

최종보고

소셜 데이터 분석을 통한 벤처기업의 안정성에
영향을 미치는 외부 환경 요인 분석: 외식 프랜
차이즈를 기반으로

제출일: 2019. 11. 11

연구책임자 : 위 희 주

과제명	2019 벤처창업 학술연구
한 글	소셜 데이터 분석을 통한 벤처기업의 안정성에 영향을 미치는 외부 환경 요인 분석: 외식 프랜차이즈를 기반으로
영 문	Analysis of External Environmental Factors Affecting Venture Business Stability through Social Data Analysis: Based on the Food Service Franchise

초록
<p>기존 벤처 기업의 결정요인 연구에서는 기업가의 정신 및 전략 등의 기업 내부 요인이나 정책이나 산업 환경 등의 외부요인만 고려하였다. 이는 무엇보다도 중요한 소비자 의견을 반영하지 않았다는 문제점이 있다. 따라서 본 연구에서는 소셜 데이터를 이용하여 대중의 의견을 시간의 흐름에 따른 키워드로 분석하고, 창업 후 성장 혹은 재정 상태를 유지하고 있는 안정성이 높은 기업과 창업 후 2-3년 내에 적자 전환 혹은 점포 수 감소에 이른 안정성이 낮은 기업 사이의 차이점을 비교 및 분석한다.</p> <p>그 결과 안정성이 높은 기업의 경우, 대부분의 키워드들의 변동 폭이 작고, 일정한 빈도를 유지한다는 것을 확인할 수 있었다.</p> <p>신생 벤처 기업의 경우 평가할 만한 지표의 수가 적거나 그 구간이 짧아 신뢰하지 못하는 경우가 많은데, 본 연구를 통해 새로운 지표를 제안한다. 또한 소셜 데이터 분석을 벤처기업 연구에 이용한 첫 번째 연구가 되어 이후 관련 연구의 발판이 될 것이다.</p>

Keyword: 빅데이터, 데이터마이닝, 소셜데이터, 벤처기업, 안정성, 결정요인
--

연구책임자				
성명	한글	위희주	소속 (학교)	KAIST
	영문	Wi Heeju	직위 (과정)	학사
e-mail		bb0711@kaist.ac.kr	연락처	01062914105
연구진척도		100% 진행	제출일	20191111

중소기업청 『대학 기업가센터 지원사업』 연구사업부문의 지원을 받아 수행한 연구의 결과를 제출합니다. 위탁기관에서 해당 결과물을 별도의 연구 및 교육콘텐츠 등의 활용에 동의합니다.

연구책임자: 위희주^{위희주} (서명, 인감)

목 차

I . 서론	3
II . 이론적 배경	4
(1) 기업의 안정성	4
(2) 소셜 데이터 분석	4
III . 분석 과정	6
(1) 분석 기업 설정	6
(2) 데이터 수집	6
(3) 데이터 전처리	7
(4) 데이터 분석 방법	8
IV . 결과 및 분석	9
(1) 시간에 따른 관련 데이터 변화량	9
(2) 시간에 따른 키워드 흐름	12
(3) 기업의 안정성과 소셜 데이터와의 관련성	15
V . 결론	15
(1) 연구 요약 및 의의	15
(2) 한계점 및 향후 연구 방향	16
Reference	17

I. 서론

혁신적인 아이디어와 기술에 기반한 새로운 기업들의 창업과 성장은 국가의 경제 발전에 많은 영향을 미친다. 이는 새로운 시장을 열기도 하고, 고용을 증가시키며, 기존의 기업들의 더 높은 경쟁력을 위한 성장을 이끌기도 한다. 이에 따라 여러 선진국들처럼 우리나라 정부 또한 벤처 기업에 대한 투자를 강화하고 있으며, 여러 기관 외에도 개인이 벤처기업에 투자하는 ‘엔젤투자’ 또한 2019년 현재 역대 최고치를 기록했다(중소벤처기업부, 2019). 지난해부터 개인투자조합 수와 결성 규모가 빠르게 성장하고 있으며, 이에 따라 벤처 기업 수도 급격히 증가하고 있는 추세이다.

하지만 벤처 기업에 대한 투자에는 많은 위험성이 따른다. 벤처기업의 경우, 설립된 지 얼마 되지 않아 기업을 평가할만한 지표의 양이나 질이 신뢰하기 어려운 경우가 많기 때문이다. 그래서 투자자들이 기업의 실적이 뒷받침되지 않은 상태에서 투자를 하게 되고, 이는 기업의 거품을 초래할 수 있고, 더 큰 위험이 따를 수 있다.

이에 따라 투자자들은 투자 기업에 대한 최소한의 안정성을 보장받고 싶어 하며, 기업 측은 이를 보장할 수 있는 지표를 통해 투자자들을 설득하고 싶어 한다. 따라서 벤처기업의 안정성에 대한 연구가 그 어느 때보다 필요한 시점이다. 기존 연구에서는 사례연구나 제한된 수의 표본을 가지고, 기업 내부의 요인들을 가지고 성공 및 실패 요인을 분석하려는 시도가 많았다. 창업자 개인에 대한 요인을 바탕으로 연구를 하기도 하며[1,2], 이에 창업 과정 및 전략, 산업환경, 외부지원 시스템도 포함하여 통합적인 영향력을 파악하려는 연구도 실시되었다[3]. 또한 최근에는 성공요인 뿐만이 아닌, 실패 사례 분석을 통한 실패요인 분석[4,5]이 이뤄지고 있다.

하지만 기존 성공 및 실패 요인 분석은 기업 내부 역량에 집중한 경우가 많으며, 외부 요인이라 해도 산업 환경 및 정부의 정책뿐이었다는 한계가 있다. 이번 연구에서는 외부 요인이 내부요인보다 성공에 큰 영향을 미쳤다는 연구[6]를 기반으로, 무엇보다도 중요한 외부요인인 대중의 의견과 기업의 성공 또는 실패와의 연관성을 파악하고자 한다. 따라서 이번 연구에서는 빅데이터인 소셜 데이터를 분석하여 대중의 반응을 파악하고, 이러한 외부요인이 벤처기업의 안정성과 어떠한 관계를 가지는지 분석해본다. 이 연구는 국내 벤처 기업의 연구에 소셜 데이터 분석을 이용한 최초의 연구이며, 기존의 지표 외에도 또 다른 지표를 제안한다는 데에 의의가 있다.

II. 이론적 배경

1. 기업의 안정성

기존 기업의 '안정성'지표는 기업의 현금흐름과 관련하여 이자를 갚을 수 있는 능력이 되는지 판단하는 데에 쓰인다. 부채비율, 이자보상비율, 유동비율, 차입금의존도, 당좌비율 등이 있다. 부채 비율을 통해 자기 자본으로 부채 비율을 감당할 수 있는지를 판단할 수 있으며, 업종 평균 및 과거 추이와의 비교를 통해 판단이 가능하다. 또한 부채비율이 높아 이자보상이 가능한지, 유동성에 문제가 없는지는 각각 이자보상비율과 유동비율을 통해 알 수 있다. 차입금의존도는 수주계약금이 부채로 산정되는 건설 및 조선업에서 쓰이며, 당좌비율은 재고가 쌓이기 쉬운 의류업계에서 현금화가 쉬운 자산을 파악하기 위해 쓰인다. 이외에도 기업의 결정요인들을 수치화하기 위한 여러 지표들이 있으며, 이러한 지표들은 기업이 직접 발표한 부채, 자기자본총계, 영업이익, 이자비용 등을 통해 계산이 가능하다[7].

이 연구에서 말하는 기업의 안정성은 이러한 지표들을 전부 고려하기보다는, 얼마나 기업이 실패하지 않고 성장해나갈지를 포괄적으로 의미하는 데에 쓰인다.

2. 소셜 데이터 분석

(1) 빅데이터

빅데이터란 기존 데이터의 양을 넘어서는, 대량의 정형 또는 비정형 데이터를 의미하며, 이러한 데이터로부터 필요한 가치를 추출하고 결과를 분석하는 것을 빅데이터 분석이라고 한다. 다양한 종류의 대규모 데이터에 대한 생성, 수집, 분석, 표현을 그 특징으로 하는 빅데이터 기술의 발전은 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동하게 하고 개인화된 현대 사회 구성원 마다 맞춤형 정보를 제공, 관리, 분석 가능하게 하며 과거에는 불가능했던 기술을 실현하기도 한다. 이는 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게 가치 있는 정보를 제공할 수 있는 가능성을 제시하며 그 중요성이 부각되고 있다.

빅데이터의 공통적 특징은 3V로 설명할 수 있다. 속도는 대용량의 데이터를 빠르게 처리하고 분석할 수 있는 속성이다. 융복합 환경에서 디지털 데이터는 매우 빠른 속도로 생산되므로 이를 실시간으로 저장, 유통, 수집, 분석처리가 가능한 성능을 의미한다. 다양성(Varity)은 다양한 종류의 데이터를 의미하며 정형화의 종류에 따라 정형, 반정형, 비정형 데이터로 분류할 수 있다. 빅데이터의 특징은 3V로 요약하는 것이 일반적이다. 즉 데이터의 양(Volume), 데이터 생성 속도(Velocity), 형태의 다양성(Variety)을 의미한다. 최근에는

가치(Value)나 복잡성(Complexity)을 덧붙이기도 한다. 이처럼 다양하고 방대한 규모의 데이터는 미래 경쟁력의 우위를 좌우하는 중요한 자원으로 활용될 수 있다는 점에서 주목받고 있다. 대규모 데이터를 분석해서 의미 있는 정보를 찾아내는 시도는 예전에도 존재했다. 그러나 현재의 빅데이터 환경은 과거와 비교해 데이터의 양은 물론 질과 다양성 측면에서 패러다임의 전환을 의미한다. 이런 관점에서 빅데이터는 산업혁명 시기의 석탄처럼 IT와 스마트혁명 시기에 혁신과 경쟁력 강화, 생산성 향상을 위한 중요한 원천으로 간주되고 있다. 데이터 속도(Velocity)은 다양한 형태의 데이터를 포함하는 것을 뜻한다. 정형 데이터뿐만 아니라 사진, 오디오, 비디오, 소셜 데이터, 로그 파일 등과 같은 비정형 데이터도 포함된다.

이번 연구에서는 빅데이터의 일종인 트위터의 소셜 데이터를 이용한다.

(2) 트위터

트위터는 소셜미디어이자 마이크로블로그 서비스이다. 소셜미디어란 사람들이 자신의 생각과 의견, 경험, 관점 등을 서로 공유하고 참여하기 위해 사용하는 온라인 툴과 미디어 플랫폼이다. 이는 수많은 개인들의 경험과 지식, 정보를 쌍방향으로 생산·소비하는 구조를 가지고 있다. 또한 마이크로블로그는 블로그 서비스의 일종으로, 인터넷에 블로거가 올린 한두문장의 단편적 정보를 해당 블로그에 관심이 있는 개인들에게 실시간으로 전달하는 서비스를 의미한다.

트위터에서는 140자 이내의 단문(트윗)만 작성할 수 있어, 다른 소셜 미디어에 비해 빠르고 간결하게 전달될 수 있다. 또한 트윗들은 개방적 성격을 지니고 있어, 무수히 많은 독자들과 화자가 소통할 수 있고, 답글로 반응할 수 있으며, 이 반응에 대해 또 다른 사용자가 답글을 달 수 있어 급속하게 메시지가 전파될 수 있다. 또한 다른 사용자들과의 실시간 소통을 중요시 여기는 트위터의 특성상, 익명이 아닌 자신의 신분을 밝히고 사용하는 경우가 많아, 신뢰도가 높은 솔직한 개인의 의견을 얻기 쉽다[8].

따라서 국내외의 많은 데이터 분석 연구에서도 트위터 데이터를 이용한 경우가 많다. 데이터의 주기별 변화량을 측정하여 영화 흥행과의 상관성을 비교하거나, 데이터에 감성값을 부여하는 오피니언 라벨링을 통해 영화 추이를 예측하기도 하였다. 하지만 이런 정량적 접근만으로는 그 근거와 속성을 파악하기 어렵기 때문에, 최근에는 이런 빈도수 측정을 통해 트윗의 내용을 대표하는 키워드를 추출하여 그 속성을 밝히고 관객들의 반응을 분석하여 영화 흥행에 영향을 미친 요인들을 제시하는 연구가 진행되었다[9].

III. 분석 과정

1. 분석 기업 선정 및 정의

분석 기업으로 총 5개의 기업을 선정하였다. 소셜 데이터를 얻기 쉬운, 외식 프랜차이즈 기업으로 제한했다. 또한 비슷한 시기에 창업하여야 유사한 소셜 데이터를 얻을 수 있기 때문에, 2006~2016년 사이에 창업한 기업들을 대상으로 하였다.

