차를 거친 후, 평가의뢰인이나 투자자들에게 설명해 주어야 한다. 더욱이 신용평가 산업에 대한 금융당국의 감독규제 방향이 신용평가 프로세스의 투명성 제고에 맞추어져 있는 점을 감안할 때 AI 모형의 활용 범위는 당분간 상당히 제한될 수 밖에 없다.

한편, 각종 규제 등으로 기업 관련 데이터 수집이 제한되는 점도 빅데이터와 AI 기반의 신용평가 모형을 개발하는 데에 걸림돌이 되고 있다. 효과적인 평가모형 개발을 위한 핵심 인프라는 다양한 원천의 데이터 확보인데, 현재로서 확보 가능한 정보는 언론기사, SNS, 인터넷 등을 통한 간접 정보가 대부분이며, 기업, 금융기관 및 정부가 생성하거나 보유한 정보에 대한 접근은 대부분 정보보호 등의 이유로 수집이 기본적으로 차단되어 있는 상태이다. 특히 비상장 중소기업의 경우 데이터 부족이 더욱 심각한 상태이다.

4. SCR의 신용평가 Up-grade 프로젝트 (빅데이터 및 AI 기술 적용)

SCR은 기업 신용평가 분야에 빅데이터와 AI 기술을 접목하여 기존 신용평가 방식을 보완하고 나아가 차별화된 신용평가 정보를 자본시장에 제공하려는 전략을 추진하고 있다. 동 전략 하에서 2017년 11월 빅데이터와 인공지능 기반의 솔루션 전문회사인 (주)SAI를 설립하였으며, SCR과 긴밀한 협업 체계를 구축하고 있다.

[그림 4] SAI와 주요 주주간 협업구조



4-1. 빅데이터 구축 및 활용

SAI는 신용평가 모형을 개발함에 있어 토대가 되는 기업 신용도와 관련성이 높은 빅데이터를 구축하고 있다. 이는 신용등급 결정에 있어 회계정보에 대한 의존도를 축소하고 다각적이고 입체적인 분석을 통하여 신용등급의 정확도를 제고하기 위함이다.

SAI의 빅데이터는 SAI의 주주사가 보유하고 있는 지적재산(IP) 데이터베이스(DB) 등을 기반으로 구축되고 있다. SAI의 주주구성에는 SCR 이외에 WISEfn, WIPS, 그리고 SCI평가정보가 참여하고 있는데, 각 주주 회사는 현물출자 또는 업무제휴의 형태로 각사가 보유한 DB를 SAI에 제공하고 있다. 이에 따라 SAI는 (주)WIPS가 보유한 국내 민간 최대의 지적재산권 및 기술 관련 DB, SCI평가정보(주)가 보유한 기술신용평가(TCB) 및 상거래 기업평가 DB, 그리고 (주)WISEfn 등이 보유한 재무 DB와 자본시장 DB를 확보하게 되었다.

향후에도 SAI는 추가 정보원천 발굴 및 다른 DB 업체와의 신규 제휴를 기반으로 빅데이터를 지속적으로 확충할 계획이다.

[빅데이터 적용 관련 진행현황]

(1) 주가 정보를 활용한 부실예측

주가를 활용하여 상장기업의 신용위험(부도확률)을 추정하는 모형은 다른 모형 대비 부도 예측에 보다 적합한 것으로 평가되면서 이미 널리 활용되고 있으며, 넓은 범주에서 빅데이터가 신용평가에 활용되는 사례 중 하나라고 할 수 있다.

SCR은 주가 정보를 활용한 두 가지 종류의 부도확률 예측 모형을 개발 완료하였다. 첫번째 부실예측 모형은 통상적인 EDF 모형(Expected Default Frequency)과는 달리 입력정보로서 시장 정보(주가)와 재무 정보를 결합한 유형으로서 부실 예측에 있어 시장가격뿐만 아니라 기업 펀더멘털과 관련된 정보를 반영하고 있으며, 시장가격 정보만을 반영하는 EDF 모형 대비 예측치의 변동성이 상대적으로 낮다는 특징을 지니고 있다.

SCR이 개발한 또 다른 부실예측 모형은 통상적인 EDF 모형(Expected Default Frequency) 처럼 입력정보로서 시장 데이터(주가)만 활용하되, 신용위험 관련 시그널을 보다 잘 추출하기 위하여 주가 데이터에서 여타 주가등락 요인(경기순환 요소 등)을 제거한 조정 데이터(Through-The-Cycle)를 활용하는 특징을 갖고 있다.

