

빅데이터 분류 기법에 따른 벤처 기업의 성장 단계별 차이 분석

정 병 호*

The Difference Analysis between Maturity Stages of Venture Firms by Classification Techniques of Big Data

Jung Byoung-ho

〈Abstract〉

The purpose of this study is to identify the maturity stages of venture firms through classification analysis, which is widely used as a big data technique. Venture companies should develop a competitive advantage in the market. And the maturity stage of a company can be classified into five stages. I will analyze a difference in the growth stage of venture firms between the survey response and the statistical classification methods. The firm growth level distinguished five stages and was divided into the period of start-up and declines. A classification method of big data uses popularly k-mean cluster analysis, hierarchical cluster analysis, artificial neural network, and decision tree analysis.

I used variables that asset increase, capital increase, sales increase, operating profit increase, R&D investment increase, operation period and retirement number. The research results, each big data analysis technique showed a large difference of samples sized in the group. In particular, the decision tree and neural networks' methods were classified as three groups rather than five groups. The groups size of all classification analysis was all different by the big data analysis methods. Furthermore, according to the variables' selection and the sample size may be dissimilar results. Also, each classed group showed a number of competitive differences.

The research implication is that an analysts need to interpret statistics through management theory in order to interpret classification of big data results correctly. In addition, the choice of classification analysis should be determined by considering not only management theory but also practical experience. Finally, the growth of venture firms needs to be examined by time-series analysis and closely monitored by individual firms. And, future research will need to include significant variables of the company's maturity stages.

Key Words : Big Data, Data Mining, Venture Firm, Firm Maturity Stage, Classification Methods

* 한국외국어대학교 경영정보학과 외래교수

I. 서론

통계청에 의하면 국내 벤처기업의 수는 1999년도 2,042개에서 2017년 35,282 대폭 증가하였으며, 2019년도 현재 37,044개로 창업 및 유지되고 있다고 한다. 국내 산업 경제 환경은 아직까지 대기업 위주의 안정 추구형 경영을 진행하다보니 성장 잠재력이 낮아지고 있는 상황이다. 그리고 4차 산업혁명으로 인한 글로벌 환경의 급진적 변화, 산업의 융·복합으로 새로운 산업 영역을 개척할 필요성이 생겼다[1]. 국내 시장의 더욱 건전한 산업 생태계를 유지하기 위해서 카카오, 배달의 민족과 같은 혁신적인 벤처기업 육성이 필요하다는 목소리가 높아져가고 있는 실정이다. 벤처 기업과 스타트업 기업들은 초기 시장의 안착이 매우 중요하며, 이를 실패할 경우 비즈니스뿐만 아니라 기술까지 시장에서 도태되거나 사라지는 경우가 발생할 수 있다[2].

한편, 4차 산업혁명에서 경쟁우위를 갖추고자 많은 기업들은 새로운 시장 진출과 마케팅, 내부 역량 파악을 위해서 빅데이터 수집 및 분석에 몰입하고 있다[3]. 기업들은 기존에 인지적으로 비즈니스를 수행하던 조직 행동들을 데이터 기반의 조직 행동으로 변화시키고, 비즈니스 위험도를 낮추기 위한 시도로서 진행되고 있다[4]. 특히, 최근의 벤처기업들은 데이터를 바탕으로 새로운 시장 진출, 고객 맞춤형 마케팅, 지역별 제품/서비스를 제시하기 위해 분류 분석을 실시하고 기업에게 유리한 비즈니스를 선택하고 있다. 분류 분석은 기업 내부의 고객과 신 시장을 탐색하는데 목적을 가진다[5].

이렇듯, 분류분석은 기업의 미래 가치를 높이기 위해서 충성고객과 이탈고객, 신 시장진출의 가능성 분석에 도움을 제공하기 때문에 기업의 미래성장 가능 상태를 판단하는데 매우 중요한 척도가 된다. 하지만 분류분석의 종류는 두 집단과 세 집단 이상으로 구분하였을 경우 분석 기법이 달라지며, 데이터 척도에

따라서도 분류 기법 선택은 다양하다. 즉, 어떤 분류 기법이 빅데이터 분석으로 더 나은지는 판단하기 어렵다. 이에 빅데이터의 분류 기법을 선택하는데 있어 분석 시나리오와 분류 결과의 해석이 매우 중요할 수 있겠다.

그래서 본 연구에서는 빅데이터의 여러 분류 기법을 통해 벤처기업들의 창업기부터 쇠퇴기까지 분류 분석하고자 한다. 그리고, 벤처기업들이 응답한 기업 현 성장 상태와 빅데이터 기법으로 분류된 결과를 비교할 것이다. 또한, 경쟁력 분석에서 기업 성장별 집단 차이가 있는지도 연구하고자 한다.

II. 관련 연구

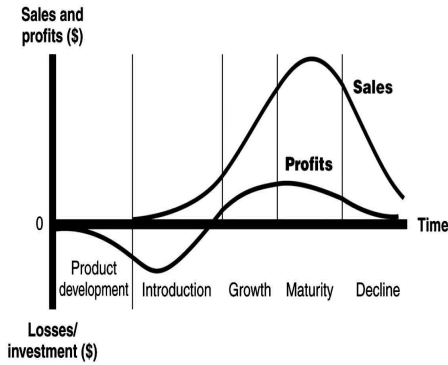
2.1 벤처기업과 기업의 성장 단계

벤처기업이란 벤처(venture)와 기업(company)의 합성어로 모험적 사업 또는 금전상의 위험을 감수한 사업을 의미한다[6]. 미국에서는 다른 기업보다 상대적으로 사업 위험성은 높으나 성공하면 높은 수익의 기회를 가진 기업이나 벤처캐피탈에서 투자를 받은 기업을 벤처기업으로 보고 있다. 그러나 보통 신사업, 기술집약 기업, 첨단기술 기업 등에서 벤처기업의 용어를 사용하고 있다.

벤처기업협회에 따르면 국내 관점에서 기존 기업에 비해 기술성과 성장성이 상대적으로 높지만, 정부 지원이 필요하다고 인정하는 기업으로서 '벤처기업 육성에 관한 특별조치법'의 세 가지 중 한 가지를 만족하는 기업을 의미한다. 그리고 국내에서 벤처확인요건은 한국벤처캐피탈협회, 기술보증기금, 중소벤처기업진흥공단에서 지정한 요건에 부합되면 벤처기업으로 등록이 된다.

벤처기업은 새로운 첨단 제품 및 서비스의 아이디어를 사업화를 실현시키고자 하는 목적을 가지고 탄

생되는 기업으로서 초기 시장의 안착과 시간이 지남에 따른 안정화가 매우 중요한 요소가 되겠다[6].



<그림 1> 제품 및 기업 수명주기

출처: <https://marketing-insider.eu/marketing-explained/part-iii-designing-a-customer-driven-marketing-strategy-and-mix/product-life-cycle-stages>

한편, 기업의 많은 제품과 서비스는 그 수요가 시간이 흐르면서 제품의 수명주기에 따라 변한다[7]. 한 제품이 개발되어 테스트되고 나면 시장에 출시되며, 제품들은 도입, 성장, 성숙, 쇠퇴 등의 4~5단계를 거치게 된다[8]. 기술 기반의 벤처기업 성장에서는 기술 개발, 제품 상업화, 시장 성장, 시장 안정화로 구분하며, 벤처 기업의 성장을 기술 개발과 자원의 성취를 1 단계로서 제품 상업화 2단계, 시장 확장과 조직적 이슈가 3단계, 수입과 미래 성장 기반의 안정화를 4단계로 보기도 한다[9].

