

유튜브 먹방과 온라인 배달 주문: 영향력 분석과 예측 모형*

최사라

한양대학교 일반대학원 경영학과 경영정보시스템전공
(sarah525@hanyang.ac.kr)

이상용

한양대학교 경영대학 경영학부 교수
(tomlee@hanyang.ac.kr)

음식 문화 및 산업과 관련한 대표적 특징들 중에는 음식 배달 주문 산업이 성장하고 있다는 것과 유튜브와 같은 1인 미디어에서의 소위 ‘먹는 방송’ (먹방)이 최고의 인기 콘텐츠로 자리 잡았다는 사실 등을 거론할 수 있다. 본 연구는 이러한 배경에 근거하여 두가지 초점을 두어 연구하고자 하였다. 먼저, 유튜브 먹방과 먹방 댓글에서 확인되는 대중들의 감성이 관련 음식의 배달 이용 건수에 영향을 미치는지를 회귀분석 모형을 통하여 확인하고자 하였다. 다음으로, 대한민국에서 대표적인 주문 음식인 치킨의 배달 이용 건수 데이터와 유튜브 먹방 댓글 데이터와 날씨 데이터를 활용하여, 머신 러닝을 통한 치킨 배달 주문 예측 모형을 구현하였다. 2015년 6월 3일부터 2019년 9월 30일까지 총 1,580개의 데이터를 활용하였고, 날씨 변수로서의 온도, 습도, 강수량과 유튜브 먹방 변수로의 영상에 달린 댓글 수, 댓글의 긍정어 수, 중립어 수, 부정어 수 등을 수집하였다. 본 연구에 활용된 데이터의 유튜브 먹방과 먹방 댓글의 감성이 배달 이용 건수에 영향 미침을 확인하기위해 선형 회귀 방법론을 사용하였으며, 예측모형을 위해 사용된 머신 러닝은 Linear Regression, Ridge, Lasso, Random Forest, Gradient Boost이다. 본 연구를 통해 유튜브 먹방과 댓글의 감성이 배달 이용 건수에 영향 미침을 확인하였고 예측 모형 또한 기존 모델보다 성능이 좋아짐을 Root Mean Square Error 값을 통하여 확인하였다. 본 연구는 먹방의 광고 효과를 확인하였으며, 배달 업종에서의 경영에 활용할 수 있는 함의를 제공하고자 하였다.

주제어 : 온라인 배달 주문, 유튜브 먹방, 감성 분석, 머신 러닝, 예측 모형

논문접수일 : 2022년 9월 4일 논문수정일 : 2022년 10월 19일 게재확정일 : 2022년 11월 3일
원고유형 : Regular Track 교신저자 : 이상용

1. 서론

우리나라의 음식 배달 주문의 역사는 매우 길다. 1768년 냉면을 시작으로 지금의 배달 문화가 생성되었다고 한다. (김정엽, 2016). 이러한 음식 배달 주문은 배달 주문 플랫폼의 등장으로 Information Technology (IT)와 결합하면서 질적인 변화를 겪었으며, 2020년 코로나19로 인해 주문량이 폭발적으로 증가하게 되었다. 뉴데이일리 경제에 따르면 코로나19가 지속되면서 2021년 배달 음식

주문이 78.6%나 증가했다고 말한다(김보라, 2019). 이렇게 배달 주문은 우리에게 없어서는 안 될 의미있는 산업으로 자리잡게 되었다.

한편, 음식과 관련된 미디어 문화로서 중요한 트렌드 중의 하나가 소위 ‘먹는 방송 (먹방)’이며, 먹방은 상업적으로 효과가 큰 것으로 나타나고 있다. 예를 들어, BBQ는 2019년에 방영한 ‘멜로가 체질’이란 드라마에서 브랜드 홍보 효과를 위해 간접광고 (product placement)를 진행하였다. 그 결과 BBQ에 따르면 2019년 8월 16일 3화에

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A3A2A02093277)

황금 올리브 치킨 PPL 먹방 장면이 방영이 되었을 때, 해당 관련 메뉴의 매출이 평균 138% 상승하였으며, 8월 31일 8회에 극한왕갈비치킨이 등장했을 때에는 해당 제품의 판매량이 125%까지 급증하였다고 한다(김보라, 2019). 이렇게 미디어가 배달에 주는 영향 또한 무시하지 못한다.

IT의 발전이 이루어지면서, 음식과 관련된 미디어 문화도 급속도로 변화하고 있다. 트렌드가 바뀌고, IT의 발전으로 요즘 미디어의 형태도 변화하고 있다. 사람들은 TV보다 유튜브와 같은 소셜미디어를 통해 정보를 소비하고 있으며, 이에 따라 수많은 1인 크리에이터들이 등장하게 되었다. 1인 크리에이터들의 콘텐츠는 뷰티, 코미디, 게임 방송, 음식, 먹는 방송(이하 먹방)까지 다양한데(송정은, 장원호, 2013), 그 중에서도 먹방의 콘텐츠의 인기가 1위를 차지하고 있다(김경진, 2019).

그렇다면, 기존 드라마와 같이 TV에서의 먹방 뿐만 아니라, 유튜브에서의 먹방도 배달 음식 주문에 영향을 미칠 수 있다고 생각할 수 있다. 따라서, 본 연구는 먹방과 먹방 영상 댓글의 감성이 배달 주문 이용 건수에 미치는 영향을 분석하고자 하였으며, 아울러 배달 주문 건수 예측 모형을 수립하고자 하였다. 본 연구를 통하여 판매자는 배달 이용 건수를 예측하여 하루에 필요한 물량을 준비하고, 1인 크리에이터들은 자신의 콘텐츠가 배달 이용에 영향이 있음을 파악하여 광고 효과를 확인할 수 있을 것으로 예상된다. 또한 ‘고객구매예측모형’은 최근 마케팅 분야에서 중요하게 생각되는 이슈 중에 하나로 여기고 있다. 고객의 구매행동을 정확하게 파악할 수 있는 능력을 기업이 보유하고 있는 경우, 그 기업은 이를 이용해 다양한 사업기회를 발굴, 육성할 수 있기 때문이다(Chiu, 2002).

