



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

벤처캐피탈리스트 투자 요인분석 및 예측 모델링

연세대학교 정보대학원
비즈니스빅데이터분석 전공
손 애 린

벤처캐피탈리스트 투자 요인분석 및 예측 모델링

지도교수 이 준 기

이 논문을 석사 학위논문으로 제출함

2021년 12월

연세대학교 정보대학원
비즈니스빅데이터분석 전공
손 애 린

손애린의 석사 학위논문을 인준함

심사위원 이 준 기 인

심사위원 장 백 철 인

심사위원 이 정 훈 인

연세대학교 정보대학원

2021년 12월

감사의 글

원하는 연구를 할 수 있게 지지해주시고, 지도해주신 이준기 교수님 감사드립니다. 교수님의 가르침 아래 더 넓은 세상을 알고 배울 수 있었습니다. 논문을 심사해주신 장백철 교수님, 이정훈 교수님께도 감사드립니다. 교수님들의 조언으로 졸업 논문을 잘 마무리 할 수 있었습니다.

ICeT 연구실 분들에게도 감사를 전합니다. 먼저, 술선수범과 아낌없는 조언을 주신 황지욱 선생님. 물심양면으로 도움을 준 민규오빠. 함께 공부하면서 큰 힘이 되어준 동기 왕수, 혜인, 수진. 동생들이지만 많은 것들을 배우고 즐겁게 공부할 수 있었던 지훈, 형록, 영진. 궁금한게 많았던 후배들에게 귀찮은 내색 없이 배움을 나눴던 선배 인아, 소정, 형주에게도 진심으로 고마움을 전하고 싶습니다. 학교에서 만난 소중한 인연인 민수, 동신, 오현, 효중, 동훈, 동현, 용준. 고맙습니다.

저를 있는 그대로 아껴주는 친구들, 윤현, 수인이에게도 고맙다는 말을 전합니다. 친구들 덕분에 힘을 내어 공부할 수 있었습니다. 힘들 때마다 위로와 격려로 이끌어준 남자친구 정현. 사랑하고, 고맙습니다. 마지막으로 세상에서 가장 사랑하고 존경하는 부모님, 늦은 나이에 새로운 공부를 하는데 아낌없는 응원과 지지를 해주셔서 감사합니다.

차 례

그림 차례	ii
표 차례	iii
국문 요약	iv
제1장 서론	1
제2장 이론적 배경	3
2.1 벤처기업 및 벤처캐피탈 정의	3
2.2 벤처캐피탈리스트의 투자 요인 선행연구	3
2.3 결정 트리 기반 앙상블 모델 기법	6
2.3.1 랜덤 포레스트(Random Forest)	6
2.3.2 Extreme Gradient Boosting(XGBoost)	6
2.3.3 Light Gradient Boosting Machine(LightGBM)	7
2.3.4 Categorical Boosting(CatBoost)	7
2.4 설명가능한 인공지능(XAI) : SHAP	8
제3장 연구 방법	10
3.1 데이터	10
3.2 분석 절차	13
제4장 예측 모델링 및 평가	15
4.1 예측 모델 개발	15
4.2 모델 성능 평가	18
4.3 모델 해석	20
제5장 토의 및 시사점	25
5.1 결과 토의 및 연구 한계	25
5.2 연구의 학술적 · 실무적 의의	26
참고 문헌	27
영문 요약	30

그림 차례

[그림 1] - 예측 모델링 과정	14
[그림 2] - LGBM model Feature Importance	20
[그림 3] - LGBM model을 설명하는 각 변수들의 SHAP value	21

표 차례

[표1] - VCs의 투자 요인 선행 연구	5
[표2] - 설문조사 개요	11
[표3] - 선행연구를 이용한 설문지 문항 선정 내용	12
[표4] - 데이터 클래스 요약	15
[표5] - 사용 변수	16
[표6] - Model Performance Metrics	19
[표7] - LGBM 모델의 SHAP value 해석	22

국 문 요 약

벤처캐피탈리스트 투자 요인분석 및 예측 모델링

국내 벤처기업 수는 꾸준히 증가하여 2020년 기준 39,511개사로 역대 최대 규모로 증가하였으며, 미국, 이스라엘, 중국에 이어 GDP 대비 벤처투자 비중 세계 4위에 진입하는 등 벤처기업과 벤처기업 투자에 대한 관심이 많아지고 있다. 그러나 벤처투자는 기밀하게 이뤄지는 특성으로 인해 연구자료가 부족하여 벤처투자요인에 대한 연구가 미비한 실정이다. 또한, 선행연구는 벤처투자 결정요인을 연구하기 위한 표본이 적다는 한계점을 지적한다. 따라서 본 연구에서는 다수의 데이터를 사용하여 벤처 투자 요인을 분석하고, 벤처 투자를 받을 기업을 예측하는 모델 개발하는 것을 연구 목적으로 한다.

LightGBM 모델이 f1-score 85.02%로 가장 우수한 성능을 보였으며, SHAP 기법을 사용하여 예측 모델에서 중요하게 사용된 설명변수들을 확인하여 투자 요인을 분석하였다. 그 결과 창업자 특성으로는 창업 당시 창업자의 최종학력이 높을수록, 단독 창업보다는 공동창업의 경우에 SHAP value가 양(+)의 영향력을 가졌으며, 제품/서비스 특성으로는 주력제품의 국내 시장 점유율이 높을수록, 해외 매출 비중이 높을수록, 사업내용이 IT 관련 분야인 경우에 SHAP value가 양(+)의 영향을 끼치는 것으로 나타났다. 반면, 재무적 특성에서는 재무 상황이 좋지 않더라도 SHAP value가 양(+)의 영향을 미치는 것으로 보아 투자할 때 크게 고려되는 상황이 아니라고 유추할 수 있다.

본 연구를 통해 VCs의 투자 요인을 파악해볼 수 있으며, 투자 유치를 원하는 벤처기업가들에게 보완할 정보를 제공할 수 있다. 벤처캐피탈에는 투자 결정에 대한 보조 수단을 제공할 수 있다는 점에 실무적 의의를 가진다.

주제어: 벤처기업 투자 예측, 벤처캐피탈리스트 투자 요인 분석, CV, LightGBM, SHAP

제 1장 서론

4차산업혁명이 글로벌 경제의 패러다임을 변화시키고 있다. 잠재적 경쟁국들이 우리나라 기술력을 크게 위협하고 있는 상황에서도 벤처기업은 도전적이고 혁신적인 기술로 글로벌 진출 및 일자리 창출 등 국민 경제에 기여하고 있다[1]. 국내 벤처기업 수는 꾸준히 증가하여 2020년 기준 3만 9천개사로 역대 최대 규모로 증가했으며, 이에 따라 국가별 벤처투자 비교지표인 GDP 대비 벤처투자 비중이 증가하여 미국, 이스라엘, 중국에 이어 세계 4위권에 진입했다[2].

