1. 주제

가짜 뉴스 판별 가능한 모델

나반, 13팀

2. 요약

- 목표: 가짜뉴스 모델을 만들어 가짜뉴스를 판별할 수 있도록 도와줍니다

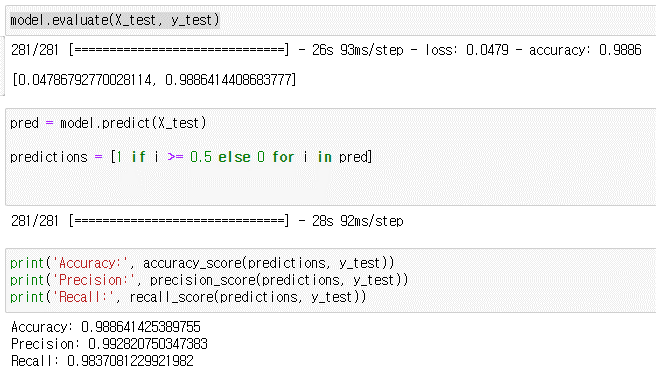
- 핵심내용: 모델을 해외의 데이터셋을 이용하여 학습 데이터와 테스트 데이터로 나누어 학습시키고 모델의 정확성을 표현합니다. 정확도를 나타내고 모델에 데이터셋 이외의 기사를 대입하여 진위를 판별합니다.

- 중요성: 최근 정보의 양이 많아 어느 것이 가짜 뉴스인지 구분하기 어려워지는 경우가 종종 발생하기 때문에 진위를 판별할 수 있는 모델의 중요성이 높을 수 있습니다.

3. 대표그림

- 개발 배경: 파이토치 기반

- 예상 결과: 모델의 정확성을 보여 성능을 보일 예정



4. 서론

- 배경 설명

가짜 뉴스는 잘못된 정보를 전파하고 사회적 혼란을 일으킬 수 있는 중요한 문제 중 하나입니다. 가짜 뉴스는 소셜 미디어와 인터넷의 보급, 빠른 정보 전달, 그리고 언론의 다양한 형태로 확산되고 있습니다. 이로 인해 사람들은 종종 신뢰할 수 없는 정보에 노출되며, 잘못된 판단을 내릴 수 있습니다. 따라서 가짜 뉴스 탐지 모델은 현재의 정보 환경에서 중요한 역할을 합니다.

- 사례 분석

정치 분야에서 후보자나 정당에 대한 공격적인 가짜 뉴스가 퍼진 예도 많습니다. 이는 선거의 결과에 영향을 미칠 수 있으며, 정치 논쟁을 더 복잡하게 만들 수 있습니다. 전쟁에서도 가짜 뉴스가 퍼져 군의 사기를 저하시키는 경우도 생겼습니다.

- 문제 정의

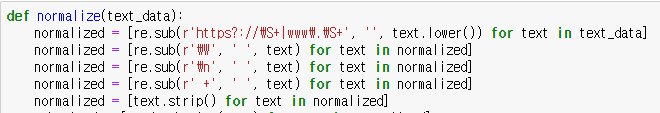
가짜 뉴스가 확산되면 기존의 뉴스 매체와 정보에 대한 신뢰도가 하락할 수 있습니다. 사람들은 어떤 정보를 믿어야 할지 혼란스러워지며, 실제로 현실적인 정보에 대한 신뢰가 훼손될 수 있습니다. 가짜 뉴스는 종종 사회적 분열을 심화시키는 역할을 합니다. 거짓 정보는 사람들 사이에 갈등을 유발하고, 정치적, 종교적, 인종적 분쟁을 촉발할 수 있습니다. 가짜 뉴스를 제작하고 퍼뜨리는 사람들은 광고 수익, 사기, 불법 상품 판매 등을 목적으로 할 수 있으며, 이로 인해 경제적 피해가 발생할 수 있습니다. 가짜 뉴스는 정치적 결정과 공공 정책에 영향을 미칠 수 있으며, 정확한 정보 없이 잘못된 결정을 내리는데 기여할 수 있습니다.

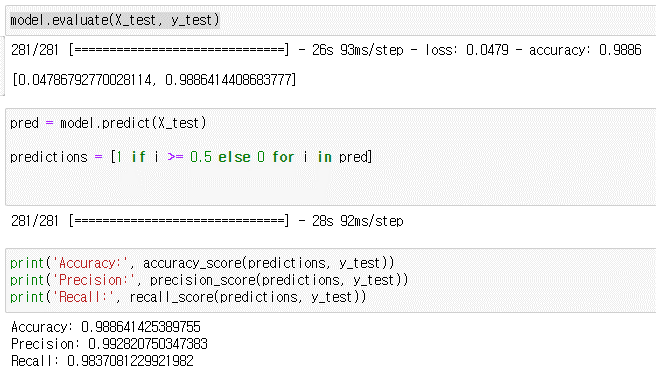
- 극복 방안

가짜 뉴스 모델을 사용하여 가짜 뉴스를 탐지하고 사용자의 피해를 줄일 수 있게 해줍니다.

5. 본론

- 시스템 개요





학습에 사용될 텍스트를 전처리하고 단어들에 고유의 숫자를 매핑하여 해당 단어의 비중이나 다른 단어들과의 연관성 분석합니다. 그 후 모델의 성능을 출력합니다.

- 필요한 기술 요소 설명

자연어 처리: 텍스트 데이터를 처리하고 이해하는데 필요한 핵심 기술입니다. 텍스트를 토큰화하고, 단어를 벡터로 표현하며, 문장 구문 및 의미를 분석하는 기술이 포함됩니다.

가짜 뉴스 모델을 훈련하기 위해 기계 학습 및 딥러닝 기술을 활용합니다. 이러한 기술은 모델을 데이터로부터 학습하고 예측을 수행하는 데 사용됩니다.

LSTM: RNN의 단점을 보완하기 위해 개발된 자연 언어 처리 방식입니다. 새로운 데이터가 들어올때마다 업데이트 가능하고 필요없어진 구데이터는 삭제가 가능하여 매일 새로운 뉴스가 업데이트 되는 모델에 적합합니다.

- 구현 방법 및 개발 방향

파이토치 환경에서 모델을 설계합니다. 그리고 가짜 뉴스와 진짜 뉴스의 데이터를 수집하고 전처리합니다. 이 데이터는 훈련 데이터와 테스트 데이터로 나눠 진행합니다.

텍스트를 전처리하고 토큰화하여 모델 학습에 이용합니다.

설계한 모델을 훈련 데이터로 학습시킵니다. 모델은 LSTM방식을 이용하였습니다.

테스트 데이터를 사용하여 모델의 성능을 평가하고 정확도나 정밀도를 측정합니다.

6. 결론

- 보고 내용 요약 : LSTM방식을 이용하여 모델을 학습시키고 가짜 뉴스를 판별할 수 있는 모델을 만들어 정확성을 파악합니다.

- 향후 할일 정리 : 나중에 나올 뉴스를 대비하여 더 많은 데이터를 수집하여 모델을 업데이트합니다. LSTM이후 새로운 언어 모델이 나온다면 실험을 합니다.

사용한 데이터셋

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset