How concept drift can impair the classification of fake news

Renato M. Silva[1,2], Tiago A. Almeida[2]

[1] Department of Computer Engineering, Facens University, Sorocaba, São Paulo, Brazil
renato.silva@facens.br
[2] Department of Computer Science, Federal University of São Carlos (UFSCar), Sorocaba, São Paulo, Brazil
talmeida@ufscar.br

Abstract. Fake news is a serious problem that can influence political choices, harm people's physical and mental health, promote treatments without scientific evidence, and even incite violence. Machine learning methods are one of the leading solutions that have been studied for filtering fake news automatically. However, most studies do not consider the dynamic nature of news, creating static models and evaluating them offline through the traditional holdout or cross-validation. These studies naively assume that news characteristics do not change over time and, therefore, the performance of offline models is preserved as time goes on. In this study, we show how concept drift can impair the classification of fake news. We aim to verify whether the conclusions obtained in studies that disregarded the dynamic nature of the news are sustained. We analyzed how the performance of methods trained in an offline fashion is affected by the news update over time, including concept drift due to impacting events like the Covid-19 pandemic and the United States presidential election. The results showed that the performance of offline models is over-optimistic. Incremental learning methods should be preferred because they can adapt to changes in textual patterns over time.

CCS Concepts: • **Computing methodologies** → **Machine learning**.

Keywords: fake news, online learning, text categorization, machine learning

1. INTRODUCTION

The popularization of social media and instant messaging apps made the production and dissemination of fake news easier, cheaper, faster, and more effective. The abrupt increase in the volume of fake news is worrying because they may cause serious harm to society since they can influence people's opinions and political choices, incite violence, and even be a risk to public health. For instance, in the Covid-19 pandemic, there was mass dissemination of fake news with inaccurate information about the disease, which stimulated behaviors that undermined the efforts of governments and health authorities to implement prevention measures [Galhardi et al. 2020; Salvi et al. 2021]. Fake news became so serious during the pandemic that the term infodemic became popular since the world faced a misinformation pandemic besides the coronavirus pandemic [Zarocostas 2020; Salvi et al. 2021].

The danger of fake news during the pandemic was evidenced by a quantitative empirical study conducted by Galhardi et. al [Galhardi et al. 2020]. Using the notifications of inappropriate content sent by users of the *Eu Fiscalizo* Brazilian application, the authors found that 65% of the fake news received between March 17 and April 20, 2020, taught homemade methods to prevent the spread of Covid-19, and 20% presented methods to cure the disease. With the evidence found in this study, Galhardi et. al [Galhardi et al. 2020, p. 4207] concluded that the spread of fake news could "discredit science and global public health institutions and weaken people's adherence to the necessary preventive care when addressing the epidemic". One of the reasons fake news is a serious problem is because they spread faster than legitimate news due to technological factors, manipulation, and also because a person is more likely to share falsity than truth [Vosoughi et al. 2018].

Machine learning methods are one of the leading solutions that have been studied to filter fake news automatically. However, most studies do not consider the dynamic nature of the problem. For example, some studies use datasets that do not contain information about the date when the fake news was posted ([Faustini and Ferreira Covões 2019; Wang 2017; Ghosh and Shah 2018]). In addition, most studies analyze the problem of fake news using approaches based on offline learning, a paradigm

We gratefully acknowledge the support provided by the São Paulo Research Foundation (FAPESP; grant 2018/02146-6).

that uses batches of data to train the models, often without regard to temporal information [Biesialska et al. 2020]. In this paradigm, methods that cannot update their model incrementally (also known as offline learning methods) are generally used [Silva et al. 2020; Biesialska et al. 2020; Pérez-Rosas et al. 2018; Zhou et al. 2020]. As news content can change considerably over time, the underlying data distribution can also change. It is known as the concept drift phenomenon [Horne et al. 2019; Ksieniewicz et al. 2020]. Therefore, studies that use offline learning and do not consider the temporal variation of news can obtain overestimated results. Problems with dynamic characteristics like fake news detection should be evaluated using an online learning paradigm that considers a continuous stream of training examples provided in sequential order [Biesialska et al. 2020]. Moreover, this kind of problem should be addressed with incremental learning methods because their training process can occur even when memory is limited, and the predictive model can be updated over time to adapt quickly to new data patterns [Silva et al. 2017].