벤처캐피탈(Venture Capital, VC)은 높은 위험성을 가지고 있지만 고수익의 가능성이 상대적으로 높은 벤처기업 또는 중소기업에 투자하는 사모투자의 한 형태를 가리킨다[3]. 벤처캐피탈리스트(Venture Capitalist, VCs)는 일정 기간 내에 높은 성장 잠재력을 지닌 기업에 투자할 자본을 제공할 수 있는 자본투자자이다[4]. VCs의 투자목적은 고성장, 고수익의 잠재력을 지닌 기업에 자금을 지원하여 VC기업 및 투자자 모두에게 높은 수익을 제공함으로써 투자 가치를 창출하는 것이다[5]. VCs는 성장 가능성이 높은 새롭고 혁신적인 벤처기업을 식별하여 대상기업에 자금을 투자하는 가장 핵심적인 역할을 하며[6], VCs가 지원하는 벤처기업은 VCs의 지원을 받지 않는 벤처기업보다 일반적으로 성공할 가능성이 높은 것으로 나타난다[3].

따라서 VCs의 벤처기업 투자 결정은 국가경제 발전과 벤처기업 성장에 중요한 역할을 한다. 해외에서는 VCs의 의사결정에 관한 연구가 1970년대 이후 활발하게 이루어져 왔으나, 국내에서는 VCs 활동에 대한 접근성의 어려움, VC기업의 활동에 대한 기밀성, 연구 자료 부족 등의 한계들로 연구가 미비한 실정이다[7]. 의사결정 기준[8], 의사결정 모형[9], VCs의 경험과 의사결정[10][11]에 대한 연구와 최근 연구로는 투자결정에 미치는 요인에 대한 연구가 있으나 많은 연구가 이뤄지고 있지 않다[7][12]. 또한, 국내에서 이뤄진 기존의 연구들은 연구 표본 한계로 인해 최대 200명 내외의 적은 표본을 대상으로 이루어졌다는 한계를 갖는다. [13]에서는 투자 결정 요인에 대한 사례가

적으며, 정량적 모형에 대한 필요성을 언급하였고, [7]에서는 한국의 벤처캐피탈 산업은 성장기에 진입하였으나 데이터 부족으로 벤처캐피탈 관련 연구가 부족한 실정이라고 언급한다.

따라서 본 연구는 투자를 받은 벤처기업과 투자를 받지 않은 벤처기업의 설문조사 6,992개를 통해 VCs의 투자 결정 요인을 분석하여 데이터 부족에 대한 선행연구 한계를 보완한다. 기존의 정량적 연구가 진행되지 못했던 선행연구의 한계도 보완한다. 또한, 벤처 투자 예측 모델을 개발하여 예측 모델을 통해 벤처 투자 중요 요인을 분석한다. SHAP 기법을 사용하여 투자 예측 모델에서 중요한 변수로 사용된 요인들을 분석하고 기계학습의 black box 문제점을 보완한다. 즉, 본 연구 결과를 통해 VCs의 투자 요인을 파악해볼 수 있으며, 투자 유치를 원하는 벤처기업가들에게 보완할 정보를 제공할 수 있다. 마지막으로 벤처캐피탈에는 투자 결정에 대한 보조 수단을 제공할 수 있다는 점에 실무적 의의를 가진다.

제 2장 이론적 배경

2.1. 벤처기업 및 벤처캐피탈 정의

우리나라에서는 ‘벤처기업’을 벤처기업 육성에 관한 특별조치법(이하 “벤처기업법”)에서 정의하고 있다. ‘벤처기업’은 중소기업으로써 벤처투자를 받았거나, 기술평가 보증 대출을 받은 기업이나, 연구개발 비중이 높은 기업 등의 요건을 갖춘 기업이다. 또한, 벤처기업은 신생기업이라는 인식때문에 창업기업의 개념과 연결되기도 한다. 우리나라에서 창업기업은 중소기업창업지원법(이하 “창업지원법”)에 ‘업력 7년이내인 중소기업’으로 정의되어 있으며, 이중 ‘업력 3년 이내인 중소기업’을 초기창업기업으로 따로 구분한다[7].

벤처캐피탈(Venture capital)은 1929년 미국 대공황 이후 자금난을 겪는 기업에 투자하는 자본의 의미로 쓰이기 시작했다. 현대적인 벤처캐피탈의 시작은 1946년 미국에서 설립된 ARDC(American Research and Development Corporation)로 알려져 있으며, 자금을 모집하여 신기술 기업에 투자한 최초의 사례이다. 최근에는 벤처캐피탈이 사모펀드 형태를 주로 활용하면서 오늘날 벤처캐피탈이라는 용어는 ‘스타트업에 투자하는 사모펀드’로 인식되고 있다[14].

2.2. 벤처캐피탈리스트의 투자 요인 선행연구

VCs의 의사결정 기준은 여러 학자들의 주요 연구 주제로서 1970년대 이후 지속적으로 관심을 받으며 다양한 관점에서 실증적 연구가 수행되어 왔다[7]. 선행 연구를 살펴보면, 연구 표본 수, 자료수집 방법, 연구 방법, 통계 분석 방법 등이 다양하게 사용되었다. 연구의 표본의 측면에서는 초기 연구는 주로 5명 내외에서 수십 명 내외의 비교적 적은 표본의 VCs를 대상으로 연구되었으며, 최근 140명의 벤처캐피탈리스트 표본이 최대 표본 연구이다. 그러나, 다른 일반적인 영역의 연구에서 사용되는 표본의 수와 비교하면 상대적으로 적은 표본이다. 연구의 자료 수집은 설문조사가 주를 이뤘고, 연구 방법에는 회귀분석과 AHP가 주로 사용되었다 <표1 참조>.

저자 (연도)	데이터 수집 방법	연구 표본	연구 방법론	주요 요인
반성식 & 송경모 (2004)[9]	설문조사	53 VCs	Regression & ANOVA	기업가와 팀 능력, 제품/서비스의 우수성, 시장과 경쟁기업의 현황, 성공시 잠재적 수익률
Franke et al(2008) [15]	컨조인트 분석	51 VCs	Probit regression	창업팀의 산업경험
구경철 외(2008) [11]	설문조사	20 VCs	AHP	기업의 성장성 및 수익성, 경영진 및 직원의 자질, 시장변화 대응력, 시장상황, 마케팅 능력, 제품 차별성, 경쟁상황, 기술력 등
차명수 (2009)[8]	설문조사	32 VCs	Regression & judgment analysis	창업자/팀의 능력, 잠재적 수익률, 시장 상황
Monika Dhodhak & Sharma (2015)[16]	설문 조사	108 VCs	Exploratory factor analysis & AHP	창업자의 특성, 재무적 고려사항, 제품/서비스 특성
한재우 외(2016) [17]	설문 조사	65 VCs	AHP	창업가(팀), 제품/서비스, 시장, 네트워크, 재무, 제품/서비스와 모기업 연관성, ICT 업종 등
조세근 & 한주희 (2018)[13]	사례 연구	17 VCs	AHP	시장규모, 진입장벽, 시장접근성, 경영자 기술/경영 경험 및 추천/평판 등
김승찬& 김홍근 (2019)[18]	설문 조사	140 VCs	AHP	재무 특성, 제품 서비스 특성, 시장 특성, 경영자 특성
구중회 외(2019) [7]	문헌 연구	-	AHP	창업자의 특성, 제품/서비스 특성, 시장 특성, 재무적 특성, 네트워크

표 1. VCs의 투자 요인 선행 연구

2.3. 결정 트리 기반 앙상블 모델 기법

2.3.1. 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest) 알고리즘은 다수의 의사결정나무 알고리즘을 랜덤으로 조합하여 예측하는 기계 학습 방법이다[19]. 의사결정나무(Decision Tree) 알고리즘은 데이터가 지닌 관계와 규칙을 파악하고 모형화하는 분석기법으로 선형성, 정규성, 등분산성 가정이 필요 없다[20]. 랜덤 포레스트는 각 의사결정나무를 만들 때 무작위로 학습 데이터와 독립 변수를 선택하여 예측을 진행하며, 개별적인 정확도는 떨어질 수 있으나 모든 의사결정나무를 종합하여 예측을 수행하므로 정확도와 안정성이 높아진다는 장점을 지닌다[21]. 즉, 랜덤 포레스트는 무작위로 독립 변수와 데이터를 무작위로 선정하는 의사결정나무 알고리즘을 N개 만들어 다수결의 원칙으로 가장 많이 도출되는 값 또는 평균값을 예측값으로 사용한다.

2.3.2. Extreme Gradient Boosting(XGBoost)

부스팅(Boosting) 기법은 초기 샘플 데이터를 이용하여 약한 학습기를 만들고 학습 결과의 오차를 줄이는 방향으로 신규 학습기를 반복적으로 추가하는 앙상블 기술이다[22]. 특히 ‘Gradient Boosting’은 이전 학습기의 잔차를 예측하는 신규 모델을 계속 추가해 나아가는 알고리즘으로[23], 학습 속도가 느리며 과대 적합이 발생하는 단점이 있다. 이를 보완한 알고리즘이 XGBoost이다. XGBoost는 의사결정나무 기반 기계 학습 알고리즘으로 병렬 처리, 결측치 처리 및 규제 등을 통해 훈련 손실을 최소화하면서 과적합을 방지하는 최

적화된 모델을 생성한다[24].

2.3.3. Light Gradient Boosting Machine(LightGBM)

LightGBM은 2016년에 마이크로소프트MSRA(Microsoft Research Lab Asia)에서 설계한 빠르고 효율적인 GBDT(Gradient Boosting Decision Tree) 기반 알고리즘이다[25]. 기존 GBDT 기반 알고리즘은 모두 가능한 분할점에 대해 정보 이득(Information gain)을 평가하기 위해 데이터 전부를 스캔해야 했기 때문에 대용량의 고차원 데이터에서는 성능을 제대로 발휘하지 못한다는 한계가 존재한다. LightGBM은 GOSS(Gradient-based One-Side Sampling)와 EFB(Exclusive Feature Bundling) 기법을 도입하여 GBDT 기존의 한계를 보완하였다. GOSS는 기울기가 큰 데이터 속성은 유지하고 기울기가 작은 데이터 속성은 일정 확률로 랜덤하게 제거한다. EFB는 변수 개수를 줄이기 위해 희소한 변수 공간의 특성에 따라 상호배타적인 변수들을 묶는다[26]. 즉, LightGBM은 이러한 기법들을 사용하여 훈련 속도를 빠르게 하고, 메모리 사용량 감소 등의 장점이 있다.

2.3.4. Categorical Boosting(CatBoost)

CatBoost 알고리즘은 범주형 변수 전처리와 과적합(Overfitting) 문제 해결에 초점을 둔 순서형 부스팅(Ordered Boosting) 기법이다[27]. 모든 잔여 오차를 순차적으로 학습하는 기존의 부스팅 모델과는 달리 순서형 부스팅은 일부 데이터로 잔여 오차를 계산하여 모델을 만든다. 이후 해당 모델을 통해 남은 데이터의 잔여 오차를 계산하는 기법이다. 또한, 순서형 부스팅에 무작

위 순열(Random Permutation)을 통해 데이터 순서를 섞어줌으로써 과적합을 방지한다. CatBoost 알고리즘은 동일한 정보 획득(Information gain)을 가진 변수들을 하나로 묶는 변수 조합을 통해 훈련 속도를 향상시킨다. 또한, 최적의 초매개변수(Hyper parameter)를 찾기 위해 Grid Search나 Randomized Search를 사용하는 다른 앙상블 알고리즘과 달리 초기 초매개변수 값이 최적화되어있어서 매개변수(Parameter) 조정 절차가 요구되지 않는다는 특징을 가진다.

2.4. 설명 가능한 인공지능(XAI) : SHAP

설명 가능한 인공지능(explainable Artificial Intelligence, XAI)은 인공지능에 의해 예측된 결과를 설명하여, 사용자가 결과를 이해할 수 있도록 돕는다. 기계학습 모델의 예측에 영향을 미치는 주요 요소들을 찾아 내어 예측 결과를 사람이 이해할 수 있도록 직관적으로 설명한다[28].

SHAP(SHapley Additive exPlanations)은 기계학습 모델의 결과를 설명하기 위해 게임이론의 샐플리 값(Shapley Values)을 사용한 알고리즘이다. 샐플리 값은 한 개의 변수 중요도를 알기 위해 다른 변수들의 조합을 구성한 후 해당 변수의 유무에 따른 평균적인 변화를 통해 얻어낸 값을 나타낸다[29]. 이를 통해 설명 변수가 예측 모델 결과 값의 공헌도를 확인할 수 있게 한다.

기존의 변수 중요도(Feature Importance) 기법은 순열(Permutation) 방법을 사용해서 변수가 모델에 미치는 영향을 측정한다. 이 방법은 계산 속도가 빠르다는 장점이 있지만, 변수들이 서로 의존적일 때는 결과가 왜곡될 수 있고, 음(-)의 영향력은 계산하지 못한다는 단점이 있다. 따라서 실제 영향력보다 특정 변수의 가치가 높게 표현될 수 있다. 반면, SHAP 기법은 변수들간에 서로 영향을 미치는 가능성을 고려하며, 음(-)의 영향력을 계산할 수 있다. 속도가 느리다는 단점이 있지만, 변수 중요도 기법보다 정확한 영향력을 측정할 수 있다[30].

제 3장 연구 방법

3.1. 데이터

본 연구는 중소벤처기업부에서 2018년부터 2020년 까지 매년 시행한 ‘벤처기업정밀실태조사’ 데이터를 이용한다. ‘벤처기업정밀실태조사’는 벤처기업에 대한 정확한 실상을 파악하고, 벤처확인기업의 변화를 살피는 것을 목적으로 한다(중소벤처기업부, 2020). 설문조사는 ‘①기술평가대출기업(중소벤처진흥공단), ②기술평가보증기업(기술보증기금), ③벤처투자기업(창업투자회사 등이 자본금 10% 이상 투자), ④연구개발기업’을 대상으로 이뤄졌다.

