

딥러닝을 활용한 크라우드 펀딩 성공 예측 모델 연구

Deep Learning Predictive Models for Crowdfunding Success

김주영 · Kim, Juyoung, 조영신 · Cho, Youngsin

크라우드 펀딩 플랫폼은 스타트업의 초기 자금충당 수단으로 성장해왔으며, 자금충당 수단 이외에도 상품의 시장성 진단을 위한 사전 판매, 창작자들의 작품 판매 등 다양한 목적으로 활용되고 있다. 크라우드 펀딩 관련 기존 연구들은 동영상 수, 이미지 수, 기간 등의 수치 데이터를 변수로 주로 사용하였다. 그중 일부 연구에서는 비정형 데이터인 텍스트 변수를 활용하고 있으며 품사의 수, 문장의 길이 등을 사용하거나 토픽분석을 통해 추출한 토픽을 변수로 활용하고 있다. 하지만 이러한 변수들은 텍스트의 맥락적 의미를 포함하지 않거나 제한적으로 반영하고 있는 실정이다.

따라서 본 연구에서는 텍스트의 활용을 확장하고 문맥적 의미를 반영하기 위해 언어모델을 활용한다. 먼저, 딥러닝 예측모델과 딥러닝 분류모델로 두 개의 모델로 연구를 진행하였다. 텍스트 변수를 분석하기 위하여 2017년 발표된 사전학습된 BERT 모델과 직접 텍스트 자료를 입력하여 학습한 트랜스포머의 인코더 모델을 활용하였다. 또한 기존 연구와 달리 종속변수를 펀딩 서포터 수로 설정하였으며, 벡터와 수치 데이터를 활용하여 딥러닝 모델을 구축하였다. 그 다음으로 서포터 수가 날짜별로 증가하는 자료도 함께 이용하였다.

본 연구는 와디즈 사이트에서 2021년 1월부터 2023년 1월까지 총 9,755개의 펀딩이 완료된 데이터를 크롤링하였다. 크롤링한 데이터는 펀딩 사이트에서 제공하고 있는 프로젝트별 카테고리, 펀딩 이름, 펀딩 설명글, 본문 텍스트, 펀딩 기간, 펀딩 금액, 펀딩 서포터 수, 달성률, 옵션 가격, 옵션 수, 메인 이미지 수, 메인 동영상 수, 본문내 이미지 수, 본문내 동영상 수, 스크롤 길이, 일일 펀딩 금액 데이터이다. 딥러닝 모델에서 텍스트를 활용하여 BERT 및 트랜스포머 인코더(Encoder)에 넣어 분석한 데이터와 수치 데이터들을 활용하여, 전통적인 회귀모델과는 다른 형태의 구조를 가지고 더 좋은 결과를 만들 수 있었다. 본 연구는 플랫폼 사용자와 운영자 모두에게 크라우드 펀딩 성공을 이해하고 예측하는 새로운 접근 방식을 제시한다.

핵심주제어: 크라우드 펀딩, 딥러닝, 언어모델, 예측모델, 텍스트 분석, 멀티모달

김 주 영 | 서강대학교 경영대학 교수(jkimsg@sogang.ac.kr), 제1저자

조 영 신 | 서강대학교 일반대학원 석사(dudtnls601@naver.com), 교신저자

ABSTRACT

Crowdfunding platforms have grown as a means of initial funding for startups, and they are also being used for various purposes beyond funding, such as pre-selling products for market assessment and selling the works of creators. Existing research on crowdfunding has mainly used quantitative data such as video views, image counts, and duration as variables. Some studies have incorporated unstructured text variables, utilizing metrics like the number of parts of speech, sentence length, or topics extracted through topic analysis. However, these variables often lack the contextual meaning of the text or provide limited reflection.

In this study, language models are employed to extend the use of text and incorporate contextual meaning. Two models, DNN prediction models and classification models, were employed for the research. For text variables, pretrained BERT models released in 2017 and Transformer Encoder models trained directly on text data were utilized. Unlike previous research, the dependent variable was set as the number of funding supporters, and DNN was constructed using both text and numeric data. Furthermore, data featuring the increase in supporters on a daily basis was also utilized.

This study collected data from Wadiz, a crowdfunding site, spanning from January 2021 to January 2023, encompassing a total of 9,755 completed funding projects. The collected data includes project categories, funding names, descriptions, main text, funding duration, funding amount, number of supporters, achievement rate, option prices, option quantities, counts of main images and videos, counts of images and videos in the main text, scroll length, and daily funding amount. By employing text in DNN models, using data analyzed by BERT and Transformer Encoder alongside numeric data, a different structural form from traditional regression models was achieved, resulting in improved outcomes. This study presents a new approach for both platform users and operators to understand and predict crowdfunding success.

Keywords: Crowdfunding, Deep Learning, Language Model, Predictive Model, Text Analysis, Multimodal

Juyoung Kim | Professor of Marketing, Sogang University, First Author

Youngsin Cho | Master in Graduate School (Business Administration), Sogang University, Corresponding Author

1. 서론

크라우드 펀딩은 스타트업의 초기 자금 조달 수단으로 사용되며, 상품의 시장성 진단, 사전 판매, 작품 판매 등 다양한 목적으로 활용되고 있다(Nielsen and Binder 2021). 이는 강력한 자금 조달 도구일 뿐만 아니라 아이디어 개발, 고객 관계 촉진, 효율적인 마케팅 등 다양한 목적을 제공한다(Gerber, Hui and Kuo 2012; Hui, Gerber and Greenberg 2012). 크게 보면 크라우드 펀딩은 투자를 필요로 하는 기업이나 이들을 대표하는 에이전트가 투자금을 모집하고 추후에 배당금을 받는 투자형 펀딩과 투자를 했을 때 재화 등으로 보상을 하는 보상형 펀딩으로 구분된다. 특히, 우리나라에서는 보상형 크라우드 펀딩 플랫폼이 급성장하였으며, 2019년에는 전년 대비 168.2% 성장한 것으로 나타났다(이구형 2020). 보상형 펀딩의 원래 의미는 새로운 사업 혹은 제품 개발 아이디어를 보여주고 이에 대한 펀딩을 받은 후 보상을 하는 개념이다. 이런 경우, 신제품의 판매에 측 및 고객들의 선확보 및 홍보효과 등도 같이 기대할 수 있으며, 보상으로 해당 제품을 주면 투자자의 입장에서 필요할 제품을 남들보다 미리 얻을 수 있는 기회가 된다. 하지만 제품에 하자가 있을 때 소비자 보상 관련된 것이 어려울 수 있으며, 이미 시장에서 판매되는 제품에 대한 펀딩을 받는 경우 등도 있다는 우려가 있다.

크라우드 펀딩도 하나의 온라인을 통한 마케팅활동의 하나라고 볼 수 있다. 온라인을 통한 제품의 판매나 구매의도에 미치는 영향은 꾸준히 이루어져 왔다. 온라인 플랫폼에서의 구매에 영향을 미치는 요소들에 의한 연구들도 꾸준히 진행되어 왔다. 예를 들면 온라인평점이 예술품과 기능품에 미치는 영향이 다르다는 연구도 있고(이서진, 최인철 2022), 온라인 쇼핑몰에 사용되는 추천 시스템의 품질 요소의 효과를 연구하기 위하여 신뢰도와

이미지의 매개효과를 연구하기도 하였다(성기범, 한상만, 김태완 2022). 또 소셜미디어를 통한 인스타그램 피드 맞춤형광고에서 지각된 개인화와 지각된 유용성이 소셜 CRM에 대한 태도와 관계지속의도에 미치는 영향(윤영진, 박정은 2022)도 온라인마케팅에서 필요로 하는 연구들이다.

크라우드 펀딩에 대한 연구는 많이 진행되고 있지만, 기존 연구들은 주로 수치 데이터를 중심으로 하여 텍스트 데이터의 활용은 제한적이었다. 이전 연구에서는 목표액과 펀딩 기간과 같은 수치 데이터가 펀딩 성공에 영향을 미치는 요인으로 확인되었다(Mollick 2014; Kuppuswamy and Bayus 2018). 또한, 이미지 수, 프로젝트 업데이트 비율, 비디오 수 등이 펀딩 성공과 양의 상관관계를 가진 것으로 나타났다(Mollick 2014). 국내 텀블벅을 대상으로 한 연구에서는 동영상, 시각 이미지가 펀딩 성공에 유의한 영향을 미치는 것으로 나타났다(이정은, 신형덕 2014). 또한, 초기 투자자의 참여 확보도 펀딩 성공에 영향을 미치는 요소로 확인되었다(Vismara 2018; Fan, Gao and Steinhart 2020). 또 크라우드 펀딩에서 사회적증거(참가자 수)와 마감까지의 시간이 투자의도에 미치는 영향을 연구하기도 하였다(허재강, 엄금철, 차동진 2022). 더 나아가서 크라우드 펀딩을 통한 기부연구에서는 기부자의 권력거리신념과 수혜자의 필요형태에 따라 기부의도가 지각된 감정이입과 책임감에 의하여 매개되는 것이 연구되기도 하였다(씨에천짜오, 최우진 2022). 그러나 이전 연구들은 텍스트의 수나 길이 등의 수치적 정보만을 사용하였으며, 텍스트의 문맥적 의미를 반영하여 크라우드 펀딩을 분석한 연구는 아직 없다.

최근 자연어 처리 작업에서 탁월한 성과를 보여주는 사전 훈련된 언어 모델인 트랜스포머(Transformer) 모델이 등장하였다(Vaswani et al. 2017). 트랜스포머

의 인코더 구조에는 Self-Attention 메커니즘이 포함되어 있으며 이를 활용하여 문장 속 모든 단어들 간의 상호작용을 파악하여 문장의 맥락을 이해할 수 있다. 이러한 트랜스포머의 인코더 부분을 활용한 모델이 BERT이며, 다양한 자연어 이해 작업에서 우수한 성능을 보인다. 본 연구에서는 BERT를 활용하는 것과 트랜스포머의 인코더를 직접 활용하여 텍스트 데이터의 문맥적 의미를 반영하는 접근 방식을 비교 분석하고자 한다.

본 연구는 딥러닝 방법을 활용하여 크라우드 펀딩 플랫폼에서 텍스트 데이터의 맥락적 의미를 포함하여 펀딩 성공을 예측하는 새로운 방법을 제시하고자 한다. 또한 이를 통하여 가장 높은 펀딩 성공을 만들어내는 펀딩 제안과 관련된 요소들도 확인해보고자 한다. 이를 통하여 스타트업 및 크라우드 펀딩 플랫폼 운영자에게 도움이 될 것으로 기대한다.

