

Classificação de falhas em processo industrial de mineração a partir de uma representação *fuzzy* de séries temporais

Gabriel Vinícios M. Fernandes*. Agnaldo Rocha Reis**. Frederico Gadelha Guimarães***

*Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto e Instituto Tecnológico Vale e Vale AS, Ouro preto, MG, Brasil Vale S.A – Brasil (e-mail: gabrielvinicios02@gmail.com).

**Depto. de Engenharia de Controle e Automação da Escola de Minas da Universidade Federal de Ouro Preto, MG, (e-mail: reis@ufop.edu.br).

***Machine Intelligence and Data Science (MINDS) Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, (e-mail: fredericoguimaraes@ufmg.br)

Abstract: Brazil is one of the most important producers of iron ore in the world and with the potential to grow over the years. The mining sector demands the adoption of new technologies for the development of specialized systems that increase the efficiency in the production process. Mining must embrace innovation for automation of repetitive tasks, systems integration, continuous process improvement, disaster risk reduction and adaptation to the global context. In this sense, the article proposes the use of *Fuzzy Time Series* (FTS) to learn a representation of the data that is more effective for the classification stage, using data originating from Plant A of the Ferro S11D Project, from Vale S.A., located in Canaã dos Carajás, Pará, Northern region of Brazil. The result demonstrated a significant improvement in failure prediction from the addition of PWFTS (Probabilistic Weighted *Fuzzy Time Series*) techniques to the XGBoost (*Extreme Gradient Boosting*) algorithm. From the proposed methodology, there was an increase in accuracy from 79,3% to 98,9%.

Resumo: O Brasil é um dos mais importantes produtores de minérios do mundo e com potencial para crescer ao longo dos anos. O setor de mineração demanda a adoção constante de tecnologia para o desenvolvimento de sistemas especializados que permitam maior eficiência do processo produtivo. A mineração deve abraçar a inovação para automação de tarefas repetitivas, integração de sistemas, melhoria contínua de processos, redução de riscos de desastres e adaptação ao contexto global. Nesse sentido, o artigo propõe o uso de *Fuzzy Time Series* (FTS) para aprender uma representação dos dados que seja mais efetiva para a etapa de classificação e detecção de falhas, utilizando dados originários da Usina A do Projeto Ferro S11D, da Vale S.A., localizada em Canaã dos Carajás, Pará, região Norte do Brasil. O resultado demonstrou uma melhoria significativa na classificação, detecção e previsão de falhas a partir da adição de técnicas da PWFTS (*Probabilistic Weighted Fuzzy Time Series*) ao algoritmo XGBoost (*Extreme Gradient Boosting*). A partir da metodologia proposta, foi um aumento na acurácia de 79,3% para 98,9%.

Keywords: *Fuzzy Time Series*, *machine learning*, mining, fault detection.

Palavras-chave: *Fuzzy Time Series*, aprendizado de máquina, mineração, detecção de falhas.

1. INTRODUÇÃO.

O Brasil é um dos mais importantes produtores de minérios do mundo e com potencial para crescer ao longo dos anos (Humphreys, 2015), ocupa a 2ª posição dentre os países de maior produção de minério de ferro do mundo. O setor de mineração demanda a adoção de novas tecnologias para o desenvolvimento de sistemas especializados que permitam uma melhor eficiência na produção. Nesse sentido, o mercado de mineração deve abraçar a inovação para automação de tarefas repetitivas, integração de sistemas, melhoria contínua de processos, redução de acidentes e de desastres. A adoção da automação inteligente permite a redução de riscos às pessoas

inseridas no processo, retirando o contato direto das pessoas com o processo industrial (Qi, 2020).

A partir de dados históricos de um sistema, mesmo que faltem dados e/ou leituras incorretas, é possível aplicar técnicas de Inteligência Artificial (IA) para análises e, assim, extrair informações dos dados. Vale pontuar que a aplicação de um sistema de classificação baseado em dados é considerada um sistema inteligente. (Satyanarayanan et al., 2009).

Sistemas inteligentes ou sistemas baseados em inteligência computacional surgiram para contornar dificuldades encontradas pelos métodos tradicionais na modelagem e entendimento de sistemas complexos. A detecção de falhas

(Ammiche, 2018) se encontra também nesse âmbito, pois em sistemas de grande porte, com muitos sensores e grande quantidade de dados constantemente gerados, o processo de diagnóstico humano torna-se impreciso. Para reduzir o tempo e melhorar os diagnósticos faz-se necessário a análise e a coleta de dados sobre o estado do sistema utilizando medições, testes e sintomas de mau funcionamento, além da migração para um processo automático (Zou; Xia; Li, 2018).