2020년에 실시된 설문조사는 2019년 12월말 기준 벤처확인기업 36,907개사 중 휴폐업 404개사를 제외한 36,503개사를 대상으로, 모집단의 대표성 확보를 위해 최종적으로 수집된 데이터는 2,500개의 유효표본이다. 2019년에 실시된 설문조사 대상은 2018년 12월말 기준 휴폐업 기업을 제외한 벤처확인기업 36,065개이다. 모집단의 대표성 확보를 위해 최종적으로 수집된 데이터는 2,052개의 유효표본이다. 2018년 설문조사는 예비 벤처기업을 제외하고, 총 2,059개의 유효표본을 사용한다. 이 중 법인 데이터만 활용하여 2020년은 총 2,200개, 2019년은 총 1,733개, 2018년은 총 2,059개의 데이터를 사용한다 <표2 참조>.

구 분	2020년	2019년	2018년
조사 대상	2019년 12월말 기준, 벤처확인기업 36,503개	2018년 12월말 기준, 벤처확인기업 36,065개	2017년 12월말 기준, 벤처확인기업 35,187개
조사 기간	2020년 8월 10일 ~ 10월 23일 (약 3개월)	2019년 7월 24일 ~ 10월 18일(약 3개월)	2018년 8월 8일 ~ 10월 19일(약 2.5개월)
유효 표본	총 2,500개	총 2,052개	총 2,059개
사용 표본	총 2,200개 법인	총 1,733개 법인	총 2,059개 법인

표 2. 설문조사 개요

본 연구 데이터의 종속변수는 벤처캐피탈(투자조합포함)로부터 투자를 받은 경험 여부로써 받은 경험이 있으면 ‘1’, 없으면 ‘0’으로 사용하였다. 사용한 설명변수는 선행연구에서 벤처캐피탈리스트가 고려하는 의사 결정 요인에 따라 3가지 공통 요인을 도출하고, 이를 바탕으로 설문지 문항을 선정하였다. 최종적으로 사용된 설문지 문항은 <표 3>과 같다.

선행연구	공통 요인	설문지 문항
Franke et al.(2008)[15]	창업자/팀의 특성	창업자 특성
차명수(2009)[8]		
Monika Dhochak & Sharma(2015)[16]		
한재우 외(2016)[17]		
구중희 외(2019)[7]		
반성식 & 송경모(2004)[9]	제품/서비스의 특성	기술 및 산업재산권 현황
구경철 외(2008)[11]		
Monika Dhochak & Sharma(2015)[16]		
한재우 외(2016)[17]		
구중희 외(2019)[7]		
Monika Dhochak & Sharma(2015)[16]	재무적 특성	재무 및 자금 현황
김승찬 & 김홍근 (2019)[18]		

표 3. 선행연구를 이용한 설문지 문항 선정 내용

3.2. 분석 절차

본 연구는 벤처캐피탈(투자조합포함)로부터 투자를 받을 기업을 예측하는 모델을 개발하고 개발된 예측 모델의 변수 영향력을 통해 투자 요인을 분석하는 것이 연구 목적이다. 데이터는 2018년부터 2020년까지 3개 년의 ‘벤처기업 정밀실태조사’에서 진행된 설문조사 데이터를 사용한다. 선행연구 조사를 통해 벤처캐피탈리스트의 투자 결정 요인 3가지를 활용하여 설명변수로 사용한다. 이후 이상치, 결측치, 파생변수 생성 등의 데이터를 전처리 진행한다. 데이터 클래스가 불균형하기 때문에 다양한 Over-Sampling 기법들을 사용하여 소수 범주의 데이터 개수를 증가시켜 비율을 맞춘다. 이후 트리 기반 앙상블 모델인 Random Forest, XGBoost, LightGBM, CatBoost을 사용하여 예측 모델을 개발한다. 그리고 K-fold Cross Validation을 통해 앙상블 모델들의 초매개변수를 조정하여 최적화를 진행한다. 모델 성능 평가 지표는 불균형 데이터를 평가하기에 적합한 Recall, F1-score, G-mean을 사용하여 종합적으로 가장 우수한 성능의 모델을 선정한다. 마지막으로 머신러닝 모델의 Blackbox 문제를 보완하고 설명 가능한 모델을 만들기 위해 주요 변수들을 규명하고 영향력을 측정하는 SHAP 기법을 적용한다. 이는 각 변수가 최종 정확도에 영향을 미치는 정도를 측정하여 머신러닝 모델의 설명력을 보충하는 데 사용된다 [31][32]. SHAP의 결과를 통해 예측 모델에서 중요하게 사용된 변수들의 영향력을 파악하고, 벤처기업 투자에 중요 요인을 분석한다. 본 연구의 전체 분석 과정은 <그림1>과 같다.

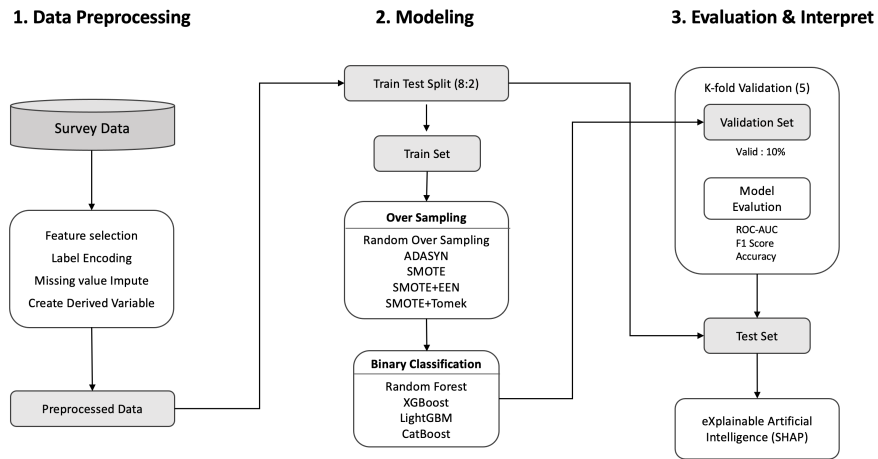


그림 1. 예측 모델링 과정

제 4장 예측 모델링 및 평가

4.1. 예측 모델 개발

전체 데이터는 총 5,992개이고, 설명 변수는 27개이다<표 5>. 예측 모델을 구축을 위해 무작위로 8:2 비율의 훈련 데이터와 테스트 데이터를 분리했다. 데이터는 약 9:1로 Target 값이 Imbalanced Data이다<표 4>. 따라서 Over Sampling 기법들을 사용하여 1:1의 비율로 맞췄다. Over-Sampling 기법에는 Random Over Sampling, ADASYN, SMOTE를 사용하였고, Hybrid-Sampling 기법에는 SMOTEEEN, SMOTETomek을 적용했다. 이를 결정트리기반 앙상블 모델인 Random Forest, XGBoost, LightGBM, CatBoost의 예측 모델을 개발하고 K-중첩 교차검증(K-fold Cross Validation)을 통해 모델을 평가하고 검증하였다. 모델 구축은 Python 모듈인 Scikit-Learn[33]을 사용했다.