일반적으로 작은 기업에서 성장 단계는 도입에서 쇠퇴기까지 구분하여 제품수명주기와 기업수명주기를 판단하고 있다[10]. 우선, 신제품이 시장에 등장하면 사람들이 호기심을 보일지는 모르지만 일반적으로 수요가 낮다. 창업기는 회사를 창업하고, 제품을 개발하는 단계를 말하고 초기 성장기는 신규 제품이 출시되어 매출이 발생하는 단계로서 초기 수요 예측이 중요한 시기이다. 고도 성장기는 후속적 신규의 제품이 출시되어 제품과 시장이 다각화되고 매출은

증폭되는 단계를 말한다. 성숙기는 경쟁 심화와 매출과 시장은 포화되어 성장이 둔화된 단계로서 마케팅을 강화가 요구된다. 쇠퇴기는 매출이 급락하고, 기업의 활동은 정체 또는 철수가 고려된 단계를 말한다. 이는 시장을 포기하거나 제품의 새로운 용도나 새로운 발굴 시도가 요구되는 시장이다[7-10].

2.2 벤처 기업의 경쟁우위 및 시장 성장성

전략은 생존에 중요한 역할을 하는 것으로 삶과 죽음의 문제이기도 하며, 안전과 존망에 영향을 미친다[11]. 그래서 기업은 장기적인 목표의 결정과 그 목표를 달성하기 위한 행동을 결정하고 경영 자원을 배분해야 한다고 설명하고 있다[12]. 전략은 경쟁자에 비해서 경쟁우위를 가질 수 있는지에 대한 문제를 설명하고 있으며 기업은 내부에서 경영자원, 핵심역량, 조직구조, 기업 문화, 리더십을 중요시 여겨 체계화시킬 필요가 있다[11]. 외부적으로는 경쟁사, 고객, 공급사의 행동과 경쟁우위를 예측할 필요가 있겠다[4].

벤처 기업은 초기 시장에서 신기술 제품의 관심을 촉발시켜서 신기술애호가와 혁신 수용층에 혁신제품의 관심을 일으켜야 한다. 또한 몇몇 부분적으로 틈새 시장을 파고들어서 틈새시장 간의 연속적 파급효과 극대화를 할 필요가 있다. 시장에서 성공하기 위해서 기업이 정신, 인적 자본, 재무(금융) 자본, 지리적 위치, 전략, 산업 현황, 조직 구조와 시스템 등이 고려되어야 한다[13]. 이들 요소는 각 성장단계별로 특징적 위치를 가지는데 예를 들어 자금의 경우 도입과 초기 성장기에는 부정적 위치에 있지만 성숙기로 이동하면서 긍정적 상황으로 변화된다[10]. 또한, 인적자본의 경우 도입과 초기 성장기에 조직이 비구조적인 형태를 가지면서 인적자본의 안정성이 낮지만 성장기로 넘어가면서 안정화된 형태로 변화되기 한다. 이렇듯, 벤처 기업의 탄생과 쇠퇴까지 자원별, 조직 및 산업 상황별로 상이한 특징을 가지게 된다[13].

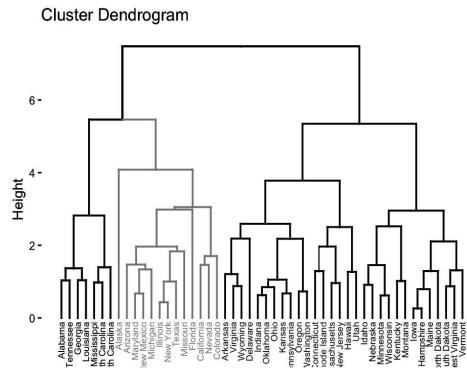
2.3 빅데이터 분류분석의 기법 종류

빅데이터는 기업 비즈니스와 사회적 이슈 등 다양한 분야에서 미래를 예측하고 분류하기 위한 목적으로 활용되고 있다[14]. 빅데이터는 문자, 음성, 이미지, 동영상, 숫자 등의 정형화와 비정형화된 데이터를 가지고 비즈니스 현안과 미래를 판단하는데 분석되어지고 있다[15-16]. 특히 예측과 분류 분석은 데이터의 유형에 따라 기법의 선택이 달라지며, 분류 분석의 종류로는 의사결정트리, 군집분석, 인공신경망, 로지스틱 회귀분석 등이 활용되며 예측 분석으로는 회귀분석, 인공신경망 등이 활용된다[14]. 또한, 사회 연결망 및 연관 규칙 등도 비즈니스에서 주로 활용되고 있다[5]. 이렇듯, 빅데이터로서 데이터 과학은 통계와 정보, 컴퓨팅, 커뮤니케이션, 사회화과 결합되어 미래의 부가가치 창출에 도움을 제공한다[17-18].

분류분석이란 분류 결과가 알려진 과거 데이터로부터 분류별 특성을 찾아내어 분류 모형 및 규칙들을 제시해 주는데 목적이 있다[14]. 분류 규칙들은 분류 결과가 알려지지 않은 새로운 데이터에 적용하여 분류값을 예측하는 것이다[5]. 우선, 군집분석은 개인이나 대상물을 군집으로 분류함으로써 같은 군집 내에서의 대상물들이 여러 군집에 속해 있는 대상물들에 비하여 매우 유사해진다. 군집 간의 이질성을 극대화하면서도 군집 내에서의 대상물들의 동질성을 극대화하는데 있다.

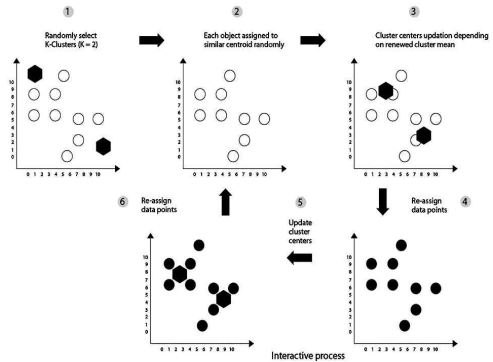
계층적 군집분석이란 나무의 구조와 같이 하나의 계층을 만드는 방법이다. 계층적 군집분석은 응집과 분열 방법을 사용된다. 응집 방법은 각 대상물을 처음부터 하나의 군집을 형성하는 것에서 출발하는 것이며 분열 방법은 가장 유사하지 않은 관찰 결과들을 분할하여 보다 적은 규모의 군집을 만드는 방식을 말한다[19]. 이에 대한 예시는 <그림 2>에 제시하였다.

비계층적 군집분석인 K-평균 군집분석은 계층적 군집과 달리 나뉘어가지와 같은 모양으로 군집을 형성



<그림 2> 계층적 군집분석 예시

출처: <https://www.datanovia.com/en/lessons/divisive-hierarchical-clustering/>



<그림 3> k-평균 군집분석 예시

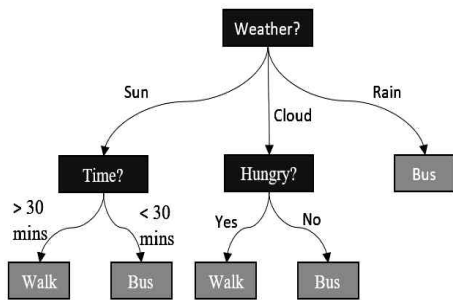
출처: <https://www.brandidea.com/kmeans.html>

하는 것이 아니다. 그 대신 형성될 군집 수가 일단 명시되면 대상물들을 군집에 할당된다[5].

K-평균 군집분석은 주어진 데이터를 K개의 클러스터로 묶는 알고리즘이다. 각 클러스터와 거리 차이의 분산을 최소화시키는 방식으로 동작한다. 이러한 알고리즘은 자율 학습의 한 형태로서 레이블이 없는 입력 데이터에 레이블을 설정하는 역할을 수행하게 된다[19]. 이에 대한 예시는 <그림 3>에 제시하였다.

한편, 의사결정 트리는 의사결정 규칙과 그 결과들을 트리 구조로 도식화한 의사결정 지원 도구의 일종이다. 의사결정 트리 학습법은 지도 분류 학습에 많

이 사용되고 있는 기법으로 인기 기법 중 하나이다 [5]. 의사결정 트리는 시각적이고 명시적인 방법으로 의사 결정 과정과 결정된 의사를 보여주는 데 사용된다. 데이터 마이닝 분야에서 결정 트리는 결정된 의사보다는 자료 자체를 표현하는데 유용하다[5]. 의사 결정 트리에 대한 예시는 <그림 4>에 제시하였다.



