

소셜 빅 데이터 분석을 통한 미용분야 대학생
창업지원 정책에 관한 연구
-탐색적 데이터 분석법을 기반으로-

윤미연^{1,†} · 박남훈²

¹배재대학교 뷰티케어학과, 교수

²한양대학교 기술경영전문대학원, 박사 연구원

(2022년 12월 3일 접수: 2022년 12월 13일 수정: 2022년 12월 16일 채택)

Study on the Policy of Supporting University Students
in the Beauty Field through Social Big Data Analysis:
Based on exploratory data analytics

Mi-Yun Yoon^{1,†} · Nam-hoon Park²

¹Professor, Department of Beauty Care, Pai Chai University

²PhD. Resrarcher, Grauate School of Thchnology Management, Hanyang University

(Received December 3, 2022; Revised December 13, 2022; Accepted December 16, 2022)

요 약 : 본 연구에서는 미용분야 창업 활성화를 위해 소셜 빅데이터 분석을 탐색적 데이터 분석(EDA)을 기반으로 하여 2019년부터 2021년 동안 각 년도별로 기간을 구분하여 '미용창업'에 대한 수요 변화와 감정 및 의미 차이의 특징적인 패턴을 도출하고자 하였다. '미용창업' 키워드를 주제로 연관된 검색어를 추출한 결과 창업에 필요한 전문적인 창업교육 보다는 미용관련 기술을 배울 수 있는 기관이나 자격증에 더 많은 관심을 보였으며, 이는 정부 및 지자체에서 여러 가지 창업지원 정책들이 마련되고 있음에도 불구하고 여전히 전문적인 창업교육의 중요성을 인식하지 못하고 있는 것으로 파악할 수 있으며, 이에 대한 대안으로 미용분야 창업을 성공적으로 이루기 위한 전공별 맞춤형 창업교육 프로그램을 개발하는 것이 필요할 것으로 사료된다. 탐색적 데이터 분석을 통해 가설을 설정하고 전통적인 확증적 데이터 분석(CDA)을 결합하여 가설을 검증한다. 미용 창업을 위한 탐색적 데이터 분석 방법이 존재한 적은 없으며, 정식 창업교육의 필요성을 언급하기보다는 미용창업에 대한 관심 변화와 예비창업자의 요구사항을 탐색적 데이터로 분석한다면 맞춤형 창업 프로그램 개발에 도움이 될 것이라고 확신한다.

주제어 : 미용창업, 빅 데이터, 탐색적 데이터 분석, 미용산업, 미용

[†]Corresponding author
(E-mail: ymy@pcu.ac.kr)

Abstract : In order to revitalize start-ups in the beauty field, this study attempted to derive characteristic patterns of changes in demand and differences in emotions and meaning for 'beauty start-ups' by dividing the period by year from 2019 to 2021 based on exploratory data analysis (EDA). Most of the search terms related to the keyword "beauty start-up" showed more interest in institutions or certificates that can learn beauty skills than professional start-up education, which still does not recognize the importance of start-up education, and as an alternative, it is necessary to develop customized start-up education programs for each major.

We establish hypotheses through exploratory data analysis and verify hypotheses by combining traditional corroborative data analysis (CDA). There has never been an exploratory data analysis method for beauty startups, and rather than mentioning the need for formal start-up education, analyzing changes in interest in beauty startups and the requirements of prospective start-ups with exploratory data will help develop customized start-up programs.

Keywords : Beauty Start-up, Big data, Exploratory data analysis, beauty field, Beauty

1. 서론

국내외 경기침체의 장기화, 기존 산업 생태계의 급 변화, 기업의 경영관리 문제 등 기업의 청년고용 감소로 인하여 청년실업이 지속적으로 증가하고 있다[1]. 경제활동에 적극적으로 참여해야 하는 청년들의 일자리 문제는 국가적 차원에서 해결해야 하는 중요한 문제로 대두되고 있다. 최근 통계청 자료에 의하면 2022년 10월 현재 우리나라 청년 실업률이 5.6%, 청년 실업자가 23만 6천명으로 집계되었다[2]. 2020년 9월 우리나라 청년 실업률이 8.9%, 청년 실업자가 36만 4천명으로 집계된 것과 비교하여 볼 때 청년의 경제활동이 증가된 양상을 보이지만 우리나라 경제고통지수(실업률+물가상승률)는 청년층에서 가장 높게 조사되었다. 청년 실업의 원인으로는 인구감소, 산업의 온라인화 및 무인화 시스템의 확산 등 산업구조의 변화로 인해 청년층 일자리 감소, 청년 연장에 따른 퇴직인구 감소, 예코세대의 청년층 진입 등 다양한 구조적 요인이 있다[2]. 따라서 이를 극복하기 위해서는 정부의 적극적 거시정책, 청년계층 맞춤형 지원을 통해 양질의 일자리 창출과 청년 실업난 해소 노력을 강화해야 하며 이를 위한 방안으로써 청년 창업의 활성화가 필요하다.

