웹툰 추천 서비스 (만화방)

1. 개요 (전체적인 설명 및 차별점 강조)

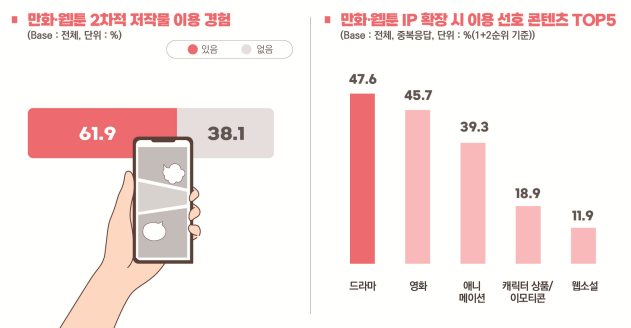
웹툰의 인기가 증가함에 따라 새로 나오는 웹툰의 양도 많아지게 되었다. 네이버 웹툰만 보더라도 베스트 웹툰, 베스트 도전 웹툰 등등 수 많은 웹툰, 여러 가지 장르의 웹툰들이 많아졌다. 이에 따라 사용자들은 자신의 취향에 맞는 웹툰을 찾기 힘들 수 밖에 없다.

우리 웹툰 추천 서비스에서는 사용자의 연령대, 즐겨보는 웹툰을 고려하여 그 사용자가 재미를 느낄만한 웹툰을 추천해주는 서비스이다.

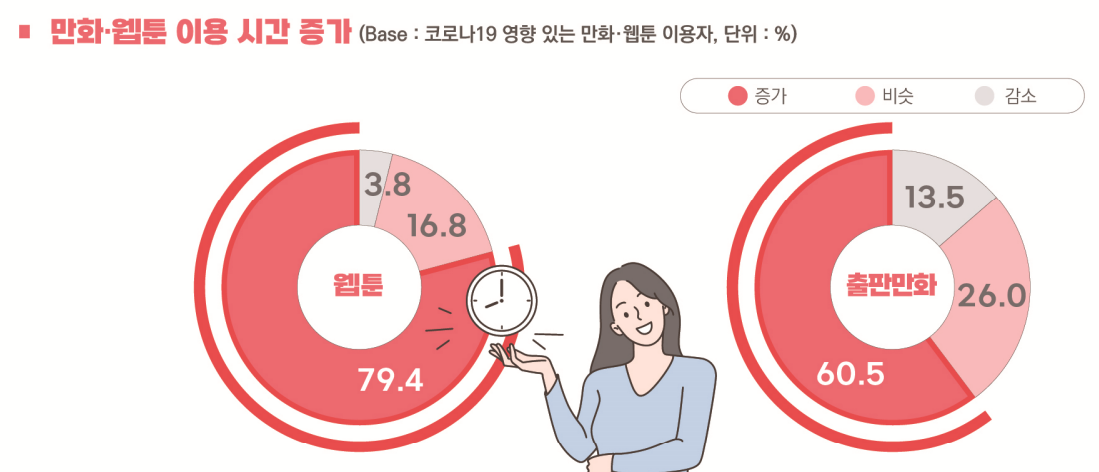
2. 프로젝트 배경 및 필요성

- (산업 동향)

최근들어 웹툰의 인기는 해마다 증가하고 있다. ‘지금 우리학교는’,’금수저’ 등 웹툰을 원작으로 하는 드라마나 영화들도 많이 만들어지고 있는데 이를 보면 웹툰의 인기를 실감할 수 있다.



위의 표는 한국콘텐츠진흥원에서 ‘2021 만화, 웹툰 이용조사’에서 발표된 결과이다. 만화, 웹툰 2차적 저작물 이용 해본 경험이 있는 사람은 61.9%, 그렇지 않은 사람은 38.1% 이다. (이용해보지 않은 사람은 38.1% 를 기록했지만 최근 sns, 유튜브 등으로 한 번쯤은 들어봤을 것이라고 생각한다.) 만화, 웹툰 IP 확장 시 이용 선호 콘텐츠 TOP5의 결과는 47.6%로 드라마가 제일 많은 수치를 기록했고 다음은 45.7%로 영화가 뒤따랐다.