A detecção de falhas (Bathelt, 2015) a partir de um sistema inteligente tornou-se um campo importante nas últimas décadas, após a adoção de práticas da Indústria 4.0 (I4.0) ou Quarta Revolução Industrial. Estão inseridas neste contexto a automação industrial, tecnologia de informação, inteligência de dados até as principais inovações de cada um desses campos (Schwab, 2017). A Quarta Revolução Industrial traz um aumento de eficiência na produção em larga escala e otimiza a utilização dos recursos necessários, integrando os sistemas de forma cooperativa. Sendo assim, a demanda de interconexão de milhares de sensores e atuadores está em constante crescimento, em um mesmo sentido a quantidade de dados gerados cresce exponencialmente na mesma proporção. Esta gama de dados transforma o processo de obtenção de informações a partir dos dados como uma iniciativa para a melhoria da produção, procedimentos industriais e um desafio na área de inteligência computacional.

Nesse contexto, este artigo tem como objetivo a adição de técnicas de *Fuzzy Time Series* (FTS). Estes métodos de previsão de séries temporais, para uma representação da série dos dados que seja mais efetiva na classificação de falhas no sensor de nível do silo 6 (TAG: SI-2012 KS-06) (Vegapuls, 2021). O mesmo está fisicamente localizado e instalado na britagem secundária da Usina A na Planta S11D (Figura 1).

Os dados utilizados na obtenção e análise da solução proposta foram extraídos do processo representado na (Figura 1). O conjunto dos dados são a leitura de nível do silo 06 de minério (SI-2012KS-06), velocidade do alimentador (AL-2012KS-06), taxa de alimentação (ton/h) e densidade de carga linear (Kg/m) da balança (WIT-1085KS-2300). Para as variáveis

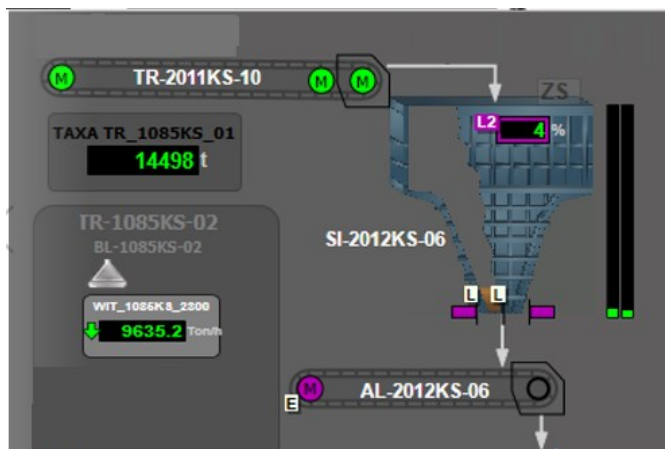


Figura 1– Processo analisado

apresentadas acima, todas foram amostradas com um intervalo de amostragem de 3 segundos (s).

Para o desenvolvimento do sistema inteligente serão utilizados modelos de lógica *Fuzzy* (Tak, 2020) e de árvores de decisão. Os dados coletados da respectiva planta para modelagem e validação do processo de classificação de falhas nos sensores do peneiramento. A condição considerada como falha se refere a 3 estados não esperado do sensor: Congelamento, variação anômala e leitura de fundo de escala do sensor. A adoção desta prática visa a redução de paradas não programadas, perdas produtivas e/ou desgastes prematuros de equipamentos, além de possibilitar o aumento do nível de segurança dos funcionários que atuam nessa área.

Este trabalho propõe a utilização de métodos estatísticos para análises, uma vez que esta prática tem se mostra eficiente no processo de detecção de falhas (Alyousifi, Y. et al, 2020) e é visível o crescimento nos últimos tempos. Nesse sentido, torna-se visível o surgimento de novas técnicas para modelagem e análise de séries temporais de acordo com Alyousifi, Y. et al (2020).

2. FUNDAMENTAÇÃO TEÓRICA.

A abordagem utilizada para análise envolve o uso do método XGBoost (Hattori, 2018) e uma composição do mesmo método adicionado com o método heurístico *Probabilistic Weighted Fuzzy Time Series* (PWFTS) (Silva, 2019) aplicados à classificação e detecção de falhas do sensor de nível (Figura 1). Com os dados reais oriundos do processo, foi gerado um modelo obtido a partir do algoritmo da PWFTS para cada variável em questão (endógenas e exógenas) (Silva, 2019). As variáveis a consideradas neste trabalho são:

- LIT-2012KS-06 - Indicação de nível do silo 06 a ser avaliado (Endógena).
- AL-2012KS-06-SpeedActual - Indicação de velocidade do alimentador (Exógena).
- WIT-1085KS-2300- Taxa da balança WIT-1085KS-2300 (Exógena).
- WIT-1085KS-2300- Densidade linear de carga (Exógena).

Para cada variável foi obtido o modelo descritivo das regras ponderadas da PWFTS. Após a obtenção das matrizes a priori e a posteriori de transição de Markov (Stülp, 2009) para cada variável considerada na análise, as matrizes foram vetorializadas e processadas pelo classificador.