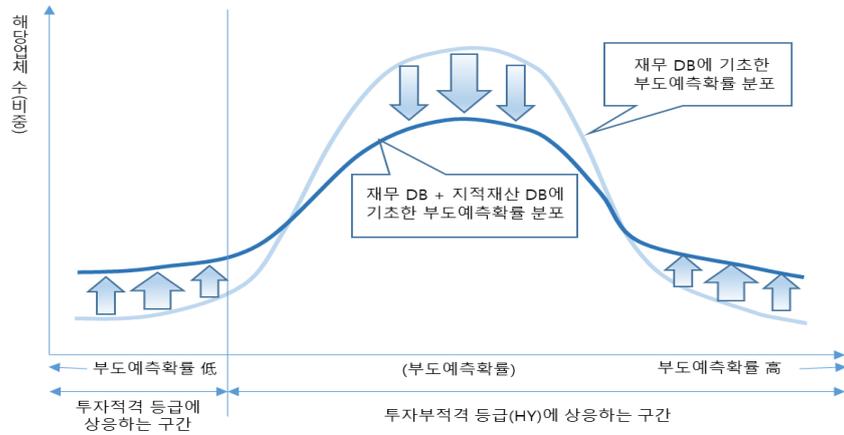
통상 주가 정보를 활용한 부도예측모형은 부도예측의 적시성은 우수하나, 주가의 변동성이 커 부도위험을 안정적으로 측정하기 어렵다는 문제를 가지고 있었다. SCR은 주가 정보와 재무 데이터를 결합한 모형과, 경기순환 요소를 배제한 Through-The-Cycle(TTC) EDF 모형을 채택함으로써 부도 위험을 적시에 발견하면서도 안정성 요건을 충족시켰다.

(2) 지적재산(IP) 정보 활용을 통한 신용등급 변별력 강화

SCR은 지적재산권(IP) 정보가 기업의 장기 실적 및 생존 여부에 대한 의미 있는 설명변수로 활용될 수 있다는 점을 밝혀 내었다.

- ① 우선 2013년부터 지적재산 DB를 기초로 기술집약적 업종에 속한 외감기업 대상으로 조사한 결과에 따르면, 기업의 보유IP 관련 특성값과 기업의 생존확률 사이에 상당한 상관관계가 존재함을 확인하였다. 즉, 보유IP와 관련된 평가등급이 우수한 기업은 그렇지 못한 기업 대비 생존확률이 차별적으로 높게 나타났다.
- ② 또한, 통계 모형을 통하여 외감기업의 부도예측확률 구간별 분포(신용도 분포)를 산출한 결과에 따르면, 재무 DB만을 사용했을 때보다 지적재산 DB를 추가적으로 반영하여 부도확률을 예측한 경우 특히 저신용등급 구간(장기등급 BB+ 이하 HY 등급)에서 부도예측확률(신용도)의 변별력이 개선되었다. 즉, 지적재산 DB를 추가적으로 통계 모형에 반영할 때 기존 HY로 분류되었던 일부 기업群은 투자등급으로 재분류되었으며, 또 다른 일부 기업群은 더 낮은 등급으로 재분류되는 등 추가 DB의 투입으로 부도예측확률 분포에 있어 실질적 변화가 발생하였다.

[그림 5] 지적재산 DB를 추가 반영했을 때의 부도예측확률 분포 변화



자료: SCR

현재 SCR은 지적재산 관련 DB를 AI 기반의 신용평가 모형에 설명변수로서 반영하는 작업을 진행하고 있다. 통계 모형에서 확인된 결과들이 AI 모형을 통하여 검증될 경우, 향후 평가 대상 기업은 실제 사업성과에 대한 과거 이력이 다소 짧더라도 지적재산권 보유 여부 및 그 특성값에 따라서는 모형 신용등급 산정에 있어 가산점을 받을 가능성이 있게 된다.

