「정보시스템연구」 제27권 제1호
한국정보시스템학회
2018년 3월, pp. 171～191

http://dx.doi.org/10.5859/KAIS.2018.27.1.171

중소기업 프로파일링 분석을 통한 기술유출 방지 및 보호 모형 연구*

유인진** · 박도형***

〈목 차〉

Ⅰ. 연구의 필요성 및 목적
Ⅱ. 이론적 배경
2.1 중소기업의 기술유출
2.2 기술유출 요인
2.3 중소기업 프로파일링
Ⅲ. 연구 모형 및 연구 가설
3.1 연구 프로세스
3.2 자료의 특성 및 분석방법
Ⅳ. 연구 결과
4.1 의사결정나무 분석
4.2 판별분석
4.3 로지스틱 회귀분석
Ⅴ. 결론 및 시사점
참고문헌
<Abstract>

Ⅰ. 연구의 필요성 및 목적

최근 제 4차 산업혁명과 함께 지능정보화 시대가 대두되면서, 그 산업의 주체인 ICT (Information and Communication Technology) 또한 급격히 부상하고 있다. 빅데이터와 인공지능, IoT 등 세상을 변화시키는 기술이 등장하면서, 개인의 소비나 여행/문화, 금융, 더 나아가 기업의 생산, 투자, 기술까지도 광범위한 영향을 미치고 있다(김건우, 박도형, 2017; 이동원 등, 2014. 전승표, 박도형, 2013; 김범수 등, 2015; 최유지, 박도형, 2017). 다가오는 4차 산업혁명에 대응하기 위하여 정부와 기업을 포함한 다양한 분야의 정책방향성 또한 진화하고 있다. 특히 과학기술 정책의 경우, 기존의 연구개발(Research and Development)에서 '사회문제 해결형 연구개발(Research and Solution Development)'의 형태로 변화하고 있다. 이처럼 다양한 산업의 핵심 경쟁력이 산업생태계의 혁신으로 나아감에 따라서 국내 중소기업을 육성하여 핵심 경쟁력을 확보하는 것이 국내 산업의 핵심과제로 강조되고 있다(미래창조과학

* 이 논문은 2017년 대한민국 정부(교육부)와 한국연구재단의 지원을 받아 수행된 연구임. (NRF-2017R1C1B5016712)
** 국민대학교 비즈니스 IT 대학원, ijyoo@kookmin.ac.kr(주저자)
*** 국민대학교 경영대학/비즈니스 IT 전문대학원, dohyungpark@kookmin.ac.kr(교신저자)

부, 2013; 김문구 등, 2015; 홍지훈, 박도형, 2013). 국내 중소기업의 경우 이미 전체 산업의 99%, 전체 고용의 87.7% 절대적 비중을 차지하고 있으며, 그 비중 또한 확대되고 있는 추세이다(미래창조과학부, 2013). 이는 국내 중소기업에 대한 정부 차원의 R&D활동과 ICT 인프라 등, 다차원적 지원에 의해 국가 경쟁력의 주축으로 성장하고 있음을 보이는 바이며, 나아가 중소기업은 발 빠른 기술혁신의 수용을 통해 4차 산업혁명의 주역이 될 것으로 기대해볼 수 있다.

그러나 중소기업은 다가오는 4차 산업혁명에 대한 대응에 걸림돌이 될 수 있음을 고려할 필요성이 있다. 이는 중소기업이 내외부적 지원과 성장을 통해 국내 ICT 환경의 활성화를 담당하는 한 축으로 자리매김 하는 과정을 거치고 있는 한편, 오히려 기업의 기술유출 피해에 의해 경쟁력 확보가 어려워지고, ICT 활성화를 저해하는 걸림돌이 될 수 있기 때문이다(서봉군, 박도형, 2017).

중소기업의 기술유출은 물질적 손실뿐만 아니라, 기업의 이미지 손상 등의 비물질적 손실과 같은 수많은 악영향을 야기하며, 나아가 산업발전과 국가 성장까지 저해할 가능성을 내포한다. 특히, 중소기업은 대기업에 비하여 핵심기술에 대한 의존도가 매우 높기 때문에 기술유출 이슈와 그에 대한 예방책 마련은 기업의 생존에 있어서 필수불가결한 요소로 볼 수 있다.

이에 최근 기술유출 이슈에 대하여 정부주도로 수행된 기술보호 정책 지원 결과, 기업의 기술유출 비율은 소폭 감소하였으나, 오히려 기업당 기술유출로 인한 피해 규모는 증가하는 추세를 보이고 있다(김문선 등, 2007). 이는 기술유출이 발생할 수 있는 기업을 보다 실증적인 데이터를 기반으로 분석하고, 사전에 정확하게 예측하는 것이 중요해지고 있음을 시사한다. 최근 빅데이터 기술의 발전으로 다양한 분야에서 데이터 분석이 활용되고 있다. 웹검색트래픽 데이터부터(Jun and Park, 2016; 2017; 박도형, 2017), SNS 상의 구전데이터(강태영, 박도형, 2016), 기업의 공시데이터(Kang et al., 2018), 정부의 개방 정책에 따라 분석 가능하게 된 공공데이터까지(박도형 등, 2014; 정재권 등, 2014; 이동원 등, 2015) 다양한 분야에서 여러 기법들이 활용되고 있다.

이런 추세에 발맞추어 본 연구에서는 데이터마이닝 기법을 통해 기술유출 가능성이 있는 기업에 대한 프로파일링과 기술유출 기업에 대한 예측 모형을 제시하고자 한다. 구체적으로 의사결정나무를 통해 기술유출이 발생한 기업에 대한 프로파일을 구축하여 실증적인 데이터 근거를 산출하고, 앞서 구축된 프로파일에 근거하여, 판별분석과 로지스틱 회귀분석을 통해 기술유출 기업에 대한 예측 모형을 구축할 것이다.

Ⅱ. 이론적 배경

2.1 중소기업의 기술유출

중소기업의 기술유출에 대한 선행연구들은 정책적 대안 제시가 대부분으로 정보보안 및 기술보호 관련 정책에 대한 접근을 통해 보다 효율적인 대안을 제시하고자 하였다. 그러나 그

중 통계적 분석이나 증거기반 실증 분석은 약 16%로 상대적으로 매우 낮은 수준에 머무르고 있는 것을 볼 수 있다(정병수 등, 2012). 구체적으로 기술유출에 대한 보안정책 관련 선행연구에서는 국가적 차원에서 중소기업 보안에 대한 지원책과 그 연구들을 비교 분석하여 국내 보안정책에 대한 현실적인 시사점을 제공하고 있다. 그 중 미국과 유럽의 경우는 기업 정보보안 교육 및 보안 전문 강사 양성을 진행하거나, 정보화 지원 정책과 일반 사용자 및 중소기업을 중심으로 한 통신만 안전 경보 시스템 구축 등에 대한 연구를 지원하고 있음을 볼 수 있다(문현정, 2009).

실증 분석의 경우 단순히 대안 제시나, 새로운 대안의 필요성을 주장하는 것에 그치지 않고 분석 결과에 근거하여 보다 의미 있는 결과를 제시하기도 한다. 하지만, 기업의 기술유출에 대한 대부분의 연구는 전산망이나 서버, 웹로그와 트래픽과 같은 전산적이고, 기술적인 분야에 중점을 두고 사내 보안 관리 및 인식 실태에 대한 연구를 통해 선행 모델 및 정책의 한계점과 개선점을 도출하고, 기업 기술유출 방지를 위한 보안 모델을 구축하거나 중소기업의 특성을 고려한 정책적 대안을 제시하는 경우가 많았으며, 특히 사이버 보안에 관한 모델에 집중하는 경향을 볼 수 있다(이희명 등, 2008; 노민선 등, 2010; 문제욱 등, 2013; 문호건 등, 2014; 홍준석 등, 2014).