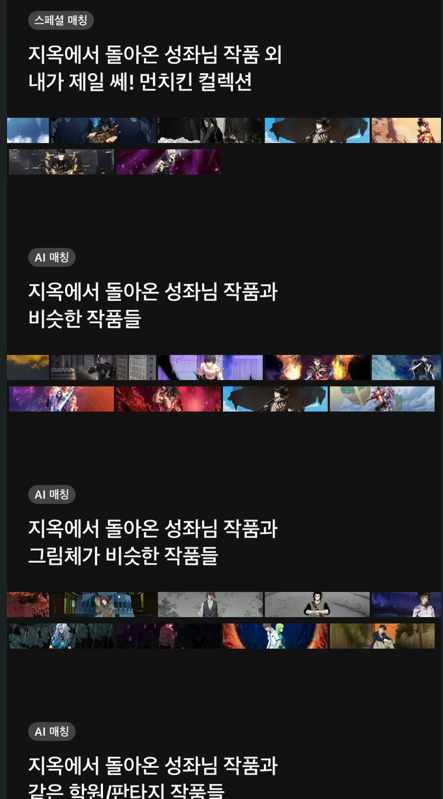


위의 표도 한국콘텐츠진흥원에서 설문조사한 결과이다. 만화, 웹툰 이용시간 증가에 대해서 웹툰은 79.4%, 출판만화는 60.5% 라는 대부분의 사람들이 웹툰을 즐기는 시간이 증가했다. 앞으로도 웹툰을 즐기는 사람들이 더욱 증가할 것이라고 전망된다. 이렇게 인기가 끝없이 증가함에 따라 당연히 새로운 웹툰들도 많이 추가될 것이다. 그렇게 되면 당연히 사용자의 취향에 맞는 웹툰을 찾기 힘들어질 수 밖에 없다.

그러므로 현재 뿐만 아니라 앞으로도 웹툰 추천 서비스의 필요성은 더욱 더 부각 될 것이라고 생각된다.

- (기존 서비스 분석(장단점))

카카오 웹툰은 다양한 AI 추천 서비스를 제공하고 있다.



네이버 웹툰은 현재 협업기반필터링만 제공하고있다.

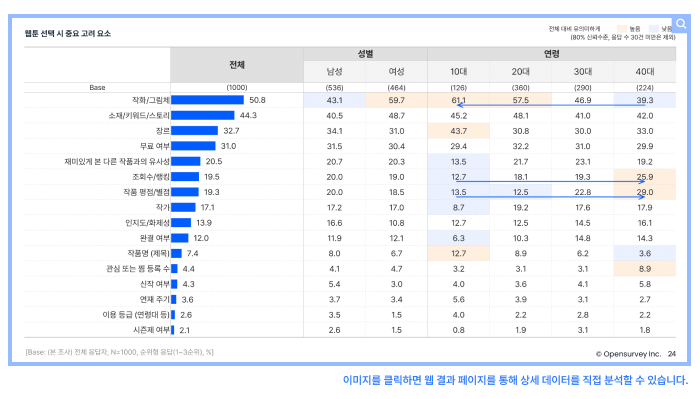


이용자가 가장 많은 플렛폼인 네이버 웹툰은 마땅한 추천 서비스가 제공되고 있지 않다.

네이버 웹툰에대한 더 구체적인 추천 시스템을 제공하고자 한다.

3. 서비스 설명

- (기능 설명)



위의 설문조사는 사람들이 웹툰을 선택할 때 고려하는 사항이다. 가장 많이 고려하는 것은 ‘작화/그림체’이고 그 뒤로 ‘소재/키워드/스토리’, ‘장르’ 가 뒤를 이었다.