본 연구는 다음과 같이 구성되어 있다. 다음

의 2장에서는 본 연구와 관련된 선행 연구 및 연구 대상에 대해 설명한다. 3장에서는 본 연구에 사용된 방법론에 대해 설명한다. 4장에서는 유튜브 먹방 영상 댓글 수, 댓글의 감성, 날씨변수들이 배달 이용 건수의 영향에 대한 선행 회귀 결과를 설명한다. 또한 배달 이용 건수예측모형에 대한 연구 결과를 설명한다. 5장에서는 본 연구의 결론을 제시한다. 끝으로 6장에서는 본 연구의 시사점과 한계점을 제시하고자 한다.

2. 선행연구

2.1. 소셜미디어에 대한 감성 분석 및 텍스트 마이닝

현대시대에는 과거의 비해 한 상품에 대한 정보와 실제 사용해본 후기들에 대한 정보를 쉽게 소셜미디어를 통해 알아볼 수 있으며 많은 사람들은 소셜미디어를 통해서 상품에 대한 정보를 확인한 후 구매를 하기 시작했다. 이에따라 경영정보학에서는 소셜미디어의 감성 분석을 활용한 연구들이 많이 존재한다. 감성분석이란 인터넷에서 블로그나 SNS등과 같은 소셜미디어에서 개인 및 집단이 가진 의견 등의 데이터를 수집하고 분석하여 그들의 태도, 의견, 성향과 같은 주관적인 데이터를 분석하는 자연어 처리 기술로서, 다양한분야에 걸쳐 연구가 수행되고 있다(Peng & Park, 2011). 정지선 등(2015)은 KOSPI200의 상위 종목들을 분석 대상으로 선정하여 개별 기업의 온라인 뉴스에 데이터에 대한 감성 분석을 바탕으로 개별 기업의 주가 변화 예측을 하였다. 조정태와 최상현(2015)은 영화 리뷰 데이터를 활용하여 평점을 예측하는 연구를 진행하였다. 이와 같이 소셜미디어의 텍스트를 통하여 사

람들의 감성을 수집하고, 소셜 감성이 기업에 어떠한 영향을 미치는지에 대한 연구가 많이 진행되고 있으며, 본 논문도 예측모형을 구현할 뿐만 아니라 추가로 소셜미디어의 감성이 소비자의 구매에 어떠한 영향을 미치는지에 관한 연구도 진행한다.

2.2. 예측모델에 관한 선행연구

해외 선진국에서의 배달 산업은 코로나19 이후에야 급증하는 추세이기 때문에 이와 관련하여 진행된 해외 연구도 아직은 많지 않은 실정이다. 국내의 경우에는 배달 이용 건수와 관련된 데이터가 공개되면서, 배달 이용 건수 예측모델 관련 연구가 시작되었다. 권재영 등(2015)이 날씨데이터를 활용한 배달 이용 건수예측모형 연구를 진행했다. 그 결과 훈련데이터로 예측했을 때 보다 테스트 데이터로 예측했을 때에 선형 회귀 모델과 랜덤 포레스트에서 Roots Mean Square Error(RMSE) 값이 적게 나와 성능이 좋아짐을 확인할 수 있었다. 하지만 그라디언트 부스팅과 서포트 벡터 머신과 신경망 모형에서는 성능이 좋아짐을 확인할 수 없었다(권재영 등, 2015). 그 외에 김대룡 등(2016)은 날씨에 따른 배달음식 주문 건수 예측 연구도 있었으며, 권태용 등(2018)의 기상 자료와 기계 학습을 통한 배달 주문 수요예측 연구도 있었다.

배달 이용 건수를 예측하는 연구는 많이 없었지만 머신 러닝을 활용한 예측모델연구는 많이 있다. 김경재 등(2005)은 서포트 벡터 머신(Support Vector machine)을 이용한 고객구매예측모형에 관한 연구를 통해 서포트 벡터 머신(Support Vector machine)이 로지스틱 회귀 모형과 인공 신경망(ANN)보다 성능이 좋음을 밝혔다. 이러한 연구

들이 있지만 아직까지 유튜브를 텍스트 마이닝(Text Mining)과 관련하여 진행한 연구는 없다. 본 연구는 기존연구들이 활용하였던 날씨데이터와 수집한 데이터인 유튜브 먹방 영상의 댓글 데이터를 통해 예측 모형을 연구하고자 한다.

2.3. 유튜브 방송의 구매 영향에 관한 선행연구

유튜브의 사용률이 높아짐에 따라 유튜브에 관한 연구가 많이 진행되고 있다. 이맏탁 등(2020)은 유튜브의 콘텐츠에 대한 즐거움에 따라 시청자들은 콘텐츠가 제공하는 정보에 대해 수용의도가 커진다고 말한다. 이러한 영향은 결과적으로 유튜브에서 인플루언서가 소개하는 제품에 대한 구매의도가 높아진다고 기술수용모델(TAM)을 통하여 확인했다(이맏탁 등, 2020). 또한 다중 회귀분석을 통하여 유튜브 광고가 고객 구매의도에 영향을 미침을 확인하는 연구도 있다(Firat, 2019).

‘먹방’콘텐츠의 시작은 아프리카TV의 ‘먹쇼’로 시작했다. ‘먹쇼’가 많은 관심을 갖게 되면서 유튜브에서도 방송 형식을 모방하여 먹방 콘텐츠를 만들었다. 그러나, 지금은 유튜브의 시장점유율이 압도적인 관계로, 먹방 크리에이터들의 다수가 유튜브에서 활동하고 있다(홍석경, 박소정, 2016). 따라서 본 연구는 다른 소셜미디어보다 유튜브에 관한 연구가 더 의미가 있다고 생각하여 유튜브데이터를 활용하여 기존 연구들의 유튜브 방송의 구매 영향을 다시 한번 확인하고 유튜브 댓글 데이터를 활용하여 배달예측모형을 연구하고자 한다.