(3) 텍스트 마이닝을 통한 정보 획득

텍스트 정보(산업분석 리포트, 언론기사, 인터넷, SNS 등)로부터 텍스트 마이닝(Text Mining) 기법을 활용하여 산업별 경기순환 사이클, 투자자/소비자의 정서지수(Sentiment Indicators) 등에 대한 정보를 추출하여 정성적 평가항목에 대한 판단근거로써 활용할 계획이다. 또한, SCR은 상장기업의 사업보고서 또는 감사보고서 주식에서도 텍스트 마이닝 기법을 통하여 유용한 정보를 추출할 수 있을 것으로 보고 있다. 예를 들어 각 업체별로 사업보고서와 감사보고서 주식 등으로부터 납품관계(Supply Chain Relationship) 정보를 추출하고 다른 기업의 동일 유형 정보와 연결 짓는다면 평가대상 업체의 거래위험 노출을 쉽게 파악할 수 있을 것이다. 또한, 사업보고서상 텍스트 정보로부터 국가별 매출분포 정보를 추출하여 환율 등락에 따른 매출 및 손익 변동과 관련된 시나리오 분석을 체계적으로 구현할 수 있다.

그러나, 텍스트 마이닝을 통한 빅데이터 정보 수집은 실제 적용에 이르기까지 향후 추가적인 시간과 분석이 필요할 것으로 예상된다. 이는 텍스트 정보의 대부분이 일부 대기업을 중심으로 생성되어 있고, 현재 텍스트 마이닝 기술 수준이 정보이용자(전문가)의 수요를 충족시킬 만큼 발전되지 않았기 때문이다. 이 분야의 수요가 많은 주식시장에서도 아직까지 구체적인 성과물이 나타나지 않고 있다. 다만, 최근 관련 기술의 빠른 발전 속도를 감안할 때, 신용평가 분야에서도 텍스트 마이닝을 통한 정보의 생성이 가능할 것으로 보인다. 한편, SCR에서는 정보 원천의 생성이 일부 대기업에 국한된 점을 감안하여 우선적으로 기업보다는 산업에 대한 빅데이터 정보 수집에 주력할 계획이다.

4-2. AI 기반의 신용평가 모형 개발

SCR이 AI 기반의 신용평가 모형을 개발하고자 하는 것은 기존 신용평가 방식이 지니는 문제점 보완과 질적 개선을 추구하는 동시에, 향후 신용평가 시장의 외연 확대(신용등급 평가대상 확대)에 대비하여 분석 역량(Capacity)을 선제적으로 확충하려는 데에 있다.

모형개발 과정은 먼저 신용평가 전문가의 등급산출 논리를 기계(컴퓨터)에 이식하고, 이 논리에 의거하여 기계가 투입된 데이터를 가공하여 (신용평가 전문가였다면 부여하였을 수준의) 신용등급을 생성하도록 학습시키는 것으로 요약된다. 기본 모형이 개발된 이후에도 다양한 데이터 베이스(입력 데이터)와 결합되면서 학습과정은 지속된다. 개발된 모형이 실제 운영되면서 검증되고, 또한 정기적인 모형 갱신을 거치면서 신용등급의 예측력이 점차 강화될 것으로 기대된다.

학습 데이터가 많을수록 예측력이 개선되는 AI의 특성을 감안할 때, SCR의 AI 모형도 향후 빅데이터가 추가되면서 보다 정확한 신용등급을 생성할 것으로 전망된다. AI 모형에 투입될 시장가격(주가, 신용 스프레드 등), 지적재산 DB, 그리고 텍스트 DB 등은 개별 데이터 독자적으로도 정교한 평가모형 구축을 위한 새로운 데이터 원천이 될 뿐만 아니라, 서로 다른 이종 데이터와의 다양한 결합을 통해서도 의미 있는 데이터가 될 수 있을 것이다.