In online learning, data is presented one at a time to the classifier, which yields a judgment. Then, the classifier receives feedback and can compare the predicted class with the gold standard class. Finally, based on the known class, the classifier can update its predictive model. This training approach used in studies that address online learning is known as *prequential approach* (also known as interleaved test-then-train) [Gama et al. 2013]. In most studies, feedback is received immediately after the data is classified. Although this type of feedback is desired, it rarely occurs in real-world scenarios. Therefore, in some studies, especially in spam filtering, other types of feedback are also considered to evaluate methods that address online learning problems, such as the delayed feedback and uncertain feedback. With delayed feedback, the gold standard is given to the classifier sometime later. When uncertain feedback is considered, the classifier only receives feedback for some data [Cormack 2007; Bittencourt et al. 2020]. Despite the importance of how the feedback is given to online learning tasks, few studies investigate its impact on classification performance.

In this study, we show how concept drift can impair the classification of fake news. We investigate the impact of using the online learning paradigm in the classification of fake news to verify whether the conclusions obtained in studies that disregarded the dynamic nature of the news are sustained. We analyze how the news update over time and how impacting events like the Covid-19 pandemic and the United States presidential election may affect the performance of methods trained on an offline learning paradigm. We also evaluate the performance of incremental learning methods in different learning scenarios, varying how feedback is received.

2. RELATED WORK

In natural language processing and related areas, fake news detection has seen some important efforts. Some studies address the problem using linguistic-based features, such as grammatical classes, semantics, spelling errors, and content diversity [Monteiro et al. 2018; Zhou et al. 2004]. Other studies use content-based features using distributive or distributed text representations [Silva et al. 2020]. In studies that use distributive representation, two types of bag-of-words (BoW) are frequently applied: term-frequency representation and term frequency-inverse document frequency (TF-IDF) representation [Silva et al. 2020]. In studies that use distributed representation, the state-of-the-art Word2Vec [Silva et al. 2020; Song et al. 2021; Wang 2017], FastText [Silva et al. 2020; Alves et al. 2019], and GloVe (global vectors) [Kaliyar et al. 2020] models are usually applied.

Fake news detection is frequently formulated as a binary classification problem where the goal is to predict if the news is fake or legitimate [Silva et al. 2020; Monteiro et al. 2018]. However, some studies considered it as a multiclass or multilabel classification problem [Kaliyar et al. 2019; Rasool et al. 2019]. For example, Rasool et. al [Rasool et al. 2019] modeled the problem as a multilabel classification task. In the first level, the document can be classified as true or false. In the last level of the positive class, the document can be classified as mostly-true, true, barely-true, or half-true. Finally, in the last level of the negative class, the document can be classified as false or pants-fire.

Some studies applied traditional machine learning algorithms, such as support vector machines [Silva et al. 2020; Rasool et al. 2019; Monteiro et al. 2018; Zhou et al. 2020], naïve Bayes [Silva et al. 2020; Zhou et al. 2020], random forest [Silva et al. 2020; Zhou et al. 2020; Horne et al. 2019] K-nearest neighbours [Silva et al. 2020], and logistic regression [Silva et al. 2020; Zhou et al. 2020]. More recently, some studies applied deep learning algorithms, such as long short term memory (LSTM) [Alves et al. 2019; Wang 2017], and convolutional neural networks (CNN) [Wang 2017; Kaliyar et al. 2020; Khan et al. 2021]. Even advanced language models (e.g., bidirectional encoder representations from transformers – BERT [Khan et al. 2021]) have been studied for the fake news problem.

Most of the existing studies ignore the chronological order of news or evaluate the methods using an offline learning paradigm [Silva et al. 2020; Zhou et al. 2020; Rasool et al. 2019; Kaliyar et al. 2020; Wang 2017]. However, as new facts occur all the time, and new terms that refer to politicians, celebrities, companies, and new technologies arise frequently, we hypothesize that offline approaches are not appropriate for fake news detection in real-world scenarios.