3. 연구 방법

3.1. 데이터

본 연구에서 활용한 데이터는 SKtelecom Bigdatahub사이트 (<http://www.bigdatahub.co.kr>)에서 제공하는 오픈 소스 데이터인 ‘치킨의 배달 이용 현황분석’에 있는 배달이용건수 데이터이다. 배달 음식의 종류가 많지만 그중 ‘치킨’의 배달 이용 건수 데이터를 사용한다. 연합뉴스 디지털스토리에 따르면 한국인이 가장 많이 시킨 배달음식 1위 메뉴는 치킨(30%)이라고 한다(전승엽 등, 2017). 따라서, 본 연구에서는 배달 음식 중 ‘치킨’의 관한 데이터를 종속변수로 활용하여 연구하고자 한다. 독립변수로 기상청에서 제공하는 날씨 데이터와 유튜브 먹방 영상 댓글 수와 댓글이 달린 날짜 데이터를 활용하였다.

먼저, ‘먹방’ 유튜버를 선정하기 위해서 유튜브 순위를 제공하는 사이트(<https://youtube-rank.com/>)를 참고하였다. 사이트에서 제공하는 순위를 활용하여 각 연도별로 조회수와 구독자 수를 기준으로 ‘먹방’ 유튜브를 확인하였다. 그렇게 유튜브 먹방 영상 데이터는 조회수가 높은 먹방 관련 영상 중에 치킨을 다룬 먹방 60개 영상의 데이터를 python 웹 크롤러를 만들어 수집하였다. 또한 대댓글에 경우 영상에 관련된 내용보다 대댓글을 한 댓글에 관련된 내용들이 많다고 판단되어 분석 범위에서 제외하였으며, 댓글 데이터만 활용하였다. 유튜브 영상의 댓글 내용, 댓글 수, 댓글 작성 날짜 등을 알려주는 Hazy 사이트(<https://hadzy.com/>)를 통해 데이터를 수집하였으며, 총 수집된 댓글

은 60,000건이다.

영상의 업로드 날짜와 관계없이, 댓글이 달린 날을 시청한 날로 간주하여 분석하고자 하였으며, 당일 뿐만 아니라, 댓글이 1일 뒤의 주문량에 미치는 영향, 댓글이 2일 뒤의 주문량에 미치는 영향, 댓글이 3일 뒤의 주문량에 미치는 영향 등을 각각 회귀분석하고자 하였다.¹⁾ 다음으로 기상청(<http://www.kma.go.kr>)에서 날씨 변수로 온도(temperatures), 강수량(precipitation), 습도(humidity), 일사량(insolation), 운량(cloud), 지면 온도(ground_temperature), 체감온도(per_temperature) 등을 수집하였다. 하지만 날씨 변수들간의 다중공선성이 확인되어, 회귀분석에 활용된 날씨 변수는 온도(temperatures), 강수량(precipitation), 습도(humidity) 등이다. 종속변수의 주문이 행해진 날의 날씨 변수가 독립변수로 활용되었다. 총 사용된 데이터의 기간은 2015년 6월 3일부터 2019년 9월 30일까지이며, 총 1,580개의 데이터를 활용하였다. 다음 Table 1은 본 연구에 사용된 종속변수와 독립변수에 대한 설명이다.

3.2. 형태소 분석

감성 분석을 진행하기에 앞서 형태소 분석을 우선적으로 행하였다. 형태소란 뜻을 가지고 있는 최소한의 말의 단위이며, 본 연구에서는 python에 라이브러리에서 가장 잘 활용되고 있는 꼬꼬마(Kkma)를 활용하여, 동사, 명사, 형용사 등을 추출하였다. 이렇게 형태소 분석을 진행한 후의 데이터의 형태는 Table 2와 같다.

1) 통계청 조사(<https://www.korea.kr/news/visualNewsView.do?newsId=148864074>)에 따르면, 유튜브 방송 시청은 21-24시대에 집중되어 있고, 연령대가 낮을수록 밤-새벽 시간대를 이용하는 것으로 확인된다. 따라서, 댓글이 다음날 또는 그 다음날 주문으로 연결될 가능성이 있기 때문에, 본 논문은 lag 효과를 살펴보았으며, 그 범위를 3일까지로 하였다.

〈Table 1〉 Summary of the key variables

Key variables	Description
Dependent variable	
Delivery	The number of deliveries
Independent variables	
Comment	The number of data on the video.
Positive	The number of positive words in the comments
Neutral	The number of neutral words in the comments
Negative	The number of negative words in the comments
Temperatures	Weather temperature
Humidity	Weather Humidity
Precipitation	Weather precipitation

〈Table 2〉 Morpheme result

Morpheme
맛있다, 예쁘다, 사랑, 좋아하다, 최고, 대박, 행복하다, 재밌다, 부럽다, 응원, 만족, 대단하다, 깔끔하다, 바삭, 웃기다, 파이팅, 잘생기다, 매력, 땡기다, 행복, 가능하다, 만족하다, 치킨, 영상, 복회, 썬양, 한국, 음식, 보이다, 처음, 생각, 나오다, 말, 구독, 방송, 치즈볼, 닭다리, 느낌, 올리다, 모르다, 모습, 다이어트, 짹다, 궁금하다, 들어가다, 먹어주다, 씹다, 버리다, 쓰레기, 안되다, 짜다, 아프다, 걱정, 겁나다, 별로, 힘들다, 싫다, 졸다, 아쉽다, 맛없다, 밍다, 안좋다, 이상하다, 죄송하다, 늦다, 느끼하다, 사라지다, 불편하다, 떨어지다

3.3. 감성 분석

감성 분석은 NLP(Natural Language Processing, NLP) 기술이다. 감성 분석은 오피니언 마이닝(Opinion Mining)으로도 불리우는데, 텍스트에 나타난 사람들의 느낌과 감정과 주관성 등을 분석하는 기술이다. 따라서 텍스트가 어떠한 감성을 나타내고 있는가에 초점을 맞춰 분석해야 한다. 이러한 감성 분석을 활용하면 문서 또는 문장에 어떠한 견해를 가지고 있는지 판별할 수 있

다. 따라서 감성 분석은 소셜 감성을 분석하는데 유용하다(김영민 등, 2014).

