

머신러닝을 활용한 TV 오디션 프로그램의 우승자 예측 모형 개발: 프로듀스X 101 프로그램을 중심으로¹

Development of a Model for Winner Prediction in TV Audition Program Using Machine Learning Method: Focusing on <ProduceX 101> Program

곽 주 영 (Juyoung Gwak) 전남대학교 일반대학원 경영학과²

윤 현 식 (Hyun Shik Yoon) 전남대학교 경영대학 경영학부³

ABSTRACT

In the entertainment industry which has great uncertainty, it is essential to predict public preference first. Thanks to various mass media channels such as cable TV and internet-based streaming services, the reality audition program has been getting big attention every day and it is being used as a new window to new entertainers' debut. This phenomenon means that it is changing from a closed selection process to an open selection process, which delegates selection rights to the public. This is characterized by the popularity of the public being reflected in the selection process. Therefore, this study aims to implement a machine learning model which predicts the winner of <ProduceX 101>, which has recently been popular in South Korea. By doing so, this study is to extend the research method in the cultural industry and to suggest practical implications. We collected the data of winners from the 1st, 2nd, and 3rd seasons of the Produce 101 and implemented the predictive model through the machine learning method with the accumulated data. We tried to develop the best predictive model that can predict winners of <ProduceX 101> by using four machine learning methods such as Random Forest, Decision Tree, Support Vector Machine (SVM), and Neural Network. This study found that the audience voting and the amount of internet news articles on each participant were the main variables for predicting the winner and extended the discussion by analyzing the precision of prediction.

Keywords: Machine learning, Supervised learning, Random forest, TV audition, Winner prediction, Entertainment industry

1) 논문접수일: 2019년 7월 28일; 1차 수정일: 2019년 8월 16일 게재확정일: 2019년 8월 23일

2) 제 1저자(j_g123@naver.com)

3) 교신저자(Dr.Yoon@jnu.ac.kr)

1. 서론

1.1 연구 배경

미래사회에는 대중문화를 바탕으로 한 엔터테인먼트 산업이 더욱 중요한 위치를 차지할 것이다. 경제·사회가 발전할수록 엔터테인먼트에 관한 사람들의 욕구가 상승하기 때문이다. 뿐만 아니라 온라인, 디지털화로 콘텐츠의 유통은 국경이 무의미해지면서 국가 경쟁력에도 큰 영향을 끼치고 있다(유진룡·최화열 2010). 엔터테인먼트 산업은 오락, 여흥 등을 대중에게 제공하는 것을 목적으로 하는 상품이나 서비스를 의미한다(장규수 2013). 최근에는 모바일과 결합된 미디어 환경이 예측하지 못했던 범위의 새로운 비즈니스 모델을 만들어 내고 있기도 하다. 일례로 1인 미디어 환경의 발전을 통한 일반인의 연예인화, 유튜브(YouTube)를 통한 K-Pop의 이례적인 흥행 등을 이야기 할 수 있다. 즉, 기존의 미디어 매체에 단순히 모바일 기술이 더해진 것이 아니라, 그로 인해 산업 구조가 재구성되거나 새로운 영역이 창조되고 있는 것이다(백현미·이새롬 2018).

그러나 기술 발달에 따른 미디어 콘텐츠의 급속한 팽창은 과도한 정보의 노출이라는 부작용을 수반하기도 한다(김주미 등 2018). 이 때, 과잉 정보 속에서 가치 있는 정보 만이 의미를 가지는데, 여기서 희소성과 가치를 지니는 정보는 인간의 관심과 주목을 끄는 것을 뜻한다(나은영·나은경 2019; Sayre 2010). 그렇기에 앞으로 엔터테인먼트 산업에서 더 큰 부가가치를 창출하기 위해서는 수많은 정보 속에서 사람들이 주목하는 희소성을 가지고 있으면서 보다 유의미한 정보를 예측해내고, 이를 효과적으로 활용하는 것이 핵심 과제가 될 것이다. 이미 국내 문화 산업 전반에서 빅데이터를 활용하고자 하는 노력이 진행 중이다. 일례로 NC소프트는 게임 사용자들의 로그기록 데이터를 이용하여 머신러닝을 통해 게임 버그와 사기 탐지 시스템을 구현했으며, 카카오 음원 서비스 플랫폼인 멜론(Melon)은 사용자들의

데이터를 활용하여 이용자 관심도와 아티스트 친밀도 등을 분석하고 이에 따른 콘텐츠를 추천하고 있다(한국정보화진흥원 2015).

1.2 연구문제

상기 한 바와 같이, 산업 전반에 걸쳐 빅데이터를 활용한 가치 창출에 혈안이 되고 있는 현실은 주지의 사실이다. 이에 빅데이터는 여타 연구 방법론과 인간의 직관에 의한 선택의 한계를 극복할 수 있는 대안이 될 수 있을 것인가에 대한 의문이 생길 수 있다.

김기홍(2016)은 빅데이터의 활용이 문화 연구의 통계적 방법론의 한계를 극복할 수 있는 대안으로 충분한 가능성이 있음을 역설하였다. 문화에 대한 근원적 질문들이 포괄적 일 수밖에 없고, 이는 문화를 ‘총체’라는 모호하고 결정론적인 정의로 수렴하는 경향으로 이끈다고 보았다. 그러나 총체성을 포기한 통계적 방법론의 적용은 모호함을 제거하고 명료함을 얻을 수는 있겠으나, 사실상 문화라는 총체적 현상을 설명하는 모든 변수를 찾아내어 계량적으로 측정한다는 것이 불가능하다는 한계가 있다는 것이다. 이로 인해 빅데이터 방법론을 활용한 문화와 기술경영 간 학제적 연구는 앞으로 더 중요해질 것으로 보인다. 이미 문화산업에 빅데이터를 적용하기 위한 향후 과제와 그 전망에 대한 연구가 활발히 수행되고 있다. 윤홍근(2013), 조정환(2012), 김해원과 이미나(2016) 등의 연구를 통해 스포츠, 음악, 미술 등 하위 분야의 빅데이터 관련 연구가 증가하는 추세이며, 문화 산업 전 범위에서 빅데이터의 활용이 중요해지고 있음을 파악할 수 있었다(박지연 등 2019).

이 중, 엔터테인먼트 분야의 경우는 역사가 짧고, 계속해서 시스템이 변화하고 있는 영역이기 때문에 관련 연구가 활발히 진행되지 못했던 측면이 있다(장규수 2013). 그러나 영화 콘텐츠의 흥행을 예측하기 위한 빅데이터 연구는 다수 존재하는데, 장리 등(2017), 장재영(2017), 우종필 등(2018)의 연구를 참고하면, 문화상품

중에서도 상대적으로 영화 콘텐츠가 고위험-고수익-고비용, 불확실성의 특징을 강하게 가지고 있기 때문이라고 추론할 수 있다. 이와 마찬가지로, 엔터테인먼트 산업 내에서 연예인 선발과정 또한 유사한 특징이 있다. 여기에 소비자의 불확실한 기호 또한 예측하기 어려운 변수로 작용하며, 이로 인해 연예인 발굴을 위한 '선별력'은 산업 내에서 중요한 요소이자 문제가 되는 것이다(장규수, 2013). 따라서 본 논문은 빅데이터를 기반으로 한 엔터테인먼트 분야에서의 예측 모형을 구현하여 문화 산업에서의 연구 방법을 확장하고, 새로운 접근 방법으로 실무적인 해결책을 제시해 보고자 한다.

1.3 연구목적

이에 본 연구의 목적을 정리하면 다음과 같다.

첫 번째, 머신러닝 알고리즘을 연구에 적용하여 문화 산업 분야에서의 연구방법의 확장을 가능하게 하는 것이다. 오늘날 다채널화 되는 방송 환경과 온라인 활동의 증가로 본래 지표들의 정확성이 낮아지고 있다. 온라인 스트리밍 기반의 다양한 채널의 증가로 기존의 방법으로는 TV 시청률이나 화제성 척도를 정확히 재단하기에 어려움이 있다. (강일권 등 2017). 연구 문제에서 언급하였던 것처럼, 빅데이터를 통한 분석이 그 보완책이 될 수 있는 바, 이를 통해 문화 산업 분야 연구의 지평을 넓히는데 기여할 수 있을 것이다.