II. 크라우드 펀딩

크라우드 펀딩(Crowd funding)은 플랫폼을 통해 회사 또는 개인 사업가 등이 일반 대중들로부터 자금을 지원 받는 형태로 판매자와 일반 대중 간의 상호 작용이 특징이다(Nielsen and Binder 2021). 크라우드 펀딩은 최근 소규모 창업자의 증가 및 온라인 판매가 활성화되면서 각광받고 있다. 제품이나 서비스 개발 과정에서 대중으로부터 아이디어와 피드백을 수집하는 크라우드 소싱

(Crowd sourcing)의 한 형태로 시작되었으며, 펀딩에 참여한 서포터와 이익을 공유한다. 일반적으로 비금융형인 기부형, 보상형과 금융형인 대출형 및 증권형의 4가지 모델로 분류한다(Belleflamme, Lambert and Schwienbacher 2014; Short et al. 2017; 윤민섭 2017).

그중 우리나라에서는 펀딩 참여 대가로 유형이나 무형의 상품을 제공하는 보상형 크라우드 펀딩 플랫폼이 급성장했으며, 2019년에는 전년 대비 168.2% 성장하여 2,103억 원에 이르고 있으며(이구형 2020), 그 이후에도 증가세를 이어가고 있다고 판단된다.

크라우드 펀딩은 강력한 자금 조달 도구일 뿐만 아니라 아이디어 개발, 잠재 고객과의 관계 촉진, 효율적인 마케팅 도구 등 다양한 목적을 제공한다(Gerber et al. 2012; Hui et al. 2012). 보상형 크라우드 펀딩의 경우, All or Nothing 방식을 채택하고 있으며, 목표액을 달성하지 못할 경우, 프로젝트가 취소된다. 이러한 이유로 크라우드 펀딩 성공과 실패를 종속변수로 하여 성공에 영향을 미치는 요인에 대한 연구가 이루어져 왔다.

성공적인 크라우드 펀딩은 일반적으로 캠페인 성공을 결정하는 주요 기준으로 작용하는 자금 조달 목표액을 달성하는 것과 관련된다. 이전 연구에 따르면 이러한 성공에 기여하는 요인으로 목표액이 높을수록, 펀딩 기간이 길수록 성공하기 어렵다는 것이 확인되었다(Mollick 2014; Kuppuswamy and Bayus 2018). 반면, 이미지가 많아질수록(Koch and Siering 2015), 프로젝트

〈표 1〉 크라우드 펀딩 종류

| 유형 | | 반대급부 |
|------|-----|-----------------|
| 비금융형 | 보상형 | 無/기념품 |
| | 기부형 | 연계성 있는 재화, 공연 등 |
| 금융형 | 대출형 | 이자 |
| | 증권형 | 이자 또는 수익 |

출처: 윤민섭 2017

업데이트 비율 및 프로젝트를 설명하는 비디오 수가 많을수록 펀딩 성공과 양의 상관관계가 있는 것으로 나타났다(Mollick 2014). 국내 크라우드 펀딩 플랫폼인 텀블벅을 대상으로 한 연구에서도 동영상, 시각 이미지가 펀딩 성공에 유의한 영향을 미치는 것으로 나타났다(이정은, 신형덕 2014).

그 외에도 초기 투자자 참여 확보(Vismara 2018; Fan et al. 2020)가 펀딩 성공에 영향을 미치는 것으로 나타났는데, 예비 서포터들은 이미 상당한 자금을 조달한 프로젝트에 투자하는 경향이 더 높은 밴드웨곤(BandWagon) 행동을 보였다. 이는 사람들이 불완전한 정보로 결정을 내릴 때 다른 사람의 행동을 관찰하는 경향이 있기 때문이다(Simonsohn and Ariely 2008).

정형 데이터에 대한 연구와 함께 비정형 데이터인 프로젝트를 설명하는 텍스트에 대한 연구도 진행되어 왔다. 초기에는 텍스트의 수나 길이 등의 수치적 정보만을 사용하였다. 이를 통해 많은 양의 텍스트가 펀딩 성공에 긍정적인 영향을 미친다는 것이 확인되었다(Wang et al. 2011; Zhou et al. 2014; Pitschner and Pitschner-Finn 2014; 이정은, 신형덕 2014; Barbi and Bigelli 2015; Koch and Siering 2015). 이후 텍스트 마이닝을 활용하여 게시글 감정분석, 가독성을 분석하거나 주제

모델링을 통해 펀딩 성공을 예측했다(Du et al. 2015; Yuan et al. 2016). Nam et al.(2018)의 연구에서 수치적 정보와 더불어 텍스트에서 추출한 감정, 가독성, 지역 정보, 명사의 수를 변수로 포함하여 성공을 예측했을 때가 수치적 정보만 포함되어 있을 때보다 프로젝트 성공에 미치는 영향이 더욱 유의한 것으로 나타났다. 또한 성공과 실패 이진분류로만 성과를 예측했던 이전 연구들과 달리 모금액을 로그화한 값을 종속변수로 하여 회귀모델을 돌렸을 때도 이전과 비슷한 결과를 얻었다.

결과적으로 이전 연구들을 바탕으로 크라우드 펀딩 전략을 수립할 때 프로젝트 제작자는 자금 조달 목표와 일정이 현실적인지 확인하고, 멀티미디어 콘텐츠로 잠재적 후원자를 참여시키고, 많은 초기 투자자를 확보해야 한다는 것이 확인되었다. 그리고 설득력 있는 문장으로 프로젝트를 설명하는 것도 크라우드 펀딩 성공 가능성을 높일 수 있는 요인이었다.

대부분의 펀딩 연구에서는 종속변수로 성공과 실패 두 가지로 분류하고 있으나 보상형 펀딩의 경우에는 성공과 실패에 따라 보상품의 제공이 결정되기에, 목표 펀딩금액을 낮추는 경향이 있는 것으로 판단되어, 본 연구는 펀딩에 참여하는 서포터 수를 종속변수로 하여 크라우드 펀딩의 결과를 좀 더 현실적으로 분석하고자 한다.

〈표 2〉 크라우드 펀딩 연구들

| 구분 | 저자 | 변수 |
|---------------|---|---|
| 수치 데이터 활용 연구 | Kuppuswamy and Bayus 2018 | 목표액, 펀딩기간 |
| | Mollick 2014 | 목표액, 펀딩기간, 이미지 수, 프로젝트 업데이트 비율, 프로젝트 설명 비디오 수 |
| | 이정은, 신형덕 2014; Koch and Siering 2015 | 동영상, 시각 이미지 |
| | Vismara 2018; Fan et al. 2020 | 초기 투자자 참여 |
| 텍스트 데이터 활용 연구 | Wang, Liu and Fan 2011; Zhou et al. 2014; Pitschner and Pitschner-Finn 2014; 이정은, 신형덕 2014; Barbi and Bigelli 2015; Koch and Siering 2015 | 텍스트의 수, 길이 |
| | Du et al. 2015; Yuan, Lau and Xu 2016 | 게시글 감정 분석, 가독성, 주제 모델링 |
| | Nam, Jin and Kwon 2018 | 감정, 가독성, 지역 정보, 명사의 수 |

III. 딥러닝

1. 개념 및 구조

딥러닝(Deep Learning)은 등장할 때부터 상당한 주목을 받은 접근 방식으로 머신러닝 분야의 혁신적인 방법론이다(Hinton and Salakhutdinov 2006). 딥러닝 기술은 계층적 학습 알고리즘을 구성하기 위해 뇌의 구조와 기능에서 영감을 얻었으며, 계층 구조의 정보 처리 단계의 많은 계층이 패턴 분류와 기능 또는 표현 학습에 활용되는 기계 학습 기술의 한 종류를 말한다(Markoff 2012). 딥러닝은 수 많은 독립변수들을 복잡한 모수들과 선형 및 비선형의 함수를 만들어 다양한 종속변수들의 값들을 만들어내는데, Universal Approximation 이론에 따르면 모든 종류의 가능한 함수를 거의 정확하게 만들어낼 수 있는 유연한 함수라고 볼 수 있다.

딥러닝을 활용한 방법론은 많은 숫자 정보를 활용해서 하는 분류 및 예측분야 이외에도 광범위한 영역에서 응용되어 왔다. 예를 들어 이미지 인식(computer vision)은 딥러닝 기술의 이점을 크게 활용하여 시각 데이터를 보다 정확하고 정교하게 분석할 수 있다(Tompson et al. 2014; Szegedy et al. 2014). 마찬가지로, 음성 인식 분야는 딥러닝 알고리즘의 통합으로 상당한 개선을 목격하여 인간 음성을 보다 정교하게 분석해왔다(Mikolov et al. 2011; Hinton et al. 2012).

또한 딥러닝은 언어분야(natural language processing, NLP) 주제 분석, 감정 분석(Bordes, Chopra and Weston 2014) 및 언어 번역(Jean et al. 2015) 및 문장생성이나 대화생성을 넘어서 인간과의 모든 지식을 나눌 수 있는 대형언어처리모델(Large Language Model, LLM)으로 발전하고 있으며, 다른 딥러닝이나 인공지능을 활용한 자동화모델(automation AI)과 사람이 소통하는데 도

움을 주는 기본모델(foundation model)이 되고 있다.

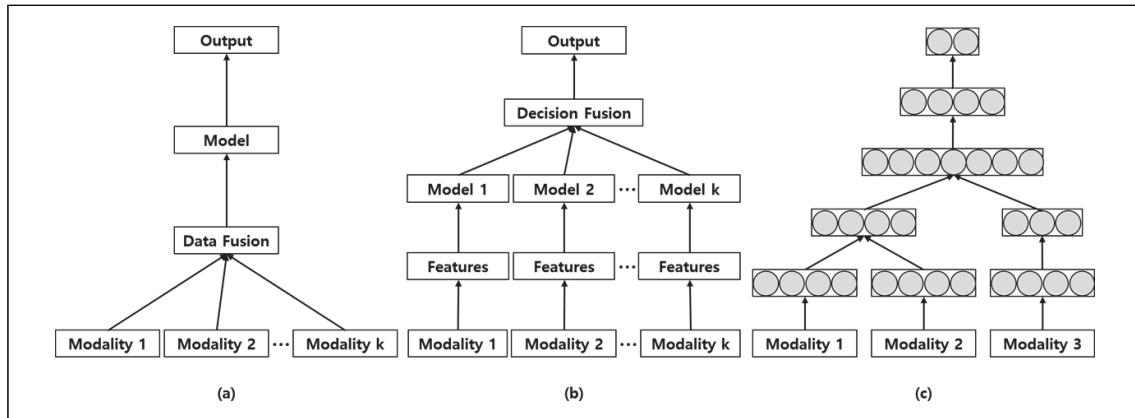
클라우드 펀딩 분야가 아닌 전체 마케팅 분야에서는 최근들어 딥러닝을 활용한 적지 않은 연구들이 진행되고 있다. 특히 댓글에 NLP를 활용하여 소비자에게 의미있는 메시지를 만들 수 있는 도구를 제공하는 연구(Hong and Hoban 2002)도 있고, 브랜드의 로고를 시각정보로 활용하고, 문자나 평가자료를 역시 포함하여 multi-modal VAE모델을 가지고 분석하여 brand arithmetic을 통해 브랜드를 평가하고 소비자 반응을 연구하기도 하고(Dew, Ansari, and Toubia 2021), 소비자들의 온라인 콘텐츠를 사용하는 정보를 분석하기 위하여 matrix factorization을 neural network로 찾아내는 연구(Dhillon and Aral 2021)도 있다. 또 딥러닝모델은 아니지만 선택자료에서 문자정보와 평가정보를 함께 컨조인트분석을 한 연구(Kim and Allenby 2022)들도 있다.

