AMCIS 2024 Proceedings

Americas Conference on Information Systems (AMCIS)

August 2024

Real-Time Fake News Detection on the X (Twitter): An Online Machine Learning Approach

Piyush Vyas
Texas A&M University-Central Texas, piyush.vyas@tamuct.edu

Jun Liu
Dakota State University, jun.liu@dsu.edu

Shengjie Xu
University of Arizona, sjxu@arizona.edu

Real-Time Fake News Detection on the X (Twitter): An Online Machine Learning Approach

Emergent Research Forum (ERF) Paper

Piyush Vyas
Texas A&M University-Central Texas
piyush.vyas@tamuct.edu

Jun Liu
Dakota State University
jun.liu@dsu.edu

Shengjie Xu
University of Arizona
sxu14@ieee.org

Abstract

Fake news, along with the speed of mass communication via social media, is having a significant impact on our social life, particularly in the political world. Fake news detection on social media has recently become an emerging research area that is attracting tremendous attention. Existing studies have proposed to employ traditional machine learning (ML) or emergent deep learning (DL) methods to detect fake news. These ML or DL methods based on batch processing and learning from a training dataset, however, are inefficient in conducting continual real-time fake news detection for incoming unseen social media data. In this research, we propose to use online machine learning (OML) to automatically identify fake news in real time. We investigated various online learning algorithms, including Approximate Large Margin Algorithm (ALMA), Passive-Aggressive (PA), etc., and compared their performance with some existing ML or DL methods for real-time fake news detection on Twitter. The preliminary results of our study demonstrate the considerable potential of OML techniques in classifying real-time fake news, thus highlighting the adaptability and robustness of OML in handling dynamic information streams such as fake news.

Keywords (Required)

Online machine learning, fake news, misinformation, X, Twitter, and social media.

Introduction

The dissemination of online misinformation carries significant repercussions for individuals, businesses, and society at large. Misinformation, such as fake news, constituting forms of deliberate propaganda, is disseminated intentionally, often through social media platforms. Fake news, a subset of online misinformation, is crafted to deceive and manipulate the public, typically for financial or ideological motives. Its manifestation can range from traditional print to online content. The circulation of fake news is shaped not only by the intentions of content creators but also by the dynamics of the propagation process, influenced by individual judgments and global circulation trends. A survey revealed a notable lack of trust in the accuracy and fairness of news among two-thirds of U.S. adults (Watson 2024). Additionally, a Pew Research (2023) study indicated that 55% to 65% of respondents urge governmental and high-tech company intervention to curb misinformation, even if it involves limiting freedom of speech. These revealing insights underscore the urgency of addressing this detrimental issue, prompting academic attention to detecting fake news, a task intricately tied to the text classification process. The utilization of online machine learning (OML) in text classification is increasing, primarily driven by the proliferation of extensive and dynamic unstructured textual data on the Internet (Barve et al. 2020). Thus, the overarching objective of this research is to automatically classify real-time fake news vs. real news using online machine learning (OML). Therefore, our research question is:

RQ1: How effective are OML techniques in automatically classifying real-time fake news vs real news on X (Twitter)?

This work employs a dataset comprised of news tweets for both training and evaluating OML techniques. OML is a different approach compared to traditional or batch machine learning (ML). In OML, the model learns and updates itself one data point at a time. It adapts to new information as it comes in, making it more responsive to real-time changes. On the other hand, in traditional ML, the model is trained using the entire dataset at once, which can be computationally intensive and might not capture immediate updates. OML is akin to learning from small, continuous streams of information like tweets, offering a more dynamic and real-time learning process to classify the news.

As a potential research gap, we identified that existing studies are integrating traditional ML and deep learning (DL) methods to counteract widespread misinformation. These algorithms prove inefficient in managing incoming data over time and executing categorization for continual misinformation detection. Notably, ML and DL techniques lack support for Online Learning, constraining their efficacy in real-time fake news detection settings (Burdisso et al. 2019). Conversely, our utilized OML techniques involve continuous data input in the form of streams, with old data discarded post-model updates, eliminating the need for repetitive model retraining to attain real-time online learning. Moreover, OML techniques learned from small streaming data batches exhibit the potential for enhanced performance in knowledge gain, requiring less time and memory space (Shan et al. 2020). Therefore, we have employed the OML techniques to classify the news tweets to utilize the aforementioned advantages.

