

Drift Detection Methods on Machine Learning Systems: a Discussion over Discrete Live Data

A.C. Rocha

andrecr03@gmail.com

Universidade Presbiteriana Mackenzie
PPG Engenharia Elétrica e
Computação
São Paulo, Brazil

M.B. Oliveira

michel_oliveira47@hotmail.com
Universidade Presbiteriana Mackenzie
PPG Engenharia Elétrica e
Computação
São Paulo, Brazil

L.A.M. Saito

lucia.saito@mackenzie.br
Universidade Presbiteriana Mackenzie
Escola de Engenharia e PPG
Engenharia Elétrica e Computação
São Paulo, Brazil

B.L.S. Lima

bruno.lima@mackenzie.br

Universidade Presbiteriana Mackenzie
Escola de Engenharia
São Paulo, Brazil

L.A. Silva

leandroaugusto.silva@mackenzie.br
Universidade Presbiteriana Mackenzie
Faculdade de Computação e
Informática e PPG Engenharia
Elétrica e Computação
São Paulo, Brazil

Abstract

Context: Machine learning has become an essential tool for addressing complex problems in information systems, encompassing industrial, commercial, and residential applications.

Problem: Machine learning systems without frequent retraining are prone to data and concept drift, compromising predictive accuracy. This issue is particularly critical in scenarios where retraining is infeasible due to high computational costs or data unavailability.

Solution: This study evaluates the performance of drift detection methods in discrete time series with controlled changes in mean and standard deviation using synthetic Gaussian signals.

IS Theory: The General Systems Theory underpins the study by emphasizing how the interplay between drift detection and adaptive systems contributes to maintaining stability and efficiency in dynamic environments.

Method: Experiments were conducted with variations in mean, standard deviation, and both parameters simultaneously in order to obtain qualitative patterns of drift detectors behaviors. The detectors ADWIN, KSWIN, and Page-Hinkley were tested under this scenario.

Summary of Results: The findings reveal that ADWIN and Page-Hinkley exhibited greater precision and robustness, while KSWIN showed excessive sensitivity, leading to a high number of false positives.

Contributions to the IS Field: This research offers a comprehensive analysis of drift detectors' performance, specifically in scenarios involving changes in mean and standard deviation, providing useful reference for designing resilient machine learning-based forecasting systems.

Impacts on the IS Field: The study advances the development of information systems that can adapt to dynamic data environments characterized by shifts in mean and standard deviation, with direct applications in industrial contexts and energy management.

CCS Concepts

- Information systems → Data stream mining.

Keywords

ADWIN, Concept Drift, Data Drift, KSWIN, Machine Learning, Page-Hinkley

1 Introdução

O aprendizado de máquina tem se consolidado como uma ferramenta essencial na resolução de problemas complexos em Sistemas de Informação (SI), abrangendo aplicações industriais [23, 53], comerciais [2, 45] e residenciais [8, 12]. Sua capacidade de analisar grandes volumes de dados e identificar padrões complexos permite a automação de processos e a tomada de decisões mais informadas [16]. No entanto, a eficácia desses sistemas depende da qualidade e da atualidade dos dados utilizados para treinamento e inferência [11].

Uma das principais preocupações em sistemas de aprendizado de máquina é a ocorrência de deriva de dados e de conceito. A deriva de dados refere-se a mudanças na distribuição dos dados de entrada ao longo do tempo, enquanto a deriva de conceito diz respeito a alterações no relacionamento entre as variáveis de entrada e a variável alvo [29]. Essas mudanças podem comprometer a precisão e a confiabilidade das previsões dos modelos, especialmente em sistemas que não podem ser retrainados continuamente devido a restrições de processamento ou disponibilidade de dados [51]. A detecção e a adaptação a essas derivas são, portanto, cruciais para manter a integridade dos sistemas de aprendizado de máquina em ambientes dinâmicos [10].

Esse problema torna-se particularmente relevante em cenários em que o retrainamento contínuo de modelos é inviável. Sistemas que processam grandes volumes de dados enfrentam desafios relacionados ao alto custo computacional e à infraestrutura necessária para armazenamento e processamento [16]. Além disso, em aplicações de tempo real, como dispositivos baseados em Internet das Coisas ou “Internet of Things” (IoT), os dados podem ser inconsistentes ou intermitentes, dificultando o treinamento recorrente [10]. Outro aspecto crítico são as rápidas mudanças no ambiente de operação, que podem alterar os padrões de entrada, resultando em uma perda de eficácia dos modelos previamente treinados [51].

Todos esses obstáculos, para que a informação mais precisa esteja disponível no momento de tomada de decisões estratégicas, são importantes percalços do ponto de vista de teoria de jogos. Essa teoria é amplamente utilizada em diferentes temas, tais como Biologia [25], Computação [32], Economia [34] e Engenharia [47]. Sumariamente, se embasa na tomada de decisão entre dois ou mais indivíduos, chamados de jogadores. Cada decisão, portanto, garante uma recompensa, benéfica ou não, para quem a tomou. Não obstante, também consideram-se dois preceitos fundamentais: que os jogadores sejam racionais, tentando maximizar as recompensas para si, e que apresentem comportamento estratégico, isto é, sabem que seus adversários também estão tentando maximizar os próprios ganhos [40, 41]. Diversos estudos se valem da teoria de jogos para abordar a necessidade de a informação estar disponível, com precisão e em tempo hábil, para melhores tomadas de decisão no âmbito empresarial [4, 7, 50].

Dessa forma, em um cenário corporativo, a falta de planejamento para lidar com as mudanças nos dados compromete diretamente a integridade das previsões, podendo levar a decisões ineficientes, de um ponto de vista estratégico, ou até mesmo prejudiciais, em que o tomador da decisão sai perdendo. Nesse caso, a empresa e toda a cadeia produtiva podem ser afetadas em um efeito em série.

Para sistemas industriais, a deriva de dados pode impactar a precisão de medições provenientes de sensores, afetando o controle de processos [16] e até a vida útil dos equipamentos [33]. Em contextos residenciais, como dispositivos de gerenciamento de energia, a falha em identificar mudanças no padrão de consumo pode levar a previsões ineficazes, reduzindo a eficiência energética [11] ou invalidando sistemas que contam com baterias e cronograma de carga e descarga [52]. Assim, a detecção precoce da deriva, especialmente em séries temporais, é o primeiro passo para garantir que os sistemas de aprendizado de máquina sejam capazes de se adaptar e manter seu desempenho em ambientes dinâmicos [29].

Os sistemas de aprendizado que podem ser afetados não se limitam apenas a aplicações financeiras, mercadológicas ou logísticas [24, 26], mas se estendem por todas as áreas de estudo [13]. Com a crescente demanda por sistemas de informação que já possuam funções de inteligência artificial nativamente [23], quase todas as profissões podem se beneficiar de técnicas de aprendizado de máquina [39]. A deriva, portanto, deve ser tratada com cautela, respeitando a criticidade que os dados apresentam à integridade do sistema.