마케팅영역에서는 실무적으로도 이미지 인식 분야에서 마트 진열장의 재고확인, 구매예정인 의류를 입은 모습 보여주기, 자료 분석 분야에서는 제품이나 영화 및 음악, 기사 등을 추천하기, 소비자의 마케팅 활동 정보의 분석, 이탈고객분석, 시장세분화, 판매예측 등에서 활용되어왔다. 하지만 최근 생성형AI 모델이 급부상하면서 광고문구, 광고포스터, 광고음악과 광고동영상 등 모든 분야에서 활용 가능성이 높아지고 있다.

2. 멀티모달(Multimodal) 딥러닝

소비자와의 다양한 커뮤니케이션을 할 때 마케팅담당자는 다양한 미디어와 도구를 사용하게 되기 때문에 이미지, 동영상, 텍스트 등 여러 가지 유형의 데이터를 함께 분석하고 처리하는 멀티모달 딥러닝(Multimodal Deep Learning)모델이 필요하게 되었다. 각 데이터 유형은

〈그림 1〉 멀티모달 데이터 처리 모델



출처: Ramachandram and Taylor 2017

고유한 정보를 제공하며, 마케팅담당자가 이를 효과적으로 활용하면 보다 풍부한 표현을 통해 더 좋은 커뮤니케이션을 할 수 있다(Mao et al. 2014).

딥러닝 모델 학습에 있어서 멀티모달 데이터 처리 방식에는 초기 융합(a), 후기 융합(b), 중간 융합(c) 세 가지의 접근법이 있다(Ramachandram and Taylor 2017). 초기 융합 모델은 데이터를 먼저 결합한 후 하나의 모델로 처리하며, 후기 융합이나 중간 융합 방식으로 처리하는 멀티 모델은 특성이 서로 다른 데이터를 각각 처리한 후 결합한다. 멀티모달 데이터 모델은 다양한 연구에서 싱글 모델에 비해 멀티 모델로 학습하는 것이 더 효과적이라고 한다(Atrey et al. 2010; Arevalo et al. 2020).

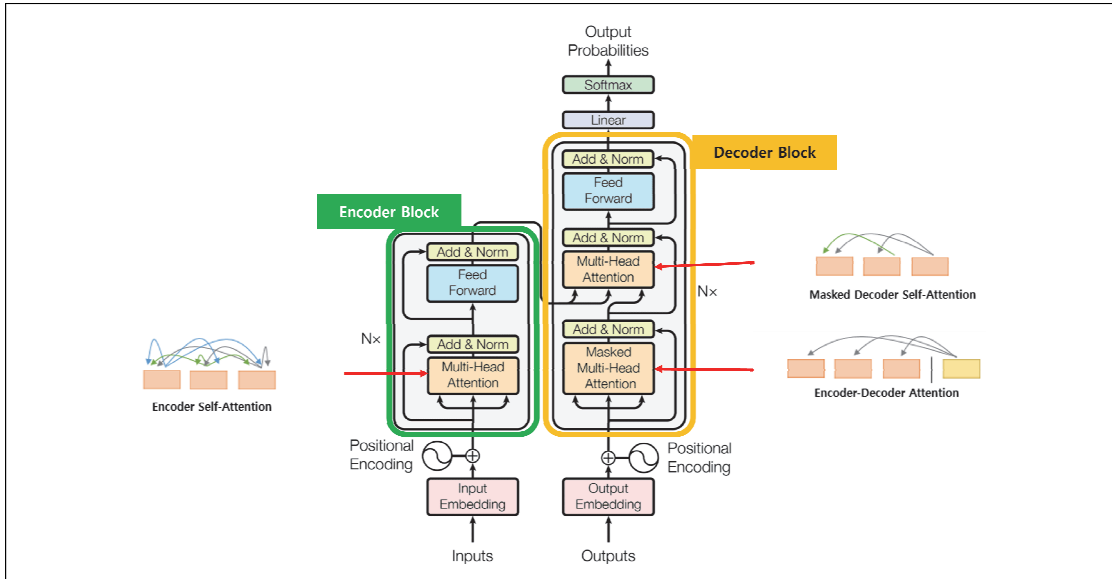
본 연구에서는 특징이 서로 다른 두 가지 유형의 데이터, 즉 수치 데이터와 텍스트 데이터를 분석하는 데 있어서, 초기 융합 방식을 통해 데이터를 직렬로 처리하는 방식과 후기 융합을 통해 데이터를 병렬로 처리하는 방식을 모두 분석하여 비교한다. 각각의 처리 방식 간에 어떤 차이점이 있는지를 이해하고 더 나은 표현 학습과 결과 예측에 있어서 가장 효과적인 데이터 처리 방식을 확인하고자 한다.

3. 트랜스포머 모델

사전 훈련된 언어 모델은 많은 자연어 처리 작업을 개선하는 데 효과적인 것으로 입증되었다(Dai and Le 2015; Peters et al. 2018; Radford et al. 2018; Howard and Ruder 2018). 2017년 Google에서 개발한 트랜스포머 모델이 발표되면서 자연어 처리에서 대표모델이 되었다. 이전까지 활용되던 RNN(Recurrent Neural Network) 모델이 순서를 기반으로 했던 것과 달리 Self-Attention 메커니즘을 활용하여 문장 속 모든 단어들 간의 상호작용을 파악하여 문장의 맥락을 이해할 수 있기 때문이다(Vaswani et al. 2017). 최근에는 자연어 처리뿐만 아니라 이미지 분류 등의 여러 분야에서 활용되고 있다.

트랜스포머는 오토인코더(autoencoder) 모델과 유사하게, 인코더와 디코더(Decoder)로 구분할 수 있다. 인코더는 입력문장의 의미를 학습하고, 그 결과를 디코더에 전달하면 디코더는 훈련시에는 이것과 출력단어들을 가지고 그다음 출력단어를 예측하는 것을 학습하며, 실제 활용시에는 인코더의 결과를 가지고 다음 단어를 순차적으로 만들어내는 모델이다. 하지만 인코더와 디코더를

〈그림 2〉 트랜스포머 구조



출처: Vaswani et al. 2017; 딥러닝을 이용한 자연어 처리 입문(<https://wikidocs.net/31379>)

모두 활용하는 모델도 있지만 인코더 부분만을 활용하는 BERT(Bidirectional Encoder Representations from Transformers)와 디코더 부분을 활용하는 GPT (Generative Pre-trained Transformer)가 더 많이 활용되고 있다(Devlin et al. 2018; Radford and Narasimhan 2018).

트랜스포머의 디코더를 활용한 GPT 또는 기타 NLP 모델은 단방향인 반면, BERT는 양방향 셀프 어텐션(self-attention)을 사용한다. 이 메커니즘을 활용하는 BERT는 단어의 앞이나 뒤뿐만 아니라 전체 문장을 이해하고 텍스트에 포함된 정보를 처리, 재구성 및 추출하는데 탁월하다(Devlin et al. 2018). 하지만 실제로 훈련시키기도 어려운 GPT계열의 모델들(예, ChatGPT, LaMDA)이 다음 단어를 만들어내는 생성형AI로서 능력을 인정받아 〈그림 2〉에서 보는 layer들의 숫자와 입력자료의 차원을 크게 하는 등, 모델의 크기를 매우 크게 하고 수많은

자료(token)들로 모델의 모수를 학습한 pre-trained 된 LLM으로서의 많이 활용되고 있다. 본연구에서는 트랜스포머의 인코더 부분만을 활용하여 수집한 자료를 활용하여 직접 모델의 모수들을 추정된 모델을 운영하고자 한다.

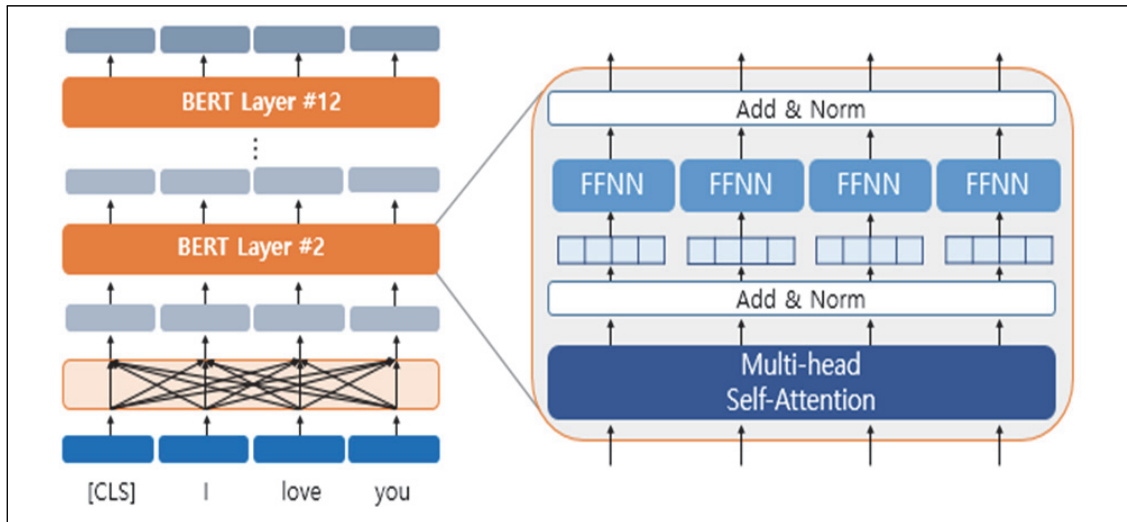
4. BERT

BERT는 트랜스포머의 인코더 부분을 활용하여 개발된 언어 모델로서 Wikipedia 또는 책 데이터와 같은 대규모 데이터 세트에서 훈련되었으며 각각 다양한 매개 변수가 있는 BERT Base(L=12, H=768, A=12, Total Parameters=110M) 및 BERT Large(L=24, H=1024, A=16, Total Parameters=340M¹⁾)의 두 가지 모델 크기로 제공된다(Devlin et al. 2018).