두 번째, 스타에 대한 대중의 선호요인에 대해 분석해보고자 한다. 리얼리티 오디션 프로그램의 등장 이후, 대중이 직접 우승자를 선발하는 시스템이 각광을 받고 있다. 이에 리얼리티 오디션 프로그램이 일반적인 연예인 선발 채널로 자리 잡았으며, 프로그램의 흥행은 연쇄적으로 우승자들을 스타의 반열에 올려놓고 있다. 즉, 흥행에 성공한 프로그램을 거쳐온 연예인 지망생들은 데뷔 전에 대중들의 선호에 의해 선별된 그룹이라는 것이다. 다시 말해, 대중들의 선호를 예측할 수 있다면, 스타 선별 과정에서의 실패 확률을 줄일 수 있다는

의미로 생각해 볼 수 있다. 이에 대중이 우승자를 선발하는 데 있어 어떠한 요인들이 영향을 미치는지 확인해 보고자 한다.

마지막으로 직관과 경험이 개입되어 진행되던 기존의 선별 시스템이 데이터 기반의 예측 모형으로 대체될 수 있을 것인지에 대한 문제를 탐구하고자 한다. 연예인의 자질은 정량적으로 측정할 수가 없었던 만큼 연예인의 선별은 그 동안 관련 분야 종사자들의 경험과 직관에 의존했다(장규수 2013). 여기서 선별력은 엔터테인먼트 산업의 오랜 경험으로 얻어질 수 있는 능력, 즉, 암묵지적 지식으로만 여겨졌기 때문에 뛰어난 인적자원의 확보가 중요했다. 그러나 정확도가 높은 예측 모형이 해당 산업에 적용될 수 있다면, 기존 시스템에서 비용을 절감하고 과정의 공정성을 높이는데 기여할 수 있을 것이다.

이에 본 논문에서는 머신러닝 알고리즘을 활용하여 리얼리티 오디션 프로그램의 불확실한 경쟁 상황에서 대중의 선택 결과를 예측해보고자 한다. 이를 위해, 본 논문에서는 국내 리얼리티 오디션 프로그램의 개념과 발전, 특성 및 흥행요인을 분석하고, 2016년부터 인기리에 방영된 <프로듀스 101 시리즈>의 시즌 1부터 시즌 3까지의 우승자 데이터를 수집하였다. 수집된 누적 데이터를 바탕으로 머신러닝을 통해 예측 모형을 구현하였다. 이를 현재 방영 중인 <프로듀스X 101>의 우승자를 예측해 보고자 한다.

2. 이론 고찰과 사례 선정

2.1 리얼리티 오디션 프로그램

2000년대 이전까지는 연예인 선발을 위해 길거리 캐스팅, 기획사 오디션, 영화·방송사 공채가 주를 이뤘다. 그러나 대중 매체가 발전하면서부터는 길거리 캐스팅은 점차 사라지고, 인터넷 매체와 잡지 등을 통한 발굴,

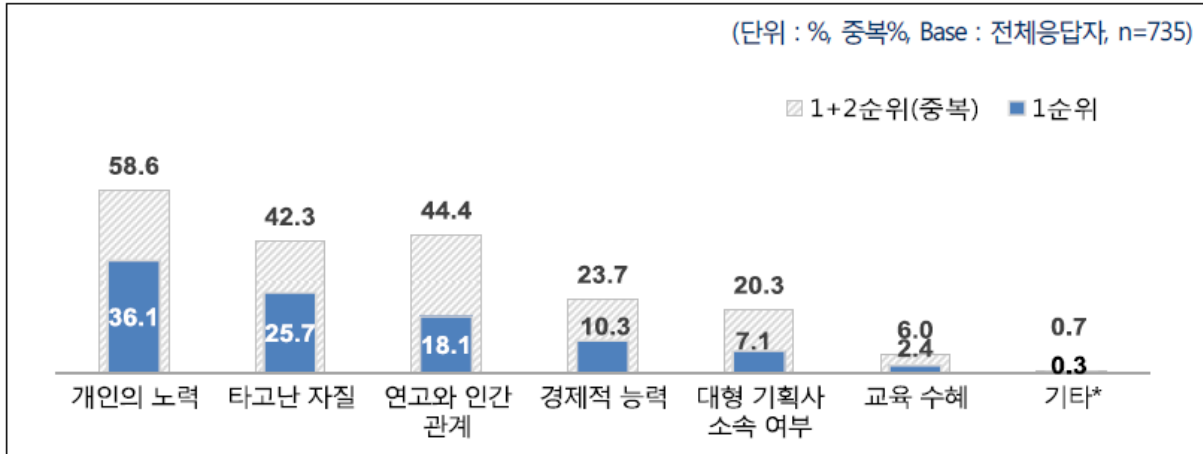
대학교, 연기학원, 에이전시를 통한 양성 등 다양한 방법으로 신인 연예인들이 배출되고 있다(배국남 2016). 그러나 이 과정들 모두 대중에게 폐쇄적이라는 특징이 있다. 오롯이 내부의 전문가, 영화감독, 제작자, 또는 방송사 PD 등의 직관에 의해 선발되는 폐쇄적인 선발 채널 만이 존재하고 있었다. 그러나 이러한 선발 과정이 리얼리티 프로그램과의 결합을 통해 리얼리티 오디션 프로그램이라는 포맷으로 대중들에게 공개되기 시작했다.

오디션 프로그램 열풍은 미디어 기술의 발달 및 세계적인 리얼리티 프로그램의 확산과 관련이 있다(이경숙 2011). 리얼리티 프로그램이란 각본은 없으나 희극, 비극, 미스터리, 액션 등의 극적인 요소들을 이끌어 낼 수 있도록 등장인물과 상황을 전략적으로 선택하는 장르를 뜻한다(나은영·나은경 2019). 리얼리티 프로그램의 가장 큰 특징은 현실과 허구라는 상반된 속성을 모두 지니고 있다는 데 있다. 대본 없이 촬영되고, 배우들이 사전에 정해진 역할을 수행하는 것이 아니라 상황에 따라 자연스럽게 행동하는 모습을 담아낸다는 점에서 현실적이다. 그러나 예측 가능한 방향의 모습을 끌어내기 위해 때로는 특별한 상황들을 설정하여 오락의 기능을 하도록 의도된다는 점에서는 허구적인 측면도 있다(Hall 2006; Nabi 2003). 초창기 리얼리티 프로그램은 범죄를 다루는 수사와 같은 다큐멘터리 취재 형식으로 발달하였는데, 점차 경쟁이 치열해지는 미디어 환경에서 시청자들을 끌어들이기 위해 새로운 시장과의 결합을 도모하는 방향으로 발전하고 있다(Baltruschat 2009). 즉, 현재 리얼리티 오디션으로 분류되는 포맷은 리얼리티와 경연 형식을 결합하여 오락성을 부각 시킨 프로그램으로 정의할 수 있다(최영준 2015). 해외의 경우, 일찍이 <아메리칸 아이돌(American Idol)>과 <브리튼스 갓 탤런트(Britain's Got Talent)>가 세계적인 성공을 거둔 바 있고, 한국에서도 이와 같은 포맷이 적극적으로 수용되었다(최문

경 등 2013).

