# A Moving Target: Detecting Concept Drift in Brazilian Portuguese Fake News

**Manuela Guedes Wanderley**[1], **Lucca Baptista Silva Ferraz**[1], **Tiago Agostinho Almeida**[2], **Renato Moraes Silva**[1]

[1]Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematics and Computer Science
University of São Paulo (USP), São Carlos, Brazil

[2]Department of Computer Science
Federal University of São Carlos (UFSCar)
Sorocaba, São Paulo, Brazil

manuela.guedes@usp.br,luccaferraz@usp.br
talmeida@ufscar.br,renatoms@icmc.usp.br

***Abstract.*** *Static fake news detectors trained on offline data degrade over time due to concept drift, an understudied phenomenon outside English. This paper presents the first large-scale analysis of concept drift in Brazilian Portuguese fake news. Combining statistical two-sample tests, semantic similarity, and non-parametric change point detection, we quantify the presence and impact of drift, providing explainability by identifying points in time where shifts occur. Our results reveal significant shifts in topical and semantic patterns, demonstrating that model performance can degrade considerably when trained on older data. These findings prove the critical need for adaptive, time-aware methods, and the curation of temporally diverse datasets to build robust defenses against online misinformation in Brazilian Portuguese. The source code for our experiments is publicly available at https://github.com/GDSMN/STIL2025_conceptdrift.*

## 1. Introduction

The efficacy of most machine learning models for fake news detection is predicated on static, offline training, which is limited in handling the temporal dynamics of real-world information ecosystems. This approach leaves models vulnerable to *concept drift*, the phenomenon where the statistical properties of data, including linguistic and topical patterns, evolve over time. A significant body of research on fake news detection disregards the chronological order of data, thereby masking the effects of temporal drift. While concept drift is a recognized challenge in dynamic text classification, a critical gap persists: there is currently no standardized benchmark for evaluating learning methods in text streams subject to drift [Garcia et al. 2025]. This scarcity of evaluation protocols and temporally-aware datasets is a major problem especially for under-resourced languages like Portuguese, impeding progress and hindering comparability across studies.

To our knowledge, no prior work has systematically investigated these temporal dynamics within Portuguese-language misinformation. This omission is critical, as concept drift is especially pronounced during high-impact events, such as elections or health crises, where disinformation tactics evolve rapidly. Prominent studies on Brazilian Portuguese fake news, including those by [Silva et al. 2020], [Chavarro et al. 2023], and [Garcia et al. 2022, Garcia et al. 2024], have overlooked the chronological aspect of the data. By treating the problem as static, these approaches risk creating overly simplified evaluation scenarios and substantially overestimating their real-world performance.

This study addresses this gap by conducting the first extensive analysis of concept drift in Brazilian Portuguese fake news from a temporal perspective. Our central hypothesis is that significant concept drift exists in these data streams, and previous studies that overlooked this may have overestimated their performance. To test this, we employ a multi-faceted methodology using complementary techniques, including two-sample drift detection, semantic similarity measures, and change point detection (CPD), to identify and characterize drift. Our study establishes a methodological baseline for analyzing temporal dynamics in this domain, guided by the following research questions (RQs):

**RQ1:** To what extent is concept drift empirically detectable in temporal streams of Brazilian Portuguese fake news?

**RQ2:** How does this drift manifest in terms of semantic change and the evolution of topical clusters over time?

Our main objective is to address these questions by systematically applying a suite of complementary drift detection techniques. We aim to provide the first quantitative characterization of this phenomenon in the domain and underscore the critical need for time-aware evaluation in the development of robust misinformation detection systems.

The remainder of this paper is structured as follows. Section 2 presents the main related work in the field. Section 3 details the materials and methods employed in the research. Section 4 covers the results. Finally, Section 5 concludes the paper and offers guidelines for future research.

## 2. Related work

Concept drift in data streams is a well-documented phenomenon where the statistical properties of data change over time. These changes can manifest in various forms, including sudden, gradual, incremental, or recurring patterns, each posing unique challenges to model stability [Barboza and de Almeida 2022, Gama et al. 2014]. Methodologies for addressing drift are often categorized as active or passive. Active approaches employ explicit drift detectors to trigger model retraining or updates, while passive approaches use online learning models that continuously adapt to new data [Silva et al. 2023, Silva and Almeida 2021]. Techniques like transfer learning can also mitigate the high computational costs of retraining in dynamic environments [Moradi et al. 2024]. Our work focuses on the prerequisite for most active adaptation strategies: the robust detection and characterization of drift.

A primary challenge in this domain is to first identify if and when drift has occurred. Statistical two-sample tests are a cornerstone of this effort. Methods such as the Kolmogorov-Smirnov (KS) test, and particularly kernel-based methods like the Kernel Two-Sample Test (KTS) [Gretton et al. 2012] and those based on density-difference estimation like the Least-Squares Density Difference (LSDD) [Sugiyama et al. 2013, Bu et al. 2018], are frequently employed to compare the distributions of high-dimensional text representations (e.g., document embeddings) over time.

Recent studies provide a strong methodological precedent for our analysis. [Feldhans et al. 2021], for instance, benchmarked several drift detectors on document embeddings, concluding that KTS and LSDD offered superior performance. In a follow-up, [Feldhans and Hammer 2025] introduced an explainable drift detection method, further validating the utility of these statistical techniques for nuanced textual analysis.

In parallel, research has explored the drivers of drift. [Sarnovský and Babič 2025] demonstrated that shifts in topic distributions within social media streams can directly induce concept drift and degrade classifier performance. This finding highlights the critical link between semantic content evolution and model stability. However, while these

studies establish powerful techniques for detecting drift in English text, they do not address the unique linguistic and cultural contexts of other languages, nor do they focus on the high-stakes domain of fake news. The specific temporal dynamics of Brazilian Portuguese misinformation, therefore, remain an open and critical research question that this paper aims to address. Additionally, most studies on concept drift in text primarily use single techniques, often focusing on two-sample statistical tests. Our study fills this gap by integrating complementary methods: two-sample drift detection, semantic similarity measures, and CPD. This combined evaluation offers a more nuanced and complete understanding of how misinformation evolves over time.

## 3. Methodology

To test our central hypothesis that Brazilian Portuguese fake news is subject to concept drift, we required a dataset with an adequate temporal span and sufficient document volume to enable a meaningful analysis. Our search identified three primary candidates with at least 5,000 documents. The Fake.Br dataset [Silva et al. 2020, Monteiro et al. 2018] consists of 7,200 articles. The FakeRecogna dataset [Garcia et al. 2022] contains 12,398 articles. Finally, the FakeRecogna 2.0 dataset [Garcia et al. 2024] consists of 52,536 news articles, including all the ones in FakeRecogna.

An analysis of the temporal distribution of these datasets (illustrated in Figure 1) reveals that Fake.Br and FakeRecogna cover narrow timeframes, rendering them unsuitable for a robust drift study. In contrast, FakeRecogna 2.0 spans a much longer period. However, it exhibits a significant concentration of fake news instances between 2020 and 2021, a period coinciding with the COVID-19 pandemic, which likely acted as a catalyst for new disinformation narratives.

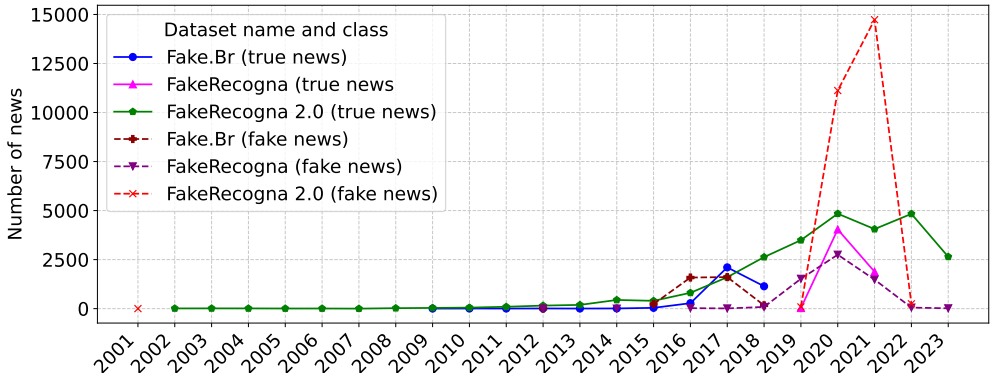


**Figure 1. News published by year.**

Based on this analysis, we selected the FakeRecogna 2.0 dataset for all experiments. To ensure a dense data stream suitable for drift analysis, we further filtered the dataset to a two-year window from January 2020 to December 2021. This specific timeframe contains 25,854 fake news articles, comprising 98% of all fake news in the corpus, thus providing a rich and temporally coherent sample for our investigation.

To convert the news articles into numerical vectors for analysis, we employed two distinct embedding strategies to ensure our findings are not model-dependent: **BERT-based embeddings** and **Word2vec**. For the former, we used the pre-trained BERTimbau Large model [Souza et al. 2020], which generates 1024-dimensional embeddings. This model was trained on the large Brazilian Portuguese Web as Corpus (brWaC) [Wagner Filho et al. 2018] and represents a strong baseline for semantic representation in this language.

For the latter, we trained a custom word2vec model using the continuous bag-of-words (CBOW) architecture, implemented via the Gensim library [Řehůřek and Sojka 2010]. For this model, the text was first preprocessed by replacing numerals, URLs, emails, and social media handles with standardized tokens. The model was trained for 50 epochs on the entire FakeRecogna 2.0 corpus, with a vector size of 100, a window size of 5, and a minimum word count of 3.

Using these two sets of text representations, we applied a series of concept drift detection techniques. The following sections detail the specific statistical tests, semantic similarity measures, and CPD algorithms employed to identify and characterize concept drift in the selected dataset.