감성 분석을 위해 사용된 감성 사전은 군산대에서 구축한 KNU한국어 감성 사전을 활용하여 구축하였다. KNU 한국어 감성 사전은 기본 한국어의 사전적인 의미를 기준으로 구축된 감성 사전이지만, 상황에 따라 단어의 감성을 모두 포괄하지는 못한다. 따라서 본 연구에서는 KNU 한국어 감성 사전을 기반으로, 먹방 상황에 맞는 감성 사전을 새로이 구축하였다. 본 연구에서 감성 사전을 구축하기 위해 먹방 댓글에서 수집한 단어들의 빈도 분석을 통하여 100번이상 빈도를 나타내는 단어들을 정리하였다. 정리한 후 본 연구자 포함 한 총 5명의 연구자들이 감성을 긍정, 부정, 중립으로 나누어 분류작업을 진행한 후 감성 사전에 추가하여 먹방에 맞는 감성 사전을 구축하였다. 그렇게 감성 사전에 구축된 단어들 예시는 Table 3과 같다.

〈Table 3〉 Sentiment dictionary example

Positive	Neutral	Negative
맛있다, 예쁘다, 사랑, 좋아하다, 최고, 대박, 행복하다, 재밌다, 부럽다, 응원, 만족, 대단하다, 깔끔하다, 바삭, 웃기다, 파이팅, 잘생기다, 매력, 땡기다, 행복, 가능하다, 만족하다	치킨, 영상, 복회, 썬양, 한국, 음식, 보이다, 처음, 생각, 나오다, 말, 구독, 방송, 치즈볼, 닭다리, 느낌, 올리다, 모르다, 모습, 다이어트, 짹다, 궁금하다, 들어가다, 먹어주다, 씹다	버리다, 쓰레기, 안되다, 짜다, 아프다, 걱정, 겁나다, 별로, 힘들다, 싫다, 졸다, 아쉽다, 맛없다, 밍다, 안좋다, 이상하다, 죄송하다, 늦다, 느끼하다, 사라지다, 불편하다, 떨어지다

3.4. 회귀 분석 모형

본 연구는 유튜브 댓글 데이터와 날씨데이터의 배달 이용 건수에 대한 영향을 확인해 보기

위해서 다중 회귀분석을 사용하였다. 회귀분석은 통계학에서 사용되는 대표적인 방법 중 하나이며, 여러 자료들 간의 관계성을 수학적으로 추정하여 그들의 관계를 설명한다. 추가로 구하려는 해와 실제 해의 오차의 제곱의 합이 최소가 되기 위해 최종적으로 최소 제곱 법(Ordinary Least Squares, OLS)을 사용하였다. 본 연구에서 회귀 분석에 사용한 독립변수는 먹방 영상 댓글 수, 긍정어 수, 중립어 수, 부정어 수, 온도, 습도, 강수량이다. 종속변수는 치킨의 배달 이용 건수이다.

3.5. 예측모델

본 연구의 예측 모형에는 선형회귀모형(Linear, Ridge, Lasso), 랜덤 포레스트(Random Forest), 그라디언트 부스팅(Gradient Boosting)을 사용하였다. Ridge는 다중 공선성을 방지하는 장점이 있고, Lasso는 예측 오차를 줄일 수 있다는 장점이 있다. Ridge와 Lasso는 회귀 계수에 벌점을 주어 계수를 축소하는 방법이라는 점에서 동일하지만, 모든 설명 변수를 고려하는 Ridge와 다르게 Lasso는 변수 선택의 기능을 가지고 있다는 점에서 차이가 있다(Tibshirani, 1996).

랜덤 포레스트(Random Forest)는 다수의 의사결정 나무를 결합하여 하나의 모형을 생성하는 방법이며 의사결정나무의 수가 증가할수록 예측 오차가 줄어든다. 또한 의사결정나무의 수가 많아도 과적합(Overfitting)이 발생하지 않는다는 장점이 있는 모델이다(Breiman, 2001). 그라디언트 부스팅(Gradient Boosting) 방법은 앙상블형태의 약한 예측모형들의 응집으로 강한 예측 모형을 생성한다. 이는 SVM (support Vector Machine)이나 뉴럴 네트워크 등과 비슷한 성능을 보여주면서도 해석할 수 있는 결과를 보여준다(Guelman, 2012).

추가로 본 연구에서는 예측모형을 일반화하기 위해 AIC(Akaike Information Criterion)와 BIC(Bayesian Information Criterion)를 사용하였다. 이는 최대 우도에서의 독립변수의 개수에 대한 손실(penalty)분을 적용하는 방법이다. 이를 정보량 기준(information criterion)이라고 한다. 손실가중치의 계산법에 따라 AIC와 BIC 두 가지가 있다. AIC는 Akaike(1974)가 제안한 모델로 예측모형 선택을 위해 일반적으로 사용되어지고 있는 기준이다. BIC는 Schwarz(1978)가 제안한 모델로 AIC의 과적합이 있는 모형에 벌점을 주는 형식으로 인해 최적합 모형을 가장 적은 AIC값을 주는 문제점을 보완한 방법이다. AIC와 BIC 값은 작을수록 좋은 것을 의미한다.

4. 분석 결과

4.1. 기초 통계와 상관관계

본 연구에서는 유튜브의 먹방 영상이 배달 이용 건수에 미치는 영향을 분석하기 위해 2015년 6월 3일부터 2019년 9월 30일까지 총 1,580일간의 데이터를 추출하였다. Table 4는 분석 기간 동안의 배달 이용 건수, 댓글 수, 긍정어 수, 중립어 수, 부정어 수, 온도, 강수량, 습도 에 대한 기초 통계량이며, Table 5는 상관관계 계수를 보여주고 있다. 댓글 수와 감성 관련 변수들이 서로 상관관계가 있는 것이 확인되기에, 강건성 검증 부분에서는 이 문제를 해결하기 위한 추가 분석을 진행하였다.

