

## 한국 웹툰의 글로벌 소비 현황과 시장 확장 전략 : 데이터 기반 계량 분석

김주희 · 공민정 · 김가영 · 이해림 · 이신형

### [국문초록]

본 연구는 한국 웹툰의 해외 소비자 데이터를 활용하여 단계적 분석을 수행하고, 이를 바탕으로 글로벌 시장 확장 전략을 수립하고자 한다. 먼저, 로지스틱 회귀분석과 네트워크 분석을 통해 해외 소비자들의 한국 웹툰 소비 실태를 파악한 결과, 외국인 한류 팬의 한국 출판물 및 애니메이션 소비 경험이 웹툰 이용에 긍정적 영향을 미쳤으며, 일반 해외 소비자의 만화·코믹스 소비 경험 또한 유의미한 관련성을 보였다. 이어 토픽 모형 분석을 통해 한국 웹툰 해외 소비자를 ‘삶-교훈 중심형’, ‘시각요소 중심형’, ‘캐릭터 중심형’, ‘로맨스 중심형’, ‘복합 장르형’의 다섯 가지 유형으로 분류하였다. 각 유형의 흥행 성과를 로지스틱 회귀분석과 머신러닝 기법인 랜덤 포레스트를 통해 분석한 결과, ‘시각요소 중심형’과 ‘복합 장르형’ 유형이 유의미하게 높은 구독자 수를 기록하였다. 이러한 경향은 인공지능 기반의 SHAP 분석 결과에서도 일관되게 확인되었다. 본 연구는 웹툰을 독립적인 문화 콘텐츠로 재조명하고, 전통적 통계 기법과 머신러닝을 결합한 데이터 분석을 통해 새로운 통찰을 제시한다. 연구 결과는 한국 웹툰의 글로벌 소비자 세분화와 타겟 마케팅 전략 수립에 중요한 산업적 시사점을 제공하며, 문화 콘텐츠 정책 수립 시 해외 소비자의 이용 패턴과 선호도를 반영하여 맞춤형 지원체계를 구축해야 한다는 정책적 함의를 제시한다.

[주제어] 웹툰, 한류, 네트워크 분석, 토픽 모형 분석, 랜덤 포레스트, SHAP

\*본 논문은 문화체육관광부 주관 ‘2024 문화 디지털혁신 및 문화데이터 활용 공모전’ 대상 수상작을 학술 논문으로 발전시킨 것이다.

\*\*모든 저자는 본 논문에 동등하게 기여하였다.

투고일: 2025. 6. 3 심사일: 2025. 6. 26. 게재 확정일: 2025. 8. 12.

<https://doi.org/10.16937/jcp.2025.39.2.111>

김주희\_숙명여자대학교 일반대학원 경영학과 석사졸업/제1저자(jo041h2@sm.ac.kr)

공민정\_숙명여자대학교 일반대학원 경영학과 석사졸업/공동저자(mjkong12@sm.ac.kr)

김가영\_숙명여자대학교 일반대학원 경영학과 석사졸업/공동저자(kk991011@sm.ac.kr)

이혜림\_숙명여자대학교 일반대학원 경영학과 석사졸업/공동저자(2331609@sm.ac.kr)

이신형\_숙명여자대학교 경영학부 부교수/교신저자(sh.lee@sm.ac.kr)

## I. 서론

웹툰(webtoon)은 ‘웹(web)’과 ‘카툰(cartoon)’의 합성어로, 한국으로부터 유래한 디지털 만화 형식을 지칭하는 고유명사이다. 기존 출판 만화와 차별화된 스크롤 방식의 연출과 연재 형식을 통해 독자들의 접근성을 증대시켜 한국 만화 산업의 새로운 성장 동력으로 자리 잡고 있다(박지혜, 2022). 한국의 웹툰 산업은 디지털 기술과 플랫폼 비즈니스의 발전에 힘입어 지속적인 확장세를 보이고 있으며, 웹툰 서비스 플랫폼은 단순한 콘텐츠 제공을 넘어 창작자와 독자 간의 유기적 연결을 촉진하고 소비자 경험을 극대화하는 방향으로 진화하고 있다(김영재, 2015; 이지혜·임성준, 2018).

2022년 한국의 만화·웹툰 산업 규모는 전년 대비 23% 성장한 2조 6,240억 원에 달했으며, 국내 웹툰 플랫폼의 해외 진출 확대에 따라 수출 규모 또한 전년 대비 31.3% 증가한 1억 764만 달러를 기록하였다(문화체육관광부, 2024). ‘2023 웹툰 사업체 실태조사’에 따르면, 웹툰 사업체의 매출 구조에서 수출이 전체 매출의 12.5%를 차지하며 유료 콘텐츠에 이어 두 번째로 높은 비중을 보이고 있다(한국콘텐츠진흥원, 2024). 이는 웹툰 산업이 국내 시장을 넘어 글로벌 콘텐츠 시장에서 주요한 문화상품으로 자리매김하고 있음을 시사한다.

‘2024 해외한류실태조사’는 한국 웹툰의 성장 잠재력을 실증적으로 입증하고 있다(한국국제문화교류진흥원, 2024). 해외 소비자들은 한국 웹툰의 높은 스토리 완성도와 섬세한 그림 표현을 핵심적 강점으로 평가하며, 이는 강력한 구전효과로 이어지고 있다. 특히, 네이버의 글로벌 웹툰 플랫폼인 ‘WEBTOON’<sup>1)</sup>의 높은 이용률과 지인 추천 의향은 한국 웹툰의 글로벌 시장 잠재력을 보여주는 주요 지표이다. 이러한 본질적 경쟁우위

1) 본 연구에서 ‘웹툰’은 한국으로부터 유래한 디지털 만화 형식을 지칭하며, ‘WEBTOON’은 네이버가 운영하는 글로벌 웹툰 플랫폼을 지칭하는 고유명사로 구분하여 사용한다.

를 기반으로 적절한 전략이 수립된다면, 현재의 마니아층 중심 소비 단계에서 벗어나 보다 광범위한 글로벌 소비자층을 확보할 가능성이 열릴 것이다.

한편, 글로벌 빅테크 기업들의 웹툰 시장 진출이 가속화되면서 산업 내 경쟁 구도는 급변하고 있다. 애플북스(Apple Books)는 미국을 시작으로 전 세계 51개국에서 웹툰 서비스를 단계적으로 도입할 계획을 발표하였으며, 아마존(Amazon)은 아마존 플립툰(Amazon Fliptoon) 서비스를 출시하며 시장 진입을 가속화하고 있다(황순민, 2024). 이러한 글로벌 기업들의 시장 참여는 한국 웹툰 산업의 성장 가능성을 입증하는 동시에, 국내 웹툰 기업에 차별화된 경쟁 전략 수립의 필요성을 제기하고 있다. 특히, 일본 만가(漫画; manga)가 여전히 글로벌 시장을 선도하는 가운데, 한국 웹툰은 글로벌 디지털 콘텐츠 시장에서 중대한 전환점을 맞이하고 있다(한국국제문화교류진흥원, 2024).

이처럼 위협과 기회가 공존하는 시장 환경 속에서, 한국 웹툰은 글로벌 콘텐츠로서 한 걸음 더 도약할 것인지, 혹은 경쟁 심화로 인한 정체기를 맞이할 것인지를 기로에 서 있다. 상황이 이러함에도, 기존 연구들은 웹툰 소비자를 K-pop이나 K-drama와 같은 한류 콘텐츠의 팬층으로 한정하는 경향이 있어 시장 확장을 통한 기회이익을 상실하고 있기에, 보다 포괄적인 글로벌 소비자 관점에서의 학술적 접근이 절실한 시점이다.

이러한 맥락에서 본 연구는 한국 웹툰의 글로벌 소비 실태를 다각도로 분석하고, 이를 토대로 시장 확장 전략을 도출하고자 한다. 구체적으로는 첫째, 한국 웹툰 해외 소비자가 어떤 콘텐츠 유형과 높은 소비 연관성을 보이는지를 실증적으로 규명하고, 둘째, 이들의 웹툰 소비 유형을 토픽 기반으로 분류하며, 셋째, 유형별로 웹툰의 글로벌 흥행 성과에 미치는 영향을 계량적으로 분석하고자 한다. 특히, 본 연구는 한국 웹툰 해외 소비자들이 실제로 남긴 댓글 데이터를 분석 기반으로 활용함으로써, 소비자 관점에서의 실증적 통찰을 제공하고 글로벌 시장 확장을 위한 전략 수립에 실효성 있는 시사점을 제공하고자 한다.

## II. 연구 문제

웹툰은 영화나 드라마처럼 완결된 서사를 제시하기보다는, 세로 스크롤 기반의 ‘무한 캔버스’ 형식을 통해 회차마다 ‘열람 지속 여부’, ‘댓글·추천·결제 참여’ 등의 반복

적 선택을 유도하는 특성을 보인다(Chen et al., 2023; Lamerichs, 2020). 이러한 특성은 이용자가 욕구 충족을 위해 능동적으로 매체를 선택한다는 이용·충족 이론(Katz et al., 1973)과 소비 선택이 다차원적 가치 평가에 기초한다는 소비가치 이론(Sheth et al., 1991)의 관점에서 해석될 수 있다. 특히, 짧은 주기의 선택-참여-보상이 순환하는 환경에서, 이용자가 각 회차에서 판단한 가치는 즉각적인 체류 또는 이탈 행동으로 연결되고, 이는 실시간 피드백 루프를 형성하게 된다. 결국 웹툰 산업의 경쟁력은 이러한 소비자 가치 추구 과정을 정밀하게 포착하고, 이를 데이터화하여 수익 모델 및 글로벌 전략에 반영하는 역량에 달려 있다고 할 수 있다(Gillespie, 2010). 즉, 소비자 개개인이 발신하는 선호 신호를 체계적으로 포착하고 분석하는 작업이 웹툰 생태계의 지속적 성장에 핵심적인 역할을 한다. 그럼에도 불구하고, 지금까지의 웹툰 연구는 웹툰을 주로 한류 콘텐츠의 하위 범주로 간주하여, 그 고유한 소비 메커니즘과 수익 구조를 충분히 조명하지 못하였다. 또한, 일본 만가나 미국 코믹스 등 기존 만화 시장과의 경쟁 구도 속에서 웹툰의 차별적 위상을 심층적으로 규명하지 못한 한계도 존재한다.