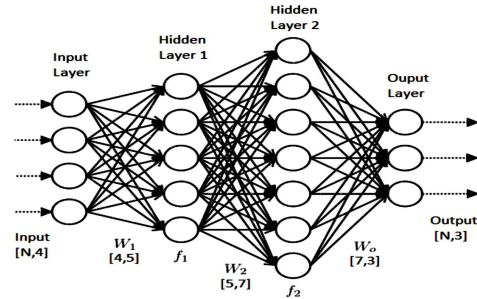
<그림 4> 의사결정트리 예시

출처: <https://www.displayr.com/what-is-a-decision-tree/>

의사결정 트리는 연구자가 설정한 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델이다[14]. 이는 통계학과 데이터 마이닝, 기계 학습에서 사용하는 예측 모델링 방법 중 하나이다. 의사결정 트리 모델 중 목표 변수가 유한한 수를 분류 트리라 한다. 이 트리 구조에서 잎은 클래스 라벨을 나타내며, 가지는 클래스 레벨과 관련 있는 특징들을 나타낸다. 의사결정 트리 중 목표 변수가 연속하는 값을 가지는 것을 회귀 트리라 한다.

인공신경망은 인간의 뇌가 패턴을 인식하는 방식을 모방한 알고리즘이다[14]. 인공 신경망은 시각, 청각 입력 데이터를 퍼셉트론이나 분류, 군집을 이용하여 해석하며, 해석 결과를 이용하여 이미지, 소리, 문자, 시계열 데이터에서 특정 패턴을 해석한다[5].

인공신경망은 기계학습 및 인지과학에서 생물학의 신경망의 영감을 얻어지는 통계적 학습 알고리즘이다. 인공신경망을 이용하여 각종 분류 및 군집화가 가능하다. 분류나 군집화를 원하는 데이터 위에 여러



<그림 5> 인공신경망 예시

출처: <https://medium.com/coinmonks/the-artificial-neural-networks-handbook-part-1-f9ceb0e376b4>

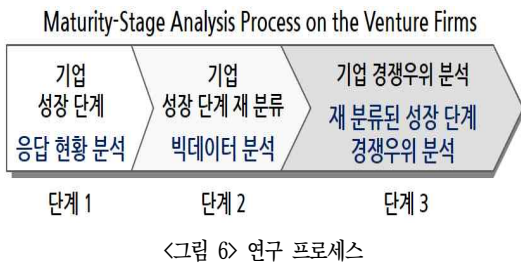
개의 레이어를 구성하여 작업을 실시하게 된다. 각 레이어에서 라벨링이 되지 않은 데이터를 서로 비교하여 유사도를 구해주고, 라벨링이 되어 있는 데이터를 기반으로 분류기를 학습하여 자동으로 데이터를 분류하도록 할 수 있다. 이러한 인공신경망은 시냅스 결합을 통해 네트워크를 형성한 인공 뉴런이 학습을 하고, 시냅스의 결합의 세기를 변화시켜, 문제 해결 능력을 보여주는 모델이라고 본다[5]. 이에 대한 예시는 <그림 5>에 제시하였다.

III. 연구 방법론

본 연구는 2017년도 중소벤처기업부에서 발간한 ‘2017년 벤처기업정밀실태조사’의 데이터를 기초로 하고 있다[20]. 이 조사는 국내 벤처 기업을 대상으로 벤처기업 운영 상황에 대한 표본 조사를 위해서 진행되었다. 설문조사 기간은 2016년에 진행되었고, 총 수집된 표본 수는 2,114개이다. 본 데이터는 중소벤처기업부에서 매년 조사하고 있지만 2018년 데이터는 아직 미 공개되어 있고, 벤처기업 현 경쟁력 상태의 비교하는데 1년 차이가 크지 않다고 판단하여 2017년 실태조사 데이터를 활용하였다.

이 데이터는 벤처기업들의 경쟁력 변수와 데이터 수가 상당히 많다. 하지만, 본 연구에서는 기업이 산업에서 성장하는데 설명될 수 있는 인적자본, 재무자본, 연구개발 등의 변수들만 선정하여 분석을 수행하였다.

서울이 480개(22.71%)로 가장 높게 나타났다. 벤처유형은 기술평가보증기업으로서 1697개(80.27%)로 높게 나타났다. 제조 및 비제조 유형에서는 일부 제조와 아웃소싱을 겸임하는 형태로 844개(39.92%)로 높게 나타났다. 이에 대한 세부적인 분석 결과는 <표 1>에 제시하였다.



특히, 본 데이터에는 기업 성장 단계를 확인할 수 있는 기업 성장 단계, 직원 수, 연구개발, 재무 데이터 등이 다수 포함되어 있다. 이 변수들은 기업 성장 단계를 판단하고 해석하는데 용이하다[7-10, 13]. 설문 변수를 토대로 벤처 기업의 성장단계를 응답자가 직접 응답한 변수와 경영 이론에서 설명하는 성장 단계를 변수들을 통해 분류 분석을 실시하여 응답과 분류 분석 간의 비교하고 차이를 확인할 것이다.

그래서 우선적으로 인지된 기업의 성장단계를 분석하고자 한다. 그 다음 분류 분석을 통해 성장 단계를 분석할 것이다. 마지막으로 분류된 집단을 통해 경쟁력의 차이를 확인하고자 한다. 이러한 내용은 <그림 1> 연구 프로세스에 표현하였다.

<표 1> 표본의 특성

	구분	빈도	구성 비율(%)
지역	강원	63	2.98%
	경기	579	27.39%
	경남	96	4.54%
	경북	77	3.64%
	광주	49	2.32%
	대구	82	3.88%
	대전	166	7.85%
	부산	106	5.01%
	서울	480	22.71%
	세종	8	0.38%
	울산	36	1.70%
	인천	72	3.41%
	전남	30	1.42%
	전북	25	1.18%
	제주	5	0.24%
	충남	112	5.30%
충북	128	6.05%	
벤처 유형	기술평가대출기업	127	6.01%
	기술평가보증기업	1697	80.27%
	벤처투자기업	96	4.54%
	연구개발기업	194	9.18%
제품 생산	자체 제조	716	33.87%
	일부제조+아웃소싱	844	39.92%
	전체 아웃소싱	64	3.03%
	비제조기업	490	23.18%

IV. 연구 결과

4.1 표본의 특성

본 연구의 표본 특성으로 인구통계학을 분석하였다. 분석 결과, 응답 비율의 내용을 살펴보면, 지역은

4.2 단계 1 분석 : 응답한 기업 성장 단계

본 연구에서는 벤처 기업 성장을 판단하기 위한 변수로서 자산 증감, 자본 증감, 매출액 증감, 영업이익 증감, 연구개발투자 증감, 운영기간, 퇴직자 변수를 활용하였다. 본 분류 분석을 위해서 앞서 7개 변수 이외에도 당기순이익, 충원필요인원, 제품개발단계 등의 3가지 변수는 의사결정트리와 인공신경망에서 유

의미한 변수로 나타나지 않아 분석에서 제외하였다.

여러 분류 분석을 실시하고자 할 때 한 번에 많은 다양한 변수를 활용하는데 제약적 상황이 발생하게 되었다. 이에 기업 성장 단계 분석에서 필요한 변수 중 설명력이 나타난 변수만 선정하여 분석하게 되었다. 그리고 기업 성장 단계별로 고려할 수 있는 변수는 다양하며 산업별, 기업별로 성장 변수는 무궁무진하다. 즉, 산업 성장, 마케팅, 시장 개척, 조직 확장, M&A, 법·제도 등의 여러 변수를 포함하여 기업 성장을 분석해야 하지만 모든 외생변수를 고려하는데 무리가 발생한다. 그럼에도 불구하고, 재무 및 인적자원, 연구개발 변수는 기업성장 분석에 필수적 변수이므로 이 일곱 가지 변수를 통해서 분류 분석에 실시하였다. 이 변수의 내용은 <표 2>에 제시하였다.