BERT Layer에서는 멀티헤드어텐션과 포지션-와이드

1. L : 트랜스포머 block의 layer 수, H : hidden size, A : self-attention heads 수

〈그림 3〉 BERT의 모델 구조



출처: 딥러닝을 이용한 자연어 처리 입문(<https://wikidocs.net/31379>)

피드 포워드 신경망을 수행하며, 입력된 단어 임베딩은 BERT 연산을 거치면 문장의 문맥을 모두 반영한 벡터로 출력된다. 이렇게 출력한 벡터는 다양한 목적의 다운스트림(down stream) 모델에 입력자료로 활용될 수 있다.

다른 NLP 모델과 비교하여 BERT는 질문 답변, 의미 추론 및 요약 등을 포함한 광범위한 자연어 이해 작업에서 우수한 결과를 달성했다(Devlin et al. 2018; Gabriel et al. 2019; Yoshimura et al. 2019; Zhang et al. 2019). BERT를 사용한 텍스트 정보 추출은 단순히 텍스트의 특정 패턴이나 키워드를 기반으로 정보를 추출하는 것보다 훨씬 뛰어난 성능을 보이며, 문장 의미를 압축하거나 요약한 정보 포함하는 벡터(embedding vector)를 생성한다. 이를 활용하여 BERT는 다른 모델에 입력자료로 활용되거나 벡터 생성 후에 몇 개의 layer를 추가하여 분류작업이나 의미추출작업 혹은 유사성 비교 작업 등을 수행할 수 있다(Reimers and Gurevych 2019).

본 논문에서는 Python 라이브러리인 Huggingface의

BERT 기본 uncased pre-trained 모델을 사용하였다(Wolf et al. 2020). 이 모델에서 단어들로 이루어진 문장의 의미를 반영하는 벡터를 출력하였으며, 모델에 활용하였다.

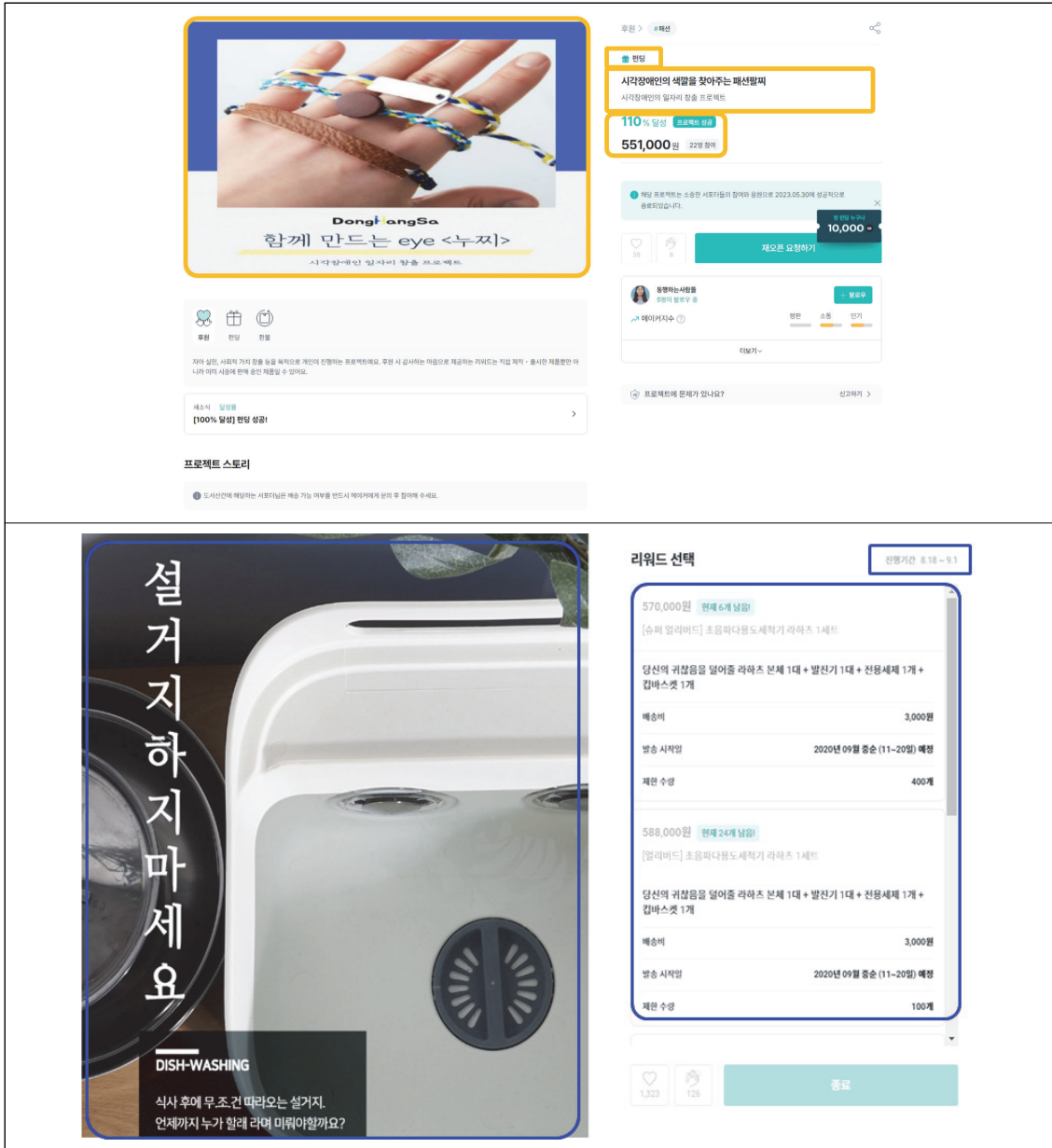
IV. 크라우드 펀딩 성공 예측 모델

1. 데이터 확보 및 특성

데이터는 국내 대표적인 펀딩 사이트인 와디즈(Wadiz)에서 수집하였다. 와디즈는 2014년 크라우드 펀딩 서비스를 시작했으며, 2020년 기준 약 2천억 원의 펀딩을 성사시킨 한국 1위 크라우드 펀딩 플랫폼이다(와디즈 2021). 현재 펀딩, 프리오더, 스토어 서비스를 제공하고 있다.

본 연구에서 수집한 데이터는 프리오더 및 스토어를 제외하고 프로젝트가 종료된 2021년 1월부터 2023년 1월까지의 펀딩 프로젝트 9,755개의 url을 수집하였다. 총 5개 카테고리의 데이터를 수집하였으며, 테크·가전

〈그림 4〉 와디즈 홈페이지 내 프로젝트 상세페이지



출처: 와디즈 홈페이지(<https://www.wadiz.kr/>)

(1,083개), 패션·잡화(1,642개), 홈·리빙(2,575개), 푸드(3,200개), 뷰티(1,255개)에서 데이터를 수집하였다.

와디즈 사이트의 상세 페이지는 〈그림 4〉와 같으며, 크롤링을 통해 각 프로젝트 별로 카테고리, 펀딩 이름, 펀딩 설명글, 본문 텍스트, 펀딩 기간, 펀딩 금액, 펀딩 서

포터 수, 달성률, 옵션 가격, 옵션 개수, 메인 이미지 수, 메인 동영상 수, 본문내 이미지 수, 본문내 동영상 수, 스크롤 길이, 일일 펀딩 금액 데이터를 수집하였다.

수집한 데이터를 바탕으로 탐색적 분석을 시행하였다. 먼저 수집한 프로젝트 중 달성률이 100% 이상으로 성

공인 프로젝트는 8,772개, 100% 미만인 실패한 프로젝트는 983개로 나타났다. 다음으로 수집한 변수들의 기술 통계량을 분석하였다. <표 3>에서 연구에서 활용하는 주요 변수들의 기본적인 통계량을 보여준다.

서포터 수, 총 모금액, 가격, 스크롤 길이는 상당한 표준편차를 보여주고 있다. '서포터 수' 변수는 평균(Mean = 179.64)에 비해 높은 표준편차(SD = 378.23)를 보인다. 이는 캠페인마다 참가자 수가 크게 다르다는 것을 나타내며, 이는 프로젝트의 성격 등에 따라 달라지기 때문에 프로젝트의 결과를 가장 잘 반영한다고 보아 본 논문에서 종속변수로 활용하였다. '총 모금액' 변수 역시 '서포터 수'에 따라 달라지므로 큰 표준편차(SD = 3.36e+07)를 가지고 있으며, 평균(Mean = 1.27e+07)

에 비해 크다. 본 연구에서 서포터 수를 종속변수로 써서 총 모금액을 활용하지 않았지만, 두 변수의 상관관계는 0.70으로 높았다.

메인 IMG 수는 landing page의 이미지 수를 의미한다. 본문내 IMG나 JPG도 수집은 하였으나, 내용에 이미지뿐만 아니라 텍스트를 이미지화 한 것들이 섞여 있고, 이미지 안의 텍스트를 구글 OCR을 활용해서 추출하였으나 내용적으로 메인페이지의 'title'와 'description'에 있는 텍스트 정보와 겹치는 것이 많아서 실제 분석에서는 활용하지 않았다.

본 연구에서는 초기 서포터 수를 고려한 예측 모델링을 추가적으로 수행하였다. 이를 위해, 펀딩 시작 날부터 14일 이후까지, 즉 초기 2주 동안의 일별 서포터 수

<표 3> 변수의 기초 통계량

| Variable | N | Mean | SD | Min | Median | Max |
|-----------------|-------|----------------------------|-----------|----------|-----------|------------|
| 펀딩 카테고리 | 9,755 | 테크·가전, 패션·잡화, 홈·리빙, 푸드, 뷰티 | | | | |
| 서포터 수 | 9,755 | 179.64 | 378.23 | 1 | 72 | 9,787 |
| 총 모금액 | 9,755 | 1.27E+07 | 3.36E+07 | 1.58E+04 | 3.82E+06 | 9.69E+08 |
| 프로젝트 기간 | 9,755 | 19.51 | 8.04 | 5 | 19 | 84 |
| 가격 | 9,755 | 1.02E+05 | 1.71E+05 | 2.75E+03 | 6.02E+04 | 7.59E+06 |
| 스크롤 높이 | 9,755 | 40,856.72 | 21,670.24 | 1,906.00 | 37,199.00 | 237,469.00 |
| 옵션 수 | 9,755 | 6.45 | 4.79 | 1 | 6 | 260 |
| 메인 IMG 수 | 9,755 | 2.82 | 1.47 | 1 | 4 | 4 |
| 메인 JPG 수 | 9,755 | 2.27 | 1.67 | 0 | 2 | 4 |
| 메인 GIF 수 | 9,755 | 0.44 | 0.93 | 0 | 0 | 4 |
| 메인 YouTube 수 | 9,755 | 0.1 | 0.3 | 0 | 0 | 1 |
| 콘텐츠 GIF 수 | 9,755 | 9.02 | 10.92 | 0 | 6 | 152 |
| 콘텐츠 YouTube 수 | 9,755 | 0.24 | 0.8 | 0 | 0 | 12 |
| title 문자수 | 9,755 | 35.65 | 5.72 | 3 | 38 | 40 |
| description 문자수 | 9,755 | 80.42 | 20.11 | 1 | 88 | 100 |