한편, 한국 웹툰의 글로벌 시장 진출이 가속화됨에 따라, 산업 경쟁력 강화를 위한 다양한 전략이 모색되어 왔다. 이에 따라 웹툰 관련 연구도 활발히 진행되고 있으며, 글로벌 경쟁력 확보 방안을 탐색한 다양한 선행연구가 존재한다. 예컨대, 박지혜(2017)는 웹툰 플랫폼별 차별화된 지원과 중소형 전문 플랫폼 육성의 필요성을 강조하면서, 불법 복제를 방지하기 위한 모니터링 시스템 구축을 제안하였다. 또한, 박지혜(2022)는 IP 세계관 확장과 슈퍼 IP 발굴이 웹툰의 핵심 비즈니스 전략으로 대두되고 있음을 언급하며, 현지화 및 정부 지원 정책의 중요성을 역설하였다. 이근아(2023)는 웹툰을 한류 콘텐츠의 원천 IP로 활용하여 독자 유입을 확대하는 전략을 제시하였고, 신준식(2023)은 콘텐츠 생산량의 한계를 지적하며 불법 유통 문제 해결이 시급함을 강조하기도 하였다. 최용석 등(2023)은 웹툰 창작자 및 지원기관을 대상으로 한 IPA 분석을 통해 금융 지원과 저작권 보호가 글로벌 경쟁력 확보의 핵심 요소임을 도출한 바 있다.

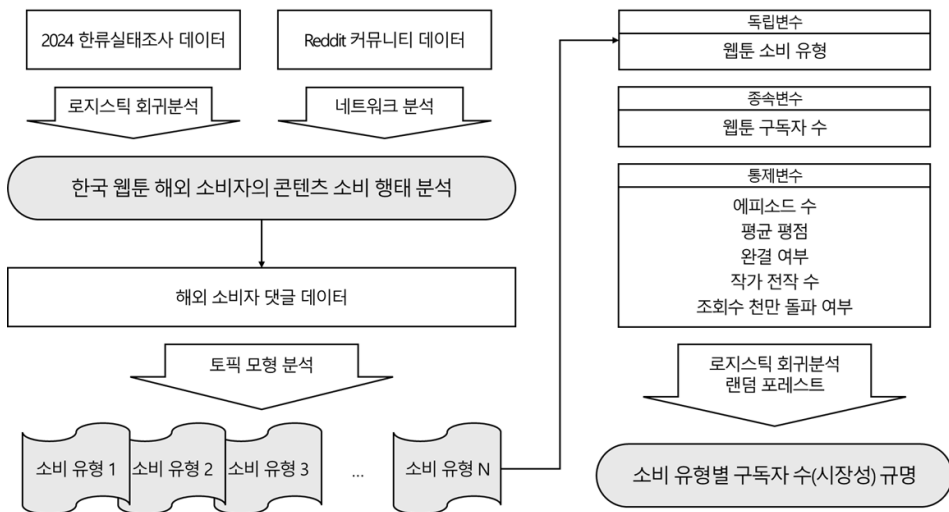
그러나 이들 연구는 주로 정책적·산업적 차원에 초점을 두고 있어, 실제 소비 주체인 글로벌 독자의 이용 행태와 선호 요인에 대한 실증적 분석은 여전히 미비한 실정이다. 최근 들어 웹툰에 관한 실증적 연구가 일부 진행되고 있지만, 웹툰의 구독자 수나 평점과 같은 정량적 데이터에 의존한 단순 통계 분석이 주를 이루거나(윤기현 등, 2015), 이용 내역 기반의 추천 시스템 구축 및 이미지 기반 머신러닝 등 기술적 접근에 치중되어

있어(오세규·강주영, 2022; 이준식·박도형, 2019), 해외 소비자의 복합적 소비 행태를 정교하게 규명하기에는 한계가 있다.

이에 본 연구는 한국 웹툰의 해외 소비 특성을 실증적으로 분석하고 효과적인 글로벌 확장 전략을 모색하기 위해 단계적 분석 설계를 채택하였다. 구체적인 연구 문제와 그에 대응되는 분석 방법은 다음과 같다. 먼저, 로지스틱 회귀분석과 네트워크 분석을 통해 한국 웹툰 해외 소비자의 소비 현황을 파악하고, 어떤 콘텐츠 유형과 밀접한 소비 연관성을 보이는지를 실증적으로 규명하고자 한다(연구 문제 1). 다음으로, 선행 단계에서 분석된 웹툰 소비자들을 대상으로 토픽 모형을 적용하여 웹툰 소비를 유형화하고자 한다(연구 문제 2). 마지막으로, 로지스틱 회귀분석 및 머신러닝·인공지능 기반 예측 모형을 활용하여 각 소비 유형의 시장성을 계량적으로 검증하고자 한다(연구 문제 3). 이처럼 본 연구는 전통적 통계 기법부터 머신러닝 기반의 예측 모델까지 아우르는 분석 체계를 통해, 각 분석이 서로 유기적으로 연결되는 설계를 도모하였다. 연구 설계의 전체적인 흐름은 [그림 1]에 제시하였다.

본 연구는 이러한 탐색-분류-검증의 3단계 분석을 통해, 한국 웹툰의 해외 소비 특성과 시장성에 대한 실증적 통찰을 제시하고자 한다. 분석 과정에서 해외 소비자들이 직접 작성한 댓글 데이터를 활용한다는 점에서 차별화되며, 이를 바탕으로 기존 연구들이 간과한 소비자 관점의 통찰을 제공하고자 한다. 나아가, 웹툰의 독자층을 한류 팬으로 한

[그림 1] 연구 흐름도



정하기보다는, 웹툰을 독립적인 문화 콘텐츠로 간주하는 시각에서 현존 논의를 확장하고, 글로벌 시장 확장을 위한 실효성 있는 전략적 시사점을 도출하고자 한다.

### Ⅲ. 실증 연구

#### 1. 글로벌 웹툰 소비 현황 분석

##### 1) 한류 콘텐츠 소비 경험과 웹툰 소비의 관계

###### (1) 연구 방법: 로지스틱 회귀분석

한 개의 종속변수와 여러 개의 독립변수 간의 관계를 검증하는 데 가장 널리 사용되는 방법은 회귀분석이다. 그러나 일반적인 회귀분석은 종속변수가 독립변수에 의해 선형적으로 변화한다고 가정하므로, 이항 사건을 분석하는 데 적절하지 않다. 이러한 경우, 로지스틱 회귀분석이 유용한 대안이 될 수 있다. 로지스틱 회귀분석은 종속변수가 이항 변수일 때 독립변수와의 관계를 분석하는 통계 기법으로, 독립변수를 바탕으로 종속변수가 특정 범주에 속할 확률을 예측하며, 그 결과는 로짓(logit) 함수로 표현된다(Peng et al., 2002; Stoltzfus, 2011). 즉, 이진형 종속변수와 독립변수 간의 관계를 분석하고 통계적 유의성을 검증하는 데 효과적이다. 본 연구는 한류 팬 데이터를 기반으로 로지스틱 회귀분석을 수행하여, 한류 콘텐츠에 대한 다양한 소비 경험이 한국 웹툰 이용에 미치는 영향을 검토하고자 한다. 이를 통해 한국 웹툰 소비가 어떠한 콘텐츠 소비와 높은 관련성을 갖는지 실증적으로 규명하고자 한다(연구 문제 1).

###### (2) 연구 자료 및 변수 설정

본 연구는 2024년 3월 한국국제문화교류진흥원에서 진행한 ‘2024 해외한류실태 조사’ 데이터를 사용하였으며, 이 데이터는 26개국 15~59세 인구를 대상으로 한 25,000개의 표본으로 구성되어 있다. 본 연구는 다양한 한류 콘텐츠의 이용 경험이 웹툰 소비에 미치는 영향을 파악하기 위해, 웹툰과 유사성이 높은 서사 중심의 시각 콘텐츠인 드라마, 영화, 출판물, 애니메이션의 이용 경험을 주요 변수로 선정하고 로지스틱 회귀분석을 수행하였다. 구체적으로, ‘드라마 이용 경험’, ‘영화 이용 경험’, ‘출판물 이용 경험’ 그리고 ‘애니메이션 이용 경험’을 독립변수로, ‘웹툰 이용 경험’을 종속변수로

설정하였다. 변수들은 이용 경험이 있는 경우 1, 없는 경우 2의 값을 가지며, 정제된 데이터를 활용하였기에 별도의 전처리 과정은 생략하였다.