<표 2> 변수의 설명

No	변수명	의미	
1	자산 증감	2016년 자산 - 2015년 자산	(기준: 백만원)
2	자본 증감	2016년 자본 - 2015년 자본	(기준: 백만원)
3	매출액 증감	2016년 자본 - 2015년 자본	(기준: 백만원)
4	영업이익 증감	2016년 영업이익 - 2015년 영업이익	(기준: 백만원)
5	연구개발 투자증감	2016년 연구개발비 - 2015년 연구개발비	(기준: 백만원)
6	운영기간	2016년 - 창업연도	연도
7	퇴직자	2016년 퇴사자 (임원급+중간관리자+사원급의 모든 퇴직자)	인원수

우선적으로 설문 응답자들이 응답한 자신의 벤처 기업이 성장 단계가 어느 정도인지를 분석하였다. 분석은 평균을 토대로 비교하였으며 <표 3>에 제시하였다.

내용을 살펴보면 창업기는 23, 초기 성장기는 338,

고도 성장기는 889, 성숙기는 841, 쇠퇴기는 23개로 나타났다. 특히, 초기 성장기부터 고도 성장기에 매출액과 영업이익, 자산의 증가는 매우 높은 수치로 높아지고 있는 형태를 보이고 있다. 이러한 형태는 성숙기에도 나타나고 있으며 영업이익이 감소되고 퇴직자가 다른 성장기에 비해 높게 나타나고 있다. 또한 운영 기간은 창업기부터 쇠퇴기까지 연차가 높아지는 것을 확인할 수 있겠다.

<표 3> 인지된 기업성장 단계 현황

성장단계	창업기	초기 성장	고도 성장	성숙기	쇠퇴기
표본수	23	338	889	841	23
자산증감	404.26	1231.7	1942.2	1955.2	-16.30
자본증감	131.26	626.38	944.19	1209.81	223.21
매출액 증감	179.78	1429.2	1340.2	806.63	-220.78
영업이익 증감	-39.82	33.54	112.44	50.77	-338.08
연구개발 투자증감	587.08	356.21	728.60	1233.77	534.60
운영기간	3.043	6.16	11.46	17.71	22.04
퇴직자	-1.69	-4.60	-5.24	-7.65	-5.73

4.3 단계 2 분석 : 기업 성장 단계 재분류

다음은 창업기, 초기 성장기, 고도 성장기, 성숙기, 쇠퇴기를 분류하기 위해서 군집분석을 실시하였다. 우선 K-평균 군집분석은 5개의 군집의 형태로 분류를 실시하였으며 반복계산은 10으로 하였다.

K-평균 군집분석은 비지도 학습방법으로 데이터마이닝과 빅데이터 방법론으로 자주 활용되며[5], 새로운 집단을 분류하고, 집단의 특성을 파악하기 위해서 주로 활용된다. K-평균 군집분석에서는 자산증감부터 퇴직자 변수까지 모형 분석에 적합하다고 나타났다.

분류된 그룹 형태를 보면, 1909, 8, 9, 183, 5개로 다섯 개의 집단으로 분류되었고, 군집 결과의 그룹 내

샘플 수는 극단적인 형태를 보이고 있다. 분류 집단은 제품 및 기업 성장별 단계 이론에서 설명하는 기업 성장별 나타나는 특징에 따라 분류를 적용하였다. 분류된 결과내용을 평균값으로 비교 살펴보면, 자산 증감부터 영업이익 급감, 운영 기간이 높은 5개 샘플은 쇠퇴기로 지정하였다. 1909개의 집단은 매출액은 없지만 자산, 자본, 연구개발투자에 일정부분 투자를 진행하고 있고, 운영 기간이 짧은 집단을 창업기로 지정하였다. 그리고 매출액과 영업이익, 자본, 자산이 가장 크고, 연구개발도 일정부분 진행하고 있는 집단을 고도 성장기로 분석하였다. 성숙기는 매출액과 영업이익이 급감하고, 운영기간이 오래되고 퇴직자가 있는 집단을 성숙기로 지정하였다. 초기 성장기는 창업기에 비해 매출액, 영업이익, 연구개발투자가 급진적으로 성장하는 집단으로 선정하였다. 이 분류의 형태는 <표 4>에 제시하였다.

<표 4> K-평균 군집 결과

성장단계	창업기	초기성장	고도성장	성숙기	쇠퇴기
군집수	1909	8	9	183	5
자산 증감	423.1	74285.3	32941.44	11632.1	-6178.6
자본 증감	263.31	48463	21497.11	5648.6	-8164.4
매출액증감	-42.74	30797.1	70323.33	10554.2	-75419
영업이익 증감	-114.24	8560.13	15337.33	1257.81	-14676
연구개발 투자 증감	572.87	8689.75	6297.56	3268.16	2732.8
운영기간	12.88	18.38	12.67	15.37	17.8
퇴직자	-5.41	-10.5	-7.33	-12.13	-24.2

즉, 이러한 분류 방식은 기업 성장 단계 또는 제품 수명 주기의 이론에서 설명되는 단계별 고려사항을 참고하여 분류를 실시하였다. 많은 변수를 고려하지 않은 상황에서 여러 외생 변수가 있을 수 있지만 분

석에 적용한 변수를 토대로 이론을 비교·일치시켜 분류하였다.

다음은 계층적 군집분석으로 군집 계산 방법은 Ward 방법으로 제곱 유클리디안 거리로 계산하였다. 계층적 군집은 거리가 가장 가까운 것 2개를 묶는 방법으로 반복적으로 시행되고, 하나의 군집으로 합쳐질 때까지 진행되는 특징을 가진다. Ward 방법은 군집 평균과 군집 내 유클리디언 최소 증가 방식으로 군집간의 거리에 따라 데이터들을 연결하기 보다는 군집 내 편차들의 제곱합에 근거를 두고 군집들을 병합시키는 방법이다. 이 방법은 군집 분석의 각 단계에서 데이터들을 하나의 군집으로 묶어서 생기는 정보 손실을 군집 평균과 데이터들 사이의 오차 제곱합으로 측정한다[19].

<표 5> 계층적 군집분석 결과

성장 단계	창업	초기 성장	고도 성장	성숙기	쇠퇴기
군집수	1968	86	11	44	5
자산 증감	698.39	8913.18	68442.3	21211.5	-6178.60
자본 증감	373.568	4431.30	44269.1	11688.5	-8164.40
매출액증감	132.660	20718.1	50948.2	2860.50	-75419.6
영업이익 증감	-94.167	3102.65	15967.8	-875.796	-14676.6
연구개발 투자 증감	627.116	2054.93	9802.27	6799.00	2732.80
운영기간	12.9746	15.0930	17.3636	14.6136	17.8000
퇴직자	-5.5254	-11.627	-10.8182	-16.0909	-24.2000

<표 5> 계층적 군집 분석 결과를 평균값으로 살펴보면, k-평균 군집분석처럼 창업기에 표본 샘플이 많이 집중되었으며, 쇠퇴기의 그룹 샘플은 동일하게 나타났다. 차이점은, 초기 성장기와 성숙기의 집단 샘플이 상이한 차이를 보이는 형태를 나타냈다. 계층적 군집 분석에서도 마찬가지로 k-평균 군집분석과 동일

한 방법으로 제품수명주기 또는 기업 성장 단계의 이론을 토대로 집단을 5개로 분류를 실시하였다.

다음은 의사결정트리 분석을 실시하였다. 의사결정 트리는 지도학습 방법으로 마케팅에서 핵심고객과 이탈고객을 찾는 데 주로 활용된다. 의사결정 트리는 단순히 예측/분류 하나의 문제를 풀 수 있는 것이 아니라 예측/분류 두 가지 문제를 풀 수 있습니다 [5]. 의사결정트리 분석은 학습은 70%, 검정은 30%로 지정하였으며 계산 방법은 CART 방법을 실행하였다.