4.2. 선형 회귀(Linear regression)

Table 6은 유튜브 먹방 영상 댓글의 감성이 배

〈Table 4〉 Describe statics

	mean	std	min	max
deliver	16410.77	4909.31	5982	34606
comment	25.53	78.51	0	1268
positive	1.38	3.97	0	43
neutral	2.542	8.99	0	134
negative	78.87	251.02	0	2948
temperatures	14.18	10.92	-14.8	33.7
humidity	58.64	14.79	21.8	97.3
precipitation	3.08	11.64	0	144.5

〈Table 5〉 Correlation of variables

	Delivery	Comment	Positive	Neutral	Negative	temperature	Humidity	Precipitation
Delivery	1.0000							
Comment	-0.1539	1.0000						
Positive	-0.1263	0.8209	1.0000					
Neutral	-0.1746	0.8989	0.7368	1.0000				
Negative	-0.1568	0.9423	0.8026	0.8995	1.0000			
temperature	-0.0976	-0.0318	-0.0362	-0.0087	-0.0331	1.0000		
Humidity	0.0742	-0.0405	-0.0429	-0.0213	-0.0420	0.3978	1.0000	
Precipitation	0.0560	-0.0189	-0.0152	-0.0150	-0.0199	0.1408	0.4538	1.0000

달 이용 건수에 미치는 영향을 보여주는 회귀분석 결과표이다. 그런데, 유튜브는 TV프로그램과 같이 정해진 시간에 방영이 되는 것이 아니며, 영상을 게시하면 게시자가 게시물을 삭제하지 않는 한 계속 시청이 가능하다. 이러한 점을 고려하여 댓글을 단 당일과 그 이후의 영향도 확인해보고자 하였다. 따라서 댓글이 달린 당일 뿐만 아니라 댓글이 달리고 나서 1일 뒤, 2일 뒤, 3일 뒤에 영향도 확인하기 위해 총 4개의 모델을 만들었다. 즉, 댓글이 당일 주문량에 미치는 영향 (Model1), 댓글이 1일 뒤 주문량에 미치는 영향 (Model2), 댓글이 2일 뒤 주문량에 미치는 영향

(Model3), 댓글이 3일 뒤 주문량에 미치는 영향 (Model4) 등의 4개 회귀 모형을 만들었으며, 이때 사용된 변수는 Table 1에서 이미 설명한 바 있다. 이렇게 모델들에 사용된 데이터 수는 Model1은 1,580개, Model2는 1,579개, Model3는 1,578개, Model4는 1,577개이다.

먼저 댓글이 당일의 주문에 영향을 미친 효과를 살펴본 Model 1의 결과를 보면 강수량을 제외하고는 대체적으로 유의함을 확인할 수 있다. 그 중 댓글 수와 댓글내에서의 긍정어의 수가 배달 이용 건수에 양(+)의 영향이 있고, 댓글 내에서의 부정어 수는 배달 이용에 음(-)의 영향이 있

〈Table 6〉 OLS result

OLS (dependent variable: number of delivery orders)				
independent variables	Model1 (same day)	Model2 (lag 1day)	Model3 (lag 2day)	Model4 (lag 3day)
(intercept)	8.9607*** <2e-16	9.2719*** <2e-16	9.7984*** <2e-16	9.7568*** <2e-16
temperature	-0.1263*** 3.78e-14	-0.1310*** 3.42e-15	-0.1310*** 8.07e-16	-0.1324*** 2e-16
humidity	0.1216** 0.0427	0.1252** 0.0034	0.1083** 0.0095	0.1004* 0.0154
precipitation	0.0382 0.1512	0.0357 0.1800	0.0433 0.0953	0.0474 0.0656
comment	0.1783*** 0.001	-0.0390** 0.0015	-0.5319*** 3.35e-08	-0.0598*** 3.00e-12
positive	0.4078*** 7.47e-07	0.3370*** 3.29e-05	0.2428** 0.0028	0.2892*** 0.0003
neutral	-0.1438* 0.0177	-0.0983 0.0755	-0.163** 0.0013	-0.1866*** 0.0006
negative	-0.2301*** 6.38e-15	-0.0945*** 2.95e-08	-0.0732*** 1.23e-05	-0.728*** 8.01e-06
R- square	0.156	0.157	0.195	0.268
Adjusted R-square	0.152	0.154	0.192	0.207
F	41.31	41.8	54.32	59.78

**** 0.00, *** 0.001, ** 0.01

음을 알 수 있다. 시차 효과를 보는 Model 2, Model 3, Model 4의 경우에도 긍정어의 수는 양의 효과가, 부정어의 수는 음의 효과가 지속적으로 유지되는 것이 확인된다. 댓글의 감성이 당일 주문에 미치는 효과 (Model1)가 대체로 가장 크게 나타나며, 1일 후(Model2), 2일 후(Model3)에서 계수의 크기가 감소하다가, 3일 후(Model4)에서 상대적으로 작기는 하나 다소 증가하는 것으로 나타났다. 결과적으로, 댓글의 감성변수들이 주문량에 미치는 효과가 며칠간 지속된다고 해석할 수 있었다.

따라서, 먹방에 대한 사람들의 긍정 감성은 배

달 주문을 증가시키는 효과가 있고, 부정 감성은 배달 주문을 감소시키는 효과가 며칠간 지속된다고 볼 수 있다. 다만, 댓글 수 자체는 이틀째부터는 음의 효과로 바뀌는 것으로 나타나는데, 이에 대하여서는 조금 더 연구가 필요하다고 판단된다. 아마도 댓글 작성하는 사람들이 당일 배달 주문을 하게 되면, 몇일 간은 동일 음식 주문을 안하게 되는 효과가 있을 것으로 예상된다. 반면에, 다른 날에 방송 및 댓글을 보는 사람들에게도 댓글의 감성은 지속적으로 영향을 미침으로써, 배달 주문에 영향을 미치는 것으로 생각된다.