두 개의 기업은 안정성이 낮은 기업의 예시로, 창업 후 2~3년 내에 적자 전환되어 더 이상 새로운 지점을 내지 않고 그 지점 수가 줄어드는 기업을 선택하였다. 해당 기업에는 ‘대만 카스테라’, ‘망고 식스’를 선택하였다.

다른 세 개의 기업은 창업한지 6~7년이 지났지만 아직도 프랜차이즈 지점 수가 유지되거나 늘고, 매출이 꾸준히 유지가 되는 기업으로 안정성이 높다고 판단하였다. 이에 해당하는 기업은 ‘명랑 핫도그’, ‘백다방’, ‘공차’를 선택하였다.

2. 데이터 수집

GetOldTweets3 API와 python을 이용하여 트위터 데이터를 수집하였다. 각 기업마다 해당 상호명을 포함한 트위터를 수집하도록 하였다. 예를 들면, 명랑핫도그의 경우 ‘명랑 핫도그’를 포함하는 트위터를 수집하였고, 대만카스테라의 경우 여러 상호명을 이용하였기 때문에 ‘대만 카스테라’ 및 ‘대왕 카스테라’를 포함하는 트위터를 수집하였다. 수집 기간은 각 기업의 설립 날짜 이후 최초 발생한 트윗부터 9000개의 트윗을 수집할 때까지로 지정하였다. 트윗 개수를 제한한 이유는 대만카스테라의 경우 창업 후 망하기까지의 기간에서 수집할 수 있는 최대 트윗 수가 9000개였기 때문이다. 동일 조건에서 비교하기 위해 다른 기업의 트윗 개수도 약 9000개가 되도록 하였다. [표 1]은 각 기업별 데이터 수집한 기간을 나타낸다.

분류	기업명	수집 기간
안정성 낮은 기업	대만카스테라	2016-08-05 ~ 2018-01-05
	망고식스	2012-01-01 ~ 2013-01-16
안정성 높은 기업	명랑핫도그	2016-08-21 ~ 2017-12-07
	백다방	2014-01-27 ~ 2016-07-30
	공차	2012-01-01 ~ 2012-10-03

[표 1] 기업 분류와 기업별 데이터 수집 기간

3. 데이터 전처리

(1) 형태소 추출

형태소는 의미를 가지는 가장 작은 말의 단위이다. 실제 글에서는 같은 형태소(예시. 같-)를 가진 단어라도, 여러 가지 형태로 존재할 수 있다(예시. 같은, 같다, 같으며, 같으면서). 이번 연구에서는 같은 의미를 가진 단어들의 흐름을 파악하기 위해, 같은 어근을 가지는 단어는 같은 의미를 가진다고 가정하였다. 따라서 기존 단어에서 형태소만을 추출해 그 단어의 기본형으로 표시하였다. 형태소 추출에는 konlpy library의 Okt module을 이용하였다.

기존 데이터:

친구랑 홍대갔다가 대만카스테라 샀당ㅎㅎㅎ pic.twitter.com/PxtXQxoKAM

전처리한 데이터:

친구 랑 홍대 가다 대만 카스테라 사다 ㅎㅎㅎ

[그림 1] 형태소 추출 및 불용어 제거한 후의 데이터 예시

(2) 불용어 제거

기존 자연어 처리(Natural Language Processing) 연구에 쓰이는 방법이다. 데이터 분석에 의미 없는 불용어(Stop words)를 제거하여, 전체 데이터의 크기를 줄일 수 있고, 의미 있는 데이터를 쉽게 얻을 수 있도록 한다. 주로 불용어로 지정되는 단어들은 조사나, 부사들이 있다. 이번 연구에서는 직접 데이터를 보면서 자주 발생하는 조사나 부사 및 상호명 등을 불용어로 처리하여, 해당 단어들은 데이터에서 제거하였다. 또한 한글이 아닌 이모티콘 및 하이퍼링크 주소 등도 불용어 처리하였다.[그림 1]은 기존 데이터와 형태소 추출 및 불용어 제거를 한 한 트윗 데이터의 예시이다.

(3) 유사 의미 단어 그룹화

소셜 데이터에는 비슷한 의미의 다양한 단어가 쓰일 수 있기 때문에, 자연어 처리(Natural Language Processing)에 자주 쓰이는 단어 벡터를 이용하여 유사한 의미를 지닌 단어를 하나의 그룹으로 만들고, 그 그룹의 대표어로 치환하였다. 기존 한글 자연어 처리에 자주 쓰이는 konlpy 라이브러리의 단순히 단어를 한 개의 토큰(정수)으로 표현하는 토큰화와 다르게, 단어 벡터는 단어를 n차원의 벡터로 나타내어 단어 벡터 자체가 단어의 의미 자체도 포함할 수 있다. 또한 사전에 등재된 형식 언어뿐만 아니라, 실제 인터넷에서 사용자들이 많이 사용하는 언어의 의미를 파악하고 학습할 수 있다는 장점이 있다.

단어 벡터 학습과 생성에는 여러 방법이 있는데, 이번 연구에서는 해외에서 단어벡터를

이용한 연구에서 가장 빈번히 쓰이는 fasttext API의 CBOW 방법을 이용하였다. 이는 학습하고자 하는 단어의 주위 단어들의 평균 벡터값을 이용해 문맥을 학습해, 그 단어를 벡터로 만들어준다. 단어 벡터 생성을 위해 충분히 많은 한글 데이터로 학습된 모델[10]을 이용하였다. 단어 벡터의 학습을 위해 한국 Wikipedia와 common crawl의 한글 텍스트 데이터를 이용하였으며, 벡터의 차원은 300, 5-grams(문맥 파악을 위해 주위 5개의 단어를 살핌)으로 설정하였다.

벡터화된 한글 단어를 가지고, 단어 유사도 검사에 자주 쓰이는 cosine 유사도를 이용하여 유사도 0.75 이상을 비슷한 의미를 가진 단어라고 판단하고 하나의 단어로 치환하였다. 이를 통해 온라인에서 자주 쓰는 단어도 학습이 잘 되는 것을 확인할 수 있었다. ‘존맛’, ‘맛남’, ‘남’, ‘졸맛’, ‘핵존맛’ 등이 같은 단어로 판단되었고, ‘증말’, ‘완전’, ‘느므’ 도 같은 그룹의 단어로 판단되었다. 이외에도 요일을 나타내는 단어들(수요일, 목요일, 금요일 등), 성별을 나타내는 단어(여자, 남자) 등이 같은 의미의 단어로 표현되었다. [표 2]에서 볼 수 있듯이, 유사한 의미를 가지는 고유 단어 수가 평균 642.8개에서 72.4개의 그룹으로 표현되어, 유사한 의미를 가지는 단어에 대해 단어의 수를 11%로 압축할 수 있었다. 해당 단어들의 경우, 실제로 자주 쓰이는 경우가 많으므로, 다음 단계에서 비슷한 경향성을 가지는 단어들을 찾는 데에 도움을 주었다. 고유 단어는 해당 관련 데이터에서 추출된 중복되지 않는 단어 수를 의미한다.