그 외에도 기업의 기술유출에 대한 재무적 측면에서, 기업의 매출액과 보안관리 투자 간의 상관관계를 규명하거나, 보안 교육과 기업 보안 역량 간의 상관관계를 도출하는 등의 연구 또한 활발히 수행되고 있는데(이강백 등, 2015; 이치석 등, 2015), 이는 통계적 분석을 통해 기술유출이 발생하는 기업을 예측하는 선행연구도 다수 수행되고 있음을 시사하고 있다(허승표 등, 2010; 홍준석 등, 2015). 다만 데이터 수집의 불균형성과 과거 기밀정보 유출 사실의 은닉, 응답 대상자의 기억에 의존한 회상 편향 문제 등의 데이터 수집에 대한 문제점이나 중소기업의 기술적, 구조적 요인을 반영하지 못하는 등의 한계점을 확인할 수 있다.

2.2 기술유출 요인

중소기업의 기술유출에 대한 선행연구들은 두 가지 관점으로 크게 나누어 살펴볼 수 있다. 첫 번째는 기업이 기술보호를 목적으로 보유한 보안수준에 따른 기술보호 역량에 의하여 기술유출이 방지되는 정도에 집중하며, 두 번째는 기업에 따른 고유한 기업 특성에 의하여 기술유출이 발생되는 정도에 집중한다. 각 관점은 기술유출을 방지하는 요인과 기술유출을 야기하는 요인에 각각 초점이 맞추어져 있으나, 공통적으로 기술유출과 관련된 개별 요인들을 상위 카테고리로 분류하여 기술유출에 대하여 보다 전사적인 측면에서 접근한 것을 볼 수 있다. 박향미 등(2015)의 연구는 전자의 관점에서 국내 정보보안 기관의 기술보호를 위한 지침서 간의 비교분석을 수행하였으며, 이때 각 지침서에서 살펴볼 수 있는 개별 보안 요소들을 각각 관리적 보안, 물리적 보안, 기술적 보안으로 응집하고 나아가 중소기업의 보안 적용에 대한 부담을 줄일 필요가 있음을 시사하였다. 문현정(2009)의 연구는 후자의 관점에서 국내 중소기업의 보안 위협에 대한 중요성을 기반으로 중

소기업의 특성에 맞는 보안의 필요함을 제시하였는데, 이때 국내 중소기업이 보유한 다양한 특성들을 경영 측면, 업무 조직적 측면, 경영 환경적 측면으로 분류하여 이를 설명하였다. 그 외에도 기업이 보유한 핵심 기술의 매력도나 유출 행위자의 특성에 따라서 기술유출 요인을 분류하거나(채정우 등, 2012), 기술유출 요인을 기업 내 종사자들의 심리적 요인의 차원에서 접근하고(황현동 등, 2006), 나아가 기업이 속한 산업군의 시장 포화도나 경쟁수준에서 기업을 하나의 개인으로 바라보며 범죄학적 차원에서 접근하는 연구가 수행되기도 하였다(J. Conklin et al. 1977).

본 연구는 중소기업이 보유한 다양한 요인들이 기술유출에 영향을 미치는지에 대한 후자의 관점에서 살펴보고자 하며, 기업이 보유한 수많은 요인들을 상위 차원으로 범주화하여 보다 전사적인 관점에서 다수의 요인들을 다루고자 하였다. 이때 각 요인들은 '기업 특성,' '조직적 특성,' '기술적 특성,' '관계적 특성,' '재무적 특성,' '기업 핵심역량'으로 범주화 되었다.

첫 번째 기업 특성은 '기업 유형,' '기업 성장단계,' '주 납품처'가 하부 요인으로 구성된다. 기업 유형의 경우 해당 중소기업의 성격상 분류로 혁신형 기업으로 분류되는 기술혁신형, 경영혁신형 기업으로 지정받은 기업과 일반기업으로 분류되는 벤처기업 및 일반중소기업으로 구성된다. 기업 성장단계는 해당 기업의 연혁 및 활성화 수준을 반영하기 위함으로 매출액 및 시장점유율의 성장성과 수익성, 경영지표에 따라서 시장진입기, 성장기, 성숙기, 쇠퇴기로 분류된다. 주 납품처는 수출기업과 내수기업으로 구분되는 기준을 보다 세분화하며, 기업에서 판매하는 재화 및 서비스의 주 고객층(납품처)을 대기업, 중소기업, 소비자, 수출로 분류하였다.

두 번째 조직적 특성은 '연구직 인원수,' '성과보상제도 시행여부,' '장비 보유수준,' '외부장비 활용수준,' '기술개발 인프라 수준'이 하부 요인으로 구성된다. 연구직 인원수의 경우 기업 내 기술개발을 위한 연구를 수행, 보조하는 연구직 종사자의 수를 나타내며 기업의 조직적 규모를 반영한다. 성과보상제도 시행여부는 직원의 업무 성과에 따른 성과급 및 인센티브 지급 제도를 포함한 물질적, 비물질적 보상제도 시행 여부를 나타낸다. 장비 보유수준과 외부장비 활용수준은 기술개발을 위하여 소모, 활용되는 장비 및 설비를 기업 내에서 보유하고 있는 수준과 기업 내에서 보유하지 않고 외부 기업 및 타 기관의 장비 및 설비를 활용하는 수준을 나타낸다. 기술개발 인프라 수준은 기업 내 기술개발을 위한 전담 조직의 유무를 나타내며 기업 부설 연구소의 유무, 상시 개발 전담 부서의 유무, 임시 개발 전담팀의 유무와 전담 부서 부재로 구성된다.

세 번째 기술적 특성은 '기업 기술군,' '세계대비 기술수준,' '기술 신규성,' '기술 모방기간,' '제품 수명주기,' '지식재산권 등록건수,' '기술개발 시도'가 하부 요인으로 구성된다. 기업 기술군은 해당 기업이 보유한 기술의 특성에 따른 분류로 고기술, 중간기술, 범용기술로 구성된다. 세계대비 기술수준은 해당 기술 및 제품의 선도 기업 대비 자사의 기술수준을 나타낸다. 기술 신규성은 해당 기업이 주력으로 연구개발한 기술의 신규성 수준을 나타내며, 세계 최초 신기술, 일부 선진국 개발 및 국내 최초 개발, 선진국 보편화 및 국내 최초 개발, 보편화

기술로 구성된다. 기술 모방기간과 제품 수명주기는 각각 기업이 보유한 주력 제품기술에 대하여 외부 경쟁업체에서 모방 및 개발하는데 소요되는 기간과 주력 제품기술의 시장도입-쇠퇴기까지의 기간을 나타낸다. 지식재산권 등록 건수는 기업 내 보유 및 출원 중인 특허 및 실용신안권 등의 산업재산권 등록 현황을 나타내며, 기술개발 시도는 과거 기술개발을 수행한 이력을 반영하며 기술개발을 시도한 횟수와 개발 성공비율로 구성된다.

네 번째 관계적 특성은 '개방형 혁신여부(Open Innovation),' '개발 수준'이 하부 요인으로 구성된다. 개방형 혁신여부는 대기업에 비하여 상대적으로 열위의 자원을 보유하는 중소기업에 있어서 기업 협력 네트워크를 통하여 부족한 자원과 역량을 확보하여 기술을 개발하는 것으로 단점을 보완하고 강점을 더욱 강화할 수 있으나, 그 과정에서 야기되는 기술유출 가능성을 반영한다. 개발 수준은 기술개발을 수행하는 과정에서 기업 온전한 자체개발 수준, 협력을 통한 공동개발 수준, 하도 및 외주를 통한 위탁개발 수준으로 구성된다.

다섯 번째 재무적 특성은 '기술개발 투자비'와 '기술개발을 통한 매출액'이 하부 요인으로 구성된다. 기술개발 투자비는 신기술이나 제품개발 및 개선과 관련하여 투입된 총비용을 나타내며, 인건비, 원재료비, 감가상각비 등 손익계산서 상의 투입된 각 비용이 포함된다. 기술개발을 통한 매출액은 신제품 출시 및 제품 개선이나 기술개발을 통하여 발생한 매출액을 나타낸다.

마지막으로 기업 핵심역량은 '제품기획,' '디자인,' '신기술개발,' '제품설계,' '공정설계,' '시험검사,' '제조가공,' '생산관리,' '유지보수,' '사업화'의 각 역량이 하부 요인으로 구성된다. 이때 각 요인은 기업이 보유하고 있는 핵심 기술에 대하여 경쟁국 및 세계선도 기술능력 대비 상대적 기술요소별 기술능력을 나타낸다.