<표 4> 일별 서포터수 히트맵 분석 결과

| 날짜 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|-----|-------|-------|------|------|------|------|------|------|------|------|------|------|------|------|
| 성공 | 38.66 | 12.13 | 7.96 | 5.32 | 4.56 | 4.38 | 4.32 | 4.10 | 3.80 | 3.79 | 3.62 | 3.56 | 3.49 | 3.52 |
| 실패 | 3.80 | 1.22 | 0.63 | 0.40 | 0.35 | 0.28 | 0.27 | 0.27 | 0.20 | 0.19 | 0.18 | 0.14 | 0.15 | 0.12 |
| 차이 | 34.86 | 10.91 | 7.33 | 4.92 | 4.21 | 4.10 | 4.05 | 3.83 | 3.60 | 3.60 | 3.44 | 3.42 | 3.34 | 3.40 |
| 유의도 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

를 분석하였다. 성공여부에 따라 일별로 서포터 수의 차이를 본 결과, 통계적으로 모두 유의한 것으로 나타났다. 특히 초기에 이미 성공할 것과 실패할 것이 명확하게 구분되는 것을 알 수 있었다. 따라서 초기 3일 동안의 서포터 수를 추후에 모델에 포함하게 되었다.

2. 전처리

수집된 데이터를 각각의 데이터 특성에 맞게 다음과 같이 데이터 전처리를 수행했다.

로그변환 및 표준화

수치 데이터인 '서포터 수', '가격', '스크롤 높이', '옵션 수', '콘텐츠 GIF 수' 변수들은 로그 변환을 적용하였으며, 펀딩 시작 후 1일부터 3일까지의 서포터 수는 로그 변환하였다. 데이터 간 변동성이 비교적 적은 변수는 로그 변환하지 않았다. 데이터의 특성들이 서로 다른 범위에 있을 수 있으므로, 데이터 특성들을 표준화하였다.

interaction term 생성

수치 데이터를 각각 두 개씩 곱하여 새로운 변수를 생성하였다. 생성한 데이터는 총 15개이며, 'log_스크롤 높이_가격', 'log_옵션 수_가격', 'log_콘텐츠 GIF 수_가격', '메인 IMG 수_가격', '메인 JPG 수_가격', 'log_옵션 수_스크롤', 'log_콘텐츠 GIF 수_스크롤', '메인 IMG 수_스크롤', '메인 JPG 수_스크롤', 'log_콘텐츠 GIF 수_option', '메인 IMG 수_option', '메인 JPG 수_option', '메인 IMG 수_congif', '메인 JPG 수_congif', '메인 JPG 수_mainimg'를 생성하였다. 이는 서로 다른 변수 간의 상호작용 효과를 반영하여 데이터 내 관계를 파악하고 숫자 간 상호작용이 종속변수에 어떤 영향을 미치는지 분석하여 데이터를 포괄적으로 이해할 수 있다. 단,

모든 변수들에 대하여 상호작용 변수를 만든 것이 아니라 변수의 값들이 거의 한가지 값을 갖는 변수들은 다른 변수와 상호작용변수를 만드는 것이 의미가 없어서 6개의 변수에 대해서만 만들었다.

'category 변수' one-hot encoding

명목형 변수 데이터인 펀딩 카테고리(Category)는 총 5개이며, 원핫인코딩(One-hot-encoding)을 통해 더미 변수로 변환하였다. category1부터 category5까지의 변수를 생성하여 각각의 카테고리에 해당되면 1, 아니면 0으로 입력하였다.

분류모델 적용을 위한 '서포터 수' 변수의 카테고리 변수화

또한 '서포터 수'를 분류모델에서 종속변수로 활용하기 위해서 로그 변환한 서포터 수를 구간 경계값을 3, 4, 5, 6을 기준으로 나누고 총 5개의 범주로 분류하였다.

학습 세트와 테스트 세트

모델링에 앞서 데이터를 학습 세트와 테스트 세트로 분할하였다. 학습 세트는 모델을 학습하는 데 사용되며, 테스트 세트는 학습된 모델의 성능을 평가하는 데 사용된다. 데이터는 무작위로 80:20의 비율로 분할하였다(학습 세트: 80%, 테스트 세트: 20%).

3. 클라우드 펀딩 성공 예측 모델

데이터의 분석과 전처리 이후, 본 연구에서는 총 5단계로 다양한 모델을 구축하여 비교하고자 하였다.

1) 먼저 총 서포터 수를 예측하는 딥러닝 예측모델과 5개로 분류한 서포터 수 구간을 예측하는 딥러닝 분류모

〈표 5〉 딥러닝 모델 구조 구축

| | BERT | | 트랜스포머 인코더 | |
|------|----------|----------|-----------|--------------|
| | 출시전 자료들 | | | 출시후 일별 자료 추가 |
| | 직렬 | 병렬 | | |
| 예측모델 | 숫자+곱 | 숫자+곱 | 좌동 | |
| | 숫자+텍스트 | 숫자+텍스트 | | |
| | 숫자+곱+텍스트 | 숫자+곱+텍스트 | | |
| 분류모델 | 숫자+곱 | 숫자+곱 | | 좌동 |
| | 숫자+텍스트 | 숫자+텍스트 | | |
| | 숫자+곱+텍스트 | 숫자+곱+텍스트 | | |

델을 설계하였다. 예측모델은 서포터 수를 예측하는 진통 통계에서 진행되는 회귀모형과 유사한 형태의 모델이다. 실제 사용하는 activation 함수도 선형을 활용하고 loss 함수도 Mean Square Error(MSE)를 사용한다. 딥러닝 모델이 많이 활용되는 분류모델을 활용해서 서포터 수를 맞추어보고자 서포터 수를 실제 자료에 맞도록 5개 집단으로 인위적으로 구분하여서 분류모델을 적용해보고자 하였다. 분류모델에서는 5개 집단을 맞추기 위하여 activation 함수는 softmax를 활용하였고 loss 함수는 일반적인 cross entropy를 활용하였다. 이는 예측모델과 분류모델의 성능을 비교하기 위한 목적이다.

두 개의 모델의 마지막 layer를 제외하고는 다수의 Dense layer와 BatchNormalization layer, Dropout layer로 구성된 순차적인(Sequential) 구조를 사용하여 과적합을 방지하였다. 중간 layer들의 activation 함수들은 'relu'를 활용하고 'He' 방법을 통한 초기 모수값을 사용하였다.

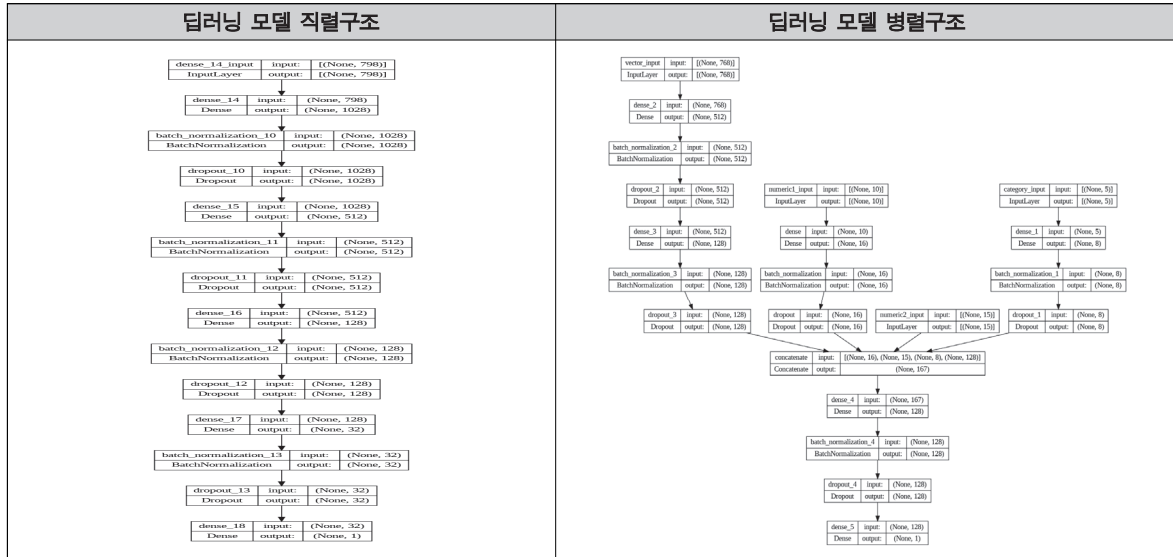
2) 입력변수들을 3가지 features(숫자 변수, 숫자 변수들의 곱, 텍스트 변수)로 생각하여 이들간의 3가지 조합(숫자+곱, 숫자+텍스트, 숫자+곱+텍스트)을 만들어 각각의 모델을 학습하여 각 feature의 영향력을 판단하고자 하였다. 여기서 숫자는 숫자 변수들을 의미하며, 텍스트는 텍스트 변수들을 총칭하는 말이다.

3) 각각의 모델링에 활용된 멀티모달 데이터를 학습하는 방식을 직렬방식과 병렬방식으로 구분하여 비교 분석하였다. 특히, 병렬방식에서는 수치 데이터, 원핫인코딩한 명목형 변수, 텍스트 데이터를 각각의 특성에 맞춰 최적화하는 레이어를 거쳐 처리한 후 합치도록 구조화하였다.

4) 다음은 초기 서포터 수를 변수로 추가하여 딥러닝 예측모델과 딥러닝 분류모델을 분석하였다. 이는 초기 참여자 수가 캠페인의 최종 성공에 큰 영향을 미칠 수 있다는 기존 연구들을 바탕으로 설계된 모델로 모델 구조는 이전의 구조에 변수를 추가한 것이며, 다른 구조는 동일하다.

5) 마지막으로 텍스트 변수 처리 방식을 BERT 방식이 아닌 트랜스포머 인코더를 직접 학습하는 모델링을 진행하였다. 4)번까지의 모델 구축에서 활용한 텍스트 변수는 BERT 모델을 사용하여 768개의 벡터로 변환한 뒤, 이를 Dense 레이어에 입력하는 방식이다. 반면, 일반적인 단어들로 사전 학습된 BERT 모델이 가지는 한계점을 인식하고, 본 연구를 위한 자료를 활용해서 OKT 형태소 분석기를 사용하여 텍스트 데이터를 토큰화하고 이를 정수 인코딩한 텍스트 데이터는 트랜스포머 모델의 인코더에 넣는 모델을 운영해보았다. 이는 텍스트 변수

〈그림 5〉 딥러닝 모델의 직렬구조와 병렬구조



의 처리 방식에 따른 모델 성능을 비교하고자 하였다.