### 3.1. Two-sample statistical drift detection

We have employed a suite of two-sample statistical tests to detect distributional shifts between temporal data windows. Our selection of methods — Kolmogorov-Smirnov (KS), Cramér-von Mises (CVM), Kernel Two-Sample Test (KTS) [Gretton et al. 2012], and Least-Squares Density Difference (LSDD) [Bu et al. 2018] — is supported by the comprehensive evaluation presented by [Feldhans et al. 2021], as discussed in Section 2.

While [Feldhans et al. 2021] also evaluated the Concept Drift Bayesian Detector (CDBD), we selected the CVM test in its place. This decision was driven by the scope of our research. Unlike CDBD, the CVM test operates directly on the data distributions, without requiring an auxiliary classifier. As our primary goal is to characterize drift itself, rather than evaluate its impact on classifier performance, the CVM test is a more direct and appropriate choice for our protocol.

The KS, CVM, and LSDD tests were implemented using the Alibi Detect library[1], using their default parameter settings. The KTS test was implemented using the code provided by Olivetti[2]. Since KS and CVM are univariate tests, they are applied to each dimension of the embedding vectors independently. To derive a single multivariate test statistic for each sample comparison, the resulting p-values from each dimension were aggregated using Fisher's method.

### 3.2. Analysis of semantic drift

To complement the statistical drift detectors, which are primarily sensitive to any change in the data distribution, we conducted a more targeted analysis of *semantic drift*. This approach uses similarity metrics to directly quantify changes in the semantic content of news articles over time. By measuring the semantic distance between document sets from different temporal windows, we can characterize the nature of the drift (e.g., whether topics are evolving or shifting in focus). We employed two distinct metrics for this purpose.

The first metric was the **centroid cosine distance**, which captures the overall shift in semantic focus between two temporal windows, $W_a$ and $W_b$. For each window, we calculated the cosine distance between their respective centroids as the mean of all document embeddings within it ($C_a = \text{mean}(\vec{d_i} \in W_a)$). The resulting distance, $1 - \cos(C_a, C_b)$, provides a single scalar value representing the magnitude of the semantic shift between the two periods.

The second metric was the **word mover's distance (WMD)**, which [Kusner et al. 2015] offers a more granular document-to-document similarity measure. It

[1]Alibi Detect library: `https://github.com/SeldonIO/alibi-detect`. Accessed on August 11, 2025.

[2]Code available at: `https://github.com/emanuele/kernel_two_sample_test/`. Accessed on August 11, 2025.

calculates the minimum cumulative distance that words from one document must 'travel' in semantic space to match the words of another. This metric is based on the earth mover's distance and requires a static word embedding dictionary. Consequently, WMD was applied exclusively to our word2vec representations. It is fundamentally incompatible with models like BERT that generate contextualized token embeddings, as they do not provide the necessary fixed, one-to-one word-to-vector mapping.

### 3.3. Change point detection (CPD)

In addition to comparing discrete temporal windows, we employed CPD to identify the specific moments in our continuous data stream where the underlying data distribution shifts significantly. Formally, given a time-ordered sequence of document embeddings $X = (\vec{d_1}, \vec{d_2}, ..., \vec{d_n})$, the goal of CPD is to estimate the set of unknown change points $\{t_1, t_2, ..., t_k\}$ that partition the series into statistically homogeneous segments [Truong et al. 2020].

To solve this problem, we employed the binary segmentation algorithm, a widely used method for offline CPD. We selected a Gaussian kernel-based cost function, as this non-parametric approach is well-suited for high-dimensional data such as text embeddings and can detect diverse distributional changes beyond simple shifts in mean or variance [Garreau and Arlot 2017]. Our implementation leverages the *ruptures* library[3], using its default Gaussian kernel setting. This procedure recursively partitions the time series of embeddings to identify points that maximize the discrepancy between adjacent segments, thus automatically pinpointing significant shifts in the fake news data stream.

While larger time windows are effective for analyzing general trends with statistical tests, CPD algorithms are designed to accurately estimate the precise moment of a distributional change. For this reason, our CPD analysis was performed on daily centroids. Using large time windows runs the risk of oversmoothing the data, which can hide short-lived drifts. Daily variation helps capture sudden changes that could be hidden in larger time windows [Truong et al. 2020].

### 3.4. Experimental protocol

To systematically address our research questions, we designed a multi-faceted experimental protocol. The core of our methodology involves comparing analyses on temporally ordered data against a randomized baseline, allowing us to isolate the effects of concept drift. The protocol was conducted independently for the true and fake news classes and replicated five times to ensure robustness, with results averaged across runs. The following paragraphs detail how this framework is used to evaluate each research question.

**RQ1: To what extent is concept drift empirically detectable in temporal streams of Brazilian Portuguese fake news?** We address our first research question by testing the hypothesis that drift scores will be significantly higher in chronologically ordered data. The two-year data corpus (2020–2021) was partitioned into 53 contiguous, bi-weekly temporal windows. From each window, we randomly sampled 40 articles. This sample size was chosen as the maximum possible to ensure a consistent number of articles across all windows, as some periods contained fewer articles.

As a control, we created a second set of 53 partitions. Each partition again contained 40 articles, but these were sampled randomly *without replacement* from the entire

[3]Ruptures library: `https://centre-borelli.github.io/ruptures-docs/`. Accessed on August 11, 2025.

two-year corpus, thus breaking any temporal ordering. This baseline simulates a scenario with no concept drift, where any detected distributional difference between partitions is due to random sampling variance alone.

Finally, we apply our suite of two-sample statistical tests (KS, CVM, KTS, LSDD) to both the time-ordered data partitions and the randomized baseline partitions. A statistically significant difference in the outcomes between these two conditions serves as direct evidence for the presence of concept drift. To ensure the robustness of our findings and mitigate sampling bias, we repeated both the time-ordered and randomized baseline sampling procedures five times. Each of the five iterations generated a new set of partitions. The drift detection techniques were applied to each iteration independently, and the final reported metrics are the mean and standard deviation calculated across these five runs.

Semantic similarity measures and CPD are applied only to the time-ordered partitions because both methods depend on the temporal sequence of the data. Randomizing the data breaks this temporal structure, preventing the capture of meaningful semantic evolution or the identificatio of change points. Therefore, applying these techniques to randomized partitions would not provide valid or interpretable results.

**RQ2: How does this drift manifest in terms of semantic change and the evolution of topical clusters over time?** Our second research question, focusing on the nature of the drift, is addressed by applying our semantic analysis techniques exclusively to the time-ordered data. We use centroid cosine distance and $MMD^2$ to quantify the magnitude of semantic shifts over time. The WMD is used to analyze the internal topical diversity of the content. Finally, our automated CPD identifies the specific moments when the most significant shifts occurred.

## 4. Results

This section presents and analyzes the results from our experimental protocol, structured to address each research question in turn.

### 4.1. Two-sample drift detection

Our primary finding, directly addressing RQ1, is the clear evidence of concept drift in the time-ordered data. This becomes apparent when comparing the aggregated p-values of the time-ordered samples, shown in Table 1(a), with those from the randomized baseline, shown in Table 1(b). For the most sensitive tests, KTS and LSDD, the mean p-values are substantially lower in the time-ordered condition. For instance, using BERT embeddings, the mean KTS p-value was 0.305 for the time-ordered data versus 0.510 for the random baseline. This discrepancy confirms that the observed distributional shifts are a function of temporal evolution, not merely random sampling variance.

An analysis of the individual detectors reveals important nuances. As shown in the aggregated results, the KS and CVM tests consistently produced high p-values, proving largely insensitive to the subtle changes in this dataset. In contrast, KTS and LSDD demonstrated much higher sensitivity. Furthermore, these sensitive tests achieved lower p-values when applied to BERT embeddings compared to word2vec, suggesting that the richer contextual representations from BERT are more effective for drift detection.

The temporal dynamics of this drift are visualized in Figure 2. The plot shows the p-value for each bi-weekly comparison, where dips below the significance level ($\alpha = 0.05$) indicate drift. Using this threshold, the KTS results indicate drift in fake news around late-March and early-December 2020. These periods likely correspond to shifts

**Table 1. P-value distributions from the two-sample drift detection techniques.**

(a) P-values obtained from time-ordered samples.

| Test | Model | Mean | STD | Quantile | | |
|---|---|---|---|---|---|---|
| | | | | 25% | 50% | 75% |
| CVM | word2vec | 0.806 | 0.164 | 0.768 | 0.841 | 0.904 |
| CVM | BERT | 0.895 | 0.143 | 0.883 | 0.927 | 0.959 |
| KS | word2vec | 0.774 | 0.156 | 0.747 | 0.815 | 0.863 |
| KS | BERT | 0.832 | 0.136 | 0.819 | 0.861 | 0.894 |
| KTS | word2vec | 0.369 | 0.170 | 0.250 | 0.389 | 0.498 |
| KTS | BERT | 0.305 | 0.157 | 0.197 | 0.309 | 0.421 |
| LSDD | word2vec | 0.446 | 0.174 | 0.343 | 0.443 | 0.559 |
| LSDD | BERT | 0.380 | 0.162 | 0.260 | 0.386 | 0.509 |

(b) P-values obtained from randomized samples.

| Test | Model | Mean | STD | Quantile | | |
|---|---|---|---|---|---|---|
| | | | | 25% | 50% | 75% |
| CVM | word2vec | 0.539 | 0.080 | 0.480 | 0.537 | 0.598 |
| CVM | BERT | 0.544 | 0.083 | 0.480 | 0.544 | 0.601 |
| KS | word2vec | 0.621 | 0.077 | 0.574 | 0.619 | 0.676 |
| KS | BERT | 0.761 | 0.068 | 0.719 | 0.764 | 0.812 |
| KTS | word2vec | 0.492 | 0.061 | 0.447 | 0.490 | 0.535 |
| KTS | BERT | 0.510 | 0.062 | 0.466 | 0.515 | 0.552 |
| LSDD | word2vec | 0.490 | 0.045 | 0.462 | 0.495 | 0.521 |
| LSDD | BERT | 0.446 | 0.040 | 0.419 | 0.445 | 0.472 |

in disinformation narratives related to the evolving COVID-19 pandemic, providing a plausible real-world context for the detected statistical changes.