### (3) 연구 결과

로지스틱 회귀분석 진행 결과, <표 1>과 같이 모든 독립변수가 통계적으로 유의하였다. 즉, 한류 콘텐츠 소비 경험은 한국 웹툰 소비 경험을 진작시키는 것으로 나타났다. 특히, 드라마와 영화보다 출판물과 애니메이션 이용 경험이 해외 소비자의 한국 웹툰 이용 경험에 더욱 큰 영향을 미치는 것으로 나타났다. 한류 콘텐츠의 대중적인 유형으로 여겨지는 드라마와 영화보다 출판물, 애니메이션과 같은 콘텐츠 유형이 웹툰 이용 경험에 더 높은 관련성을 가진다는 점이 흥미로운데, 이는 해당 콘텐츠와 웹툰의 장르적 유사성에 기인하는 것으로 추론할 수 있다.

<표 1> 로지스틱 회귀분석 결과

변수명	$\beta$	Exp( $\beta$ )	p-value
드라마 이용 경험	0.732	2.080	<0.001
영화 이용 경험	0.597	1.817	<0.001
출판물 이용 경험	1.391	4.021	<0.001
애니메이션 이용 경험	1.067	2.907	<0.001
상수항	-2.725	0.066	<0.001

## 2) 웹툰과 글로벌 문화 콘텐츠의 연관 구조 분석

### (1) 연구 방법: 네트워크 분석

네트워크 분석은 네트워크의 구조적 특징을 도출하고 이를 계량적으로 분석하는 방법론으로, 노드(node) 간의 엣지(edge)를 시각적으로 표현하여 전체 네트워크의 구조를 파악할 수 있다(문진희 · 김영정, 2016). 네트워크의 구조적 특징을 살펴보기 위한 주요 지표로는 연결 정도(degree), 밀도(density), 중심성(centrality) 등이 있다. 연결 정도는 특정 노드가 관계 맺고 있는 다른 노드의 개수를 의미하며, 연결 정도가 높을수록 정보의 흐름에서 중심적인 역할을 수행한다(최정아, 2023). 밀도는 네트워크 내에서 형성될 수 있는 총관계 수 대비 실제로 형성된 관계 수의 비율을 나타내며(최정아,

2023), 밀도가 높은 네트워크일수록 정보의 확산이 빠르게 이루어진다(고세욱 등, 2018). 중심성은 네트워크에서 권력과 영향력을 측정하는 중요 지표로, 특성에 따라 연결 중심성(degree centrality), 매개 중심성(betweenness centrality), 근접 중심성(closeness centrality) 등으로 구분하여 분석할 수 있다(신용욱 · 박상혁, 2013). 연결 중심성은 특정 노드가 다른 노드들과 연결된 정도를 의미하며(우연주 등, 2013), 연결 중심성이 높을수록 다른 노드와 많은 관계를 맺고 있어 정보 또는 자원을 얻을 가능성이 크다(김용학, 2004). 다음으로, 매개 중심성은 특정 노드가 네트워크 내의 다른 노드 간에 위치하는 정도로서, 해당 노드가 매개자 또는 중개자 역할을 하는 정도를 의미한다(고세욱 등, 2018). 특정 노드가 두 개의 노드 간 최단 경로에 위치할 경우, 그 노드는 정보 전달 과정에서 중요한 영향력을 가지게 된다. 매개 중심성이 높은 노드는 네트워크 내에서 핵심적인 브로커 역할을 수행한다(김용학, 2004; 손동원, 2002). 마지막으로, 근접 중심성은 네트워크 내에서 특정 노드가 다른 모든 노드와 맺고 있는 최단 거리의 합으로 정의된다(고세욱 등, 2018). 근접 중심성이 높을수록 네트워크 내에서 다른 노드들과 가까이 위치하고 있으며, 이는 정보 권력, 영향력, 지위에 대한 확보와 접근성이 뛰어남을 의미한다(손동원, 2002). 본 연구는 한류 팬을 대상으로 한 앞선 로지스틱 회귀분석을 확장하여, 한류 콘텐츠 소비가 전제되지 않은 일반 해외 소비자를 대상으로 네트워크 분석을 수행하고자 한다. 이를 통해 한국 웹툰이 글로벌 문화 콘텐츠 중 어떤 범주와 깊은 연관성을 가지는지를 추가적으로 규명하고자 한다(연구 문제 1).

## (2) 데이터 설명 및 전처리 과정

본 연구는 미국의 대표적인 온라인 커뮤니티 플랫폼인 Reddit에서 웹툰 관련 게시판(서브레딧) 간의 이용 행태를 네트워크 분석 기법으로 살펴봄으로써, 게시판 간 연관 구조와 이용자 이동 패턴을 규명하고자 하였다. Reddit은 하루 평균 약 7,310만 명이 방문하는 대규모 다국적 커뮤니티로서(오현우, 2024), 다양한 국가의 이용자를 포괄한다는 점에서 웹툰 소비자 활동을 탐색하기에 적절한 분석 환경을 제공한다.

우선, 2025년 7월 10일 기준 webtoons 게시판의 최신 게시물 1,000건을 수집한 뒤, 해당 게시물에 글 또는 댓글을 남긴 이용자 계정을 전수 확보하였다. 이 중 webtoons 게시판에서 누적 글 · 댓글 활동이 100회 이상인 계정을 '한국 웹툰 해외 소비자'로 조작적으로 정의하고, 이들의 게시판 이동 행태를 웹툰 이용자의 관심 콘텐츠

영역 및 소비 성향을 추정하기 위한 대리변수(proxy variable)로 활용하였다. 이는 게시판 이용 행태를 관심·인식·행동의 대리변수로 해석한 선행연구들(Olson & Neal, 2015; Segneri et al., 2024)과 방법론적 궤를 같이한다.

이후, webtoons 게시판에서 100회 이상 활동한 이용자들이 동일한 수준(100회 이상)의 활동을 보인 타 게시판 데이터를 선별하고, 이를 기반으로 네트워크 시각화 기법을 적용하여 게시판 간 연관 구조를 도출하였다. 이를 통해 한국 웹툰 해외 소비자가 웹툰 외에 어떠한 콘텐츠 유형에 관심을 보이는지, 전반적 콘텐츠 소비 성향을 규명하고자 하였다. 시각화에서는 글·댓글 활동량이 가장 많은 30개 게시판을 노드로 선별하였으며, 노드의 크기는 연결 중심성을, 엣지(연결)의 굵기는 두 게시판 간 겹치는 사용자 수를 나타낸다. 주요 게시판에 대한 설명과 데이터의 기술통계량은 각각 <표 2> 및 <표 3>에 제시하였다.

<표 2> 글·댓글 활동이 가장 많은 30개 게시판 목록과 설명

게시판(서브레딧)	설명
webtoons	웹툰 전반에 대한 추천, 감상, 토론이 이루어지는 커뮤니티
WebtoonCanvas	인디 혹은 아마추어 작가들의 웹툰에 대한 커뮤니티
Otomelsekai	여성향 로맨스·이세계물 장르의 소설을 중심으로 형성된 팬 커뮤니티
AskReddit	다양한 주제에 대해 자유로운 질문과 토론이 이루어지는 커뮤니티
lookismcomic	웹툰 《외모지상주의(Lookism)》를 중심으로 한 팬 커뮤니티
comics	미국식 전통 만화(마블, DC 등)에 관한 정보 공유 및 토론 커뮤니티
OriginalCharacter	직접 창작한 캐릭터(Original Character)를 소개하고 공유하는 커뮤니티
teenagers	10대 이용자들이 학교, 연애, 일상 등을 주제로 교류하는 커뮤니티
webtoon	웹툰 전반을 다루는 일반적인 토론 커뮤니티
anime	일본 애니메이션에 대한 감상, 추천, 토론이 활발히 이루어지는 커뮤니티
manhwa	한국 만화 및 웹툰에 대한 관심을 중심으로 형성된 해외 이용자 커뮤니티
Genshin_Impact	게임 《원신(Genshin Impact)》 커뮤니티
AmltheAsshole	도덕적 딜레마 상황을 공유하고 정당성에 대한 판단을 받는 커뮤니티
webcomics	서양권 웹코믹 작품을 중심으로 창작물 및 감상평이 공유되는 커뮤니티

<표 계속>

게시판(서브레딧)	설명
OmniscientReader	웹소설 <전지적 독자 시점>에 관한 팬 커뮤니티
LoveAndDeepspace	SF 연애 시뮬레이션 게임 <Love and Deepspace>에 대한 팬 커뮤니티
animequestions	애니메이션 관련 정보 교환이 이루어지는 질의응답형 커뮤니티
AO3	팬픽션 플랫폼 AO3에 대한 추천, 작품 공유 등이 이루어지는 커뮤니티
manga	일본 만화 전반에 대해 토론, 추천, 발매 정보 등이 공유되는 커뮤니티
shoujo	소녀 만화 장르에 특화된 일본 만화 · 애니메이션 중심 팬 커뮤니티
writing	창작 글쓰기와 관련된 조언, 피드백이 공유되는 글쓰기 중심 커뮤니티
HonkaiStarRail	모바일 RPG 게임 <붕괴: 스타레일> 관련 팬 커뮤니티
CharacterAI	AI 캐릭터와의 상호작용을 중심으로 시나리오 등을 공유하는 커뮤니티
AITA	AmItheAsshole와 유사한 도덕적 판단 사례가 공유되는 커뮤니티
HungryArtists	일러스트 등 시각 창작물을 공유하고 피드백을 주고받는 창작자 커뮤니티
NoStupidQuestions	간단하지만 솔직한 질문을 자유롭게 나누는 질의응답 커뮤니티
Fantasy	판타지 장르 전반(문학, 게임, 영화 등)을 주제로 논의하는 커뮤니티
CookieRunKingdoms	모바일 게임 <쿠키런: 킹덤> 관련 팬 커뮤니티
artcommissions	일러스트 외주 작업을 주제로 한 커뮤니티
cartoons	미국식 카툰에 대한 감상 및 추천이 이루어지는 커뮤니티

