AI 기반 유망 인플루언서 발굴 및 매칭 플랫폼 개발

조유라, 김은수, 김주연, 김예진, 한채원
서울여자대학교 데이터사이언스학과 학부생
stxllaaa03@swu.ac.kr, lyra8@swu.ac.kr, kimin3939@swu.ac.kr, yejinkim00@swu.ac.kr,
hxchaeny@swu.ac.kr

The development of an AI-Based Platform for Identifying and Matching Promising Influencers

Yura Cho, Eunsoo Kim, Juyeon Kim, Yejin Kim, ChaeWon Han
Dept. of DataScience, Seoul Women's University

요 약

본 논문은 인플루언서 매칭 서비스를 위한 데이터 수집 및 분석 방법, 매칭 및 성과 모니터링 알고리즘, 웹 서비스 구현에 대해 말하고 있다. 매칭 알고리즘은 사용자가 원하는 인플루언서를 맞춤형으로 제공하며, 성과 모니터링을 이용해 유의미한 자료를 시각화하여 제공한다.

1. 서론

최근 인플루언서를 통한 브랜드 마케팅이 활성화되며 유명 인플루언서의 영향력이 확대되고 있다. 그러나 인플루언서 마케팅은 이미 인지도가 있는 인플루언서들을 기준으로 진행되어 [1] 마케팅 비용이 많이 발생한다는 것과 기업이 예상한 반응을 끌어오지 못할 경우 그에 대한 벌충비용은 기업이 져야 한다는 치명적인 단점이 존재한다. 또한, 기존의 마케팅 방식은 어떤 인플루언서와 협업할 것인지 정하는 과정에서 기업이 직접 비용과 인플루언서의 행실, 계정의 아이덴티티 등을 수집하여 분석해야 하는 번거로움이 있다. 따라서 인지도가 높진 않지만, 성실한 업로드로 신뢰성이 높은 유망 인플루언서를 발굴하여 매칭시키고, 협업 이후의 성과 이행도까지 한 번에 파악할 수 있는 서비스 "스타 캐처"를 제안한다. 본 서비스는 기업의 마케팅 시간과 비용 절감의 효과를 제공하고, 검증된 인플루언서가 스스로를 PR할 수 있는 플랫폼을 만들고자 한다.

2. 소프트웨어 구조

스타캐쳐의 매칭 서비스는 최소 비용으로 최대 효율을 낼 수 있는 유망 인플루언서들과의 매칭을 원하는 프리미엄 고객들을 대상으로 진행되는 서비스이다. 그러므로 Web도 다양한 연령층의 프리미엄 고객들이 사용하기 쉬운 UI로 구성하였고, 데이터베이스 또한 너무 무겁지 않고, 간결하지만 다양한 기능과 우수한 보안성을 가진 MariaDB로 선정했다. 그래서 클라이언트가 Web에서 API 호출을 통해 서버로 데이터를 요청하면, 서버는 설계된 API의 기능에 따라 요청에 대한 작업을 수행한다. 서버는 필요한 경우 DB와 통신하여 데이터를 가져온 후 가공하여 클라이언트의 요청에 대한 응답에 데이터를 포함하여 보내도록 설계하였다.

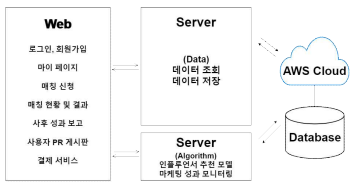

그림1) 소프트웨어 구조도

3. 웹 서비스 구조 및 구현

3-1 알고리즘

1)데이터 수집

Python의 라이브러리인 BeautifulSoup, Selenium을 이용해 인스타그램 웹 크롤링을 진행했다. 먼저 인스타그램에서 인기 게시물 상단에 노출되는 인플루언서 목록을 추출하기 위한 자동화 코드를 작성하였다. 웹 버전 인스타그램에서 해시태그를 검색했을 때 추천 게시물이 제대로 표시되지 않는 이슈가 있기 때문에, 인플루언서 계정에서 자주 사용하는 것으로 알려진 #광고, #협찬 등의 해시태그를 주기적으로 검색하여 게시글을 작성한 계정의 정보를 모두 수집했다. 이후 수집된 인플루언서들의 게시물 수, 팔로워 및 팔로우 수, 게시글과 댓글의 내용 및 포함된 해시태그 등의 정보를 인플루언서 매칭 모델에 학습시킨다. 인스타그램에서는 원칙적으로 데이터 크롤링 시도를 허용하지 않고 있기에, 일정 횟수마

다 접속 계정을 변경하며 계정 제한 조치를 피하는 방식의 자동화 코드를 작성하여 반복적으로 새로운 데이터를 수집할 수 있도록 했다. 이렇게 수집된 데이터 중 intro(소개글) 데이터는 TF-IDF로 변환하여 벡터화시킨 뒤, 알고리즘에 사용된다.