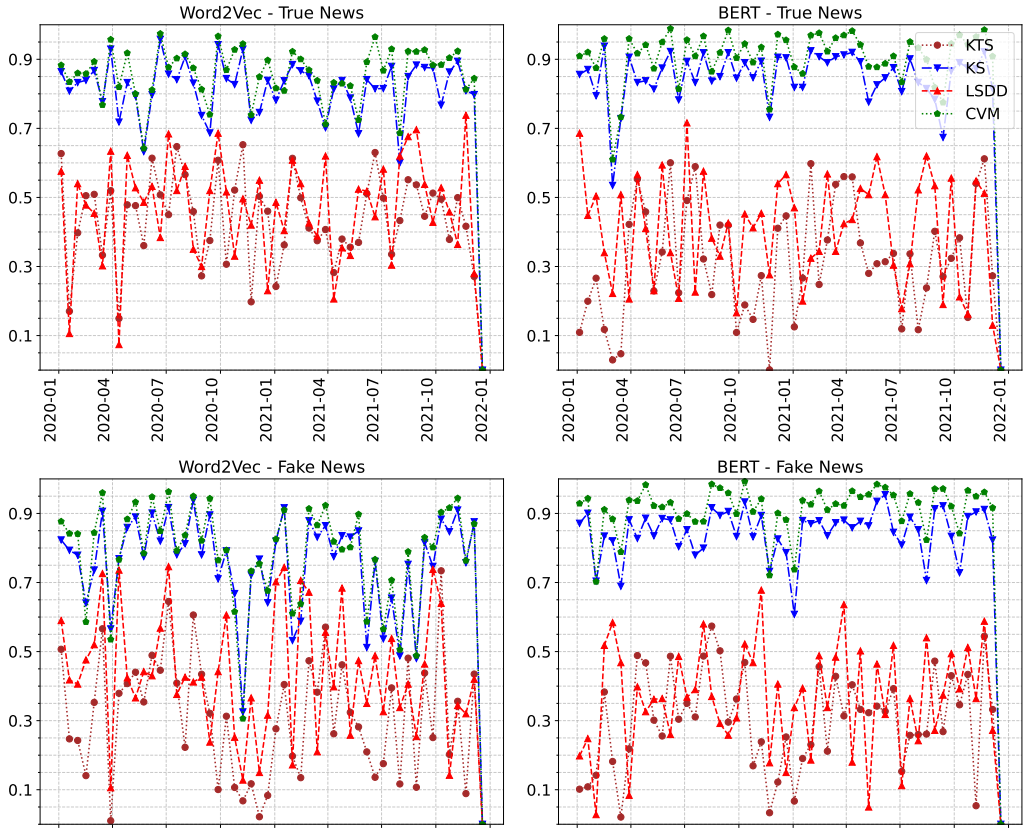


**Figure 2. P-values from two-sample tests comparing each bi-weekly period to the previous one.**

## 4.2. Characterizing drift with semantic measures

Having established the presence of drift, we now address RQ2 by characterizing its nature using semantic measures. This analysis aims to pinpoint when significant shifts occurred and to understand how true and fake news content evolves differently over time. We employed several metrics, including cosine distance between window centroids, a cumulative version of this distance, and the Maximum Mean Discrepancy ($MMD^2$). We conducted these analyses using BERT embeddings because the results in Section 4.1 showed BERT was more effective for drift detection than word2vec.

Figure 3 visualizes the inter-window dissimilarity scores. The results for fake news show pronounced peaks in cosine distance between March and April 2020, and again around January 2021, indicating significant semantic shifts at these times. The cumulative distance metric confirms these periods as major divergence points from the historical semantic mean. Crucially, the $MMD^2$ results corroborate this trend, confirming that the full embedding distributions, not just their centroids, differ significantly during these intervals. While true news also exhibits drift, the magnitude of these shifts is consistently smaller.