Como exemplo, um sistema que faz previsões de acordo com dados de células tumorais [21] pode ser bastante impreciso em suas classificações devido à ocorrência de deriva de conceito. Outro exemplo crítico, agora se valendo de séries temporais, que é a situação abordada no presente trabalho, pode ser a de sistemas que fazem previsões contínuas e em tempo real de sinais vitais de pacientes, de modo a ajudar a prever eventos de descompensação clínica, como parada cardíaca, choque séptico ou insuficiência respiratória [54].

Além de todo o exposto, ainda devem ser consideradas disparidades culturais e idiomáticas [15], que podem ser uma fonte passiva de deriva. Nesse caso, uma deriva de conceito poderia surgir simplesmente por não considerar variáveis importantes de outro cenário sócio-cultural do qual o *software* seria acessado.

Ademais, este trabalho também se alinha aos desafios atuais, estabelecidos pela comunidade científica, à área de SI. Em particular, relaciona-se aos desafios de Inteligência Artificial e Ciência de Dados em SI, conforme discutido por Silva e colegas [28], e à necessidade de uma perspectiva sistêmica e socialmente consciente em SI, conforme investigado em Pereira e Baranauskas [36]. Essa conexão reforça a pertinência de abordar detecção de deriva no contexto de SI, pois a adaptação contínua a dados dinâmicos e realidades socioculturais diversas figura entre os principais problemas apontados pela comunidade de SI.

A Seção 2 do presente trabalho relaciona a revisão da literatura acerca de trabalhos semelhantes que abordam aprendizado de máquina em séries temporais, bem como técnicas de detecção de deriva. A Seção 3 relaciona a metodologia, desde a geração dos sinais discretos para o modelo até os testes de previsão em tempo real. A Seção 4 apresenta os resultados obtidos, discutindo as implicações práticas, se valendo de referências expostas na revisão bibliográfica da Seção 2. Por fim, a conclusão relaciona as principais contribuições do presente trabalho para a área de SI, além de sugerir possíveis desdobramentos como escopo futuro.

2 Revisão da literatura

Este capítulo apresenta uma revisão detalhada dos principais conceitos, métodos e abordagens relacionados ao aprendizado de máquina aplicado a séries temporais, com foco na detecção e mitigação de deriva de dados e de conceito. Serão exploradas contribuições acadêmicas de diversas áreas para compreender os desafios e avanços nesse tema, contextualizando o problema do presente trabalho com base em aplicações práticas e perspectivas teóricas.

2.1 Aprendizado de máquina e séries temporais

Os estudos acerca de previsões em séries temporais não são algo recente. Tampouco o são os estudos sobre detecção de deriva em séries temporais. Os métodos Page-Hinkley e soma cumulativa (conhecido como “CUSUM”), por exemplo, datam de 1954, originalmente publicados por Page [35]. Um pouco mais tarde, em 1970, o modelo estatístico “ARIMA”, sigla em inglês para “auto-regressivos integrados de médias móveis”, desenvolvido por G. Box e G. Jenkins em 1970 [5], foi concebido especialmente para trabalhar com séries temporais. Desde então, foi aplicado aos mais variados cenários, como modelagem preditiva para pandemias [1], tráfego de pessoas em cidades [58], séries temporais financeiras [55] e demanda de energia elétrica [46]. No entanto, o modelo ARIMA possui algumas limitações, como não ser eficiente para capturar não-linearidades ou para processar grandes conjuntos de dados ou de muitas dimensões [43]. Essas limitações são determinantes em um mundo cada vez mais conectado e que produz um volume de informação exponencialmente maior em relação ao final do século XX [49].

O aprendizado de máquina tem se mostrado eficaz na superação das limitações dos métodos tradicionais, como o ARIMA, especialmente em contextos que envolvem não-linearidades, alta dimensionalidade dos dados e volumes crescentes de informações [20, 38]. A introdução de algoritmos como redes neurais, máquinas de vetores de suporte (método conhecido também por sua sigla em inglês

“SVM”) e árvores de decisão possibilitou a captura de padrões complexos em séries temporais, resultando em previsões mais robustas [11, 18].

A transição dos métodos estatísticos para os baseados em aprendizado de máquina também foi impulsionada pela explosão de dados provenientes de dispositivos conectados, como sensores IoT. Essa nova realidade exige modelos capazes de processar dados em tempo real e lidar com mudanças dinâmicas nos padrões dos dados [11, 49].

Redes neurais recorrentes (RNN), incluindo as técnicas de LSTM (sigla em inglês para “memória de longo e curto prazo”) e GRU (ou traduzido para “unidade recorrente com porta”), têm demonstrado desempenho superior em diversas tarefas de previsão de séries temporais. Esses modelos são capazes de reter informações de longo prazo, uma característica que métodos estatísticos como o ARIMA não conseguem replicar eficientemente [14, 37].

Modelos baseados em árvores de decisão, como o “XGBoost” e o “LightGBM”, têm sido aplicados em tarefas de previsão em séries temporais multivariadas, demonstrando alta precisão em contextos com múltiplas variáveis de entrada. Esses modelos oferecem maior eficiência computacional e interpretabilidade, tornando-se escolhas frequentes em cenários industriais e comerciais [18]. Avanços recentes, como os “transformers” [27], modelo introduzido originalmente para processamento de linguagem natural, têm sido adaptados para séries temporais, gerando resultados promissores [48]. É importante ressaltar, entretanto, que a eficácia pode ser limitada em certos tipos de séries temporais, indicando que *transformers* podem não superar consistentemente métodos tradicionais em cenários específicos [56].

Por fim, é importante citar a existência de técnicas híbridas, a fim de cobrir as limitações de cada uma das técnicas separadamente [55]. Conforme estudado por Joseph et al. em 2024 [17], que propuseram utilizar um método estatístico como o ARIMA em conjunto com aprendizado de máquina, houve melhorias significativas com o uso dessa combinação. O ARIMA é, então, usado para capturar componentes lineares nos dados, como tendências e sazonais. Esse passo é essencial para decompor a série temporal em padrões previsíveis com base em dependências históricas bem definidas. Em seguida, os resíduos do modelo ARIMA, que representam os padrões não capturados (como não-linearidades), são processados por uma técnica de aprendizado de máquina, como uma rede neural artificial ou um modelo SVM. Os autores constataram uma melhora na precisão das previsões superior a 20%, em termos de erro médio quadrático [17].

2.2 Deriva em previsão de séries temporais

A previsão do consumo de energia elétrica, especialmente em sistemas industriais e redes de distribuição, é fundamental para otimizar a operação, reduzir custos e melhorar a eficiência energética. Modelos de previsão de consumo, como aqueles baseados em aprendizado de máquina, têm mostrado grande potencial para fornecer estimativas precisas em tempo real, permitindo uma gestão mais eficiente dos recursos. No entanto, esses modelos podem enfrentar desafios significativos quando submetidos à deriva de dados e à deriva de conceito, fenômenos que podem comprometer sua acurácia e eficácia ao longo do tempo.

