Domain Background

Our world is highly interconnected and it is of paramount importance that citizens are informed objectively about issues that influence and shape our world (like issues on geopolitics or climate change). The internet lead to a rise of news media (e.g. social media, news portal, etc.) to report on these stories. The vast amount of (online) articles available yield a new phenomenon called fake news. Fake news is false or misleading information presented as news and can reduce the impact of real news [1].

Problem Statement

How can fake news be distingiushed from reliable and trustworthy information? In the following capstone project, a machine learning model shall be developed whose aim is to detect fake news. The underlying problem is to bring words and sentences into a mathematical representation for machine learning to be applied.

Datasets and Inputs

Articles classified as fake and real news are needed in order to train machine learning models. Such a dataset is taken from kaggle [2] and was collected from real world sources. Truthful articles were obtained from Reuters and fake news articles were gathered from unreliable websites that were flagged by Politifact which is a fact-checking organization. These datasets contain different types of articles on different topics, the majority of articles focus on political and world news topics [3], [4], [5]. There are about 21,417 real news articles and about 23,481 fake news articles in the dataset. So data is roughly balanced. Furthermore, the data consists of four columns: title, text, subject and date.

Solution Statement

Natural language processing (NLP) models are used to bring words and sentences into a mathematical representation. Often applied is the n-gram model which is a contiguous sequence of items with length n (such as words or characters) to be analyzed. Based on the n-gram model, features can be extracted as discussed by Hadeer et al. [4]:

- Term frequency (TF): TF defines the number of times a words appears in a document with respect to the number of total words in the document.
- Term frequency-inverted document frequency (TF-IDF): TF-IDF is a statistical metric used to measure how important a term is to a document compared to a set of documents.

Benchmark Model

Hadeer et al. [5] achieved an accuracy of 92 % when a linear support vector machine (LSVM) combined with 1-gram model and a 50,000 top feature selection TF-IDF method was used.

Evaluation Metrics

The evaluation metric for this formulated problem is accuracy. This means how many labels are correctly classified (either as fake or real news) in respect to all classified labels.

Project Design

Machine learning models shall be developed on AWS SageMaker [6]. The following project design is proposed:

- Article exploration: The nature of the articles is investigated and a text processing strategy is derived.
- Text processing: Articles need to be processed with respect to formation, stop word removal, stemming, etc.
- Feature extraction: Dictionaries and n-gram models are built and together with techniques such as TF and TF-IDF features are extracted.
- Data preparation: Training, test and validation data for the machine learning models are prepared, for example 60 % training, 20 % test and 20 % validation data.
- Modeling: Four machine learning models are evaluated sourcing from different frameworks such as support vector machine (SVM) from scikit-learn, linear learner (LL) and xgboost (XGB) from SageMaker built-in models and recurrent neural network (RNN) based on PyTorch.
- Tuning and validation: The test data is used to tune the machine learning hyperparameters and the validation data is used to evaluate the final metrics and to clarify the best modeling strategy.

References

[1] Fake news, (2020, December 22), wikipedia.org, https://en.wikipedia.org/wiki/Fake_news

[2] Fake and real news dataset, (2020, December 22), kaggle.com, https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

[3] ISOT fake news detection datasets, (2020, December 22), University of Victoria, https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php

[4] Ahmed, H, Traore, I, Saad, S. Detecting opinion spams and fake news using text classification, Security and Privacy, 2018, https://doi.org/10.1001/spy2.9

[5] Ahmed H., Traore I., Saad S. (2017) Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham. https://doi.org/10.1007/978-3-319-69155-8_9

[6] Amazon SageMaker, (2020, December 22), amazon.com, https://aws.amazon.com/sagemaker