2.1 XGBoost.

O XGBoost (extreme gradient boosting) é um algoritmo de Machine Learning supervisionado que utiliza como base árvores de decisão (Chen e Guestrin, 2016). O algoritmo do XGBoost implementa um processo denominado boosting, para a geração de modelos mais precisos. Boosting refere-se à técnica de aprendizado para construir modelos sequencialmente, em que o modelo que sucede, tenta corrigir as deficiências do anterior. À medida que a árvore aumenta, cada novo modelo adicionado ao conjunto é uma nova árvore de decisão. Este modelo treina as árvores de decisão impulsionado por um gradiente (Sahin, 2020), o que traz alta escalabilidade para o algoritmo.

O algoritmo fornece reforço para as árvores sucessoras que vão sendo criadas durante a sua execução, sendo este o gradiente proposto pelo algoritmo. A adoção deste algoritmo no trabalho proposto é devido ao fato deste algoritmo ser um benchmark entre os algoritmos de classificação. Convém sinalizar que o modelo tem sido eficaz na redução do tempo de processamento do algoritmo e pode ser usado tanto para tarefas de regressão quanto de classificação (Qiu, Y, 2021).

2.2 PWFTS-Probabilistic Weighted Fuzzy Time Series.

A partir de uma série temporal numérica Y e seus valores individuais $y(t)$, uma previsão de valores futuros para esta série normalmente são valores pontuais ou escalares (Luo, 2019). Este tipo de previsão geralmente é o melhor “palpite” para Y no tempo de $t+1$, considerando o que se sabe de Y e seus valores passados. Assim, temos para o valor previsto de Y , uma incerteza associada a essa previsão. Existem formas de lidar com essa limitação das previsões pontuais, e são as previsões probabilísticas, descritas pelo modelo da PWFTS (Silva, 2019).

Também de acordo com (Silva, 2019), uma série temporal pode ser descrita pela Equação 1 e representada graficamente de acordo com a Figura 2. Representado no eixo y estão os conjuntos nebulosos (A_0 até A_5) que compõem essa série temporal.

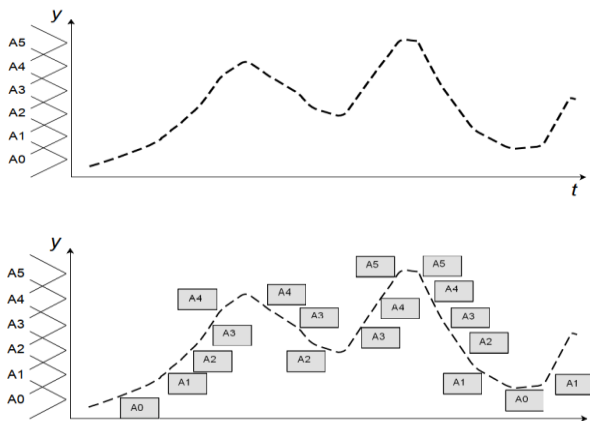


Figura 2 – Série temporal e seu particionamento em conjuntos fuzzy (Silva, 2019)

A PWFTS é um método heurístico e invariante no tempo que pode fazer previsões pontuais, intervalares e probabilísticas para um ou muitos passos à frente (Silva, 2019). O método é composto por um conjunto de regras de transição entre os conjuntos nebulosos e os pesos delas (Equação 1 e 2). O método incorpora todos os hiperparâmetros explorados em uma série temporal fuzzy normal.

$$\begin{cases} \pi_1 A_1 \rightarrow \omega_{11} A_1, \dots, \omega_{1k} A_k \\ \vdots \\ \pi_k A_k \rightarrow \omega_{k1} A_1, \dots, \omega_{kk} A_k \end{cases}, \quad (1)$$

$$\begin{bmatrix} \pi_1 \\ \vdots \\ \pi_n \end{bmatrix} x \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}^t = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} x \begin{bmatrix} \omega_{11} & \dots & \omega_{1k} \\ \vdots & \ddots & \vdots \\ \omega_{k1} & \dots & \omega_{kk} \end{bmatrix} \quad (2)$$

$$[LHS] x \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}^t = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} x [RHS] \quad (3)$$

[LHS] e [RHS] são matrizes compostas por pesos (Yu, 2015) e probabilidades de transição entre os conjuntos nebulosos nas regras do modelo da PWFTS (1)(2), respectivamente. Estas regras descrevem o mais provável comportamento no futuro (RHS-Lado direito da regra) a partir de algum comportamento passado (LHS-Lado esquerdo da regra).

Estas matrizes de pesos foram utilizadas neste trabalho e foram obtidas a partir do processo de fuzzificação da PWFTS descrito em (Silva, 2019) para cada variável considerada da Figura 1.