국내의 리얼리티 프로그램은 <슈퍼스타-K>, <K-pop Star>, < 짹>, <복면가왕>, <위대한 탄생>, <아빠 어디가>, <슈퍼맨이 돌아왔다>, <나혼자 산다> 등이 있으며, 이들은 리얼리티 프로그램을 기반으로 육아, 오디션, 로맨스 등 다양한 분야와 결합하며 새로운 변용이 이루어졌음을 알 수 있다. 특히, 국내 리얼리티 오디션이 외국과 크게 다른 점은 '감동 서사'의 재현이다. 냉정한 경쟁보다 참가자의 개인적인 사연과 성장 과정, 감동적인 스토리가 강조되었다는 것이다(최문경 등 2013). 최근에는 <미스 트롯>, <프로듀스 101>, <더 유닛> 등과 같이 일반인보다는 기존의 연습생, 무명 연예인들을 전면에 내세워 꿈에 대한 열정과 갈망을 한층 강한 서사로 풀어내며, 시청자들의 감정에 호소하는 모습을 볼 수 있다. 2010년대에 들어서는 리얼리티 오디션 프로그램의 약진이 눈에 띈다. 최근 '2019년 7월 둘째 주 비드라마 부문 TV화제성 순위'를 살펴보면, 10위 안에 든 주요 예능 프로그램의 50% 이상이 리얼리티 프로그램이라는 것을 확인할 수 있다. 더 주목할 만한 점은 1위에 이름을 올린 오디션 프로그램 <프로듀스X 101>이 점유율 27.69%를 차지하고 있다는 점이다(굿데이터코퍼레이션 2019). 이 외의 프로그램들이 5%를 상회하거나 그 미만이라는 점을 감안한다면, 매우 높은 인기를 구가하고 있음을 알 수 있다.

삼성경제연구소(2011)는 오디션 프로그램의 성공 요인 3가지를 주제·소재의 공감성, 경연 방식의 공정성, 경연 참가자들의 독창성으로 분석하였다. 이 외에 많은 연구들이 리얼리티 오디션 프로그램의 흥행에서 사회 맥락적 함축의미를 파악하기 위해 논의한 바 있다. 정치사회적 관점에서의 연구는 신자유주의와의 관계성 또는 참여적 특성이 주를 이룬다(나은경 등 2012). 리얼리티 오디션 프로그램의 참여·투표행위가 시청자들에게 정치 행위와 비슷한 맥락을 경험하게 한다는 것이다.



* 기타에는 기회/운, 사회의 문화적 수준 등이 있음.

<그림 1> 대중문화예술인의 성공요인(한국콘텐츠진흥원 2015)

2.2 오디션 우승을 위한 요건

오디션 프로그램에서 우승자를 예측하는 것은 대단히 어렵다. 최근 흥행에 성공한 TV조선의 <미스 트롯>의 문경태 PD는 우승 후보였던 참가자 지원이 양의 탈락에 대해 “군 장병들의 투표가 결과에 반영됐다. 점수가 퍼포먼스가 좋은 쪽으로 치중할 줄 알았는데, 마음을 울리는 노래에 더 반응한 듯하다”고 답변한 바 있다(뉴스1 2019). 또한 <슈퍼스타-K>의 경우에는 실력 이외에 가능성과 참가자들의 인간관계까지 평가에 반영하는 심사방식을 가지고 있었고, 감성을 자극하는 연출 방식으로 사적이고 드라마틱 한 이야기들이 화제성을 불러일으킨 바 있다(최문경 등 2013).

이를 보면, 참가자들이 통제할 수 없는 외부 요인과 운이라는 요소 또한 오디션에서 작용하고 있다는 것을 알 수 있다. 그러나 외부적 요인보다는 참가자의 내부적인 요인이 그보다 크게 작용 될 것이라는 주장도 있다. 삼성경제연구소(2011)는 시청자가 참가자의 독자적인 음악 세계를 기대하고 있으며, 연속되는 미션에서 창조적인 무대 변신을 연출하는 참가자 만이 생존한다고 분석했다. 또한 한국콘텐츠진흥원(2015)에서 발간한 대중문화예술산업 실태보고서를 살펴보면, 대중문화예술

인이 꼽은 성공요인에서 기회와 운이 해당하는 기타의 수치가 가장 낮음을 알 수 있다.

이에 반해 개인의 노력이 중복 선택으로는 50%, 단독 1순위 선택으로는 30%를 차지하며, 가장 큰 요인으로 꼽혔다. 장규수(2013) 또한 스타가 되기 위해 필요한 것은 학력이나 기술이 아니라 외모, 끼 등 타고난 조건과 개인적인 창의적 자질이라고 언급하였다. 최근 한류의 중심에 있는 방탄소년단의 인기 요인에 대해 미국 매체 CNBC(Consumer News and Business Channel)은 진정성으로 설명한 바 있다. SNS를 통한 팬들과 진실된 소통, 윤리적 기준을 벗어나지 않기 위한 개인적인 노력들이 여러 나라의 팬들의 마음을 움직였다는 것이다(시선뉴스 2019). 그러나 연예인의 자질이나 성공요인을 한마디로 명확히 열거하기는 어렵다. 위의 요인들은 모두 예시일 뿐이며 연예인의 성공에 대한 모든 상황을 설명하지는 못한다고 할 수 있다. 때문에 오디션 프로그램에서 우승자를 예측한다는 것은 큰 불확실성을 내포하고 있다.

2.3 프로그램 소개: 프로듀스 101

2016년 처음 시작한 <프로듀스 101 시리즈>는 3개

<표 1> 프로듀스 101 시즌 별 방영정보

시 즌	방 영 기 간	최고시청률 (닐슨코리아 제공)
시즌1: PRODUCE 101(I.O.I)	2016.01.22~2016.04.01 (11부작)	4.4%
시즌2: PRODUCE 101(Wanna One)	2017.04.07~2017.06.16 (11부작)	5.2%
시즌3: PRODUCE 48(IZ*ONE)	2018.06.15~2018.08.31 (12부작)	3.1%
시즌4: PRODUCE X 101(XI)	2019.05.03~2019.07.19 (12부작)	3.9%

시리즈를 거쳐, 현재는 시즌 4인 <프로듀스X 101>이 방영 중이다. <프로듀스 101 시리즈>는 기획사, 나이, 데뷔 유무와 관계없이 참가한 101명의 연습생들 중 시청자 투표를 통해 선발된 최종 11명이 프로젝트 그룹으로 데뷔하는 포맷이다. 101명의 참가자는 시청자를 ‘국민 프로듀서님’이라 칭하며, 시청자들의 표를 얻기 위해 노력한다. 해당 시리즈는 방영 내내 높은 시청률을 기록하며 대중문화의 새로운 열풍을 몰고 왔다.

기본 포맷은 4개 시즌 모두 동일하지만, 세부적인 구성은 각기 차이를 가지고 있다. 기본적으로는 시청자 투표가 반영되어 최종 우승자가 결정된다. 연습생 101명을 대상으로 레벨 테스트 후 일정한 등급(A~F)이 부여 되고, 등급에 따라 각종 혜택 또는 패널티를 받는다. 즉, 높은 등급의 연습생일수록 시청자들에게 자신을 어필할 기회를 더 가질 수 있는 것이다. 또한 매주 방송을 통해 온라인 투표가 진행되고, 일정 단계에서는 순위에 따라 정해진 인원이 방출된다. 이러한 기본 포맷에 세부 구성은 각 시즌 별로 근소한 차이가 있는데, 현재 진행 중인 <프로듀스X 101>은 기존의 등급제를 변형하여 X등급을 추가했다.

X등급 연습생들은 트레이닝 센터에 입소하지 못하는

제약을 받게 된다. 또한 미션에서 1등 팀 또는 개인에게 주어지는 혜택이 더욱 확대되었으며, 기존 11명을 모두 현장투표로 결정했던 것과는 달리 11명 중 1명은 총 4번 진행되는 투표를 합산한 누적 투표수로 선발된다.

투표기간은 크게 4차로 나뉜다. 5월 13일부터 5월 25일(1차), 6월 1일부터 6월 16일(2차), 6월 29일부터 7월 6일(3차), 7월 13일부터 7월 19일(4차)의 해당 기간 동안 1일 1회 중복투표(총 1일 2회)가 가능하다. 1,2차는 1회 투표 시 11명을 선택할 수 있으며, 3차는 2명, 4차는 1명을 지정해야 한다. 4기간 동안 개인은 최대 88회 투표가 가능하다. 또한 방영 기간 동안에 실제 방송에서는 현장 그룹 배틀이 있는 회차를 제외한 나머지 회차에서 온라인투표 결과를 공지하고 있다. 12회차 방송 중 투표 경과를 9번 공유하여 현재 상황을 확인할 수 있게 하고, 시청자들의 투표를 독려했다.