이번 만화방의 서비스는 사용자가 원하는 ‘작화/그림체’, ‘소재/키워드/스토리’, ‘장르’로 나누어 원하는 취향에 맞는 웹툰을 추천하는 서비스이다.

1. 데이터 수집 방법
2. 데이터 분석 방법
3. 데이터 모델링
4. 데이터 분석
5. 테스트
6. 시험 및 운영 적용

* 업무 분장

서정우

김재현

김태형

정승환

- (기술 스택)

운영체제 : Windows OS, Mac OS

IDE : Jupyter Notebook, Visual Studio Code

데이터 수집 : BeautifulSoup, Selenium

Front-end : HTML5,CSS,JavaScript

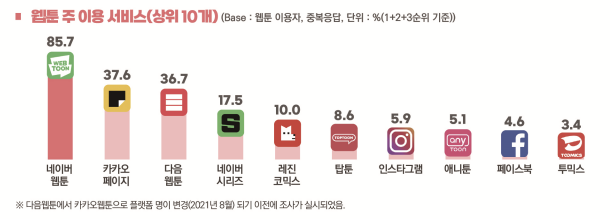
Server : Flask

DBMS : MySQL

데이터 시각화 : matplotlib

- (개발 과정)

< 데이터 수집 대상 결정 >

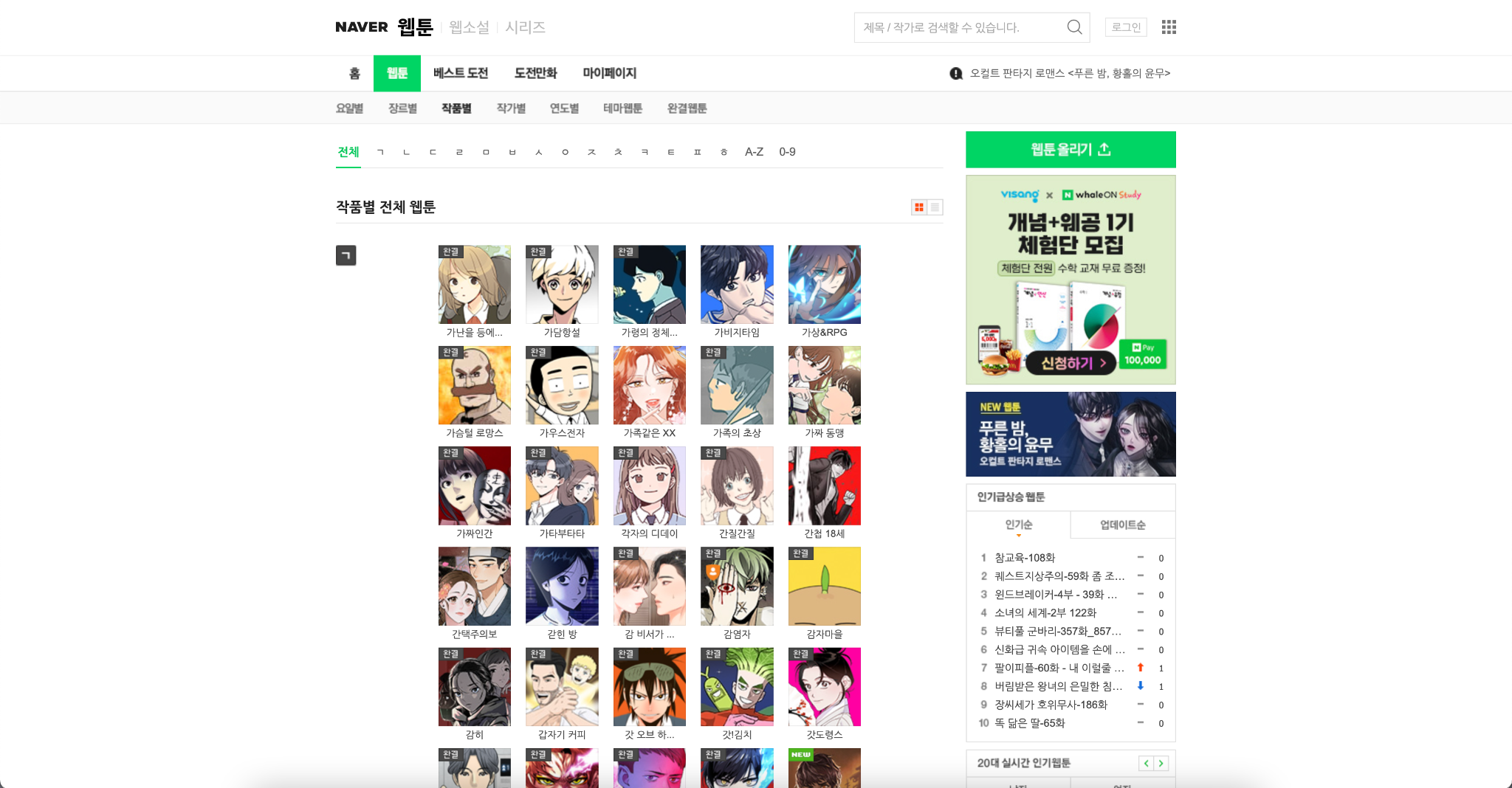
 한국콘텐츠진흥원에서 설문조사 한 내용에서 웹툰 서비스 중 네이버 웹툰을 가장 많이 사용 하므로 추천 대상은 네이버 웹툰으로 결정했다.