Other studies in the literature support our hypothesis. For example, Horne et. al [Horne et al. 2019] evaluated the impact of time in fake news classification. They show that the performance of the classifiers slowly degrades as time progress, which can be mitigated using online learning. However, they simulated an online learning scenario using an offline learning method (random forest). Moreover, they only used linguistic-based features instead of the full content. Zhang and Kejriwal [Zhang and Kejriwal 2019] also analyzed the underlying data distribution changes in two tasks related to fake news detection: bias and sensationalism detection. However, they performed experiments using an offline learning paradigm and offline learning classifiers (logistic regression and support vector machine).

Ksieniewi et. al [Ksieniewicz et al. 2020] proposed novel classification methods based on feature extraction techniques to address fake news detection in streaming data from social media. To evaluate the methods, they applied the prequential methodology. The classifiers they applied in the experiments were Gaussian naive Bayes, multi-layer perceptron, and Hoeffding tree. According to the authors, their experiments showed that the online learning strategies helped preserve the classification performance over time. Their study has some goals similar to ours, but they used a dataset called *Getting Real about Fake news*[1] that contains 13,000 articles scraped from 244 websites tagged as "bullshit" by the BS Detector Chrome Extension. The articles from this dataset were published between October 25, 2016, and November 25, 2016. To properly analyze how concept drift can impair the classification of fake news, we believe that the conclusions obtained by Ksieniewi et. al [Ksieniewicz et al. 2020] in data from just one month need to be confirmed by a study that considers a more extended period.

None of the related work presented in this section evaluated concept drift using distributive text representation. The only work that did something similar was Ksieniewi et. al [Ksieniewicz et al. 2020], but they considered a period excessively short. Moreover, to the best of our knowledge, no study analyzed the fake news classification using the online learning paradigm with different types of feedback.

3. MATERIALS AND METHODS

In this study, we intend to fill the gaps in the literature and answer the following research questions:

(1) How concept drift affects the fake news classification problem?
(2) Is the performance of methods trained with the offline learning paradigm sustained by changes in the news patterns?
(3) How different types of feedback impact the performance of incremental learning methods applied to the classification of fake news?

Unlike most studies in the literature, which used datasets that do not contain temporal information or span a brief period, we have used a large and public set of news published over a long period, including impacting events that ensure the concept drift. We performed experiments with the followings datasets:

—NELA-GT-2019 [Gruppi et al. 2020]: it contains 1,200,000 news articles from 260 sources published between January 1st, 2019, and December 31st, 2019.
—NELA-GT-2020 [Gruppi et al. 2021]: it contains 1,779,127 news articles from 519 sources published between January 1st, 2020, and December 31st, 2020. This dataset includes a subset of news about Covid-19 and another subset with 2020 United States presidential election-related articles.

The news from NELA-GT-2019 and NELA-GT-2020 are labeled as unreliable, mixed, or reliable. We performed experiments only with the news labeled as unreliable or reliable.

[1] *Getting Real about Fake news.* Available at `https://www.kaggle.com/mrisdal/fake-news` (accessed on September 23, 2021).

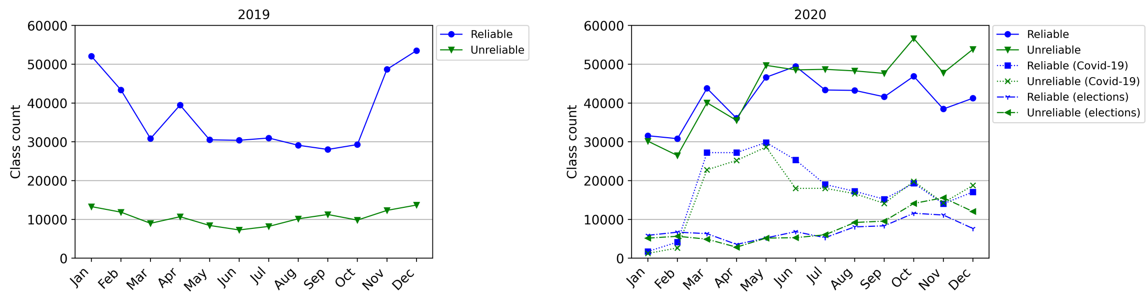


Fig. 1. Number of news from each class.