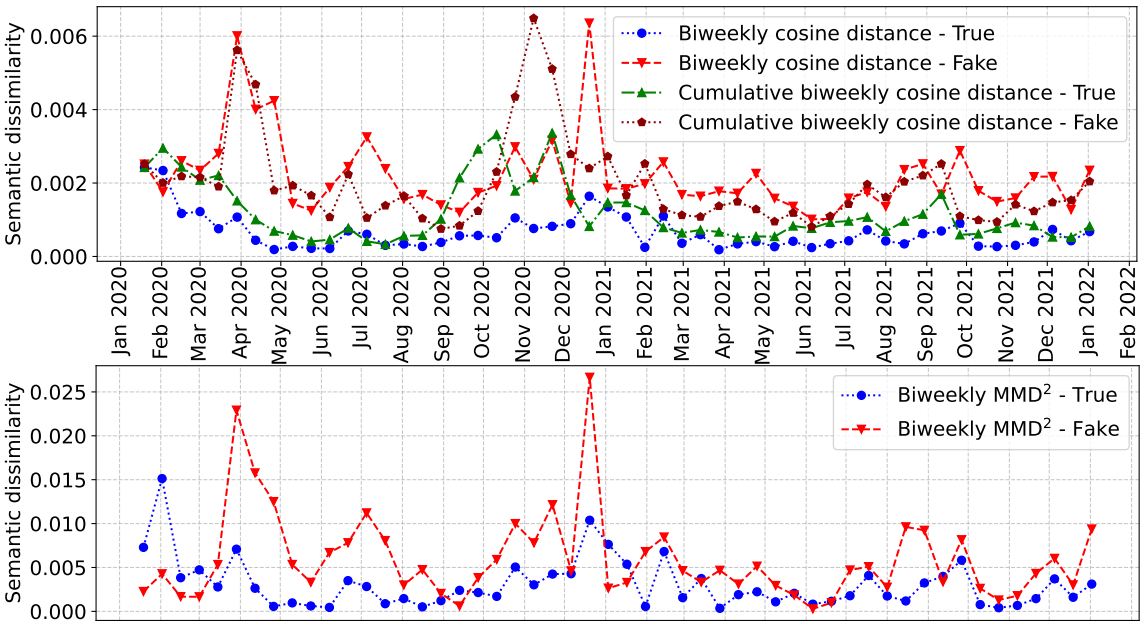


**Figure 3. Semantic dissimilarity between consecutive bi-weekly windows. Top: Cosine distances (consecutive and cumulative). Bottom: $MMD^2$.**

To further investigate the content's evolution, we analyzed the WMD using our word2vec embeddings. As explained in Section 3.2, we used word2vec embeddings in WMD because it requires static embeddings. The distributions, shown in Figure 4, reveal a striking contrast between the two classes. False news consistently exhibits a higher mean and variance in WMD, suggesting a greater degree of semantic novelty and broader topic coverage from one period to the next. Conversely, true news displays a significantly lower mean WMD. This indicates that true news narratives are more semantically stable over time, likely because they consistently report on factual events and follow established journalistic standards, which limits abrupt thematic shifts.

### 4.3. Automated change point detection (CPD)

To complement our previous analyses and objectively identify drift points without manual thresholding, we applied an automated CPD protocol. Based on our findings that BERT embeddings are more sensitive to drift, we focused exclusively on these representations for this analysis. To prepare the data, we first computed the daily centroid of all news embeddings. We then used Principal Component Analysis (PCA) to project this sequence of high-dimensional centroids onto its first principal component, yielding a single univariate signal that captures the dominant mode of variation over time.

The results of applying the binary segmentation algorithm to this signal are shown in Figure 5. The analysis automatically identified three distinct change points for fake