CART 방법은 의사결정 트리 방법으로 대중적으로 사용하는 방법론으로 Classification And Regression Tree라는 뜻을 가집니다. CART는 범주형 목표변수는 지니 지수를 연속형 목표변수의 경우 부산의 감소량을 이용하여 이진 분리를 수행하는 알고리즘이다. 지니 지수는 불순도를 측정하는 하나의 지수로서 역할을 수행한다.

의사결정트리 분석은 응답자들이 인지한 기업 성장 단계인 범주형 변수를 목표 변수로 선택하고, 7개의 변수를 통해서 분류를 실시하였다. 그 결과, 5개의 집단을 예측하고자 하였으나 3개의 집단의 분류로 판단하는 것이 옳바르다고 나타났다. 그리고 학습에서 정확도는 63%, 검정에서는 58.8%로 분류의 정확도는 그리 높게 나타나지는 않았다. 이러한 결과는 <표 6>에 제시하였다.

분석에서 사용된 독립 변수의 중요도를 살펴보면 운영기간을 시작으로 퇴직자증감, 연구개발투자 증감의 순으로 나타났으며, 이들 변수를 통해서 집단을 분류하는데 활용 및 가지치기가 되었다. 정규화의 중요도를 살펴보면 퇴직자 증감외에 다른 변수들의 중요도는 매우 낮아지는 형태를 보이고 있다. 이에 대한 세부 내용은 <표 7>에 제시하였다.

의사결정트리의 분류결과는 3가지 집단으로 초기, 성장기, 쇠퇴기로 구분하였다. 창업기는 1193, 성장기는 215개, 성숙기는 706개의 표본수로 구성되었다. 매출액과 영업이익이 증폭되는 시점을 성장기로 구분

<표 6> 의사결정트리 분류 예측표

표본	관측	예측					정확도 퍼센트
		1.0	2.0	3.0	4.0	5.0	
학습	1.0	0	13	2	0	0	0.0%
	2.0	0	102	138	7	0	41.3%
	3.0	0	34	485	111	0	77.0%
	4.0	0	5	232	368	0	60.8%
	5.0	0	1	5	12	0	0.0%
	전체 퍼센트	0.0%	10.2%	56.9%	32.9%	0.0%	63.0%
검정	1.0	0	5	3	0	0	0.0%
	2.0	0	32	59	0	0	35.2%
	3.0	0	20	178	61	0	68.7%
	4.0	0	3	91	142	0	60.2%
	5.0	0	0	0	5	0	0.0%
	전체 퍼센트	0.0%	10.0%	55.3%	34.7%	0.0%	58.8%

성장방법: CRT

종속변수: 장기적 성장 전망성(5점 척도)

<표 7> 의사결정트리 독립변수 중요도

독립변수	중요도	정규화 중요도
운영기간	.144	100.0%
퇴직자	.028	19.2%
연구개발투자 증감	.012	8.7%
매출액 증감	.010	7.0%
자본 증감	.007	5.1%
자산 증감	.006	4.0%
영업이익 증감	.004	2.6%

하였고 215개 샘플로 나타났다. 앞서 군집분석의 5개의 집단으로 분류할 경우 매출액, 영업이익 등에서 음의 값을 가질 수 있는 형태를 보였지만 3개의 집단으로 분류된 의사결정트리에서는 음의 값을 보이는 집단은 나타나지 않았다. 그럼에도 불구하고 특정 하나의 집단에 1000개가 넘는 샘플이 집중되는 형태는 군집 분석과 동일한 결과를 보이고 있겠다. 이에 대한 세부 내용은 <표 8>에 제시하였으며 비교 용이하기 위해서 평균값으로 비교 분석하였다.

<표 8> 의사결정트리 분류결과

성장단계	창업기	성장기	성숙기
표본수	1193	215	706
자산증감	1067.4895	1861.7488	3006.3144
자본증감	556.6932	916.8651	1721.6006
매출액증감	832.3965	2033.9535	1305.4986
영업이익 증감	17.6521	105.5116	143.7535
연구개발 투자증감	528.0168	374.8279	1587.8513
운영기간	10.0972	3.6884	21.1275
퇴직자	-4.0763	-4.214	-9.9901

다음은 인공지능경망을 토대로 분류를 예측하였다. 인공지능경망은 지도학습 방법으로 부동산 가격 예측, 자율주행차 등 분야에서 많이 활용된다[5]. 이 연구에서 인공지능경망은 다층 퍼셉트론으로서 의사결정트리와 마찬가지로 인지한 기업 성장 단계인 범주형 변수를 종속변수로 7개의 변수를 독립변수를 토대로 설정하여 분석하였다. 학습은 70%, 검정은 30% 방법의 형태로 분석을 하였다.

<표 9>를 살펴보면 인공지능경망은 5개의 집단이 아닌 3개의 집단으로 분류를 실시하였으며 학습은 60.7%, 검정은 58.8%로 크게 높지 않은 정확도를 보여주고 있다. 이러한 내용은 7개의 변수로 5개의 집단으로 분류는 올바르지 않다고 보고 있는 것이다.

인공지능경망의 독립변수 중요도를 살펴보면 의사결정 트리와 다른 결과를 보여주고 있다. 정규화에 제일 중요한 변수는 운영기간, 자산 증감, 매출액 증감 순으로 나타났으며 퇴직자의 변수가 가장 낮은 중요도를 나타내고 있었다. 세부 내용은 <표 10>에 제시하였다.

인공지능경망의 분류 결과를 살펴보면 의사결정트리 분석과 동일하게 약 1000개의 샘플이 하나의 그룹으로 형성된 모습을 볼 수 있으며 창업기 1039개, 성장기 16개, 성숙기 914개로 분류되었다. 인공지능경망의 세부 분류 결과를 평균값으로 비교하기 위해 <표 11>

<표 9>인공지능경망 분류 예측표

표본	관측	예측					정확도 퍼센트
		1.0	2.0	3.0	4.0	5.0	
학습	1.0	0	11	4	0	0	0.0%
	2.0	0	69	153	13	0	29.4%
	3.0	0	26	406	190	0	65.3%
	4.0	0	6	154	413	0	72.1%
	5.0	0	0	2	17	0	0.0%
	전체 퍼센트	0.0%	7.7%	49.1%	43.2%	0.0%	60.7%
검정	1.0	0	6	2	0	0	0.0%
	2.0	0	28	71	4	0	27.2%
	3.0	0	14	166	87	0	62.2%
	4.0	0	0	81	187	0	69.8%
	5.0	0	1	0	3	0	0.0%
	전체 퍼센트	0.0%	7.5%	49.2%	43.2%	0.0%	58.6%

종속변수: 장기적 성장 전망성(5점 척도)

<표 10> 인공지능경망 독립변수 중요도

독립변수	중요도	정규화 중요도
운영기간	.276	100.0%
자산 증감	.207	75.0%
매출액 증감	.190	68.7%
영업이익 증감	.106	38.3%
자본 증감	.093	33.8%
퇴직자	.071	25.7%
연구개발투자 증감	.056	20.4%

종속변수: 장기적 성장 전망성(5점 척도)

로 제시하였다.

다양한 분류 분석을 종합하여 <표 12>에 정리하였다. 내용을 비교해보면, 최초 인지된 응답으로 창업기부터 쇠퇴기까지 그룹 샘플과 빅데이터 기법을 통한 분류 분석한 결과의 집단의 샘플은 큰 차이를 보이고 있다.