이 같은 과정을 거치면 최종 우승자 11명이 선발된다. 이들은 프로젝트 그룹인 엑스원(XI)으로 데뷔하는 특전을 누릴 수 있다. 지금까지 우승자들로 구성되었던 그룹은 프로그램의 흥행에 힘입어 국내 정상급 아이돌로서의 인기를 구가했다. 각 시즌 별 우승자들은 다음과 같다.

<표 2> 프로듀스101 시즌 별 우승자

시 즌	우 승 자
시즌1: PRODUCE 101(I.O.I)	전소미, 김세정, 최유정, 김청하, 김소혜, 주결경, 정채연, 김도연, 강미나, 임나영, 유연정
시즌2: PRODUCE 101 (Wanna One)	강다니엘, 박지훈, 이대휘, 김재환, 웅성우, 박우진, 라이관린, 윤지성, 황민현, 배진영, 하성운
시즌3: PRODUCE 48(IZ*ONE)	장원영, 미야와키 사쿠라, 조유리, 최예나, 안유진, 야부키 나코, 권은비, 강혜원, 혼다 히토미, 김채원, 김민주, 이채연

<프로듀스 101 시즌1>은 최종 우승자로 구성된 아이오아이(I.O.I)를 비롯해 여러 스타를 배출했다. 프로젝트 그룹 활동 종료 후에도 세정, 청하 등은 솔로 활동을 이어가고 있으며, 위키미키, 구구단 등 각자의 그룹에서 활발한 활동을 이어가고 있다. <프로듀스 101 시즌2>는 남성 연습생들을 대상으로 하여, 여성들의 열렬한 호응을 끌어냈다. 프로젝트 그룹 워너원(Wanna One)은 한국 아이돌그룹 팬덤⁴의 최대치를 이끌어냈다는 평을 받는다. 이는 데뷔 시점부터 멤버 11명이 각자의 팬덤을 가진 상태에서 출발했기 때문이다(강일권 등 2017). 프로그램의 흥행이 그룹의 인기에도 영향을 미친 것이다. <프로듀스 101 시즌3>가 탄생시킨 아이즈원(IZ*ONE)은 일본에서 55억 매출을 달성하며, 한국 신인으로서는 최초 매출 1위를 기록하는 등 인기 고공행진을 이어가고 있다(MBN스타 2019; 미디어SR 2019). <프로듀스 101 시리즈>의 성공은 현실을 압축적이고 극적으로 보여주는 것에 있다. 시청자들은 참가자들의 고통과 좌절에 공감하고 몰입하기 때문이다(한국국제문화교류진흥원 2018). 과열된 경쟁, 결과 등급에 따른 불이익은 치열한 경쟁사회를 반영하며, 시청자들은 같은 사회의 일원으로서 공감한다. 특히, 타 오디션 프로그램과 <프로듀스 101 시리즈>의 차이점은 게임 육성 시뮬레이션 장르의 특성을 차용하여, 시청자들에게 게임과 비슷한 경험을 하도록 유도했다는 점이다. 육성 시뮬레이션에서 게이머들은 캐릭터를 성장시키고 목표를 달성하여 정서적 교감과 보상을 얻는다는 것을 비추어 살펴보면 더 의미가 있다고 할 수 있다(한국콘텐츠진흥원 2017).

3. 연구설계

3.1 데이터 소개

본 연구에서는 2016년부터 2018년까지 이미 방송된 프로듀스 101 프로그램의 시즌 1부터 시즌 3까지 공개된 데이터를 수집하여 <프로듀스X 101>의 최종 우승자 예측에 활용하였다. 우선 머신러닝의 트레이닝을 위한 데이터들을 수집하기 위하여 각종 인터넷 매체를 통해 대중들에게 공개된 303명(1개 시즌 당 101명의 출연자)의 출연자에 대한 정보를 수집하였다. 이를 위해 출연자의 최종 순위 결정에 영향을 미칠 것으로 예상되고 인터넷을 통해 얻을 수 있는 15개의 예측 변수들을 두 가지의 범주로 구분하여 수집하였다.

첫 번째 범주는 개인특성에 관한 변수로써 출연자의 소속사, 출생지, 나이, 키, 몸무게, 혈액형, 방송출연 경험 유무, 출연 당시의 연습생 기간 등 총 7개 변수에 대한 자료를 수집하였다. 두 번째 범주는 스타성에 관한 지표로써 1차부터 4차까지의 시청자 투표에 의한 순위(시즌 별 총 11회차의 방송으로 구성되며, 매 방송 이후 시청자 투표에 의해 종합순위를 발표함. 단, 4회차 순위는 비공개이며, 5회차 공개를 기점으로 첫 번째 탈락자 그룹이 발생함. 본 연구에서는 첫 번째 탈락자 그룹이 발생하는 5회차까지의 자료를 활용함.), 팀 득표수, 개인 득표수, 첫 번째 탈락자 그룹 발생 전까지의 개인별 인터넷 뉴스 기사 건수(국내 인터넷 사용자들이 주로 사용하는 검색 매체인 네이버 뉴스에서 검색하여 기사 건수 확인) 등 총 8개 변수에 대한 자료를 수집하였다. 자세한 예측 변수는 <표 3>과 같다.

본 연구에서는 <프로듀스 101>의 네 번째 시리즈인 <프로듀스X 101> 출연자 101명 중에 최종 우승자 11명을 예측하고자 한다. 따라서 상기 서술한 시즌 별 예측 변수들과 그 최종결과를 학습시켜 네 번째 시리즈 출

4) 특정한 인물이나 분야를 열정적으로 좋아하는 사람들 또는 그러한 문화현상을 말한다. 흔한 말로 '오빠(누나) 부대'로 불리기도 하며, '워너비(Wanna Be)' 혹은 '그루피(Groupie)'라는 표현도 사용된다(시사상식사전).

<표 3> 예측 변수

변수 범주	변 수 명		변수 소개 및 설명
개인특성	1	Age	나이
	2	Height	키
	3	Weight	몸무게
	4	Blood type	혈액형
	5	Hometown	고향
	6	Experience	방송경험유무 (데뷔 혹은 미디어 노출 경험)
	7	Trained Period	출연 당시 연습생 기간
스타성	1	1st Vote	-총 11회차 방송이며, 매 방송에서 종합순위를 발표
	2	2nd Vote	-4회차 순위는 비공개이며 5회차 공개를 기점으로 첫번째 탈락자 그룹 발생
	3	3rd Vote	- 5회, 8회, 10회에서 각각 탈락 그룹 발생.
	4	5th Vote	-모든 출연자의 데이터를 활용하기 위해 첫번째 탈락자 그룹 발생 전까지의 데이터만 사용
	5	Group Battle Part	-총 11회차 방송 중 3차례의 평가 시스템이 존재 (“그룹배틀”, “포지션”, “콘셉트” 까지 3차례)
	6	Group Battle Team Vote	-이 중 “포지션”, “콘셉트” 평가는 1차 탈락자 발생 이후 시행 -따라서 1차 탈락자 발생 전인 “그룹배틀” 경연 관련 자료 만 사용 -Group Battle Part: 그룹배틀에서 담당할 역할(보컬/래퍼)
	7	Personal Vote in Group Battle	-Group Battle Team Vote: 그룹배틀 결과 팀 득표량 -Personal Vote in Group Battle: 팀 득표량 중 개인 득표량
	8	1st Period News	첫 방송 시작 후 1번째 탈락 그룹 발생 시까지의 개인 별 기사량(네이버 뉴스 내 검색 기준)

연자 101명의 최종 11인 선발 여부를 타겟 변수로 설정하였다.

10회를 반복 수행하였다. 네 가지 모형 중 최종 우승자 예측을 위한 정확도가 가장 높은 모형을 선정하였다.