3. METODOLOGIA.

O controle de processos industriais modernos é baseado em dados e na inteligência que se pode extrair dos mesmos. Não sendo possível controlar o que não se pode medir. Muito menos melhorar ou otimizar um processo sem os seus dados históricos (Esmail Jahanshahi, 2015).

Foram desenvolvidos modelos baseados em dados reais do processo, buscando a detecção de falhas dos sensores de nível processo industrial descrito na Figura 1 (Benner, 1985). Para o processamento dos dados, foi utilizada a linguagem Python em conjunto com a biblioteca PyFTS (Silva, 2018), para a construção dos modelos. A plataforma, utilizada foi o Google Colab Pro (Google, 2021).

Foi utilizado a princípio o algoritmo do XGBoost (Sahin, 2020) para realizar esta classificação, aplicação considerando 8 “lags”, como descrito na seção 3.1 (Ismail, 2011). Foi avaliado o desempenho de taxas de acerto e falsos positivos, com os dados reais provenientes da britagem secundária da Usina A da mina do S11D-Vale.

O modelo de comparação proposto neste trabalho foi desenvolvido em duas etapas: uma de modelagem e outra de classificação. Na etapa de modelagem, foram utilizadas janelas de 100 amostras de dados espaçadas igualmente ao longo do tempo e foi obtido o modelo treinado (80%) e testado (20%) na PWFTS. Após a obtenção do primeiro modelo, a janela de 100 dados foi deslizada até o final da série temporal, e para cada uma dessas janelas novas, um novo modelo treinado e testado de PWFTS foi obtido, com seus hiperparâmetros.



Figura 3 – Sequência de processamento dos dados com melhoria proposta

Para a modelagem, foram utilizados dados tanto de falhas quanto de operação normal do nível dos silos da britagem secundária, sendo formadas duas bases. Para os dois casos, foram utilizados 80% dos dados para treinamento e 20% para teste. Realizados os tratamentos, os pesos das regras dos modelos foram colocados em um classificador XGBoost para definir a existência ou não da falha e em qual situação melhor se enquadra. Ao final deste trabalho, será apresentado um comparativo com o XGBoost clássico, composto de 8 “lags” de atraso e a metodologia proposta descrita na seção 3, onde foi possível de forma comparativa, demonstrar os ganhos obtidos no processo de classificação.

3.1 Universo dos dados.

No universo de discurso proposto, foi possível classificar tanto situações em que temos falha dos sensores de nível dos silos, quanto situações de um cenário de operação normal dos sensores.

Para este objetivo, foram coletados dados tanto de operação normal quanto de operação em que o sensor apresenta falha. Foi extraída a mesma quantidade de dados para as duas situações em que temos falhas e o comportamento normal. Para se ter uma maior abrangência do processo, foram coletadas a variável endógena e 3 variáveis exógenas. São elas:

- Variável endógena: Nível do silo (LIT_2012KS_0251);
- Variáveis exógenas: Velocidade do alimentador (AL_2012KS_06M1_SpeedActual), taxa do alimentador em Ton/h (WIT_1085KS_2300_Rate) e carregamento do transportador móvel em Kg/m (WIT_1085KS_2300_BeltLoad).

Comportamento das variáveis observadas onde não se encontram falhas (dados de operação normal):

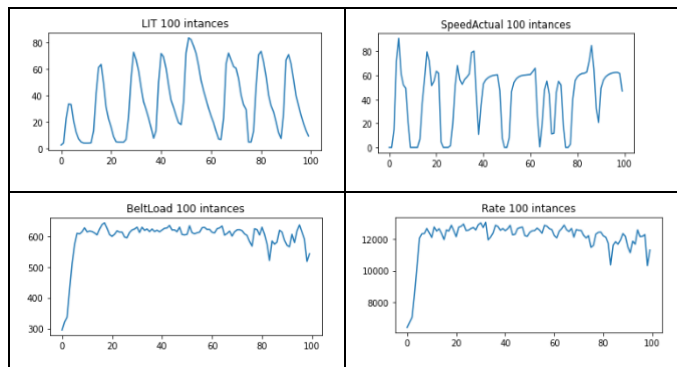


Figura 4 – Visualização dos dados de Operação Normal.