Literature Review

Several recent studies contribute to the comprehensive understanding of fake news detection using machine learning approaches. Salh and Nabi (2023) delve into the application of Support Vector Machine (SVM), Random Forest Model (RF), and Convolutional Neural Network (CNN) for detecting fake news, particularly focusing on the Kurdish language. Hamsheen and R. Flah (2023) emphasize the identification process of fake news using the Passive-Aggressive Classifier (PAC) and highlight the inadequacy of attention given to less-resourced languages, along with the absence of labeled fake corpora and fact-checking websites. Farooq et al. (2023) propose a fake news classifier for Urdu news employing machine learning techniques, utilizing Term frequency-inverse document frequency (TF-IDF) and Bag of Words (BoW) with n-grams. They also note the existing approaches' lack of robustness for multi-domain datasets in Urdu news. Reddy et al. (2023) focus on a content-based approach using machine learning ensemble techniques for detecting fake news, defining fake news as information produced to deceive readers for monetary gain. Collectively, these studies assert insights into the diverse applications of machine learning for fake news detection across different languages and domains. Table 1 further shows the more relevant studies' focus, datasets used, and key findings.

Authors	**Focus**	**Dataset**	**Key findings**
(Salh and Nabi 2023)	To detect fake news in Kurdish.	Kurdish Dataset (KDFND)	Deep learning-based models perform well.
(Hamsheen and R.Flah 2023)	To build a system to detect any type of news in Kurdish.	RFkurdish news data	Passive-Aggressive Classifier (PAC) has the potential to perform well.
(Farooq et al. 2023)	To develop a fake news classifier for Urdu news.	Urdu language news data	Feature stacking with different machine learning models improved the performance.
(Reddy et al. 2023)	To automate the detection of fake news.	ISOT news articles dataset	ML has the potential to automate the detection of fake news content.

Table 1. The aim, data, and findings of the existing study.

Moreover, research conducted by Shalev-Shwartz et al. (2007) has delved into the domain of online machine-learning techniques. Han et al. (2021) emphasize the continuous growth of online data, paralleled by the dissemination of fake news, necessitating recurrent retraining of fake news detection models on new data—a potentially expensive endeavor. However, methodologies such as Gradient Episodic Memory (GEM) and Elastic Weight Consolidation (EWC) offer viable solutions, allowing the integration of historical and new data for model training with minimal computational overhead. Studies by Babu et al. (2021) and Nikam and Dalvi (2020) concentrate on the adoption of the passive-aggressive (PA) algorithm, which is recognized for online learning and commonly employed in classification tasks. The PA algorithm operates passively with correct predictions and adjusts aggressively in the face of incorrect predictions, earning its designation as a passive-aggressive algorithm.

Given the continuous stream of data on the internet, treated as infinite word sequences, it becomes imperative to train models continuously, one data point at a time. Nikam and Dalvi (2020) specifically employ real-time Twitter streamed data, utilizing the PA technique and extracting TF-IDF features for fake news detection, while Babu et al. (2021) showcase the efficacy of PA in detecting large-scale fake news on social media.

Method

Data collection and preprocessing

To assess the effectiveness of our proposed model in classifying news tweets, we relied on labeled datasets prepared by (Han et al. 2019) that were originally provided by PHEME (Zubiaga et al. 2016), a reliable source. This dataset comprises short news tweets related to real-world events like the Boston attack. Tweets in this dataset were categorized as either "rumor" or "non-rumor." Specifically, there are 1,238 rumours and 3,714 non-rumours tweets. This openly accessible dataset is structured in JavaScript Object Notation (JSON) format, featuring various directories containing multiple tweets. The labeling convention designates 1 for rumors and 0 for non-rumors, facilitating a comprehensive analysis and classification of news tweets.

Preprocessing plays a pivotal role in readying textual data for model learning due to the frequent presence of redundancies and irregularities. Employing Natural Language Processing techniques (NLP), the textual data underwent systematic preprocessing for subsequent analysis. Various components, including stop words, special characters ('!', '&', '$'), emojis, symbols, repetitive period signs ('..' or '...'), white spaces, line breaks, blank rows, and additional variables like tweet ids and user ids, were removed. Specifically, stop words, comprising articles, pronouns, and prepositions, were eliminated, as they hold grammatical significance in English but lack semantic importance in the model's learning process. Further, we lowercase all words to ensure uniform treatment during the encoding process. Tokenization has been performed to segment sentences into distinct tokens (words). We have used the River Python-based library for the OML implementation. River supports various data sources, including in-memory data, file-based data, and streaming data. As our data is file-based (comma-separated file (CSV)), River usually processes it in streams (one record at a time) to feed into the OML technique for training. Whereas in traditional ML techniques, data is fed into batches (a group of records) prior to training an ML technique.