3.2 예측 모형 생성 및 선정

본 연구에서는 예측 모형 생성을 위해 머신러닝의 지도학습 분류기법을 활용하였다. 다양한 지도학습 방법 중 Random Forest, Decision Tree, Support Vector Machine(SVM), 그리고 Neural Network 기법을 이용하여 타겟 변수를 예측하였다. 예측을 위한 분석 도구로는 Orange(Ver. 3.2)를 사용하였다. 예측 모형의 신뢰도를 높이기 위해 Random Forest, Decision Tree, SVM, Neural Network 기법에 k-fold cross-validation을 사용하였으며 k-fold cross-validation은

4. 예측 모형 평가 및 선택

4.1 예측 모형 성과 검증 지표

본 연구의 모형 별 성과 검증지표로써 구현된 분석기법 중 가장 높은 예측 정확도를 이용하여 가장 높게 나온 분석기법을 선택한 후, 변수 선택법들을 이용해 새로 도출된 변수들의 성능을 검증할 수 있도록 한다. 이를 위해 이진분류의 오차 행렬<표 4>를 활용하여 각 모형의 정확도(Precision)와 재현율(Recall)을 확인하였다.

<표 4> 이진분류의 오차 행렬(Confusion Matrix)

		예측 값	
		Positive	Negative
실제 값	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

True와 False는 실제 값을 예측한 값이 정확하게 맞았는지 또는 그렇지 아니한지를 나타낸다. 즉, True의 경우는 예측 값이 실제 값과 같은 것을 나타내며, False는 이와는 달리 예측 값이 실제 값과 다른 경우를 나타낸다. Positive와 Negative는 예측된 값이 양성인지 아니면 음성인지를 나타낸다. 따라서 True Positive(TP)는 실제 값이 양성인 것을 양성이라고 정확하게 예측한 경우를 나타낸다. True Negative(TN)은 실제 값이 음성인 것을 음성이라고 정확하게 예측했음을 의미한다. 이와는 반대로, False Positive(FP)는 실제 값이 음성인데 양성이라고 예측한 경우를 나타내며, False Negative(FN)은 실제 값이 양성인데 음성이라고 예측한 경우를 나타낸다(Muller and Guido 2016). 이와 같은 오차 행렬을 기반으로 하여 예측 모형을 평가할 수

있는 분석지표를 만들 수 있다. 이는 아래와 같은 <표 5>로 나타낼 수 있다.

4.2 예측 모형의 검정력 평가

위에 서술한 분석지표를 기반으로 하여 네 개 모형의 예측력을 평가하였다. 예측 모형에 대한 검정력을 평가하기 위하여 ROC(Receiver Operating Curve) 분석을 수행하였다. ROC는 x축에 FP(실제 음성을 양성으로 분류) 비율을, y축에 TP(실제 양성을 양성으로 분류) 비율을 플롯팅(plotting)한 것이다. 따라서 FP 비율 값이 낮으면서 TP 비율 값이 높은 것이 더 좋은 정확도를 나타내는 모형이라 할 수 있고, 이 때의 곡선은 위로 보다 볼록한 형태를 나타낸다. 이 곡선(ROC)의 아랫부분을 AUC(Area under ROC)라 칭한다(Keller et al.

<표 5> 오차 행렬 분석지표

구 분	의 미	수 식
정확도 (Accuracy)	전체 예측 값 중, 정확하게 실제 값을 예측한 비율 (실제 양성을 양성으로 실제 음성을 음성으로 예측)	$\frac{TP + TN}{TP + TN + FP + FN}$
정밀도 (Precision)	양성으로 예측 한 값 중, 실제 값이 양성인 비율	$\frac{TP}{TP + FP}$
재현율 (Recall)	실제 양성인 값을 양성으로 예측한 비율	$\frac{TP}{TP + FN}$
F1-Score	정밀도와 재현율의 조화 평균을 사용하여 두 요소를 모두 고려하여 평가한 지표	$2 \times \frac{\text{정밀도} \times \text{재현율}}{\text{정밀도} + \text{재현율}}$

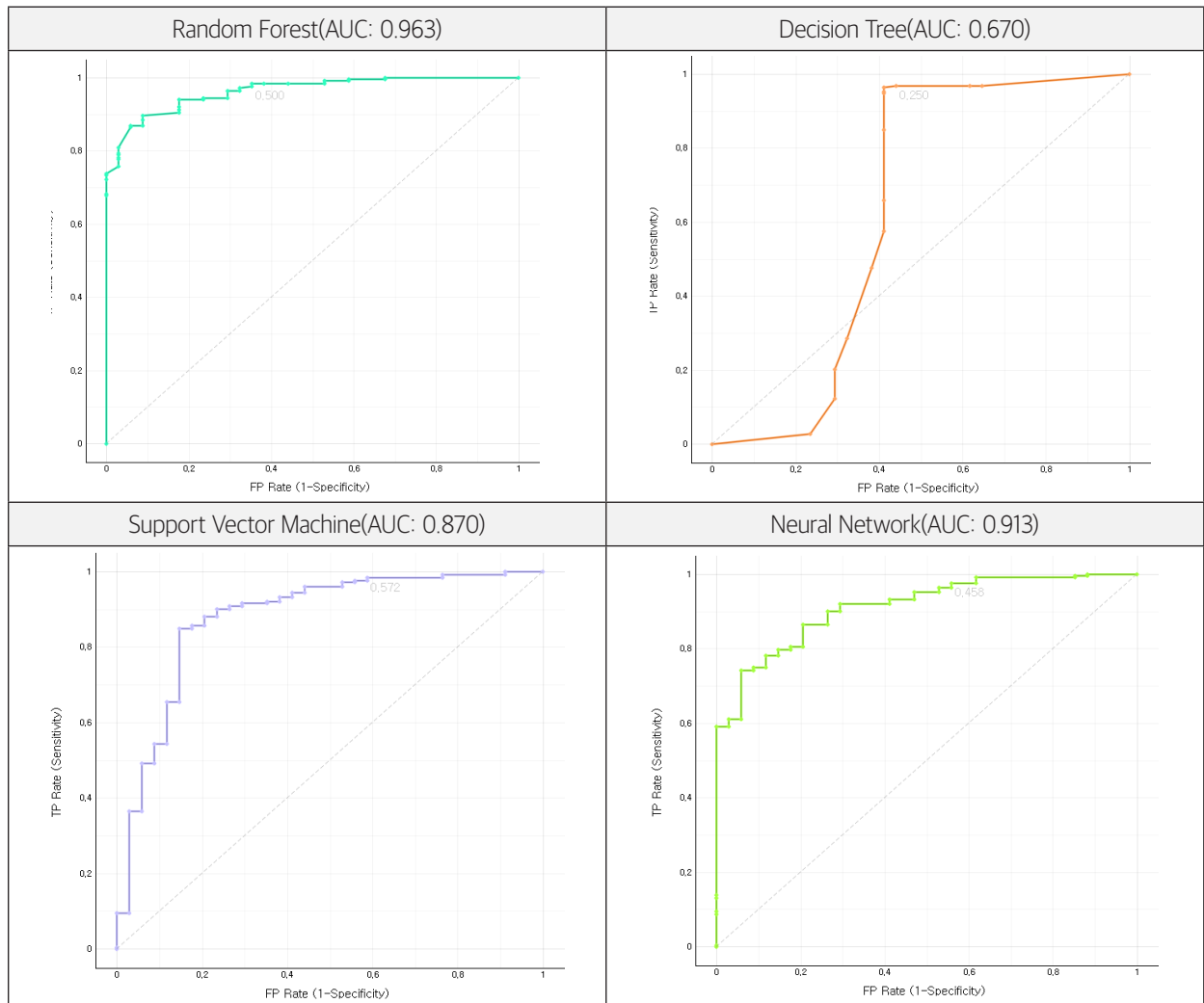
<표 6> AUC에 따른 평가 기준(Keller et al. 2015; Muller et al. 2005; Swets 1988)

AUC = 0.5	비정보적 예측
0.5 < AUC ≤ 0.7	덜 정확한 예측
0.7 < AUC ≤ 0.9	중간 정도의 정확한 예측
0.9 < AUC < 1	매우 정확한 예측
AUC = 1	완벽한 예측

<표 7> 모형 별 예측력

Model	AUC	F1	정밀도(Precision)	재현율 (Recall)
Random Forest	0.963	0.935	0.934	0.937
Decision Tree	0.670	0.917	0.915	0.920
Support Vector Machine(SVM)	0.870	0.878	0.894	0.902
Neural Network	0.913	0.898	0.895	0.906

<표 8> 모형 별 ROC비교



2015). AUC의 값은 최대 1이며 최소 0.5이다. AUC에 따른 모형의 검정력 평가는 <표 6>과 같다.