BERT를 통한 text embedding 값 추출

첫 번째 방법에서는 BERT의 풀링 레이어를 활용하여 768개의 벡터를 추출하였다. 그러나 BERT의 한국어 버전인 KoBERT는 pre-train된 단어의 수가 약 3,000 개 정도로 적어 텍스트 내용의 대부분을 반영하지 못하였다. 이에 한글로 된 텍스트 데이터를 Google 번역을 활용하여 영어로 번역하였으며, 이를 BERT Base('bert-base-uncased')를 활용하여, 768개의 벡터로 추출하였다. 이 벡터들은 문장 내 모든 단어간 상호작용을 통해 텍스트의 모든 의미 정보를 포함하고 있으며, 이를 Dense 레이어를 포함한 딥러닝 예측모델과 딥러닝 분류 모델링에 활용하였다.

트랜스포머 인코더 함수

두 번째 방법에서는 Open Korean Text(OKT) 토큰라이저를 사용하여 한국어 텍스트를 토큰화하였다. OKT는 한국어 문장을 개별 단어나 형태소로 분리할 때 활용

하는 인기 있는 토큰라이저이다. 토큰화 후에는 10번 이상 출현한 단어를 기반으로 4,588개의 어휘를 생성하고 각 토큰을 해당하는 정수 임베딩으로 변환하였다. 이 정수 임베딩들은 첫 번째 방법에서 사용한 Dense 레이어가 아닌 트랜스포머 인코더에 입력하여 분석을 진행하였다.

V. 모델 결과 및 평가

1. 기초분석

딥러닝 모델 분석에 앞서 종속변수인 서포터 수와 독립변수 간의 상관관계를 분석하였다. 분석 결과, log 콘텐츠 GIF가 0.249로 비교적 강한 양의 상관관계를 나타냈다. 자료의 수가 많기 때문에 유의성은 모두 높게 나타나기에 상관관계의 크기를 보고 영향의 정도를 파악해야 한다. 반면, 옵션 가격, 메인 IMG와 JPG는 서포터 수와 낮은 음의 상관관계를 보이고 있다.

〈표 6〉 독립변수와 종속변수 간 상관관계 분석

| | 서포터 수 | 유의도 |
|---------------|--------|-------|
| 프로젝트 기간 | 0.102 | 0.000 |
| log_가격 | -0.012 | 0.248 |
| log_스크롤 높이 | 0.171 | 0.000 |
| log_옵션 수 | 0.132 | 0.000 |
| log_콘텐츠 GIF 수 | 0.249 | 0.000 |
| 메인 IMG 수 | -0.027 | 0.008 |
| 메인 JPG 수 | -0.104 | 0.000 |
| 메인 GIF 수 | 0.131 | 0.000 |
| 메인 YouTube 수 | 0.042 | 0.000 |
| 콘텐츠 YouTube 수 | 0.034 | 0.001 |

〈표 7〉 딥러닝 예측모델 성능 비교

| 구조 | 직렬 | | | | 병렬 | | |
|-----|-------|-------|-------|----------|-------|--------|----------|
| | 숫자 | 숫자+곱 | 텍스트 | 숫자+곱+텍스트 | 숫자+곱 | 숫자+텍스트 | 숫자+곱+텍스트 |
| MSE | 1.136 | 1.129 | 1.351 | 1.131 | 1.123 | 1.073 | 1.078 |
| R2 | 0.294 | 0.298 | 0.160 | 0.296 | 0.302 | 0.332 | 0.330 |

2. 딥러닝 예측모델과 전통모델과의 비교

본 연구에서는 딥러닝 예측모델의 성능을 평가하기 위해 모델의 Mean Squared Error(MSE)와 R Squared (R2) 값을 활용하였다. MSE와 R2 값을 활용하여 각 모델을 비교한 결과, 직렬 구조에서는 숫자 데이터를 입력값으로 넣은 데이터의 성능이 높았다. 그리고 직렬 구조에 비해 병렬 구조를 갖는 모델의 성능이 더 높게 나타났다.

이외에도 딥러닝 모델과 기존의 전통적인 통계모델과 비교해보았다. 가장 기본이라고 할 수 있는 선형회귀분석과 숫자자료 혹은 Tabular 자료들에 대해서는 딥러닝 모델들보다도 성능이 좋다고 알려진 XGBoost 모델을 숫자와 숫자+곱 자료에 대하여 적용해보았다. 아래

〈표 8〉에서 보면, 현재 딥러닝의 직렬인 예측모델이 MSE와 R2에서 선형회귀분석과 XGBoost의 성능보다 좋은 것을 알 수 있다.

병렬 구조를 갖는 모델 중에서도, 숫자와 텍스트 데이터를 병렬로 처리한 모델이 가장 높은 성능을 보였다. 이는 각 데이터의 특성을 반영하여 높은 성능을 보이기 위해서는 각 데이터의 특성에 따라 모델링을 해야 한다는 것을 의미한다. 이를 확인하기 위해 병렬모델을 활용하여 각 모델을 30번씩 반복적으로 실행하여 평균 성능을 계산하였고, 이후 t-test를 진행하여 성능 차이가 통계적으로 유의미한지를 확인하였다. 분석 결과, 모든 모델간 차이가 유의한 것으로 나타났으며 숫자, 곱, 텍스트를 모두 입력값으로 넣은 모델의 성능이 가장 높은 것으로 나타났다.

〈표 8〉 딥러닝 예측모델 성능 비교

| LinearRegression | 숫자 | 숫자+곱 | xgboost | 숫자 | 숫자+곱 |
|------------------|-------|-------|---------|-------|-------|
| MSE | 1.157 | 1.154 | MSE | 1.254 | 1.236 |
| R2 | 0.28 | 0.282 | R2 | 0.22 | 0.231 |

〈표 9〉 딥러닝 예측모델 성능 비교(30번 평균)

| 구조 | 병렬 | | |
|------------------|--------------------------|----------------------|------------------------|
| | 숫자+곱 | 숫자+텍스트 | 숫자+곱+텍스트 |
| MSE | 1.113 | 1.073 | 1.057 |
| R2 | 0.302 | 0.327 | 0.337 |
| 유의도 (p-value) | (숫자+곱+텍스트)와의 차이 0.000 | (숫자+곱)과의 차이 0.000 | (숫자+텍스트)와의 차이 0.031 |

3. 딥러닝 분류모델

딥러닝 분류모델의 성능을 평가하기 위해 모델의 Accuracy와 Loss 값을 활용하였다. Accuracy의 값을 계산하는 식은 다음과 같다.

- $ACC = (TP + TN) / (TP + TN + FP + FN)$

- TP: True Positive. 실제 값이 Positive(긍정)인 대상 중에서 예측과 실제 값이 모두 Positive(긍정)로 일치한 대상의 수.
- TN: True Negative. 실제 값이 Negative(부정)인 대상 중에서 예측과 실제 값이 모두 Negative(부정)로 일치한 대상의 수.
- FP: False Positive. 실제 값이 Negative(부정)인 대상 중에서 예측 값이 Positive(긍정)인 대상의 수.

- FN: False Negative. 실제 값이 Positive(긍정)인 대상 중에서 예측 값이 Negative(부정)인 대상의 수.

Accuracy와 Loss 값을 활용하여 각 모델을 비교한 결과, 직렬 구조에서는 숫자 데이터만 입력값으로 넣은 데이터의 성능이 높았다. 그리고 직렬 구조에 비해 병렬 구조를 갖는 모델의 성능이 더 높게 나타났다.

병렬 구조를 갖는 모델 중에서도, 숫자와 텍스트 데이터를 병렬로 처리한 모델이 가장 높은 성능을 보였다. 이를 확인하기 위해 각 모델을 30번씩 반복적으로 실행하여 평균 성능을 계산하였고, 이후 t-test를 진행하여 성능 차이가 통계적으로 유의미한지를 확인하였다. 분석 결과, 숫자, 곱만 입력값으로 넣은 데이터와 숫자, 곱, 텍스트를 모두 입력값으로 넣은 데이터의 차이만 유의한

〈표 10〉 딥러닝 분류모델 성능 비교

| 구조 | 직렬 | | | | 병렬 | | |
|----------|-------|-------|-------|----------|-------|--------|----------|
| | 숫자 | 숫자+곱 | 텍스트 | 숫자+곱+텍스트 | 숫자+곱 | 숫자+텍스트 | 숫자+곱+텍스트 |
| Loss | 1.370 | 1.373 | 3.829 | 1.369 | 1.384 | 1.743 | 1.640 |
| Accuracy | 0.381 | 0.372 | 0.305 | 0.368 | 0.386 | 0.365 | 0.378 |

〈표 11〉 딥러닝 분류모델 성능 비교(30번 평균)

| 구조 | 병렬 | | |
|------------------|--------------------------|----------------------|------------------------|
| | 숫자+곱 | 숫자+텍스트 | 숫자+곱+텍스트 |
| Loss | 1.397 | 1.725 | 1.697 |
| Accuracy | 0.358 | 0.362 | 0.365 |
| 유의도 (p-value) | (숫자+곱+텍스트)와의 차이 0.038 | (숫자+곱)과의 차이 0.128 | (숫자+텍스트)와의 차이 0.451 |

것으로 나타났으며 숫자, 곱, 텍스트를 모두 입력값으로 넣은 데이터의 성능이 높았다.

4. 출시후 일별 자료 추가 모델

딥러닝 예측모델과 딥러닝 분류모델에 각각 초기 서포터 수 1일차부터 3일차 누적 값을 차례로 입력하였다. 여기서는 앞의 모델 비교 결과 가장 좋은 성능을 보여준 숫자+곱+텍스트 자료를 모두 넣은 병렬모델을 가지고 비교하였다. 그 결과 기존 연구들에서 초기 서포터 수가 성공 가능성을 높인다는 것과 마찬가지로 선행요인인 숫자와 텍스트 데이터만을 가지고 예측했을 때보다 편딩 시작 후 일별 서포터 수를 추가로 분석했을 때 설명력이 더 높게 나타났다. 또한 딥러닝 분류모델에서도 일별 서포터 수를 추가로 분석했을 때, 정확도가 약 2배 높아진 것으로 나타났다.