OML classifiers and evaluation

Any OML approach that is being trained uses data streams, which are different from batch learning and enable real-time text classification tasks like detecting fake news. Using previously obtained labeled datasets as training data, an OML approach may be utilized as a classifier to predict the label for new unknown data. According to Bifet et al. (2023), OML classifiers learn about new input sequentially, discarding previous knowledge. Online learning, according to Hoi et al. (2021), is a collection of machine learning approaches where a model learns from a succession of sequential data sequences in an “on-the-go” manner (i.e., learn as new data arrive).

Enhancing the accuracy of the OML model's sequence of predictions is the only goal of online learning. Learning from continuous data streams has been shown to be a viable strategy for online learning (Hoi et al. 2021). We have used a range of techniques from the family of supervised binary classifiers, including Ensemble - Adaptive Boosting (AdaBoost), Bagging, Passive-Aggressive binary classifier (PA-BC), Multinominal Naïve Bayes (MNB), Approximate Large Margin Algorithm (ALMA), Bagging, and Tree-based - Hoeffding Tree Classifier (HDT), to determine which OML techniques are best to implement in any business setting for early fake news detection. Here, ALMA utilizes algebraic representations, emphasizing robustness to statistical properties and focusing on reducing bias(error). PA adjusts its learning incrementally based on prediction accuracy. MNB proves effective in text classification, assuming conditional independence of features and displaying adaptability to evolving data distributions. Ensemble methods like Adaboost and Bagging demonstrate the power of combining weak learners and incrementally updating and aggregating predictions. The Tree-based HDT excels in incremental decision tree algorithms, efficiently adapting to streaming data with a focus on memory efficiency. Collectively, these techniques contribute to the dynamic landscape of online learning, particularly in scenarios where data arrives sequentially as news tweets.

The OML models were applied to the test datasets, and the classification performance was evaluated using classical metrics such as precision, recall, F1 score, and accuracy. In this context, precision measures the accuracy of positive predictions, recall quantifies the model's ability to capture all actual positive classes, and the F1 score balances precision and recall. Accuracy assesses overall correctness by considering both true positives and true negatives. To calculate the values of all these above-mentioned measures, the values of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were used. These metrics provide a comprehensive overview of the classification effectiveness of each OML technique in the news classification task.

Preliminary results and discussion

Techniques	**Accuracy**	**Precision**	**Recall**	**F1 Score**
ALMA	96.30%	96.33%	96.30%	96.31%
PA-BC	99.15%	99.16%	99.15%	99.15%
MNB	97.46%	97.67%	97.46%	97.49%
HDT	74.84%	61.60%	74.84%	64.32%
Adaboost	74.78%	67.59%	74.29%	67.12%
Bagging	75%	56.25%	75.00%	64.29%

Table 2. Evaluation measure's scores for OML techniques for news tweet PHEME dataset.

When we fed data for traditional ML techniques, we used the Sklearn python library that usually processes fixed datasets in batches, splitting data into training and test sets upfront prior to training and assessing the model. Whereas, River (python library for OML) handles the provided data as streaming data, learns sequentially from one observation at a time, and then evaluates the performance on the fly.
In response to our research question concerning the efficacy of OML techniques in real-time fake news classification, the results obtained from various OML models are shown in Table 2. By utilizing River (Montiel et al. 2020) Python-based library, the OML techniques were trained on fake news tweet data. The datasets were supplied to techniques in the form of the data stream in a one-at-a-time fashion. Table 2 depicts the news classification performance metrics for various OML techniques. Each technique is evaluated based on four key metrics: Accuracy, Precision, Recall, and F1 Score. For ALMA, the results show an accuracy of 96.30%, precision of 96.33%, recall of 96.30%, and an F1 Score of 96.31%. The PA-BC technique demonstrates exceptional performance with an accuracy of 99.15%, precision of 99.16%, recall of 99.15%, and an F1 Score of 99.15%. MNB achieves an accuracy of 97.46%, precision of 97.67%, recall of 97.46%, and an F1 Score of 97.49%. HDT exhibits an accuracy of 74.84%, precision of 61.60%, recall of 74.84%, and an F1 Score of 64.32%. Adaboost and Bagging techniques yield 74.78% and 75.00% accuracy rates, respectively, with varying precision, recall, and F1 Score values. Further, we compared that our best performer, PA-BC, has outperformed the existing approaches, such as (Farooq et al. 2023) achieved 93.39% accuracy, whereas our PA-BC received 99.15% accuracy.

Conclusion

Overall, our preliminary work demonstrates the considerable potential of OML techniques in classifying real-time fake news on Twitter, surpassing existing ML-based approaches such as (Farooq et al. 2023) . The preliminary results highlight the adaptability and robustness of OML in handling dynamic information streams such as fake news generation. This research not only contributes to the practical implementation of robust systems for combating misinformation but also offers valuable insights for academia. The proposed OML approach serves as a foundation for future research, showcasing its efficacy in dynamic fake news detection and inspiring further exploration in the realm of innovative online learning methodologies. As a limitation, the datasets used have label imbalance, and all models were trained on small-scale data.

