

http://dx.doi.org/10.17703/JCCT.2022.8.5.507

JCCT 2022-9-63

머신러닝을 활용한 VOD 이용건수 예측

Machine Learning Approach for Prediction of VOD Usage

전종석*, 장하은**, 오주희***

Jeon Jong Seok*, Jang Ha Eun**, Oh Joo Hee***

요약 본 연구는 영화 산업에서 온라인 시장인 IPTV의 VOD 이용 건수 예측 모델을 개발하였다. 한국영화진흥위원회에서 수집한 2017년부터 2021년까지 VOD 이용건수 데이터를 활용하여 머신러닝 기반 예측모델을 구축했다. 문헌 조사와 군집분석을 통하여 오프라인 시장과 온라인 시장의 차이를 밝히고, VOD 이용 건수의 새로운 범주를 제안한다. 머신러닝 기반의 VOD 이용 건수 예측 모델 개발을 통해 IPTV 기업들의 의사결정 지원 뿐 아니라 마케팅 전략 수립을 돕는 것을 목적으로 한다.

주요어 : IPTV, 온라인 VOD, 한국영화진흥위원회, 예측모델, 머신러닝

Abstract This study developed a model for predicting the number of VOD uses of IPTV, an online market in the film industry. A machine learning-based prediction model was established using the VOD usage data collected by the Korean Film Council from 2017 to 2021. Through literature research and cluster analysis, the difference between the offline market and the online market is revealed, and a new category of VOD usage is proposed. The purpose is to help IPTV companies establish marketing strategies as well as support decision-making by developing a machine learning-based VOD usage prediction model.

Key words : IPTV, Online VOD, Korea Film Council, Predictive Analytics, Machine Learning

1. 서론

한국 영화 산업의 시장구조는 오프라인 시장인 극장과 TV 방영권, DVD/블루레이, TV VOD, 인터넷 VOD를 포함하는 온라인 시장, 그리고 완성작 수출 또는 서비스 수출을 의미하는 해외시장 부문으로 구분한다. 2008년에 등장하여 꾸준히 성장해 오던 온라인 시장은 2021년 영화 산업 1위를 달리던 극장 부문 수입을 넘어서 한국 영화 산업에서 수익 창출 기여도 시장 1위를

기록했다[1]. 코로나19 시대의 갑작스러운 도래로 국내 온라인 기반 문화예술이 급격히 확장되면서[2] 미디어 시장이 빠르게 발전된 것으로 보인다.

영화 콘텐츠는 소비자가 직접 경험하기 이전에 선호도 및 품질을 판별하기 어려운 경험재로서, 고수익, 고위험의 특징을 지니고 있으므로, 가능한 초기 단계에서 영화 흥행을 정확히 예측하는 것은 영화산업 관련 업계 종사자들에게 매우 중요한 과제이다[3]. IPTV, OTT 등의 확산으로 대중들의 영화 시청 방법이 달라지고 있고, 영화산업의 수익모델도 극장 일변도를 벗어나 VOD

*준회원, 한동대학교 ICT창업학부 학부생 (제1저자)
**준회원, 한동대학교 커뮤니케이션학부 학부생 (참여저자)
***정회원, 한동대학교 경영경제학부 조교수 (교신저자)
접수일: 2022년 7월 11일, 수정완료일: 2022년 9월 1일
게재확정일: 2022년 9월 9일

Received: July 11, 2022 / Revised: September 1, 2022

Accepted: September 9, 2022

***Corresponding Author: jooheeh@handong.edu

Dept. of Management and Economics, Handong Global University, Korea

등의 창구로 다변화되어가고 있다[4]. 극장을 기반으로 영화 제조 및 생산에 집중했다면, 이제는 영화의 유통, 권리 보호 및 소비 향유에 집중하는 전략이 필요해졌다 [5].