창업의 활성화는 국가 경제 활성화는 물론 국가 경쟁력과 직결되는 만큼 전 세계적으로 창업지원 제도 및 정책을 적극적으로 확대하고 있다. 스타트업 게놈(startup genome)은 2022년 글로벌 스타트업 생태계 현황을 분석하여 보고서[3]를 발

표하였는데, 2022년 세계에서 가장 우수한 창업 생태계를 가진 5개 지역은 실리콘밸리, 뉴욕과 런던, 보스턴, 베이징 순으로 보고하였다. 서울은 2020년 20위에서 2021년 16위로 순위를 높인 것에 이어 2022년 사상 처음으로 세계 10대 스타트업 생태계에 이름을 올렸다. 그럼에도 불구하고 여전히 창업지원이 적극적으로 이루어지고 있는 북미와 유럽의 도시가 약 30% 중반대로 가장 많이 차지하였다. 북미 국가의 창업지원 정책 중 대학의 창업교육은 1945년 하버드대학의 창업 과정이 개설되면서 시작되었으며 이후 창업관련 교과목 개설, 학과설치, 창업센터 운영, 창업동아리 지원, 창업 경진대회 개최, 온라인 창업 강좌 개설 등 창업 친화적인 학사제도를 통해 대학생들의 창업에 대한 의지를 강화시키고 있다[4]. 창업 교육의 중요성은 이미 많은 연구자들에 의해 밝혀졌으며, 특히 이제훈[5]은 대학생을 대상으로 운영하는 창업교육 및 다양한 창업관련 프로그램이 대학생들의 창업의지에 영향을 미친다고 보고하였으며, 배병운[6]은 목표 지향성과 개인 기업가 지향성이 창업동기에 영향을 미치는 것을 대학생 대상 실증연구를 통해 밝혔다. 이와 같이 청년창업의 활성화를 위해 우리나라 대학에서는 창업지원센터를 비롯하여 창업관련 교과목 개설, 창업동아리 등 대학생들에게 창업에 대한 인식을 긍정적으로 개선시키기 위해 다양한 프로그램을 운영하고 있다. 우리나라에서 미용분야는 미용관련 국가자격증을 취득하거나 관련분야 전공으로 대학을 졸업하게 되면 창업을 할 수 있는 면허증

을 발급받게 된다. 또한 소규모·소자본 창업이 가능해 전공 관련 기술만 보유하고 있다면 누구나 쉽게 창업할 수 있다. 그러나 기업가정신 및 창업의지의 결여, 창업교육의 부재 등으로 인해 창업이 실패할 확률이 높아지게 된다. 따라서 국가나 지방자치 또는 대학에서 이루어지고 있는 창업지원 정책을 잘 활용해 창업에 대한 인식개선, 창업지원 및 창업교육의 실효성을 증가시켜 창업에 대한 관심을 증가시키는 것이 무엇보다 중요하다.

탐색적 데이터 분석(Exploratory Data Analysis, EDA)은 데이터 간의 상관관계, 기본적인 분석, 패턴 또는 특징 등을 파악하여 여러 변수 간 트렌드나 관계성을 찾기 위해 수집된 방대한 양의 데이터를 공식적인 모델링 또는 가설을 테스트하는 접근에서 벗어나 다양한 각도에서 관찰하고 이해하며 데이터 속에 내포된 새로운 의미를 발견하고 적용할 수 있는 접근법이다[7]. 목적을 가지고 데이터를 확보하여 분석하는 확증적 데이터 분석(Confirmatory Data Analysis, CDA)은 사전에 가설을 설정한 후 수집한 데이터로 가설을 평가하고 증명하는 전통적인 분석기법[8]이라면, 탐색적 데이터 분석은 수집된 데이터를 기반으로 인사이트를 발견하고 데이터의 구조와 특징을 파악하여 얻은 정보를 바탕으로 통계 모형을 만드는 단계에서 사용되며, 빅데이터 분석에 활용되는 분석법이다. 또한 확증적 데이터 분석 결과를 해석하는데 도움을 준다[9]. 탐색적 데이터 분석법은 기업의 마케팅 전략을 수립하기 위해서도 많이 사용되는데 대표적인 사례로는 온라인 쇼핑몰 아마존(amazon.com)에서 고객 리뷰의 유용성과 리뷰 패턴을 식별하기 위하여 탐색적 데이터 분석 방법을 활용하였다[10].

따라서 본 연구는 미용분야 예비창업자들에 대한 요구도 조사를 통해 창업지원 정책을 보완하여 정책에 대한 수요자들의 인식 개선뿐만 아니라 지속적인 빅데이터 추적관찰을 통해 향후 정부의 효율적인 창업지원정책을 수립하고자 하였

다. 이를 위해 최근 3년간 미용창업 수요 변화와 감정 및 의미 차이의 특징적 패턴을 빅데이터를 통한 넓은 의미의 분석을 위하여 탐색적 데이터 분석(Figure 1)을 통해 도출하였다. 탐색적 데이터 분석 프로세스 4단계에 따라 텍스트(Textom)을 활용하여, ‘미용창업’을 주제로 대표 포털 사이트인 네이버(블로그, 카페, 뉴스), 다음(블로그, 카페, 뉴스), 구글(뉴스, 페이스북) 내 최근 3년 데이터를 수집하였으며 데이터 형태소 분석을 기초로 정제 후, TF-IDF분석, 감성분석, 네트워크 분석을 수행하였으며 최종적으로 향후 창업지원정책에 대한 인사이트를 도출하였다.