Para este conjunto de dados, foi também feita a análise de autocorrelação (ACF) e autocorrelação parcial (PACF), conforme as figuras 5 e 6, que auxilia na tomada de decisão do número de parâmetros autorregressivos e do número de parâmetros de médias móveis, foi obtido:

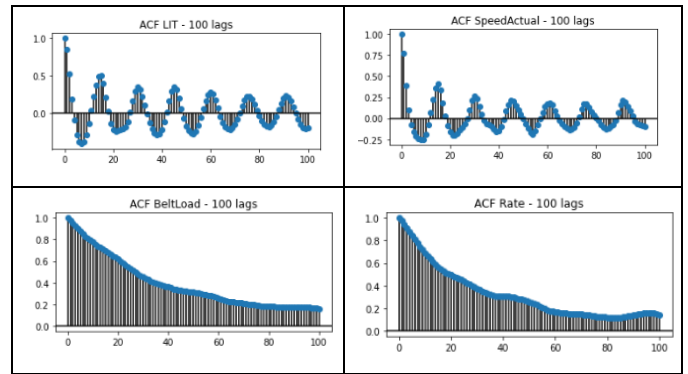


Figura 5 – Análise de Autocorrelação dos dados de operação normal.

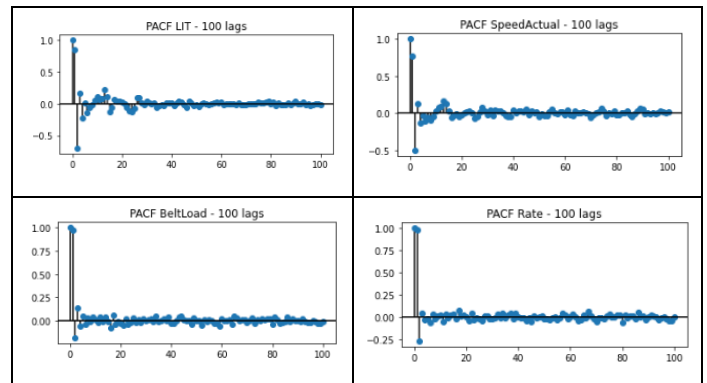


Figura 6 – Autocorrelação parcial dos dados de operação normal.

Comportamento das variáveis onde se observa falhas:

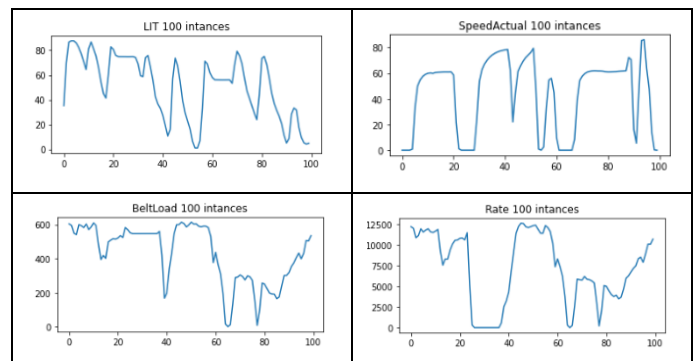


Figure 7-Visualização dos dados com falhas de operação.

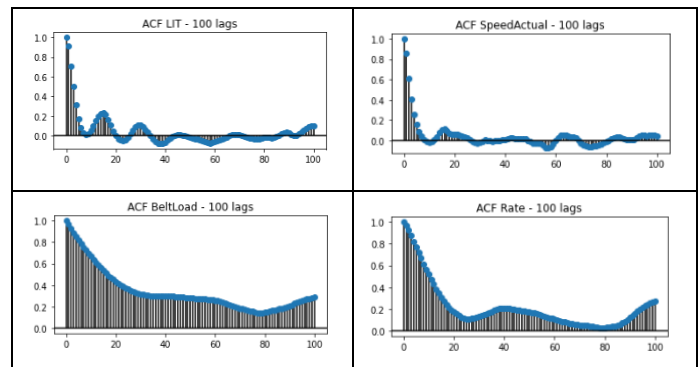


Figura 8 – Autocorrelação dos dados com falhas.

Para o conjunto de dados com falhas, foi também feita a análise de autocorrelação e autocorrelação parcial, conforme as Figuras 8 e 9.

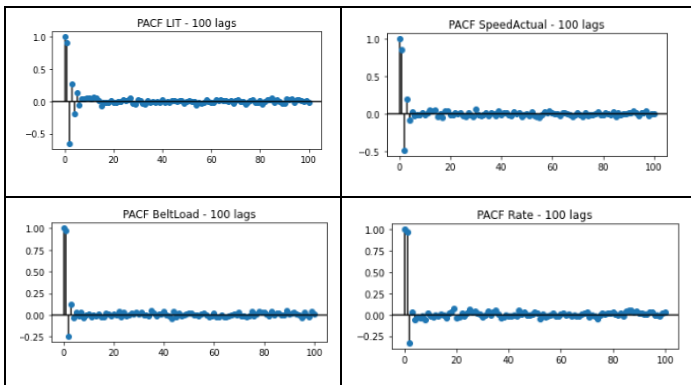


Figure 9 – Autocorrelação parcial dos dados com falhas.

Com base nas análises de ACF e PACF (Denig, 2016) apresentadas para as 4 variáveis consideradas nesta análise, vemos que a utilização de 4 ‘lags’ seria suficiente para realizar, sem perdas significativas (Demir, 2020) para processos de classificação e previsão das variáveis consideradas. Como margem de confiança, vamos utilizar um total de 8 ‘lags’ (100% a mais) nas análises.