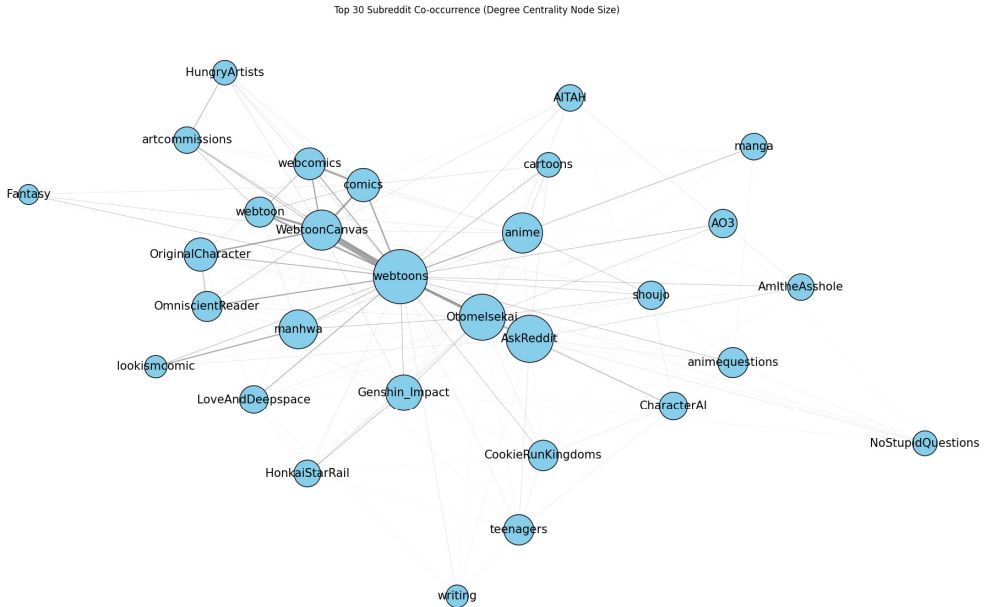
<표 3> 기술통계량

변수	N	고유값	최빈도	평균	표준편차	최솟값	중앙값	최댓값
사용자	105,129	2,095	1,635	N/A	N/A	N/A	N/A	N/A
게시판	105,129	23,139	1,762	N/A	N/A	N/A	N/A	N/A
글 · 댓글 활동	105,129	N/A	N/A	10.14	42.43	1	2	1,755

### (3) 연구 결과

[그림 2]의 네트워크 시각화 결과는 한국 웹툰 해외 소비자들이 주로 어떤 커뮤니티와 연관되어 있는지를 보여준다. 우선, 'AskReddit' 등 Reddit 전체 이용자가 범용적으로 사용하는 게시판을 제외하고 살펴보면, webtoons 노드는 동아시아 만화 · 애니메이션 권역, 서구 코믹스 · 그래픽노블 권역, 독립 창작 · 팬아트 권역이 중첩되는 지점에 위치해 있다. 특히, 웹툰 작가나 특정 작품에 기반한 커뮤니티(예: 'WebtoonCanvas',

[그림 2] Reddit 내 webtoons 게시판 기반 네트워크 시각화



‘lookismcomic’)와 가장 밀접하게 연결되어 있으며, ‘OtomeIsekai’, ‘manhwa’, ‘manga’, ‘anime’ 등 동아시아 만화·애니메이션 커뮤니티와 ‘comics’, ‘webcomics’, ‘cartoons’ 등 서구 코믹스 커뮤니티가 webtoons 노드를 중심으로 양측에 반원형으로 분포해, 웹툰을 매개로 두 문화권이 교차 결합하는 양상이 관찰되었다. 이와 더불어, ‘Genshin\_Impact’, ‘HonkaiStarRail’, ‘CookieRunKingdoms’ 등 게임 커뮤니티와 ‘OriginalCharacter’, ‘AO3’, ‘writing’ 등 창작·스토리텔링 커뮤니티도 webtoons 노드와 인접해 있어, 웹툰 소비자가 게임·독립 창작·2차 창작 생태계와도 활발히 상호작용하고 있음을 나타낸다.

반면, ‘kpop’, ‘kdramas’, ‘kpopthoughts’, ‘kdramarecommends’ 등 한류 관련 커뮤니티는 원 데이터에는 포함되었으나, 이들 노드가 webtoons 노드와 공유하는 이용자의 활동량이 임계치에 미달하여 최종 네트워크 시각화에서는 표현되지 않았다. 이들 노드의 연결 중심성과 webtoons 노드와의 연결 강도 역시 낮은 수준이었다. 이러한 결과는 한국 웹툰의 글로벌 확장 전략이 기존의 한류 팬층에만 의존하기보다는, 만화·그래픽노블 애호가 및 창작 커뮤니티 전반을 아우르는 다층적이고 통섭적인 접근을 요구함을 시사한다.

결론적으로, 네트워크 분석 결과는 한국 웹툰 해외 소비자들이 K-pop이나 K-drama와 같은 대표적인 한류 콘텐츠보다 만화·코믹스 기반 출판물과 더 밀접한 관련성을 맺고 있음을 보여주며, 앞서 진행한 로지스틱 회귀분석의 결과를 확증한다. 이에 따라, 웹툰의 글로벌 시장 확장을 위해서는 한류 팬층에 국한된 타깃 전략보다는, 보다 폭넓은 만화 독자층을 포괄하는 전략 수립이 필요하다. 다음 단계에서는 한국 웹툰의 해외 소비자가 작성한 댓글 데이터를 바탕으로 토픽 모형 분석을 수행하여, 글로벌 소비자 유치 전략을 구체화하고자 한다.

## 2. 한국 웹툰 해외 소비자의 소비 유형 분석

### 1) 연구 방법: 토픽 모형 분석

텍스트 마이닝(text mining)은 신문 기사, 리뷰, 인터뷰 텍스트, 소셜 미디어 게시글 등의 비정형 데이터를 분석하는 기법으로 텍스트 데이터 내의 유용한 정보를 추출하는 과정에서 연구자의 주관은 최대한 배제하고, 통계적 방법론을 통해 의미 패턴과 구조를 도출하는 기법이다(이삼열 등, 2016). 텍스트 마이닝 기법의 일종인 토픽 모형 분석(topic modeling)은 텍스트 문서를 다수의 잠재 토픽이 공존하는 혼합체로 보고, 이러한 잠재 토픽의 할당은 단어들의 확률 분포에 근거하여 결정된다는 개념에서 출발한다(오정심, 2021; Steyvers & Griffiths, 2007). 초기에는 LSA(latent semantic analysis), pLSA(probabilistic latent semantic analysis)등의 알고리즘이 제시되어 왔으나(Blei et al., 2003; Hofmann, 1999), 본 연구는 최근 널리 활용되고 있는 LDA(latent Dirichlet allocation)를 분석 모형으로 채택하였다.

LDA는 블레이 등(Blei et al., 2003)이 고안한 모형으로 디리클레 분포를 활용하여 문서 내의 토픽의 분포와 개별 토픽의 단어 분포를 모형화하는 방식이다(오민정, 2020). 텍스트 데이터 내에 내재된 주제인 토픽을 파악하고 각 문서가 어떤 토픽들을 어느 정도 포함하고 있는지를 텍스트 데이터의 문맥, 문장 구조 등을 통해 추론한다. 이를 통해 전체 문서 집합의 토픽 및 개별 문서의 토픽 비율과 각 토픽에 속한 단어들의 확률을 도출할 수 있다(김혜진·이명구, 2021; Blei et al., 2003). LDA 모형을 추론하는데 있어 라플라스 근사(Laplace approximation), SVI(stochastic variational inference) 등의 방법이 있으며, 본 연구에서는 변분 추론(variational inference) 방법을 사용하였다. 본 연구는 LDA 모형을 통해 해외 소비자들이 한국 웹툰을 소비하는

이유를 파악하고, 이를 기반으로 웹툰의 소비 유형을 분류하고자 한다(연구 문제 2).

## 2) 데이터 설명 및 전처리 과정

한국 웹툰 해외 소비자들이 Reddit에 직접 작성한 댓글 데이터를 대상으로 토픽 모형 분석을 진행하였다. 긍정적인 리뷰는 특정 콘텐츠를 통해 얻고자 하는 추구 효용을 반영한다는 점에 착안하여(김종대 · 이유석, 2019), 2024년 6월 20일 기준 webtoons sub-Reddit에서 ‘좋아하는 웹툰과 그 이유’ 게시물에 작성된 모든 댓글 338개를 Reddit 개발자 API를 활용하여 수집하였다. 수집된 데이터는 웹툰 단위로 분리하여 각 행에 단일 웹툰 정보만 포함되도록 전처리하였다. 이후, Python의 자연어 처리 라이브러리인 ‘NLTK’와 ‘spaCy’를 활용하여 텍스트를 토큰화하고, 불용어, 이모티콘, 문장 부호를 제거하였다. 또한, 분석의 일반화를 위해 특정 웹툰에만 국한된 단어와 전체 문장 중 15% 미만 출현 빈도를 보이는 단어를 제외하였다.