2. 연구방법

2.1. 데이터 수집 및 정제

본 연구는 2019년부터 2021년까지 3년간 국내 대표 포털 사이트 네이버(블로그, 카페, 뉴스), 다음(블로그, 카페, 뉴스), 구글(뉴스, 페이스북)의 ‘미용창업’을 주제로 데이터를 수집하였으며, 수집과 분석은 각 1년 단위로 실행하였다(Table 1). 데이터 수집 및 전처리과정은 텍스트(Textom 3.0)을 사용하였으며, 텍스트는 빅데이터 분석 프로그램으로 키워드와 관련된 단어들의 네트워크를 시각화 및 단어의 유사군집을 도출하기 위해 많은 연구에서 사용되고 있다[11]. 데이터 수집 결과 2019년은 6,156건, 2020년은 6,347건, 2021년에는 8,405건의 단어가 수집되었다. 텍스트의 정제·형태소 분석을 활용하여 1차 정제하였으며, 정제된 결과는 2019년은 6,152건, 2020년은 6,341건, 2021년은 8,401건으로 수집되었다. 2차 정제에서는 1차에서 정제되지 않은 비정형화 텍스트를 중점으로 정제 작업을 수행하였으며 ‘의’, ‘것’, ‘등’, ‘위’ 등 불필요한 데이터를 삭제하고 형용사, 동사 등을 명사화 하였다. 2차 정제 후 2019년은 6,840건, 2020년은 8,152건, 2021년은 7,045건으로 최종 데이터를 확보하였다.

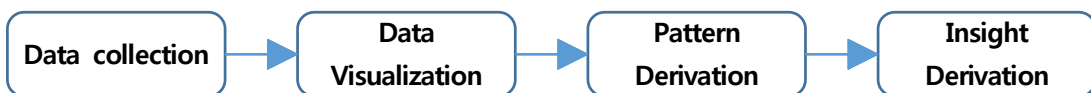


Fig. 1. Exploratory Data Analysis (EDA) process.

Table 1. data Collection

Category	Content	
Collect keywords	Beauty start-up	
Collection period	2019	2019.01.01. - 2019.12.31
	2020	2020.01.01. - 2020.12.31
	2021	2021.01.01. - 2021.12.31
Collection Channels	Naver	Blogs, cafes, news
	Daum	Blogs, cafes, news
	Google	News, Facebook
Collection · Analysis tools	Textom 3.0(www.textom.co.kr)	

2.2 데이터 분석기법

2.2.1. TF-IDF(Term Frequency-Inverse Document Frequency) 분석

TF-IDF는 텍스트 분석에 일반적으로 사용되는 데이터 변환 및 점수체계를 말하며, 문서 내에서 중요도가 높을 것 같은 단어에는 높은 가중치를 부여(penalizing)하고, 중요도가 낮은 단어에는 작은 가중치를 부여하여[12]. 수집대상이 존재하는 경우에 주어진 단어가 문서 안에서 얼마나 중요한지를 나타내는 수치이다[13,14]. 주로 단어출현 빈도(Term Frequency)와 역문서 빈도(Inverse Document Frequency)로 구성되는데, 단어의 출현빈도는 특정 단어가 해당 문서에서 얼마나 자주 출현하는지를 확인할 수 있는 값이며, 역문서 빈도는 다른 문서에서는 많이 출현하지 않고, 해당 문서에서만 얼마나 자주 출현하는지를 의미하게 된다[15]. TF-IDF는 단어 빈도와 역문서 빈도의 곱으로 산정하여 값을 나타내는데, 이러한 분석법은 광고 및 투자 수익 추적을 위한 비즈니스 및 콘텐츠 마케팅에서 많이 사용되고 있다[16]. 본 연구에서는 최종 정제된 단어를 TF-IDF로 분석하여 단어별 가중치를 측정하여 측정된 값에 따라 순위를 정하였으며 TF-IDF 순위 중 상위 50개를 기준으로 네트워크 분석에 사용하였다.