	전체 고유 단어 수	유사한 의미를 가지는 단어 수	유사 의미 단어 그룹 수
대만카스테라	12761	770	76
망고식스	11417	288	69
뽕다방	11767	498	74
공차	12223	412	75
명랑핫도그	11576	1246	68

[표 2] 유사한 의미로 그룹화된 고유 단어 수

4. 데이터 분석 방법

데이터 분석은 크게 두 가지로 이뤄진다. 이는 두 가지 가설을 검증하기 위해 설계되었다. 첫 번째 가설은 기업의 안정성이 발생한 트윗 수와 관련이 있을 것이란 것이다. 이를 확인하기 위해 안정성 있는 기업과 그렇지 않은 기업 그룹간의 시간의 흐름에 따른 트윗 발생수를 비교해본다. 두 번째 가설은 특정 키워드 단어들의 시간의 흐름에 따른 경향성이 안정성 있는 기업 그룹과 그렇지 않은 기업 그룹 간에 차이가 있을 것이란 것이다. 이를 검

증하기 위해 특정 구간에서 빈번히 발생한 키워드들을 추출하고, 각 기업별 시간에 따라 비슷한 흐름을 갖는 키워드들끼리 분류한다. 그룹 내의 키워드 단어들의 의미를 파악하여, 해당 그룹의 의미를 찾고, 기업별로 각 키워드 그룹의 흐름이 어떤지 비교분석한다.

(1) 시간에 따른 관련 발생 트윗 수 분석

각 기업별로 시간의 흐름에 따른 발생한 관련 데이터량을 알기 위해, 트윗 및 리트윗 수를 분석한다. 각 구간을 7일로 설정하고, 각 해당 상호 명을 포함하는 트위터 데이터들의 개수를 분석하였다. 트위터에는 본인이 직접 작성하는 트윗 외에도 리트윗(retweet)이라는 다른 사람의 의견을 자신의 의견처럼 퍼뜨릴 수 있는 기능이 있다. 리트윗 수를 통해 해당 의견에 대한 대중들의 공감 정도를 파악할 수 있다. 따라서 단순히 작성한 트윗 수뿐만 아니라, 리트윗수 또한 고려하였다.

(2) 시간에 따른 키워드의 흐름 분석

시간에 따른 키워드의 흐름 분석은 다음과 같이 이뤄진다.

기업별 소셜 데이터에서 자주 발생하는 키워드를 추출하기 위해 각 단어별로 빈도를 측정한다. 이 빈도를 기반으로 각 기업별 상위 20개 정도의 키워드를 추출한다. 이 키워드들에 대해 구간을 트윗 150개 단위로 나누고, 시간의 흐름에 따른 빈도를 확인하고, 유사한 증감을 보이는 키워드들을 같은 그룹으로 묶는다. 이후 해당 그룹의 의미를 찾고, 기업별로 각 그룹의 흐름을 비교 분석한다. 그룹의 키워드들의 데이터 상 의미를 파악하여, 각 그룹의 의미를 파악하고, 안정성 높은 기업과 낮은 기업의 차이점을 알아내고, 같은 분류 내에서 공통점을 찾아낸다.

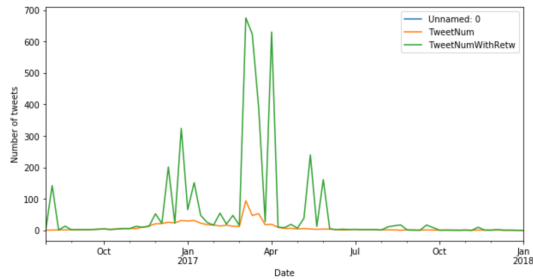
IV. 결과 및 분석

1. 시간에 따른 관련 데이터 변화량

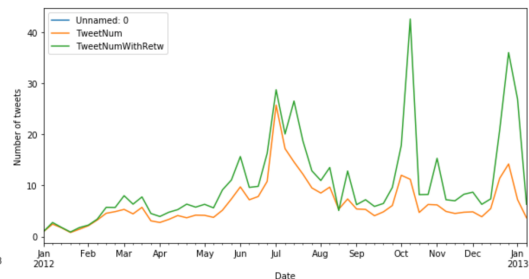
다음 [그림 3]~ [그림 6]에서 초록색 꺾은선 그래프는 리트윗을 모두 고려한 트윗 수를 의미하며, 노란색 꺾은선 그래프는 트윗 수만을 의미한다.

안정성이 있는 기업과 불안정한 기업 모두 특정 구간에 대해 리트윗 수가 급증한 구간이 있음을 알 수 있다. 또한 ‘대만카스테라’, ‘뽕다방’, ‘명량핫도그’는 트윗 수에 비해, 리트윗수가 급증한 구간이 꽤나 많으나, ‘망고식스’와 ‘공차’의 경우 트윗 수 그래프와, 리트윗수 그래프가 유사하게 증감함을 확인할 수 있다. 또한 ‘대만카스테라’와 ‘뽕다방’은 급증한 구간이 많으나, 이에 반해 나머지 기업들은 급증한 구간이 적음을 알 수 있다. 따라서 안정한

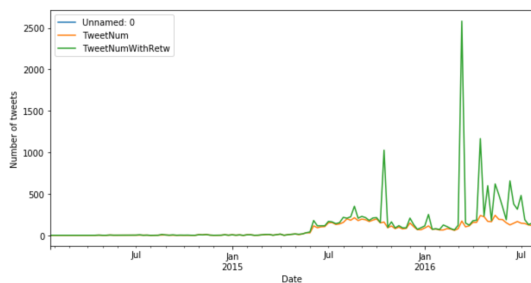
기업과 불안정한 기업 간의 차이를 소셜 데이터 변화량, 즉 시간별 트윗 및 리트윗 수만으로 판단하기는 어렵다.