## 3) 연구 결과

토픽 모형 분석 실행 시 연구자는 토픽의 개수  $K$ 를 사전에 지정해야 하며, 토픽 수 결정을 위한 다양한 방법론이 존재한다. 토픽 간 혼잡도 분석, 토픽 응집도 평가 등이 대표적이다. 그러나 이러한 정량적 방법들은 전처리 과정의 잠재적 오류와 알고리즘의 확률적 특성을 고려해야 하는 한계를 지닌다. 이에 따라 연구자가 데이터의 맥락과 주제를 직접 분석하여 토픽 수를 결정하는 정성적 접근이 보다 효과적인 것으로 평가되며, 다수의 선행연구에서도 이러한 방식이 채택되고 있다(남춘호, 2016; DiMaggio et al., 2013). 따라서 본 연구 역시 추출된 토픽별 단어들의 맥락적 연관성을 연구자가 직접 분석하고, 반복적인 검토와 수정 과정을 통해 최적의 토픽 수를 5개로 도출하였다.

토픽의 수를 5개로 설정하고 LDA 모형으로 분석한 결과는 <표 4>와 같다. 본 분석에서는 토픽 모형 결과를 통해 각 유형 소비자의 성향을 추론하고, 토픽을 명명하였다. 이후, 토픽별 키워드를 통해 각 유형의 소비자가 웹툰을 소비하며 얻고자 하는 효용을 해석하였다. 이러한 과정을 통해 글로벌 웹툰 소비의 다양한 측면을 이해하고 분석 결과의 신뢰성을 확보하고자 하였다.

첫 번째 유형(토픽 1)의 소비자들은 갈등과 그 해결 과정에서 도출되는 교훈적 가치(lesson)를 중시한다. 이들은 주인공이 직면한 외재적 갈등과 그 극복 과정에서 나타나

〈표 4〉 토픽별 키워드

토픽 구분	1	2	3	4	5
토픽 1	world	people	lesson	heart	ending
토픽 2	story	art	style	author	ending
토픽 3	character	story	day	game	thriller
토픽 4	romance	year	life	fantasy	comedy
토픽 5	action	reader	genre	comedy	people

는 인격적 성장을 중요한 평가 요소로 고려한다(world). 특히, 등장인물들이 역경을 통해 자아를 발견하고 타인과의 관계를 회복해 가는 과정에서 독자들이 공감할 수 있는 삶의 교훈을 제시하는 작품들을 선호한다(people, heart). 이에 근거하여 첫 번째 유형을 ‘삶-교훈 중심형’으로 명명하였다.

두 번째 유형(토픽 2)의 소비자들은 웹툰의 시각적 완성도를 핵심 평가 기준으로 활용한다. 이들은 작가의 독창적인 그림체, 효과적인 컷 분할, 채색의 조화 등 작품의 미학적 요소들을 중점적으로 평가한다(author, style, art). 또한, 이러한 요소들이 서사 전달력을 높이는 데 얼마나 기여하였는지를 중요하게 고려한다. 이를 기반으로 두 번째 유형을 ‘시각요소 중심형’으로 명명하였다.

세 번째 유형(토픽 3)의 소비자들은 캐릭터의 입체적 묘사와 성장 서사에 주목한다(character, story). 이들은 등장인물의 과거 서사, 내면의 갈등, 가치관의 변화 등 캐릭터 서사의 깊이를 중시하며, 캐릭터의 결정과 행동에 대한 명확한 동기 부여, 일관된 성격 묘사, 개성 있는 대사 등을 통해 캐릭터의 진정성이 얼마나 잘 구현되었는지를 평가한다. 이러한 성향을 고려하여 세 번째 유형의 소비자들을 ‘캐릭터 중심형’으로 명명하였다.

네 번째 유형(토픽 4)의 소비자들은 다양한 형태의 로맨스 서사를 선호한다(romance). 이들은 전통적인 러브스토리의 매력뿐만 아니라, 판타지 세계관이 결합된 로맨스 판타지, 일상적 소재에 유머를 가미한 로맨스 코미디, 미스터리나 스릴러 요소가 가미된 장르 혼합형 로맨스 등 다채로운 로맨스 서사를 긍정적으로 평가한다(fantasy, comedy). 그중에서도 남녀 주인공 간의 관계 발전 과정, 갈등 해결 방식, 감정선의 자연스러운 전개 등을 중요한 평가 요소로 고려한다. 이에 근거하여 네 번째 유형을 ‘로맨스 중심형’으로 명명하였다.

다섯 번째 유형(토픽 5)의 소비자들은 장르 간 혁신적 결합을 통한 신선한 서사 경험을 추구하는 유형의 소비자들이다. 이들은 판타지와 무협, SF와 로맨스, 호러와 코미디 등 이질적 장르 간의 실험적 결합을 통해 만들어지는 예측 불가능한 전개를 높이 평가한다(genre). 또한, 작품의 복잡한 세계관과 다층적 서사 구조를 해석하고 예측하는 과정에서 독자들 간의 활발한 토론과 이론 제기가 이루어지며, 이러한 집단적 향유 방식이 작품의 재미 요소로 작용한다. 장르 간 크로스오버를 통해 기존 장르의 한계를 뛰어넘는 새로운 서사적 가능성을 제시하는 작품들을 특히 선호하는 경향을 보이기도 한다. 이에 따라 다섯 번째 유형 소비자들을 ‘복합 장르형’으로 명명하였다.

### 3. 소비 유형별 흥행 성과 분석

#### 1) 소비 유형별 구독자 수 회귀분석

##### (1) 데이터 수집 및 변수 설정

앞서 토픽 모형 분석을 통해 도출한 웹툰 소비 유형이 실제 웹툰의 흥행 성과에 어떤 영향을 미치는지 검증하기 위해 로지스틱 회귀분석을 실시하였다. 이를 위해 WEBTOON에서 연재 중이거나 완결된 웹툰을 대상으로 데이터를 수집한 뒤, 데이터 수집일인 2024년 6월 30일 기준으로 에피소드 수가 10개 이하인 작품을 제외하고, 구독자 수 기준 이상치를 제거하여 최종적으로 663개의 웹툰을 분석 대상으로 선정하였다. 다음으로, Reddit 댓글을 기반으로 학습한 LDA 모형을 각 웹툰의 최신 10개 에피소드에서 최다 추천을 받은 댓글에 적용하여, 해당 웹툰의 주요 소비 유형(dominant topic)을 추출하였다. 이는 동일한 주제 영역에서는 토픽 구조를 재학습하지 않고도 새로운 문서 집합에 적용할 수 있다는 LDA의 이론적 특성에 근거한 것이다(Blei et al., 2003). 실제로 유사한 도메인 간에, 학습된 토픽 구조를 서로 다른 출처·언어·규모의 데이터에 성공적으로 적용한 다수의 선행연구가 존재하며(Glenny et al., 2019; Yao et al., 2009), 본 연구 역시 이러한 방법론적 타당성에 기반해 분석을 수행하였다.

토픽 모형 분석에서 도출된 소비 유형을 본 회귀분석의 독립변수로 설정하였으며, 각 웹툰은 ‘삶-교훈 중심형’, ‘시각요소 중심형’, ‘캐릭터 중심형’, ‘로맨스 중심형’, ‘복합 장르형’ 중 하나로 분류되었다. 종속변수는 구독자 수로 설정하여, 실질적인 소비 가능성이 있는 독자층을 반영하고자 하였다. 이와 같은 분석 설계는 리뷰 텍스트에 토픽 모형을 적용하여 도출된 토픽 비율을 바탕으로, 제품의 매출이나 평점 등을 예측한 기존

연구들과 분석 방향을 같이한다(Büschken & Allenby, 2016; Li et al., 2019). 즉, 소비자 행동을 예측하기 위해 기구축된 토픽을 독립변수로 활용한 기존 연구를 참고하여, 웹툰 소비 맥락에서 각 토픽의 시장성을 탐색하고자 하였다.

또한, 에피소드 수, 평균 평점, 완결 여부, 작가 전작 수, 조회수 천만 돌파 여부를 웹툰 성과의 간접 지표로 간주하여 통제변수로 설정하였다. 에피소드 수는 연재 회차를 의미하며, 회차가 늘어날수록 완독률이 급감한다는 선행연구를 반영하였다(윤기현 등, 2015). 평균 평점은 이용자들의 품질 인식을 대변하는 지표로, 흥행指数와 강한 양(+)의 관계를 보인다는 연구 결과에 따라 포함하였다(양지훈 등, 2016). 완결 여부는 작품이 완결(1) 또는 연재 중(0) 상태인지 구분하는 변수로, 완결된 작품이 독자 만족도에서 우위를 보인다는 실증 결과를 반영하였다(Mun, 2025). 작가 전작 수는 동일 작가가 이전에 연재한 작품 수를 의미하며, 작가 브랜드 자본이 흥행에 미치는 영향을 통제하기 위해 포함하였다(양지훈 등, 2016). 조회수 천만 돌파 여부는 천만 이상(1) 또는 미만(0)으로 구분되며, 플랫폼과 업계가 흥행 기준으로 활용하는 ‘사회적 증거 효과’에 기반하여, 작품 간 초기 노출 편차를 보정하기 위한 목적에서 설정하였다(김휘정, 2015; Park et al., 2019). 통제변수 중 에피소드 수, 평균 평점, 작가 전작 수와 같은 수치형 변수들은 Min-Max Scaling을 통해 0과 1 사이의 값으로 정규화하였으며, 이를 통해 독립변수가 시장 성과에 미치는 순수 효과를 보다 정확하게 추정하고자 하였다. 이와 같은 분석 설계를 통해 웹툰 소비 유형별 구독자 수 차이를 계량적으로 검증하고, 각 유형의 시장성을 실증적으로 평가하고자 한다(연구 문제 3).

## (2) 연구 결과

다섯 가지 소비 유형(토픽)이 웹툰의 글로벌 흥행에 미치는 영향을 분석하기 위하여, 소비 유형을 독립변수로, 구독자 수를 종속변수로 두고 회귀분석을 실시하였다. 다섯 가지 유형 중 첫 번째 유형인 ‘삶-교훈 중심형’을 기준 범주로 두고 분석한 결과는 <표 5>와 같다.