5. 텍스트 변수 처리 방식에 따른 모델

마지막으로 텍스트 데이터의 처리 방식(pre-train 된 BERT, 자체 학습시킨 트랜스포머 인코더)에 따른 모델을 평가한 결과, 텍스트 데이터만을 가지고 한 경우에는 트랜스포머 인코더를 통해 분석하고 서포터 수를 예측한 모델의 성능이 더 높게 나타났다. 그러나 숫자 데이터와 텍스트 데이터를 모두 입력한 모델에서는 BERT를 적용한 모델의 성능이 더 높게 나타났다. 이는 현재 텍스트 자료만으로는 본연구가 가진 자료로 직접 훈련한 트랜스포머 인코더가 더 잘 맞추지만, 다른 숫자 데이터와 가진 정보와 합쳐졌을 때는 정보가 겹쳐지는 현상이 일어나는 것으로 판단된다. 좀 더 많은 자료를 가지고 일반화를 할 수 있는 연구가 필요하다.

〈표 12〉 딥러닝 병렬모델에 초기 서포터 수를 추가한 모델 성능 평가

| 구분 | 평가 | 숫자+곱+텍스트 | 숫자+곱+텍스트+1일차 | 숫자+곱+텍스트+2일차 | 숫자+곱+텍스트+3일차 |
|------|----------|----------|--------------|--------------|--------------|
| 예측모델 | MSE | 1.078 | 0.466 | 0.290 | 0.248 |
| | R2 | 0.330 | 0.710 | 0.820 | 0.846 |
| 분류모델 | Loss | 1.640 | 1.189 | 0.954 | 0.886 |
| | Accuracy | 0.378 | 0.589 | 0.6472 | 0.679 |

〈표 13〉 딥러닝 예측모델의 텍스트 처리 방식에 따른 모델 성능 평가

| 텍스트모델 | BERT | | 트랜스포머 인코더 | |
|-------|-------|--------------|-----------|--------------|
| | 텍스트만 | 숫자+곱+텍스트(병렬) | 텍스트만 | 숫자+곱+텍스트(병렬) |
| 자료 | | | | |
| MSE | 1.351 | 1.078 | 1.301 | 1.133 |
| R2 | 0.160 | 0.330 | 0.221 | 0.295 |

〈표 14〉 딥러닝 분류모델의 텍스트 처리 방식에 따른 모델 성능 평가

| 텍스트모델 | BERT | | 트랜스포머 인코더 | |
|----------|-------|--------------|-----------|--------------|
| | 텍스트만 | 숫자+곱+텍스트(병렬) | 텍스트만 | 숫자+곱+텍스트(병렬) |
| 자료 | | | | |
| Loss | 3.829 | 1.640 | 6.040 | 4.882 |
| Accuracy | 0.305 | 0.378 | 0.324 | 0.304 |

VI. 결론

1. 연구 요약

본 연구에서는 클라우드 펀딩 성공을 예측하기 위한 딥러닝 모델을 설계하고 성능을 평가하였다. 다양한 모델 아키텍처와 변수 처리 방식을 비교 분석함으로써 다음과 같은 결론을 도출하였다.

첫째로, 딥러닝 예측모델의 성능 평가 결과, 직렬 구조보다 병렬 구조를 갖는 모델의 성능이 더 우수했다. 특히, 병렬 구조에서는 숫자 데이터와 텍스트 데이터를 병렬로 처리한 모델이 가장 높은 성능을 보였다. 이는 각 데이터의 특성을 고려하여 모델링을 수행하는 것이 성능 향상에 기여한다는 점을 나타낸다.

둘째로, 딥러닝 분류모델의 성능 평가 결과, 역시 병렬 구조를 갖는 모델이 직렬 구조보다 우수한 성능을 보였다. 특히, 숫자와 텍스트 데이터를 병렬로 처리한 모델이 가장 높은 정확도를 나타냈다. 이는 다양한 데이터 유형을 종합적으로 고려하는 병렬 구조의 모델이 펀딩 성공 여부를 더욱 정확하게 예측할 수 있다는 것을 보여준다.

셋째로, 초기 서포터 수를 변수로 추가한 딥러닝 모델의 성능 평가 결과, 초기 서포터 수가 펀딩의 성공 가능성을 높이는 것으로 나타났다. 초기 서포터 수를 변수로 추가한 예측모델에서는 숫자와 텍스트 데이터만을 사용한 모델보다 설명력이 더 높게 나타났다. 또한, 딥러닝 분류모델에서는 초기 서포터 수를 추가로 고려하였을 때, 정확도가 약 2배 상승하는 것으로 나타났다. 이는 초기 서포터 수가 클라우드 펀딩의 성공 여부에 중요한 영향을 미치며, 딥러닝 모델에 이를 반영함으로써 성능을 향상시킬 수 있다는 것을 보여준다.

마지막으로, 텍스트 변수 처리 방식에 따른 모델의 성능 평가 결과, 텍스트 데이터를 트랜스포머 인코더를 통

해 분석하는 방식이 더 높은 성능을 보였다. 텍스트 데이터를 토큰화하여 정수 인코딩한 후 트랜스포머 인코더 모델에 입력하는 방식은 BERT 모델에 비해 텍스트의 의미를 더욱 잘 반영하여 예측 성능을 향상시켰음을 나타냈다. 그러나 숫자 데이터와 텍스트 데이터를 모두 입력한 모델에서는 Dense 레이어를 적용한 모델의 성능이 더 높게 나타났다. 자체적으로 학습한 트랜스포머 인코더는 가지고 있는 텍스트 정보를 최대한 이용하지만, 사전 학습된 BERT는 그보다는 범용적이어서 숫자 정보 등과 같이 사용하는 것에 더 효과적일 것이라고 판단된다.

결과적으로, 클라우드 펀딩 성공을 예측하기 위해 다양한 딥러닝 모델을 설계하고 평가하였다. 병렬 구조와 초기 서포터 수를 고려한 모델은 펀딩의 성공 가능성을 더욱 정확하게 예측하는 데 도움이 되었다. 특히, 텍스트 데이터의 처리 방식을 고려한 모델은 텍스트의 의미를 잘 반영하여 예측 성능을 향상시킬 수 있었다. 이러한 결과는 클라우드 펀딩 플랫폼에서 예측모델을 활용하여 펀딩 성공 가능성을 더욱 높일 수 있음을 시사한다.

2. 결과 토의

본 연구에서 제안한 딥러닝 모델을 통해 클라우드 펀딩의 성공 가능성을 예측하였다. 다양한 모델 아키텍처와 변수 처리 방식을 비교 분석한 결과, 병렬 구조와 초기 서포터 수를 고려한 모델이 가장 높은 성능을 보였다. 특히, 텍스트 데이터의 처리 방식을 고려한 모델은 텍스트의 의미를 더욱 잘 반영하여 예측 성능을 향상시켰다.

그러나 본 연구에는 몇 가지 한계점이 존재한다. 첫째로, 사용한 데이터셋은 특정 시간 범위와 분야에 국한되어 있으며, 일반화가 어려울 수 있다. 그리고 자료의 숫자 또한 부족하다고 판단될 수 있다. 따라서 더 다양하고 많은 데이터셋을 수집하고 분석함으로써 모델의 일반

화 성능을 더욱 향상시킬 필요가 있다.

둘째로, 텍스트 데이터의 처리 방식에는 BERT 모델과 트랜스포머 인코더를 비교하였으나, 다른 텍스트 처리 모델의 활용 가능성을 고려하지 않았다. 다중언어도 같이 효과적으로 처리할 수 있는 다양한 LLM 모델을 활용한 연구가 필요하다.

셋째로, 본 연구에서는 딥러닝 모델만을 사용하여 클라우드 펀딩의 성공 가능성을 예측하였으나, 다른 머신러닝 알고리즘과의 비교를 통해 최적의 모델을 찾는 노력을 계속해야 한다. 또한 전통적인 베이지언이나 통계 모델들도 계속 개량모델들이 나오고 있기에 항상 최상의 모델로 수정하고 보완해야 하는 일을 계속해야 한다.

넷째로, 딥러닝이 가지고 있는 한계를 완화할 수 있는 보완연구가 필요하다. 딥러닝은 결과만 있고 과정이 없다는 평가를 받고 있는데, 이를 해결하기 위한 explainable AI(XAI) 기법들이 나오고 있다. 특히 어렵지 않게 결과가 나온 이후에 적용할 수 있는 SHAP와 같은 XAI 기법들이 개발된다면 이를 적용해서 숫자자료뿐만 아니라 문자자료까지 중요도나 영향력을 판단 할 수 있도록 모델개발이 필요하다.

다섯째로, 다양한 종류의 자료들, 텍스트, 이미지, 숫자 정보들을 얻을 수 있고, 동영상도 마케팅에서 많이 활용하고 있다. 이미지 자료에서 배열된 사물이나 사람의 구도라던가 매력도와 같은 정보를 추출하고 이와 비슷한 정보를 동영상광고에서도 추출할 수 있으면 더욱 좋은 연구가 될 것이다.

다섯째, explainable AI가 필요한 시기가 되었다. 특히 SHAP와 같은 모델은 모델에서 활용하는 자료들의 중요도를 찾아주는데, 본 연구에서와 같이 숫자와 텍스트가 같이 있는 경우에도 각각의 변수군의 중요도를 찾아주고, 더 나아가서 의미있는 단어를 찾아주는 연구들이 필요하다. 더 나아가서 중요도뿐만 아니라 가장 좋은

성과를 낼 수 있는 변수들을 찾아내는 연구도 필요하다.

3. 시사점

본 연구에서 제안한 딥러닝 모델은 클라우드 펀딩 플랫폼에서 예측모델을 활용하여 펀딩의 성공 가능성을 예측하는 데 큰 도움을 줄 수 있다. 특히, 병렬 구조와 초기 서포터 수를 고려한 모델은 예측 성능을 더욱 향상시켜 플랫폼 사용자들이 펀딩의 성공 가능성을 더욱 정확하게 파악할 수 있도록 도와준다.

또한, 텍스트 데이터의 처리 방식을 고려한 모델은 텍스트의 의미를 잘 반영하여 예측 성능을 향상시킬 수 있다. 클라우드 펀딩 플랫폼에서 프로젝트 설명이 중요한 요소이므로, 텍스트 데이터를 효과적으로 분석하고 활용하는 것은 플랫폼의 성공에 큰 도움이 될 수 있다.

또한, 초기 서포터 수를 고려한 모델은 프로젝트의 초기 단계에서 예측 성능을 향상시키는데 기여한다. 프로젝트 기획단계에서부터 사전 관심도를 측정할 수 있는 방법을 개발해서 이에 따라 최종 프로젝트 시작전에 프로젝트의 내용을 변경할 수 있게 하는 것이 필요하다.

마케팅의 커뮤니케이션 활동에서 사진이나 문자, 또 음악, 동영상과 중요한 숫자정보를 모두 활용한 활동들이 많아지고 있으며, 이들을 모두 고려한 예측모델이 필요하고, 이를 반영하여 AI가 최적의 사진, 음악, 문자, 동영상, 숫자정보의 조합을 자동으로 여러개 만들어서 최종적으로 마케팅담당자의 판단에 따라 실행하게 하는 것이 필요하다. 아직은 부족하지만 본 연구와 같은 multi modal 연구들이 AI 기술의 발전에 따라 지속적으로 개발된다면 마케팅담당자들에게는 좋은 도구가 될 것이다.