Thus, to overcome these limitations, in our next iteration, we aim to enhance OML models through improvements and hyperparameter optimization, including data scaling and advanced textual feature extractions.

REFERENCES

Babu, A. M., Divya, K. J., and Joffi, A. 2021. *HAWKSEYE-A Machine Learning-Based Technique for Fake News Detection with IoT*, IJISRT, pp. 620–624.

Barve, Y., Mulay, P., and others. 2020. "Bibliometric Survey on Incremental Learning in Text Classification Algorithms for False Information Detection," *Libr. Philos. Pract* (2020), pp. 2388–2392.

Bifet, A., Gavalda, R., Holmes, G., and Pfahringer, B. 2023. *Machine Learning for Data Streams: With Practical Examples in MOA*, MIT press.

Burdisso, S. G., Errecalde, M., and Montes-y-Gómez, M. 2019. "A Text Classification Framework for Simple and Effective Early Depression Detection over Social Media Streams," *Expert Systems with Applications* (133), Elsevier, pp. 182–197.

Farooq, M. S., Naseem, A., Rustam, F., and Ashraf, I. 2023. "Fake News Detection in Urdu Language Using Machine Learning," *PeerJ Computer Science* (9), PeerJ Inc., p. e1353. (https://doi.org/10.7717/PEERJ-CS.1353/SUPP-2).

Hamsheen, E. S., and R.Flah, L. 2023. "Improvement Performance by Using Machine Learning Algorithms for Fake News Detection," *Zanco Journal of Pure and Applied Sciences* (35:2), Salahaddin University - Erbil. (https://doi.org/10.21271/ZJPAS.35.2.6).

Han, S., Gao, J., and Ciravegna, F. 2019. *Data Augmentation for Rumor Detection Using Context-Sensitive Neural Language Model With Large-Scale Credibility Corpus*. (https://openreview.net/forum?id=SyxCysRNdV).

Han, Y., Karunasekera, S., and Leckie, C. 2021. "Continual Learning for Fake News Detection from Social Media," in *International Conference on Artificial Neural Networks*, pp. 372–384.

Hoi, S. C. H., Sahoo, D., Lu, J., and Zhao, P. 2021. "Online Learning: A Comprehensive Survey," *Neurocomputing* (459), Elsevier, pp. 249–289.

Montiel, J., Halford, M., Mastelini, S. M., Bolmier, G., Sourty, R., Vaysse, R., Zouitine, A., Gomes, H. M., Read, J., Abdessalem, T., and Bifet, A. 2020. "River: Machine Learning for Streaming Data in Python," *Journal of Machine Learning Research* (22), Microtome Publishing. (https://arxiv.org/abs/2012.04740v1).

Nikam, S. S., and Dalvi, R. 2020. "Machine Learning Algorithm Based Model for Classification of Fake News on Twitter," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 1–4.

PRC. 2023. "Misinformation - Research and Data from Pew Research Center," *Pew Research Center*, , July. (https://www.pewresearch.org/topic/news-habits-media/media-society/misinformation/, accessed February 19, 2024).

Reddy, B. Ravinder, Reddy, B. Rohit, Abhinay, D., and Saisuma, K. 2023. "Fake News Detection an Effective Content-Based Approach Using Machine Learning Ensemble Techniques," *International Journal of Scientific Research in Science and Technology* (10:2), Technoscience Academy, pp. 81–89. (https://doi.org/10.32628/IJSRST52310212).

Salh, D. A., and Nabi, R. M. 2023. "Kurdish Fake News Detection Based on Machine Learning Approaches," *Passer Journal of Basic and Applied Sciences* (5:2), University of Garmian, pp. 262–271. (https://doi.org/10.24271/PSR.2023.380132.1226).

Shalev-Shwartz, S., Singer, Y., and Srebro, N. 2007. "Pegasos: Primal Estimated Sub-Gradient Solver for Svm," in *Proceedings of the 24th International Conference on Machine Learning*, pp. 807–814.

Shan, G., Xu, S., Yang, L., Jia, S., and Xiang, Y. 2020. "Learn\#: A Novel Incremental Learning Method for Text Classification," *Expert Systems with Applications* (147), Elsevier, p. 113198.

Watson, A. 2024. "Statistics & Facts about Fake News | Statista," *Statista*, , January 9. (https://www.statista.com/topics/3251/fake-news/#dossier-chapter2, accessed February 19, 2024).

Zubiaga, A., Liakata, M., and Procter, R. 2016. "Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media," *ArXiv Preprint ArXiv:1610.07363*.