2.3 중소기업 프로파일링

본 연구에서는 '프로파일링(Profiling)'을 통하여 기술유출이 발생하는 기업의 프로파일을 구축하고자 한다. 프로파일링은 범죄학에서 개인의 특성을 계량화하여 분석하는 기법으로 데이터마이닝 및 분석의 관점에서는 정보 수집을 통한 서술 및 설명으로 정의된다. 따라서 중소기업을 하나의 '개인'으로 바라보면, 개인에게 문제가 발생하는 원인을 라이프 스타일이나 성격, 행동 등으로 설명할 수 있으며, 나아가 중소기업의 경우 기술유출이 발생한 기업이 보유한 기술적, 경영적, 조직적 특성 등으로 설명할 수 있을 것이다.

프로파일링을 사용한 선행연구 중 중소기업에 대하여 시사점을 제공한 연구로 전승표 등(2016)의 연구를 확인해볼 수 있다. 해당 연구에서는 표준정책에 대한 수요가 있는 중소기업을 '표준 대응'과 '표준 정책'의 성격으로 분류하고 각 중소기업 집단의 특징을 프로파일링을 통해 도출하였다. 특히 이론적 배경을 토대로 한 연역적 접근과 함께 데이터 분석을 사용함으로써 하여금 귀납적 접근 또한 적용하여 각 접근 방법의 장점을 적극 활용한 장점이 있다. 또한 데이터에 기반 한 기업 특성을 도출하는 과정에서 이론적 연구에서 다루어지지 못하던 새로운 기업 특성을 확인한 시사점을 가지기도 한다.

이처럼 중소기업에 대한 프로파일링은 이론적 배경과 함께 데이터 분석을 통해 기업에 대한 보다 다양한 관점에서 접근할 수 있는 이점이 있다. 특히 기업이 보유한 수많은 특성들 중 목적에 부합하는 요인들을 도출하고, 관심 받지 못하던 유의한 영향요인들 또한 확인해볼 수 있다. 따라서 본 연구에서는 중소기업이 보유한 다수의 요인을 기반으로 기술유출이 발생하는 기업의 특징들을 프로파일링 분석을 통해 확인하고, 중소기업의 각 특징들을 상위 차원으로 범주화하여 전사적 관점에서 기술유출에 영향을 미치는 요인들을 도출하여 보고자 한다.

Ⅲ. 연구 모형 및 방법

3.1 연구 프로세스

본 연구에서는 위험관리를 통한 ICT 환경의 활성화 측면에서 기술 유출 기업 프로파일과 그 예측 모형을 구축하고자 한다. 이때 각 중소기업을 개인의 관점에서 바라보며 분석을 수행하는 '프로파일링' 방식을 통하여 기술유출 예측모형에 접근하였으며, 전술한 선행 연구에 근거하여 중소기업이 보유한 다수의 요인들을 여섯 가지 범주로 분류하여 전사적인 관점에서 중소기업의 기술유출을 다루고자 하였다. 모형의 종속변수인 '기술유출 여부'는 과거 기업의 기술유출 여부를 확인하는 문항을 통하여 설정하였고, 이에 영향을 미치는 기업의 특성들은 각각 '기업 특성,' '조직적 특성,' '기술적 특성,' '관계적 특성,' '재무적 특성,' '기업 핵심역량'으로 29개의 독립변수에 대하여 상위 범주화를 수행하였다.

도출된 기술유출 영향요인의 상위 범주에 기반 하여 '중소기업기술통계조사'의 국가주도 설문을 통하여 중소기업 표본과 함께 정량적으로 각 하위 요인들을 확보하고자 하였다. 이후

<표 1> 기업 범주별 기술유출 하부요인

변수 상위 범주	하부 요인
기업 특성	기업 유형
	기업 성장단계
	주 납품처
조직적 특성	연구직 인원수
	성과보상제도 시행여부
	장비 보유수준
	외부장비 활용수준
	기술개발 인프라 수준
기술적 특성	기업 기술군
	세계대비 기술수준
	기술 신규성
	기술 모방기간
	제품 수명주기
	지식재산권 등록건수
	기술개발 시도

변수 상위 범주	하부 요인
관계적 특성	개방형 혁신여부
	개발 수준
재무적 특성	기술개발 투자비
	기술개발을 통한 매출액
기업 핵심역량	제품기획
	디자인
	신기술개발
	제품설계
	공정설계
	시험검사
	제조가공
	생간관리
	유지보수
	사업화

<그림 1> 연구 프로세스

분석의 용이성과 각 표본 간 정규성의 확보를 위하여 데이터 전처리와 제곱근 변환을 수행하였다. 또한 확보된 중소기업의 유출-비유출 기업 표본 상 비율의 불균형으로 인하여 표본에 대한 혼합 샘플링을 통하여 각 표본 간 1:1 대응을 이루어 모형의 타당성을 확보하고, 편향을 제거하고자 하였다. 나아가 데이터마이닝의 기법 중 하나인 의사결정나무를 통해 기술 유출이 발생한 혹은 발생할 가능성이 높은 기업에 대한 프로파일을 도출하여 실증적인 데이터 근거를 산출하고, 이어 도출된 프로파일에 근거하여, 판별분석과 로지스틱 회귀분석을 통해 기술유출 기업에 대한 예측 모형을 구성하였다.

3.2 자료의 특성 및 분석 방법

3.2.1 자료 수집과 전처리

본 연구에서 사용한 분석 자료는 중소기업청과 중소기업중앙회에서 공동 수행, 설문한 '중소기업기술통계조사'의 2011년, 2013년, 2014년, 2015년의 4개년 조사 데이터를 통합하여 구성하였다. 이때 각 연도별로 각각 3,500개, 2,000개, 2,200개, 3,300개의 기업 표본에 대한 설문이 수행되었다. 상기 조사는 중소기업의 R&D 실태를 파악하고, 중소기업에 대한 기술개발 활동, 기술개발 조직, 기술보호 등과 같은 중소기업 지원에 대한 기획 및 추진을 위한 자료이며, 중소기업기술혁신촉진법 제 8조(중소기업 기술통계의 작성)에 근거하여 실시되고 있다. 조사 대상 기업은 '한국표준산업분류'인 KSIC2 상에서 제조업 및 제조업 외 기업을 포함하며 면접조사, E-mail, 유선조사 등의 방법으로 수집되었다.

이후 분석의 용이성과 각 표본 간 정규성의 확보를 위하여 수집된 데이터에 대한 전처리가 수행되었다. 본 연구에서 사용된 프로파일링 중 의사결정나무 기법의 경우 비모수형 검정 방식으로 명목형 변수와 연속형 변수의 종류에 따른 선택의 문제가 발생하지 않으며 원 데이터에 대한 전처리를 요구하지 않는다. 하지만 이후 수행 될 판별분석과 로지스틱 회귀분석에 적용되는 다변량 데이터의 경우 변수의 정규성이 고려되며, 연속형 변수를 사용해야 한다. 이때 '중소기업기술통계조사'를 통해 수집된 각 변수들은 각각 범주형 변수와 연속형 변수로 분류된다. 따라서 변수의 정규성과 규모의 문제를 해결하기 위하여 범주형 변수에 대해서는 더미변수 생성 및 일반적인 정규화 방식을 통하여 변수 변환을 수행하였고, 연속형 변수들 중 연구직 인원수, 지식재산권 등록건수, 기술개발 투자비, 기술개발을 통한 매출액과 같은 연속형이며, 기업의 규모에 대한 의존적인 변수들에 대하여 제곱근 변환을 수행하였다.