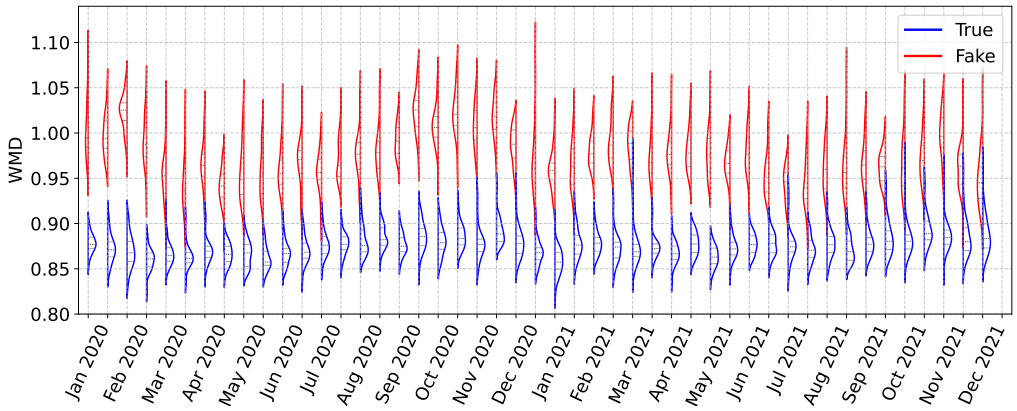


**Figure 4. Distribution of WMD between consecutive bi-weekly windows.**

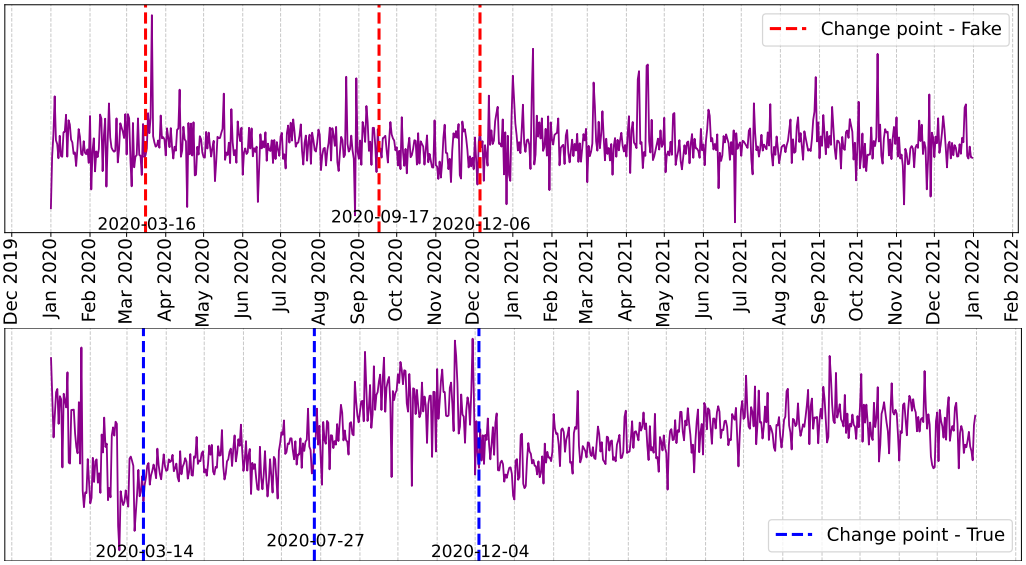


**Figure 5. Primary PCA component of CPD applied to the daily news centroids. Top: values based on fake news. Bottom: values obtained from true news.**

and true news. Crucially, these points align precisely with the periods of high semantic dissimilarity observed in Figure 3, providing objective corroboration for our earlier findings:

- **The COVID-19 pandemic onset (March 2020):** Both true news (March 14) and fake news (March 16) exhibit an abrupt change point, coinciding with the first major peak in cosine distance. This signals the start of a sudden drift in news early in the COVID-19 pandemic.
- **Sustained drift on true news (July 27, 2020):** This point marks the beginning of a sustained drift period that was also identified by the cumulative distance metric and corresponds to a significant jump in $MMD^2$.
- **Sustained drift on fake news (September 17, 2020):** Analogously, the fake news stream undergoes a similar period of sustained drift, but with a notable delay compared to true news. This change point is also identified by the cumulative distance metric and $MMD^2$ jump.
- **The late 2020 stabilization point (December 2020):** Both streams show a final major point in early December (December 4 for true news and December 6 for fake news). This point effectively closes the prolonged drift period of late 2020,

after which the semantic distance metrics in our prior analysis began to stabilize.

Notably, no significant change points were detected in 2021. This automated analysis provides strong evidence that the most significant concept drift in the dataset was concentrated in 2020. It is also noteworthy that the most critical change points — the initial pandemic onset and the late-2020 stabilization — occurred almost simultaneously for both true and fake news, indicating they were driven by the same real-world events.

## 5. Conclusion

This paper presented the first large-scale analysis of concept drift in Brazilian Portuguese fake news. By applying a multi-faceted methodology using statistical tests, semantic dissimilarity measures, and automated CPD, we provided appropriate answers to our research questions. Our results empirically confirm that concept drift is a significant and detectable phenomenon in this domain (RQ1). We further characterized this drift, identifying specific periods of abrupt semantic shifts corresponding to real-world events, providing interpretability to our analysis. Additionally, we uncovered a striking behavioral difference between fake and true news, with the former exhibiting significantly greater semantic variability and temporal drift (RQ2). This pattern may result from the dynamic nature of fake news dissemination. Narratives, whether emerging from coordinated efforts or organically within online ecosystems, are often adapted to align with recent events and maximize audience engagement, or replaced entirely when their relevance declines.

Crucially, our findings demonstrate the inadequacy of static evaluation protocols for this task. The significant discrepancy between results on time-ordered versus randomized data proves that studies ignoring the temporal nature of fake news are likely reporting overly optimistic performance metrics. This highlights a critical methodological flaw in current research practices and underscores the danger of deploying models that cannot adapt to evolving disinformation narratives.

Our analysis is, nevertheless, constrained by the current state of available data. The primary limitation is the scarcity of large-scale, longitudinal fake news datasets for Brazilian Portuguese. While FakeRecogna 2.0 is the most suitable corpus to date, its temporal range is still limited, with a heavy concentration of articles within just a few years. This imbalance poses challenges for a more granular and extended temporal analysis.

These limitations show clear directions for future research. There is a need to curate richer, more diverse, and temporally balanced datasets, potentially incorporating multimodal features to reflect the complexity of misinformation. Building on our findings that Brazilian Portuguese fake news exhibits concept drift, future work should also move from detecting it to mitigating it. Exploring the efficacy of online and incremental learning methods on this data is a critical next step toward building truly adaptive and resilient fake news detection systems for Portuguese and other under-resourced languages.