Figure 1 presents the number of news from each class in each month of 2019 and 2020. The number of unreliable news in the year 2019 is much smaller than the number of reliable ones. Nevertheless, in almost every month of 2020, the proportion of true news and fake news is similar. Another interesting point is that the number of fake news about the Covid-19 pandemic was higher in the first months of the pandemic, while the number of fake news about the presidential election increased with the approach of the elections.

We converted all documents to lowercase and used non-alphanumeric characters as delimiters (except underscore) in the tokenization process. Moreover, in all experiments, we used TF-IDF [Salton and Buckley 1988] to create a vector representation of each document. We used passive-aggressive (PA) [Crammer et al. 2006], an established incremental learning method, to create the classification model. For this, we used the implementation from the *scikit-learn library*[2] with the default parameters. To compare the results, we employed the traditional F-measure.

To investigate how concept drift may affect the fake news classifier, we analyzed how non-ordinary or periodic impacting events like the Covid-19 pandemic and the United States presidential election impair the performance of methods trained on an offline learning paradigm. To properly answer the research questions, we performed experiments with the offline and online learning paradigms, using the protocol presented in Figure 2.

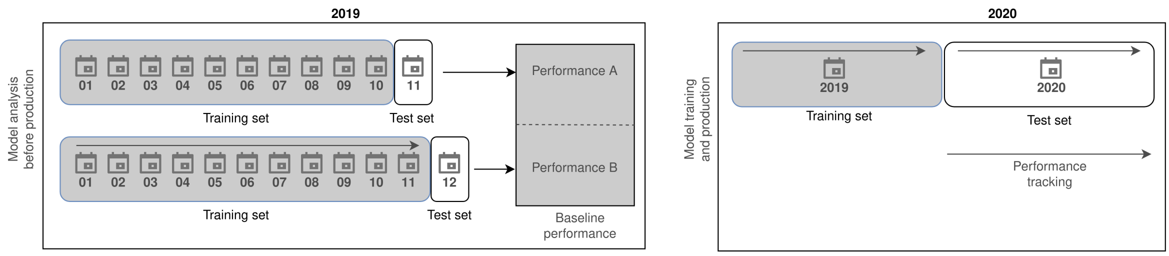


Fig. 2. Experimental protocol.

We first used PA to train a classification model with news from January to the end of October 2019 and test it with the news from November 2019. Next, we used PA to train another classification model with news from January to the end of November 2019 and test it with the news from December 2019, comparing both performances. In this case, if there is no concept drift on these data, the performance achieved in December is expected to be close to that obtained in November. In this way, it is safe to use this performance as an expected baseline in future data. Finally, we trained a model with all news from 2019 and test it with the news from 2020, tracking the model performance continuously.

These experiments follow the traditional classification paradigm, where the model is first statically trained, evaluated, and then put into production. The results obtained with the two test sets, composed of news from November (Performance A) and December 2019 (Performance B), are used as a

[2]*Scikit-learn*. Available at `http://scikit-learn.org/stable/` (accessed on September 23, 2021).

baseline (expected performance). If the expected performance is kept when classifying the news from January to December 2020, we can safely conclude that the model is not affected by the concept drift.

In the offline paradigm, the model is kept static during the year 2020. Thus, it is not robust against concept drift. On the other hand, the model is updated continuously in the online paradigm, allowing it to adapt to changes in the data patterns. In this case, the model can be updated with different types of feedback, as explained in the next section.

3.1 Online learning

To simulate an online learning scenario, we used the news from 2019 for training the classifier. In the test stage, we used the news from 2020 through the *prequential approach* [Gama et al. 2013]: the documents are presented one at a time to test the classifier, which makes a prediction; then, the classifier can receive feedback and, based on the gold standard class, it can update its predictive model. Although the predictive model was updated over time in this scenario, the word dictionary obtained during training was not updated because PA, like several other state-of-the-art online learning methods, expects feature vectors with a fixed size.