4. 개발구현방법

가. 데이터 수집 방법

1. 네이버 웹툰 크롤링(추가예정)

네이버 웹툰 사이트의 작품별 을 기준으로 크롤링(<https://comic.naver.com/webtoon/creation>)



1. 네이버 웹툰 선호도 설문조사

총 5가지 질문으로 구성

1. 연령대

연령대는 10대부터 80대까지 8개의 항목으로

10대 3명, 20대 95명, 30대 2명, 40대 1명

1. 성별

남성, 여성,비공개 3가지 항목으로 설정

남성 76명, 여성 21명, 비공개 4명

추후 비공개 4명은 데이터 변환시 남자로 변환

1. 현재 보는 웹툰, 점수

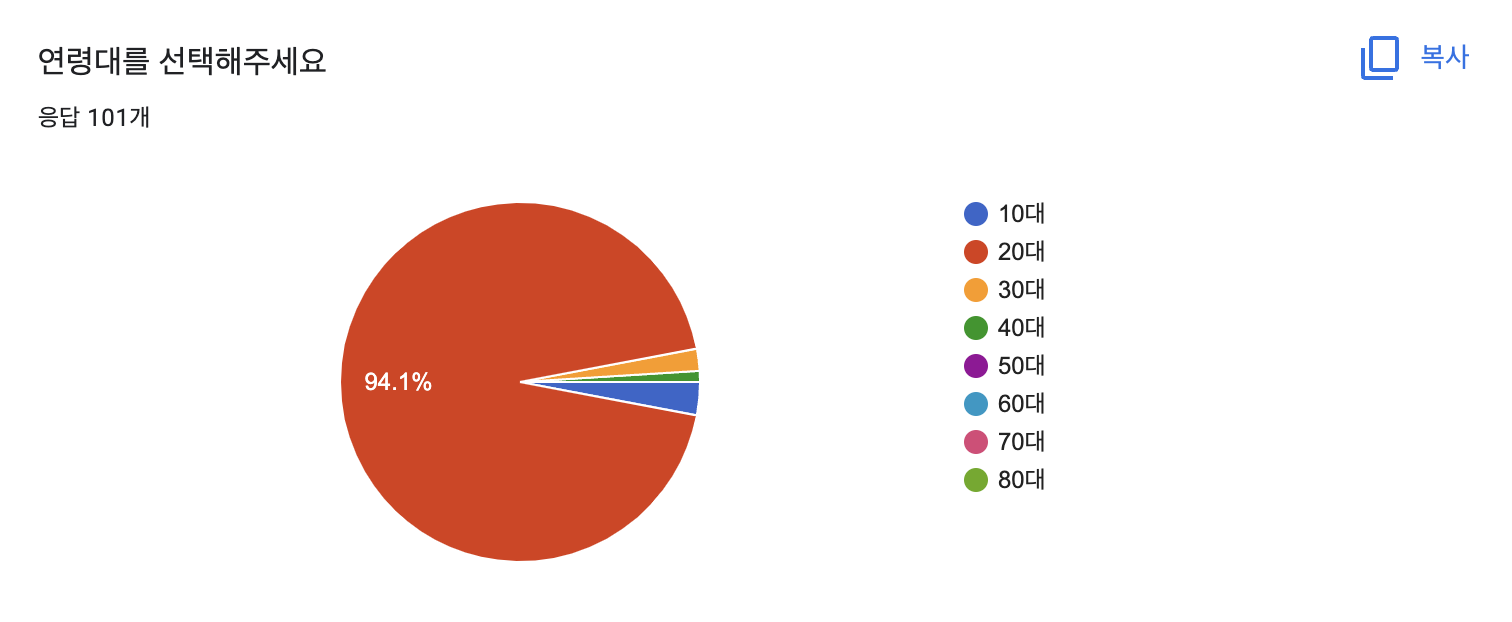
참여자가 보는 웹툰들과 그에따른 점수를 입력받음

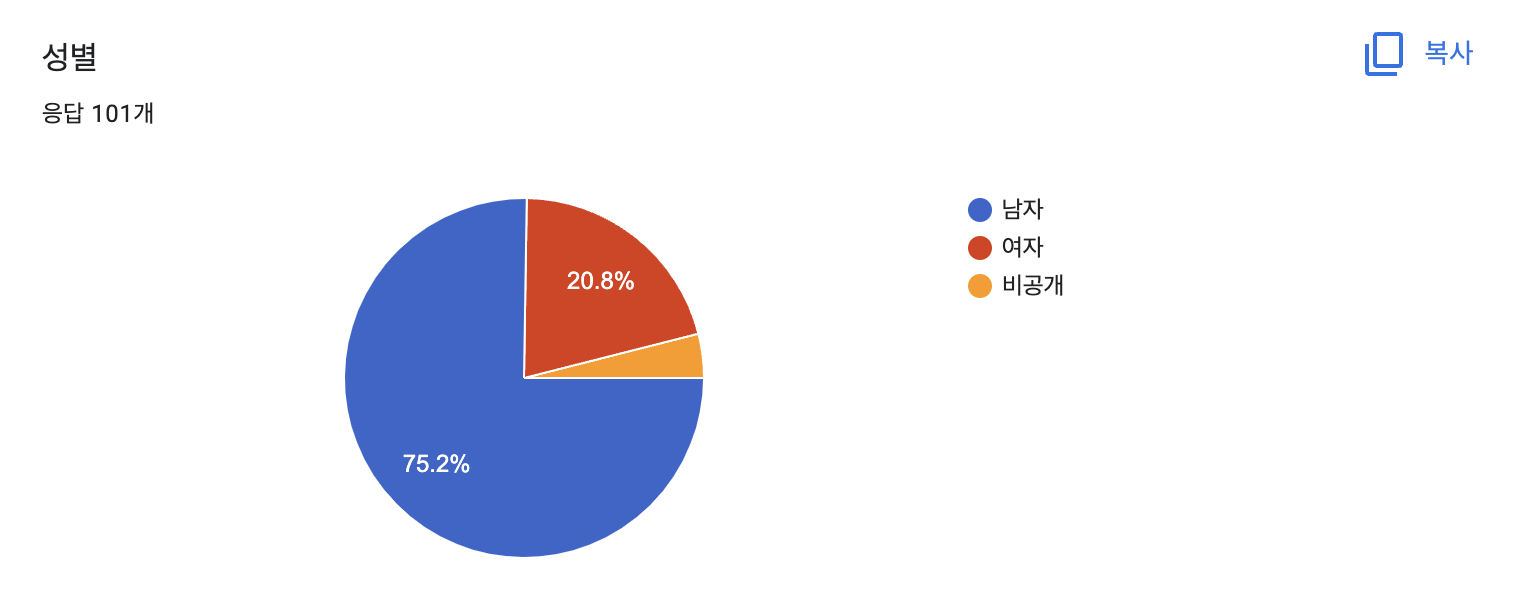
1. 재미 없었던 웹툰

참여자가 보는 웹툰 중 재미없는 웹툰을 입력받음

1. 추천해주고싶은 웹툰