Em adição, foi feita uma análise de correlação cruzada utilizando coeficiente de Pearson para verificar o quanto estes dados estão relacionados e foi identificado uma baixa correlação entre as variáveis consideradas na avaliação dos modelos (Capaci, 2019). Ou seja, ambas as variáveis irão contribuir na obtenção dos modelos, trazendo ganhos positivos aos mesmos, comportando-se como variáveis independentes entre si.

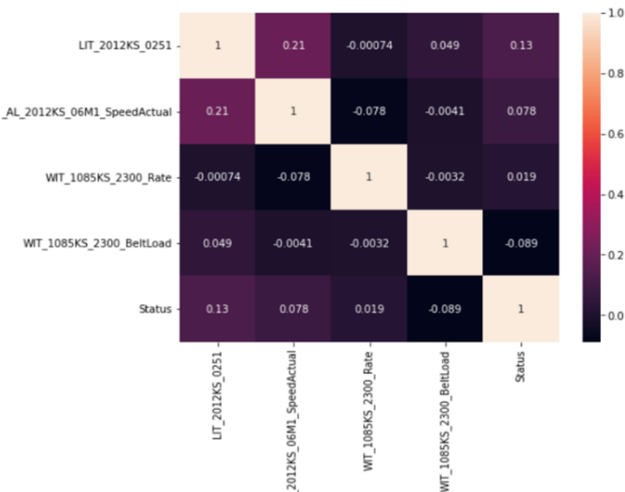


Figure 10 – Análise de correlação cruzada das variáveis endógenas e exógenas do processo em estudo.

4. RESULTADOS E DISCUSSÃO.

Na proposta para a PWFTS, em uma representação gráfica de suas probabilidades e os pesos das regras aplicadas para 10 conjuntos nebulosos (Silva, 2019). Na Figura 11, é demonstrada uma representação gráfica dos modelos obtidos pelo algoritmo da PWFTS. Obtivemos os seguintes modelos para as variáveis avaliadas (Figura 11):

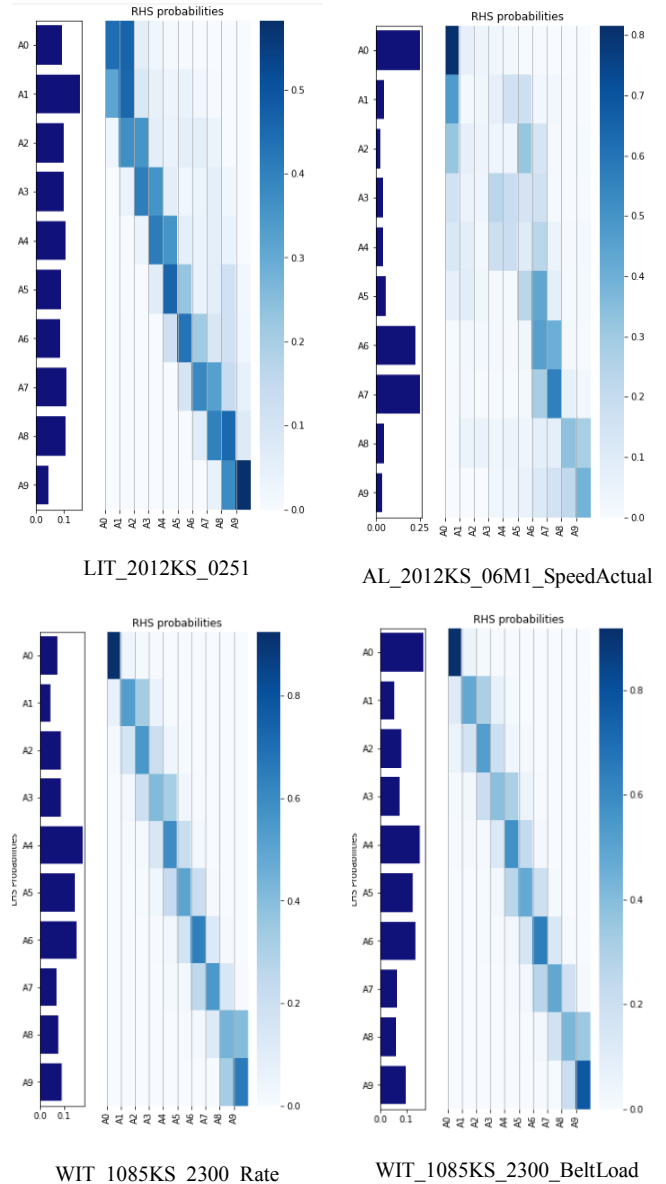


Figure 11 – Representação Gráfica das variáveis com o Algoritmo da PWFTS (Silva, 2019).