2.2.2. 감성분석(Sentiment Analysis)

감성분석은 기본적인 문서 또는 의견, 평가와 같은 텍스트에 내포된 긍정, 중립, 부정 등 어떤 견해를 갖는지에 대한 주관적인 정보를 컴퓨터를 통해 식별, 분석하는 방법으로 소셜미디어에서 발

생하는 방대한 데이터를 분석할 수 있는 유용한 대안이 되고 있다[17,18]. 최근 팬데믹 상황과 더불어 비대면 문화가 확산되면서 소셜미디어, 온라인 쇼핑 등 웹사이트에서 데이터 기반 매체에 적극적으로 참여하는 사용자들이 늘어나고 있으며 이에 따라 온라인 환경에서 수집할 수 있는 데이터의 양도 기하급수적으로 증가하고 있다. 감성분석은 머신러닝 기반으로 한 감성분석과 감성사전 기반으로 한 감성분석으로 나눌 수 있는데, 머신러닝을 기반으로 한 감성분석은 사용자가 작성한 리뷰 데이터와 그에 대응하는 감성 라벨 값을 같이 모델에 학습시키는 방법이다[19]. 소셜미디어 사용자는 어떤 주제나 기사에 대한 자신의 감정을 의견과 견해의 형태로 온라인상에 공유할 수 있는 최고의 의사소통 수단이고, 이런 과정을 통해 많은 양의 비정형 데이터가 생성된다. 감성분석의 기본적인 접근 방법은 긍정적 어휘('좋아요', '최고야' 등)와 부정적 어휘('싫어요', '별로', '슬픈' 등)를 극성으로 분류하고 문장의 패턴을 정량화한 후 그에 따라 데이터를 분류하고 분석하는 것이다. 이러한 감정을 분석하기 위해 과거에는 다양한 머신러닝 및 자연어 처리 기반 접근 방법을 통해 분석하였으나 최근에는 딥러닝 기술을 활용한 감성 분석이 증가하고 있다[20]. 본 연구는 Textom 3.0 을 활용해 감성분석을 수행하였으며 문장의 내용을 긍정과 부정으로 구분할 수 있는 감성 분류 분석과 원문 데이터 안에 감성과 관련된 키워드가 몇 번 들어갔는지 알려주는 '감성단어빈도분석'을 사용하였다. 감성어 어휘사전은 긍정과 부정이라는 카테고리 안에 긍정의 키워드는 '흥미, 호감, 기쁨' 3개의 단어가 있으며, 부정의 키워드에는 '통증, 슬픔, 분노, 두려움, 놀

람, 거부감' 6개의 단어가 있다. '흥미' 라는 단어에는 흥미를 표현하는 수많은 단어를 강도에 따라 감성강도 7점 만점으로 표준화하였다. 예시로 '호감' 이라는 단어 안에 '행복하다' 와 '그저 그렇다' 라는 단어가 있을 경우 '행복하다' 단어에는 5점 '그저그렇다' 단어에는 1점을 부여해 '호감' 안에 들어가는 단어라도 감성 강도를 다르게 반영하도록 하였다.

2.2.3. 네트워크 분석

네트워크 분석은 텍스트에 출현하는 키워드와 키워드 사이를 하나의 연결선으로 관계를 표시하여 네트워크를 구축하여 그들 간에 형성되는 의미적 관계의 속성들을 파악하고, 이를 통해 현상을 해석하는 분석기법이다.[21]. 한 가지 주제를 설명하는 텍스트에서 함께 출현하는 키워드들을 하나의 선으로 연결하여 키워드 사이의 관계를 의미 네트워크 형태로 시각화할 수 있으며[22], 특정 키워드와 함께 출현하는 단어가 무엇인지, 함께 출현하는 단어의 빈도를 산출하여 키워드 간의 구조적 관계를 파악할 수 있다[23]. 따라서 텍스트에서 단어가 전달하고자 하지만 명백히 드러나지 않지만, 그 단어의 의미를 파악하는데 유용한 분석법이다[24]. 본 연구는 'Gephi 0.9' 프로그램을 사용하였으며, 각 단어의 중심성 및 구조적 등위성을 중심으로 기간별로 네트워크 분석법을 적용하였다. 네트워크 구성요소는 노드(node)와 엣지(Edge)가 있는데, 노드는 고유한 속성을 가지는 키워드를 의미하고 엣지는 노드들 간 연결 관계를 나타낸다[25]. 본 연구에서는 주된 키워드인 미용창업과 관련된 각각의 키워드들이 노드가 되며, 각 키워드 간의 상관관계는 엣지로 표현하였다.

3. 결과 및 고찰

3.1. TF-IDF 분석결과

'미용창업'을 키워드로 2019년부터 2021년의 TF와 IDF를 곱으로 나타내는 TF-IDF 가중치를 산출한 결과(상위 50위)는 Table 2 와 같다. 연도별 특징을 살펴보면 2019년은 '청년', '기업', '센터', '교육', '기술', '혁신' 등이 미용창업에 있어서 핵심단어인 것으로 나타났으며, 특히 '청년', '벤처', '스타트업' 등 창업과 직접적으로 관련된 단어들이 TF-IDF 순위가 높게 나타난 것으로 보