To simulate real-world scenarios, we performed experiments varying the way the feedback is presented, following the protocol defined in previous studies to evaluate online learning approaches [Almeida and Yamakami 2012; Cormack 2007; Bittencourt et al. 2020]:

—**Immediate feedback:** it simulates an ideal scenario, where after receiving and classifying a document, the classifier immediately receives feedback and updates its predictive model;

—**Uncertain feedback:** the classifier receives feedback only for some documents; and

—**Delayed feedback:** the classifier receives feedback with a delay.

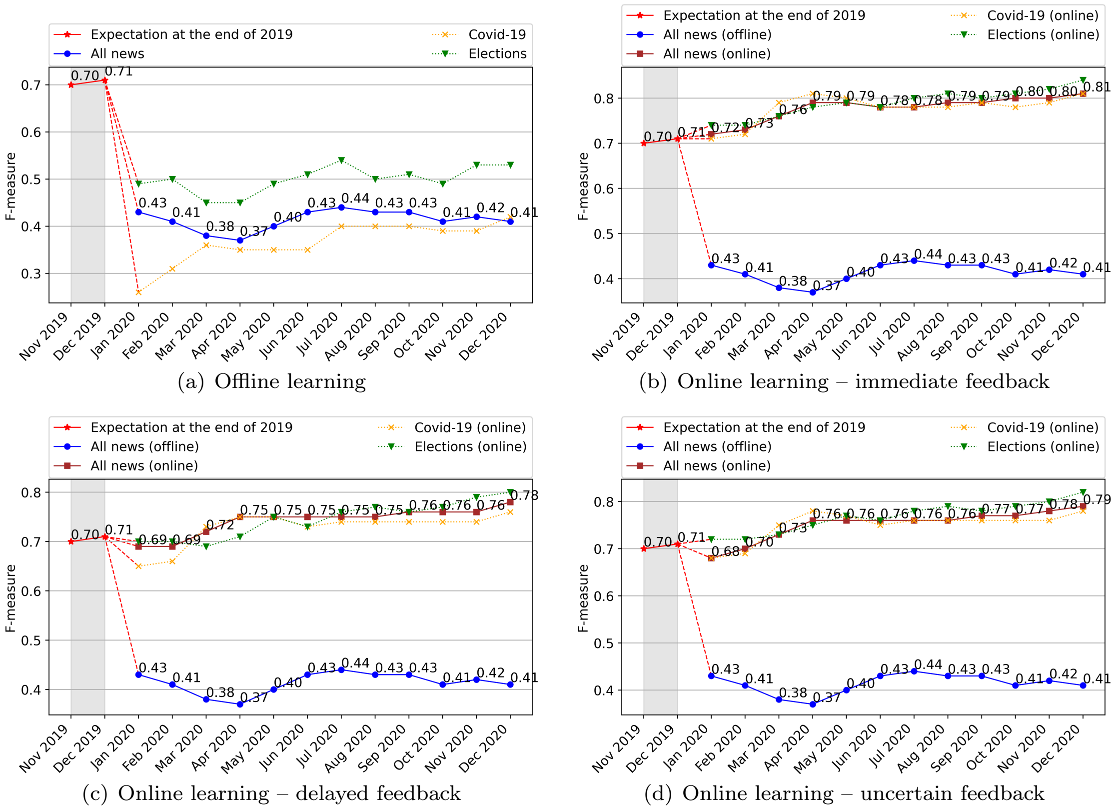
We performed experiments that consider a scenario where the types of feedback presented above are provided only when the classifier predicts a wrong label (i.e., training on error). In experiments with uncertain feedback, the decision whether to send feedback or not is defined randomly. In the delayed feedback, the delay is randomly set between 0 and 20.