Os modelos obtidos graficamente apresentam um modelo descritivo das regras para a transição entre os conjuntos nebulosos. Estes modelos são descritos em (Silva, 2019). Para cada variável, em cada janela de dados de 100 elementos foi gerado um modelo com os pesos e probabilidades para cada conjunto nebuloso.

Para a variável endógena, o modelo obtido para as últimas 100 amostras, apresentadas no formato [[LHS][RHS]], foram:

- LIT_2012KS_0251,

0,095	0,441	0,464	0,057	0,022	0,01	0,004	0	0	0	0
0,159	0,316	0,459	0,087	0,045	0,031	0,04	0,019	0,002	0	0
0,1	0,029	0,37	0,356	0,058	0,041	0,054	0,059	0,032	0	0
0,102	0	0,04	0,408	0,358	0,059	0,022	0,035	0,054	0,022	0
0,107	0	0,003	0,034	0,413	0,357	0,055	0,039	0,055	0,04	0,002
0,09	0	0,001	0,002	0,067	0,463	0,23	0,037	0,061	0,12	0,019
0,088	0,002	0,002	0	0,001	0,122	0,43	0,213	0,089	0,12	0,022
0,109	0	0	0	0	0,001	0,099	0,385	0,329	0,14	0,045
0,108	0	0	0	0	0	0	0,068	0,401	0,456	0,075
0,043	0	0	0	0	0	0	0	0,042	0,376	0,582

Para as variáveis exógenas, os modelos obtidos para as últimas 100 amostras, apresentadas no formato [[LHS][RHS]], são:

• AL_2012KS_06M1_SpeedActual

0,249	0,815	0,07	0,042	0,034	0,025	0,009	0,004	0,001	0	0
0,042	0,478	0,021	0,034	0,069	0,167	0,174	0,029	0,024	0,004	0
0,026	0,321	0,068	0,043	0,04	0,041	0,32	0,142	0,013	0,004	0,008
0,036	0,159	0,048	0,036	0,238	0,193	0,143	0,161	0,013	0,004	0,006
0,037	0,124	0,083	0,021	0,18	0,184	0,096	0,233	0,046	0,012	0,02
0,056	0,069	0,09	0,033	0,012	0,014	0,236	0,437	0,079	0,015	0,016
0,225	0,008	0,024	0,018	0,017	0,011	0,034	0,458	0,406	0,02	0,005
0,248	0,005	0,017	0,009	0,006	0,009	0,02	0,282	0,564	0,065	0,024
0,046	0,006	0,049	0,054	0,038	0,045	0,039	0,075	0,077	0,34	0,279
0,036	0	0	0,001	0,03	0,04	0,055	0,113	0,15	0,22	0,391

• WIT_1085KS_2300_Rate

0,165	0,918	0,053	0,016	0,005	0,003	0,003	0,002	0	0	0
0,053	0,118	0,484	0,307	0,075	0,015	0,002	0	0	0	0
0,081	0,055	0,175	0,526	0,194	0,035	0,013	0,003	0,001	0	0
0,072	0,015	0,026	0,205	0,387	0,306	0,054	0,004	0,001	0	0,002
0,152	0,006	0,014	0,013	0,131	0,577	0,236	0,02	0,001	0	0,002
0,124	0,002	0,003	0,01	0,027	0,267	0,478	0,195	0,016	0,002	0
0,135	0,002	0,002	0,01	0,005	0,028	0,157	0,645	0,139	0,013	0
0,063	0	0	0	0	0,012	0,026	0,271	0,485	0,193	0,013
0,059	0,003	0,001	0	0,001	0,008	0,006	0,031	0,179	0,421	0,351
0,097	0,002	0,001	0	0	0	0,005	0,009	0,208	0,776	

• WIT_1085KS_2300_BeltLoad

0,071	0,924	0,043	0,017	0,004	0,002	0,005	0,005	0	0	0
0,044	0,068	0,532	0,329	0,06	0,01	0,001	0	0	0	0
0,086	0,018	0,169	0,555	0,205	0,042	0,009	0,002	0	0	0
0,088	0,003	0,021	0,196	0,412	0,323	0,037	0,005	0	0	0,001
0,178	0,003	0,006	0,017	0,142	0,593	0,218	0,019	0,001	0	0,001
0,146	0	0,001	0,008	0,033	0,233	0,504	0,208	0,012	0,002	0
0,153	0	0,001	0,003	0,007	0,03	0,175	0,643	0,128	0,013	0
0,069	0	0	0,001	0,001	0,016	0,024	0,256	0,551	0,145	0,006
0,075	0	0	0,002	0,001	0,002	0,005	0,024	0,122	0,438	0,407
0,091	0	0	0,002	0,002	0,002	0,001	0,001	0,008	0,327	0,658