A deriva de dados refere-se a mudanças nas distribuições das variáveis de entrada que alimentam um modelo preditivo. Em outras palavras, quando o padrão de dados que o modelo lidou durante o treinamento começa a se modificar ao longo do tempo, sua capacidade de fazer previsões precisas pode ser prejudicada. No contexto de previsão de consumo de energia elétrica, a deriva de dados pode ser causada por mudanças sazonais no comportamento do consumo, flutuações na carga de máquinas, variações na eficiência dos equipamentos, ou até mesmo mudanças no perfil de demanda dos consumidores [10]. A deriva de conceito, por outro lado, ocorre quando a relação entre as variáveis de entrada e a variável de saída se altera. Isso pode ser particularmente crítico em cenários como a previsão de consumo de energia, em que fatores externos, como mudanças nas condições operacionais de uma planta ou na utilização de novas tecnologias, podem alterar as dinâmicas de consumo sem que o modelo tenha sido treinado para essas novas condições [29].

Em sistemas de previsão de consumo de energia, esses fenômenos são de grande relevância, pois as condições que influenciam o consumo de energia (como fatores climáticos, modificações nos processos de produção ou inovações tecnológicas) podem variar ao longo do tempo, tornando o modelo inicialmente treinado obsoleto. Como resultado, a capacidade do modelo de gerar previsões cessa a decair, e a necessidade de adaptação se torna crítica [51].

O impacto da deriva de dados e conceito na previsão de consumo de energia pode ser abrangente. Para começar, um modelo que não se adapta adequadamente às mudanças nas condições pode gerar previsões imprecisas, o que pode resultar em planejamento inadequado de recursos, desperdício de energia e aumento nos custos operacionais. Além disso, a falha em detectar a deriva pode levar a um aumento gradual no erro de previsão, tornando o sistema cada vez mais ineficiente. A adaptabilidade do modelo é, portanto, um dos principais desafios ao lidar com a deriva, pois as condições operacionais podem mudar rapidamente, e os modelos precisam ser capazes de se ajustar em tempo real para manter sua precisão [11, 16].

Detectar a deriva de dados é um passo crítico, mas também desafiador. Embora seja possível identificar desvios estatísticos nos dados (por exemplo, por meio da monitorização de métricas como a média ou a variância das variáveis de entrada), a deriva de conceito é mais difícil de ser identificada. A mudança nas relações entre as variáveis pode ser sutil e gradual, tornando difícil perceber o momento exato em que um modelo se torna obsoleto [51]. Além disso, a deriva de conceito pode ocorrer em múltiplas dimensões, o que complica ainda mais o processo de detecção e adaptação.

As estratégias de adaptação à deriva são fundamentais para garantir que o modelo continue a fornecer previsões precisas. Existem diversas abordagens que podem ser utilizadas para mitigar os efeitos da deriva, como algoritmos de aprendizado *online* e atualizações incrementais do modelo. Em vez de treinar um modelo completamente novo a cada alteração, esses métodos permitem que a atualização seja contínua, o que possibilita que ele se adapte de forma eficiente a mudanças nas condições de operação [29]. Uma abordagem complementar consiste em usar detecção de deriva, em que o sistema é monitorado constantemente para identificar quando a performance do modelo cessa a se degradar devido à mudança nos dados ou no conceito [10].

O uso de modelos híbridos, que combinam técnicas tradicionais com métodos de aprendizado de máquina, também pode ser vantajoso. Esses modelos podem tirar proveito de dados históricos para capturar as dinâmicas de longo prazo, enquanto as técnicas de aprendizado de máquina podem ser utilizadas para ajustar as previsões em tempo real [16]. Além disso, o uso de algoritmos de aprendizado profundo tem mostrado resultados promissores em tarefas de previsão com grandes volumes de dados, em que a complexidade e a variabilidade das entradas exigem uma adaptação constante dos modelos para lidar com a deriva de dados de maneira eficaz [11].

Por fim, a integração de sensores inteligentes e a computação em borda têm o potencial de fornecer dados em tempo real e permitir uma atualização constante dos modelos, especialmente em sistemas industriais em que a flexibilidade e a capacidade de adaptação são essenciais. Nesse contexto, a adoção de sistemas de monitoramento contínuo de energia, com *feedback* imediato para ajustar as previsões, pode ser uma maneira eficaz de lidar com a deriva de dados [10]. Não obstante, o uso de estratégias de modelagem baseada em caos e análise de séries temporais pode ser útil para entender o comportamento da deriva e antecipar as mudanças que afetam as previsões de consumo [10].

2.3 Previsões em tempo real

A deriva em previsões de tempo real apresenta um desafio superior em relação a contextos estáticos ou de processamento em lote. Nesse cenário, as decisões precisam ser tomadas imediatamente, com base em dados que chegam continuamente ao sistema. Essa dinâmica exige métodos eficientes para detectar e corrigir alterações no padrão de entrada, mantendo a precisão e a confiabilidade das previsões.

Modelos utilizados para previsões em tempo real podem contar com aprendizado também em tempo real, e são projetados para atualizar seus parâmetros continuamente à medida que novos dados chegam. Essa abordagem é particularmente útil em aplicações como previsão de consumo energético, monitoramento de tráfego, controle de processos industriais ou monitoramento de equipamentos [33]. No entanto, esses modelos podem ser suscetíveis a *overfitting*, isto é, quando mudanças transitórias e/ou isoladas são interpretadas como padrões significativos, ou podem falhar em se adaptar rapidamente quando alterações drásticas ocorrem [51].

Métodos de detecção de deriva em tempo real são frequentemente baseados em técnicas estatísticas ou algoritmos de aprendizado adaptativo. Por exemplo, a “janela de estatísticas deslizantes” ou, do original em inglês, “Sliding Window Statistics”, e o “detector adaptativo de janela deslizante” (mais conhecido por sua sigla em inglês “ADWIN”) são ferramentas populares para monitorar mudanças nos dados de entrada [3]. Essas técnicas ajustam dinamicamente a janela de observação para identificar desvios significativos na distribuição dos dados ou nas relações entre as variáveis. Outra técnica que se vale do mesmo princípio de janelamento, em combinação com técnicas estatísticas, é conhecida pela sigla “KSWIN” e foi introduzida em 2009 [9]. Essa sigla é a combinação das iniciais de Kolmogorov A. e Smirnov N., dois matemáticos soviéticos que introduziram o modelo estatístico em 1933 [19], e da abreviação da palavra “windowing” como sufixo da sigla, referente à tradicional

técnica de segmentar um fluxo contínuo de dados para facilitar, ou até mesmo viabilizar, diversos processos computacionais [30].

Outra estratégia consiste no uso de *ensembles*, em que múltiplos modelos são treinados para capturar diferentes aspectos dos dados ou cenários específicos, e o que melhor performar fica encarregado pela previsão [57]. Em sistemas de tempo real, um *ensemble* dinâmico pode ativar ou desativar modelos com base no desempenho recente, garantindo maior resiliência às mudanças nos dados [22]. Por exemplo, em um sistema de previsão de tráfego, um modelo específico pode ser ativado para lidar com picos de tráfego em horários de maior movimento, enquanto outro se concentra em condições normais.

