impact of false news in rag

Saumya Dabhi1, Joseph Martínez2, Brian Llinás3, and Faryaneh Poursardar4

1Dept. of Industrial Eng. and Operations Research, Columbia University, New York, NY, USA

2Dept. of Civil and Environmental Eng., Carleton University, Ottawa, ON, CANADA

3Dept. of Supply Chain Mgmt. and Analytics, University of Nebraska-Lincoln, Lincoln, NE, USA

4Dept. of Systems Eng., City University of Hong Kong, Hong Kong SAR, CHINA

ABSTRACT

# INTRODUCTION

# methodology

A dataset of 100 news with 50 fake and 50 real news was used, split into training and testing with 80-20.

We will train Random Forest Classifier (RFC), Support Vector Machine (SVM), Deep Convolutional Neural Network (DCNN), Bidirectional LSTM models are Deep Learning based, 3 models are based on the Transformer's [1] architecture: BERT[[1]](#footnote-2) [2], roBERTa[[2]](#footnote-3) [3] and MiniLM[[3]](#footnote-4)[4]

# results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Architecture family | Model | Accuracy | Recall | Precision | F1 |
| Machine Learning | RFC |  |  |  |  |
| SVM |  |  |  |  |
| Deep Learning | DCNN | 89.99 ~ 90 |  |  |  |
| Bidirectional LSTM | 94.99 ~ 95 |  |  |  |
| Transformers | BERT\* | 0.95 | 0.95 | 0.95 | 0.95 |
| roBERTa\* | 0.90 | 0.91 | 0.90 | 0.89 |
| MiniLM\* | 1.00 | 1.00 | 1.00 | 1.00 |

\*The transformer models were stopped at 7, 7, and 13 epochs respectively. All of them are probably overfitted because of the high metric performances and a huge drop in the loss during the training and evaluation through the epochs.

# conclusions

REFERENCES

1. Vaswani, A., et al., *Attention is all you need.* Advances in neural information processing systems, 2017. **30**.

2. Devlin, J.a.C., Ming-Wei and Lee, Kenton and Toutanova, KristinaJacob Devlin, *BERT: Pre-training of Deep Bidirectional Transformers forLanguage Understanding.* arXiv, 2019.

3. Barbieri, F., et al. *TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification*. Association for Computational Linguistics.

4. Wang, W., et al., *Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers.* Advances in Neural Information Processing Systems, 2020. **33**: p. 5776-5788.

1. [google-bert/bert-base-uncased · Hugging Face](https://huggingface.co/google-bert/bert-base-uncased) [↑](#footnote-ref-2)
2. [cardiffnlp/twitter-roberta-base-sentiment · Hugging Face](https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment) [↑](#footnote-ref-3)
3. [microsoft/MiniLM-L12-H384-uncased · Hugging Face](https://huggingface.co/microsoft/MiniLM-L12-H384-uncased) [↑](#footnote-ref-4)