3.2.2 데이터 샘플링

‘중소기업기술통계조사’ 상의 4개년 간, 총 10900개 표본 기업 중 오직 454개 기업이 기술유출에 의한 피해를 입은 것으로 수집되었다. 이때 기술유출 피해를 입지 않은 기업은 10446개로 전체 기업 표본의 약 96%의 비중을 차지하며, 이에 따라서 본 데이터를 그대로 사용할 시 비 유출 기업에 대한 편향된 예측 모형이 구축되어 모형의 타당성을 확보할 수 없다. 이는 구축 시 모두 기술유출이 일어나지 않을 것으로 예측되는 문제점이 발생되며 모두 비 유출로 예측하더라도 96%의 분류 정확도를 확보할 수 있기 때문이다. 따라서 기술유출이 발생한 기업 454개를 기준으로 샘플링을 수행하여 편향된 예측 모형을 방지하고, 각 표본 간 1:1 대응을 유지할 수 있도록 하였다.

이때 언더 샘플링(Under-Sampling)의 경우, 과적합화(Over-Fitting)의 문제는 해소할 수 있는 장점이 있으나, 전체 기업에서 언더 샘플링의 결과로 추출되는 기업이 매우 소수이기 때문에 많은 유용한 데이터가 소실되는 문제점이 발생할 수 있다. 반면 오버 샘플링(Over-Sampling)은 목적 표본인 기술유출 기업의 수가 상대적으로 작아(전체의 4%), 다수 범주에 대응시키기 위하여, 소수 범주 전체를 복제할 시 소수 범주 표본의 중복 문제화 함께 과적합화 문제가 발생할 가능성이 높다. 따라서 본 연구에서는 각 샘플링 방법을 혼합해 각 샘플링 시 발생할 수 있는 문제점을 완화하는 혼합 샘플링을 통하여 모형에 사용될 데이터 표본을 구성하였다.

이때, 혼합 샘플링 과정에서 적합한 각 샘플링의 수준을 개념적으로 추정하기 어렵기 때문에, 소수 범주 표본에 대하여 0% 수준에서 10%씩 증가시키며 100% 수준에 도달할 때 까지 오버 샘플링을 진행하여 최대 2배수의 기술유출 기업이 확보되도록 하였고, 동시에 기술유출이 발생하지 않은 중소기업 표본에 대하여 오버 샘플링 된 표본과 동일한 수의 기업 표본을 언더 샘플링하였고, 최종적으로 비 유출 기업을 기술유출 기업과 일대일 대응시켜 혼합 샘플링 표본을 구성하였다. 이때 수행한 각 오버샘플링 수준과 복제된 소수 집단 표본의 수, 총 소수 집단 표본 수(원 표본+복제 표본), 전체 표본 수(기술유출 기업+비 기술유출 기업)를 포함하는 단계적 혼합 샘플링 결과는 다음 <표 2>와 같다.

<표 2> 단계적 혼합 샘플링 결과

Sampling Rate	Sampled Number	Total Over Sample	Total Sample
10%	45	499	998
20%	91	545	1090
30%	136	590	1180
40%	182	636	1272
50%	227	681	1362
60%	272	726	1452
70%	318	772	1544
80%	363	817	1634
90%	409	863	1726
100%	454	908	1816

<표 3> 샘플링 수준 별 위험도 추정값

Sampling Rate	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
위험도	.300	.305	.294	.313	.321	.324	.306	.307	.310	.311

이후 각 샘플링 단계 수준에 따라 구성된 표본의 적합성을 확인하기 위해, 의사결정나무 분석을 통해 위험도 추정값을 확인하였다. 그 결과 기술유출 기업에 대한 30%수준의 오버샘플링 혼합 모형에서 위험도 추정값이 감소하는 것을 확인되었으며, 그 이상의 오버샘플링 수준에서는 의사결정나무의 구조도 깊이가 증가하며 지나친 가지 세분화와 함께 과적합화 문제가 발생하는 것이 확인되었다.

따라서 기술유출 케이스에 대한 30% 수준의 오버 샘플링과 비 유출 케이스에 대한 언더 샘플링을 혼합하여 1:1 대응 표본이 구성되었고, 총 1180개 분석용 표본이 확보되었다. 이후 각 중소기업의 설문 수집 연도와 KSIC2 산업 분류에 따라서 유출 및 비유출 기업에 대하여 그 수를 1:1로 대응시켜 모형을 구성하였고 모형의 검증을 위해 전체 표본을 학습용 표본과 검정용 표본으로 70:30의 비율로 분류하여 분석을 수행하였다.

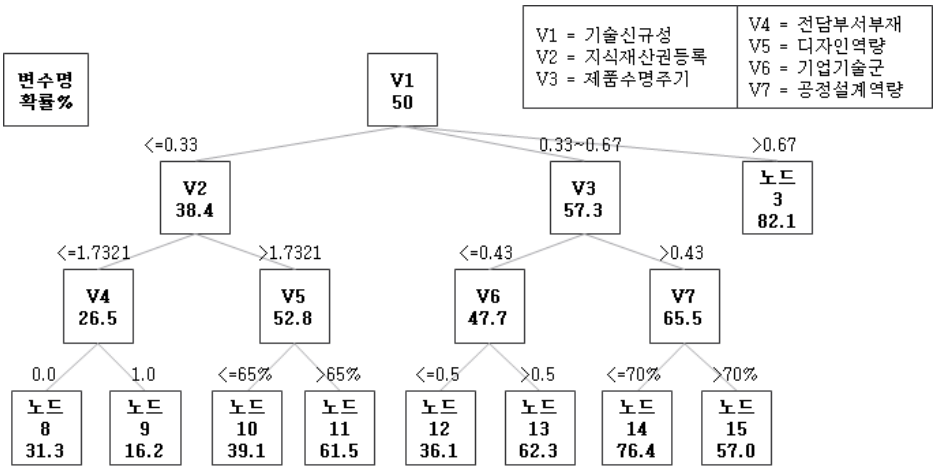
Ⅳ. 연구 결과

4.1 의사결정나무 분석

의사결정나무 분석은 일련의 변수들에 대하여 분류 및 예측을 수행하는데, 최종 결과물이 일련의 규칙으로 표현되어 결과의 대한 해석이 용이할 것으로 판단하였다. 분석에 앞서 데이터 샘플링을 통해 전체 표본을 학습용 표본과 검정용 표본으로 70:30의 비율로 분류하여 모형 분석의 타당성을 확보하고자 하였다. 모형 분석 간 반복적 분할을 통하여 기술유출 기업 분류를 시작하였으며, 지나친 세분화에 의한 과적합 문제를 해결하기 위해 가지치기를 수행하였다. 이때 CHAID는 독립변수와 종속변수 간의 관계를 확인할 때 유용한 방법으로 알려져 있다(Pyle, 1998). 따라서 CHAID 성장방법을 통해 적정 수준에서 나무 모형의 성장을 중단하였다.

4개년 기술유출 기업에 대한 의사결정나무 분석결과, 투입한 29개 변수 중, 7개의 변수를 통해 모형이 모두 구축되었다<그림 2>. 이때 기술 신규성, 기업 기술군, 지식재산권 등록건수, 제품 수명주기, 기술개발 인프라 수준(전담조직 부재), 기업 핵심역량(디자인)과 기업 핵심역량(공정설계)이 기술유출이 발생한 기업을 분류하는데 유의한 영향 요인으로 도출되었다.

특히 기술 신규성에 따라서 기술유출이 두드러지게 변화하는 것을 확인 할 수 있는데, 세계 최초의 가까운 기술 신규성을 보유한 경우, 기술 신규성 하나의 조건만 충족되었음에도 불구하고 기술유출이 82.1%까지 증가하였다. 그 외 기술 신규성이 일부 선진국에서만 개발 및 보편화 되었으나 국내에서는 최초 개발이면서 제품 수명주기가 짧은 경우, 기업 기술군이 첨단 및 고기술 기업일 때 기술유출이 상대적으로 빈번하게 발생하였고(62.3%), 제품 수명주기가 상대적으로 긴 경우, 기업 핵심역량(공정설계)



<그림 2> 의사결정나무 분석 결과

이 70% 이하의 수준을 갖춘 기업일 때 76.4%까지 기술유출 빈도가 증가하는 것이 관측되었다. 반면 기술 신규성은 보편화된 기술이며, 지식재산권 등록건수이 약 3건 이하이면서 기술개발 전담부서(기술개발 인프라 수준)가 존재하지 않는 경우 기술유출 빈도가 16.2%로 현저히 감소하였다.