4.3. 강건성 검증(Robustness Check)

〈Table 7〉 Regression result with polarity

OLS (dependent variable: number of delivery orders)	
independent variable	
(Intercept)	8.97607*** <2e-16
temperature	-0.12619*** 5.24e-14
humidity	0.13424** 0.002
precipitation	0.03486 0.193
comment	0.18286*** 1.62e-06
polarity	0.29927*** 4.58e-16
R- square	0.141
Adjusted R-square	0.138
F	51.41

**** 0.00, *** 0.001, ** 0.01

앞서 상관관계 분석에서 감성 변수들 간에 상관관계가 큰 것을 확인하였다. 따라서 본 연구는 이를 확인하기 위해 추가로 강건성 검증을 하였다. 강건성 검증은 감성변수를 감성지수변수로 계산하여 분석하였다. Table 7은 감성지수변수로 회귀분석을 한 결과이다. 여기서 사용된 감성지수변수는 다음 [1] 식을 통해 산출하였다.

$$\frac{(\text{긍정어 수}) - (\text{부정어 수})}{(\text{긍정어 수}) + (\text{부정어 수})} \quad [1]$$

다음 강건성 확인을 위한 회귀 결과는 강수량을 제외하고 모든 변수가 유의함을 확인할 수 있다. 따라서 감성이 서로 영향이 있어도 감성 지수를 통해 유티브의 먹방 영상과 영상의 댓글 감성이 배달이 영향이 있다고 볼 수 있다.

4.4. 배달 이용 건수 예측 모델

배달 이용 건수 예측모형을 구현하는데 사용된 모델은 Linear, Ridge, Lasso, 랜덤 포레스트, 그라디언트 부스트를 사용하였다. 예측 모형에 사용된 데이터는 치킨 배달 이용 건수 데이터와 유티브 댓글 수, 긍정어 수, 부정어 수, 중립어 수, 온도, 습도, 강수량 데이터로 선형회귀분석에서 사용한 4개의 model들의 변수들을 모두 사용하였다. 회귀분석에서는 종속변수인 주문량을 기준으로 당일, 1일전, 2일전, 3일전 댓글을 분석하는 모델을 제시하였지만, 예측모델에서는 더 좋은 예측을 위해 당일 댓글, 1일전 댓글, 2일전 댓글, 3일전 댓글 등을 모두 다 활용한 한가지 모델만 활용하였다.

선행연구 모델은 선행연구(권재영 등 2015, 김대룡 등 2016, 권태용 등 2018)에 근거하여 날씨 변수들을 활용한 모델이다. 본 연구 모델은 앞에서 회귀분석을 통해 유의성을 확인한 유티브 먹방에 관한 각 모델의 변수들과 날씨변수들을 함께 활용한 모델이다. 이 두 예측 모델들의 성능을 실제값과 예측값의 차이를 확인하는 Root Mean Square Error (RMSE) 값과 모든 절대 오차의 평균을 확인하는 Mean Absolute Error(MAE)를 활용하여 비교하였다. Table 8은 선행연구 모델과 본 연구 모델2의 예측 성능을 비교하는 표이다

〈Table 8〉 RMSE of delivery order predictive models

RMSE(MAE)	선행연구 모델	본 연구 모델
Linear	4959.626(3885.450)	4764.665(3852.250)
Ridge	4959.626(3885.448)	4746.336(3801.342)
Lasso	4961.435(3884.623)	4840.497(3824.268)
Random Forest	5306.144(4212.605)	4800.431(3922.791)
Gradient Boost	5467.977(4333.280)	5016.367(4156.106)

선행연구 모델에서는 Linear, Ridge, Lasso, Random Forest, Gradient Boost 순으로 4959.626(3885.450), 4959.626(3885.448), 4961.435(3884.623), 5306.144(4212.605), 5467.977(4333.280) 이며, 본 연구 모델은 각각 4764.665(3852.250), 4746.336(3801.342), 4840.497(3824.268), 4800.431(**3922.791**), 5016.367(4156.106) 이다. 모든 결과값에서 본 연구 모델이 기존 연구에 기반한 모델보다 우수함이 확인되었다. 특히 Random Forest의 RMSE, MAE값 둘다 성능이 가장 좋아진 것을 확인할 수 있었으며, 그 다음으론 Gradient Boost, Ridge, Linear, Lasso 순이었다. 따라서 선행연구의 기존 변수인 날씨 변수만 사용했을 때보다 날씨 변수와 유튜브 댓글 데이터를 동시에 사용하였을 때의 예측의 성능이 좋아짐을 확인할 수 있었다.

RMSE 이외에도 예측 모형의 성능을 비교할 때 사용되는 AIC (Akaike Information Criterion) 과 BIC (Bayesian Information Criterion)를 이용하여 본 연구 모델의 예측력을 비교하여 보았다. Table 9와 Table 10은 각각 AIC와 BIC의 비교값을 보여주고 있는데, 공히 본 연구에서의 예측 모델이 더 작은 값을 갖고 있음이 확인된다. 따라서, 본 연구의 예측력이 더 우수하다고 할 수 있었다.

〈Table 9〉 AIC(Akaike Information Criterion) result

	선행연구 모델	본 연구 모델
Linear	8107.630	8069.533
Ridge	8107.631	8065.871
Lasso	8107.977	8084.533
Random Forest	8168.366	8077.563
Gradient Boost	8203.195	8118.069

〈Table 10〉 BIC(Bayesian Information Criterion) result

	선행연구 모델	본 연구 모델
Linear	8157.589	8119.493
Ridge	8157.591	8115.831
Lasso	8157.937	8134.493
Random Forest	8218.326	8127.523
Gradient Boost	8253.155	8168.029

5. 결론

본 논문에서는 유튜브 먹방 영상의 댓글 데이터를 활용하여 배달 이용 건수에 미치는 영향을 확인해보고 날씨 데이터와 수집한 유튜브 댓글 데이터를 활용하여 배달 이용 건수 예측 모델을 구현하였다.