아 팬데믹 이전 창업에 참여하고자 하는 청년층에 의해 미용창업이 활성화되고 있었다는 것을 미루어 짐작할 수 있다. 또한 '취업' 이 25위를 차지함으로써 취업에 대한 가중치는 다소 낮은 것으로 나타났다. 그러나 2019년에 대체로 '창업'에 대한 TF-IDF 가중치가 높게 나타난, 반면 창업지원 정책의 주관기관을 의미하는 '기관'은 46위, '정부'는 42위, '대학'은 37위, '중소벤처'는 35위, '경진대회'는 34위, '대학교'는 31위로 TF-IDF 가중치가 낮은 것으로 나타난 것으로 보아 청년 및 대학생들의 창업에 대하여는 관심은 많으나 정부 및 지자체, 대학에서의 창업지원 정책에 따른 창업 활성화라고 볼 수 없는 결과값을 나타내었다. 2020년은 '취업'이 6위로 TF-IDF 가중치가 높게 나타났으며, '학원', '피부', '자격증', '아카데미', '수강' 등의 단어가 핵심단어인 것으로 보여진 결과 팬데믹 상황이 진행이 되면서 비대면 교육의 확산으로 제대로된 창업 프로그램을 제공 받을 수 없어 창업을 위해 개인이 스스로 기술을 습득하는 것을 중요하게 생각한 것으로 해석할 수 있다. 이 시기에는 창업 보다는 취업을 더 우선적으로 생각했으며 대부분 미용관련 기술에 관한 단어들이 수집되었다. 2021년은 팬데믹 상황의 연장으로 위드 코로나 시기에 발맞춰 반려동물에 대한 관심이 증가되면서 '애견' 키워드가 1위로 TF-IDF 가중치가 가장 높게 나타났다. 또한 2020년과 유사하게 '학원', '뷰티', '취업', '자격증', '피부', '헤어', '아카데미' 등 전문적인 창업교육 및 지원 보다는 각자의 전공관련 기술교육이 더 중요하다고 느끼며 창업을 준비하는 것을 알 수 있었다. 미용분야로 창업을 하기 위한 전공분야 기술은 반드시 필요하지만 창업교육, 기업가정신 등의 결여로 창업에 실패하는 확률이 높아 미용분야 창업을 위해 기업이 정신을 함양시킬 수 있는 창업교육 활성화 방안을 적극적으로 모색해야 할 것으로 사료된다.

3.2. 감성분석결과

'미용창업' 키워드를 주제어로 Textom 3.0 을 활용해 감성분석을 수행하여 문장의 내용을 긍정과 부정으로 구분할 수 있는 감성 분류 분석과 원문 데이터 안에 감성과 관련된 키워드가 몇 번 들어갔는지 알려주는 '감성단어빈도분석'을 한 결과는 Table 3 과 같다. 조사한 기간에 따라 유의미한 차이는 나타나지 않았으나 앞선 연구결과에서 2019년, TF-IDF 분석결과 팬데믹 이전 창업

Table 2. TF-IDF Analysis Results (Top 50)

2019			2020			2021		
rank	word	TF-IDF	rank	word	TF-IDF	rank	word	TF-IDF
1	youth	3517.551	1	school	4526.579	1	pet dog	9780.563
2	entrepreneurial	3217.385	2	skin	4105.889	2	school	6111.865
3	center	2451.496	3	Certification	3439.536	3	skin	5931.326
4	cafe	2374.366	4	Beauty	3208.481	4	Beauty	4360.863
5	success	2161.029	5	education	3134.266	5	Supplies	4319.923
6	Seoul	2071.976	6	employment	2884.947	6	employment	4088.561
7	education	2037.903	7	pet dog	2726.368	7	Certification	3977.796
8	technique	1977.576	8	academy	2499.402	8	education	3963.221
9	Venture	1802.525	9	Hair	2412.821	9	Nail	3555.874
10	innovation	1795.619	10	companion animal	2407.906	10	Hair	3153.526
11	funding	1679.902	11	beauty shop	2093.578	11	academy	3018.262
12	item	1614.876	12	specialist	2082.223	12	National Certificate	2820.306
13	Franchise	1487.79	13	wonju	2023.118	13	companion animal	2760.924
14	Startup	1462.467	14	cosmetologist	1882.183	14	beauty shop	2655.463
15	cost	1431.53	15	Course	1840.208	15	specialist	2635.137
16	Busan	1396.92	16	acquisition	1830.156	16	Art	2591.893
17	women	1395.178	17	makeup	1724.982	17	Course	2527.216
18	space	1358.658	18	cosmetology	1694.83	18	acquisition	2510.455
19	industry	1353.146	19	management	1679.702	19	makeup	2208.741
20	Incheon	1346.388	20	relation	1609.22	20	eyelashes	2194.343
21	loan	1269.481	21	technique	1575.992	21	management	1947.779
22	Recruitment	1269.323	22	university	1488.548	22	brand	1926.956
23	start	1227.781	23	recruitment	1457.453	23	class	1846.389
24	region	1225.91	24	Nail	1443.281	24	university	1819.998
25	employment	1206.176	25	Nail Art	1377.345	25	Busan	1799.192
26	opinion	1185.294	26	representative	1325.873	26	Bandi	1795.556
27	prospective entrepreneur	1182.672	27	class	1322.682	27	technique	1792.973
28	market	1128.808	28	industry	1287.649	28	Companion	1776.887
29	representative	1124.383	29	entrepreneurial	1281.659	29	subject matter	1773.815
30	application	1115.346	30	subject matter	1261.678	30	relation	1765.507
31	university	1105.813	31	Entrance exam	1256.303	31	Bandi House	1722.56
32	investment	1083.394	32	Suwon	1230.409	32	daegu	1719.057
33	Founder	1083.166	33	Shared beauty salon	1224.026	33	Incheon	1705.334
34	Competitions	1065.187	34	daegu	1195.081	34	cosmetologist	1608.772
35	small and medium-sized Venture	1035.672	35	cost	1165.138	35	industry	1603.896
36	economy	1009.703	36	youth	1143.602	36	cafe	1560.479
37	university	1007.388	37	Gwangju	1134.789	37	cosmetology	1555.831
38	idea	993.8891	38	designer	1114.92	38	pet	1515.087
39	Small capital	989.3327	39	start	1094.323	39	playground	1514.697
40	relation	967.4991	40	daejeon	1089.859	40	Semi-perm anent	1504.606
41	restaurant	954.0383	41	eyelashes	1088.908	41	opinion	1487.133
42	government	947.9835	42	device	1080.299	42	success	1476.091
43	korea	942.8106	43	opinion	1077.726	43	hello	1459.447
44	society	931.2036	44	success	1071.616	44	recommendation	1449.954
45	trouble	930.8027	45	campus	1047.154	45	start	1411.951
46	agency	923.3063	46	center	1027.512	46	Entrance exam	1410.625
47	coffee	919.4943	47	Startup	1019.159	47	shopping mall	1382.306
48	competition	900.5859	48	medical care	1012.265	48	fair	1316.37
49	Employment announcement	899.5907	49	subjects	1000.931	49	share	1313.283
50	ecology	892.0386	50	agency	978.7234	50	career	1300.308