4개년 기술유출 기업에 대한 분석 결과 학습용 표본에서의 위험도 추정값은 32.7%이고 검정용 표본에서는 35.9%로 나타났다. 검정용 표본을 기준으로 의사결정나무의 전체 분류 정확도는 64.1% 이고, 학습용 표본의 분류 정확도는 67.3%로 나타났다.

<표 4> 의사결정나무 분석 결과

표본	관측	예측		
		비유출	유출	분류정확도
학습	비유출	262	151	63.4%
	유출	119	294	71.2%
	(%)	46.1%	53.9%	67.3%
검정	비유출	102	75	57.6%
	유출	52	125	70.6%
	(%)	43.5%	56.5%	64.1%

모형 분석 결과, 기술 신규성에 의해서 기술유출 빈도가 크게 변화하는 것을 확인할 수 있으며, 특히 기술 신규성이 세계 최초 및 국내 최초일수록(신규성이 높은 경우), 기업 기술군이 첨단 및 고수준 기업일수록, 기업 핵심역량(공정설계)은 낮을수록 기술유출이 빈번하게 발생하였다.

4.2 판별분석

앞서 판별과 회귀분석에 사용될 변수에 대하여 정규성을 확보하기 위해 정규화 및 제곱근 변환을 수행하였다. 또한 본 연구는 분석 모형에 투입될 변수들을 앞선 의사결정나무 분석을 통해 선정하였다.

이때 판별분석은 Fisher의 선형판별함수를 사용하며, 변수 진입은 단계선택을 통해 선정하며, 진입 기준은 Wilks's λ를 사용하였다. 의사결정나무 기법을 통해 구축된 모형에서 유의하게 관측된 7개의 변수인 기업 기술군, 기술 신규성, 지식재산권 등록현황, 제품 수명주기, 기

<표 5> 집단평균의 동질성에 대한 검정

변수명	Wilks's λ	F 통계량	유의확률
기업 기술군	.979	17.682	.000
기술 신규성	.942	50.768	.000
지식재산권 등록건수	.940	52.337	.000
제품 수명주기	.998	1.776	.183
전담부서 부재여부	.977	18.986	.000
기업 핵심역량(디자인)	.990	8.253	.004
기업 핵심역량(공정설계)	.996	3.344	.068

<표 6> 단계진입 선택별 판별분석 결과

단계	진입된 변수	Wilks's λ		
		통계량	정확한F	
			통계량	유의확률
1	지식재산권 등록건수	.940	52.343	.000
2	기술 신규성	.903	44.279	.000
3	전담부서 부재여부	.897	31.335	.000

<표 7> 정준 판별함수 요약(고유값, Wilk's λ)

함수	고유값	분산의%	누적%	정준상관	Wilks's λ	카이제곱	자유도	유의확률
1	0.114	100	100	0.320	0.897	89.062	3	.000

술개발 전담부서 부재여부, 기업 핵심역량(디자인), 기업 핵심역량(공정설계)을 사용하였다.

<표 5>의 집단 평균의 동질성에 대한 검정 결과, 기업 기술군, 기술 신규성, 지식재산권 등록건수, 전담부서 부재여부, 기업 핵심역량(디자인)이 95%의 신뢰수준 하에서 기술유출 기업에 대한 평균 차이가 유의한 것으로 나타났다. 그 중 지식재산권 등록건수는 Wilks's λ값은 가장 작고, F 통계량은 가장 큰 값을 나타내어 전체 분산의 비율이 높아 기술유출 기업에 대한 판별력이 높음을 알 수 있다. 반면 제품 수명주기와 기업 핵심역량(공정설계) 수준은 Wilks's λ값이 1에 가까우며 95%의 신뢰수준 하, 기술유출 기업에 대한 집단 간 유의한 차이가 없는 것을 확인할 수 있었다.

<표 6>을 통해 기술유출 기업에 대한 판별식 적합도를 확인한 결과 99%의 신뢰수준 하에서 Wilks's λ값이 낮아 판별식이 유의한 것을 확인할 수 있었다.

이후 산출된 정준 판별함수의 요약은 <표 7>과 같다. 분석을 통해 도출된 판별함수와 기술유출 기업 간의 관련 정도는 0.320로 관측되었으며, 정준상관계수의 제곱인 0.1024를 통해 종속변수인 기업 기술유출의 판별 점수 분산의 10.24%가 단계선택에 의해 선정된 3개의 독립변수에 의해 설명됨을 알 수 있다. 이때 판별함수의 고유값이 0.114로 집단 총분산의 100%를 설명하고 있다. 판별함수의 Wilks's λ는 0.897이며, Wilks's λ값과 자유도를 고려한 카이제곱 통계량은 89.062의 환산 값을 가지며, 99%의 신뢰수준 하에서 본 판별함수는 유의한 것으로 나타났다.

<표 8> 표준화 정준 판별함수 계수

변수명	표준화계수	계수
기술 신규성	0.602	1.895
지식재산권 등록건수	0.598	5.129
전담부서 부재여부	-0.249	-0.694
(상수)	-	-1.204

<표 8>의 분석결과 제시된 분류 함수의 계수표를 통해 기술유출 기업에 대한 다음과 같은 판별식을 구성할 수 있다.

*D=0.602*X1(기술 신규성) + 0.598*X2(지식재산권 등록건수) - 0.249*X3(전담부서 부재여부)*

단 이때, 판별식에 대입하기 위한 데이터는 표준화 되어야 하며, 이 판별식에서 계수의 절대값 크기는 변수들 간의 상대적 중요도를 나타낸다. 즉, (기술 신규성) > (지식재산권 등록건수) > (전담부서 부재여부) 순으로 설명력이 높은 것을 알 수 있다.

<표 9> 구조행렬

변수명	함수
지식재산권 등록건수	0.745
기술 신규성	0.734
전담부서 부재여부	-0.449
기업 기술군	0.303
기업 핵심역량(디자인)	0.233
기업 핵심역량(공정설계)	0.207
제품 수명주기	-0.066

이후 구조행렬을 통해 판별함수와 변수들 간의 상관관계를 관측할 수 있으며, 상관관계가 높을수록 판별점수도 높아진다. 이때 지식재산권 등록건수가 상관계수 0.745로 가장 크며, 본 판별함수에서 가장 영향력이 큰 변수로 도출되었다.

<표 10> Fisher의 선형판별함수

변수명	기술유출여부	
	비유출	기술유출
기술 신규성	3.056	4.336
지식재산권 등록건수	6.201	9.666
전담부서 부재여부	2.395	1.926
(상수)	-1.676	-2.489

이후 각 집단별 판별점수를 결정하는 Fisher의 선형판별함수를 통해 기술유출 집단에 대한 판별점수를 산출할 수 있으며, 그 판별식은 다음과 같이 산출되었다.

<표 11> 기업 기술유출 여부 별 판별식

기술유출 여부	판별식
기술유출 (O)	*Y = -1.676+3.056*X1(기술 신규성)+6.201*X2(지식재산권 등록건수) +2.395*X3(전담부서 부재여부)*
기술유출 (X)	*Y = -2.489+4.336*X1(기술 신규성)+9.666*X2(지식재산권 등록건수) +1.926*X3(전담부서 부재여부)*

마지막으로 사전 추출된 1180개의 기업에 대한 기술유출 집단을 판별한 결과, 도출된 판별함수와 판별식을 통해 학습용 표본에 대해서는 63%를 정확히 분류하였고, 검증용 표본에 대해서는 65%를 정확히 분류하였다. 이때 학습용 표본에서 실제 기술유출이 발생하지 않았으나, 모형에서는 기술유출이 발생할 것으로 예측된 표본이 그렇지 않은 표본보다 더 많이 관측되었다. 이는 실제 기술유출이 발생하지는 않았으나, 기업의 현 상황을 통해 앞으로 기술유출이 발생할 가능성이 있음을 시사하며, 기술유출 예방 차원에서 경각심을 일깨워 줄 수 있을 것으로도 해석해 볼 수 있다.