ROC 분석 결과, <표 7>과 <표 8>에 제시된 바와 같이 Random Forest의 AUC 값이 0.963으로 Decision Tree의 AUC 값 0.670, SVM의 AUC 값 0.870, Neural Network의 AUC 값 0.913에 비하여 가장 높음을 알 수 있다. 또한 본 연구의 예측에 있어서 Random Forest의 예측 알고리즘의 정밀도, 재현율, 그리고 F1 값이 높음을 볼 때 다른 세 개 모형의 알고리즘에 비하여 Random Forest의 전반적인 예측 검정력이 보다 우수하다는 것을 알 수 있다.

5. 우승자 예측 실행 및 평가

5.1 우승자 예측

상기 장에서 설명한 바와 같이 네 가지 예측 모형 중에서 가장 예측 검정력이 우수한 것으로 평가된

Random Forest 알고리즘을 활용한 예측 모형을 통해 <표 9>와 같이 실제 <프로듀스X 101>의 최종 우승자를 예측하였다.

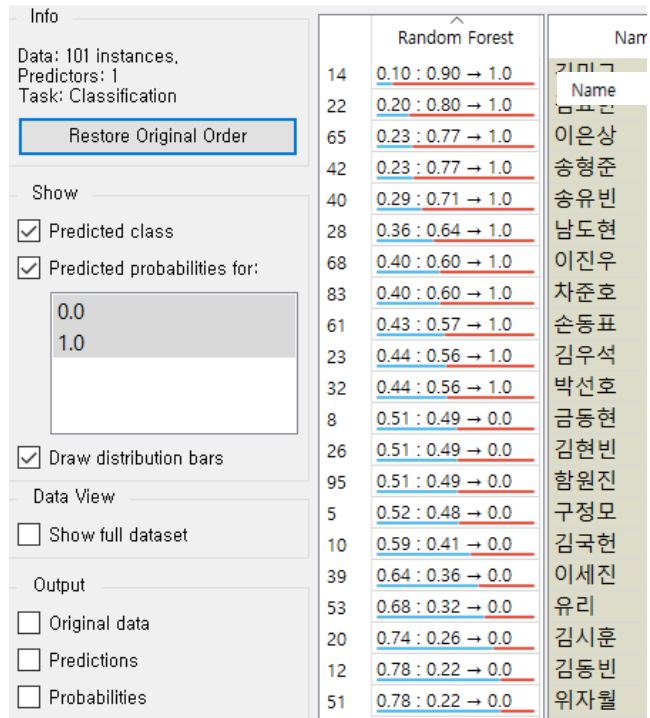
5.2 우승자 예측에 영향을 미친 변수 평가

상기 서술한 바와 같이, 우승자 예측을 위해 총 15개의 변수를 개인특성과 스타성 범주로 나누어 예측에 활용하였다. 총 15개의 독립변수 중 어떤 변수가 우승자를 예측하는데 보다 많은 정보를 담고 있는지 추가로 분석하였다.

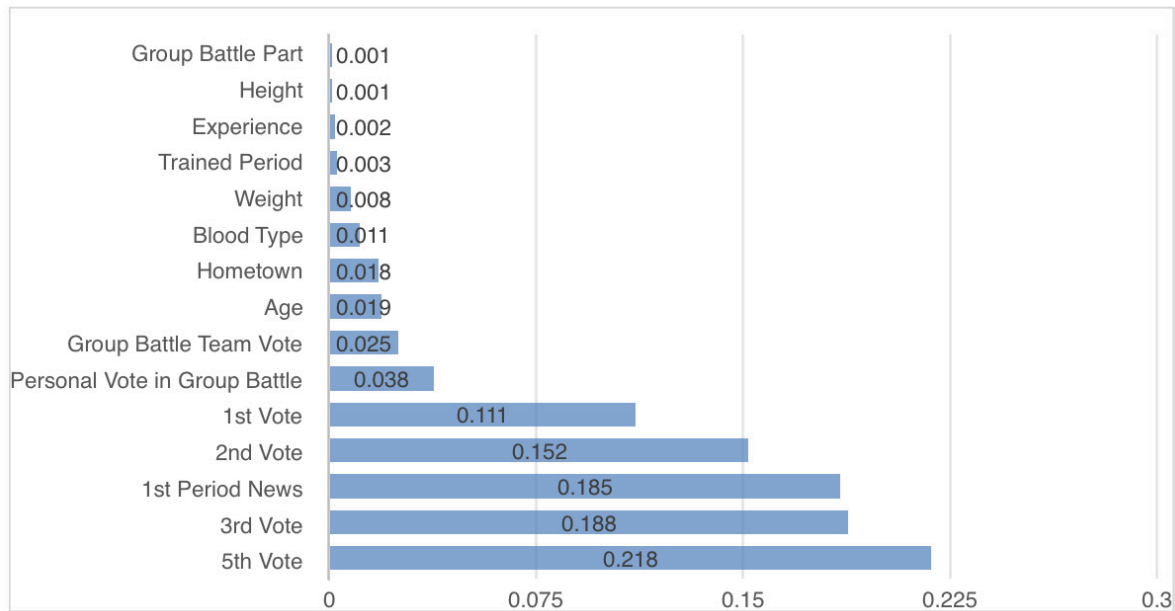
그 결과, <그림 3>에서 보는 바와 같이 우승자 예측을 위한 가장 중요한 변수로 5회차 방송 후 시청자 투표수(21.8%)가 가장 높게 나타났으며, 이어서 3회차 방송 후 시청자 투표수(18.8%)가 중요한 변수로 작용했음이 밝혀졌다. 첫 방송 시작 후 첫 번째 탈락 그룹 발생 시까지의 개인 별 인터넷 뉴스 기사량(18.5%) 또한 세 번째로 중요한 변수로 나타났다. 이 세 가지 변수의 중요도 합계가 약 60%를 넘어 우승자 예측에 있어서 대단

<표 9> Random Forest 모형을 통한 우승자 예측 결과

예측 순위	이름	우승 가능성	모형을 통한 예측	실제 우승 여부 (2019년 7월 19일 방송)
1	김민규	90%	우승	탈락
2	김요한	80%	우승	우승
3	이은상	77%	우승	우승
4	송형준	77%	우승	우승
5	송유빈	71%	우승	탈락
6	남도현	64%	우승	우승
7	이진우	60%	우승	탈락
8	차준호	60%	우승	우승
9	손동표	57%	우승	우승
10	김우석	56%	우승	우승
11	박선호	56%	우승	탈락



<그림 2> 예측 모형 결과 산출 스크린 캡처



<그림 3> 우승자 예측을 위한 변수의 중요도

히 중요한 역할을 했다고 할 수 있다. 나머지 개인특성 범주의 변수들(나이, 고향, 신체조건, 방송경력 등)은 우승자 예측에 있어서 그 효과가 비교적 미미했음을 알 수 있었다.

6. 결과 토의 및 시사점

최근 산업 전반에 걸쳐 빅데이터를 활용한 가치 창출 및 문제 해결에 박차를 가하고 있는 바, 본 연구에서는 머신러닝 알고리즘을 활용하여 리얼리티 오디션 프로그램의 불확실한 경쟁 상황 하에서 대중의 선택 결과를 예측해 보았다. 이를 위해, 본 논문에서는 국내 리얼리티 오디션 프로그램의 개념과 발전, 특성 및 흥행 요인을 분석하고, 2016년부터 인기리에 방영된 <프로듀스 101 시리즈>의 시즌1부터 3까지의 우승자 데이터를 수집하였다. 수집된 누적 데이터를 바탕으로 머신러닝을 통해 네 가지 알고리즘에 기반한 예측 모형 중 Random Forest 알고리즘을 활용한 예측 모형을 구현

하여 최근 종영한 <프로듀스X 101>의 우승자를 예측해 보았다. 이를 통해 본 연구는 중요한 학술적/실무적 시사점을 얻을 수 있었다.