## Acknowledgments

We gratefully acknowledge the support provided by São Paulo Research Foundation (FAPESP; grants #2024/17834-6 and #2025/13608-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – grant #444933/2024-7). Some researchers involved in this study are also affiliated with the Center for Artificial Intelligence of the University of São Paulo (C4AI – `http://c4ai.inova.usp.br/`), with support by FAPESP (grant #2019/07665-4), the IBM Corporation, and the Ministry of Science, Technology and Innovation, with resources from Law No. 8,248 of October 23, 1991, under the scope of PPI-SOFTEX, coordinated by Softex and published as Residence in TIC 13, DOU 01245.010222/2022-44.

## References


Barboza, E. and de Almeida, P. (2022). Challenges on classifying data streams with concept drift. In *Anais Estendidos do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 126–132, Porto Alegre, RS, Brasil. SBC.

Bu, L., Alippi, C., and Zhao, D. (2018). A pdf-free change detection test based on density difference estimation. *IEEE Transactions on Neural Networks and Learning Systems*, 29(2):324–334.

Chavarro, J., Carvalho, J., Portela, T., and Silva, J. (2023). FakeTrueBR: Um corpus brasileiro de notícias falsas. In *Anais da XVIII Escola Regional de Banco de Dados*, pages 108–117, Porto Alegre, RS, Brasil. SBC.

Feldhans, R. and Hammer, B. (2025). Towards reliable drift detection and explanation in text data. In Julian, V., Camacho, D., Yin, H., Alberola, J. M., Nogueira, V. B., Novais, P., and Tallón-Ballesteros, A., editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2024*, pages 301–312, Cham. Springer Nature Switzerland.

Feldhans, R., Wilke, A., Heindorf, S., Shaker, M. H., Hammer, B., Ngonga Ngomo, A.-C., and Hüllermeier, E. (2021). Drift detection in text data with document embeddings. In Yin, H., Camacho, D., Tino, P., Allmendinger, R., Tallón-Ballesteros, A. J., Tang, K., Cho, S.-B., Novais, P., and Nascimento, S., editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2021*, pages 107–118, Cham. Springer International Publishing.

Gama, J. a., Žliobaitundefined, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Comput. Surv.*, 46(4).

Garcia, C. M., Abilio, R., Koerich, A. L., Britto, A. d. S., and Barddal, J. P. (2025). Concept drift adaptation in text stream mining settings: A systematic review. *ACM Transactions on Intelligent Systems and Technology*, 16(2).

Garcia, G. L., Afonso, L. C., and Papa, J. P. (2022). Fakerecogna: a new brazilian corpus for fake news detection. In *International Conference on Computational Processing of the Portuguese Language*, pages 57–67. Springer.

Garcia, G. L., Paiola, P. H., Jodas, D. S., Sugi, L. A., and Papa, J. P. (2024). Text summarization and temporal learning models applied to Portuguese fake news detection in a novel Brazilian corpus dataset. In Gamallo, P., Claro, D., Teixeira, A., Real, L., Garcia, M., Oliveira, H. G., and Amaro, R., editors, *Proceedings of the 16th International Conference on Computational Processing of Portuguese - Vol. 1*, pages 86–96, Santiago de Compostela, Galicia/Spain. Association for Computational Lingustics.

Garreau, D. and Arlot, S. (2017). Consistent change-point detection with kernels. *arXiv*, arXiv:1612.04740.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(25):723–773.

Kusner, M. J., Sun, Y., Kolkin, N. I., and Weinberger, K. Q. (2015). From word embeddings to document distances. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 957–966. JMLR.org.

Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., de Almeida, T. A., Ruiz, E. E. S., and Vale, O. A. (2018). 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)*, pages 324–334, Canela, Rio Grande do Sul, Brazil. Springer International Publishing.

Moradi, M., Rahmanimanesh, M., and Shahzadi, A. (2024). Transfer learning for concept drifting data streams in heterogeneous environments. *Knowledge and Information Systems*, 66(5):2799–2857.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA.

Sarnovský, M. and Babič, F. (2025). Concept drift influenced by topic change in data streams from social media. In *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 000337–000340.

Silva, R. M. and Almeida, T. A. (2021). How concept drift can impair the classification of fake news. In *Proceedings of the 9th Symposium on Knowledge Discovery, Mining and Learning (KDMiLe'21)*, pages 1–8, Rio de Janeiro, RJ, Brazil. Brazilian Computing Society.

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

Silva, R. M., Pires, P. R., and Almeida, T. A. (2023). Incremental learning for fake news detection. *Journal of Information and Data Management*, 13(6):566–579.

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, Proceedings, Part I*, page 403–417.

Sugiyama, M., Kanamori, T., Suzuki, T., Plessis, M. C. d., Liu, S., and Takeuchi, I. (2013). Density-difference estimation. *Neural Computation*, 25(10):2734–2775.

Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167:107299.

Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brWaC corpus: A new open resource for brazilian portuguese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).