본 연구의 결과는 다음과 같이 요약할 수 있다. 첫째, 본 유튜브 먹방 영상이 배달 이용에 미치는 영향을 회귀분석을 통하여 확인하였다. 유튜브의 영상은 언제 어디서나 볼 수 있는 시간적 제한이 없기 때문에 TV나 라디오처럼 정해진 시간에 영향을 미치는 것을 확인하기 힘들다. 이러한 이유로, 본 연구에서는 먹방 영상이 업로드한 당일의 댓글 및 감성의 영향력 뿐만 아니라, 영상 업로드 1일, 2일, 3일 뒤까지도 영향도 확인하였다. 긍정 감성은 당일 뿐만 아니라 3일 뒤까지도 배달 이용을 증가시켰으며, 부정 감성은 반대로 배달 이용을 감소시키는 것이 확인되었다. 긍정 부정을 포괄하는 polarity 점수에서도 긍정성이 높을수록 배달 주문이 증가됨이 확인되었다. 또한 댓글 자체는 당일의 주문을 증가시키는 효과가 있으나, 그 다음날부터는 음의 효과가 있음이 확인되었다.

둘째, Linear, Ridge, Lasso, Random Forest, Gradient Boost 이렇게 총 5가지의 머신 러닝 기법을 사용하여 배달 이용 건수 예측 모델 실험을 진행하였다. 전반적으로 모든 모델이 날씨 변수만 포함했을 때 보다 유튜브 댓글 데이터를 포함했을 때에 RMSE 값과 MAE값이 작아지는 것을 확인할 수 있었으며, Random Forest 모델의 성능이 가장 크게 좋아지는 것을 확인할 수 있었다. 예측모형들의 일반화를 확인하기 위하여 AIC와 BIC의 값을 비교하였을 때에도, 기존 연구에 기반한 모델보다 유튜브 먹방 영상 데이터를 포함했을 때의 성능이 개선됨을 확인하였다.

본 연구는 날씨 데이터만 활용하여 배달 이용 건수 예측 모형을 만든 기존 연구와 다르게 유튜브 먹방 영상의 댓글을 텍스트 마이닝을 통하여 연구한 차이점이 있다. 또한 유튜브 댓글 데이터를 활용하여 댓글 수와 감성 분석을 통한 데이터를 통해 배달 이용 건수에 미치는 영향을 확인하고 배달 이용 건수 예측 모형을 구현하였다는 차이점이 있다. 이는 새로운 미디어 형태의 영향을 시사하게 되는 것과, 먹방이라는 새로운 관점으로 먹방 영상의 영향을 확인할 수 있는 연구이다.

본 연구는 대표적으로 두 가지 실무적인 이익을 줄 수 있다. 첫 번째로, 최근에 1인 가구가 늘어나고 경제활동 인구의 범위가 늘어남에 따라 배달 이용 건수가 증가하고 있다. 더불어 코로나19로 인해 배달 이용 건수가 급증하게 되었다. 반면에 배달원 부족 사태가 발생함에 따라 배달 플랫폼의 입장에서는 배달 주문 예측이 필요한 시점이 되었다. 본 연구는 배달 플랫폼 또는 배달업에서 활용할 수 있는 배달 주문 예측 모형을 제공함으로써, 효율적인 배달원 확보 및 발전에 도움을 줄 수 있을 것이다.

두번째로, 개인이 스스로 자신있는 콘텐츠를

제작하여 쉽게 동영상을 올릴 수 있고, 그 영상을 통해 구독자들의 관심을 받게 되면 연예인만큼의 인기를 얻을 수 있게 되었다. 이렇게 유튜버들의 활동이 증가하면서 그들이 쓰는 화장품 리뷰, 먹방, 패션 등 여러 가지 콘텐츠를 구성한 영상들이 늘어나게 되었고, 이러한 영상들을 구독하는 사람들이 증가하면서 이에 대한 구매 효과도 무시할 수 없게 되었다. 이러한 이유로 기업들은 유명 유튜버에게 협찬을 제공하며 광고를 하고 있다. 그 중 먹방은 가장 큰 관심을 받고 있는 콘텐츠이다. 따라서 본 연구의 결과는 TV, 라디오와 같은 기존 매체의 광고 효과를 벗어나 ‘먹방’ 영상이 갖는 주문 증대 효과를 측정함으로써, 비용 대비 좋은 성과를 기대하는 음식 업주로 하여금 유튜브 먹방과 같은 곳을 광고의 대상으로 선택할 수 있는 근거를 제시하였다.

다른 연구와 마찬가지로 본 연구도 한계점을 가지고 있다. 우선, 본 연구는 날씨의 영향이 다소 적은 치킨 데이터를 활용하였다는 한계점을 지니고 있다. 본 연구는 2015년부터 2019년까지 구독자 수가 높으면서 조회수도 높은 영상들을 가진 유튜버를 선정하여 해당 유튜버의 영상 중 치킨에 관련된 영상의 댓글을 수집하여 데이터로 활용하였다. 따라서, 조회수 높은 영상이 우선순위가 아닌 유명 유튜버의 영상을 우선순위로 선정하여 데이터를 수집하였다는 점에서 데이터 수집의 한계점을 가지고 있다.

또한, 데이터의 기간이 2015년부터 2019년까지의 데이터로 배달이용이 급증한 코로나19의 상황을 반영하지 못한 한계점을 가지고 있다. 특히, 코로나19라는 특수 상황에서 배달 주문이 급격히 증가하고, 또한 주문의 패턴도 과거와 다른 양상을 보일 가능성이 있을 것으로 보인다. 따라서, 코로나19 이후의 먹방의 효과와 배달 주문이

과거와 어떻게 차이를 보일지에 대한 연구는 또 다른 향후 과제가 될 것으로 판단된다.

향후 연구에서는 유튜버와 상관없이 해당 먹방에 관련된 영상의 조회수 순위로만 영상을 선택한 후 데이터를 수집하여 연구할 필요가 있을 것으로 보인다. 또한 2020년 유튜버들의 협찬을 받아 광고를 했음에도 불구하고 광고라고 표시를 안하고 영상을 게시한 ‘뒷광고 사건’으로 인해 2020년도부터 해당 영상에 광고인 영상과 광고가 아닌 영상으로 구분을 지어 명시하게 되었다. 따라서 광고인 영상과 광고가 아닌 영상을 각각 수집하여 광고인 영상들과 광고가 아닌 영상들의 영향을 확인해보는 것이 향후 연구과제가 될 것이다.

