

## 유튜브 쇼츠의 조회수 예측 - 뷰티 광고 콘텐츠를 중심으로

김슬<sup>○</sup>, 김수인\*, 박훈\*, 임수빈\*, 정재우\*

<sup>○</sup>서울교육대학교 컴퓨터교육학과

e-mail:{niceonesuri<sup>○</sup>, suinkim864, growth.da.hoon, sue9927, joowoo777}@gmail.com

## Predicting the number of views for YouTube Shorts - Focusing on beauty advertising contents

Seul Kim<sup>○</sup>, Suin Kim\*, Hoon Park\*, Subin Lim\*, Jae-Woo Jung\*

<sup>○</sup>Dept. of Computer Education, Seoul National University of Education

### 요약

본 연구에서는 유튜브 쇼츠 내 뷰티 광고 콘텐츠의 메타 데이터와 시청각 데이터를 수집하여 조회수를 예측하는 모델을 제안한다. 모델링 이후 SHAP 분석을 통해 변수들의 기여도를 확인한 결과, 쇼츠 출시(2021. 07) 이후 개설된 신생 채널의 조회수가 더 높은 경향을 보였다. 또한 유튜버의 외모와 함께 창의적인 콘텐츠 구성 등의 다양한 특성이 조회수에 복합적인 영향을 미쳤으며, 발화 속도가 느린 영상의 조회수는 낮게 예측되는 경향이 있었다. 더불어 보다 간접적인 광고 콘텐츠의 조회수가 높은 경향도 확인되었다. 본 연구는 뷰티 광고 유튜브 쇼츠 제작에 참고할 수 있는 방향성과 기준을 제시하고자 한다.

▶ Keyword : 쇼츠(Shorts), 뷰티(Beauty), 광고(Advertising), 유튜브(Youtube)

실정이다. 따라서 본 연구는 뷰티 광고 중 유튜브 쇼츠에 특화된 조회수 예측 모델을 제안하고자 한다.

### I. Introduction

2023년 기준 국내 시청자가 사용하는 숏폼 콘텐츠 플랫폼으로 유튜브 쇼츠(74.8%)가 1위를 기록한 동시에 국내 뷰티 업종의 디지털 광고비는 전년 대비 4% 이상 증가했다[1, 2]. 이러한 배경에서 뷰티 산업에서 유튜브 쇼츠 광고 전략 수립에 참고할 수 있는 방법론의 필요성이 대두된다.

### II. Preliminaries

유튜브 동영상의 조회수를 예측하는 모델 연구는 활발하게 이루어져 왔으나[3, 4] 2021년 7월 출시된 쇼츠의 조회수를 예측하는 선행 연구는 비교적 부족한

### III. The Proposed Scheme

#### 1. 데이터 수집, 전처리 및 모델링

Table 1. Data Category

수집 방식 및 변수	
대상	<ul style="list-style-type: none"><li>국내 뷰티 유튜버 192명 선정, YouTube Data API로 유튜버의 광고 쇼츠 추출</li><li>최종 2184개 쇼츠 선정</li></ul>
메타 데이터	<ul style="list-style-type: none"><li>영상 게시일, 구독자 수, 광고 표현 방식, 채널 생성 기간 등</li><li>총 7개 변수 수집</li></ul>
시청각 데이터	<ul style="list-style-type: none"><li>썸네일의 HSV, 유튜버 PrettyScale 값, 목소리 발화 속도 평균값 등</li><li>총 9개 변수 수집</li></ul>

모델의 종속 변수는 쇼츠의 조회수로, 편차에 따른 학습 저하를 예방하고자 자연로그를 취했다[3]. 전체 데이터 중 1747개를 학습, 437개를 평가 데이터로 분리했으며 모든 모델의 목적함수로는 평균 오차 제곱(MSE)을 채택했다. 최종적으로 XGBoost 회귀 모델이 MSE 1.36으로 가장 우수한 성능을 기록하였으며, 이는 기존 유튜브 영상(롱폼)의 조회수를 예측한 선행 연구[3]와 비교할 때 유의미하게 개선된 수치라는 점에서 의의가 있다.

