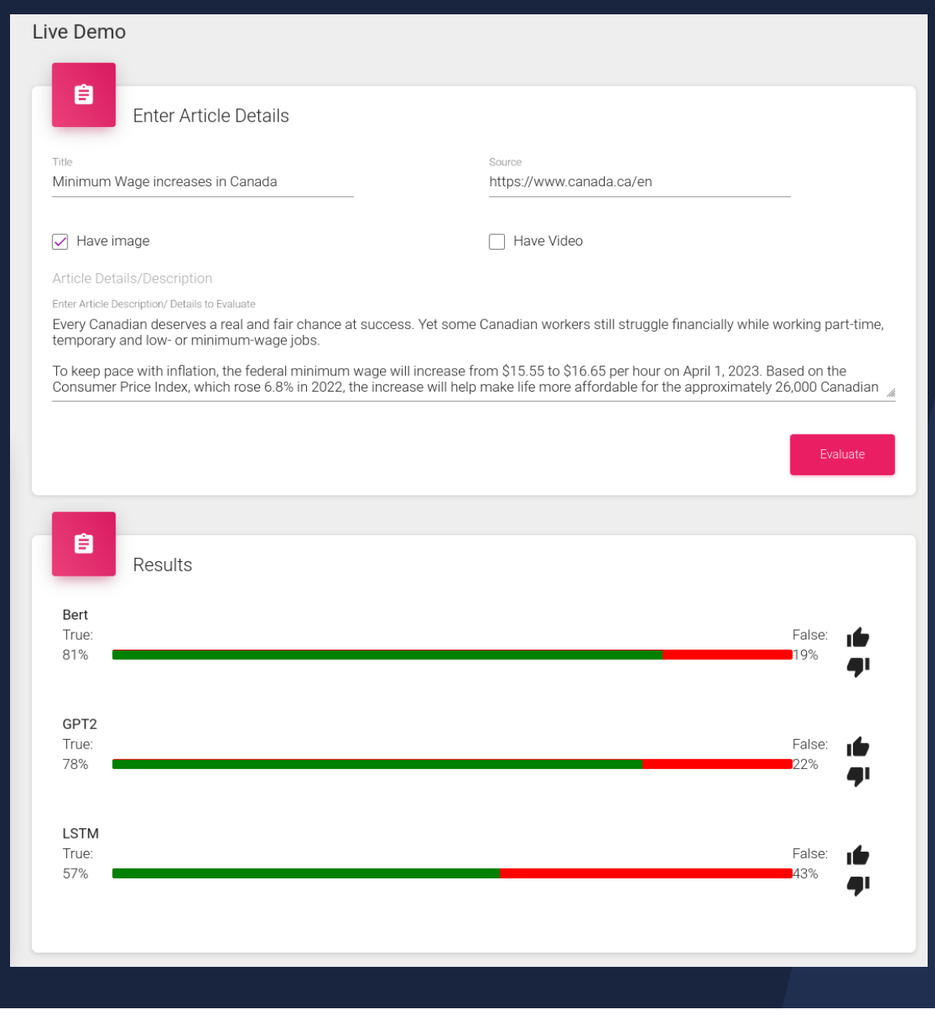
The Truth Behind Fake News: Tools and Techniques for Detection

By 3_datamen

A Deep Learning tool that uses cutting-edge technology to detect authenticity of a news article, simply by entering its details.



1. Introduction

- The spread of fake news on social and online media can intensify conflicts, promote discrimination, and undermine democracy, making fake news detection important.
- Detecting fake news is challenging due to the huge volume of information online, diverse formats, topics, and sources, and fake news makers getting better at disguising it.

2. Methods

- Integrated dataset from different sources to form training, validation, and test sets
- Transfer learning and fine tuning on three popular general purpose NLP models
- Live data and feedback collection for updating models continuously

3. Results

- Developed and deployed three fake news detectors
- All three models can achieve >86% accuracy on test set
- GPT-2-based model achieved a 93% accuracy on test set, with a f1-score of 91% on "Fake" class

4. Benefits

- Protecting credibility
- Promoting accuracy
- Preventing misinformation
- Saving time and resources