〈표 5〉 회귀분석 결과

변수	<i>t</i>	Wald	<i>p</i> -value
시각요소 중심형	2.528	6.389	<0.05
캐릭터 중심형	1.339	1.794	>0.1
로맨스 중심형	1.680	2.823	<0.1
복합 장르형	2.413	5.826	<0.05
에피소드 수	-3.958	16.287	<0.001
평균 평점	3.794	14.396	<0.001
완결 여부	1.255	1.574	>0.05
작가 전작 수	-0.499	0.202	>0.1
조회수 천만 돌파 여부	23.206	540.540	<0.001
Constant	-1.276	1.628	>0.1
$R^2$ (Adj. $R^2$ )	0.515(0.509)		
<i>F</i>	77.14		
Durbin-Watson	1.759		

분석 결과, 웹툰 소비 유형 중 ‘시각요소 중심형’과 ‘복합 장르형’이 구독자 수에 유의미한 정(+)의 영향을 미치는 것으로 나타났다. 구체적으로, 이 두 유형으로 분류된 웹툰은 다른 유형에 비해 통계적으로 유의하게 높은 구독자 수를 기록하였다. ‘로맨스 중심형’의 경우 제한적이거나 긍정적인 영향을 보였다. 한편, 통제변수 가운데 에피소드 수, 평균 평점, 조회수 천만 돌파 여부 역시 구독자 수에 유의미한 영향을 미치는 것으로 확인되었다. 즉, 에피소드 수가 적을수록, 평균 평점이 높을수록, 그리고 총 조회수가 천만 이상일 경우, 해당 웹툰의 구독자 수가 증가하는 경향을 보였다.

## 2) 머신러닝 기법을 활용한 추가 검증

### (1) 연구 방법: 랜덤 포레스트

머신러닝은 지도 학습과 비지도 학습으로 구분되는데, 본 연구에서 사용한 랜덤 포레스트(random forest)는 지도 학습에 해당하며, 대표적인 앙상블 학습(ensemble learning) 모형이다. 랜덤 포레스트는 브리만(Breiman, 2001)이 제안한 모형으로, 양

상블 기법 중 하나인 배깅(bagging; bootstrap aggregating)을 통해 다수의 의사결정 나무를 생성한 뒤 종합한 모형이다. 의사결정나무(decision tree)란 종속변수를 효과적으로 설명하는 독립변수를 기준으로 가지(branch)를 확장하는 기법이다. 연속형, 범주형 모두 설명변수로 활용할 수 있으며, 그 결과를 시각적으로 표현할 수 있어 모형을 직관적으로 이해하는 데 유용한 장점이 있다(Siroky, 2009). 그러나 의사결정나무는 예측력이 낮고, 과적합과 데이터 변형에 따른 모형 변화의 위험이 존재한다(유진은, 2015). 이를 보완하기 위해 앙상블 학습 기법인 배깅으로 다수의 표본을 추출하여 의사결정나무를 생성한 후, 각 모형의 예측 결과를 종합하여 성능을 향상시킨 모형이 바로 랜덤 포레스트이다. 랜덤 포레스트는 의사결정나무들 간의 상관관계를 낮춤으로써 예측 성능을 향상시키고, 과적합 문제를 줄여 적은 데이터에서도 효과적으로 활용될 수 있다(Breiman, 2001). 본 연구는 토픽 모형 분석을 통해 분류한 웹툰의 소비 유형이 웹툰의 흥행 성과에 미치는 영향을 추가 검증하기 위해, 랜덤 포레스트와 SHAP을 활용하여 변수 중요도 분석을 수행하고자 한다. 이러한 과정을 통해 앞서 회귀분석에서 얻은 결과를 확증하고, 각 유형의 시장성을 재검증하고자 한다(연구 문제 3).

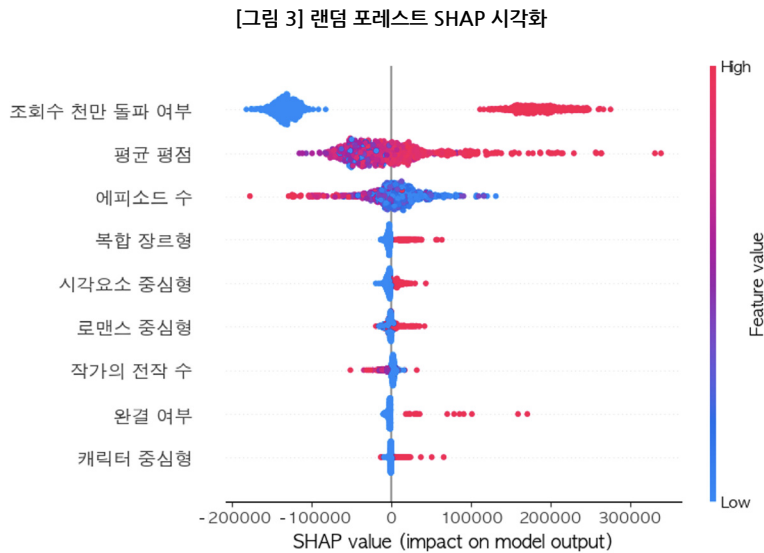
## (2) 모형 설계 및 변수 설정

글로벌 웹툰 소비 유형 중 ‘시각요소 중심형’ 및 ‘복합 장르형’이 웹툰의 구독자 수에 미치는 영향의 방향과 크기가 다른 방법론에서도 여전히 유의한지 확인하기 위해, 머신러닝 기법인 랜덤 포레스트와 인공지능 기법인 SHAP(Shapley additive explanations)을 활용하여 추가 검증을 진행하였다(김주영 · 조영신, 2023).

먼저, 회귀분석과 동일한 데이터 및 변수를 활용하여 한국 웹툰의 해외 구독자 수를 예측하는 머신러닝 모형을 구축하였다. 이를 위해 랜덤 포레스트 기법을 적용하였으며, 모형의 성능을 최적화하기 위해 주요 하이퍼파라미터를 다음과 같이 설정하였다. 랜덤 포레스트 모형의 트리 개수( $n_{estimators}$ )는 1,000으로 설정하여 충분한 학습이 이루어지도록 하였으며, 트리의 최대 깊이( $max\_depth$ )는 10으로 제한하여 과적합(overfitting)을 방지하고 일반화 성능을 확보하고자 하였다. 또한, 데이터 분할 시 훈련 데이터 셋과 테스트 데이터 셋의 비율을 7:3으로 설정하여, 모형이 전체 데이터의 다양한 패턴을 학습할 수 있도록 하였다.

### (3) 연구 결과

구독자 수를 예측하는 랜덤 포레스트 모형을 구축하고, 이를 기반으로 독립변수가 예측값에 미치는 영향을 분석하는 SHAP 시각화를 수행하였다. 결과는 [그림 3]에 제시되어 있다. SHAP 시각화에서는 변수의 중요도가 세로축 기준으로 위쪽에 위치할수록 높으며, 가로축을 기준으로 붉은색이 오른쪽에 위치할 경우 양(+)의 상관관계를, 왼쪽에 위치할 경우 음(-)의 상관관계를 갖는 것으로 해석된다.



분석 결과, 변수의 중요도와 영향력의 방향이 회귀분석 결과와 상당 부분 일치하는 것으로 나타났다. 구체적으로, 회귀분석에서 통계적으로 유의했던 조회수 천만 돌파 여부, 평균 평점, 에피소드 수, ‘복합 장르형’, ‘시각요소 중심형’, 그리고 ‘로맨스 중심형’ 변수의 중요도가 높게 나타났다. 또한, 조회수가 천만 이상인 경우와 에피소드 수가 적은 경우 구독자 수가 증가하는 경향을 보였으며, 소비 유형에 따라 ‘복합 장르형’, ‘시각요소 중심형’, ‘로맨스 중심형’으로 분류된 웹툰이 그렇지 않은 웹툰보다 높은 구독자 수를 기록하는 것을 재확인할 수 있었다.<sup>2)</sup>

2) 과적합 가능성을 완화하기 위하여 전체 자료 633건 중 약 15%(n≈95)를 무작위 추출한 뒤, 이를 다시 학습:평가=7:3으로 분할하여 추가 분석을 수행하였다. 학습에 포함되지 않은 538건에 대하여 SHAP 값을 산출한 결과 [그림 3]의 변수 중요도 및 패턴을 재현하였다.

종합적으로, 회귀분석과 머신러닝 기법을 통해 웹툰의 소비 유형별 글로벌 흥행 성과를 분석한 결과, ‘시각요소 중심형’과 ‘복합 장르형’ 소비 유형이 공통적으로 구독자 수에 유의미한 영향을 미치는 것으로 나타났다. 구체적으로, ‘시각요소 중심형’ 및 ‘복합 장르형’으로 분류된 웹툰은 그렇지 않은 웹툰에 비해 글로벌 구독자 수가 유의하게 높은 것으로 확인되었다. 이는 두 소비 유형이 글로벌 시장에서 긍정적인 소비자 반응을 이끌고 있으며, 한국 웹툰의 글로벌 흥행을 견인하는 주요 요인으로 작용할 수 있음을 시사한다.