<표 12> 판별분석 결과

표본	관측	예측		
		비유출	유출	분류정확도
학습	비유출	249	164	60.3%
	유출	142	271	65.6%
	(%)	47.3%	52.7%	63.0%
검정	비유출	115	62	65.0%
	유출	62	115	65.0%
	(%)	50.0%	50.0%	65.0%

4.3 로지스틱 회귀분석

앞선 판별분석을 통하여 중소기업의 기술유출에 영향을 미치는 7개의 변수의 영향력 수준과 함께 판별식을 산출하여 변수간의 관계와 기술유출 기업을 예측하는 모형을 구성하였다. 다만 이때 종속변수인 기술유출이 이분형으로 구성됨에 따라서 이분형 척도로 구성된 종속변수를 예측하는데 매우 강건하고 해석력 또한 제공하는 로지스틱 회귀분석을 추가로 수행하였다.

로지스틱 회귀분석에 사용될 변수들은 앞서 정규화 및 제곱근 변환을 수행한 상태로 추가적인 조작은 수행되지 않았다. 또한 다변량 정규분포나 공분산 행렬의 가정이 요구되지 않아 의사결정나무 분석을 통해 확인된 7개 변수를 모두 모형에 투입하고, 모형의 검증을 위하여 훈련용 표본과 검정용 표본을 70:30으로 구성하여 기술유출 기업을 분류하였다.

이때 Hosmer와 Lemeshow 적합도 검정을 함께 수행하였고, 변수 진입은 모든 변수를 입력하여 유의하게 관측된 7개의 변수인 기업 기술군, 기술 신규성, 지식재산권 등록현황, 제품 수명주기, 기술개발 전담부서 부재여부, 기업 핵심역량(디자인), 기업 핵심역량(공정설계)이 모형에 투입되었다.

모형 요약 통계량과 Hosmer와 Lemeshow 검정 결과, Nagelkerke R-제곱이 0.143으로 해당 로지스틱 회귀 모형의 설명력이 14.3%인 것을 확인할 수 있으며, 적합도 검정의 유의확률이 0.136으로 0.05보다 크게 나타나 도출된 회귀모형이 적합한 것이 확인되었다.

분석 결과 도출된 로지스틱 회귀모형은 다음과 같다. 모형에 투입된 7개 변수 중 기술 신규성, 지식재산권 등록건수, 전담부서 부재여부의 3개 변수가 90% 신뢰수준에서 유의한 것이 확인되었다. 기술 신규성의 경우 99%의 신뢰수준 하에서 유의하며, 독립변수의 기울기를 나타내는 B의 경우 1.125로 (+)값을 가지는 것이 확인되었다. 지식재산권 등록건수 또한 99%의 신뢰수준 하에서 유의하며, B의 경우 0.161로 (+)값을 가지는 것이 확인되었다. 반면 전담부서 부재여부는 90%의 신뢰수준에서 유의하며, B는 -0.395로 유의하게 도출된 변수들 중 유일하게 (-)값을 가지는 것이 확인되었다.

Exp(B)의 경우, 기술 신규성이 보편적인 기술인 경우에 비하여 세계 최초의 신규성을 가지는 경우 3.083배의 기술유출 위험도가 증가

<표 13> 로지스틱 회귀분석 모형 요약 및 Hosmer와 Lemeshow 검정 결과

단계	모형 요약			Hosmer와 Lemeshow 검정		
	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱	카이제곱	자유도	유의확률
1	1051.575	0.107	0.143	12.355	8	0.136

<표 14> 로지스틱 회귀분석에 포함된 변수

변수명	B	S.E	Wald	유의확률	Exp(B)
기업 기술군	.265	.226	1.367	.242	1.303
기술 신규성	1.126	.249	20.368	.001	3.083
지식재산권 등록건수	.161	.031	26.224	.001	1.175
제품 수명주기	-.153	.244	.393	.531	.858
전담부서 부재여부	-.395	.222	3.168	.075	.673
기업 핵심역량(디자인)	.225	.335	.451	.502	1.252
기업 핵심역량(공정설계)	-.143	.320	.200	.655	.867
상수항	-.918	.327	7.867	.005	.399

하고, 지식재산권 등록건수는 제곱근 변환 수행 후 1 증가할수록 기술유출 위험도가 1.175배 증가하며, 기술개발을 전담하는 부서가 없는 경우 기술유출 위험도가 0.673배 감소하는 것으로 나타났다.

정(+)의 B값을 가지는 기술 신규성은 중소기업이 보유한 기술의 신규성 수준이 신규할수록, 즉 세계 최초의 신기술을 개발한 경우 기술유출의 위험도가 증가하는 것으로 나타났다. 기술의 신규성은 해당 기술의 매력도를 상승시킬 수 있으며, 타 기업의 경제적, 기술적 욕구를 만족시킬 수 있는 기회로 작용하여 기술유출의 위험도를 증가시키는 것으로 해석된다. 또한 지식재산권 등록건수의 경우 중소기업이 보유한 지식재산권의 건수가 증가할수록 기업 기술유출 위험도가 증가하는 것으로 나타났다. 이는 기업이 보유한 기술에 대한 특허 및 산업재산권을 통하여 기술 소유권 주장 및 법적 우위를 달성하여 기술유출을 사전에 방지할 수 있다는 일반적 통념과는 사뭇 대립되는 결과로 볼 수 있다. 다만, 중소기업이 지식재산권에 대하여 등록 절차와 비용에 대한 애로사항이 존재하고, 특허로 등록 시 기술 공개가 이루어짐을 고려할 필요성이 있다. 따라서 중소기업이 자사에 보유한 역량 이상의 지식재산권을 획득하는 과정에서 기업 자원이 치중될 위험과 기술 공개에 따른 모방 및 원천 기술에 대한 매력도 증가로 인하여 기술유출 위험도가 증가하는 것으로 해석해 볼 수 있다. 반면, 부(-)의 B값을 가지는 전담부서 부재여부는 중소기업이 보유한 기술개발 전담부서가 없는 경우 기술유출 위험도가 감소하는 것으로 나타났다. 즉, 기술개발을 위한 전담 부서나 연구소가 조직되어 있는 경우 기술유출 위험이 증가하는 것을 의미한다.

마지막으로, 로지스틱 회귀모형을 통한 기술유출 기업 분류 결과는 다음과 같다. 추출된 1180개 기업에 대한 기술유출 여부를 분류한 결과, 학습용 표본에 대해서는 62.3%를 정확히 분류하였고, 검정용 표본에 대해서는 66.9%를 정확히 분류하였다. 학습용 표본의 경우 판별분석에 비하여 다소 낮은 분류정확도를 나타내었으나, 검정용 표본에 대해서는 상대적으로 높은 분류정확도를 나타내는 것이 확인되었다.

<표 15> 로지스틱 회귀분석 분류표

표본	관측	예측		
		비유출	유출	분류정확도
학습	비유출	254	159	61.5%
	유출	152	261	63.2%
	(%)	49.2%	50.8%	62.3%
검정	비유출	124	53	70.1%
	유출	64	113	63.8%
	(%)	53.1%	46.9%	66.9%

Ⅵ. 결론 및 시사점

국내 중소기업은 정부 차원의 지원에 의해 국가 경쟁력의 주축으로 성장하고 있으며, 나아가 4차 산업혁명의 주역이 될 것으로 기대해볼 수 있다. 그러나 중소기업이 4차 산업혁명과 ICT 환경의 활성화를 담당하는 한 축으로 성장하는 과정에서 기술유출 피해에 의해 물질적 손실뿐만 아니라, 기업의 이미지 손상 등의 비물질적 손실을 야기할 수 있으며, 나아가 산업발전과 국가 성장까지 저해할 위험에 노출되어 있다. 특히, 중소기업은 대기업에 비하여 핵심기술에 대한 의존도가 매우 높기 때문에 기술유출 이슈와 그에 대한 예방책 마련은 기업의 생존에 있어서 필수불가결한 요소로 볼 수 있다.