In a real-world application, the feedback can vary. For example, the frequency of feedback on a fake news filtering application installed on mobile devices of regular users would likely be different from the frequency obtained on an application used by a news agency. We believe the feedback in mobile applications is likely to be similar to that given by email users [Cormack 2007]. Hardly a user would present the behavior simulated by the immediate feedback scenario, immediately correcting a wrong prediction given by the fake news filter. Thus it offers an optimistic overall performance. In a more realistic setting, some users would present the behavior simulated in the delayed feedback scenario, correcting the wrong predictions with a delay. Finally, some users would also present the behavior simulated in the uncertain feedback scenario, correcting the wrong predictions only for some news. In an application used by a news agency, we believe that prediction errors would be corrected more often but would probably follow the scenario simulated by the delayed feedback scenario due to limited human resources.

4. RESULTS

Figure 3(a) presents the F-measure obtained in the experiments with the offline learning paradigm. The highlighted area with a gray background corresponds to the expected performance (baseline) computed in the last two months of 2019. In this case, the expected F-measure for the next months is around 0.7. If we naively assumed that the future data has no concept drift, we would expect the classifier performance throughout 2020 to be maintained.

The classifier performance suffered a considerable drop in 2020, probably because of a concept drift caused by the appearance of news about Covid-19. The content of these documents may have different data patterns like medical terminology, drugs, and treatments, not seen in the training set. For example, the performance in January 2020 was 39% lower than in December 2019, and in the month with the worst performance (April 2020), the difference to December 2019 was 47%. Therefore, the expected performance at the end of 2019 was overestimated. With these results, we can safely conclude that a static classification model, which does not adapt to changing news patterns over time,



(a) Offline learning
(b) Online learning – immediate feedback
(c) Online learning – delayed feedback
(d) Online learning – uncertain feedback

Fig. 3. F-measure obtained by the classification method.

is unsuitable for detecting fake news in real-world scenarios. These results are also evidence that studies based on offline learning models can present overestimated results.

If we analyze the news about the two major events of 2020 (Covid-19 and United States presidential election), we can see that the results obtained in the classification of news about the Covid-19 were lower than the results related to the election. Although the United States presidential election is an event that only takes place every four years, political news is published frequently. Pandemic-related news uses rarer terms and therefore had more impact on the performance of the classifier. The difference in the F-measure obtained with all the news, with the news of the pandemic, and with the news of the elections is strong evidence of the concept drift phenomenon in the fake news classification task, which reinforces the importance of considering the chronological order of news and using the online instead of the offline learning paradigm.

Figures 3(b), 3(c), and 3(d) presents the performance obtained by the classifier in the online learning scenarios considering the three types of feedback: immediate, delayed, and uncertain.

While the performance of the offline learning paradigm presented a significant drop when classifying the 2020 news, the results in the online learning paradigm were initially similar to the results obtained at the end of 2019 and even increased over time. In December 2020, the classifier performance trained with the offline learning paradigm was about 37% lower than the classifier incrementally updated considering all evaluated feedback scenarios.

The best results were obtained when the immediate feedback scenario was used, which was expected since it is the ideal but over-optimistic scenario. The delayed feedback was the one that most negatively affected the performance. It was lower than the one obtained in the immediate feedback in all months but higher than uncertain feedback in January 2020. However, the differences in performance between the types of feedback are slight. For example, the highest difference in the same month was observed in January 2020, when the F-measure obtained in the uncertain feedback was 5.5% lower than that obtained in the immediate feedback. These results indicate that updating the classifier on error is enough to overcome the concept drift phenomenon. Even in scenarios where the model is updated sporadically, it can adapt to changes in data patterns over time.

If we analyze the performance in the news related only to Covid-19 or the United States presidential election, we can see that the behavior was similar to that observed with the offline paradigm. The results obtained in the classification of news about the Covid-19, in general, were lower than the results related to the election, probably because news about the election has terms or other patterns that are more similar to the news from 2019, as politics is a more explored topic. However, the difference between the performance achieved in dealing with the news of those topics was significantly smaller in the online learning scenarios than in the offline one. For instance, in the offline learning, the performance obtained in the news related to Covid-19 was on average 27% smaller than in the news related to the election, while in the online learning scenario, the difference did not even reach 5% for any of the three types of feedback we evaluated. This slight difference is further evidence that the online learning paradigm has decreased the effect of the concept drift phenomenon.