본 연구는 온라인 시장 중 IPTV를 대상으로, 2017년부터 2021년까지 국내 IPTV 3사에 등록된 영화 데이터를 바탕으로 머신러닝을 활용하여 VOD 이용건 수 예측 모델을 구성하고자 한다. 본 연구의 예측 모델을 통해, 온라인 시장에서 IPTV 기업들의 영화 관련 계약 및 마케팅 전략 수립 및 개선에 함의점을 제공한다.

II. 선행연구

1. 영화 온라인 시장 성장

코로나19는 지난 2년 동안 한국의 극장 산업을 20년 전의 규모로 축소하였다. 반면, 온라인 플랫폼에서 상영한 작품의 세계적 성과는 지속해서 이어지고 있다[1].

온라인 시장의 다음과 같은 특징들은 변화를 더욱 빠르게 확산시키고 있다. 첫째, 콘텐츠 소비의 유연성을 크게 높였다. 사용자들은 원하는 시간과 장소에서 TV, PC 또는 스마트폰을 포함한 다양한 기기를 통해 콘텐츠를 시청할 수 있다. 이는 콘텐츠의 시간, 장소 및 기기 이동이 모두 가능함을 의미한다[6]. 둘째, 수십 개의 실시간 채널에 비해 VOD의 폭넓은 콘텐츠 라이브러리를 제공함으로써 시청자들이 원하는 콘텐츠를 증가시켰다 [7]. 셋째, 상호 작용할 수 있음에 따라 검색, 브라우징, 일시 정지, 빨리 감기, 되감기 등의 편리한 기능을 추가하여 사용자 편의성을 높였다[8]. 마지막으로 시청자가 능동적으로 선택한 콘텐츠에 Push 방식으로 광고가 제공되는 형태가 가능해지면서, 채널 변경을 통한 Zapping이 불가능하여 광고의 주목률이 높아짐으로[9] 더욱 효과적으로 광고를 배치할 수 있게 되었다.

2. IPTV 시장

IPTV(Interactive Network Television) 서비스는 온라인 플랫폼 중에 한 분야로, 광대역 인터넷의 급속한 발전과 함께 부상하고 있는 새로운 인터넷 부가가치 서비스이다[10]. IPTV의 콘텐츠들은 중소 제작사 및 수입사(CP)로부터 전문 유통사(MCP)가 받아온 콘텐츠의 판권을 IPTV가 계약하는 구조로 이루어져 있으며, 수익의 경우 총매출을 배분하는 방식 RS(Revenue Share

(수익배분)) 혹은, CP에게 콘텐츠 매출액의 일정 금액을 보장해 주는 형태의 계약인 MG(Minimum Guarantee (최소보장금)) 방식이 주로 이루어진다[11].

한국에서는 디즈니의 거울왕국이 국내 최대 IPTV 플랫폼인 KT 올레 TV에서 하루 동안 6억 원을 달성했다[12]. 또한 한국 영화'관상', '광해, 왕이 된 남자', 그리고 '7번 방의 선물'은 100만 VOD 다운로드에 도달했는데, 이는 극장 시장의 천만 관객과 비슷한 규모이다 [13]. 이러한 성장에 따라 VOD 중심 배급 전략, IPTV 우선 개봉 등 영화 제작사와 배급사의 채널 전략 또한 변화하고 있다[14].

유료 방송 플랫폼의 온라인 영화 매출 비율은 2021년 기준 전체 극장 외 시장 매출의 64.6% 차지하고 있고[1], 이중 IPTV 3사가 54.1%를 차지하고 있다[15]. 방송산업 관련 최신 통계인 「2021년 방송산업 실태조사 보고서」에 따르면, 2020년 말 기준 IPTV 가입자는 1854만 가구로 2019년 1713만 가구보다 141만 가구 (8.2%) 증가하였다[16]. 이용자가 감소 추세인 다른 유료 방송 들과는 달리, IPTV 이용자 수가 증가할 수 있었던 이유는 이동통신 3사가 국내 유료 방송 시장을 주도하는 상황에서 유료 방송 상품이 인터넷과 모바일 결합 상품으로 묶여 있다 보니 해지가 쉽지 않은 가입구조이며[17], 국내 IPTV 3사의 적극적인 OTT 따라잡기 전략(태블릿 IPTV 등)과 오리지널 콘텐츠 투자 강화 [18]가 있기에 가능한 현상으로 보인다.