Table 3. Sentiment analysis results

Collection period	2019		2020		2021	
sentiment	positivity	negation	positivity	negation	positivity	negation
Frequency Ratio (%)	76.43	23.57	73.06	26.94	78.57	21.43
Sentimental intensity Ratio(%)	76.35	23.65	73.4	26.6	80.2	19.8
Keywords	want good best Innovative recommendation thank special Grow up fun new	ache cry Ignored worry Amazing overwhelming Hard vague embarrassing difficult	want good best Innovative recommendation thank new Grow up happy special	stuffy cry hateful worry unexpected overwhelming ache Hard Ignored afraid	want good best Innovative recommendation happy new fast thank special	itch cry hateful worry shock difficult ache Hard irritation vague

에 대한 가중치가 높게 나타난 것과 유사하게 긍정의 감정빈도비율이 76.43 % 로 높게 나타났으며, 긍정에 대한 주요 키워드는 ‘원하다, 좋다, 혁신적이다, 추천, 감사하다, 특별하다, 성장하다, 즐겁다, 새롭다’ 등으로 분석되었다. 반면 2020년에는 팬데믹 상황으로 비대면 사회로 전환되면서 창업에 대한 긍정적인 감정이 저년도인 2019년에 비해 다소 낮아진 결과를 나타냈으나 유의미한 차이는 나타나지 않았다. 2020년 부정감정의 주요 키워드는 ‘답답하다, 울다, 밉다, 걱정하다, 부담스럽다’ 등 대체적으로 창업에 대한 두려움과 미래에 대한 답답함이 내포되어 있는 부정적인 키워드가 출현되었다. 2021년에는 긍정의 감성강도 비율이 80.2 % 로 2020년에 비해 약 6.8 % 상승하였으며, 예비창업자들에게 창업에 대한 인식이 사회 분위기에 따라 변화되는 것을 알 수 있다. 미용창업 키워드에서의 부정적인 감정 키워드는 ‘힘들다, 무시당하다, 아프다, 어렵다, 막막하다, 답답하다’ 등 대체로 미용기술 습득에 대한 어려움과 습득한 기술을 직접 사람에게 적용해야 하는 부담감이 강하게 작용하는 것을 알 수 있었다.

3.3. 기간별 네트워크 분석결과

텍스트마이닝을 통해 추출된 기간별 키워드를 대상으로 그들 간에 형성되는 의미적 관계의 속성들을 파악하기 위해 네트워크 분석법을 실시하

였다. 분석을 위해 노드(node)와 엣지(Edge) 파일을 생성하였고, Gephi 분석 도구를 활용하여 네트워크의 시각화로 나타난 결과를 미시적 관점으로 분석하였다. 기간별 추출된 전체 키워드를 대상으로 분석하였으며, 추가로 상세한 분석이 필요한 키워드는 예고 네트워크 분석을 하여, 노드와 연결된 이웃 노드의 연결 중심성을 나타내어 기간별 연관 키워드에 대한 중요도를 분석하였다. 노드는 각각의 세부 단위 단어로 구성되며, 중요도에 따라 노드의 크기가 구분되어 진다. 엣지는 노드 사이의 상관관계를 나타내고 있으며, 밀집도가 높을수록 굵은 선으로 표현되어 키워드에 미치는 영향력을 파악할 수 있고, 서로 다른 색으로 구성된 군집은 근접한 노드들을 동일한 색으로 표현하여 키워드 속성별 분석과 인사이트 도출이 신속하게 이루어질 수 있다.