구 분	2020년	2019년	2018년	합계
0	2,094	1,450	1,864	5,408
1	106	283	195	584
합계	2,200	1,733	2,059	5,992

표 4. 데이터 클래스 요약

구분	대영역	소영역	변수명	변수 내용
1	창업자 특성		B2_3	창업 당시 창업자 최종학력
2			B2_3_1	창업 당시 창업자 전공
3			B2_4	창업 직전 근무지
4			B2_6_1	창업방식
5			B2_6_2	단독창업 여부
6			B2_7	창업 당시 목표시장
7			B2_8	창업 당시 창업자 실무경험 년수(년)
8	재무 및 자금 현황	매출 구조 (%)	D4_1	매출구조(%)_B2B(대기업 또는 대기업그룹사 소속사)
9			D4_2	매출구조(%)_B2B(대기업에 납품하는 1·2차 벤더)
10			D4_3	매출구조(%)_B2B(중소기업 및 벤처기업, 중견기업)
11			D4_4	매출구조(%)_B2C(소비자 매출)
12			D4_5	매출구조(%)_B2G(정부 포함 공공부문)
13			D4_6	매출구조(%)_해외
14		재무 건전성	PR	수익성(%)
15			BIS	자기자본비율(%)
16			ROE	자기자본이익율(%)
17	기술 및 산업재산권 현황		E1	IT 관련 기업
18			E2_1	특허권
19			E2_2	실용신안권
20			E2_3	디자인권
21			E2_4	상표권

22		E2_5	해외
23		E4	제품 및 서비스 구조
24		E5	국내 유일기술 보유여부
25		E7	기술연구소나 연구개발 전담부서 보유여부
26		G3_1	주력제품 국내 시장점유율(%)
27		G3_2	주력제품 해외 시장점유율(%)
28	Target	I3	벤처캐피탈(투자조합포함) 투자 경험여부

표 5. 사용 변수

4.2. 모델 성능 평가

본 모델은 벤처투자를 받을 기업을 예측하는 모델을 개발하고자 하였다. 투자 받은 기업은 '1', 투자 받지 않은 기업은 '0'으로 분류하는 문제이며 Imbalanced Data이기 때문에 모델 분류 성능은 Recall, F1 Score, G-mean을 사용하여 종합적으로 평가하였다. 모델별 예측 성능 비교는 <표 6>과 같다. 3가지 성능 지표에서 균형있게 높은 성능을 보인 모델을 우수한 모델로 선정하였다. 따라서, RandomForest와 XGB 모델은 SMOTEEEN Over-Sampling에서 가장 우수한 성능을 보였으며, LGBM 모델은 Random Over Sampling에서 CatB 모델은 SMOTE에서 가장 좋은 성능을 보였다. 이 중 Recall과 F1-score에서 가장 우수한 성능을 나타낸 LGBM 모델을 최종 모델로 선정하였다.

Model	Metrics	Over Sampling			Hybrid Sampling	
		ROS	ADASYN	SMOTE	SMOTE+ENN	SMOTE+Tomek
RF	Recall	58.61	63.67	63.20	71.22	62.48
	F1_score	61.60	63.83	63.35	64.86	62.87
	G-mean	49.08	59.08	58.45	71.22	57.33
XGB	Recall	64.56	62.99	62.84	69.76	62.84
	F1_score	65.25	63.97	64.20	63.74	64.20
	G-mean	60.33	57.71	57.29	68.57	57.29
LGBM	Recall	82.81	73.72	74.06	68.22	72.31
	F1_score	85.02	78.85	79.01	74.67	77.73
	G-mean	70.59	75.13	72.26	74.13	72.08
CatB	Recall	71.64	78.14	79.05	72.81	66.21
	F1_score	77.25	81.45	82.24	78.01	59.27
	G-mean	72.84	62.74	65.97	68.81	65.11

표 6 Model Performance Metrics

4.3. 모델 해석

성과 지표가 종합적으로 높은 점수를 가진 Random Over Sampling을 적용한 LGBM 모델을 SHAP 기법을 사용하여 모델을 설명한다. SHAP value 시각화는 모델에 미치는 변수 중요도에 따라 20개의 변수가 출력된다. 변수 중요도 그림은 Python의 Feature Importance 와 유사하지만<그림 2>, SHAP 기법은 각 변수가 모델에 긍정적인 영향을 미치는지, 부정적인 영향을 미치는지까지 확인이 가능하다<그림 3>. 따라서, LGBM 모델을 SHAP 기법을 통해 모델에 미치는 변수들의 긍/부정 영향력을 아래 <표 6>과 같이 해석하였다.