두 가지 군집분석 방법을 활용하여 해석해보면, 인지된 응답에서 창업기는 23개이고, 고도성장파 성숙기에 샘플이 집중되어 있다. 하지만 군집분석의 경우

<표 11> 인공신경망 분류결과

성장단계	창업기	성장기	성숙기
표본수	1039	161	914
자산증감	2329.2926	308.2857	1451.291
자본증감	1288.6516	131.0497	784.1389
매출액증감	1634.4264	1927.1553	375.919
영업이익 증감	136.3311	-32.9876	9.7341
연구개발 투자증감	745.1684	190.9255	1123.1575
운영기간	8.7892	3.3292	19.7888
퇴직자	-5.2868	-2.8696	-7.5131

<표 12> 분류 분석별 정리 현황

성장 단계	창업기	초기성장	고도성장	성숙기	쇠퇴기
인지 응답	23	338	889	841	23
K평균 군집	1909	8	9	183	5
계층 군집	1968	86	11	44	5
의사결정트리	1193	215	706		0
인공신경망	1039	161	914		0

7개의 변수를 토대로 5개의 그룹으로 집단을 분류하였다. 그 결과로 약 1900여개의 샘플이 한 그룹으로 묶이는 현상을 확인할 수 있었다. 그리고 5개의 그룹으로 지정하기에는 나머지 집단의 샘플 수가 10개미만으로 형성되면서 차후 분석에 문제점으로 나타났다. 이는 7개의 변수를 통해서 5개의 집단으로 분류하기에는 어려움이 있다는 것으로 해석될 수 있다. 또는 설문 응답자가 인지하는 기업 성장 단계가 실질적인 기업 성장 단계가 아닐 수 있다는 것이다. 즉, 성숙기라고 응답하였지만 쇠퇴기에 접어든 형태일 수도 있으며, 초기 성장기라고 응답하였지만 고도 성장기에 접어든 형태일 수도 있다는 것이다.

또한, 의사결정트리와 인공신경망의 경우는 5개의 집단이 아닌 3개의 집단으로 분류하였다. 즉, 7개의 한정된 변수를 토대로 기업 성장 단계를 분류해보면 분류 분석 기법마다 상이한 집단 그룹을 제시하고 있다. 이러한 집단 분류의 결과는 경영 이론과 실무적

경험치가 함께 고려하여 분류 분석의 결과를 채택할 필요성이 있겠다. 또한, 경영 이론이 없는 상태에서 분류할 경우 분류 오류를 더욱 크게 하는 경향을 보일 수 있겠다.

4.4 단계 3 분석 : 경쟁력 차이 분석

다음은 각 빅데이터 분류 분석으로 구분된 집단별마다 기업 경쟁력 차이가 어떠한지를 분석을 실시하였다. 기업 성장 단계별로 기술부터 마케팅까지 역량이 기업 성장에 맞추어 올라가고 있는지 여부를 판단하기 위해서 분석을 실시하였다.

집단 비교분석은 두 가지 방법을 활용하였다. 우선 인지된 응답과 의사결정트리, 인공신경망 분석은 정규분포가 가정된 일원배치 분산분석을 활용하였다. 하지만 샘플 수의 균형의 차이가 많아 정규분포를 가정할 수 없는 K-평균과 계층적 군집분석은 비모수 통계분석 기법인 Kruskal Wallis 방법을 활용하였다.

경쟁력 변수는 기술, 가격, 디자인, 품질, 조직관리, 마케팅으로 분석하였다. 우선 응답자들이 인지한 기업 성장 단계를 집단으로 분류하여 각 분야별 경쟁력을 분석하였다. 분석 결과, 가격과 디자인, 품질의 경우에는 기업 성장단계와 상관없이 모두 경쟁력이 보통 이상이라고 응답하였고, 경쟁력에서 큰 차이는 없다고 나타났다. 하지만 기술, 조직관리, 마케팅에서 고도성장기, 성숙기, 쇠퇴기에 위치한 기업이 창업기와 초기성장기보다는 조금 더 높은 경쟁력을 가지고 있다고 나타났다. 기술, 가격, 디자인, 품질, 조직관리, 마케팅에서 전반적으로 고도성장기, 성숙기에 위치한 기업들이 경쟁력이 높게 나타났으며 창업기에 위치한 기업들의 점수는 다른 성장 단계보다 낮은 점수로 나타났다. 이러한 세부 내용은 <표 13>에 제시하였다.

K-평균 군집으로 분석된 일부 집단들은 정규분포의 가정이 어렵기 때문에 비모수 통계기법인 Kruskal Wallis 일원배치 분석을 실시하였다.

분석 결과, 기술, 기술, 가격, 디자인, 품질, 조직관리, 마케팅에서 기업 성장 단계별로 차이를 보인다고 나타났다. 인지된 응답에서 성장 점수와 유사하게 창업기에 위치한 1909개 기업들은 성장기, 성숙기, 쇠퇴기의 기업보다는 낮은 경쟁력을 보인다고 나타났다. 특히, 쇠퇴기에 위치한 5개의 샘플 기업들은 자산 감

소, 자본 감소, 매출액 감소, 영업이익 감소, 퇴직자 증가가 매우 높은 특징을 가지고 있다. 그러나 쇠퇴기에 위치한 기업들 경쟁력 수준의 응답은 다른 성장 단계의 집단보다 높게 나타났다. 이는 산업 내 오랜 운영 기간과 경영 노하우에서 비롯된 응답으로 판단된다. 이에 대한 상세 내역은 <표 14>에 제시하였다.

<표 13> 기업성장 단계별 경쟁력 차이 : 인지된 응답

변수	집단	n	평균	표준편차	F	P	사후검정
기술	1	23	3.69	.875	6.72	.000	3>2 4>2
	2	338	3.65	.702			
	3	889	3.85	.656			
	4	841	3.85	.606			
	5	23	3.73	.619			
가격	1	23	3.56	.843	0.30	.872	
	2	338	3.50	.727			
	3	889	3.54	.651			
	4	841	3.51	.647			
	5	23	3.47	.665			
디자인	1	23	3.17	.886	2.02	.089	1=2 =3 =4=5
	2	338	3.36	.755			
	3	889	3.43	.710			
	4	841	3.46	.666			
	5	23	3.47	.730			
품질	1	23	3.52	.947	1.57	.180	
	2	338	3.67	.693			
	3	889	3.73	.661			
	4	841	3.75	.619			
	5	23	3.78	.671			
조직관리	1	23	3.00	.797	9.69	.000	2,3,4,5>1 3,4>2
	2	338	3.37	.716			
	3	889	3.53	.632			
	4	841	3.57	.624			
	5	23	3.52	.593			
마케팅	1	23	2.91	.949	6.00	.000	4>1 3>1 3>2 4>2
	2	338	3.16	.765			
	3	889	3.31	.661			
	4	841	3.33	.632			
	5	23	3.26	.619			

집단: 창업기(1) 초기성장(2) 고도성장(3) 성숙기(4) 쇠퇴기(5)

<표 14> 기업성장 단계별 경쟁력 차이 : K-평균 군집

변수	집단	n	평균	표준편차	F	P	사후검정
기술	1	1909	1043	15.81	.003	4>1 3>1	3>2 4>2
	2	8	1227	.702			
	3	9	1397	.656			
	4	183	1171	.606			
	5	5	1350	.619			
가격	1	1909	1047	10.96	.027	2>1 4>1	
	2	8	1467	.727			
	3	9	1036	.651			
	4	183	1134	.647			
	5	5	1435	.665			
디자인	1	1909	1040	23.56	.000	4>1	1=2 =3 =4=5
	2	8	876	.755			
	3	9	1373	.710			
	4	183	1216	.666			
	5	5	1486	.730			
품질	1	1909	1044	18.26	.001	5>1 4>1	
	2	8	1195	.693			
	3	9	1287	.661			
	4	183	1162	.619			
	5	5	1736	.671			
조직관리	1	1909	1040	23.14	.000	5>1 4>1	2,3,4,5>1 3,4>2
	2	8	1150	.716			
	3	9	1082	.632			
	4	183	1216	.624			
	5	5	1628	.593			
마케팅	1	1909	1038	30.55	.000	5>1 4>1	4>1 3>1 3>2 4>2
	2	8	1351	.765			
	3	9	1390	.661			
	4	183	1205	.632			
	5	5	1784	.619			

집단: 창업기(1) 초기성장(2) 고도성장(3) 성숙기(4) 쇠퇴기(5)

다음은 계층적 군집에 의해 분류된 집단들을 비모수 통계기법으로 일원배치 분산분석을 실시한 결과이다. K-평균 군집의 집단 결과와 유사하게 기술, 가격, 디자인, 품질, 조직관리, 마케팅 영역에서 기업 성장단계별로 차이가 있다고 나타났다. 이에 대한 상세 내역은 <표 15>에 제시하였다.