2019년 네트워크 시각화 결과는 Fig. 1 과 같이 크게 6개의 군집으로 분류되었으며, 각 군집은 ‘청년’, ‘미용사’, ‘교육’, ‘학원’, ‘창업’, ‘미용’으로 분류되었다. 가장 크게 분포된 군집인 ‘청년’은 창업지원에 대한 정보 및 대학에서의 창업과 관련된 프로그램 키워드가 포함되었으며 그 대표적인 키워드로는 ‘대학교’, ‘동아리’, ‘학부’ 등이 있다. 해당 기간에는 대부분 첫 번째 군집에 나타난 키워드가 주를 이루고 있었고, 그 외의 군집들은 유의미한 차이를 나타내지 않았으나 ‘창업’, ‘미용’ 키워드 검색 분포가 높게 나타난

국가 경제활성을 위한 대안으로써 창업을 적극적으로 권장하며, 확대시키기 위해 정부 및 지자체, 대학교에서 다양한 창업지원 정책을 마련하고 있다. 특히 대학에서는 창업지원 센터 등을 운영하여 학생들이 대학을 졸업하는 것과 동시에 창업을 시작할 수 있도록 다양한 창업교육을 실시하고 있으며, 학과에서도 창업과 관련된 일부 교과목을 운영하여 학생들로 하여금 창업에 대한 친근감이 높아질 수 있도록 노력하고 있다. 하지만 그럼에도 불구하고 창업을 시도하거나, 창업에 성공하는 사례가 많지 않으며 특히 미용분야에서는 졸업과 동시에 창업을 시도하는 경우가 극히 드물게 나타나고 있는데, 이러한 이유 중 하나는 미용분야의 특성상 기술능력이 충분히 갖추어진 상태에서 창업을 해야만 한다는 편견이 대학생들에게 창업을 주저하게 만드는 원인이 되고 있다. 따라서 미용분야 예비창업자들에 대한 요구도 조사를 통해 창업지원 정책을 보완할 필요성이 있으며, 수요자들의 인식 개선 뿐만 아니라 지속적인 빅데이터 추적관찰을 통해 향후 효율적인 맞춤형 창업지원 정책을 수립하는 것이 무엇보다 필요한 실정이다. 본 연구의 결과를 토대로 살펴보면 ‘미용창업’ 키워드를 주제로 연관된 검색어를 추출한 결과 대부분 창업에 필요한 전문적인 창업교육 보다는 미용관련 기술을 배울 수 있는 기관이나 자격증에 더 많은 관심을 보였으며 이는 여러 가지 창업지원 정책들이 마련되고 있음에도 불구하고 여전히 창업교육의 중요성을 인식하지 못하고 있는 것으로 파악할 수 있으며, 이에 대한 대안으로 미용분야 창업을 성공적으로 이루기 위한 전공별 맞춤형 창업교육 프로그램을 개발하는 것이 필요할 것으로 사료된다. 또한 ‘미용창업’을 위해 기술을 배울 수 있는 뷰티 아카데미에 대한 관심도가 높은 것을 토대로 뷰티 아카데미에서 기술적인 부분만 교육하는 것이 아니라 창업에 필요한 전문적인 이론적 교육, 즉 기업가정신 및 창업의지를 함양시킬 수 있는 이론 중심적인 교육을 병행하는 것이 새로운 대안이 될 수 있다고 사료된다. 연구에서 사용한 탐색적 데이터 분석은 데이터에 대한 신속한 통찰력을 제공하며, 가설을 정형화하고 검증을 돕기 위한 목적을 갖는다. 따라서 탐색적 데이터 분석을 통해 가설을 수립하고 전통방식의 확증적 데이터 분석(CDA)을 결합하여 가설을 검증하게 된다. 미용창업과 관련하여 탐색적 데이터 분석법을 적용한 사례는 전무후무하며, 미용분야의 특수성을

고려할 때 정형화된 창업교육의 필요성을 언급하는 것이 아닌 탐색적 데이터 속에서 미용분야 창업 관심도의 변화와 예비 창업자들의 요구사항들을 분석한다면 미용분야 맞춤형 창업프로그램 개발 많은 도움이 될 것이라 확신한다.

감사의 글

이 논문은 2019년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구임(NRF-2019S1A5A2A01046093).