5. CONCLUSIONS

The problems caused by fake news have drawn people's attention in recent years, especially in major events such as the United States presidential election and the Covid-19 pandemic. Among several studies to mitigate this problem, machine learning methods are one of the biggest promises since the large daily mass of news makes manual analysis infeasible. However, most studies ignore the chronological order of the news and are based on the offline learning paradigm that has no protection against the concept drift phenomenon.

We considered the hypothesis that studies based on the offline paradigm might have overestimated results. Then, in this study, we analyzed: (i) how the concept drift phenomenon affects the fake news classification problem; (ii) if the performance of methods trained with the offline learning paradigm is sustained by impacting events like the Covid-19 pandemic and the United States presidential election; and (iii) how different types of feedback impact the performance of incremental learning methods. We performed experiments with PA, a state-of-the-art incremental learning classification method, considering the offline and the online learning paradigm. We evaluated the classification method in news from 2019 and 2020.

In the experiments with the offline learning paradigm, we observed that the performance of offline classification models degrades over time. The expected performance for the classifier based on the results obtained in the last months of 2019 was very overrated compared with the reality presented with the 2020 news, where two impacting events (Covid-19 and the United States presidential elections) caused a significant change in the data patterns. Moreover, we noticed a big difference between the pandemic-related news results and the election-related news. These results indicated that the underlying data distribution changed over time, and methods trained with the offline learning paradigm may present an overestimated performance. Therefore, we do not recommend offline fake news filters in real-world applications.

In the experiments with the online learning paradigm, the evidence that the fake news classification problem suffers from the concept drift phenomenon was reinforced. Unlike what we observed in offline learning, the performance improved over time, thanks to the incremental update of its predictive model periodically. These results indicate that considering the chronological order of the news and using the online learning paradigm is crucial for adequately filtering fake news.

Another important observation from our study is that the type of feedback used in the online learning paradigm is less critical. This assumption is based on the slight difference between the performance observed in the three types of feedback (immediate, uncertain, and delayed). The results indicated that even in scenarios where the classifier was updated sporadically, it overcame the concept drift phenomenon and sustained its performance.

REFERENCES

Almeida, T. A. and Yamakami, A. Facing the spammers: A very effective approach to avoid junk e-mails. *Expert Systems with Applications* 39 (7): 6557–6561, June, 2012.

Alves, J. L., Weitzel, L., Quaresma, P., Cardoso, C. E., and Cunha, L. Brazilian presidential elections in the era of misinformation: A machine learning approach to analyse fake news. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, I. Nyström, Y. Hernández Heredia, and V. Milián Núñez (Eds.). Springer International Publishing, Cham, pp. 72–84, 2019.

Biesialska, M., Biesialska, K., and Costa-jussà, M. R. Continual lifelong learning in natural language processing: A survey. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), pp. 6523–6541, 2020.

Bittencourt, M. M., Silva, R. M., and Almeida, T. A. ML-MDLText: An efficient and lightweight multilabel text classifier with incremental learning. *Applied Soft Computing* vol. 96, pp. 106699, Nov., 2020.

Cormack, G. V. Trec 2007 spam track overview. In *Proceedings of the Sixteenth Text REtrieval Conference (TREC'2007)*. Gaithersburg, MD, USA, pp. 1–9, 2007.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. Online passive-aggressive algorithms. *Journal of Machine Learning Research* vol. 7, pp. 551–585, Dec., 2006.

Faustini, P. and Ferreira Covões, T. Fake news detection using one-class classification. In *Proceedings of the 8th Brazilian Conference on Intelligent Systems (BRACIS'19)*. IEEE, Salvador, BA, Brazil, pp. 592–597, 2019.

Galhardi, C. P., Freire, N. P., Minayo, M. C. d. S., and Fagundes, M. C. M. Fato ou fake? uma análise da desinformação frente à pandemia da COVID-19 no Brasil. *Ciência & Saúde Coletiva* vol. 25, pp. 4201 – 4210, 10, 2020.