2)인플루언서 추천 모델

크롤링 데이터에서 인플루언서 소개글인 "intro" 칼럼안에서 키워드를 추출하여 "key"칼럼을 만든다. 이후에 두 칼럼간의 코사인 유사도를 측정하여 가장 높은 값을 가지는 "key"값을 해당 인플루언서의 카테고리로 설정한다. clustering으로 K-means 군집화를 하여 카테고리로 분류한 뒤, 각 카테고리에 해당하는 번호를 부여하여 "num_group"에 추가한다. 인스타그램 "follower" 칼럼값을 바탕으로 상, 중, 하 레벨로 인플루언서를 분류한다. 이를 통해 기업이 원하는 예산의 인플루언서를 추천해 줄 수 있으며. 최종적으로, 기업이 원하는 예산 내에서 가장 통일성 있게 자신의 계정을 운영한 인플루언서를 3명 추천한다.

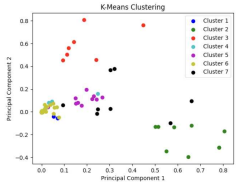

그림2) K-Means Clustering 모델링 결과

3)마케팅 성과 모니터링

매칭 알고리즘의 평가는 위에서 사용한 모델링의 평가지표를 사용하여 분석한다. 코사인 유사도 분석은 유사도가 1에 가까울수록 높은 유사도를 가졌음을 나타내고 k-means 클러스터링은 데이터를 2D로 차원 축소한 뒤, 그룹핑이 얼마나 잘 되었는가를 시각화하여 기업에 신뢰성 있는 매칭 알고리즘을 제공한다.

마케팅 성과 모니터링은 게시물의 지속성과 게시물의 영향력을 평가하여 제공한다. 첫째, 게시물의 지속성은 인플루언서가 게시한 게시물의 url에 접속하는 코드를 이용하여 게시물이 열리지 않는다면 게시물이 삭제된 것으로 간주한다. 둘째, 기업은 매출 데이터에 민감하기 때문에 직접적으로 데이터를 구할 순 없지만 기업의 재무제표에 명시된 인플루언서 광고 지출과 매출 등을 비교하여 매출을 짐작하여 인플루언서의 [2] 영향력을 측정할 수 있다. 마지막으로 인플루언서가 게시한 게시물의 평균 좋아 요수와 댓글 수를 수치상으로 게시물의 호응도를 측정할 수 있고 댓글의 긍·부정 분석을 통해 게시물의 호감도를 측정할 수 있다.

3-2 Web

그림3) 웹사이트 서비스 화면

그림 3에서와 같이 본 웹사이트는 크게 로그인, 기본 정보 입력, PR LOUNGE, 매칭 신청, 마이페이지로 구성되었다. 웹의 UI는 본 서비스를 이용하는 다양한 연령층을 고려하여 손쉬운 조작으로 이용할 수 있도록 기획했다. 구체적으로 PR LOUNGE는 인플루언서와 기업이 각각 홍보하는 게시글을 작성해 양 방향적 이용을 도모한다. 매칭 신청은 알고리즘을 토대로 레벨과 분야를 선택해 매칭을 할 수 있도록 구성했다. 마이페이지는 사용자 정보, 프로젝트 이행률, 매칭 히스토리를 확인할 수 있게 구성했다.

4. 결론 및 향후 연구 과제

본 논문에서는 기업이 원하는 카테고리의 유망한 인플루언서를 발굴하고 매칭시켜 주는 서비스인 '스타캐처'에 대해 기술하였다.

연구를 진행하며 기업이 매출 데이터를 오픈하지 않아 수치적 평가에서 어려움을 겪었던 마케팅 성과 모델은 향후 기업이 직접 매출 데이터를 입력하는 단계를 만들어 해당 데이터를 보충하는 방식으로 모델을 발전시킬 수 있을 것이다.

또한 크롤링을 통해 데이터를 수집한 방식에 따른 데이터의 양이 적다는 한계가 있어, 인스타그램 개발자 API를 통해 데이터를 수집하는 것으로 발전시켜 향후 더 많은 인플루언서의 데이터를 수집할 수 있기를 기대한다.

따라서 마케팅 비용이 많이 발생하는 인플루언서 중심으로 이루어지는 기존 마케팅 시장을 개편하고 유망 인플루언서에게는 기회를 기업에는 비용 절감을 이뤄내길 기대한다.

Acknowledgment

※ 본 프로젝트는 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] Xiao Han, Leye Wang, Weiguo Fan Cost-Effective Social Media Influencer Marketing. INFORMS Journal on Computing 35(1):138-157. (2022)

[2] Salminen, Kalle. "Creating Influencer Marketing Campaign In Social Media.2018(".).