따라서 본 연구에서는 다음과 같은 이유로 IPTV에 대한 연구를 진행하고자 한다.

첫째, 여전히 한국 영화 산업의 극장 외 시장은 케이블TV, 위성방송, IPTV 등 VOD 서비스 목적으로 판매되는TV VOD 시장의 비중이 64.6%로 가장 크고, 웹하드 플랫폼, OTT 플랫폼에 VOD 서비스 목적으로 판매되는 인터넷 VOD 시장은 27.8%로, TV VOD 시장과 2배 이상 차이가 난다[1]. 둘째, VOD 매출은 2020년부터 3사 유료 방송플랫폼이 역성장하고 있어서 이용자들이 글로벌OTT 2개, 국내 OTT 1개의 OTT에 가입하는 단계까지는 국내 유료 방송이OTT보다 경쟁력이 있다 [19]. 셋째, LG 유플러스와 디즈니 플러스, SK브로드밴드와 애플TV플러스, KT와 넷플릭스 등 한국의 IPTV 사업자들은 세계적으로 영향력 있는 글로벌 OTT 서비스의 제휴사로 역할하고 있다[1].

3. 영화 산업에서 머신러닝 활용

최근 문화예술 콘텐츠 전반에 빅데이터의 활용은 다양하게 적용되고 있고, 영화산업에서는 영화의 인지도와 호감도 및 버즈량 분석에 기반한 소셜 데이터와 연동하여 영화 마케팅 및 흥행 예측에 크게 활용되고 있다[20].

Song과 Han(2013)은 선형모형, random forests model, gradient boosting model을 예측모형으로 사용하여 흥행 수익을 예측하기 위한 가장 중요한 변수는 감독의 평점이란 것을 증명했다[21]. Chang(2017)은, 머신러닝 기법을 이용하여 영화의 흥행성적을 예측하는 기법을 제안하였고 배우, 감독과 같은 정적 데이터와 영화평 평점, 뉴스 기사 수, 블로그 수 등 동적 데이터를 조합한 모델이 가장 좋은 성능을 보였다는 점을 보였다[22]. Jeon과 Son(2016)은 네이버 및 다음 포털의 평점과 평가자 수, 블로그 수, 뉴스 수와 같은 OWOM 변수들을 활용하여 다중회귀모형 혹은 MLP 신경망 모형에 의해 최종 OWOM 변수들이 예측 정확도 개선에 도움이 된다는 점을 밝혀냈다[23].

III. 연구방법

1. 데이터 수집

본 연구는 온라인 VOD 이용건 수를 수집하기 위해 영화진흥위원회 산업 통계에서 2017년부터 2021년까지의 “IPTV 및 디지털 케이블 TV 영화 VOD 이용순위 top 200” 엑셀 데이터를 활용했다. 영화진흥위원회는 영화정보 시스템을 운영함으로써 국내외 영화산업 지식정보와 신뢰성 있는 영화 관련 통계자료를 제공한다. 2017년부터 2021년까지 이용 중 계절과 관련된 사용자의 이용형태를 파악하기 위해 1년을 사계절로 나누어 수집하였다. 해당 데이터에는 온라인 VOD 이용건수 뿐만 아니라, FIMS(영화정보 표준 식별코드), 영화제목, 출시 년도, 제작국가, 극장 개봉일, 감독, 배우, 관람 연령 정보가 포함되었다. 각 영화의 세부적인 특징 및 정보를 파악하기 위해 한국 영화진흥위원회의 KOBIS API를 사용하여 영화 상세정보를 완성하였다. 위에서 수집한 영화진흥위원회의 VOD의 FIMS 코드를 바탕으로 영화의 상세정보(상영시간, 제작상태 명, 장르, 배우, 상영 형태 명, 참여 영화사, 스태프)를 수집하였다. 마지막으로, 오프라인 박스오피스 시장이 온라인 VOD 이용건