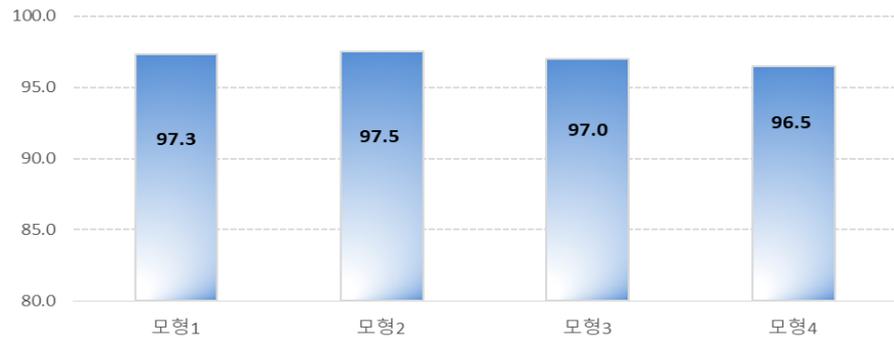
SCR은 AI 기반의 신용평가 모형 결과를 1차적으로 기존 신용평가에 참고 지표(Reference)로 활용하면서 등급의 정확도를 상호 점검할 계획이다. 또한, AI 모형 개발 과정에서 획득한 다양한 테스트 및 검증 결과를 반영하여 기존의 신용평가 방식을 개선시켜 나갈 계획이다. SCR은 현재 AI 모형을 개발하는 과정에서 다양한 데이터 원천과 신용등급 사이의 상관관계를 검증하고 있다. 이와 더불어 향후에는 기존 평가방법론과는 차별적으로 비선형(Non-linear) 평가요소를 추가하거나, 선형 평가요소의 비선형 결합 등 평가모형의 정확성을 제고하기 위하여 다양한 테스트 작업을 진행할 계획이다.

[AI 기반의 평가모형 개발 진행현황]

기존 시장 유효 신용등급(2013년부터 공시된 등급)을 대상으로 다양한 분석방법론에 기반한 여러 AI 모형을 적용한 결과, 각 모형이 생성한 신용등급은 97% 수준의 예측 정확도(신용등급 Notch 기준)를 시험하였다.

[그림 6] SCR이 개발한 AI 모형의 신용등급 예측 정확도

(단위: %)



자료: SAI

SCR은 향후 신용평가 대상이 중소기업을 중심으로 크게 확대될 것에 대응하기 위하여 다양한 AI 분석방법론을 적용한 복수의 신용평가 모형을 개발하여 테스트하고 있다. 향후에도 신규 평가대상이 기존 시장 유효등급 보유기업군과 유사한 규모 또는 업종일 경우에는 현재 SCR의 AI 모형으로도 평가사들의 애널리스트가 부여할 정도의 신용등급을 거의 생성해낼 수 있다. 기존 평가대상과 유효등급에 대해서는 AI 모형이 학습을 충분히 완료했기 때문이다. 다만, 향후 신용평가 시장이 개편되어 평가대상이 기존 우량 대기업 일변도에서 중소기업으로 크게 확산되거나, 신규 업종에 속한 평가대상이 크게 확대될 경우를 대비하여 AI 모형의 기업규모별 및 산업별 학습을 지속적으로 진행하고 있다.

5. 신용평가 Up-grade 프로젝트의 기대 효과

상장기업 및 외감기업 각각 특화되어 개발된 SCR의 AI 평가모형과 부실예측모형에 대한 운용 방안과 관련하여, 먼저 AI 모형은 기존 신용평가에 참고 지표(Rating Reference)로 활용하면서, 동시에 평가방법론과의 다양한 피드백(테스트 및 검증) 과정을 거치면서 상호 개선시켜 나갈 계획이다. 또한, AI 모형에 의하여 생성된 신용등급 정보는 기존 신용등급과는 독립된 별도의 정보로서 투자자에게 제공하는 방안을 모색해 나갈 것이다. 한편, 부실예측 모형은 BIR(Bond Implied Rating)과 더불어 우선적으로 신용등급 사후관리 강화를 위하여 투입되어 운용될 계획이다.

(1) 기존 신용평가 방식의 질적 개선

SCR이 신용평가 Up-grade 프로젝트를 추진함에 있어 그 1차적인 목표는 기존 신용평가 방식이 지니는 문제점들에 대하여 구체적인 해결 방안을 제시함으로써 기업 신용평가를 질적으로 한층 더 개선시키는 데에 있다.

① 신용등급의 적시성 개선

빅데이터와 AI 기반의 신용평가 모형을 가동하게 되면, 평가대상 업체에 대한 실시간 모니터링이 가능해질 뿐만 아니라, 실질적으로 신용평가의 빈도가 증가하면서 신용등급의 적시성이 크게 개선될 것으로 기대된다.

