

[REDACTED]

# 기계학습의 이해 프로젝트

## 최종 보고서

기계학습의 이해 202402295 성유현

# 프로젝트 개요

- 프로젝트
  - 머신러닝 모델을 이용하여 사용자가 관심 있을 만한 뉴스를 자동으로 분류하고, 디스크로 전송하는 서비스 시스템을 구현하고자 하는 프로젝트
- 해결하고자 하는 문제
  - 매일 쏟아지는 뉴스 중에서 사용자가 직접 뉴스를 하나하나 선택하기 어려운 문제를 해결하기 위해, 사용자의 관심 뉴스만 자동으로 필터링하는 시스템을 개발
- 전체 시스템 흐름
  - > 뉴스 수집 → 전처리 → 모델 예측 → 확률 기반 필터링 → 디스크 전송

# 진행과정

- 주제 선정 및 문제 정의
  - 너무 많은 뉴스들이 있고 대부분은 내가 관심없는 뉴스들이라는 문제가 있음
  - 이를 해결하기 위해 내가 관심있는 뉴스들을 추천해주는 시스템을 만들기로 함
- 데이터 수집 및 분석
  - 네이버 검색 API를 통해 학습 및 평가 데이터 수집
  - 수집된 뉴스 데이터의 데이터 형태 및 통계량(클래스 분포, 제목 길이, 요약문 길이 등) 분석

# 진행과정

- ML 모델 학습 및 평가
  - 수집한 뉴스 데이터를 제목+ 요약 형태로 바꾼 후 TF-IDF 벡터화
  - 수집한 데이터에서 Train / Validation / Test = 70% / 15% / 15%
  - Fully Connected Layer로 구성된 Feedforward Neural Network 모델 사용
  - 출력 레이어는 2 unit ( 관심 있음 / 관심 없음 )
  - 평가는 Accuracy, Precision, Recall, F1-score로 진행

# 진행과정

- 서비스 개발
  - 기존 계획은 일정 시간마다 자동으로 뉴스 전송이였으나 코드 실행 시 가장 최근 뉴스들 중 가장 관심 있을 만한 뉴스 몇 개 전송으로 변경
  - 학습 시킨 모델이 구글 뉴스 RSS를 통해 받은 뉴스에 대해 판단 후 디스코드 전송
  - 기존에는 argmax를 이용하여 관심 여부만 판단하였으나, 최종 구현에서는 뉴스 관심 정도를 판단하기 위해 softmax를 적용하여 각 뉴스의 관심 확률을 계산하고 확률 기반 정렬 방식으로 개선

# 모델을 서비스로 만든 구조

- 시스템 구조

- RSS를 통해 들어온 데이터를 전처리(제목+요약 형태)
- 벡터라이저(.pkl)를 이용하여 TF-IDF 벡터화
- 기존에 학습된 모델(.pth)를 통해 예측
- 예측 결과를 정리 후 디스코드에 전송

[RSS 뉴스 수집]



[전처리 / 텍스트 정제]



[TF-IDF 벡터화]



[PyTorch 모델 예측]



[확률 필터링 (0.6 이상)]



[상위 5개 선택]



[Discord Webhook 전송]

# 실제 사용 결과

The image shows a series of messages from a user named 'Captain Hook' in a messaging application. The messages are timestamped and include news links from Google News and other sources. The news topics cover various subjects like space exploration, technology, and real estate.