Além disso, métodos baseados em aprendizado por reforço têm demonstrado potencial para lidar com deriva em tempo real. Nesse contexto, o modelo aprende a otimizar suas previsões ao interagir diretamente com o ambiente, recebendo recompensas ou penalidades com base em seu desempenho [44]. Essa abordagem permite que o sistema se ajuste continuamente às mudanças no padrão de dados sem a necessidade de supervisão explícita.

A complexidade da deriva em tempo real aumenta ainda mais quando os sistemas operam em ambientes distribuídos, como redes IoT ou infraestruturas industriais. Nesses casos, as fontes de dados podem ser inconsistentes ou apresentar atrasos, dificultando a detecção de alterações em tempo real. Técnicas baseadas em computação de borda têm sido exploradas para mitigar esses problemas, permitindo o processamento local de dados e a detecção de deriva próxima à origem [42].

A análise de deriva em tempo real não se limita apenas à identificação de mudanças, mas também ao impacto dessas alterações no desempenho do sistema. Modelos dinâmicos baseados em sistemas de equações diferenciais, como o “SINDy” (traduzido do inglês para “identificação esparsa de dinâmica não-linear”), têm sido utilizados para caracterizar as alterações em séries temporais e prever padrões de mudança, fornecendo uma abordagem promissora para a análise de deriva em tempo real [6].

É importante salientar que, em aplicações de previsões em tempo real, os dados chegam em *stream*, ou seja, em fluxo contínuo e potencialmente infinito. Isso reforça a necessidade de detectores de deriva serem capazes de trabalhar sob atualizações incrementais e com recursos computacionais limitados, configurando um importante desafio em meio à crescente vertiginosa de dados gerados por sistemas.

3 Metodologia

Este trabalho apresenta uma análise sistemática do impacto de alterações na média e no desvio-padrão de sinais gaussianos discretos em detectores de deriva de conceito. Três experimentos distintos foram conduzidos para avaliar a sensibilidade dos detectores: variação apenas na média, apenas no desvio-padrão e simultaneamente em ambos. A seguir, detalhamos os passos para a geração dos sinais, a configuração dos detectores de deriva e as métricas utilizadas para análise.

3.1 Geração de sinais discretos

Os sinais discretos considerados no presente trabalho são séries temporais de ruído gaussiano. Em outras palavras, os sinais são

gerados com valores aleatórios limitados em um intervalo real, que obedecem a uma distribuição normal, sendo descritos matematicamente como apresentado em 1.

$$x_k \sim N(\mu, \sigma^2), \quad k \in \{1, 2, \dots, K\}, \quad (1)$$

em que:

- x_k representa o valor da amostra no tempo k ,
- μ é a média da distribuição (valor esperado),
- σ^2 é a variância da distribuição,
- K é o número de amostras por experimento.

Cada experimento foi realizado concatenando três séries temporais. A primeira série é o sinal-base, a segunda série é o sinal alterado e a terceira série é o sinal-base novamente. Percebe-se que o sinal intermediário da sequência é alterado intencionalmente, a fim de causar pontos de variação abrupta para os testes com os detectores de deriva (a ser detalhado na próxima Subseção). Uma vez que os sinais são distribuições normais, a alteração se dá exclusivamente na média, exclusivamente no desvio-padrão, ou em ambos concomitantemente.

Dessa forma, elencam-se:

- (1) Série 1 – (μ_1, σ_1) ;
- (2) Série 2 – (μ_2, σ_2) ;
- (3) Série 3 – (μ_3, σ_3) .

Neste trabalho serão consideradas apenas variações positivas, tanto para a média quanto para o desvio-padrão. Logo, necessariamente, $\mu_2 > \mu_1$ e $\mu_3 < \mu_2$ e/ou $\sigma_2 > \sigma_1$ e $\sigma_3 < \sigma_2$, obedecendo a um passo crescente, conforme detalhado nas Subseções 3.1.1 a 3.1.3.

O número total de amostras por teste foi fixado em $K = 150.000$, e todos os sinais foram concatenados em uma série única. Ainda, foi fixado em 10 passos de tempo para que o sinal sofra a alteração gradual de uma ou das duas variáveis. O valor inicial é fixado em 0,05 e o valor final em 0,50, com passo de 0,05.

3.1.1 Experimento 1: Variação da média.

- **Parâmetro fixo:** $\sigma = 0,02$;
- **Variação:** μ variou de 0,05 a 0,50 com passo 0,05;
- **Sinais gerados:** para cada valor de μ_2 , foram concatenadas as séries de acordo com o exposto:

$$S_1 : \mu_1 = 0,01, \sigma = 0,02;$$

$$S_2 : \mu_2 \in [0,05; 0,50], \sigma = 0,02;$$

$$S_3 : \mu_3 = 0,01, \sigma = 0,02.$$

O primeiro experimento é feito com o sinal sendo alterado em sua média μ . A Figura 1 ilustra o sinal resultante dessa geração.

3.1.2 Experimento 2: Variação do desvio-padrão.

- **Parâmetro fixo:** $\mu = 0,1$;
- **Variação:** σ variou de 0,01 a 0,20 com passo 0,02;
- **Sinais gerados:** para cada valor de σ_2 , foram concatenadas as séries de acordo com o exposto:

$$S_1 : \mu = 0,1, \sigma_1 = 0,01;$$

$$S_2 : \mu = 0,1, \sigma_2 \in [0,01; 0,20];$$

$$S_3 : \mu = 0,1, \sigma_3 = 0,01.$$

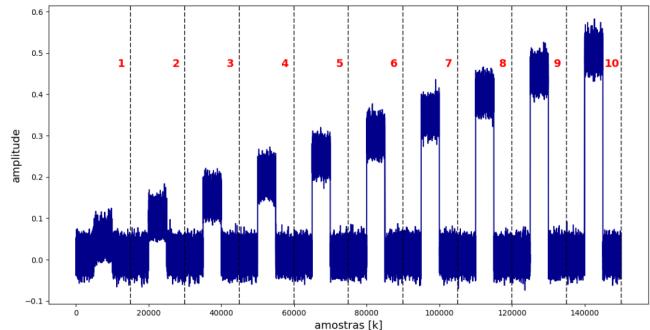


Figura 1: Série de sinais concatenados com alteração na média μ

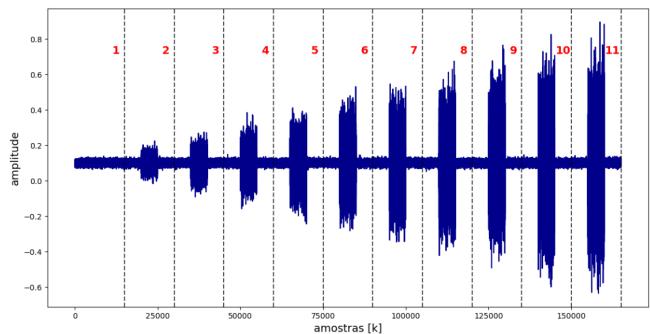


Figura 2: Série de sinais concatenados com alteração no desvio-padrão σ

O segundo experimento é feito a partir da alteração do desvio-padrão σ do sinal. A Figura 2 ilustra o sinal resultante deste experimento.

3.1.3 Experimento 3: Variação de média e desvio-padrão.

- **Variação conjunta:**
 - μ variou de 0,05 a 0,50 com passo 0,05;
 - σ variou de 0,01 a 0,20 com passo 0,02.
- **Sinais gerados:** para cada valor de μ_2 e σ_2 , foram concatenadas as séries de acordo com o exposto:

$$S_1 : \mu_1 = 0,01, \sigma_1 = 0,01;$$

$$S_2 : \mu_2 \in [0,05; 0,50], \sigma_2 \in [0,01; 0,20];$$

$$S_3 : \mu_3 = 0,01, \sigma_3 = 0,01.$$

Por fim, o terceiro experimento é feito a partir da alteração da média μ e do desvio-padrão σ concomitantemente. A Figura 3 ilustra o sinal resultante deste experimento.

3.1.4 Considerações acerca das séries utilizadas. Neste trabalho, cada experimento gera 150.000 amostras ao todo, divididas em três partes concatenadas (S_1 , S_2 e S_3). Para cada valor μ ou σ , é gerada uma sequência que se junta às demais, criando uma única sequência final. A escolha de 150.000 visa garantir volume suficiente para observar várias mudanças e permitir análises estatisticamente robustas. Testes preliminares mostraram que reduções (p.ex., 50.000) ou

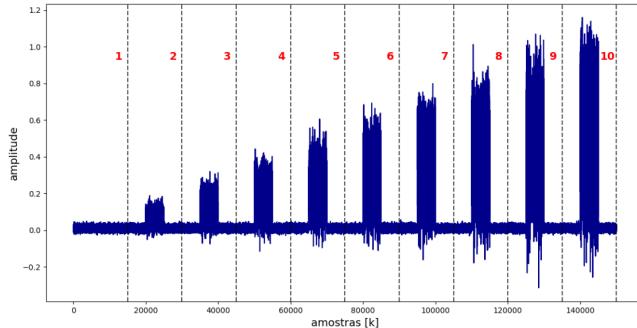


Figura 3: Série de sinais concatenados com alteração concomitante na média μ e no desvio-padrão σ

ampliações (p.ex., 200.000) mantêm comportamentos semelhantes na comparação dos detectores.

3.2 Detectores de deriva e métrica de desempenho

O presente trabalho utilizou três detectores de deriva explorados na revisão da literatura; são eles: ADWIN, KSWIN e Page-Hinkley. Todos os parâmetros para os detectores de deriva foram fixados de acordo com as recomendações originais da biblioteca Scikit-Multiflow [31], que contém diversas ferramentas para tratamento e processamento de dados em *streaming*. Essas técnicas, bem como as métricas de desempenho utilizadas neste trabalho, estão descritas brevemente a seguir.

3.2.1 ADWIN (Adaptive Windowing). O ADWIN (“Adaptive Windowing”) é um método robusto para detecção de mudanças baseado em janelas adaptativas. Ele calcula continuamente a diferença absoluta entre as médias de duas sub-janelas consecutivas, ajustando dinamicamente o tamanho das janelas conforme o fluxo de dados. A derivação é baseada na seguinte condição:

$$|\mu_{\text{win}1} - \mu_{\text{win}2}| > \epsilon \quad (2)$$

em que:

- $\mu_{\text{win}1}$ e $\mu_{\text{win}2}$ são as médias das duas sub-janelas consecutivas;
- ϵ é um valor limiar definido pela precisão estatística e é calculado em função da significância desejada.

O ADWIN se destaca por ser sensível a mudanças abruptas nos dados, o que o torna útil em cenários de fluxo contínuo de informações. Seu principal benefício é a capacidade de ajustar automaticamente o tamanho das janelas para capturar a mudança com maior precisão, minimizando falsos positivos.

3.2.2 KSWIN (Kolmogorov-Smirnov Windowing). O KSWIN utiliza o teste estatístico de Kolmogorov-Smirnov para detectar mudanças na distribuição dos dados em fluxo. Ele avalia a hipótese nula H_0 , que assume que as distribuições de duas janelas consecutivas são idênticas:

$$H_0 : F(X_{k-w}) = F(X_k) \quad (3)$$

em que:

- F representa a função de distribuição acumulada dos dados;
- X_{k-w} é o conjunto de dados anterior (janela histórica);
- X_k é o conjunto de dados atual (janela corrente).

A hipótese nula H_0 é rejeitada se:

$$D > \alpha \quad (4)$$

em que:

- D é a estatística de Kolmogorov-Smirnov, que mede a diferença máxima entre as distribuições acumuladas das duas janelas;
- α é o nível de significância escolhido.

O KSWIN é especialmente eficaz para detectar mudanças não apenas em médias, mas também em outras propriedades estatísticas dos dados, como dispersão e assimetria.

Neste trabalho, o KSWIN foi utilizado com $\alpha = 0.005$ (valor padrão da biblioteca Scikit-Multiflow [31]). Esse nível de significância estatística impacta diretamente a sensibilidade do detector: valores maiores ($\alpha > 0.01$) ocasionam mais falsos positivos, enquanto valores muito baixos reduzem a taxa de detecção em mudanças sutis.

3.2.3 Page-Hinkley. O método Page-Hinkley calcula a soma cumulativa dos desvios em relação a um valor esperado μ_0 , monitorando mudanças em tendências de longo prazo. A derivação ocorre quando a seguinte condição é satisfeita:

$$\sum_{k=1}^K (x_k - \mu_0) > \lambda \quad (5)$$

em que:

- x_k é o valor observado no instante k ;
- μ_0 é o valor esperado de referência;
- λ é um limiar fixo que controla a sensibilidade do método.

O detector Page-Hinkley é eficiente para detectar mudanças graduais em fluxos de dados contínuos. Sua abordagem é particularmente útil em cenários em que desvios acumulados indicam variações significativas no comportamento do sistema.

3.2.4 Métricas de desempenho. Os detectores serão avaliados se conseguiram detectar com precisão e em tempo hábil a ocorrência de deriva. Por “tempo hábil”, refere-se ao número de amostras de tempo k necessárias para que o detector perceba uma alteração de média μ ou desvio σ . Dessa forma, um detector que aponte deriva em até cinco amostras de tempo após a alteração em algum parâmetro pode ser mais rápido que outro detector que só detecte uma alteração no passo 50.

A precisão do modelo foi avaliada de acordo com a definição clássica, considerando apenas as detecções positivas (verdadeiras e falsas). A métrica de precisão (P) está descrita em 6.

$$P = \frac{VP}{VP + FP} \quad (6)$$

em que:

- VP: número de pontos de deriva corretamente detectados (verdadeiros positivos);
- FP: número de pontos de deriva falsos positivos, ou derivas detectadas erroneamente.