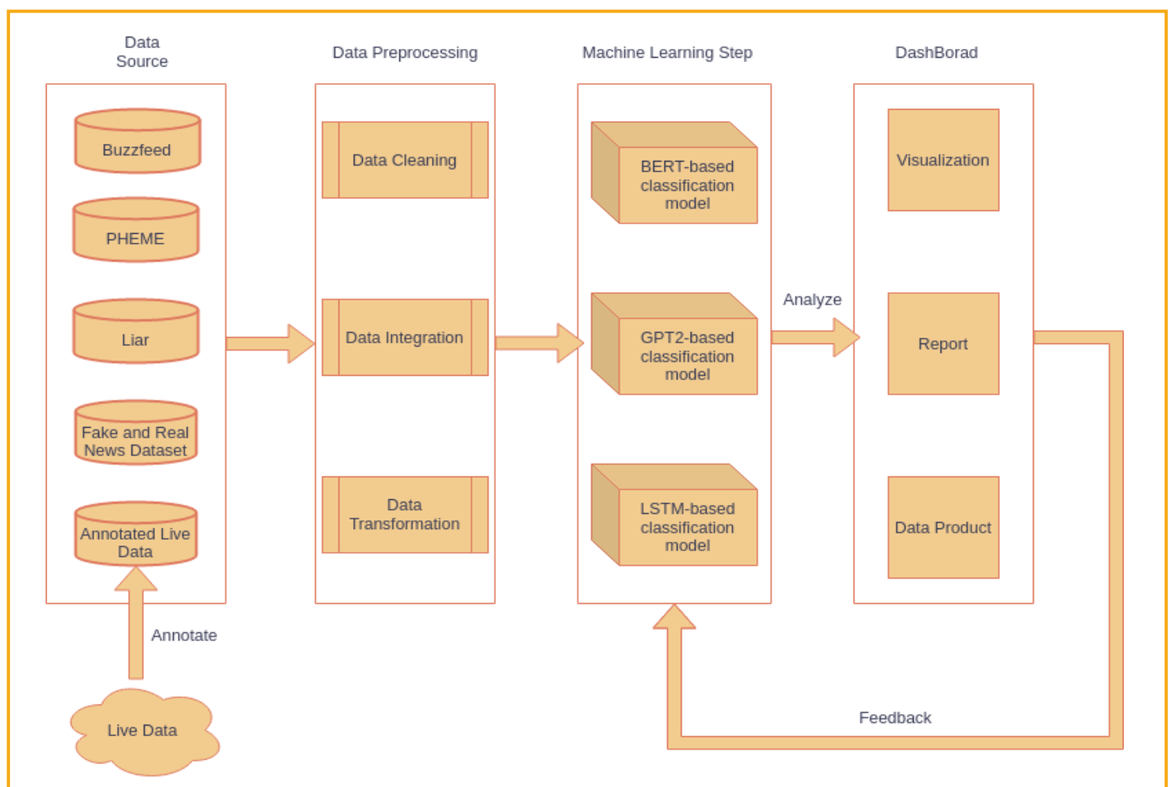
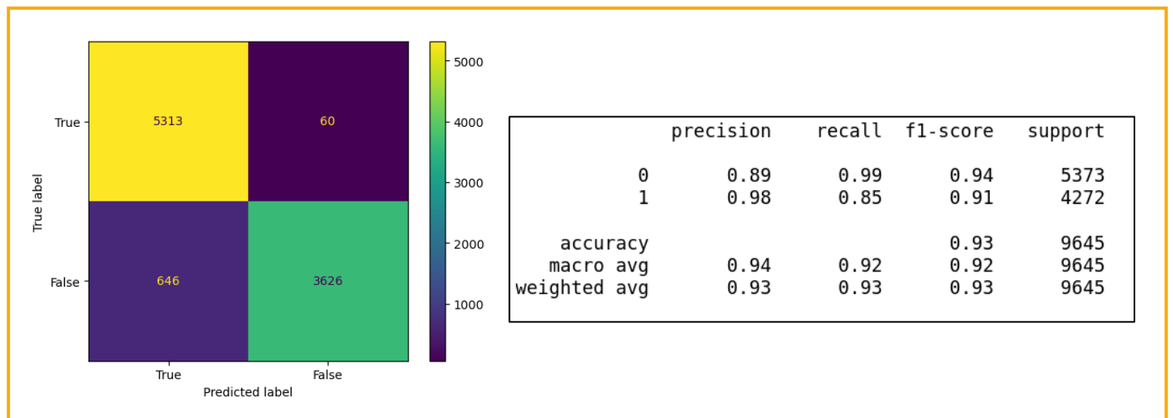


Figure 1. Project architecture. Data from different sources will be fed into three models after cleaning, integration, and transformation. A frontend dashboard will display demo and analysis results



	precision	recall	f1-score	support
0	0.89	0.99	0.94	5373
1	0.98	0.85	0.91	4272
accuracy			0.93	9645
macro avg	0.94	0.92	0.92	9645
weighted avg	0.93	0.93	0.93	9645

Figure 2. Confusion matrix and classification report of GPT-2-based fake news detector on test set

References

1. E. Kochkina, M. Liakataand A. Zubiaga, "PHEME dataset for Rumour Detection and Veracity Classification". 10-Jun-2018.
2. "Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate". BuzzFeedNews, October 20, 2016.
3. W. Y. Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection". vol 2, pages 422–426, Association for Computational Linguistics, 2017.
4. H. Ahmed, I. Traoré, & S. Saad, "Detecting opinion spams and fake news using text classification". Security and Privacy, 2018.
5. H. Ahmed, I. Traoré, & S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques". International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 2017

3_datamen
Song Lin
Shlok Nangia
Jialiang Guo

Scan the QR code to checkout our detailed **Medium article.**