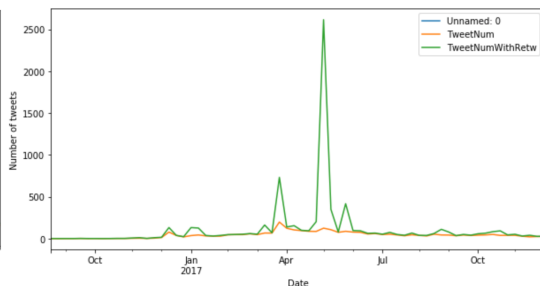
[그림 3] 대만카스테라의 구간별 트윗 수



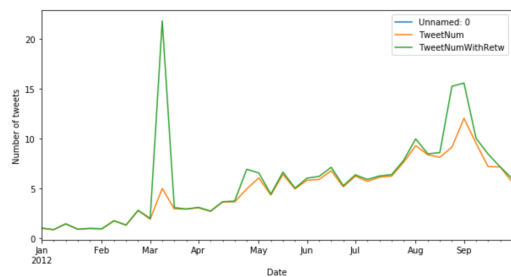
[그림 4] 망고식스의 구간별 트윗 수



[그림 5] 백다방의 구간별 트윗 수



[그림 6] 명랑햇도그의 구간별 트윗 수



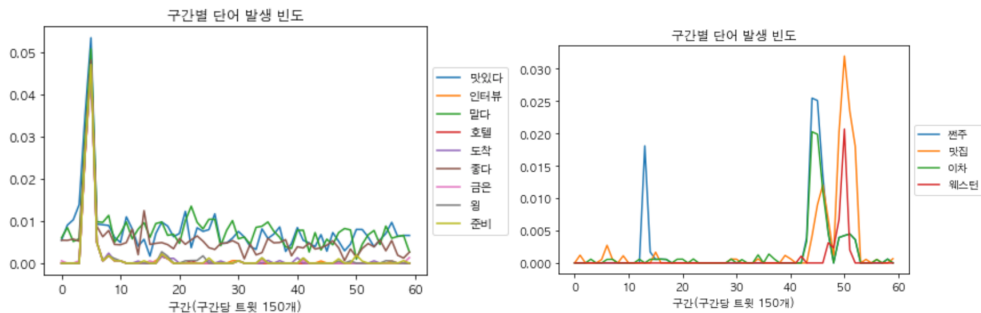
[그림 7] 공차의 구간별 트윗 수

2. 시간에 따른 키워드 흐름

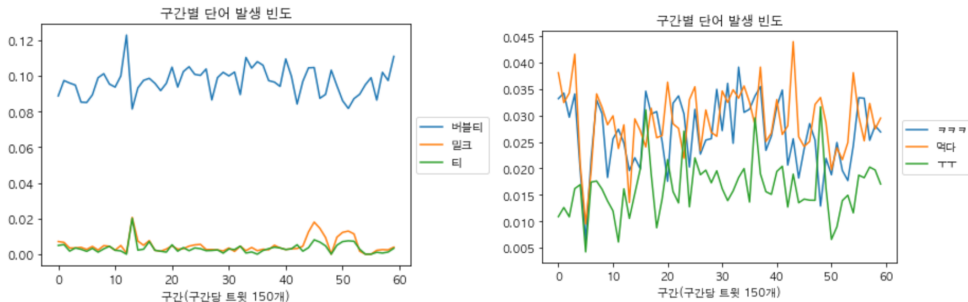
안정한 기업과 불안정한 기업으로 분류한 총 5가지 기업들에 대해 각각 키워드들의 흐름을 알아보고, 그 분류된 그룹 내부에서의 공통점 및 차이점을 찾는다.

(1) 안정한 기업의 경우

‘공차’의 경우 [그림 8]에서 볼 수 있듯이, 특정 구간에 급증 후 급락한 키워드들이 있다. 해당 키워드는 ‘맛있다’, ‘좋다’, ‘금은’, ‘왕’, ‘맛집’ 으로 대부분 긍정적인 어휘이거나 해당 장소를 의미하는 ‘호텔’, ‘웨스턴’, ‘도착’ 등이 있다. 또한 [그림 9]에서 볼 수 있듯이 해당 상품 명인 ‘버블티’, ‘밀크’, ‘티’ 등은 매우 일정한 빈도를 유지하며, 자모음 이모티콘인 ‘ㅋㅋㅋ’와 ‘ㅌㅌ’ 그리고 ‘먹다’는 변동이 꽤 있지만 꾸준히 변동이 있는 상태를 유지한다는 특징이 있다.

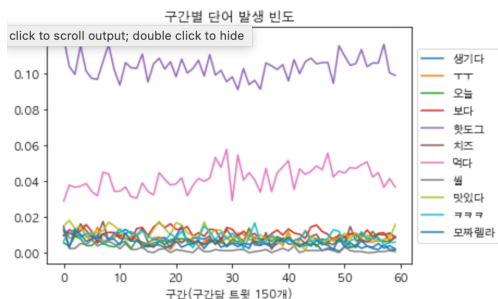


[그림 8] 공차- 급증 후 급락



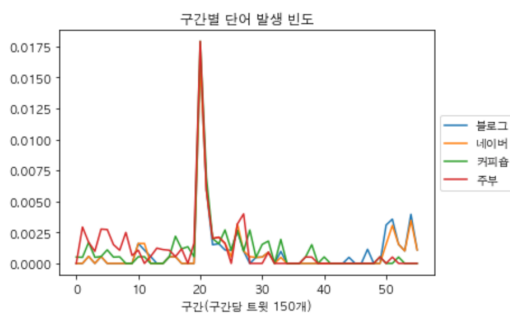
[그림 9] 공차- 일정

‘명랑핫도그’의 경우는 모든 키워드들이 아주 유사한 빈도를 가진다는 특징이 있었다. 이는 [그림 10]에서 확인할 수 있다.