본 연구의 학술적 시사점은 다음과 같다. 첫째, 머신러닝 알고리즘을 연구에 적용하여 문화 산업 분야에서의 연구방법의 확장을 시도했다는 점이다. 머신러닝 기반의 예측 모형을 개발하여 문화 관련 상품과 서비스에 대한 소비자들의 행동 패턴이나 수요 예측 등을 가능케 하는 이론적 가능성을 제시하였다. 서론부에서 기술한 바와 같이 오늘날 다채널화 되어 가는 방송 환경과 시청자의 온라인 활동의 증가로 기존의 TV 시청률이나 화제성 척도를 정확히 재단하기에 어려움이 있다. 연구 문제에서 언급하였던 것처럼, 빅데이터를 통한 분석이 그 보완책이 될 수 있는 바, 이를 통해 문화 산업 분야 연구의 지평을 넓히는데 기여하였다. 즉, 그동안 인간의 경험이나 직관에 의존하여 특정 경연의 우승자를 예측해 오던 방식을 머신러닝 기법을 활용하여 보다 향상된 정확도와 신뢰도로 이를 예측할 수 있는 이론적 방안을 제시하였다.

둘째, 우승자 예측에 어떠한 요인들이 보다 강한 영향을 미치는지에 대한 단서를 제공하였다. 우승자 예측에 영향을 미칠 것으로 예상되는 15개의 변수들을 개인특성과 스타성 두 가지의 범주로 나누어 각각의 영향력을 살펴 보았다. 실제 우승자 예측에 있어서는 참가자의 신체조건이나 개인적인 특성보다는 실제 방송 후 시청자들로부터 받은 투표수에 훨씬 강한 영향을 받는다는 것을 알 수 있었다. 이는 추후 여타의 예능 프로그램의 우승자 예측 모형을 개발하고 확장하는데 의미 있는 근거로 고려될 수 있을 것으로 사료된다.

본 연구의 실무적 시사점은 다음과 같다. 첫째, 본 연구에서 사용한 머신러닝 기법을 활용하여 우수한 예능인을 선발하고자 하는 방송국이나 문화 관련 사업 주체의 선발 프로세스의 실패 확률을 줄일 수 있을 것으로 기대된다. 최근 리얼리티 오디션 프로그램의 등장 이후, 대중이 직접 우승자를 선발하는 시스템이 각광을 받고 있다. 이에 리얼리티 오디션 프로그램이 일반적인 연예인 선발 채널로 자리 잡았으며, 프로그램의 흥행은 연쇄적으로 우승자들을 스타의 반열에 올려 놓고 있다. 즉, 흥행에 성공한 프로그램을 거쳐온 연예인 지망생들은 데뷔 전에 대중들의 선호에 의해 선별된 그룹이라는 것이다. 다시 말해, 대중들의 선호를 머신러닝 기법으로 사전에 예측하여, 우수한 예능인의 선별 과정에서의 실패 확률을 줄일 수 있을 것으로 기대된다.

둘째, 산업 전반에서 수행되고 있는 인적자원 선발 시스템의 신뢰성 검증에 본 연구에서 사용한 머신러닝 기법이 사용될 수 있는 단초를 마련하였다. 우승자 예측 전, 본 예측 모형의 정밀도는 93.4%를 나타냈다. 그 의미는 우승자로 예측한 결과가 실제 우승자일 확률이 93.4%임을 의미한다. 다시 말해서, 총 11명의 우승자를 예측한다면 93.4%인 약 10명은 실제 우승자와 일치해야 한다는 것이다. 그러나 Random Forest 기반의 우승자 예측 모형을 통한 예측 결과(93.4%)보다 실제 우승자를 비교한 정밀도는 $63.6\%(7/11=0.636)$ 로 훨씬 낮

았다. 모형 평가 시의 정밀도 93.4%와는 약 30% 포인트의 큰 차이가 발생하였다.






그러나 2019년 7월 19일에 방영된 해당 프로그램의 마지막 방송에서 발표한 최종 우승자 11명에 대하여 인터넷 상에서의 논쟁이 치열하게 진행 중이다. 많은 시청자들이 우승자 선정 과정에 치명적인 오류가 있었음을 제기하고 있는 것이다(헤럴드경제 2019). 심지어 한 국회의원이 직접 기자회견을 자처하며 사회적 정의의 문 제임을 지적하고 검찰 수사를 촉구하고 있기도 하다. <그림 4>는 실제 해당 방송국의 해당 프로그램 웹사이트에 작성된 항의 댓글을 캡처한 내용이다. 본 연구의 우승자 예측 결과에서는 우승자에 포함되지 않았으나 실제 마지막 방송에서 우승자에 포함된 4명을 언급하며 그들이 우승자로 선정된 근거에 의문을 제기하고 있다. 실제 본 연구의 우승자 예측 모형에서는 한모군의 경우 탈락 확률을 96%로 예측했으며, 강모군은 93%, 이모군은 86%, 그리고 조모군은 81%로 탈락 확률을 예측했었다. 이러한 예측 결과는 많은 시청자들이 갖는 의혹에 대한 강한 지지를 해주고 있다. 특히, 본 연구 결과, <그림 2>에서 1등으로 우승할 것으로 예측되었던 김민규(우승자 당선 가능성 90%)의 경우, 실제 최종 결과 12등을 기록하여 아쉽게 한 등수 차이로 탈락하였다. 본 연구 결과에서 1등으로 우승자에 당선될 것으로 예측되었던 대상자가 한 등수 차이로 아쉽게 탈락한 사실도 대단히 석연치 않은 부분이며, 실제 일부 매체에서 김민규의 탈락을 대단히 아쉬워하고 있다(마이데일리 2019). 그러나 일부 언론 매체나 시청자들의 주장처럼 우승자 선정과정에 부조리 또는 문제가 있었음을 밝히는 일은 대단히 어려운 일이다. 어떠한 기관에서 인적자원 선발과 관련한 의혹이 제기될 경우, 본 연구의 예측 모형과 유사한 모형을 개발하여 선발된 자원에 대한 검증을 시행하여 조직원들의 동요를 막거나 또는 실제 문제점을 점검하게 하는 전략적 도구로 활용할 수 있을 것이다. 따라서 본 연구는 특정 조직의 인적 자원

선발 프로세스에 문제가 발생했을 경우, 머신러닝 기법을 활용한 예측 모형이 제기된 문제점을 검증하고 이를 개선할 수 있는 의미 있는 전략적 도구가 될 수 있음을 제시하였다.

본 연구의 한계는 정형 데이터로는 분석하기 어려운 요인들이 존재한다는 데 있다. 선행연구에서 언급하였던 오디션 선발 요인 중, 운이나 개인 노력, 인간 관계 등은 수치로 측정이 어려운 정보로써 반영되기 어려웠다. 또한 주요 요인으로 분석되었던 3차와 5차 시청자 투표는 투표 직전 방송의 퍼포먼스와 음악성 등에 영향을 받았을 것으로 생각된다. 이때, 퍼포먼스나 음악성 등의 중요도를 따로 분석하기 어렵다는 점은 본 연구의 한계일 것이다. 이렇듯 정밀도에 영향을 주는 요인을 정확히 파악하기 어렵기 때문에 문제의 추론을 통해 지속적인 검증과 발전을 도모해야 할 것이다. 그러나 이러한 논의들을 차치하고서, 측정 가능한 요인들 만으로도 실제와 결과를 가깝게 예측 할 수 있었다는 점은 고무적이라고 할 수 있다. 때문에 향후 본 연구를 바탕으로 이미지나 음성과 같은 비정형 데이터를 활용하여 정형 데이터의 한계를 보완한다면 보다 정교한 분석을 기대해 볼 수 있을 것이다.

 프로듀스X101

지* · 2019.07.25 12:57

조  한  이  강  4명이 붙고 붙어야 할 민규랑 진혁이 황윤성? 이런애들이 떨어지고 이유불문하고 진혁이 떨어진게 켈충격이다 제일미스터리인게 10위 강  가 아니라 이진혁자리인데 해명이라고한게 소수점 반올림? 말도안되는 멍 소리하네



'프듀X101' 제작진이 순위 차이에 대해 좀 더 자세...

 프로듀스X101

. · 2019.07.25 12:53

10위



'프듀X101' 제작진이 순위 차이에 대해 좀 더 자세...