Tabela 4.1: Contagem de ocorrências de deriva para cada detector durante o experimento 1

Seção do sinal	ADWIN	KSWIN	Page-Hinkley
1	2	15	2
2	2	9	2
3	2	23	2
4	2	28	2
5	2	19	2
6	2	20	2
7	2	19	2
8	2	20	2
9	2	24	2
10	2	22	2

Tabela 4.2: Precisão de cada detector para variação da média (Experimento 1)

	ADWIN	KSWIN	Page-Hinkley
FP	1	184	1
FN	0	0	0
Detecções	20	203	20
Precisão (%)	95	9,4	95

Essa análise foi realizada individualmente para cada experimento, cujos resultados estão detalhados na Seção 4 e discutidos com base nas evidências apresentadas.

4 Resultados e discussão

Os resultados estão organizados por experimento, conforme relacionado nas Subseções 3.1.1 a 3.1.3.

4.1 Experimento 1 - Média variável

O primeiro experimento teve a média μ incrementada com passo de 0,05 a cada uma das dez seções do sinal. Nesse caso, o desvio-padrão σ foi fixado em 0,02. A Tabela 4.1 relaciona a contagem de pontos de deriva detectados por cada método, no caso do primeiro experimento. Por sua vez, a Tabela 4.2 relaciona a precisão de cada detector durante o presente experimento (até a amostra final), conforme explicado na Subseção 3.2.4.

Os resultados indicam que ADWIN e Page-Hinkley detectaram o número exato de transições esperadas, enquanto KSWIN apresentou excesso de detecções. Esse comportamento sugere que KSWIN é excessivamente sensível, resultando em um número elevado de falsos positivos.

ADWIN e Page-Hinkley demonstraram desempenho ideal ao detectar exclusivamente as transições previstas, garantindo maior precisão e confiabilidade no cenário analisado. Já KSWIN, apesar de eficiente em detectar mudanças, apresentou limitações significativas para este experimento devido à sensibilidade elevada a flutuações menores.

Tabela 4.3: Contagem de ocorrências de deriva para cada detector durante o experimento 2

Seção do sinal	ADWIN	KSWIN	Page-Hinkley
1	1	23	1
2	0	18	0
3	0	23	0
4	0	18	0
5	0	26	0
6	0	20	0
7	0	29	0
8	0	17	0
9	0	19	0
10	0	23	0

Tabela 4.4: Precisão de cada detector para variação do desvio-padrão (Experimento 2)

	ADWIN	KSWIN	Page-Hinkley
FP	1	225	1
FN	0	0	0
Detecções	1	244	1
Precisão (%)	0	7,8	0

4.2 Experimento 2 - Desvio-padrão variável

No segundo experimento, o desvio padrão (σ) foi variado progressivamente, enquanto a média (μ) foi mantida constante. Este cenário buscou avaliar a eficácia dos detectores em identificar mudanças na dispersão do sinal. Os resultados estão compilados nas Tabelas 4.3 e 4.4.

Os detectores ADWIN e Page-Hinkley apresentaram baixo desempenho, com um número de detecções inferior ao esperado. Esse comportamento sugere insensibilidade às alterações isoladas no desvio padrão. Por outro lado, KSWIN apresentou novamente um número excessivo de detecções, evidenciando sua alta sensibilidade a variações menores, o que compromete sua aplicabilidade em situações que demandam precisão.

Nenhum dos detectores conseguiu resultados ideais neste experimento. ADWIN e Page-Hinkley foram pouco responsivos, enquanto KSWIN gerou um número elevado de falsos positivos, destacando a necessidade de ajustes nos parâmetros ou o uso de métodos alternativos para esse tipo de variação.

4.3 Experimento 3 - Média e desvio variáveis

Por fim, o terceiro experimento combinou alterações simultâneas na média (μ) e no desvio padrão (σ). Este cenário representa uma situação mais complexa para os detectores por se tratar de uma análise bivariada. No entanto, era esperado que essa variação dupla gerasse uma maior precisão. Os resultados do terceiro experimento estão compilados nas Tabelas 4.5 e 4.6.

Os detectores ADWIN e Page-Hinkley foram capazes de identificar com precisão as transições esperadas. O KSWIN, por outro lado, manteve seu padrão de alta sensibilidade, detectando um número excessivo de pontos de deriva e comprometendo sua precisão.

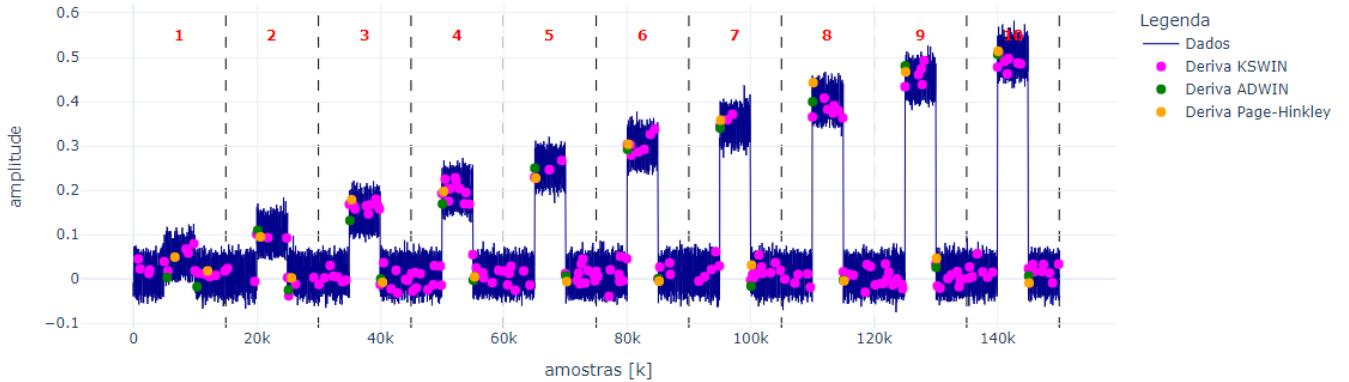


Figura 4: Série de sinais concatenados com alteração na média μ

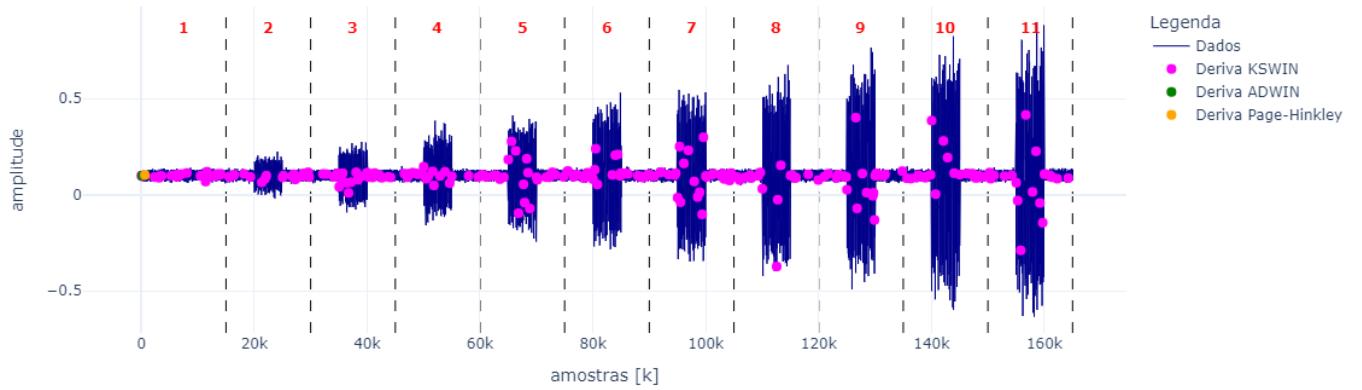


Figura 5: Série de sinais concatenados com alteração no desvio-padrão σ

Tabela 4.5: Contagem de ocorrências de deriva para cada detector durante o experimento 3