빅데이터 구축 이후에는 신용평가 과정에 있어서 회계정보 이외 다양한 정보(자본시장 정보, IP 정보 등)가 반영됨에 따라, 시차문제(Time lag)를 구조적으로 안고 있는 회계정보에 대한 의존도가 상대적으로 축소될 것으로 기대된다. 예를 들어 회계정보 이외 특히 자본시장 정보(주가, 신용 스프레드 등)를 활용한 기업부실 예측 확률이 매우 짧은 주기(매월 또는 매주)로 업데이트되어 애널리스트에 제공될 것이다. 다만, 일정 수준의 신용등급 안정성 유지 차원에서 부실예측 확률에 기반하여 투자자에게 조기경보 신호가 전달되려면, 부실예측 확률이 사전 설정된 기준치(Threshold)를 넘고, 동시에 담당 애널리스트의 추가 검증절차를 거쳐야 할 것이다.

② 정성적 평가요소(특히 지적재산) 반영을 통한 신용등급 변별력 제고

신용평가에 있어 회계정보 이외에 IP 정보를 필두로 한 다양한 결합정보가 반영됨에 따라 정성적 평가 요소에 대한 객관적 근거를 확보할 수 있을 뿐만 아니라, 신용등급의 정확도를 향상시킬 수 있을 것으로 기대된다.

SCR은 재무적 정보 이외에 특히 지적재산(IP) 정보가 설명변수로서 부실예측 모형에 추가로 반영되었을 때에 기업 부실예측에 있어서 변별력이 강화된다는 점을 발견하였다. 특히 저신용등급 구간(장기등급 기준 BB+ 이하 High Yield 등급)에서 변별력이 크게 제고된 것으로 나타났다. 또한, SCR은 국내 산업구조가 고도화되면서 기업 활동에서 지적재산(IP)의 보유

및 활용이 중요해지고 있는 점을 감안하여, 향후 지적재산 정보를 신용평가에 반영함으로써 기술 강소기업을 중심으로 차별적인 신용등급 정보를 제공할 계획이다.

한편, 기존 신용평가에서 기업 평판, Peer 대비 상대적 시장지위, 브랜드 가치 등은 정성적 평가요소로서 신용등급 산정에 반영되도록 되어 있으나, 이에 대한 객관적인 판단 근거는 부족한 상태였다. 향후 이들 정성적인 요소에 대한 평가 정보가 빅데이터를 통하여 생성되어 애널리스트에게 제공될 계획이다. 빅데이터 정보의 제공을 통하여 정성적 평가에 대한 객관적 근거를 보강할 수 있으며, 동시에 애널리스트의 주관적 판단 여지를 축소하여 평가의 일관성을 제고시킬 수 있을 것이다.

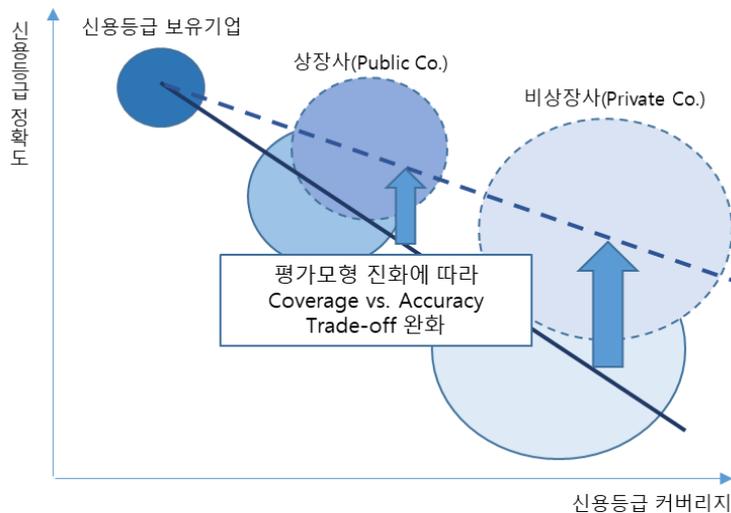
③ 기존 신용평가 모형의 선형성 보완

평가방법론은 신용등급 산출을 위한 기본적인 논리(Logic)를 담고 있으나, 평가대상 기업 고유의 평가요소 및 이에 대한 실제 숙련된 애널리스트의 평가 Know-how를 모두 반영하고 있는 것은 아니다. 특히 개별 평가요소와 신용도 사이의 매칭 관계는 현재 공시된 평가방법론에 나타난 것처럼 반드시 선형적인 관계에 있는 것이 아니다.