Table 2. Metrics

	RF	XGBoost	연구 [3]
MSE	1.45	1.36	5.03

## 2. 예측 결과 해석

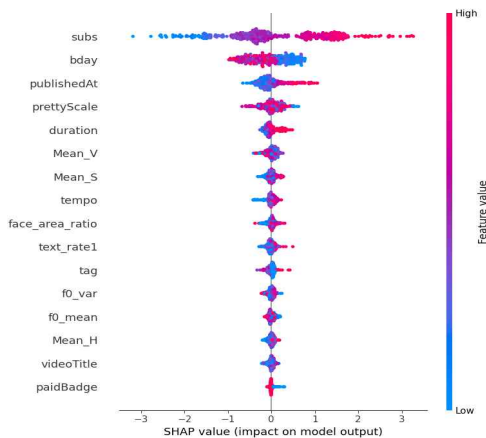


Fig. 1. SHAP Analysis Results

SHAP분석[Fig. 1.]을 통해 각 변수가 최종 모델에 미치는 영향을 측정한 결과 주목할 변수는 [Table 3.]과 같다. 상위 변수 중 유튜브의 구독자 수(subs)와 쇼츠 게시 시간(publishedAt)은 조회수와 양(+)의 상관관계가 있음을 직관적으로 파악할 수 있어 [Table 3.]에서 제외하였다.

Table 3. Interesting Features

범주	주목할 변수
메타	bday : 채널 생성 이후 경과한 일수 paidBadge : 광고 직/간접 표시 여부
시청각	prettyScale : 유튜브의 외모 매력도 tempo : 유튜브 발화 속도의 평균값

먼저 쇼츠 출시 이후 개설된 신생 채널의 예측 조회수가 높은 경향이 나타났다. 이는 전통적 롱폼 영상을 꾸준히 생산해온 기존 채널보다 쇼츠를 전문적으로 타겟팅하는 신생 채널에 주목할 필요성을 시사한다.

다음으로 '유료 광고 포함' 여부를 공식 표기한 쇼츠보다 제목, 태그, 설명 등에 간접적으로 광고를 표기한 쇼츠의 조회수가 다소 높게 예측되는 경향이 있었다. 이는 직접적으로 광고임을 명시한 콘텐츠에 시청자가 거부감을 느낄 가능성이 있어 자연스러운 제품 광고 기획이 중요함을 의미한다.

시청각 데이터로는 유튜브의 외모 매력도와 발화 속도에 주목하였는데, 외모 매력도가 낮음에도 불구하고 높은 조회수가 예측된 데이터를 분석한 결과 창의적이고 흥미로운 콘텐츠 구성, 제품 설명의 명확성 등의 특성을 확인할 수 있었다. 이는 유튜브의 외모와 함께 다양한 요소가 영상의 조회수에 복합적인 영향을 미치는 것으로 해석할 수 있다. 또한 유튜브의 발화 속도가 느릴수록 예측 조회수가 낮은 경향을 발견했으며, 이는 쇼츠 광고가 적당한 빠르기의 음성으로 제품의 정보를 전달할 필요가 있음을 시사한다.

## IV. Conclusions

본 연구 결과를 바탕으로 뷰티 유튜버, 광고주, 마케터는 메타 데이터와 시청각 데이터를 고려하는 전략적인 쇼츠를 기획할 수 있을 것이다. 본 연구에 적용한 방법론이 뷰티 이외 다른 산업 분야의 쇼츠 광고에도 적용될 수 있는지 후속 연구를 진행하고자 한다.

## References

- [1] Open Survey, <https://blog.opensurvey.co.kr/trendreport/socialmedia-2023/>
- [2] Mezzomedia, [https://www.mezzomedia.co.kr/data/insight\\_m\\_file/insight\\_m\\_file\\_1666.pdf](https://www.mezzomedia.co.kr/data/insight_m_file/insight_m_file_1666.pdf)
- [3] Haeyeon Park, "A New YouTube View Count Prediction Method using Deep Audio-Video Multimodal Learning," 2019
- [4] Lavanya Sunder, "Predictive Model for Views in YouTube Beauty Community," 2016