Seção do sinal	ADWIN	KSWIN	Page-Hinkley
1	1	19	1
2	2	23	2
3	2	17	2
4	2	24	2
5	2	26	2
6	2	17	2
7	2	21	2
8	2	23	2
9	2	22	2
10	2	23	2

Tabela 4.6: Precisão de cada detector para variação da média e do desvio-padrão (Experimento 3)

	ADWIN	KSWIN	Page-Hinkley
FP	0	196	0
FN	0	0	0
Detecções	19	215	19
Precisão (%)	100	8,8	100

5 Conclusão e escopo futuro

Este trabalho avaliou o desempenho de três detectores de deriva — ADWIN, KSWIN e Page-Hinkley — frente a alterações controladas na média e no desvio-padrão de sinais gaussianos sintéticos. Os resultados mostraram que os detectores ADWIN e Page-Hinkley apresentaram maior robustez e precisão na detecção de deriva, alinhando-se ao comportamento esperado em todos os experimentos. KSWIN, embora eficiente em identificar mudanças, demonstrou sensibilidade excessiva para o experimento proposto, resultando em falsos positivos em grande quantidade, o que compromete sua confiabilidade em cenários que demandam precisão.

Os experimentos evidenciam que a escolha de um detector depende diretamente da natureza das alterações nos dados. Em cenários

Nesse contexto, ADWIN e Page-Hinkley provaram ser opções confiáveis para aplicações práticas envolvendo mudanças simultâneas nos parâmetros do sinal. A alta sensibilidade de KSWIN, embora útil em certos casos, torna-o inadequado para cenários que exigem detecções com maior controle e menos ruído.

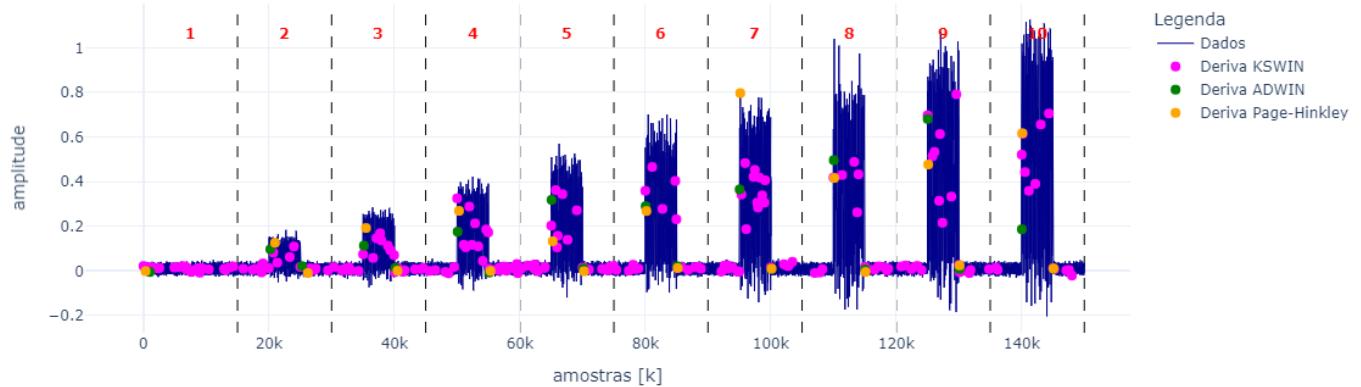


Figura 6: Série de sinais concatenados com alteração tanto na média μ quanto no desvio-padrão σ

com mudanças graduais e discretas, ADWIN e Page-Hinkley são opções mais adequadas. Por outro lado, KSWIN pode ser útil em aplicações que toleram maior sensibilidade e têm menor impacto dos falsos positivos.

A abordagem proposta contribui para a compreensão do comportamento dos detectores de deriva em diferentes cenários controlados, fornecendo uma base sólida para o desenvolvimento de sistemas de previsão robustos e resilientes a mudanças nos padrões de dados.

5.1 Escopo futuro

A seguir estão relacionados possíveis desdobramentos do presente trabalho como escopo futuro.

- Explorar a velocidade de detecção em termos do número de amostras necessárias para identificar eventos de deriva, tanto em média quanto em desvio-padrão.
- Testar os detectores em sinais reais, como medições de sensores industriais, para validar os resultados obtidos com dados sintéticos.
- Incorporar técnicas de ajuste dinâmico de parâmetros nos detectores, visando aumentar a precisão em cenários com características não previamente conhecidas e/ou não-lineares.
- Desenvolver abordagens híbridas que combinem diferentes detectores para melhorar a previsão em cenários complexos.
- Avaliar o custo computacional de cada detector, o tempo de processamento de cada algoritmo e o consumo energético de cada técnica, visto que pode se tratar de um cenário de computação de borda ou “edge computing”. Nesse tipo de aplicação, sobretudo os que utilizam baterias, otimizações energéticas são valiosas.

Acknowledgements

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), ao MackPesquisa e à Atech, empresa do Grupo Embraer, pelo suporte financeiro por meio de bolsa de pesquisa.

References

- [1] Rajesh Arora et al. 2021. Prediction and forecasting of COVID-19 outbreak using regression and ARIMA models. *Journal of Interdisciplinary Mathematics* 24 (2021), 227–243.
- [2] Reuben E. Bawack et al. 2022. Artificial intelligence in e-commerce: A bibliometric study and literature review. *Electronic Markets* 32 (2022), 1–42.
- [3] Albert Bifet and Ricard Gavaldà. 2007. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*. 443–448.
- [4] Ken Binmore. 1994. Game theory and business ethics. *Business Ethics Quarterly* 4 (1994), 125–143.
- [5] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley, Hoboken, USA.
- [6] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Applied Mathematics* 113 (2016), 3932–3937.
- [7] Kalyan Chatterjee and William Samuelson (Eds.). 2014. *Game Theory and Business Applications* (2 ed.). Springer, New York, USA.
- [8] Laura Bragante Corssac and Juliano Araújo Wickboldt. 2022. A digital twin-based smart home: A proof of concept study. *arXiv preprint arXiv:2212.14238* (2022).
- [9] João Gama et al. 2014. A survey on concept drift. *ACM Computing Surveys (CSUR)* 46 (2014), 1–37.
- [10] João Gama et al. 2014. A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)* 46 (2014), 1–37.
- [11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press, Cambridge, USA.
- [12] Xiao Guo et al. 2019. Review on the application of artificial intelligence in smart homes. *Smart Cities* 2 (2019), 402–420.
- [13] Stefan Hajkowicz et al. 2023. Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960-2021. *Technology in Society* 74 (2023), 102260.
- [14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9 (1997), 1735–1780.
- [15] Michael C. Horowitz et al. 2024. Adopting AI: how familiarity breeds both trust and contempt. *AI & Society* 39 (2024), 1721–1735.
- [16] Michael I. Jordan and Tom M. Mitchell. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349 (2015), 255–260.
- [17] Lionel P. Joseph et al. 2024. Short-term wind speed forecasting using a hybrid model. *Energy* 119 (2024), 122624.
- [18] Guolin Ke et al. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.
- [19] Andrej Kolmogorov. 1933. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4 (1933), 83–91.
- [20] Vaios I. Kontopoulos et al. 2023. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet* 15 (2023), 255.
- [21] Konstantina Kourou et al. 2015. Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal* 13 (2015), 8–17.
- [22] Bartosz Krawczyk et al. 2017. Ensemble learning for data stream analysis: A survey. *Information Fusion* 37 (2017), 132–156.
- [23] Jay Lee et al. 2018. Industrial artificial intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters* 18 (2018), 20–23.
- [24] Tae Kyun Lee et al. 2019. Global stock market investment strategies based on financial network indicators using machine learning techniques. *Expert Systems with Applications* 117 (2019), 228–242.