참여자가 보는 웹툰 중 가장 재미있는 웹툰을 입력받음









나. 데이터 전처리

1. Pandas
2. 필요없는 컬럼 삭제
3. 범주형 변수를 가변수로 변환

( genre1 컬럼 )

| 변경 전 | 변경 후 |
| --- | --- |
| 스토리 | 0 |
| 에피소드 | 1 |
| 옴니버스 | 2 |

( genre2 컬럼 )

| 변경 전 | 변경 후 |
| --- | --- |
| 로맨스 | 0 |
| 드라마 | 1 |
| 판타지 | 2 |
| 스릴러 | 3 |
| 액션 | 4 |
| 개그 | 5 |
| 일상 | 6 |
| 무협 / 사극 | 7 |
| 감성 | 8 |
| 스포츠 | 9 |

( age 컬럼 )

| 변경 전 | 변경 후 |
| --- | --- |
| 전체이용가 | 0 |
| 12세 | 1 |
| 15세 | 2 |
| 18세 | 3 |

( likes 컬럼 : int형으로 변환 )

| 변경 전 | 변경 후 |
| --- | --- |
| 99,999+ | 100,000 |

(설문조사 중 점수 데이터)

| 변경 전 | 변경 후 |
| --- | --- |
| 재미 O | 5 |
| 재미 X | 1 |
| 0점 | 1 |
| 추천해주고 싶은 웹툰 | 6 |

다. 데이터 분석

보류

5. 향후발전방향 및 기대효과

- (활용방안, 기대효과)

웹툰 추천 서비스는 네이버 웹툰 뿐만 아니라 다른 웹툰 서비스에서도 충분히 활용할 수 있다. 현재 네이버 웹툰에는 협업 필터링 추천 서비스를 제공한다. 하지만 이 추천 방법은 소외되는 웹툰 콘텐츠가 생긴다는 단점이 있다. 하지만 우리는 협업 필터링 추천 방법의 단점을 보완하기 위해 콘텐츠 기반 필터링도 같이 이용할 예정이다. 그렇게 된다면 소외되는 웹툰 콘텐츠 없이 많은 사용자들이 다양한 웹툰을 즐길 수 있을 것이라 예상한다.

웹툰의 인기는 더욱 더 증가할 것이라고 본다. 그에 따라 웹툰을 기반으로 하는 드라마나 영화가 많이 제작될 것으로 전망된다. …..