<표 15> 기업성장 단계별 경쟁력 차이 : 계층 군집

변수	집단	n	평균	표준편차	F	P	사후검정
기술	1	1968	1046	14.47	.006	2>1 3>1	3>2 4>2
	2	86	1193	.702			
	3	11	1366	.656			
	4	44	1186	.606			
	5	5	1350	.619			
가격	1	1968	1046	16.72	.002	4>1 3>1 3>2	
	2	86	1128	.727			
	3	11	1481	.651			
	4	44	1254	.647			
	5	5	1435	.665			
디자인	1	1968	1044	18.69	.001	2>1	1=2 =3 =4=5
	2	86	1268	.755			
	3	11	1240	.710			
	4	44	1139	.666			
	5	5	1486	.730			
품질	1	1968	1050	12.60	.013	5>1 5>2 5>4	
	2	86	1108	.693			
	3	11	1291	.661			
	4	44	1153	.619			
	5	5	1736	.671			
조직관리	1	1968	1047	14.04	.007	5>1 4>1	2,3,4,5>1 3,4>2
	2	86	1139	.716			
	3	11	1161	.632			
	4	44	1251	.624			
	5	5	1628	.593			
마케팅	1	1968	1044	25.14	.000	5>1 4>1 4>2 5>4	4>1 3>1 3>2 4>2
	2	86	1144	.765			
	3	11	1448	.661			
	4	44	1278	.632			
	5	5	1784	.619			

집단: 창업기(1) 초기성장(2) 고도성장(3) 성숙기(4) 쇠퇴기(5)

계층적 군집분석에서도 창업기에 위치한 기업들이 다른 성장 단계에 위치한 기업들에 비해 낮은 점수를 보이고 있었으며, 쇠퇴기에 위치한 기업들이 다른 집단에 비해 경쟁력이 높게 나타났다. 쇠퇴기에 위치한 기업들의 경쟁력이 높게 나타난 것은 오랜 기업 활동에 따른 경영 노하우에서 비롯되었다고 판단된다.

다음은 의사결정 트리로 분류된 집단들에 대한 일원배치 분산분석을 실시하였다. 의사결정 트리 분석에서는 5개의 집단의 의미가 높지 않다고 나타났으며 3개의 집단으로 재분류된 형태로 나타났다. 이에 대한 상세 내용은 <표 16>에 제시하였으며, 분석은 세 개의 집단으로 기술, 디자인, 품질, 조직관리, 마케팅에서 경쟁력 차이를 보인다고 나타났다. 그리고 가격 경쟁력은 세 집단 간에 큰 차이는 없다고 나타났다.

대다수 경쟁력 변수에서는 성숙기에 있는 기업들이 창업기, 성장기의 집단에 비해서 높은 경쟁력을 가진다고 나타났다. 하지만 경쟁력의 점수들이 3.4~3.9 사이에 분포되어 있어 대다수 보통이상 수준이라고 나타났고, 매우 높은 수준의 경쟁력을 보유하고 있지는 않다고 나타났다. 또한, 성장기의 집단들은 창업기 집단들과 유사한 경쟁력의 점수를 나타내고 있다. 즉, 세 개의 집단으로 구분되었을 때, 창업기에 위치한 기업들과 성장기로 접어든 기업들의 경쟁력 수준은 매우 유사한 수준이라고 판단할 수 있다.

다음은 인공신경망으로 분류된 집단들의 일원배치 분산분석을 실시한 결과이다. 인공신경망 분석도 5개의 집단의 의미가 높지 않다고 나타났으며 3개의 집단으로 재분류된 형태로 나타났다. 이에 대한 상세 내용은 <표 17>에 제시하였다. 지시하였으며, 분석은 세 개의 집단으로 기술, 가격, 디자인, 품질, 조직관리, 마케팅에서 경쟁력의 차이가 있다고 나타났다. 그리고 기업이 창업기에서 성숙기로 이동할수록 경쟁력이 있다고 나타났다. 그러나, 성장기에 접어든 기업들은 창업기보다 기술, 디자인, 품질, 조직관리, 마케팅에서 낮은 경쟁력이 있다고 나타났다. 이는 의사결정

<표 16> 기업성장 단계별 경쟁력 차이 : 의사결정 트리

변수	집단	n	평균	표준편차	F	P	사후검정
기술	1	1193	3.76	.658	13.1	.000	3>1 3>2
	2	215	3.78	.712			
	3	706	3.91	.605			
가격	1	1193	3.50	.659	2.37	.094	1=2 =3
	2	215	3.50	.729			
	3	706	3.56	.651			
디자인	1	1193	3.40	.703	4.56	.011	3>1 3>2
	2	215	3.37	.761			
	3	706	3.49	.682			
품질	1	1193	3.68	.659	8.04	.000	3>1
	2	215	3.74	.698			
	3	706	3.81	.624			
조직관리	1	1193	3.48	.650	12.8	.000	3>1 3>2 1>2
	2	215	3.37	.737			
	3	706	3.60	.610			
마케팅	1	1193	3.26	.666	12.6	.000	3>1 3>2 1>2
	2	215	3.14	.787			
	3	706	3.38	.639			

집단: 창업기(1) 초기성장(2) 고도성장(3) 성숙기(4) 쇠퇴기(5)

<표 17> 기업성장 단계별 경쟁력 차이 : 인공신경망

변수	집단	n	평균	표준편차	F	P	사후검정
기술	1	1039	3.77	.656	14.0	.000	3>1 3>2
	2	161	3.66	.715			
	3	914	3.89	.623			
가격	1	1039	3.49	.659	3.52	.030	3>1
	2	161	3.49	.734			
	3	914	3.56	.655			
디자인	1	1039	3.40	.709	6.59	.001	3>1 3>2
	2	161	3.30	.766			
	3	914	3.49	.681			
품질	1	1039	3.72	.652	4.23	.015	1>2 3>2
	2	161	3.60	.734			
	3	914	3.76	.639			
조직관리	1	1039	3.50	.659	18.5	.000	3>1 3>2 1>2
	2	161	3.24	.772			
	3	914	3.57	.602			
마케팅	1	1039	3.28	.682	16.9	.000	3>1 3>2 1>2
	2	161	3.01	.794			
	3	914	3.34	.629			

집단: 창업기(1) 초기성장(2) 고도성장(3) 성숙기(4) 쇠퇴기(5)

<표 18> 집단별 평균값 비교 현황

변수	집단	인지 응답	K평균 군집	계층군집	의사결정 트리	인공 신경망
기술	1	3.69	3.80	3.80	3.76	3.77
	2	3.65	4.00	3.99	3.78	3.66
	3	3.85	4.22	4.18	3.91	3.89
	4	3.85	3.96	3.98		
	5	3.73	4.20	4.20		
가격	1	3.56	3.51	3.51	3.50	3.49
	2	3.50	4.00	3.62	3.50	3.49
	3	3.54	3.44	4.00	3.56	3.56
	4	3.51	3.62	3.77		
	5	3.47	4.00	4.00		
디자인	1	3.17	3.41	3.42	3.40	3.40
	2	3.36	3.25	3.72	3.37	3.30
	3	3.43	3.89	3.73	3.49	3.49
	4	3.46	3.65	3.55		
	5	3.47	4.00	4.00		
품질	1	3.52	3.72	3.72	3.68	3.72
	2	3.67	3.88	3.80	3.74	3.60
	3	3.73	4.00	4.00	3.81	3.76
	4	3.75	3.87	3.86		
	5	3.78	4.60	4.60		
조직관리	1	3.00	3.49	3.50	3.48	3.50
	2	3.37	3.63	3.63	3.37	3.24
	3	3.53	3.56	3.64	3.60	3.57
	4	3.57	3.70	3.73		
	5	3.52	4.20	4.20		
마케팅	1	2.91	3.27	3.27	3.26	3.28
	2	3.16	3.63	3.43	3.14	3.01
	3	3.31	3.67	3.73	3.38	3.34
	4	3.33	3.48	3.57		
	5	3.26	4.20	4.20		

트리와 유사한 형태의 값을 보인다고 할 수 있다.