#### IV. 결론

본 연구는 한국 웹툰의 글로벌 시장 확장을 위한 전략적 방향을 모색하고자, 한국 웹툰의 해외 소비자 데이터를 활용하여 단계적 분석을 수행하였다. 먼저, 한국 웹툰의 해외 소비 현황을 파악하기 위해 로지스틱 회귀분석을 실시한 결과, 외국인 한류 팬의 한국 출판물 및 애니메이션 이용 경험이 한국 웹툰 소비에 긍정적인 영향을 미치는 것으로 나타났다. 나아가, 한류 콘텐츠 소비를 전제로 하지 않는 일반 해외 소비자의 소비 양상을 살펴보기 위해 네트워크 분석을 수행한 결과, 만화와 코믹스 등 출판물 이용 경험이 한국 웹툰 소비에 영향을 미친다는 것을 확인하였다. 다음으로, 한국 웹툰 해외 소비자의 소비 유형을 분류하기 위한 토픽 모형 분석을 수행한 결과, ‘삶-교훈 중심형’, ‘시각요소 중심형’, ‘캐릭터 중심형’, ‘로맨스 중심형’, ‘복합 장르형’의 다섯 가지 유형이 도출되었다. 이후, 도출된 소비 유형에 따라 실제 웹툰을 범주화하고, 각 유형이 글로벌 흥행에 미치는 영향을 규명하기 위해 회귀분석을 실시하였다. ‘삶-교훈 중심형’을 기준으로 분석한 결과, ‘시각요소 중심형’과 ‘복합 장르형’이 유의미하게 높은 구독자 수를 보였으며, ‘로맨스 중심형’ 역시 제한적인 인과관계를 나타냈다. 마지막으로, 회귀분석 결과의 타당성을 높이기 위해, 머신러닝 기반의 랜덤 포레스트 모형을 구축하고, 인공지능 기법인 SHAP을 활용하여 변수별 영향력을 분석하였다. 그 결과, ‘복합 장르형’, ‘시각요소 중심형’, ‘로맨스 중심형’으로 분류된 웹툰에서 상대적으로 높은 구독자 수가 나타났으며, 회귀분석 결과와 일관된 경향을 재확인하였다.

이러한 연구 결과는 다음과 같은 시사점을 제공한다. 첫째, 그간 한류의 하위문화로

간주되었던 웹툰을 독립적인 문화 콘텐츠로 재조명하고, 글로벌 수용 맥락에서 웹툰의 학문적·산업적 의의를 규명하였다는 점에서 의미가 있다. 웹툰은 단순한 디지털 만화를 넘어 글로벌 문화 콘텐츠로 진화하고 있으며, 본 연구는 이를 실증적으로 분석하여 기존 문헌에서 간과된 영역에 기여하였다. 둘째, 본 연구는 전통적 통계 기법과 함께 머신러닝 및 인공지능 기반 분석을 병행함으로써, 한국 웹툰 해외 소비자의 특성을 다차원적으로 분석하였다. 특히, 실제 플랫폼 이용자들이 자발적으로 생성한 텍스트 데이터를 활용함으로써, 분석의 신뢰성과 타당성을 제고하고 웹툰 이용에 영향을 미치는 핵심 요인을 보다 정밀하게 도출할 수 있었다. 셋째, 토픽 모형을 통해 도출된 웹툰 소비 유형은 단순한 분류를 넘어, 한국 웹툰의 글로벌 시장 세분화(segmentation)를 위한 실질적인 기준으로 활용될 수 있다. 이 같은 세분화 기준은 소비자 중심의 타겟 마케팅 전략 수립에 실질적인 기반을 제공하며, 각 유형에 맞춘 맞춤형 콘텐츠 기획을 통해 독자의 몰입도와 충성도를 높이고 장기적 팬덤 형성에도 기여할 수 있다. 넷째, 본 연구는 데이터 기반 의사결정을 통해 문화 콘텐츠 정책 수립의 정밀도를 높이는 데 기여할 수 있다. 특히, 문화체육관광부, 한국콘텐츠진흥원 등 유관 기관이 해외 진출 지원사업, 창작자 육성 프로그램, 플랫폼 육성 정책 등을 설계하는 과정에서, 본 연구 결과를 바탕으로 '시각적 독창성'과 '장르 혁신성'에 대한 평가 비중을 높이는 방식으로 정책 자원을 전략적으로 배분할 수 있다. 예를 들어, 웹툰 플랫폼은 시각적 매력을 강조하는 '비주얼 큐레이션'을 강화하고, 다양한 장르를 결합한 실험적 작품을 집중적으로 소개하는 '복합 장르 전문관' 등을 운영함으로써 핵심 소비층을 공략할 수 있다. 전통적으로 강세를 보여 온 '로맨스 중심형'에 대해서는 로맨스 판타지, 로맨스 코미디 등 세분화된 취향을 공략하는 '서브 장르 추천 시스템'을 고도화하는 것이 효과적일 수 있다. 다섯째, 본 연구는 글로벌 플랫폼과의 협력 및 해외 유통망 확대를 위한 국제 협력 정책의 필요성을 제기한다. 분석 결과, 한국 웹툰은 한류 콘텐츠 소비 경험이 없는 일반 해외 소비자에게도 충분한 매력을 지니는 콘텐츠임이 확인되었으며, 이에 따라 정부는 현지 시장에 특화된 콘텐츠 기획과 맞춤형 유통 전략을 수립할 필요가 있다. 특히, 번역 및 통역 지원의 확대는 언어 장벽을 해소함으로써 웹툰의 접근성을 높이는 핵심 전략으로 기능할 수 있다. 이처럼 현지 수요에 기반한 콘텐츠 설계 및 공급 전략은 한국 웹툰이 한류의 틀을 넘어 글로벌 대중문화로 자리매김하는 데 중요한 전환점이 될 것이다.

한편, 본 연구는 다음과 같은 한계를 지닌다. 첫째, 실제 한국 웹툰의 해외 시장을 다

루었음에도 국가별·문화권별 차이를 충분히 반영하지는 못하였다. 예를 들어, 동아시아 지역의 경우 전통 출판 만화와의 경쟁 구도가, 북미·유럽의 경우 그래픽노블과의 차별성이, 신흥 시장에서는 플랫폼 접근성과 번역 지원 여부가 소비 결정에 영향을 미칠 수 있다. 이러한 지역적 맥락은 소비자 행동과 시장 전략에 중요한 변수가 될 수 있으므로, 후속 연구에서는 국가·지역별 소비 양상, 선호 장르, 콘텐츠 접근성 등을 고려한 비교 분석이 필요하다. 둘째, 본 연구에서 도출한 한국 웹툰의 다섯 가지 소비 유형은 해외 소비자 데이터를 기반으로 얻은 결과로, 이러한 분류 양상이 국내 소비자와 유사한지 혹은 상이한지 여부는 직접적으로 검토되지 않았다. 향후 연구에서 국내 웹툰 소비자 데이터를 활용한 병행 분석이 수행된다면, 웹툰 소비에 내재된 범문화권적 보편성과 문화권별 특수성을 보다 명확히 구분할 수 있을 것이다. 이러한 접근은 해외 시장 진출 전략의 수립뿐만 아니라 내수 시장 전략의 정교화에도 실질적으로 기여할 수 있다. 셋째, 본 연구는 Reddit 커뮤니티 게시판 데이터를 활용하여 한국 웹툰과 기타 문화 콘텐츠 간의 연관 구조를 파악하였으나, 이러한 간접 지표만으로 웹툰의 실제 소비 양상을 완전히 설명하는 데에는 한계가 존재한다. 향후 연구에서는 실제 결제 이력 등 보다 직접적인 소비 데이터를 확보·활용함으로써, 웹툰 소비 구조에 대한 해석의 정밀도와 타당성을 한층 제고할 수 있을 것이다.○

## [참고문헌]

- 고세욱·박상혁·박정선(2018), 사회네트워크분석을 통한 시니어 기술창업교육참여자들의 네트워크 활성화 사례 연구, 「한국창업학회지」, 제13권 5호, 139-177.
- 김영재(2015), 웹툰 콘텐츠 생산자 비즈니스모델, 「애니메이션연구」, 제11권 3호, 7-26.
- 김용학(2004), 「사회 연결망 이론」, 개정판, 서울: 박영사.
- 김종대·이유석(2019), 토픽 모형 분석 기반 시장 세분화 접근법: 영화 시장을 중심으로, 「마케팅연구」, 제34권 5호, 53-72.
- 김주영·조영신(2023), 딥러닝을 활용한 클라우드 펀딩 성공 예측 모델 연구, 「마케팅연구」, 제38권 4호, 23-44.
- 김혜진·이명구(2021), 마케팅 분야의 머신러닝 연구 동향 분석, 「마케팅연구」, 제36권 1호, 1-25.
- 김휘정(2015), 한류지수 및 개선과 정책 활용도 제고 방안, 「문화정책논총」, 제29권 1호, 216-243.
- 남춘호(2016), 일기자료 연구에서 토픽모델링 기법의 활용가능성 검토, 「비교문화연구」, 제2권 1호, 89-135.
- 문진희·금영정(2016), 특허 네트워크 분석을 활용한 사물인터넷 기술융합 분석: 동시분류 분석 중심으로, 「한국경영과학회 학술대회논문집」, 2460-2466.
- 문화체육관광부 대중문화산업과(2024.01.23), 「만화·웹툰 산업 발전 방향」, Available: [https://www.mcst.go.kr/kor/s\\_notice/press/pressView.jsp?pSeq=20805](https://www.mcst.go.kr/kor/s_notice/press/pressView.jsp?pSeq=20805)
- 박지혜(2017), 국내 웹툰산업의 현황 및 시장발전을 위한 제언, 「월간 KIET 산업경제」, 제229권, 18-27.
- 박지혜(2022), K-웹툰산업의 IP 확대와 글로벌 진출 시사점, 「월간 KIET 산업경제」, 제290권, 78-87.
- 손동원(2002), 「사회네트워크 분석」, 서울: 경문사.
- 신용욱·박상혁(2013), 지역연고사업(RIS)을 통한 지역 중소기업 생태계의 네트워크 구조 변화 연구, 「벤처창업연구」, 제8권 3호, 77-84.
- 신준식(2023), 한국 웹툰의 해외시장 진출 현황을 통한 한국 웹툰의 과제 발견, 「만화애니메이션 연구」, 131-159.