- 2023-10-12 11:00**  
오늘의 추천 뉴스(확률 기반)  
韓 배터리 소재사, '유럽산' 찾는 中 기업 공략 - 지디넷코리아  
관심 확률: 1.00  
한국타이어, 인니 수자원 인프라 지원 프로젝트 마무리 - 매일일보  
관심 확률: 0.99  
서울 아파트 '증여 전환' 급증...1년 새 25%↑, 강남3구에 집중 - 뉴스1  
관심 확률: 0.97
- 2023-10-12 14:50**  
오늘의 추천 뉴스(확률 기반)  
"남편 빌려드립니다"... 라트비아서 '남편 대여 서비스' 인기, 이유는? - 조선일보  
관심 확률: 0.98  
한국타이어, 인니 수자원 인프라 지원 프로젝트 마무리 - 매일일보  
관심 확률: 0.97  
서울 아파트 '증여 전환' 급증...1년 새 25%↑, 강남3구에 집중 - 뉴스1  
관심 확률: 0.96
- 2023-10-12 15:00**  
오늘의 추천 뉴스(확률 기반)  
[초점] 스페이스X, 몸값 8000억 달러 조준...2026년 상장·군불·땐다 - 글로벌이코노믹  
관심 확률: 0.99  
[초점] 스페이스X, 몸값 8000억 달러 조준...2026년 상장 '군불' 땐다 글로벌이코노믹일론 머스크, 전기차·로봇 보다 우주사업에 옮인하나 매일경제스페이스X 구주 매각 추진...회사 가치 1180조 상정 뉴시스오픈AI 추월하나...스페이스X, 기업가치 1180조 원 '정조...'  
https://news.google.com/rss/articles/CBMiiAFBVV95cUxQdnZJVnlZWjhbyDRpVzQ0MThjYmlHY3daRkJOWWhTVXRvTE0xZnhLMkdOVXRlsru5WVZZbTc0TDFpeVk0RW5TUIR1TlJhYVJoamZiNHBYbTBNeFZHWDZmdFdvTFFfMDhNZVFpUGN1cUtJTGRSUkZebURHSXZJY1Q0RU9TSndCQmg4?oc=5  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G
- 2023-10-12 15:10**  
오늘의 추천 뉴스(확률 기반)  
프랑스 혁명수함 기지 상공서 드론 5대 포착...격추 사격 - 연합뉴스  
관심 확률: 0.99  
프랑스 혁명수함 기지 상공서 드론 5대 포착...격추 사격 연합뉴스Google 뉴스에서 전체 콘텐츠 보기...  
https://news.google.com/rss/articles/CBMiWOFVX3lxTFBtMDQ5a2k2MWRKbVZLeIyRVQ5Wm1UOUE3elp3cU1VM2pPQRSMzBNtMzmMkdTUkVVxzhmXIFIX1cyRTVnbFJWTHpfsm53Vm5NxzVpekRkdG5ib3fSAWBBVV95cUxPdGnINIIInGeWvDlIzMy5ktiOfn7a17mMVnns/N2Nk51laV01/m7SrdFwWlndovFOVh1VT1H7aVklcr  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G
- 2023-10-12 15:20**  
오늘의 추천 뉴스(확률 기반)  
뉴욕증시, 금리 인하 기대감에 3대 지수 상승 마감 - 마켓인  
관심 확률: 1.00  
뉴욕증시, 금리 인하 기대감에 3대 지수 상승 마감 마켓인[뉴욕증시] 인플레 둔화 소식에 3대 지수 상승 지속...팔란티어, 2.16% 상승 글로벌이코노믹뉴욕증시, 금리 인하 기대감에 상승 마감 v.daum.net[뉴욕마켓워치] 美 인플레이션 안도감에 증시 강세...국채·달러...  
https://news.google.com/rss/articles/CBMic0FVX3lxTFBzY1RfcFQtQ1INTVZEQXhtSzlfUDEyNm1sci13SEhGT1dYU0MzX3pJLJtdSRXJ4TJlXVGp6a2xsMzBFamU2d0RfVDRuZjA3TFVpMDF3M2IPTjVlcEZsbnUzUVZOekpBaVh4dlVNcjhlR043aOk?oc=5  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G
- 2023-10-12 15:30**  
오늘의 추천 뉴스(확률 기반)  
[초점] 스페이스X, 몸값 8000억 달러 조준...2026년 상장·군불·땐다 - 글로벌이코노믹  
관심 확률: 0.99  
[초점] 스페이스X, 몸값 8000억 달러 조준...2026년 상장 '군불' 땐다 글로벌이코노믹일론 머스크, 전기차·로봇 보다 우주사업에 옮인하나 매일경제스페이스X 구주 매각 추진...회사 가치 1180조 상정 뉴시스오픈AI 추월하나...스페이스X, 기업가치 1180조 원 '정조...'  
https://news.google.com/rss/articles/CBMiiAFBVV95cUxQdnZJVnlZWjhbyDRpVzQ0MThjYmlHY3daRkJOWWhTVXRvTE0xZnhLMkdOVXRlsru5WVZZbTc0TDFpeVk0RW5TUIR1TlJhYVJoamZiNHBYbTBNeFZHWDZmdFdvTFFfMDhNZVFpUGN1cUtJTGRSUkZebURHSXZJY1Q0RU9TSndCQmg4?oc=5  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G
- 2023-10-12 15:40**  
오늘의 추천 뉴스(확률 기반)  
한국타이어, 인니 수자원 인프라 지원 프로젝트 마무리 - 매일일보  
관심 확률: 0.99  
한국타이어, 인니 수자원 인프라 지원 프로젝트 마무리 매일일보한국타이어, 인니 수자원 인프라 지원 '워터펌프' 설치 마무리 연합뉴스한국타이어, 인도네시아 '워터 펌프 설치 프로젝트' 마무리 v.daum.net한국타이어, 물 부족' 인니 지역에 워터펌프·물탱크 설치 인...  
https://news.google.com/rss/articles/CBMiY0FVX3lxTE52cHpQbFA1QzRvYklMV9XcJvNIN5dGVuTXJhMFI3TEhUeUJndnRXdTFkWHDWlduTFIQCSPeHVNdVNTbS1SbmVSHF2bzE1bU9Cc0lYaHdZWwxYelFRWDA5Yw?oc=5  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G
- 2023-10-12 15:50**  
오늘의 추천 뉴스(확률 기반)  
서울 아파트 '증여 전환' 급증...1년 새 25%↑, 강남3구에 집중 - 뉴스1  
관심 확률: 0.97  
서울 아파트 '증여 전환' 급증...1년 새 25%↑, 강남3구에 집 사기 어려워...대출 낌 아파트 증여 알아봤더니 - 조선비즈 Chosunbiz'엄빠 찬스' 3년 만 최다...강남4구·마용성 전수검증 KBS 뉴스압구정 60억 아파...  
https://news.google.com/rss/articles/CBMiWOFVX3lxTE9WemU4SGJ6M2p3UjB1eJiNWxCQTLyUzA5THJISHA10Ww4VIBYd3JJUkwVW9nQXRyW9PcTJoV0IMcU5hWXA5RTNMbmhLS05tZniUHrlkjld3c?oc=5  
Google News  
Comprehensive up-to-date news coverage, aggregated from sources all over the world by Google News.  
G

# 배운점 및 개선 방향

- 배운점

- 데이터 수집부터 서비스까지의 전체적인 진행 과정(처음부터 끝까지의 흐름?)
- 머신러닝 모델을 파일 형태로 저장하고 로드하는 방법
- 모델을 실제 서비스 코드로 연결하는 방법(디스코드)

- 개선 방향

- 성능 개선(모델 파라미터 변경, 데이터 수 증가 등등)
- 자동 실행 스케줄러 적용
- 웹 페이지 또는 앱 형태로 확장