

Софийски Университет
”Св. Климент Охридски”
Извличане на информация

16.12. 2024 г.

Класификатор на фалшиви новини(чрез Spark)

Ради Радев

8MI3400507

Разпределени Системи и Мобилни технологии

Съдържание:

1. Мотивация

2. Кратък обзор

- Работа в областта на откриване на фалшиви новини
- Възможни методи за решаване на задачата

3. Архитектура на системата и основни моменти от реализацията

- Основен работен процес
- Ролята на Apache Spark
- Предварителна обработка на данни
- Извличане на характеристики
- Обучение на модела

4. Резултати и предизвикателства

- Основни постигнати показатели
- Забележителни предизвикателства и решения

5. Заключение

- Обобщение на решението
- Насоки за бъдеща работа
- Източник на данни и характеристики

Мотивация

Фалшивите новини са нарастващ глобален проблем. Влияят негативно на обществото, демокрацията и доверието. Те съдържат подвеждаща информация, целяща да променя мнения, за да обслужи различни интереси - политически, икономически, финансови и т.н.



Работа в областта на откриване на фалшиви новини

Откриването на фалшиви новини е активно изследвана област, която обхваща теми като обработка на естествен език (NLP), машинно обучение и когнитивни науки. Ето някои важни трудове и подходи, които са оформили тази сфера:

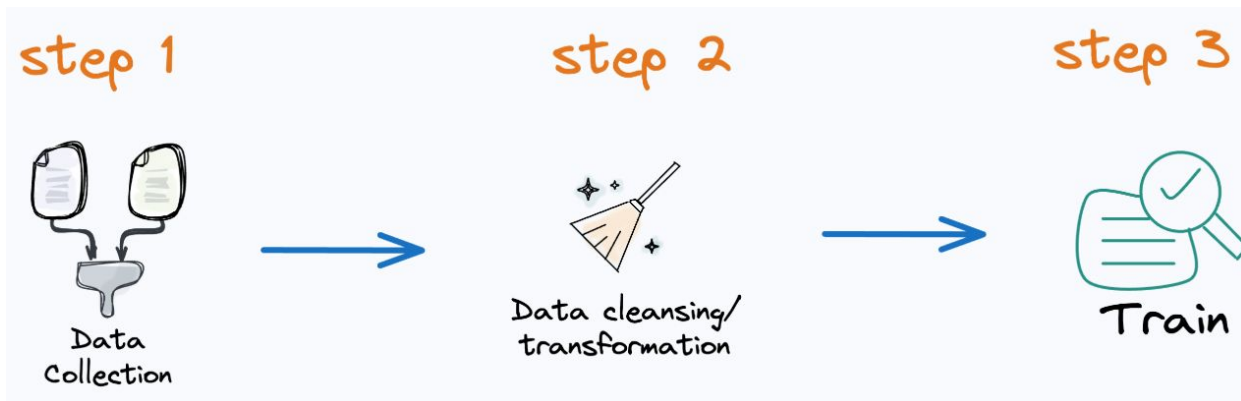
- "Fake News Detection on Social Media: A Data Mining Perspective" (Shu et al., 2017)
- "Fake News Detection Using Deep Learning" (Wang, 2017)
- "A Survey on Fake News Detection: Data, Methods, and Opportunities" (Zhou & Zafarani, 2018)

Възможни методи за решаване на задачата

- Лингвистичен анализ:
 - Лексикален анализ
 - Анализ на семантиката(Word2vec, Glove, Sentence Transformers)
- Методи на машинно обучение
 - Supervised learning(Logistic Regression, Random Forest)
 - Unsupervised learning(откриване на аномалии)
- Deep learning(Bert, GPT, RoBERTa)
- Мултимодален анализ
- Графови методи
- Методи за проверка на факти(Автоматизирана проверка(PolitiFact), Ретроспективен анализ)
- Анализ на потребителско поведение

Решение

- Архитектура и основни моменти на реализация



Архитектура на системата

1. Събиране на данни (CSV).
2. Предварителна обработка (премахване на шум, токенизация).
3. Извличане на характеристики (TF-IDF, Word2Vec).
4. Обучение на модел (логистична регресия, random forest).
5. Оценка и прогноза.
6. Контейнеризация и преносимост на прокета

Основни моменти от реализацията

Предварителна обработка:

- Премахване на стоп думи и пунктуация.
- Токенизация и стеминг.

Характеристики:

- TF-IDF (важност на думите).
- N-грам модели и анализ на дължината на текста.

Основни моменти от реализацията

Spark MLlib:

- Логистична регресия за класификация.
- Random Forest за по-сложни модели.

Оценка:

- Метрики: Точност, прецизност, F1-скор.

Резултати и предизвикателства

- **Резултати:**

- Постигната точност: 95+%.
- Подобрение след оптимизация на данните.

- **Възможни подобрения:**

- Балансиране на класовете.
- Мащабируемост на обработката на данни.

Заклучение

Постигнато:

- Изградена система за откриване на фалшиви новини.
- Подобрени резултати чрез използване на Spark.

Бъдещи подобрения:

- Въвеждане на реални времеви прогнози.
- Използване на модели като BERT за по-точно разпознаване.

Источници и данни

Примерни източници:

- *LIAR Dataset*, Kaggle Fake News Dataset.

<https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets?resource=download>

<https://www.kaggle.com/code/vovanthuong/pyspark-nlp-detect-fake-news-97-acc-99-auc>

<https://www.kaggle.com/datasets/doanquanvietnamca/liar-dataset>

<https://medium.com/walmartglobaltech/saving-and-retrieving-ml-models-using-pyspark-in-cloud-platform-d8b1db9e91b1>

https://www.researchgate.net/publication/369467973_Big_Data_ML-Based_Fake_News_Detection_using_Distributed_Learning

Декалрация за плагиатство

- Тази курсова работа е моя работа, като всички изречения, илюстрации и програми от други хора са изрично цитирани.
- Тази курсова работа или нейна версия не са представени в друг университет или друга учебна институция.
- Разбирам, че ако се установи плагиатство в работата ми ще получа оценка “Слаб”.

Ради Стефчев Радев