수에 미치는 영향을 확인하기 위해 KOBIS 박스오피스 통계에서 2004년부터 2021년까지 개봉한 영화들의 박스오피스 정보를 수집했다. 해당 데이터에는 개봉일, 수익, 박스오피스, 스크린 수, 상영 횟수, 국적, 배급사 정보가 포함되었다. KOBIS 박스오피스 정보가 2004년부터 존재했기 때문에, 온라인 VOD 이용건 수 데이터 중 개봉 연도가 2004년 전인 영화는 제외한 8,960개의 데이터를 활용했다.

2. 데이터 전처리

데이터 수집 단계에서 수집한 28개의 변수의 데이터 항목 중 수치형 데이터를 제외한 범주형 변수들에 대해 다음과 같은 전처리를 수행하였다. 우선, “장르”와 “상영 형태” 변수는 하나의 영화가 다중 장르에 속한 것으로 분류하고 있다. 따라서, 수집한 “장르”와 “상영 형태” 변수를 단일 범주형 변수로 구성하지 않고, 더미화(dummy)하여 예측 모형의 입력변수로 사용하였다. “감독”, “배우”, “제작사”, “배급사” 변수는 다중 범주를 갖고 있을 뿐만 아니라, 범주가 매우 다양하여 더미화하였을 때 유의미한 변수로 사용될 가능성이 적다고 판단했다. 그러므로 “감독 역량”, “배우 역량”, “제작사 역량”, “배급사 역량” 변수라는 변수들을 별도로 생성했다. 해당 변수들은 각각 데이터를 수집한 기간인 2017년부터 2021년까지 영화의 감독, 출연 배우, 제작사, 배급사가 이용 건수 데이터에 포함된 횟수를 수치화한 변수이다.

3. 종속변수 범주화

본 연구는 영화 흥행 예측을 분류 문제로 정의한다. 영화의 흥행 정도는 영화를 시청한 관객의 수인 연속값으로 제공된다. 제한된 수의 범주로 연속형 변수를 변환하는 이 과정을 일반적으로 이산화라고 한다. 이산값은 연속된 범주에서 제한된 수의 간격인 반면, 연속값은 무한히 많이 있을 수 있다. 일반적으로 이산값이 지식 수준 표현에 가깝고 이산화를 통해 데이터를 축소하고 단순화하여 많은 학습 알고리즘을 더 빠르고 정확하게 만든다. 또한 이산형 변수는 연속형 변수보다 이해하기 쉽기 때문에 예측 모델을 개발할 때 이산값을 사용하는 것을 선호한다[24,25,26]. 한국 영화에 대한 흥행을 예측하는 연구는 대부분 관객 수에 대해 표 1과 같은 범주를 적용한다[27,28,29].

표 1. 박스오피스 범주

Table 1. Box-office Category

section	the number of spector
1	less than 500,000
2	500,000 ~ 1,000,000
3	1,000,000 ~ 3,000,000
4	3,000,000 ~ 5,000,000
5	more than 5,000,000

하지만 VOD 시장의 이용건 수에 대해서는 아직 구분되어진 범주가 많이 존재하지 않는다. 그래서 해당 연구에서는 VOD 시장 이용건수를 데이터 클러스터링 기법 중 하나인 Jenks natural breaks을 활용하여 범주화했다. 이 알고리즘은 Jenks에 의해 1967년 제안되었고, 각 범주에서 다른 범주의 평균에 대한 편차를 최대화하는 동시에 각 범주의 범주 평균에 대한 평균 편차를 최소화하는 알고리즘이다[30,31]. 해당 알고리즘을 통해 VOD 이용 건수에 대한 범주를 구분 지었는데, 표 2와 같은 범주가 생성되었다.

표 2. VOD 이용건수 범주

Table 2. VOD Usage Category

section	thenumber of spector
1	less than 47,737
2	47,737 ~ 166,309
3	166,309 ~ 423,566
4	423,566 ~ 936,060
5	more than 936,060

범주가 분류된 VOD 이용 건수에 대해, 오프라인 박스오피스의 범주에 대해 정오 행렬 형태로 구성 후, 카이 제곱 검정을 통해, 표 3과 같이 비교를 해보았다. 카이제곱 검정은 관심 있는 범주의 빈도가 기대되는 범주의 빈도와 의미 있게 다른지의 여부를 검정하기 위해 사용되는 검정 방법이다. 정오 행렬을 바탕으로 한 두 집단의 범주에 대한 일치율은 31.66%였으며, 카이제곱(chi-squared test) 검정 결과 p-value가 2.2e-16보다

표 3. VOD Boxoffice 범주 비교

Table 3. Comparing VOD and Boxoffice Categories

vod boxoffice	1	2	3	4	5
1	2319	300	618	1335	327
2	497	180	277	2192	125
3	92	114	253	44	60
4	0	2	71	71	69
5	0	0	0	0	14

작으므로, 귀무가설이 기각되었고, 오프라인 박스오피스 관객 수와 VOD 이용 건수는 동일하지 않다는 대립가설을 채택하게 되었다.

종속변수의 범주까지 정하고, 모델링에 사용된 변수들은 표 4와 같다.

표 4. 변수 목록

Table 4. List of Variables

변수명	설명
vod	vod 이용 건수 (종속변수)
boxoffice	오프라인 박스오피스
screen	상영관 수
play	상영횟수
genre	영화 장르
season	영화 시청 계절
grade	영화 관람 연령
showTm	러닝타임
dist score	배급사 점수
artor score	배우 점수
director score	감독 점수
company score	제작사 점수
Countries	제작 국가
Showtypes	상영 형태

IV. 실험 및 결과

1. VOD 이용건수 예측 모형

예측을 위한 머신러닝 기법으로 분류를 위한 알고리즘인 종속변수가 3개 이상의 범주를 가질 때, 적용할 수 있는 회귀 모형인 다항 로지스틱 회귀(Multinomial Logistic Regression)와 앙상블 기반의 트리 모형 중 배깅을 활용한 랜덤 포레스트(Random Forest), 부스팅을 활용한 엑스뷰부스트(XGBoost), 비확률적 선형 분류 모델인 서포트 벡터 머신(Support Vector Machine)을 활용하였다. 일반적인 로지스틱 회귀 모형은 종속 범주형 변수에 두 개의 범주가 있을 때 사용되는 반면, multinomial logistic regression은 클래스가 두 개 이상 있는 다중 분류 문제에 사용되는 알고리즘이다. 데이터 집합에 있는 변수 간의 관계를 이해하는 데 도움이 되며, 로지스틱 회귀 모형을 별도로 적합 시킬 때보다 모수 추정치에 대한 표준 오차가 적다[32]. Random Forest 알고리즘은 데이터의 하위 집합에서 생성되어 최종 출력을 평균 또는 다수 순위를 기반으로 하는 앙상블 기법 중 Bagging이 적용된 트리 기반의 모형으로, 트리 기반 모델에서 발생할 수 있는 과적합 문제를 방지한다[33]. XGBoost는 확장 가능한 Distributed Gradient

Boosting Machine Learning 라이브러리이다. 병렬 트리 부스팅을 제공하며, 회귀, 분류 및 순위 문제를 위한 선도적인 머신러닝 학습 라이브러리이다. 약한 예측 모형들의 오차에 가중치를 두고 순차적으로 다음 학습 모델에 그 가중치를 반영하여 강한 예측 모형을 만든다 [34]. Support Vector Machine은 분류와 회귀 모두에 사용되는 지도 학습 알고리즘이다. 데이터의 지점을 명확하게 분류하는 n차원 공간에서 초평면을 찾는 목적을 갖는다. 데이터가 고차원일 때 노이즈에 대한 영향이 적은 장점이 있는 알고리즘이다[35].