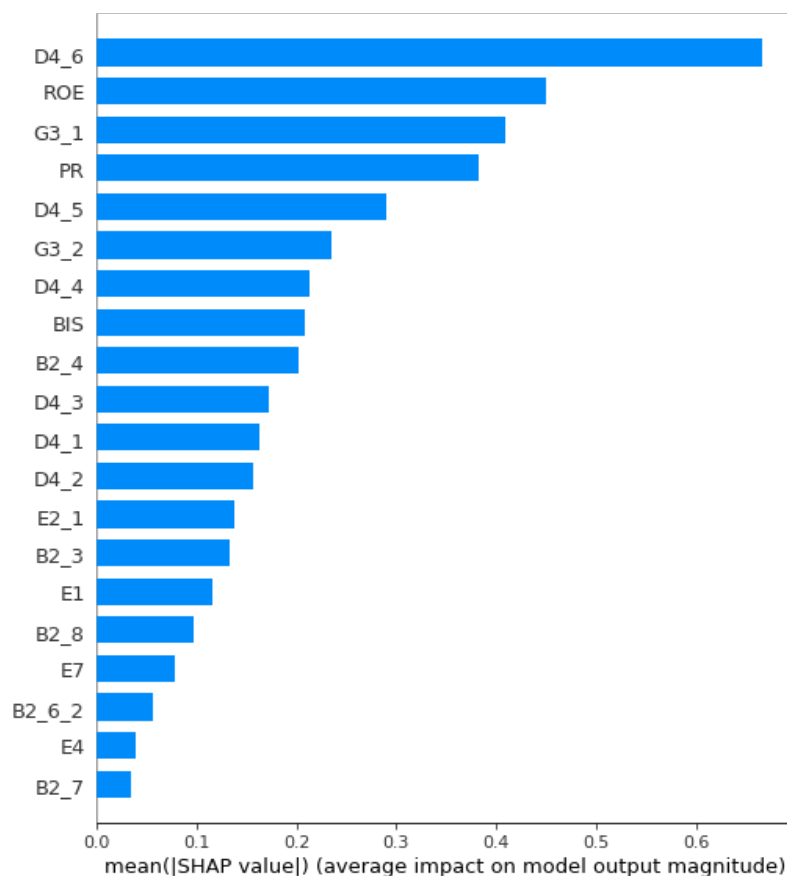


그림 2. LGBM Model Feature Importance

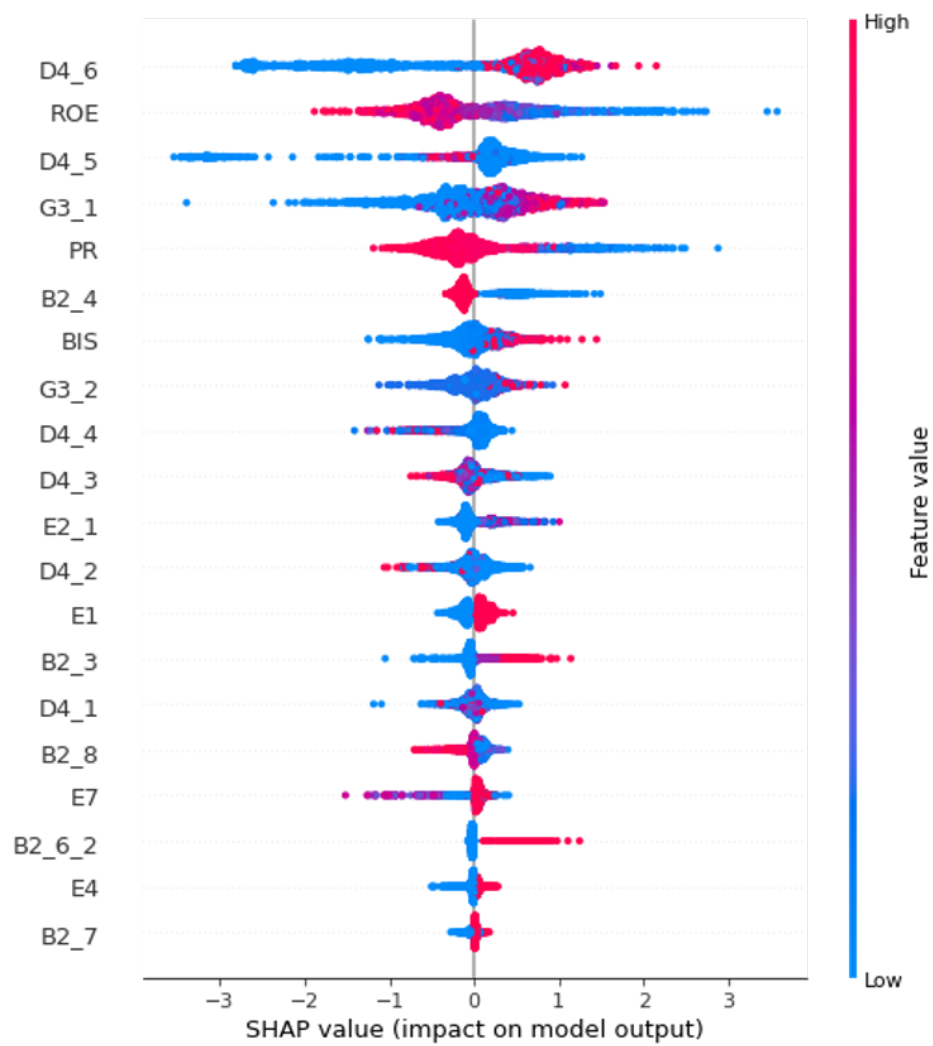


그림 3. LGBM model을 설명하는 각 변수들의 SHAP value

구 분	변수명	변수 설명	SHAP Value 해석
1	D4_6	매출구조(%)_해외	매출이 해외에서 발생하는 비중이 많을수록 SHAP value에 양(+)의 영향력을 갖는다. 즉, 전체 매출 비중 중 해외 매출 비중이 높을수록 벤처 투자를 받을 가능성이 높다.
2	ROE	자기자본이익율(%) =당기순이익/자본 총계 * 100	자기자본이익율이 높을수록 SHAP value에 음(-)의 영향력을 갖는다. 즉, 자기자본이익율이 낮으면 벤처 투자를 받을 가능성이 높아진다는 것인데, 이는 벤처투자의 특성상 신생기업을 대상으로 투자를 하고, 미래 성장성을 고려하기 때문에 재무관련하여서는 투자시 덜 고려되고 있다고 해석이 가능하다.
3	D4_5	매출구조(%)_B2G (정부 포함 공공부문)	전체 매출 중에 정부를 포함하여 공공부문 매출 비중을 나타낸 변수이다. 해당 변수 값의 낮은 것이 SHAP value의 양(+)과 음(-)의 영향력 모두 끼친다. 따라서 해당 변수가 모델에 중요한 요인이지만 긍정과 부정이 명확하게 해석되지는 않는 한계가 있다.
4	G3_1	주력제품 국내 시장점유율(%)	주력 제품의 국내 시장 점유율을 최대 100% 기준으로 나타낸 것으로 해당 값이 높을수록 SHAP value에 양(+)의 영향력을 갖는다. 즉, 주력제품의 국내 시장 점유율이 높을수록 벤처 투자를 받을 가능성이 높다.
5	PR	수익성(%) = 당기순이익/ 매출액 * 100	수익성이 낮을수록 SHAP value에 양(+)의 영향력을 갖는다. 수익성이 좋지 않은 기업이 투자를 받은 경우가 많다는 것을 유추해볼 수 있다.
6	B2_4	창업 직전 근무지	창업 직전 근무지는 변수 값이 작을수록 SHAP value에 양(+)의 영향력을 갖는다. 해당 변수는 1~10까지 라벨인코딩이 되어 있고, 1은 교수, 2

			는 학생, 3은 공공부문 연구소 근무지이다. 즉, 교수와 학생, 연구소에 근무했을 경우 벤처 투자 받을 경우가 높다고 유추할 수 있다.
7	BIS	자기자본비율(%) = 자본/자산 * 100	자산 중에 자본이 차지하는 비율에 대한 변수이다. 자기자본비율이 높을수록 SHAP value에 양(+)의 영향력을 갖는다.
8	G3_2	주력제품 해외 시장점유율(%)	주력 제품의 해외 시장 점유율이 높을수록 SHAP value에 양(+)의 영향력을 갖는다. 즉, 주력제품의 해외 시장 점유율이 높을수록 투자받을 확률이 높다.
9	D4_4	매출구조(%)_B2C (소비자)	전체 매출 중 소비자 매출 비중이 적을수록 SHAP value에 양(+)의 영향력을 갖는다. 다만, SHAP value에 음(-)의 영향력을 끼치기도 하였기 때문에 해당 변수는 명확하게 해석하기에는 한계가 있다.
10	D4_3	매출구조(%)_B2B (중소/중견/벤처기업)	전체 매출 중 중소기업, 중견기업, 벤처기업에 판매하는 비중이 높은 기업은 SHAP value에 음(-)의 영향력을 갖는다. 즉, 중소기업, 중견기업, 벤처기업에 판매하는 비중이 높은 기업은 벤처투자를 받기 어려울 수 있다고 유추할 수 있다.
11	E2_1	특허권	특허권 개수가 많을수록 SHAP value에 양(+)의 영향력을 갖는다. 즉, 특허권 수가 많을수록 투자 받는데 이득이 된다고 해석할 수 있다.
12	D4_2	매출구조(%)_B2B (대기업에 납품하는 1·2차 벤더)	대기업에 납품하는 1,2차 벤더를 대상으로 한 매출 비중이 낮다면 SHAP value의 양(+)과 음(-)의 영향력을 모두 갖는다. 해당 변수는 명확하게 해석하기에 한계가 있다.
13	E1	IT 관련 기업	사업내용이 4차 산업과 관련 있거나, IT 관련 기업의 경우 SHAP value에 양(+)의 영향력을 갖는