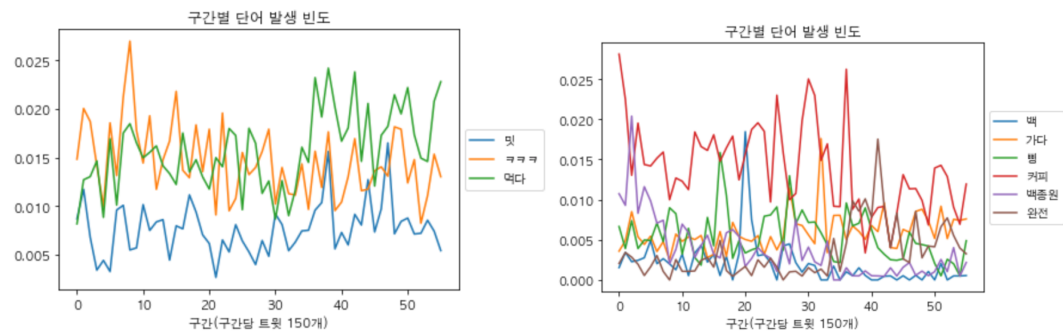


[그림 10] 명랑핫도그 - 일정

‘백다방’의 경우, [그림 11]에서 볼 수 있듯이 ‘네이버’, ‘블로그’, ‘커피숍’, ‘주부’ 등의 특정 키워드들에 대해 급증 후 급락하는 것을 확인할 수 있다. 해당 단어들은 네이버 블로그의 글들을 트위터로 퍼오면서 출처를 남기면서 발생하였다. 따라서 회사에서 직접 SNS를 통해 광고를 하였다기 보다는, 해당 기업이 특별한 트렌드를 만들면서 이를 앞다투어 알리려는 트윗으로 인해 발생했을 것이라고 추측할 수 있다. 또한 해당 기업 역시 ‘맛’, ‘ㅋㅋㅋ’, ‘먹다’ 라는 키워드들이 대해 꾸준히 변동하는 경향을 보였다. [그림 12]의 ‘가다’, ‘빵’, ‘커피’, ‘완전’ 등의 관련 아이템 명사 및 동사에 키워드들에 대해서도 꾸준히 변동하는 경향을 보였다.



[그림 11] 백다방- 급증 후 급락



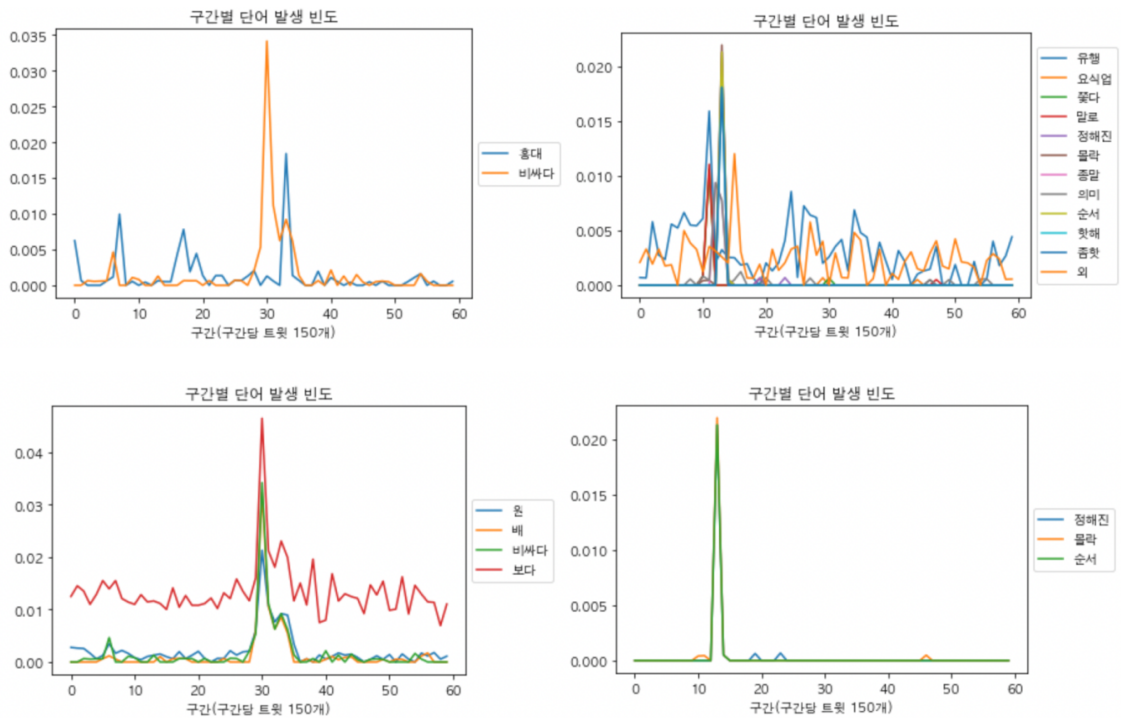
[그림 12] 백다방- 일정

(2) 불안정한 기업의 경우

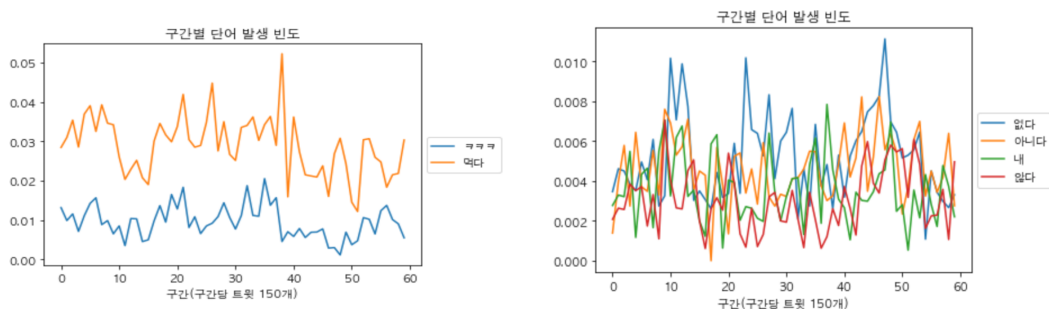
‘대만 카스테라’의 경우, 평상시에 매우 낮은 빈도를 유지하다가 특정 구간에 급증 후 급락하는 키워드들이 많았다. [그림 13]는 이런 급증 후 급락하는 키워드들을 비슷한 경향성을 가진 키워드들 끼리 묶어 그래프로 나타낸 것이다. 해당 키워드들은 긍정적, 부정적 의미를 골고루 갖고 있으며, 유사한 의미를 가지고 있다고 보기 어렵다. 하지만 [그림 13]의 두 번째 그래프의 키워드들에 ‘유행’, ‘쫄다’, ‘정해진’, ‘몰락’이 있는데, 이를 보면 어떤

맥락에서 쓰였고, 대중들이 주로 어떤 반응을 했는지 확인할 수 있다.

[그림 14]에서 볼 수 있듯이, 불안정한 기업의 경우에도 비교적 꾸준히 높은 빈도를 유지하는 단어들도 있음을 알 수 있다. 예를 들면 ‘ㅋㅋㅋ’와 ‘먹다’는 꾸준히 비슷한 빈도를 유지하였다. 또한 [그림 15]의 부정적인 언어인 ‘아니다’, ‘않다’, ‘없다’와 나의 의견을 피력할 때 자주 쓰는 1인칭 대명사인 ‘내’ 또한 변화가 있지만 그래도 꾸준히 변화가 일어나고 그 평균은 유사함을 확인할 수 있다. 이를 통해 소셜 데이터 내의 부정적인 단어의 발생 빈도 및 1인칭 대명사의 빈도와 기업의 안정성은 큰 상관성이 없음을 알 수 있다.