본 연구는 네 가지 머신러닝 모형을 이용하여 학습과 검증 실험을 수행하고, 각 모형마다의 변수 중요도를 파악하고, 예측 성능을 비교 평가하였다. 수집한 데이터 중 2017년부터 2019년까지의 5,819개의 데이터를 학습 데이터로 활용했으며, 2020년부터 2021년까지의 3,141개의 데이터를 검증 데이터로 분류했다. 데이터는 머신러닝 모형의 예측 성능에 대한 평가를 위해 분류 기법의 성능 평가에 활용되는 정밀도(precision)과 재현율(recall)이 함께 고려된 F1-score를 기준으로 그리드 서치(grid-search)를 통해 최적의 하이퍼파라미터 값을 선정하였다. 또한, 각각의 모형이 제공하는 변수 중요도를 토대로 최적의 변수를 선정하였다.

2. 모형 성능 평가

각 모형의 예측 결과에 대해, 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score 4개의 평가 척도를 통해 성능 평가를 진행했다. 정확도는 전체 데이터 수에 걸쳐 올바르게 분류된 데이터 수를 나타낸다. 정확도는 데이터가 균형을 이룰 때 좋은 평가 지표이지만, 데이터가 불균형할 때는 좋은 평가 지표가 아니다. 정밀도는 예측을 positive로 한 대상 중에 예측값과 실제값이 positive로 일치한 데이터의 비율을 뜻한다. 재현율은 실제값이 positive인 대상 중 예측값과 실제값이 positive로 일치한 데이터의 비율을 뜻한다. 정밀도와 재현율은 상호 보완적인 평가 지표이기 때문에 두 지표에는 트레이드오프(trade-off)가 발생한다. F1-score는 정밀도와 재현율의 조화평균으로, 정밀도와 재현율이 비슷할수록 높아진다. 모형별 최적 하이퍼파라미터 탐색과 변수 선택을 거쳐 최종 구축된 4개 모형을 통해 학습 데이터와 평가 데이터의 영화들에 대한 VOD 이용건 수를 예측한 결과는 표 5와 같다.

표 5. 모델 성능 평가

Table 5. Model Performance Assessment

	Data	Accuracy	Recall	Precision	F1-Score
LR	Train	0.631	0.698	0.860	0.771
	Test	0.502	0.551	0.846	0.667
RF	Train	0.773	0.833	0.847	0.840
	Test	0.748	0.789	0.823	0.806
XGB	Train	0.643	0.717	0.838	0.773
	Test	0.772	0.938	0.693	0.799
SVM	Train	0.666	0.687	0.876	0.770
	Test	0.668	0.844	0.540	0.659

Recall과 Precision을 반영한 결과값인 F1 Score의 측면에서 RandomForest가 가장 높은 성능을 보였다. Accuracy의 경우 XGBoost가 가장 높은 성능을 보였다.