AI 기반의 신용평가 모형은 자기학습 과정을 지속적으로 거치면서 실제 숙련된 애널리스트의 평가 Know-how를 구현하여 신용등급을 산출할 것으로 전망한다. 즉, 숙련된 애널리스트의 오랜 평가경험을 통하여 형성된 직관을 학습하여 개별 평가요소와 신용도 사이의 비선형적인 관계까지도 모형 내에서 반영할 수 있을 것이다.

④ 진화된 평가 모형으로 신용등급 정확도 훼손 없이 커버리지 확대

[그림 7] 평가모형 진화에 따른 등급 커버리지와 정확도 사이의 Trade-off 완화



자료: SCR

주: 그림내 각 원(圓)의 크기는 평가대상의 수를 개념적으로 나타내고 있음.

기존에는 신용등급의 정확도와 커버리지는 상당한 트레이드 오프(Trade-off) 관계를 지니고 있었다. 신용등급 커버리지를 획기적으로 확대하려면, 결국 신용평가 모형에 의존하여 신용등급을 산출할 밖에 없는데, 모형으로 산출한 신용등급은 “정규” 신용평가에 의한 신용등급 대비 정확도 측면에서 열위에 있었다. 정규 신용평가는 평가대상으로부터 직접적으로 내부 자료를 수취함으로써 모형에 의한 평가와 대비하여 많은 기초 정보를 획득할 수 있으며, 평가의 Know-how 측면에서도 종전에는 크레딧 애널리스트가 모형 대비 확실한 우위를 점하고 있었기 때문일 것이다.

그러나, 빅데이터와 AI 기술 기반의 진화된 신용평가 모형의 도입으로 이러한 트레이드 오프 관계가 약화될 것으로 보인다. 즉, 신용등급 정확도 측면에서 큰 훼손 없이도 커버리지를 획기적으로 확대할 수 있다는 의미이다. 진화된 모형은 빅데이터를 활용함으로써 평가 기초정보 확보 측면에서 “정규” 신용평가와의 격차를 축소할 수 있으며, 또한 AI 분석방법론을 활용함으로써 실제 애널리스트의 평가 Know-how를 상당부분 구현 가능할 것으로 예상되기 때문이다. SCR은 빅데이터와 AI 기반의 평가 모형을 기반으로 신용등급 커버리지를 크게 확대하여 채권시장에서의 정보비대칭 해소 및 자금배분 효율화에 기여하고자 한다.

(2) 신용평가 시장의 신규 외연 확대

SCR은 빅데이터와 AI 기술 기반의 신용평가 모형을 바탕으로 기술력을 갖춘 강소기업의 자본시장 진입을 지원하고자 하며, 이는 SCR이 기존 신용평가 방식의 보완과 함께 특히 주안점을 두고 추진하는 사업 방향이다.

기존 신용평가 방식 하에서는 평가를 위한 기초 정보와 경험 부족으로 인해 비상장 중소기업의 신용등급 산정에 어려움이 많았다. 중소기업의 경우 재무정보의 Track Record가 짧고 이에 대한 신뢰성 또한 크지 않다는 점이 정량적 평가에 있어 상당한 애로요인이었다. 또한, 정성적 평가 측면에서도 대기업과 비교할 때 객관적인 평가 근거의 확보가 더욱 어렵다. 결국 평가대상 중소기업이 설령 경쟁력 있는 기술력을 보유하고 있더라도 이를 입증할 객관적 근거 또는 평가 경험의 부족을 이유로 인정 받기 힘든 실정이다.

자본시장을 통한 자금조달이 어려운 가운데, 기술 강소기업 입장에서 자발적으로 신용평가를 받을 유인이 없어 결과적으로 신용평가의 사각지대에 처해 있는 상황이다. 투자자(채권자 등) 입장에서 이들 중소기업은 신용위험이 높은 투자대상로서 인식되는 반면, 이들 중소기업에 대한 신용평가 정보는 대기업 대비 오히려 크게 부족한 상태에 있어 자본시장에서 중소기업의 자금조달은 더욱 위축될 수 밖에 없다.

SCR은 진화된 신용평가 모형을 기반으로 신용평가 시장에서 소외된 중소기업, 특히 기술 강소기업 등에 대하여 차별화된 신용등급 정보를 제공하고자 한다. 재무정보 이외에 지적재산, 기술 등에 관한 다각적인 정보원천(빅데이터)을 신용평가에 반영함으로써 중소기업 신용등급의 정확도를 높이고, 신용등급이 상대적으로 낮은 기업군 내에서의 변별력을 강화할 계획이다. 이를 위하여 SCR은 비상장/외감 기업에 특화된 AI 기반의 평가모형을 개발하고 있다.