이에 최근 기술유출 이슈에 대하여 본 연구는 데이터마이닝 기법을 통해 기술유출 가능성이 있는 기업에 대한 프로파일링을 수행하고자 하였다. 기업이 보유한 다수의 요인들에 대하여 '기업 특성,' '조직적 특성,' '기술적 특성,' '관계적 특성,' '재무적 특성,' '기업 핵심역량'으로 상위 범주화하였고, 각 범주에 속하는 요인들이 기업의 기술유출에 어떠한 영향을 미치는지 확인하고자 하였다. 이때 본 연구에서는 범죄학에서 개인의 특성을 분석하는 기법인 '프로파일링(Profiling)'을 통하여 기업을 개인으로 보는 관점에서 기술유출이 발생하는 기업의 프로파일을 구성하였다.

2011년부터 2015년의 '중소기업기술통계조사' 4개년 데이터를 통합하여 10900개의 기업 표본을 획득하여 분석을 수행하고자 하였으며, 이때 기술유출 기업 표본과 그렇지 않은 기업 간의 불균형으로 인하여 기술유출이 발생한 기업 454개를 기준으로 샘플링을 수행하여 편향된 예측 모형을 방지하고, 각 표본 간 1:1 대응을 유지할 수 있도록 하였다. 혼합 샘플링 결과 30%의 언더 샘플링 상에서 적합성이 확인되어 분석을 위한 최종 표본이 구성되었다.

의사결정나무 분석 결과 기술 신규성, 기업 기술군, 지식재산권 등록건수, 제품 수명주기, 기술개발 인프라 수준(전담조직 부재), 기업 핵심역량(디자인)과 기업 핵심역량(공정설계)이 기술유출이 발생한 기업을 분류하는데 유의한 영향 요인으로 확인되었다. 특히 기술 신규성이 높을 때(세계 최초 및 국내 최초의 기술일 때) 기술유출 빈도가 82.1%로 기존의 균형화 표본(50%)보다 30% 이상 기술유출이 증가하는 것을 확인할 수 있으며, 기술 신규성은 보편화된 기술이고,, 지식재산권 등록건수이 약 3건 이하이면서 기술개발 전담부서(기술개발 인프라 수준)가 존재하지 않는 경우 기술유출 빈도가 16.2%로 현저히 감소하는 것을 확인할 수 있었다.

판별분석에는 의사결정나무의 분석 결과로 유의하게 관측된 7개의 변수가 투입되었으며, 집단 평균의 동질성에 대한 검정 결과, 기업 기술군, 기술 신규성, 지식재산권 등록건수, 전담부서 부재여부, 기업 핵심역량(디자인)이 95%의 신뢰수준 하에서 기술유출 기업에 대한 평균 차이가 유의한 것으로 나타났으며, 그 중 지식재산권 등록건수가 기술유출 기업에 대한 판별력이 가장 높은 것이 확인되었다.

로지스틱 회귀분석 또한 앞선 의사결정나무 결과로 유의하게 관측된 7개의 변수가 투입되었으며, 회귀 모형의 설명력은 14.3% 로 확인

되었고, 적합도 검정 결과 또한 0.136으로 확인되었다. 투입된 변수들 중 기술 신규성, 지식재산권 등록건수, 전담부서 부재여부의 세 변수가 각각 99%, 99%, 90%의 신뢰수준 하에서 중소기업의 기술유출에 유의한 영향을 미치며, 그 중 기술 신규성이 기술유출에 가장 큰 영향력을 보유하는 것이 확인되었다.

본 연구는 다음과 같은 시사점을 가진다. 첫 번째, 중소기업의 기술유출에 관련한 선행 연구들은 기술유출에 대한 보안정책 관련 연구나 전산적, 기술적, 재무적 측면을 고려한 반면, 본 연구는 중소기업의 기술적, 구조적 측면과 함께 기업 전면의 다측면에 대한 요인들을 모형에 포함하여 기술유출 영향 요인의 범위를 확대하고, 이론적 한계 또한 완화하고자 하였다. 두 번째, 정부주도의 정량적인 대량의 정밀통계조사를 통해 다수의 기업 표본에 대하여 연구 모형을 검증하였다. 정책적으로는 첫 번째, 기업을 '개인'의 관점에서 접근하고, 다수의 요인을 모형에 포함하여 실무자의 경우 기업의 특성에 따른 기술유출 위험도를 파악할 수 있고, 유출 방지에 대한 전략적인 의사결정을 지원하는데 도움을 줄 수 있을 것으로 기대된다. 두 번째, 연구의 결과로 제시한 기업 프로파일과 예측모형은 중소기업에 대한 정보보안 지침 및 정책의 수립과 이행 과정에서 실증적인 근거를 제공할 수 있으며, 향후 ICT 환경에 대한 위험 관리적 측면의 활성화에 기여할 수 있을 것으로 기대된다.

앞서 기술한 시사점에도 불구하고, 본 연구는 다음과 같은 한계점을 가진다. 첫 번째, 수집된 전체 기업 표본에 비하여 실제 기술유출이 발생한 기업 표본의 수가 상대적으로 적어 혼합 샘플링을 통하여 모형의 편향과 타당성을 확보하고자 하였지만, 언더 샘플링 과정에서 유의미할 수 있는 다수의 기업 표본이 모형에서 소거된 것이다. 두 번째, 기술유출이 발생한 기업에 대하여 해당 기술유출이 발생한 기간이 명시되어있지 않았으며, 따라서 기술유출과 영향요인 간 선후관계에 대한 문제가 발생할 수 있다는 것이다.

참고문헌

강태영, 박도형, "전문가 제품 후기가 소비자 제품 평가에 미치는 영향," 지능정보연구, 제22권, 제1호, 2016, pp. 63-82.

김건우, 박도형, "소비자 가치기반 디자인 평가 모형," 지능정보연구, 제23권, 제4호, 2017, pp. 57-76.

김문구, 박종현, "국내 ICT 중소기업의 성장요인과 글로벌 히든 챔피언 육성방향," Electronics and Telecommunications Trends. 제30권, 제2호, 2015, pp. 79-86.

김범수, 최유지, 박도형, "웹검색트래픽 정보를 활용한 투자모형 개발: KOSPI 지수를 중심으로," Entrue Journal of Information Technology, 제14권, 제3호, 2015, pp. 63-81.

김문선, 최영득, 박희정, 천새롬, "중소기업 핵심기술 유출실태에 관한 연구," 한국경영학회 통합학술발표논문집, 2007, pp. 1-10.

노민선, 이삼열, "중소기업의 산업보안 역량에

대한 영향요인 평가," 한국행정학보, 제44권, 제3호, 2010, pp. 239-259.

문제욱, 나영섭, 장항배, "중소기업 맞춤형 보안환경 구축을 위한 정보화 발전모델 심층연구," 한국인터넷정보학회 학술발표대회 논문집, 2013, pp. 23-24.

문현정, "우리나라 중소기업의 정보 보호 역량 강화를 위한 교육 훈련 현황과 문제점," 정보보호학회지, 제19권, 제1호, 2009, pp. 29-39.

문호건, 박성철, "기업보안 강화를 위한 취약점 진단 통합관리 체계 구축," 한국통신학회지(정보와통신), 제31권, 제5호, 2014, pp. 39-45.

미래창조과학부, 한국인터넷진흥원, "중견기업 성장사다리 구축방안," 2013, pp. 5-15.

박도형, "여행자 관심 기반 스마트 여행 수요 예측 모형 개발: 웹검색 트래픽 정보를 중심으로," 정보시스템연구, 제26권, 제3호, 2017, pp. 171-185.

박도형, 정재권, 정여진, 이동원, "자기조직화 지도를 활용한 성장모형 기반의 시장 성장패턴 지도 구축," 지능정보연구, 제20권, 제4호, 2014, pp. 1-23.

박향미, 유지연, "중소기업 산업보안 강화를 위한 한국과 미국의 관리체계 비교・분석 연구", 한국사회안전학회지, 제10권, 제2호, 2015, pp. 119-140.

서봉군, 박도형, "중소기업 기술 유출에 대한 조기경보시스템 개발에 대한 연구," 지능정보연구, 제23권, 제1호, 2017, pp. 143-159.

이강백, 김태환, 이상용, "기업보안투자가 기업성과에 끼치는 영향," 한국경영정보학회 학술대회, 2015, pp. 354-359.