- 양지훈·이지영·이상우(2016), 웹툰(Webtoon)의 흥행 결정요인 연구, 「한국콘텐츠학회 논문지」, 제16권 5호, 194-204.
- 오민정(2020), 빅데이터를 통해 살펴본 기부연구, 「마케팅연구」, 제35권 4호, 53-74.
- 오세규·강주영(2022), CartoonGAN 알고리즘을 이용한 웹툰(Webtoon) 배경 이미지 생성에 관한 연구, 「한국빅데이터학회지」, 제7권 1호, 173-185.
- 오정심(2021), 빅데이터 분석 기반 문화콘텐츠 학위논문 연구 동향(2004년~2020년), 「문화정책논총」, 제35권 3호, 185-218.
- 오현우(2024.03.11), ‘미국판 디씨인사이드’ 레드, 증시 데뷔 임박…벌써부터 ‘들썩’, 「한국경제」, Available: <https://www.hankyung.com/article/202403110463i>
- 우연주·유민환·이동규(2013), 네트워크 분석을 활용한 위기관리 네트워크 연구 가능성 고찰: 미국 전문가들의 네트워크 분석 연구 발전 방안 논의를 중심으로, 「국가위기관리학회 학술대회」, 33-36.
- 유진은(2015), 랜덤 포레스트: 의사결정나무의 대안으로서의 데이터 마이닝 기법, 「교육평가연구」, 제28권 2호, 427-448.
- 윤기현·정규하·최인수·최해솔(2015), 웹툰 통계 분석을 통한 한국 웹툰의 특징, 「만화 애니메이션 연구」, 38호, 177-194.
- 이근아(2023.11.05), 글로벌도 흥린 K웹툰, 드라마 흥행 치트키가 되다, 「한국일보」, Available: <https://www.hankookilbo.com/News/Read/A2023110210160003912>
- 이삼열·천영준·곽규태(2016), 한국인의 문화예술소비 심리 분석: ‘취향 절충주의’ 이론 기반 소셜 오피니언 마이닝, 「문화정책논총」, 제30권 1호, 204-226.
- 이준식·박도형(2019), 웹툰 콘텐츠 추천을 위한 소비자 감성 패턴 맵 개발, 「지능정보연구」, 제25권 4호, 67-88.
- 이지혜·임성준(2018), 한국 웹툰 플랫폼의 경쟁력과 전략, 「예술경영연구」, 제47호, 31-63.
- 최용석·정유나·고정민(2023), 웹툰산업의 글로벌경쟁력 요인 연구, 「문화산업연구」, 제23권 2호, 85-93.
- 최정아(2023), 정보화사업 성과와 이해관계자에 관한 탐색적 연구: 네트워크 분석을 활용하여, 「정책분석평가학회보」, 제33권 2호, 157-183.

- 한국국제문화교류진흥원(2024.04.02), 「2024 해외한류실태조사 결과보고서(제13차)」, 서울: 한국국제문화교류진흥원.
- 한국콘텐츠진흥원(2024), 「2023 웹툰 사업체 실태조사」, 서울: 한국콘텐츠진흥원.
- 황순민(2024.06.27), 웹툰 영향력 키우는 애플·아마존… 빅테크와 경쟁 가시밭길 예고, 「매일경제」, Available: <https://www.mk.co.kr/news/it/11053075>
- Blei, D. M., Ng, A. Y., & Jordan, M. I.(2003), Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 993-1022.
- Breiman, L.(2001), Random forests, *Machine Learning*, 45, 5-32.
- Büschken, J., & Allenby, G. M.(2016), Sentence-based text analysis for customer reviews, *Marketing Science*, 35(6), 953-975.
- Chen, C. S., Lin, Y. J., & Jeng, W.(2023), Flipping vs. scrolling in digital narratives: Reader engagement and reading patterns in comic layouts, In *Proceedings of the 86th Annual Meeting of the Association for Information Science and Technology (ASIS&T '23)*, London, UK: Association for Information Science and Technology, 914-916.
- DiMaggio, P., Nag, M., & Blei, D.(2013), Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding, *Poetics*, 41(6), 570-606.
- Gillespie, T.(2010), The politics of 'platforms,' *New Media & Society*, 12(3), 347-364.
- Glenny, L., Smith, J., & Witte, K.(2019), A framework for streamlined statistical prediction using topic models, *Journal of Data Science*, 17(3), 471-492.
- Hofmann, T.(1999), Probabilistic latent semantic analysis, In *UAI*, 99, 289-296.
- Katz, E., Blumler, J. G., & Gurevitch, M.(1973), Uses and gratifications research, *The Public Opinion Quarterly*, 37(4), 509-523.
- Lamerichs, N.(2020), Scrolling, swiping, selling: Understanding Webtoons and the data-driven participatory culture around comics, *Participations*, 15(2), 211-229.

- Li, X., Wu, C., & Mai, F.(2019), The effect of online reviews on product sales: A joint sentiment-topic analysis, *Information & Management*, 56(2), 172-184.
- Mun, J.(2025), Exploring content characteristics and consumer preference factors for webtoon development and recommendation: A focus on Naver Webtoon, *International Journal of Innovative Research and Scientific Studies*, 8(2), 4180-4188.
- Olson, R. S., & Neal, Z. P.(2015), Navigating the massive world of reddit: Using backbone networks to map user interests in social media, *PeerJ Computer Science*, 1, e4.
- Park, J. H., Lee, J., & Lee, Y.(2019), Do webtoon-based TV dramas represent transmedia storytelling? Industrial factors leading to webtoon-based TV dramas, *International Journal of Communication*, 13, 2179-2198.
- Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M.(2002), An introduction to logistic regression analysis and reporting, *The Journal of Educational Research*, 96(1), 3-14.
- Segneri, L., Babina, N., Hammerschmidt, T., Fronzetti Colladon, A., & Gloor, P. A.(2024), Too much focus on your health might be bad for your health: Reddit user's communication style predicts their Long COVID likelihood, *PLOS ONE*, 19(8), e0308340.
- Sheth, J. N., Newman, B. I., & Gross, B. L.(1991), Why we buy what we buy: A theory of consumption values, *Journal of Business Research*, 22(2), 159-170.
- Siroky, D. S.(2009), Navigating random forests and related advances in algorithmic modeling, *Statistics Surveys*, 3, 147-163.
- Steyvers, M., & Griffiths, T.(2007), Probabilistic topic models, In *Handbook of latent semantic analysis*, Psychology Press, 439-460.
- Stoltzfus, J. C.(2011), Logistic regression: A brief primer, *Academic Emergency Medicine*, 18(10), 1099-1104.
- Yao, L., Mimno, D., & McCallum, A.(2009), Efficient methods for topic model

inference on streaming document collections, In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, Paris, France: Association for Computing Machinery, 937-946.

[Abstract]

## **Korean Webtoons in the Global Market : A Data-Driven Analysis of Consumption Patterns and Expansion Strategies**

Kim, Juhee · Gong, Minjeong · Kim, Gayoung ·  
Lee, Hyerim · Lee, Shinhyoung

This study aims to develop strategies for the global expansion of Korean webtoons by conducting a multi-step analysis based on data from global webtoon consumers. First, logistic regression and network analyses were employed to examine global consumption patterns of Korean webtoons. The findings revealed that experiences with Korean publications and animations positively influenced webtoon use among foreign Hallyu fans. Concurrently, the experience of comics consumption among global consumers was also significantly related to webtoon usage. Next, topic modeling analysis was applied to classify global webtoon consumers into distinct types. The analysis identified five consumer segments: “Life-Lesson-Oriented,” “Visual-Oriented,” “Character-Oriented,” “Romance-Oriented,” and “Multi-Genre-Oriented.” Logistic regression and a machine learning technique, random forest, were employed to evaluate the market performance of each segment. The analysis revealed that the “Visual-Oriented” and “Multi-Genre-Oriented” types were associated with significantly higher subscriber counts. The Shapley additive explanations (SHAP) further confirmed the robustness of these findings. This study repositions webtoons as a distinct category of cultural content, and offers novel insights through an integrated data analysis approach combining traditional statistical methods with machine learning techniques. These findings have practical implications for segmenting the global webtoon market and for

developing targeted marketing strategies. Moreover, this study underscores the policy implication that cultural content policies should reflect global webtoon consumer usage patterns and preferences in order to facilitate more tailored and effective support systems.

**[Keywords]** webtoon, Korean wave (Hallyu), network analysis, topic modeling, random forest, SHAP

---

Kim, Juhee\_M.S. Division in Business Administration, Sookmyung Women's University / First Author (jo041h2@sm.ac.kr)

Gong, Minjeong\_M.S. Division in Business Administration, Sookmyung Women's University / Co-Author (mjkong12@sm.ac.kr)

Kim, Gayoung\_M.S. Division in Business Administration, Sookmyung Women's University / Co-Author (kk991011@sm.ac.kr)

Lee, Hyerim\_M.S. in Business Administration, Sookmyung Women's University / Co-Author (2331609@sm.ac.kr)

Lee, Shinhyoung\_Associate Professor, Division of Business Administration, Sookmyung Women's University / Corresponding Author (sh.lee@sm.ac.kr)