- [25] Olof Leimar and John M. McNamara. 2023. Game theory in biology: 50 years and onwards. *Philosophical Transactions of the Royal Society B: Biological Sciences* 378 (2023), 20210509.
- [26] Jing Li et al. 2019. A machine learning based method for customer behavior prediction. *Tehnički vjesnik* 26 (2019), 1670–1676.
- [27] Tianyang Lin, Yuxin Wang, and Xiangyang Liu. 2022. A survey of transformers. *AI Open* 3 (2022), 111–132.
- [28] Fábio Silva Lopes, Leandro Augusto da Silva, and Vivaldo José Breternitz. 2018. Research and Education in Data Science: Challenges for the Area of Information Systems. In *Grandes Desafios da Pesquisa em Sistemas de Informação no Brasil 2016–2026*, Henrique Freitas, Rita Suzana Pitanguera Maciel, and Davi Viana (Eds.). SBC—Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil.
- [29] Jing Lu et al. 2018. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering* 31 (2018), 2346–2363.
- [30] M.J. Mlaenik et al. 2004. State-of-the-art of the windowing technique. *Journal of Canadian Petroleum Technology* 43 (2004).
- [31] Jacob Montiel et al. 2018. Scikit-Multiflow: A Multi-output Streaming Framework. *Journal of Machine Learning Research* 19 (2018), 1–5. <http://jmlr.org/papers/v19/18-251.html> Accessed: 2024-08-12.
- [32] José Moura and David Hutchison. 2019. Game theory for multi-access edge computing: Survey, use cases, and future trends. *IEEE Communications Surveys & Tutorials* 21 (2019), 255–312.
- [33] Pham Van Nam et al. 2024. Using Artificial Intelligence (AI) for Monitoring and Diagnosing Electric Motor Faults Based on Vibration Signals. In *2024 International Conference on Information Networking (ICOIN)*. 1–6.
- [34] Nima Norouzi, Maryam Fani, and Atefeh Behzadi Forough. 2022. Green tax as a path to greener economy: A game theory approach on energy and final goods in Iran. *Renewable and Sustainable Energy Reviews* 156 (2022), 111968.
- [35] E. S. Page. 1954. Continuous inspection schemes. *Biometrika* 41 (1954), 100–115.
- [36] Roberto Pereira and Maria Cecília Calani Baranaukas. 2018. Systemic and Socially Aware Perspective for Information Systems. In *Grandes Desafios da Pesquisa em Sistemas de Informação no Brasil 2016–2026*, Henrique Freitas, Rita Suzana Pitanguera Maciel, and Davi Viana (Eds.). SBC—Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil.
- [37] Agbessi Akué P. et al. 2023. Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU approaches. *Energies* 16 (2023), 4739.
- [38] Frédy Pokou, Jules Sadefo Kamdem, and François Benhmad. 2024. Hybridization of ARIMA with learning models for forecasting of stock market time series. *Computational Economics* 63 (2024), 1349–1399.
- [39] Jose Prieto-Gutierrez, Francisco Segado-Boj, and Fabiana da Silva França. 2023. Artificial intelligence in social science: A study based on bibliometrics analysis. *Human Technology* 19 (2023), 149–162.
- [40] André C. Rocha and Luiz H. A. Monteiro. 2023. On the spread of charitable behavior in a social network: A model based on game theory. *Networks and Heterogeneous Media* 18 (2023), 842–854.
- [41] Stephen Schecter and Herbert Gintis. 2016. *Game Theory in Action: An Introduction to Classical and Evolutionary Models*. Princeton University Press, Princeton, USA.
- [42] Weisong Shi et al. 2016. Edge computing: Vision and challenges. *IEEE Internet of Things Journal* 3 (2016), 637–646.
- [43] Sima Siami-Namini, Nasim Siami Tavakoli, and Akbar Siami Namin. 2019. A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 1–7.
- [44] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press, Cambridge, USA.
- [45] Andreas Svoboda. 2023. The impact of artificial intelligence on the banking industry. *Journal of Banking and Finance Management* 4 (2023), 7–13.
- [46] James W. Taylor. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society* 54 (2003), 799–805.
- [47] Kifayat Ullah et al. 2024. Application of game theory in modern electrical power system (a review). *AIP Advances* 14 (2024), 010701.
- [48] Ashish Vaswani et al. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, Vol. 30. 5998–6008.
- [49] Melvin M. Vopson. 2020. The information catastrophe. *AIP Advances* 10 (2020), 085014.
- [50] Keith Weigelt and Colin Camerer. 1988. Reputation and corporate strategy: A review of recent theory and applications. *Strategic Management Journal* 9 (1988), 443–454.
- [51] Gerhard Widmer and Miroslav Kubat. 1996. Learning in the presence of concept drift and hidden contexts. *Machine Learning* 23 (1996), 69–101.
- [52] Janith R. Wijesingha et al. 2021. Smart Residential Energy Management System (REMS) Using Machine Learning. In *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. IEEE, 1–6.
- [53] Thorsten Wuest et al. 2016. Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research* 4 (2016), 23–45.
- [54] Yanbo Xu et al. 2018. RAIM: Recurrent Attentive and Intensive Model of Multi-modal Patient Monitoring Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2565–2573.
- [55] G. Peter Zhang. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50 (2003), 159–175.
- [56] Qi Zhang, Yao Li, Jing Tan, Lei Lei, and Xi Xiao. 2023. Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 9 (2023), 10901–10909. <https://doi.org/10.1609/aaai.v37i9.26317>
- [57] Zhi-Hua Zhou. 2012. *Ensemble Methods in Machine Learning*. Springer, Boca Raton, USA. 317–336 pages.
- [58] Cui-Li Zong and Lei Wang. 2018. Prediction of urban residents' travel rate in China based on ARIMA models. *Journal of Interdisciplinary Mathematics* 50 (2018), 159–175.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009