V. 결론

본 연구는 영화산업에서 온라인 VOD 이용건 수를 예측하는 머신러닝 기반의 예측 모형을 구축하였다. 이를 위하여 먼저 최근 주목할 수 있는 영화시장의 수익 구조 변화에 초점을 맞추어, VOD 이용건수와 오프라인 박스오피스와의 차이점을 문헌 조사와 군집분석을 통해 비교하였다. 이를 기반으로, 온라인 VOD 수요예측 모델로 활용할 수 있는 종속변수의 새로운 범주를 생성하였다. 영화진흥위원회에서 2017년부터 2021년 사이 VOD 이용건 수 데이터를 수집하여 8,960개의 영화 데이터에 대해 2017~2019년 VOD에서 이용된 영화를 훈련 데이터로, 2020~2021년 VOD에서 이용된 영화를 평가 데이터로 설정하여, 4개의 예측 모형을 구축하여, 향후 VOD 이용 건수를 예측하였다. 기존의 오프라인 극장에서 새롭게 확산되고 있는 온라인 시장에서 IPTV 기업들의 의사결정과 마케팅 전략 수립에 도움을 줄 수 있는 머신러닝 기법의 예측 모델을 제안한다.

References

- [1] KOFIC, 2021 Korea Film Industry Settlement Report
- [2] Lee, S. H. "A Study on the Cultural and Creative Industry in the COVID-19 Era.", The Journal of the Convergence on Culture Technology (JCCT), 6(4), 567-573, 2020. DOI : 10.17703/JCCT.2020.6.4. 567

- [3] Jeon, S., & Son, Y. S. "Prediction of box office using data mining", *The Korean Journal of Applied Statistics*, 29(7), 1257-1270, 2016. DOI : 10.5351/KJAS.2016.29.7.1257
- [4] Choi, Y. Predicting VOD Success of Motion Pictures in Pay-TV, 2018. (Doctoral dissertation, Seoul Graduate School).
- [5] Lee, G. W., & Bang, M. Y. "A study on Innovation Methods for the Content Industry in the post-OVID-19 Era", *International Journal of Advanced Culture Technology*, 9(3), 142-151, 2021. DOI : 10.17703/IJACT.2021.9.3.142
- [6] Bury, R., & Li, J. "Is it live or is it timeshifted, streamed or downloaded? Watching television in the era of multiple screens", *New media & society*, 17(4), 592-610, 2015. DOI : 10.1177/1461444813508368
- [7] Lotz, A. D. "What is US television now?", *The annals of the American academy of political and social science*, 625(1), 49-59, 2009. DOI : 0.1177/002716209338366
- [8] Subramanya, S.R., & Yi, B. K. "Utility model for on-demand digital content", *IEEE Computer*, 38(6), 95&98, 2005. DOI : 10.1109/mc.2005.206
- [9] Kim, S. H., & Lim, S. H. "Study on Status of VOD Viewer Ratings and Improvement of Measurement", *Advertising Studies*, 26(2), 139-61, 2015. DOI : 10.14377/kja.2015.2.28.139
- [10] Yuan, Y., & Wang, C. "IPTV video quality assessment model based on neural network", *Journal of Visual Communication and Image Representation*, 64, 102629, 2019. DOI : 10.1016/j.jvcir.2019.102629
- [11] Kim, H. M. *Understanding digital video platforms*. Communication Books, 2015.
- [12] Yoo, J.H., 2014, May 30. The power of home video VOD... Frozen scores 600 million household sales in a day, *The Korea Economic Daily*. Retrieved from, <http://www.hankyung.com/news/app/newsview.php?aid=2014052999831>. Accessed March 30, 2015, March 30. (In Korean)
- [13] NasMedia Digital Broadcasting Report, 2014. Increase of film VOD on TV platforms and cases of IPTV film advertisement. Retrieved from, <http://www.ad.co.kr/lit/report/show.do?ukey=90034&oid=@392014%7C8%7C1> (In Korean)
- [14] Kim, E., & Kim, S. "Online movie success in sequential markets: Determinants of video-on-demand film success in Korea", *Telematics and Informatics*, 34(7), 987-995, 2017. DOI : 10.1016/j.tele.2017.04.009
- [15] Ministry of Science and ICT (2021. 11. 10.). Number of pay-TV subscribers and market share in the first half of 2021
- [16] [Broadcasting Industry Survey Report for 2021], Korea Communications Commission Broadcasting Statistics Portal, 2021, p. 17
- [17] *The Economist*, 2021.11.14., "OTT is growing... Why doesn't pay TV codecutting happen in Korea?" <https://economist.co.kr/2021/11/14/it/general/20211114120605187>
- [18] *Asian Economy*, 2021.9.6., I don't watch pay TV. Concerned about CodeNever... IPTV that follows OTT, <https://www.asiae.co.kr/article/2021090520493049460>
- [19] Remarks from a discussion at the Digital Media Law Scheme Forum, 2021.5.6.
- [20] Kim, J. W. "A study on the use of Big Data in Film Industry - Focused on Netflix Analytical Tools", *The Korean Journal of Arts Studies*, (25), 51-64, 2019. DOI : 10.20976/kjas.2019..25.003
- [21] Song, J. W. & S. J. Han, "Predicting gross box office revenue for domestic films", *Communications for Statistical Applications and Methods*, Vol.20, 301-309, 2013. DOI : 10.5351/csam.2013.20.4.301
- [22] Chang, J. Y., "An Experimental Evaluation of Box office Revenue Prediction through Social Big data Analysis and Machine Learning", *The journal of the institute of Internet, broadcasting and communication*, Vol.17, No.3, 167-173, 2017. DOI : 10.7236/JIIBC.2017.17.3.167
- [23] Jeon, S. H., and Y, S, Son, "Prediction of boxoffice using data mining", *The Korea Journal of Applied Statistics*, Vol.29, No.7, 1257-1270, 2016. DOI : 10.5351/KJAS.2016.29.7.1257
- [24] Simon, H. A. 1969. *The sciences of the artificial intelligence*, 1981. MIT Press. [aHH], 1990. *Invariants of human behavior*. *Annual Review of Psychology*, 41, 119.
- [25] Liu, H., Hussain, F., Tan, C. L., & Dash, M. "Discretization: An Enabling Technique", *Data Mining and Knowledge Discovery*. 6: 4, 393-423, 2002.
- [26] Dougherty, J., Kohavi, R., & Sahami, M. "Supervised and unsupervised discretization of continuous features", In *Machine learning proceedings*, p. 194-202, 1995. Morgan Kaufmann. DOI : 10.1016/b978-1-55860-377-6.50032-3
- [27] Yim, J., & Hwang, B. Y. "Predicting movie success based on machine learning using twitter", *KIPS transactions on Software and Data Engineering*, 3(7), 263-270, 2014. DOI : 10.3745/ksde.2014.3.7.263