References

1. T. U. Ahn, T. W. Kang, "A Study on difficulty factors of youth startups for activation local startups", *Asia-pacific journal of business venturing and entrepreneurship*, Vol.15, No.2, pp.67-80, (2020).
2. Statistics. (2022)
3. Startup Genome. Global startup ecosystem report 2022 [Internet]. Global Entrepreneurship Network, 2022. <http://www.startupgenome.com/report/gser2022>. (accessed Jun, 14, 2022)
4. G. D. Ahn, T. K. Lee, "A study of facilitating korean youth startup by analyzing the U.S. and Korean university startup ecosystem", *Asia-pacific journal of business venturing and entrepreneurship*, Vol.33, No.2, pp.104-122, (2018).
5. J. H. Lee, "A study on the ways to vitalize students' entrepreneurship", *Asia-pacific journal of business venturing and entrepreneurship*, Vol.8, No.2, pp.63-74, (2013).
6. B. Y. Bae, "A study on the relationships of individual entrepreneurial orientation, goal orientation, entrepreneurial motivation of korean college students", *Asia-pacific journal of business venturing and entrepreneurship*, Vol.16, No.4, pp.59-70, (2021).
7. W. L. Rice, S. Y. Park, B. Pan, P. Newman, "Forecasting campground demand

- in US national parks”, *Annals of Tourism Research*, Vol.75, pp.424-439, (2019).
8. C. H. Choi, Y. W. Yoo, “The Study on the comparative analysis of EFA and CFA”, *Journal of Digital Convergence*, Vol.15, No.10, pp.103-111, (2017).
 9. J. T. Behrens, “Principles and Procedures of Exploratory Data Analysis”, *American Psychological Association*, Vol.2, No.2, pp.131-160, (1997).
 10. M. G. Majumder, S. D. Gupta, J. Paul, “Perceived usefulness of online customer reviews: A review mining approach using machine learning & exploratory data analysis”, *Journal of Business Research*, Vol.150, pp.147-164, (2022).
 11. M. N. Chang, Y. S. Heo, H. S. Lim, “Moving to Jeju’: An Exploratory Keyword Analysis Using Social Big Data”. *Journal of Tourism & Industry Research*, Vol.39, No.1, pp.15-26, (2019).
 12. J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets*. Cambridge: Cambridge University Press, (2014).
 13. J. H. Lee, T. H. Um, H. J. Lee, “A Study on Legal Document Classification Systems Using TF-IDF and CNNs”. *The Journal of Korean Institute of Communications and Information Sciences*, pp.982-983, (2019).
 14. Y. H. Ji, J. H. Han, “A Study on the Exploration of Paradigm and Future Development Plan for Tourism Sector NFT: Focused on TF-IDF and Semantic Network Analysis”. *Journal of Internet Electronic commerce research*. Vol.22, No.2, pp.131-146, (2022).
 15. H. S. Kim, Y. M. Kim, “A Validation of Effectiveness for Intrusion Detection Events Using TF-IDF”. *Journal of The Korea Institute of Information Security & Cryptology*, Vol.28, No.6, pp.1489-1497, (2018).
 16. U. Erra, S. Senatore, F. Minnella, G. Caggianese, “Approximate TF-IDF based on topic extraction from massive message stream using the GPU”. *Information Sciences*. Vol.292, No.20, pp.143-161, (2015).
 17. J. Y. Kim, S. J. Jeong, S. J. Lee, “A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media”. *Entrue Journal of Information Technology*, Vol.13, No.3, pp.59-70, (2014).
 18. J. A. Lee, J. J. Yang, “A social issue analysis using franchise big data: construction of emotional dictionary through sentimental analysis”, *Journal of hospitality & tourism studies*, Vol.25, No.5, pp.53-61, (2022).
 19. H. Y. Hwang, K. C. Lee, J. Y. Yu, Y. H. Lee, “Multimodal Sentiment Analysis Using Review Data and Product Information”, *the journal of society for e-business studies*, Vol.27, No.1, pp.15-28, (2022).
 20. Y. Ashima, K. V. Dinesh, “Sentiment analysis using deep learning architectures: a review”, *Artificial Intelligence Review*, Vol.53, pp.4335-4385, (2020).
 21. R. Popping, *Computer-assisted text analysis*, SAGE Publications, Ltd. (2000).
 22. Y. S. Kim, H. S. Kwon, “An Comparative Study of Articulation on Science Textbook Concepts and Extracted Concepts in Learning Objectives Using Semantic Network Analysis - Focus on Life Science Domain -”, *Journal of Korean Elementary Science Education*. Vol.35, No.3, pp.377-387, (2016).
 23. G. J. Han, “Significance and Research Tasks of Applying Network Analysis Techniques in Social Science Methodology Situation: Examining Meaning and Network Analysis (Semantic Network Analysis)”, *Research in Social Studies Education*, Vol.10, No.2, pp.219-235, (2003).
 24. C. S. Park, J. W. Jung, “Text Network Analysis: Detecting Shared Meaning through Socio-cognitive Networks of Policy Stakeholders”, *Journal of Governmental Studies*, Vol.19, No.2, pp.73-108, (2013).