다음 <표 18>은 개별적으로 계산된 인지된 응답, 계층군집, K군집, 의사결정트리, 인공신경망의 결과를 통합 비교한 현황이다. 대다수 분류 분석에서 샘플이 극단적으로 창업기에 몰리는 경향을 보이고, 나

머지 단계에는 샘플이 작아지는 이슈가 있었다. 그리고 의사결정 트리와 인공신경망의 경우는 3개의 집단으로 분류된 결과를 제공하였다. 집단 분류의 해석 시 제품 개발, 매출과 자산의 증폭, 시장 둔화, 기업 운영 기간 등의 성격을 참조하여 분류하는데 무리는 없었다. 하지만, 분류된 샘플 수의 차이는 분류 방법에 따라 상이하였다. 이에 벤처기업 성장 단계별로 각 분류기법을 통한 세부 해석은 조금씩 차이가 있어 해석하는데 유의해야 할 필요성이 있겠다.

그리고 분류한 집단을 토대로 인지응답을 제외한 대다수 분류된 집단에서는 기술, 가격, 디자인, 품질, 조직관리, 마케팅에서 경쟁력의 집단별 차이를 보인다고 나타났다.

V. 연구 결론 및 시사점

본 연구는 벤처기업들의 기업 성장 단계를 빅데이터에서 주로 많이 사용되는 분류 분석 기법들을 통해 재분류하는 것이었다. 벤처기업의 성장을 인지된 상황에서 어느 단계에 있는지 응답자들의 판단과 통계적 재분류의 차이를 통해서 상호 격차가 있는지를 확인하였다.

분류 분석의 결과를 살펴보면, 각각의 분류 분석마다 계산 방법의 차이가 있으며, 또한, 인공신경망, 군집분석은 거리를 계산하는 다양한 방법이 존재하고, 의사결정 트리 또한 CHAID, CART, C 5.0 등 계산 방법이 있다[5, 14]. 이처럼 다양한 계산 절차와 다양한 방법으로 집단 분류가 이루어진다. 그러므로 분류 분석의 결과 해석은 경영 이론 및 조직 이론, 실무적 관점을 다각도에서 해석하고, 이해할 수 있어야 빅데이터 기법을 올바르게 활용할 수 있을 것이다.

그리고 분류결과를 분석하기 이전에 기업성장단계에서 설명되는 매출액, 마케팅방법, 연구개발, 사업의 다각화 등의 다양한 변수를 수집하고 활용하여 벤처

기업의 성장단계를 분석하는 것이 올바르나[9-13] 일반적으로 모든 변수를 수집하고 분석하는데 제약이 발생한다. 이에 기업 성장 단계를 살펴보는데 있어 변수 증가의 한계점이 발생할 수 있다. 즉, 기업의 성장을 관찰할 수 있는 여러 가지 변수가 있을 수 있다. 생산, 조직, 마케팅, 재무 등 시계열 데이터를 모두 활용한 분석이 이루어져야 정확한 분류가 이루어질 수 있음에도 불구하고 이 모든 변수를 고려하고 수집하는데 애로사항을 발생할 수 있다. 그러므로 변수의 선택과 샘플 수에 따라서 분류 분석의 집단 차이를 보일 수 있다고 판단된다.

정리하면, 벤처 기업의 성장 단계를 분석하는데 있어서 빅데이터 기법의 계산 방법 선택, 다양한 변수의 선택이 샘플 사이즈의 결과가 다르게 나타났다. 그러므로 비즈니스 시나리오를 작성하고, 경영이론을 토대로 올바른 해석을 실시해야 할 필요성이 있겠다[21]. 여러 기업 성장 문헌에서 나타났듯이 자원과 기업이 정신, 산업적 맥락의 카테고리 안에서 여러 변수들이 있으므로[13], 분석 시 모든 변수를 고려할 수 없으므로 경영 이론적 판단이 요구된다[9-13]. 즉, 분석된 결과만 가지고 경영이론 없이 판단하게 되면 해석의 오류가 발생할 여지가 높아질 수 있겠다.

References

- [1] S. Klaus, The 4th industrial revolution World Economic Forum. New York: Crown Business, 2016.
- [2] Laudon, Kenneth C., and Jane P. Laudon. Management information system, Pearson Education India, 2016.
- [3] Turban, Efraim, et al. Electronic commerce 2018: a managerial and social networks perspective, Springer, 2017.

- [4] Robbins, S. P. and Judge, T., Essentials of organizational behavior, NJ: Prentice Hall, 2012.
- [5] Berry, Michael JA, and Gordon S. Linoff. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2004.
- [6] Gompers, Paul Alan, and Joshua Lerner. The venture capital cycle. MIT press, 2004.
- [7] Klepper, S. "Entry, exit, growth, and innovation over the product life cycle," The American economic review, 1996, pp.562-583.
- [8] Scott, M., & Bruce, R. "Five stages of growth in small business. Long range planning," Vol. 20, No. 3, 1987, pp.45-52.
- [9] Kazanjian, R. K. "Relation of Dominant Problems to Stages of Growth in Technology-Based New Ventures," Academy of Management Journal, Vol. 31, No. 2, 1988, pp.257-279.
- [10] Churchill, N. C., & Lewis, V. L. "The five stages of small business growth," Harvard business review, 1983, Vol. 61, No. 3, pp.30-50.
- [11] Hesterly, W., & Barney, J. Strategic management and competitive advantage. Upper Saddle River, NJ, 2010.
- [12] Chandler, Alfred Dupont. "Strategy and structure: Chapters in the history of the industrial enterprise," Vol. 120, MIT press, 1990.
- [13] Gilbert, B. A., McDougall, P. P., & Audretsch, D. B. "New venture growth: A review and extension," Journal of management, Vol. 32, No. 6, 2006, pp.926-950.
- [14] 데이터분석전문가 가이드 2016, 한국데이터진흥원
- [15] Erevelles, S., Fukawa, N., & Swayne, L. "Big Data consumer analytics and the transformation of marketing," Journal of Business Research, 2016, Vol. 69, No. 2, pp.897-904.
- [16] Hassani, H., & Silva, E. S. "Forecasting with big data: A review," Annals of Data Science, Vol. 2, No. 1, 2015, pp.5-19.
- [17] Cao, L. "Data science: a comprehensive overview," ACM Computing Surveys (CSUR), 2017, Vol. 50, No. 3, pp.43:1-43:42.
- [18] Gandomi, A., & Haider, M. "Beyond the hype: Big data concepts, methods, and analytics," International journal of information management, Vol. 35, No. 2, 2015, pp.137-144.
- [19] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. Multivariate data analysis, 2006.
- [20] 2017년 벤처기업정밀실태조사 - 중소벤처기업부
- [21] 정병호, 김병초, "빅데이터 분석의 역량 강화를 위한 거꾸로 교실 설계 연구," 디지털산업정보학회 논문지, 제13권, 제2호, 2017, pp.127-145.

■ 저자소개 ■



정 병 호
Jung Byoung-ho

2018년 3월 한국외국어대학교 외래교수
2015년 9월 한국외국어대학교 경영학 박사
2011년 3월 한국외국어대학교 경영학 석사

관심분야 : IT투자, 정보윤리, 빅데이터, 신기술 혁신
E-mail : jung.hm@s@gmail.com

논문접수일 : 2019년 11월 27일
수 정 일 : 2019년 12월 9일
게재확정일 : 2019년 12월 10일