[유의사항]

- 1) 신용등급은 특정 금융투자상품, 금융계약, 발행자 등의 상대적인 신용위험에 대한 현재시점에서의 서울신용평가사의 의견입니다. 신용등급은 신용위험을 제외한 다른 위험(시장가치 변동위험, 유동성위험, 운영위험 등의 여타 위험)에 대해서는 설명하지 않습니다. 특히, 구조화금융상품의 신용등급은 법령, 조세제도의 변경 등에 따라 발생하는 위험을 반영하지 않습니다. 또한 신용등급 및 당사 간행물에 포함된 견해는 현재 또는 과거 사실에 관한 서술이 아니며, 당사 고유의 평가기준에 따라 신용위험에 대하여 예측한 의견으로서, 이러한 예측정보는 실제 결과치와 다르게 나타날 수 있습니다. 신용등급 및 당사 간행물에 포함된 견해는 환경변화 및 당사가 정한 기준에 따라 변경 또는 취소될 수 있습니다.
- 2) 신용등급, 평가의견 및 간행물은 금융투자상품에 대한 투자자문을 의미하지 않으며, 특정 증권에 투자의사결정(매매, 보유 등)을 권유하는 것이 아닙니다. 신용등급이나 평가의견, 간행물 등은 정보이용자의 투자결정을 대신할 수 없으며, 정보이용자의 투자결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 정보이용자들은 유가증권, 발행자, 보증기관 등 투자대상의 위험에 대해 자체적인 분석과 평가를 통해 투자에 대한 의사결정을 하여야 합니다. 특히 당사의 신용등급과 간행물은 개인투자자들이 이용하는 것을 전제로 하고 있지 않으며, 따라서 개인투자자들이 당사의 신용등급과 간행물을 이용하여 투자의사결정을 하는 것은 바람직하지 않습니다.
- 3) 신용평가서, 평가의견, 간행물 등 보고서는 평가대상자 또는 의뢰인이 제출한 자료와 각종 공시자료 등 당사가 객관적으로 정확하고 신뢰할 수 있다고 판단한 자료원으로부터 수집한 자료에 근거하고 있으며, 당사는 평가대상자 또는 의뢰인으로부터 제출자료에 거짓이 없고, 중요사항이 누락되지 않았으며 중대한 오해를 유발하는 내용이 들어있지 않다는 확인서를 받고 있습니다. 당사는 평가대상자 또는 의뢰인이 정확하고 완전한 정보를 적시에 제공한다는 전제하에 신용평가업무를 수행하고 있으며, 신용평가 과정에서 이용하는 정보에 대해 별도의 실사나 감사를 실시하고 있지 않습니다. 또한 신용평가의 특성 및 자료원의 제한성, 분석자, 분석도구 또는 기타 요인에 의한 오류의 발생가능성이 있기 때문에 신용등급 결정시에 활용되었거나 본 보고서에 주요 판단근거로서 제시된 어떠한 정보에 대해서도 그 정확성, 적합성 또는 충분성을 보증하지 않으며, 제공된 정보의 오류, 사기 및 허위, 미제공 등에 따른 결과에 대하여 당사는 어떠한 책임도 부담하지 않습니다.
- 4) 신용등급 및 본 보고서에 반영되었거나 포함되어 있는 정보 및 기재내용의 정확성, 완전성 및 적합성을 당사가 보증하거나 확약하지 않습니다. 본 보고서에 포함된 모든 정보들은 신용등급 부여에 필요한 주요 판단 근거로서 제시된 것이며, 평가대상에 대한 모든 정보가 기재된 것은 아닙니다. 또한, 당사는 고의 또는 중대한 과실에 기인한 사항을 제외하고 신용등급 및 본 보고서에 포함된 정보의 이용으로 발생하는 어떠한 손해 및 결과에 대해서도 책임지지 않습니다.
- 5) 본 보고서에 수록된 모든 정보의 저작권은 서울신용평가사의 소유입니다. 따라서 당사의 사전서면동의 없이는 무단 전재되거나 복사, 인용, 재판매 또는 유포될 수 없습니다.