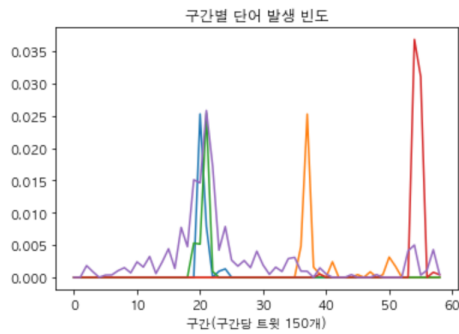
[그림 13] 대만카스테라- 특정 구간 급등하는 키워드 4



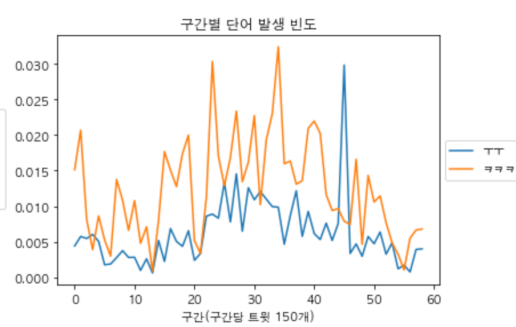
[그림 14] 대만카스테라- 큰 변화 없는 키워드

[그림 15] 대만카스테라- 부정적 어휘 및 1인칭

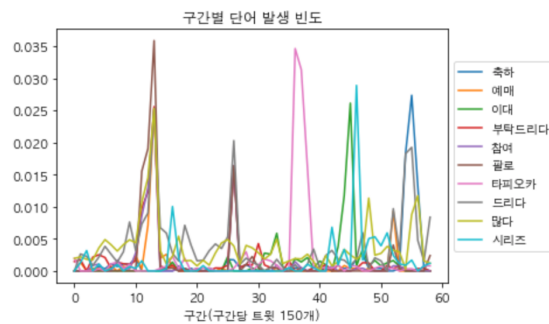
‘망고식스’의 경우 거의 모든 키워드들에 대해 그래프가 매우 큰 폭으로 요동치는 것을 확인할 수 있다. 일부 키워드들이 대부분 낮은 빈도를 유지하다가 일부 구간에서 매우 큰 폭으로 요동치는 것에 반해, ‘ㅌㅌ’, ‘ㅋㅋㅋ’ 등의 자모음 이모티콘은 다른 양상을 보임을 확인할 수 있다. 따라서 이러한 자모음 이모티콘은 기업의 안정성과는 연관 짓기 어렵다.



[그림 16] 망고식스- 연예인 관련 키워드



[그림 17] 망고식스- 자모음 이모티콘



[그림 18] 망고식스- 특정 구간에 급변하는 키워드

‘망고식스’의 경우 유일하게 연예인 이름의 키워드들이 많이 등장하였다. 해당 이름을 포함하는 트윗들을 확인해보니, 망고식스 측에서 연 이벤트 관련 트윗이 많았다. 이벤트 관련 키워드들이 전체 키워드의 상위권을 차지했다는 것은, 망고식스 측에서 트윗을 활용한 광고를 활발히 했으나, 소비자들의 반응보다 이벤트 내용의 비중이 많았다는 것을 의미한다. 동시에 이는 대중의 반응이 크지 않았음을 의미하고, 이러한 이벤트를 통해 대중의 더 큰 반응을 이끌기 어려운 경우도 있다는 것을 의미한다. 또한 정확한 의미를 파악하기 위해, ‘이벤트’라는 키워드를 가진 데이터들은 제거했음에도 불구하고, 해당 내용의 트윗은 사람이 직접 보지 않는 이상 이벤트 내용을 포함한다는 것을 알기 어려웠다. 따라서 추후에 소셜 데이터를 이용해 연구를 할 경우, 이 이벤트 관련 데이터들을 제거하는 방안을 생각해 볼 필요가 있다.

(3) 기업의 안정성과 소셜 데이터와의 관련성

하나의 안정한 기업('명랑핫도그')을 제외하고, 모든 기업들의 데이터가 평소 매우 낮은 빈도를 유지하다가 특정 구간에 급증 후 급락한 키워드 그룹을 가지고 있었다. 이러한 키워드들은 하나가 아니라, 비슷한 구간에 유사한 변화를 가지는 여러 키워드들을 동반한다는 공통점이 있다. 하지만 그 키워드들의 의미에 차이가 있었다. 안정한 기업의 경우 긍정적인 의미를 지닌 키워드들이 많았으며, 반면에 불안정한 기업은 홍보용으로 기업에서 배포하였기에 연예인 이름 같이 아주 관련이 없는 키워드들이거나, '몰락', '비싸다', '유행', '쫓다' 등의 부정적인 키워드들이 많았다.

또한 빈도가 높으나 기업의 안정성과 관계없는 키워드들도 있었다. 예를 들면 국내 인터넷에서 자주 사용하는 자모음 이모티콘인 'ㅋㅋㅋ', 'ㅠㅠ' 와 '먹다'는 안정성 및 시기랑 관련 없이 일정한 변동 폭과 평균 빈도를 가졌다. 따라서 이러한 이모티콘과 직접적으로 아 이템과 관련된 동사는 큰 관련성이 없음을 알 수 있었다.

안정성이 높은 기업의 경우 부정적인 단어들이 상위에 위치하지 않았다. 하지만 '대만 카스테라'의 경우에 '않다', '아니다', '없다' 등의 부정적인 단어가 상위에 위치하였으며, 하지만 그 시기와 상관 없이 높은 변동 폭으로 꾸준히 변동함을 알 수 있었다.

V. 결론

(1) 연구 요약 및 의의

소셜미디어는 실시간으로 자신의 생각과 의견을 생산하고 공유할 수 있는 환경을 사용자에게 제공해줌으로써, 사용자가 특정 대상을 소비하고 난 뒤의 솔직한 의견과 피드백을 즉각적으로 받을 수 있는 정보원의 역할을 한다. 따라서 소셜 미디어를 통해 올라오는 소비자의 의견과 반응은 기업의 성공 추이를 예측하고 분석하는데 매우 중요한 데이터이다.