			것을 볼 때, 4차 산업 또는 IT 관련 기업은 투자 받을 확률이 높다고 해석할 수 있다.
14	B2_3	창업 당시 창업자 최종 학력	최종학력은 높을수록 SHAP value에 양(+)의 영향력을 갖는다고 명확하게 보여진다. 즉, 고학력자일수록 투자 받을 확률이 높아진다고 해석해볼 수 있다.
15	D4_1	매출구조(%)_B2B (대기업 또는 대기업그룹 소속사)	전반적으로 변수값이 작음에 따라 SHAP value가 정해진다. 해당 변수는 어떠한 값이 명확하게 SHAP value에 영향을 끼쳤는지 확인이 어렵다,
16	B2_8	창업 당시 창업자 실무 경험 년수(년)	창업 당시 실무 경험 년수가 적을수록 SHAP value에 양(+)의 영향력을 갖는다.
17	E7	기술연구소나 연구개발 전담부서 보유여부	해당 변수는 값에 따라 양(+)과 음(-)의 영향 모두 끼치고 있다. 즉, 어떠한 값이 명확하게 SHAP value에 영향을 끼쳤는지 확인이 어렵다,
18	B2_6_2	단독창업 여부	해당 변수는 혼자 창업할 경우가 0, 공동창업할 경우가 1로 라벨링 되어 있는데, 공동창업의 경우 SHAP value에 양(+)의 영향력을 갖는다. 즉, 공동창업의 경우 투자 받을 가능성이 단독 창업에 비해 높다고 해석할 수 있다.
19	E4	제품 및 서비스 구조	제품 및 서비스가 여러개일수록 SHAP value에 양(+)의 영향력을 갖는다.
20	B2_7	창업 당시 목표시장	해당 변수의 값은 창업 당시 목표시장이 '국내+해외'는 3, 해외는 2, 국내는 1로 라벨링 되어 있다. 즉, 목표 시장에 해외 시장이 포함된 경우 투자 받을 가능성이 있다고 유추가 가능하다.

표 7. LGBM 모델의 SHAP value 해석

제 5장 토의 및 시사점

5.1. 결과 토의 및 연구 한계

본 연구는 벤처기업실태조사 설문 데이터를 사용하여 벤처 투자를 받을 기업을 예측하는 모델을 개발하고, SHAP 기법을 사용하여 모델에서 중요 투자요인을 분석하였다. 27개의 설명변수 중 중요도가 높은 순으로 20개의 변수를 확인하였다. SHAP의 결과가 분명하고 양(+)의 영향을 미치는 변수들을 확인해 봤을 때, 창업자 특성으로는 창업 당시 창업자의 최종학력이 높을수록, 단독 창업보다는 공동창업일 때 투자 중요 요인으로 확인되었다. 재무적으로는 재정상황이 건전하지 않더라도 투자를 받는다는 것을 유추할 수 있다. 이는 [34]가 벤처기업의 재무적인 특성은 중요한 의사결정 기준이 되지 않는다는 연구를 뒷받침한다. 제품/서비스 특징으로는 특허권 개수가 많을수록, 주력 제품이 국내에서 높은 시장 점유율을 갖을수록 양(+)의 영향을 미친다. 이는 [9]가 시장 경쟁자 수가 많을수록, 경쟁의 강도가 높을수록 벤처기업 성공가능성이 낮다는 결과와 유사하다. 또한, IT 관련 사업일 경우 SHAP value에 양(+)의 영향력을 미쳤는데, 이는 [13]이 CVC 투자유치를 위해서는 ICT서비스 분야로 창업하는 것이 유리하다는 연구결과와 연관이 있다. 선행 연구와 다른 결과로는 매출 중 해외에서 차지하는 매출비중이 높거나, 주력 제품의 해외시장이 높은 경우 양(+)의 영향을 미치는 것으로 확인되었다. 한국의 국내총생산(GDP)에서 수출이 차지하는 비중이 40% 정도 되며, 세계 평균 30% 보다 10% 높으며, 수출입총액을 GDP로 나눈 무역의존도는 60%가 넘는다[35]. 따라서 기업의 글로벌 성장성은 중요 요인으로 참작될 가능성이 있기때문으로 그 결과를 해석해볼 수 있다.

본 연구의 한계점으로 SHAP 기법을 사용하여 예측 모델에서 사용된 중요한 설명변수를 바탕으로 벤처투자 중요 요인을 분석했으나, 변수가 SHAP value에

양(+)과 음(-)의 영향력을 모두 끼치는 경우 명확하게 설명하기 어렵다는 한계점을 가진다.

5.2. 연구의 학술적 · 실무적 의의

본 연구가 갖는 학술적 시사점은 다음과 같다. 첫째, 투자받은 기업 584개와 투자 받지 않은 기업의 데이터 5,408개를 사용하여 선행연구에서 갖는 투자 결정 요인 연구의 적은 표본 한계를 극복하였다는 점이다. 선행연구에서는 벤처캐피탈리스트를 대상으로 설문조사나 인터뷰를 통해 투자 요인을 분석하였기 때문에 대량의 표본을 모집하기가 어렵다는 한계점이 있었다.

둘째, 본 연구는 선행연구의 투자 결정 요인을 바탕으로 벤처 투자 예측 모델을 개발했다는 데 의의가 있다. 국내 벤처 투자 분야에서 예측 모델 기법을 적용한 연구는 없었다. 또한, 투자 예측 모델 개발 뿐만 아니라 SHAP 기법을 활용하여 예측 모델을 해석 가능하게 했다는 점이다. 이는 예측 모델의 Black box 문제를 보완하고, 벤처 투자 중요 요인 분석을 가능하게 했다.

실무적 시사점으로는 투자 유치를 원하는 벤처기업가들에게 보완할 정보를 제공할 수 있으며, 벤처캐피탈에는 투자 결정에 대한 보조 수단을 제공할 수 있다. 이는 많은 기업들을 대상으로 투자를 고려할 때 기업을 사전 선별하여 업무 효율을 향상시킬 수 있다 데 실무적 의의가 있다.