이러한 결과는 클라우드 펀딩 플랫폼에서 예측모델을 활용하여 펀딩의 성공 가능성을 높일 수 있다는 것을 시사한다. 플랫폼 운영자는 제안된 딥러닝 모델과 해당 방

법론을 활용하여 프로젝트의 성공 가능성을 예측하고, 펀딩을 지원할 만한 프로젝트를 선별하는 데 활용할 수 있다. 이런 활동들을 통해서 플랫폼의 효율적인 운영과 사용자들의 만족도 향상에 기여할 것으로 기대된다.

〈최초투고일: 2023년 8월 29일〉

〈수정일: 1차: 2023년 10월 19일〉

〈게재확정일: 2023년 10월 26일〉

참고문헌

성기범, 한상만, 김태완(2022), “추천시스템 품질요소가 구매 의도에 미치는 영향: 신뢰도와 온라인 스토어 이미지의 이중매개효과,” *마케팅연구*, 38, 75-91.

씨에천짜오, 최우진(2023), “기부자 권력거리신념이 크라우드 펀딩 기부의도에 미치는 영향: 수혜자 필요형태의 조절효과를 중심으로,” *마케팅연구*, 38, 25-42.

와디즈 홈페이지(<https://www.wadiz.kr/>)

유원준, 안상준(2023), “딥러닝을 활용한 자연어처리 입문,” Retrieved from [<https://wikidocs.net/31379>].

윤민섭(2017), “전자상거래법이 보상형 크라우드 펀딩에 적용되는 문제에 관한 연구,” *한국소비자원*.

윤영진, 박정은(2023), “인스타그램상에서의 소셜CRM중 피드백형 세일즈프로모션 광고의 성과에 영향을 미치는 요인에 관한 연구: CRM에 대한 소비자 인식과 소셜CRM 빈도의 조절효과를 중심으로,” *마케팅연구*, 38, 91-118.

이구형(2020), “보상형 크라우드펀딩의 투자자보호 및 시장발전을 위한 정책과제,” *국회입법조사처 경제산업연구실*.

이서진, 최인철(2022), “온라인 쇼핑환경에서 리뷰 평점의 방향성이 소비자 구매의도에 미치는 영향: 예술품과 기능품을 중심으로,” *마케팅연구*, 37, 1-19.

이정은, 신형덕(2014), “크라우드펀딩 사이트의 게시글 정보가 펀딩 성공에 미치는 영향,” *한국콘텐츠학회 논문지*, 14(6), 54-62.

허재강, 엄금철, 차동진(2022), “보상형 크라우드 펀딩에서의 사회적 증거와 마감까지의 시간적 거리가 투자의도에 미치는 영향: 달성가능성의 매개효과를 중심으로,” *마케팅연구*, 37, 45-63.

Arevalo, J., Solorio, T., Montes-y-Gómez, M., et al.(2020), “Gated Multimodal Networks,” *Neural Comput & Applic* 32, 10209-10228.

Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S.(2010), “Multimodal Fusion for Multimedia Analysis: A Survey,” *Multimedia Systems*, 16(6), 345-379.

Belleflamme, P., Lambert, T., & Schwienbacher, A.(2014), “Crowdfunding: Tapping the Right Crowd,” *Journal of Business Venturing*, 29(5), 585-609.

Bordes, A., Chopra, S., & Weston, J.(2014), “Question Answering with Subgraph Embeddings,” arXiv preprint arXiv:1406.3676.

Dai, A. M., & Le, Q. V.(2015), “Semi-supervised Sequence Learning,” *Advances in Neural Information Processing Systems*, 28.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K.(2018), “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv preprint arXiv:1810.04805.

Dew, Ryan, Asim Ansari, and Olivier Toubia(2021), “Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design,” *Marketing Science*, 41(2), 401-425.

Dhillon, Paramveer S. and Sinan Aral(2021), “Modeling Dynamic User Interests: A Neural Matrix Factorization Approach,” *Marketing Science*, 40(6), 1059-1080.

Du, Q., Fan, W., Qiao, Z., Wang, A. G., Zhang, X., & Zhou, M. (2015), “Money Talks: A Predictive Model on Crowdfunding Success Using Project Description,” *Americas Conference on Information Systems*.

Fan, T., Gao, L., & Steinhart, Y.(2020), “The Small Predicts Large Effect in Crowdfunding,” *Journal of Consumer Research*, 47(4), 544-565.

Gabriel, S., Bosselut, A., Holtzman, A., Lo, K., Celikyilmaz,

- A., & Choi, Y.(2019), “Cooperative Generator-Discriminator Networks for Abstractive Summarization with Narrative Flow,” arXiv preprint arXiv:1907.01272.
- Gerber, E. M., Hui, J. S., & Kuo, P.-Y.(2012), “Crowdfunding: Why People Are Motivated to Post and Fund Projects on Crowdfunding Platforms,” *Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics*, 2, 11.
- Hinton, G. E., & Salakhutdinov, R. R.(2006), “Reducing the Dimensionality of Data with Neural Networks,” *Science*, 313, 504-507.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B.(2012), “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups,” *IEEE Signal Processing Magazine*, 29(6), 82-97.
- Hong, Jiyeon, Paul R. Hoban(2022), “Writing More Compelling Creative Appeals: A Deep Learning-Based Approach,” *Marketing Science*, 41(5), 941-965
- Howard, J., & Ruder, S.(2018), “Universal Language Model Fine-tuning for Text Classification,” arXiv preprint arXiv:1801.06146.
- Hui, J. S., Gerber, E., Greenberg, M.(2012), “Easy money? The demands of crowdfunding work,” *Segal Design Institute*, Northwestern University, Evanston, IL.
- Jean, S., Cho, K., Memisevic, R., & Bengio, Y.(2014), “On Using Very Large Target Vocabulary for Neural Machine Translation,” arXiv preprint arXiv:1412.2007.
- Kim, Hyowon, and Greg M. Allenby(2022), “Integrating Textual Information into Models of Choice and Scaled Response Datam,” *Marketing Science*, 41(4), 815-830.
- Koch, J. A., & Siering, M.(2015), “Crowdfunding Success Factors: The Characteristics of Successfully Funded Projects on Crowdfunding Platforms,” *Proceedings of the 23rd European Conference on Information Systems(ECIS 2015)*; Muenster, Germany.
- Kuppuswamy, V., & Bayus, B. L.(2018), “Crowdfunding Creative Ideas: The Dynamics of Project Backers,” *The Economics of Crowdfunding: Startups, Portals and Investor Behavior*, 151-182.
- LeCun, Y., Bengio, Y., & Hinton, G.(2015), “Deep Learning,” *Nature*, 521(7553), 436-444.
- Mao, J., Xu, W., Yang, Y., Wang, J., & Yuille, A. L.(2014), “Explain Images with Multimodal Recurrent Neural Networks,” arXiv preprint arXiv:1410.1090.
- Markoff, J.(2012), “Scientists See Promise in Deep-Learning Programs,” NY Times. <http://nyti.ms/sgcVec>.
- Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J.(2011), “Strategies for Training Large Scale Neural Network Language Models,” 2011 *IEEE Workshop on Automatic Speech Recognition & Understanding*, Waikoloa, HI, USA, 196-201.
- Mollick, E.(2014), “The Dynamics of Crowdfunding: An Exploratory Study,” *Journal of Business Venturing*, 29(1), 1-16.
- Nam, S., Jin, Y., & Kwon, O.(2018), “Online Document Mining Approach to Predicting Crowdfunding Success,” *지능정보 연구*, 24(3), 45-66.
- Nielsen, K. R., & Binder, J. K.(2021), “I Am What I Pledge: The Importance of Value Alignment for Mobilizing Backers in Reward-Based Crowdfunding,” *Entrepreneurship Theory and Practice*, 45(3), 531-561.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L.(2018), “Deep Contextualized Word Representations,” *In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1(Long Papers), pages 2227-2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Pitschner, S., & Pitschner-Finn, S.(2014), “Non-Profit Differentials in Crowd-Based Financing: Evidence From 50,000 Campaigns,” *Economics Letters*, 123(3), 391-394.
- Radford, A., & Narasimhan, K.(2018), “Improving Language Understanding by Generative Pre-Training.”

- Ramachandram, D., & Taylor, G. W.(2017), "Deep Multimodal Learning: A Survey on Recent Advances and Trends," *IEEE Signal Processing Magazine*, 34(6), 96-108.
- Reimers, N., & Gurevych, I.(2019), "Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks," arXiv preprint arXiv:1908.10084.
- Short, J. C., Ketchen Jr, D. J., McKenny, A. F., Allison, T. H., & Ireland, R. D.(2017), "Research on Crowdfunding: Reviewing the(Very Recent) Past and Celebrating the Present," *Entrepreneurship Theory and Practice*, 41(2), 149-160.
- Simonsohn, U., & Ariely, D.(2008), "When Rational Sellers Face Nonrational Buyers: Evidence from Herding on eBay," *Management Science*, 54(9), 1624-1637.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A.(2014), "Going Deeper with Convolutions," arXiv(Cornell University),
- Tompson, J., Jain, A., LeCun, Y., & Bregler, C.(2014), "Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation," *NIPS*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I.(2017), "Attention is All You Need," *Advances in Neural Information Processing Systems*, 30.
- Vismara, S.(2018), "Information Cascades Among Investors in Equity Crowdfunding," *Entrepreneurship Theory and Practice*, 42(3), 467-497.
- Wang, G., Liu, X., & Fan, W.(2011), "A Knowledge Adoption Model Based Framework for Finding Helpful User-Generated Contents in Online Communities," *In International Conference on Information Systems*.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. M.(2020), "Transformers: State-of-the-Art Natural Language Processing."
- Yoshimura, R., Shimanaka, H., Matsumura, Y., Yamagishi, H., & Komachi, M.(2019), "Filtering Pseudo-References by Paraphrasing for Automatic Evaluation of Machine Translation."
- Yuan, H., Lau, R. Y., & Xu, W.(2016), "The Determinants of Crowdfunding Success: A Semantic Text Analytics Approach," *Decision Support Systems*, 91, 67-76.
- Zhang, J., & Liu, P.(2012), "Rational Herding in Microloan Markets," *Management Science*, 58(5), 892-912.
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y.(2019), "Bertscore: Evaluating Text Generation with Bert," arXiv preprint arXiv:1904.09675.
- Zhou, M., Lei, L., Wang, J., Fan, W., & Wang, A. G.(2015), "Social Media Adoption and Corporate Disclosure," *Journal of Information Systems*, 29(2), 23-50.