\*\*대충 써봄띠 사용할 문장은 쓰고 버릴건 버려용\*\*

현재 네이버 웹툰에는 협업 필터링 추천 서비스를 제공한다. 사용자들의 웹툰 이용 데이터를 바탕으로 새로운 사용자에게 추천을 해주는 서비스이다. 이러한 추천 방식은 콜드스타트 현상이 나타나게 된다. 기존에 없던 새로운 웹툰이 추가되는 경우 해당 웹툰에 대한 데이터가 쌓이기 전에는 추천 대상에서 소외된다는 단점이 있다. 이러한 문제점을 개선하기 위해 만화방 프로젝트에선 콘텐츠 기반 필터링도 함께 사용할 예정이다. 사용자들의 이용 데이터 뿐아니라 실제 웹툰의 작화, 키워드, 장르, 별점등의 데이터를 이용해 소외되는 웹툰 없이 다양하게 추천 해 줄 예정이다. 콘텐츠 기반 필터링 사용시 또 다른 장점은 네이버 웹툰 뿐 아니라 다른 웹툰 플랫폼에서도 데이터만 있다면 충분히 사용할 수 있다는 점이다.

웹툰을 기반으로 하는 드라마나 영화와 같은 컨텐츠가 많이 제작됨에 따라 더욱 많은 사람들이 웹툰을 접하게 될 것이다. 기존의 수많은 웹툰과 새로운 작품들이 쏟아져 나오는 오늘날, 웹툰 추천 서비스는 처음 웹툰을 접하는 사람들의 진입장벽을 낮춰주고 기존의 사용자들에게는 신선한 컨텐츠를 추천해줄 수 있는 해결방안이 될것이다.

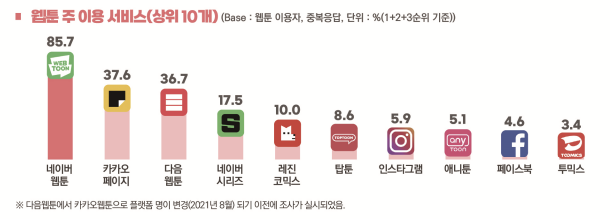
6. 소요예산 및 경비

- (? 사전 조사 필요)

6. 개발 과정

1. 데이터 수집

* 네이터 웹툰



한국콘텐츠진흥원과 오픈서베이에서 ‘웹툰 주 이용 서비스’를 설문조사한 결과, 네이버웹툰이 1위를 차지하여 가장 많이 사용하는 플랫폼으로 선정되어 네이버웹툰의 데이터를 크롤링하기도 결정함

* 구글 폼을 사용하여 네이버웹툰 선호도 설문조사지를 만들고 답변 수집(100명)

1. 데이터 전처리

* Pandas 사용
* 필요없는 컬럼 삭제
* 범주형 변수를 가변수로 변환
  + genre1

| 변경 전 | 변경 후 |
| --- | --- |
| 스토리 | 0 |
| 에피소드 | 1 |
| 옴니버스 | 2 |

* + genre2

| 변경 전 | 변경 후 |
| --- | --- |
| 로맨스 | 0 |
| 드라마 | 1 |
| 판타지 | 2 |
| 스릴러 | 3 |
| 액션 | 4 |
| 개그 | 5 |
| 일상 | 6 |
| 무협 / 사극 | 7 |
| 감성 | 8 |
| 스포츠 | 9 |

* + age

| 변경 전 | 변경 후 |
| --- | --- |
| 전체이용가 | 0 |
| 12세 | 1 |
| 15세 | 2 |
| 18세 | 3 |

* + likes

| 변경 전 | 변경 후 |
| --- | --- |
| 99,999+ | 100,000 |
| 문자형 | 정수형 |

* + rates

| 변경 전 | 변경 후 |
| --- | --- |
| 문자형 | 실수형 |

* + 설문조사 점수 데이터

| 변경 전 | 변경 후 |
| --- | --- |
| 재미 O | 5 |
| 재미 X | 1 |
| 0점 | 1 |
| 추천해주고 싶은 웹툰 | 6 |

* + 설문조사 성별 데이터

| 변경 전 | 변경 후 |
| --- | --- |
| 남자 | 0 |
| 여자 | 1 |

* + 설문조사에서 재미있다고 생각한 웹툰 : 5
  + 설문조사에서 재미없다고 생각한 웹툰 : 1
  + 설문조사에서 추천해주고 싶은 웹툰 : 6

1. 데이터 분석

…..