Gama, J., Sebastião, R., and Rodrigues, P. P. On evaluating stream learning algorithms. *Machine Learning* 90 (3): 317–346, Mar., 2013.

Ghosh, S. and Shah, C. Towards automatic fake news classification. *Proceedings of the Association for Information Science and Technology* 55 (1): 805–807, 2018.

Gruppi, M., Horne, B. D., and Adali, S. NELA-GT-2019: A large multi-labelled news dataset for the study of misinformation in news articles. *CoRR* vol. abs/2003.08444, 2020.

Gruppi, M., Horne, B. D., and Adali, S. NELA-GT-2020: A large multi-labelled news dataset for the study of misinformation in news articles. *CoRR* vol. abs/2102.04567, 2021.

Horne, B. D., Nørregaard, J., and Adali, S. Robust fake news detection over time and attack. *ACM Transactions on Intelligent Systems and Technology* 11 (1), Dec., 2019.

Kaliyar, R. K., Goswami, A., and Narang, P. Multiclass fake news detection using ensemble machine learning. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*. pp. 103–107, 2019.

Kaliyar, R. K., Goswami, A., Narang, P., and Sinha, S. FNDNet – a deep convolutional neural network for fake news detection. *Cognitive Systems Research* vol. 61, pp. 32–44, 2020.

Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., and Iqbal, A. A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications* vol. 4, pp. 100032, 2021.

Ksieniewicz, P., Zyblewski, P., Choraś, M., Kozik, R., Giełczyk, A., and Woźniak, M. Fake news detection from data streams. In *2020 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8, 2020.

Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., de Almeida, T. A., Ruiz, E. E. S., and Vale, O. A. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *13th International Conference on Computational Processing of the Portuguese Language (PROPOR'2018)*. Springer International Publishing, Canela, Rio Grande do Sul, Brazil, pp. 324–334, 2018.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, pp. 3391–3401, 2018.

Rasool, T., Butt, W. H., Shaukat, A., and Akram, M. U. Multi-label fake news detection using multi-layered supervised learning. In *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*. ICCAE 2019. Association for Computing Machinery, New York, NY, USA, pp. 73–77, 2019.

Salton, G. and Buckley, C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24 (5): 513–523, Aug., 1988.

Salvi, C., Iannello, P., Cancer, A., McClay, M., Rago, S., Dunsmoor, J. E., and Antonietti, A. Going viral: How fear, socio-cognitive polarization and problem-solving influence fake news detection and proliferation during covid-19 pandemic. *Frontiers in Communication* vol. 5, pp. 127, 2021.

Silva, R. M., Almeida, T. A., and Yamakami, A. MDLText: An efficient and lightweight text classifier. *Knowledge-Based Systems* vol. 118, pp. 152–164, Feb., 2017.

Silva, R. M., de Sales Santos, R. L., Pardo, T. A. S., and Almeida, T. A. Towards automatically filtering fake news in portuguese. *Expert Systems with Applications* vol. 146, pp. 1–48, May, 2020.

Song, C., Ning, N., Zhang, Y., and Wu, B. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing & Management* 58 (1): 102437, 2021.

Vosoughi, S., Roy, D., and Aral, S. The spread of true and false news online. *Science* 359 (6380): 1146–1151, 2018.

Wang, W. Y. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Vancouver, Canada, pp. 422–426, 2017.

Zarocostas, J. How to fight an infodemic. *The lancet* 395 (10225): 676, 2020.

Zhang, S. and Kejriwal, M. Concept drift in bias and sensationalism detection: An experimental study. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ASONAM'19. Association for Computing Machinery, New York, NY, USA, pp. 601–604, 2019.

Zhou, L., Burgoon, J., Twitchell, D., Qin, T., and Nunamaker Jr., J. A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems* 20 (4): 139–165, 2004.

Zhou, X., Jain, A., Phoha, V. V., and Zafarani, R. Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice* 1 (2), June, 2020.