참고문헌(References)

- 권재영, 김시내, 박은지, & 송종우. (2015). 국내 배달음식 이용 건수 분석 및 예측. *응용통계연구*, 28(5), 977-990.
- 권태용, 권재영, 신주일, 양다영, 여현구, 이영재, & 윤상후. (2018). 기상자료와 기계학습을 통한 배달주문 수요예측. *한국기상학회 학술대회 논문집*, 540-540.
- 김경재, 한인구, & 안현철. (2005). Support Vector Machine 을 이용한 고객구매예측모형. *지능정보연구*, 11(3), 69-81.
- 김대룡, 김다영, & 변수지. (2016). 날씨에 따른 배달음식 주문건수 예측. *한국기상학회 학술대회 논문집*, 480-481.
- 김영민, 정석재, & 이석준. (2014). 소셜 미디어 감성분석을 통한 주가 등락 예측에 관한 연구. *Entrue Journal of Information Technology*, 13(3), 59-70.
- 송정은, & 장원호. (2013). 유튜브(YouTube) 이용자들의 참여에 따른 한류의 확산: 홍콩의 10-20대 유튜브(YouTube) 이용자조사를 중심으로. *한국콘텐츠학회논문지*, 13(4), 155-169.
- 이매탁, 이준영, & 심성욱. (2020). 유튜브 (YouTube) 뷰티 인플루언서 속성이 콘텐츠 태도, 제품 태도, 구전의도, 구매의도에 미치는 영향 연구. *광고학연구*, 31(5), 117-142.
- 정지선, 김동성, & 김종우. (2015). 온라인 언급이 기업 성과에 미치는 영향 분석: 뉴스 감성분석을 통한 기업별 추가 예측. *지능정보연구*, 21(4), 37-51.
- 조정태, & 최상현. (2015). 영화리뷰 감성 분석을 통한 평점 예측 연구. *경영과 정보연구*, 34(3), 161-177.
- 홍석경, & 박소정. (2016). 미디어 문화 속 먹방과 해게모니 과정. *언론과 사회*, 24(1), 105-150.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Chiu, C. (2002). A case-based customer classification approach for direct marketing. *Expert systems with Applications*, 22(2), 163-168.
- Firat, D. (2019). YouTube advertising value and its effects on purchase intention. *Journal of Global Business Insights*, 4(2), 141-155.
- Guelman, L. (2012). Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, 39(3), 3659-3667.
- Peng, W., & Park, D. H. (2011, July). Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization. In *Fifth International AAAI*

- Conference on Weblogs and Social Media.*
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- 김경진. (2019.04.25). [ONE SHOT] 유튜브 분야별 최고 인기 채널 & 인기 유튜버 톱3. 중앙일보 <https://dcnewsj.joins.com/article/2343250>
- 김보라. (2019.09.27). ‘멜로가체질’뜨자BBQ매출 경충_PPL전략’눈에띄네’. 시장경제. <https://www.meconomynews.com/news/articleView.html?idxno=33472>
- 김보라. (2019.09.26).BBQ치킨, JTBC 드라마 ‘멜로가 체질’ 등장에 매출 평균 138% 상승. 백세시대. <http://www.100ssd.co.kr/news/articleView.html?idxno=64279>
- 김정엽. (2016.05.04). 배달음식1호, 1968년7월냉면. 서울엔. http://www.seouland.com/arti/culture/culture_general/354.html
- 전승엽, 박성은, 김유정. (2017.09.11). [디지털스토리] 한국인이 가장 많이 시키는 배달음식은. 연합뉴스. <https://www.yna.co.kr/view/AKR20170905129800797>

Abstract

Youtube Mukbang and Online Delivery Orders: Analysis of Impacts and Predictive Model

Sarah Choi* · Sang-Yong Tom Lee**

One of the most important current features of food related industry is the growth of food delivery service. Another notable food related culture is, with the advent of Youtube, the popularity of Mukbang, which refers to content that records eating. Based on these background, this study intended to focus on two things. First, we tried to see the impact of Youtube Mukbang and the sentiments of Mukbang comments on the number of related food deliveries. Next, we tried to set up the predictive modeling of chicken delivery order with machine learning method. We used Youtube Mukbang comments data as well as weather related data as main independent variables. The dependent variable used in this study is the number of delivery order of fried chicken. The period of data used in this study is from June 3, 2015 to September 30, 2019, and a total of 1,580 data were used. For the predictive modeling, we used machine learning methods such as linear regression, ridge, lasso, random forest, and gradient boost. We found that the sentiment of Youtube Mukbang and comments have impacts on the number of delivery orders. The prediction model with Mukban data we set up in this study had better performances than the existing models without Mukbang data. We also tried to suggest managerial implications to the food delivery service industry.

Key Words : Online Delivery, Youtube Mukbang, Sentiment Analysis, Machine learning, Predictive model

Received : September 4, 2022 Revised : October 19, 2022 Accepted : November 3, 2022

Corresponding Author : Sang-Yong Tom Lee

* School of MIS, Hanyang University

** Corresponding author: Sang-Yong Tom Lee

School of MIS, Hanyang University

220 Wangsimni-ro Seongdong-gu Seoul, Korea

Tel: +82-2-2220-1814, E-mail: tomlee@hanyang.ac.kr

저 자 소 개



최사라

현재 한양대학교 일반대학원 경영학과 경영정보시스템 박사과정에 재학중이다. 주요 관심 분야는 소셜미디어, 이미지분석, 빅데이터애널리틱스 등이다. 딥러닝, 머신러닝, 데이터마이닝 기법 등을 통하여 정보기술의 영향을 측정하는 연구를 진행하고 있다.



이상용

현재 한양대학교 경영대학 교수로 재직중이다. 주요 관심 분야는 정보경제, 개인정보보호(privacy) 및 보안, 소셜미디어, 빅데이터애널리틱스 등이다. 관련 연구들을 MIS Quarterly, Management Science, Journal of Management Information Systems를 비롯한 다수의 저널에 관련 논문을 게재하고 있다.