하지만 기존 벤처기업의 성공 및 실패 요인의 분석은 기업가나 전략 등의 기업 내부 요인, 정책 및 산업 환경 등의 기업 외부 요인에 한해서만 진행되었다. 이는 기업에 무엇보다도 중요한 소비자의 의견을 반영하지 않았다는 한계점이 있다. 따라서 본 연구는 이러한 소비자의 의견을 실시간으로 수집할 수 있는 소셜 데이터를 이용해 소비자의 의견 흐름을 파악하고자 했으며, 이와 기업의 안정성과의 상관성을 파악하고자 하였다. 창업 후 인기를 끌다가 2년 이내에 급격한 몰락을 견게 된 '대만 카스테라'와 비슷한 기간에 창업되었지만 꾸준히 번창하고 있는 다른 외식 프랜차이즈를 비교 분석하였다.

예상과는 다르게 관련 데이터의 변화량이 기업의 안정성에 상관 관계가 없음을 알아냈다. 하지만 발생 빈도가 높은 키워드들의 시간에 따른 경향성은 안정성이 높은 기업과 그렇지 않은 기업 사이에 분명한 차이가 있었다.

안정성이 낮은 기업의 경우 부정적인 단어들이 상위 빈도를 차지하지만, 그 변동 폭이 크고 변동이 주기적으로 일어난다. 또한 특정 구간에 갑자기 발생 빈도가 급증 후 급락하는 단어들이 대부분 부정적인 어휘이거나 홍보를 위해 사용된 전혀 관련 없는 단어였다는 특징이 있다. 반면에 안정성이 높은 기업의 경우, 대부분의 키워드들의 변동 폭이 크지 않았고, 특정 구간에 급증 후 급락한 일부 키워드들의 경우에는 긍정적이 어휘이거나, 특정 장소를 의미하는 단어였다. 또한 예상 외로 기업 아이템과 직접적인 연관이 있는 동사인 '먹다'와 자모음 이모티콘의 경우에는 큰 관련이 없는 것으로 나타났다.

따라서 이 연구 결과를 바탕으로, 분석하고자 하는 기업의 소셜 데이터에서, 특정 단어가 매우 빠르게 급증하고 이후 짧은 기간 내에 급락하고, 해당 단어가 긍정적인 의미를 지니고 있지 않다면 기업의 안정성에 유의해야할 필요가 있다. 대부분의 키워드들이 작은 변동 폭과 일정한 발생 빈도를 가진다면 안정성이 높은 기업이라고 평가할 수 있다.

본 연구는 최초로 벤처기업의 성공 및 실패 요인 분석에 소셜 빅데이터를 이용한 연구이며, 추후 관련 연구의 발판이 될 수 있다. 또한 기존의 지표 외에도 기업의 성공과 실패를 예측하는 새로운 지표를 제안했다는 점에서 의미가 있다. 이를 통해 특정 키워드를 가지는 여론이 기업의 안정성에 미치는 영향력을 파악할 수 있었으며, 여론 분석에 대한 사업가의 주관적인 의견이 아닌 객관적인 수치화된 자료를 제공할 수 있다.

(2) 한계점 및 향후 연구 방향

앞에서 설명된 분명한 시사점에도 불구하고, 여러 한계점을 가지고 있으며 따라서 다음은 한계점과 이를 해결하기 위한 향후 연구 방향을 제시한다.

첫째, 연구대상표본의 수와 특징이 제한적이다. 따라서 이를 모든 기업에 일반화하기 어려우며, 일부 기업의 경우 관련 소셜 데이터의 수가 부족해 소셜 데이터 분석을 하기 어려운 경우도 있을 것이다. 따라서 충분한 소셜 데이터가 있는 벤처 기업을 대상으로, 외식프랜차이즈가 아닌 다양한 분야의 벤처기업들을 선정하여 연구를 진행하면 좀 더 일반적인 결과를 얻을 수 있을 것이다.

둘째, 우리나라에서는 트위터 말고도, 페이스북, 인스타그램 등의 다양한 소셜 미디어를 이용한다. 각 소셜 미디어의 특징이 다르기 때문에, 사용자의 분포도 차이가 있을 것이다. 따라서 다양한 소셜 미디어의 데이터를 이용해 분석을 하면 더 폭 넓은 사용자들의 의견을 수집할 수 있다.

셋째, 분석하려는 대상 기업이 비상장기업인 경우가 많아, 회사의 자세한 재정 및 운영 상태를 알기 어려웠고 따라서 기업의 안정성을 비교할만한 지표가 부족했다는 한계점이 있다. 따라서 추후 연구에서는 재정 및 운영상태가 투명한 기업에 대해 여러 지표를 구하고, 이 지표들을 이용하면 소셜 데이터의 직접적인 연관성을 파악할 수 있을 것이다.

Reference

- [1] 이병기. (2002). 학술연구 : 벤처기업 성장요인 실증분석. 기업가정신과 벤처연구 (JSBI)(구벤처경영연구), 5(2), 93-109
- [2] 이장우, 장수덕. (2001). 벤처기업의 성공요인 : 성공한 벤처기업가들의 관점. 중소기업연구, 23(4), 23-49.
- [3] 용세중, 이상천, 고봉상. (2003). 학술연구 : 벤처기업의 성과 결정요인에 관한 실증연구. 기업가정신과 벤처연구(JSBI)(구 벤처경영연구), 6(2), 3-33.
- [4] 장수덕(Chang Soo-Duck).(2005). 실패한 벤처기업의 특성 : 사례연구. 경영연구, 20(2): 21-46
- [5] 권기환, 최종인. (2015). 기술벤처의 실패요인 분석. 벤처창업연구, 10(4), 27-37.
- [6] 용세중, 이상천, 고봉상. (2003). 학술연구 : 벤처기업의 성과 결정요인에 관한 실증연구. 기업가정신과 벤처연구(JSBI)(구 벤처경영연구), 6(2), 3-33.
- [7] 김병기. (2018). 상장기업 부채비율의 안정성에 관한 연구. 금융공학연구, 17(1), 1-25.
- [8] 하일규 (2014), “소셜 네트워크 서비스의 연구경향 분석:Twitter 관련 연구 중심”, 한국콘텐츠학회논문지, 제14권, 제9호, pp.567-581.
- [9] 이오준, 박승보, 정다울, 유은순. (2014). 소셜 빅데이터를 이용한 영화 흥행 요인 분석. 한국콘텐츠학회논문지, 14(10), 527-538.
- [10] Edouard Grave and Piotr Bojanowski and Prakhhar Gupta and Armand Joulin and Tomas Mikolov (2018), Learning Word Vectors for 157 Languages, grave2018learning, cs.CL