- [28]Kim, Y. H., & Hong, J. H. "A study for the development of motion picture box-office prediction model. Communications for statistical applications and methods", 18(6), 859-869, 2011. DOI : 10.5351/ckss.2011.18.6.859
- [29]Lee, G. J., & Jang, U. J. Predicting Financial Success of a Movie Using Bayesian Choice Model. In Proceedings of the Korean Operations and Management Science Society Conference(pp. 1851-1856). The Korean Operations Research and Management Science Society, 2006.
- [30]Jenks, G. F. The data model concept in statistical mapping. International yearbook of cartography, 7, 186-190, 1967.
- [31]McMaster, R. "In memoriam: George f. jenks (1916-1996). Cartography and Geographic Information Systems", 24(1), 56-59, 1997. DOI : 10.1559/152304097782438764
- [32]Bayaga, Anass. "Multinomial Logistic Regression: Usage and Application in Risk Analysis", Journal of applied quantitative methods 5(2), 2010.
- [33]Biau, Grard, and Erwan Scornet. "A random forest guided tour", Test 25(2) 197-227, 2016.
- [34]Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining. 2016. DOI : 10.1145/2939672.2939785
- [35]Noble, William S. "What is a support vector machine?", Nature biotechnology, 24(12), 1565-1567, 2006.

※ 본 연구는 2022년도 한동대학교 ICT기업가
정신센터 대학혁신지원사업의 지원을 받아
수행하였음.