참고 문헌

- [1] 중소벤처기업부 <2019년 벤처기업정밀실태조사>
- [2] 중소벤처기업부 <2020년 벤처기업정밀실태조사>
- [3] Gompers, P. A., & Lerner, J.(2004). The venture capital cycle, MIT press.
- [4] Gordon, J.(2014). A stage model of venture philanthropy, Venture Capital, 16(2), 85-107.
- [5] de Treville, S., Petty, J. S., &Wager, S. (2014). Economies of extremes: Lessons from venture-capital decision making. Journal of Operations Management, 32(6), 387-398
- [6] Sharma, A. K., Monika(2015). Venture Capitalists' Investment decision criteria for new ventures: A Review, Procedia-Social and Behavioral Sciences, 189, 465-470.
- [7] 구중회, 김영준, 이수용, 김도현, &백지연. (2019). 한국 벤처캐피탈리스트의 투자결정에 미치는 요인 연구. 벤처창업연구, 14(4), 1-18.
- [8] 차명수(2009). 학술연구: 벤처캐피탈의 투자의사결정 연구: 판단분석의 활용, 기업가정신과 벤처연구 (JSBI)(구 벤처경영연구), 12(4), 19-43.
- [9] 반성식 · 송경모(2004). 한국 벤처캐피탈리스트의 투자의사결정 요인과 투자평가 모형, 대한경영학회지, (42), 267-291.
- [10] 반성식(2004). 한국 벤처캐피탈리스트의 투자결정 과정: 경험 많은 벤처캐피탈리스트가 보다 좋은 의사결정을 하는가?, 산업경제연구, 17(4), 1339-1358.
- [11] 구경철 · 이철규 · 유왕진(2008). 학술연구: 한국 벤처캐피탈의 투자결정모형 도출 및 타당성 분석, 기업가정신과 벤처연구 (JSBI)(구 벤처경영연구), 11(4), 1-20.

[12] 조세근(2019). 기업벤처캐피탈의 투자 결정요인에 관한 실증 연구:기술창업기업 투자를 중심으로, 박사학위논문. 중앙대학교 대학원.

[13] 조세근, &한주희. (2018). CVC 의 투자동기 및 투자 결정요인에 대한 사례연구: CVC 9 개社의 투자 사례를 중심으로. 벤처창업연구, 13(6), 27-38.

[14] 강준영, 김혜진(2019), 벤처생태계의 스케일업(scale-up)과 패러다임 변화, KDB 산업은행

[15] Franke, N., Gruber, M., Harhoff, D., &Henkel, J. (2008). Venture capitalists' evaluations of start-up teams: Trade-offs, knock-out criteria, and the impact of VC experience. *Entrepreneurship Theory and Practice*, 32(3), 459-483.

[16] Dhochak, M., Sharma, A. K.(2015), Impact of global financial crisis on Indian venture capital firms: an empirical evaluation, *Journal for International Business and Entrepreneurship Development*, 8(4), 330-345.

[17] 한재우, 김병관, &김완기. (2016). 벤처기업 투자를 위한 실무 평가 지표와 해석. *기술혁신연구*, 24(3), 1-28.

[18] 김승찬, &김홍근. (2019). 개인투자조합 출자자의 출자의사결정요인 실증연구: 벤처기업 투자의사결정요인을 중심으로. *대한경영학회지*, 32(11), 2051-2084.

[19] Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

[20] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.

[21] S. J. Kim, H. C. Ahn, "Application of Random Forests to Corporate Credit Rating Prediction", *The Journal of Industrial Innovation*, Vol. 32, No. 1, pp. 187-211, March 2016.

[22] J. Y. Oh, D. H. Ham, Y. G. Lee, and G. Kim, "Short-term Load Forecasting Using XGBoost and the Analysis of Hyperparameters", *The transactions of The Korean Institute of Electrical Engineers*, Vol. 68, No. 9, pp. 1073-1078,

September 2019.

[23] Ogunleye, A., & Wang, Q. G. “XGBoost model for chronic kidney disease diagnosis”, IEEE/ACM transactions on computational biology and bioinformatics, Vol 17, No. 6, pp. 2131-2140. 2019.

[24] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[25] Ma, X., Sha, J., Wang, D., yu, Y., Yang, Q., and Niu, X., “Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGBoost algorithms according to different highdimensional data cleaning”, Electronic Commerce Research and Applications, Vol. 31, pp. 24-39. 2018.

[26] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ..., and Liu, T. Y., “Lightgbm: A highly efficient gradient boosting decision tree”, Advances in neural information processing systems, Vol. 30, pp. 3146-3154, 2017.

[27] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A., “CatBoost: unbiased boosting with categorical features”. arXiv preprint arXiv:1706.09516. 2017.

[28] M. Turek, “Explainable Artificial Intelligence (XAI)”, Defense Advanced Research Projects Agency

[29] S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions”, Advances in Neural Information Processing Systems, Vol. 2017-December, No. Section 2, pp. 4766-4775, Dec, 2017.

[30] J. H. Ahn, “Explainable artificial intelligence and artificial intelligence.”, wikibooks Pub, ch 05, pp 168-171, 2020.

[31] Antwarg, L., Miller, R. M., Shapira, B., and Rokach, L., “Explaining Anomalies Detected by Autoencoders Using SHAP,” arXiv: 1903.02407, 2019.

[32] Mokhtari, K. E., Higdon, B. P., and Başar, A., “Interpreting Financial Time Series with SHAP Values,” Proceedings of t^{he} 29th Annual International Conference on Computer Science and Software.

[33] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[34] Silva, J.(2004). Venture capitalists’ decision-making in small equity markets: a case study using participant observaion, Venture capital, 6(2-3), 125-145.

[35] 한국경제, “위기 돌파, 해외시장 개척은 기업인의 숙명”, 신용훈 기고, 2021.10.03.

ABSTRACT

Predictive Modeling of Venture Capitalist Investment and Factor Analysis

The number of venture companies(VC) of South Korea has steadily increased, reaching 39,511 companies as of 2020, the largest number ever, has become the fourth largest venture investment ratio in the world after the United States, Israel, and China. Regardless of the fact that the rising interest in venture companies and venture company investment, the fact is a shortage of research materials and research on the venture investment because of its character: confidential property, remains inadequate. Previous studies also point out the same limitation that few specimens for studying venture investment determinants. Therefore, the purpose of this research is to analyze venture investment factors using a large amount of data and to develop a model that predicts companies that will receive venture investment.

LightGBM model showed the best performance achieved the f1-score of 85.02%, and the investment factors were analyzed using the SHAP method through the explanatory variables used in the predictive model. As a result, in the part of the founder characteristics, as the higher the final educational background of the founder at the time of founding, the higher the SHAP value had a positive (+) influence, and it was the same positive relationship in the case of a co-founder rather than a sole proprietorship. For product/service characteristics, the positive SHAP value relationships were found as the higher the domestic market share of the main product, the higher the proportion of overseas business sales, and the case where the business content is IT-related. On the other hand, considering that the SHAP value still has a positive effect even if the financial situation is not affirmative, it can be inferred that the financial characteristics are not highly considered for venture investment.

Through this study, we can identify the investment factors of VCs and provide information to supplement venture entrepreneurs who want to attract investment. Venture capital has practical significance in that can provide auxiliary means for investment decisions.

Key words : Predicting investment in venture companies, analyzing investment factors in venture capitalists, CV, LightGBM, SHAP