 프로듀스X101

단아**** · 2019.07.25 12:53

피디랑 제작진은 죄가 없지요 대기업 윗대가리가 시키는대로 노예질 했을테니 ㅈ 씨제이 엠넷 이런 비리의 온상들이 깊이 반성하고 사죄하고 처벌받고 보상해내라 그냥 안넘어간다 갑질하는 미개한 것들 회사의 명운이 바람앞에 촛불임을 똑똑 알거라 ㅈ

<그림 4> 해당 방송국 웹사이트 향의 댓글

참고 문헌

[국내 문헌]

1. 강일권, 권석정, 차우진, 정덕현, 모신정 2017. *대중 문화 트렌드 2018 뉴미디어와 콘텐츠의 결합*, 서울:마리북스.
2. 김기홍 2016. “문화연구에서 빅데이터의 효용과 의미,” *인문콘텐츠* (-:2), pp. 103-215.
3. 김주미, 최정화, 박동진 2018. “미디어 이용에 따른 건강정보 과부하가 건강정보 불신에 미치는 영향” *한국광고홍보학보* (20:2), pp. 37-63.
4. 김해원, 이미나 2016. “빅데이터와 엔터테인먼트 콘텐츠: 사례연구 및 전망,” *인터넷정보학회논문지* (17:2), pp. 109-118.
5. 나은경, 손영준, 김옥태 2012. “리얼리티 서바이벌 오디션 프로그램과 소셜 미디어 이용을 통한 시청자 참여의 효과,” *언론과학연구* (12:4), pp. 179-220.
6. 나은영, 나은경 2019. *엔터테인먼트 심리학*, 서울:컬처룩.
7. 박지연, 김효근, 서현주 2019. “경영예술 개념 및 경영예술 창작 모형의 타당성 평가 연구: 근거이론적 접근,” *지식경영연구* (20:2), pp. 155-182.
8. 배국남 2016. *스타란 무엇인가*, 서울:논형.
9. 백현미, 이새롭 2018. “문화에 따른 개방형 협업 지식공유 활동 비교 연구; 집단주의 문화와 개인주의 문화를 중심으로,” *지식경영연구* (19:2), pp. 133-150.
10. 삼성경제연구소 2011. “SERI 경영노트-경연 프로그램 열풍과 시사점” *삼성경제연구소* (-:132), pp. 1-11.
11. 우종필, 이응환 2018. “빅데이터 분석을 통한 천만 관객 영화 예측 모델,” *한국빅데이터학회지* (3:1), pp. 63-71.
12. 유진룡, 최화열 2010. “엔터테인먼트 산업의 개념과 형성,” *서비스마케팅저널* (3:1), pp. 17-34.
13. 윤홍근 2013. “문화산업에서 빅데이터의 활용방안에 관한 연구,” *글로벌문화콘텐츠* (0:10), pp. 157-179.
14. 이경숙 2011. “오디션 프로그램의 리얼리티와 경쟁 그리고 참가자의 순환적 위치,” *미디어,젠더&문화* (-:20), pp. 107-136.
15. 장규수 2013. *스타시스템*, 서울:커뮤니케이션북스.
16. 장리, 최강준, 이재영 2017. “온라인 구전량 및 평점과 시기별 영화 흥행과의 관계,” *지식경영연구* (18:2), pp. 65-83.
17. 장재영 2017. “소셜 빅데이터 분석과 기계학습을 이용한 영화흥행예측 기법의 실험적 평가,” *한국인터넷방송통신학회 논문지* (17:3), pp. 177-173.
18. 조정환 2012. “스포츠 빅데이터 활용과 전망,” *한국체육측정평가학회지* (14:3), pp. 01-11.
19. 최문경, 박지훈, 박진선 2013. “경쟁보다 성장을 강조하는 한국 리얼리티 오디션 프로그램,” *언론과학연구* (13:4), pp. 590-618.
20. 최영준 2015. “리얼리티 오디션 프로그램 수용자들이 느끼는 ‘재미(fun)’에 관한 고찰,” *한국콘텐츠학회논문지* (15:6), pp. 13-23.
21. 한국국제문화교류진흥원 2018. *한류나우(Hallyu Now)-9+10월호 격월간 한류 심층 보고서*(26), pp. 1-89.
22. 한국정보화진흥원 2015. *2015년 빅데이터 글로벌 사례집*, pp. 1-152.
23. 한국콘텐츠진흥원 2015. *2015 대중문화예술산업 실태 보고서*, pp. 1-208.
24. 한국콘텐츠진흥원 2017. *KOCCA 이슈분석, <프*

로듀스 101>으로 본 ‘팬-콘텐츠’의 가능성과 진단(17:06), pp. 1-7.

diagnostic systems,” *Science* (240:4857), pp. 1285-1293.

[국외 문헌]

1. Baltruschat, D. 2009. “Reality TV formats: the case of Canadian Idol,” *Canadian Journal of Communication* (34:1) pp. 41-59.
2. Hall, A. 2006. “Viewers’ perceptions of reality programs,” *Communication Quarterly* (54:2), pp. 191-211.
3. Kelleher, J. D., Namee, B. M., and D’Arcy, A. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, MIT press.
4. Muller, A. C., and Guido, S 2016. *Introduction to Machine Learning with Python*, New York: O’Reilly Media.
5. Muller, M. P., Tomlinson, G., Marrie, T. J., Tang, P., McGeer, A., Low, D. E., and Gold, W. L. 2005. “Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia?” *Clinical infectious diseases* (40:8), pp. 1079-1086.
6. Nabi, R. L., Biely, E. N., Morgan, S. J., and Stitt, C. R. 2003. “Reality-based television programming and the psychology of its appeal,” *Media psychology* (5:4), pp. 303-330.
7. Sayre, S., and King, C. 2010. *Entertainment and society: Influences, impacts, and innovations*, New York: Routledge.
8. Swets, J. A. 1988. “Measuring the accuracy of

[URL]

1. 굿데이터코퍼레이션 2019. “2019년 7월 2주타 비드라마 순위”, <https://mygooddata.tistory.com/786>.
2. 뉴스1. 2019 “미스트롯’PD’종편예능최고시청률? 상상못했죠,” <http://news1.kr/articles/?3600878>.
3. 미디어SR. 2019. “아이즈원, 日 인기 고공행진...韓걸그룹 최초 신인 매출 1위 기록”, <http://www.mediasr.co.kr/news/articleView.html?idxno=53202>.
4. 시선뉴스 2019. “[이슈체크] 웹블리 스타디움 입성 ‘방탄소년단(BTS)’...인기 요인 분석”, <http://www.sisunnews.co.kr/news/articleView.html?idxno=106477>.
5. 헤럴드경제 2019. “하태경”프로듀스 X101’투표조작’거의 확실...검찰 수사해야”, <http://news.heraldcorp.com/view.php?ud=20190724000021>.
6. MBN스타 2019. “아이즈원, 55억 매출 달성...日데뷔 가수 중 최고”, <http://star.mbn.co.kr/view.php?year=2019&no=443045&refer=portal>.
7. 마이데일리 2019. “프로듀스X 101 김민규, 12등으로 안타깝게 탈락”, http://www.mydaily.co.kr/new_yk/html/read.php?newsid=201907200021537604&text=da.

● 저 자 소 개 ●



곽주영 (Juyoung Gwak)

현재 전남대학교 일반대학원 경영학과 석사 과정에 재학 중이다. 주요 관심분야는 지식 경영, 머신러닝, 빅데이터, 문화산업, 문화기술(CT) 등이다.



윤현식 (Hyun Shik Yoon)

현재 전남대학교 경영대학 경영학부 조교수로 재직 중이다. University of Missouri(Columbia)에서 박사학위를 취득하였고, 동 대학 경영대학에서 겸임교수를, Oklahoma State University(Stillwater)에서 교수를 역임하였다. 주요 관심분야는 머신러닝을 활용한 소비자행동분석, 데이터마이닝, 정보보안, 기술경영 등이다. 지금까지 International Journal of Information Management, Journal of Computer Information Systems 등 국제 학술지에 논문을 발표하였다.