이동원, 강태영, 박도형, "소비자 구매의사결정 단계에 따른 PC 기반 쇼핑과 모바일 기기 기반 쇼핑 행태 차이에 관한 연구: 과업 기술 적합 이론을 중심으로," Entrue Journal of Information Technology, 제13권, 제3호, 2014, pp. 107-122.

이동원, 정여진, 정재권, 박도형, "신규시장 성장모형의 모수 추정을 위한 전문가 시스템," 지능정보연구, 제21권, 제4호, 2015, pp. 17-35.

이치석, 김양훈, "보안교육과 보안관리 역량의 상관관계 분석," 한국전자거래학회지, 제20권, 제1호, 2015, pp. 27-36.

이희명, 임종인, "기업의 정보보호수준 측정모델 개발에 관한 연구," 정보보호학회논문지, 제18권, 제5호, 2008, pp. 161-170.

전승표, 박도형, "웹검색 트래픽 정보를 활용한 지능형 브랜드 포지셔닝 시스템," 지능정보연구, 제19권, 제3호, 2013, pp. 93-111.

전승표, 정재웅, 최산, "데이터마이닝을 이용한 표준정책 수요 중소기업의 프로파일링 연구," 기술혁신학회지 제19권, 제3호, 2016, pp. 511-544.

정병수, 류상일, 김화수, "산업보안의 연구경향 분석-학술연구정보서비스 (2000년~2011년)를 중심으로," 한국치안행정논집, 제9권, 제2호, 2012, pp. 195-215.

정재권, 박도형, 이동원, 정여진, "A study on

the applicability of a growth curve model for the Korean theater industry: A case of 'Why? π!'," 한국경영공학회지, 제19권, 제4호, 2014, pp. 95-107.

채정우, 고영희, "전문경영인의 기업정보 보호를 위한 산업기술 유출요인과 대응전략에 대한 탐색적 사례연구," 전문경영인연구, 제15권, 제1호, 2012, 87-113.

최유지, 박도형, "웹검색 트래픽 정보를 활용한 유커 인바운드 여행 수요 예측 모형 및 유커마이닝 시스템 개발," 지능정보연구, 제23권, 제3호, 2017, pp. 155-175.

허승표, 이대성, 김귀남, "군집화를 이용한 기업 핵심기술 유출자 분류에 관한 연구," 융합보안논문지, 제10권, 제3호, 2010, pp. 23-28.

홍준석, 박원형, 김양훈, 국광호, "로지스틱 회귀분석을 이용한 중소기업 기술보호 요인 분석," 한국전자거래학회지, 제20권, 제3호, 2015, pp. 1-10.

홍준석, 임영환, 박원형, 국광호, "중소기업 유해트래픽 분석을 통한 보안관제 개선방안," 한국전자거래학회지, 제19권, 제4호, 2014, pp. 195-204.

홍지훈, 박도형, "네트워크 효과에 따른 모바일 컨텐츠 플랫폼의 경쟁 전략에 관한 연구," e-비즈니스연구, 제14권, 제5호, 2013, pp. 113-130.

황현동, 이창무, "산업기술유출과 자기통제력, 조직애착도의 관계에 관한 연구," 한국경호경비학회지, 제47권, 2016, pp. 119-137.

Conklin, J. E., "Illegal but not criminal: Business crime in America," Englewood Cliffs, NJ: Prentice-Hall, 1977.

Jun, S.-P., and Park, D.-H., "Visualization of brand positioning based on consumer web search information: Using social network analysis," *Internet Research*, Vol. 27, No. 2, 2017, pp. 381-407.

Jun, S.-P., and Park, D.-H., "Consumer information search behavior and purchasing decisions: Empirical evidence from Korea," *Technological Forecasting and Social Change*, Vol. 107, 2016, pp. 97-111.

Kang, T., Park, D.-H., and Han, I. "Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics," *Telematics and Informatics*, Vol. 35, No. 2, 2018, pp. 370-381.

Pyle, Dorian. "Putting data mining in its place," *Database Programming and Design*, Vol. 11, 1998, pp. 32-37.

유 인 진 (Yoo, In-Jin)

국민대학교 경영정보학부에서 학사 학위를 취득하였으며, 현재 국민대학교 비즈니스 IT전문대학원에서 CXLab.에 소속되어 Customer Experience, Business Analytics 트랙으로 석사과정에 재학 중이다. 주요 관심 분야는 Customer Analytics로 SME, R&D, Trading area에 대한 정량적 분석과 함께 사용자 감성에 기반한 정성적 분석 등을 수행하고 있다.

박 도 형 (Park, Do-Hyung)

KAIST 경영대학원에서 MIS 전공으로 석사/ 박사학위를 취득하였다. 현재 국민대학교 경영대학 경영정보학부/ 비즈니스 IT 전문대학원 부교수로 재직 중이며, 고객경험연구실(CXLab.)을 책임지고 있다(www.cxlab.co.kr). 현재 주요 관심분야는 사회심리학 기반의 소비자 행동 이론(Customer Behavior Theory), 통계 및 인공지능 기법 기반의 소비자 애널리틱스(Customer Analytics), 디자인사고(Design Thinking) 기반의 소비자 경험 디자인(Customer Experience Design)이다. 한국과학기술정보연구원(KISTI)에서 유망아이템 발굴, 기술가치 평가 및 로드맵 수립, 빅데이터 분석 등을 수행하였고, LG전자에서 통계, 시선/뇌파 분석, 데이터 마이닝을 활용한 소비자 평가 모형 개발을 담당했었고, 스마트폰, 스마트TV, 스마트Car 등에 대한 Technology, Business, Market Insight 기반 컨셉 도출 프로젝트를 다수 수행하였다.

<Abstract>

A Study on Empirical Model for the Prevention and Protection of Technology Leakage through SME Profiling Analysis

Yoo, In-Jin · Park, Do-Hyung

Purpose

Corporate technology leakage is not only monetary loss, but also has a negative impact on the corporate image and further deteriorates sustainable growth. In particular, since SMEs are highly dependent on core technologies compared to large corporations, loss of technology leakage threatens corporate survival. Therefore, it is important for SMEs to "prevent and protect technology leakage". With the recent development of data analysis technology and the opening of public data, it has become possible to discover and proactively detect companies with a high probability of technology leakage based on actual company data. In this study, we try to construct profiles of enterprises with and without technology leakage experience through profiling analysis using data mining techniques. Furthermore, based on this, we propose a classification model that distinguishes companies that are likely to leak technology.

Design/methodology/approach

This study tries to develop the empirical model for prevention and protection of technology leakage through profiling method which analyzes each SME from the viewpoint of individual. Based on the previous research, we tried to classify many characteristics of SMEs into six categories and to identify the factors influencing the technology leakage of SMEs from the enterprise point of view. Specifically, we divided the 29 SME characteristics into the following six categories: 'firm characteristics', 'organizational characteristics', 'technical characteristics', 'relational characteristics', 'financial characteristics', and 'enterprise core competencies'. Each characteristic was extracted from the questionnaire data of 'Survey of Small and Medium Enterprises Technology' carried out annually by the Government of the Republic of Korea. Since the number of SMEs with experience of technology leakage in questionnaire data was significantly smaller than the other, we made a 1: 1 correspondence with each sample through mixed sampling. We conducted profiling of

companies with and without technology leakage experience using decision-tree technique for research data, and derived meaningful variables that can distinguish the two. Then, empirical model for prevention and protection of technology leakage was developed through discriminant analysis and logistic regression analysis.

Findings

Profiling analysis shows that technology novelty, enterprise technology group, number of intellectual property registrations, product life cycle, technology development infrastructure level(absence of dedicated organization), enterprise core competency(design) and enterprise core competency(process design) help us find SME's technology leakage. We developed the two empirical model for prevention and protection of technology leakage in SMEs using discriminant analysis and logistic regression analysis, and each hit ratio is 65%(discriminant analysis) and 67%(logistic regression analysis).

Keyword: SME, Technology Leakage, Profiling Analysis, Data Mining, ICT

* 이 논문은 2018년 3월 9일 접수, 2018년 3월 22일 1차 심사, 2018년 3월 28일 게재 확정되었습니다.