Tabela 1. Matriz de Confusão XGBoost

	Positivo Estimado	Negativo Estimado
Verdadeiro Real	281	197
Falso Real	65	721

Tabela 2. Matriz de Confusão XGBoost + PWFTS

	Positivo Estimado	Negativo Estimado
Verdadeiro Real	47	0
Falso Real	1	43

Tabela 3. Comparação de Métricas

	XGBoost	XGBoost + PWFTS
Sensibilidade	0,588	1,0
Taxa de Falso Positivo	0,083	0,023
Taxa de Falso Negativo	0,412	0,000
Especificidade	0,917	0,977
Acurácia	0,793	0,989
Precisão	0,588	0,979
F Score	0,588	0,989

A diferença do número de amostras na Tabela 1 e Tabela 2 é devido ao processo de geração dos exemplos de treinamento e teste para cada avaliação. Na Tabela 1, foram utilizadas amostras de 8 elementos. Na tabela 2, foram utilizadas janelas de 100 elementos.

Com base nos resultados obtidos nas matrizes de confusão, foram calculados índices dos processos de classificação como sensibilidade, especificidade, acurácia e precisão (Tabela 3). Em ambos é possível ver um desempenho melhor com a metodologia proposta nesse artigo. Para a acurácia, tivemos um aumento de 37,8% para 51,6%. Nos demais casos citados, os valores apresentaram aumento, o que representa uma melhora de desempenho de classificação com o novo algoritmo adotado.

É possível observar uma redução de falsos positivos e falsos negativos em ambos os casos (Tabela 3). Para os falsos positivos, no modelo apenas com o XGBoost obtivemos uma taxa de 7,1% e com a adição do modelo da PWFTS, passamos para uma taxa de 2,3%. Em relação aos falsos negativos, a taxa para o modelo do XGBoost foi de 208,4% e com a nova metodologia proposta, o resultado obtido foi de 89,6%. Essas indicações sinalizam que a metodologia proposta apresenta um menor índice de erros ao fazer as suas classificações.

5. CONCLUSÃO.

Este trabalho apresentou a primeira etapa de implementação de um modelo de série temporal *fuzzy* para realizar previsões/detecção de falhas fazendo uso de classificações e avaliou seu desempenho com a adição de técnicas de FTS do modelo PWFTS a um classificador clássico da literatura (XGBoost). O modelo probabilístico proposto alcançou bons resultados quando comparados ao método clássico, conforme descrito anteriormente, com maior acurácia e menor índice de falsos positivos, apresentado nas matrizes de confusão. Isso demonstra o potencial das previsões probabilísticas, sendo estes competitivos com os métodos mais avançados existentes na atualidade.

Com os dados das matrizes de confusão, também foi possível a obtenção de métricas de desempenho comparativas, em que em todas, o conjunto PWFTS + XGBoost apresentou desempenho superior ao XGBoost de forma isolada.

6. TRABALHOS FUTUROS.

Como trabalhos futuros, utilizando as técnicas propostas e apresentadas neste artigo:

- Previsão das falhas aplicando metodologia similar.
- Otimização do número de conjuntos nebulosos aplicado as FTSS.
- Discriminação dos tipos de falhas.
- Melhoria no processo de ‘vetorialização’ das matrizes geradas no modelo da PWFTS.

REFERÊNCIAS.

- Alyousifi, Y., Othman, M., Sollalingam, R., Faye, I., and Silva, P. C. (2020). Predicting daily air pollution index based on fuzzy time series Markov chain model. *Symmetry*, 12(2), 293.
- Ammiche, M.; Kouadri, A.; Bakdi, A. A combined monitoring scheme with fuzzy logic filter for plant-wide Tennessee Eastman Process fault detection. *Chemical Engineering Science*. Vol.187, p.269-279, 2018. <https://doi.org/10.1016/j.ces.2018.05.001>
- Bathelt, A.; Ricker, N. L.; Jelali, M. Revision of the Tennessee Eastman Process Model. *IFAC-Papers Online*, Vol.48, Issue 8, p.309-314, 2015. ISSN 2405-8963. <https://doi.org/10.1016/j.ifacol.2015.08.199>
- Benner JR, L. Rating Accident Models and Investigation Methodologies *Journal of Safety Research*, v. 16, p. 105-126, 1985.
- Capaci, F.; Vanhatalo, E., Kulahci, M.; Bergquist, Bjarne. The Revised Tennessee Eastman Process Simulator as Testbed for SPC and DoE Methods. *Quality Engineering*, Volume 31, p. 212-229, 2019. <https://doi.org/10.1080/08982112.2018.1461905>.
- Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p 785–794. <https://doi.org/10.1145/2939672.2939785>
- Denig, M. R. D. S. (2016). Análise Hierárquica Bayesiana com Aplicação à Qualidade da Água do Rio dos Sinos
- Demir, V.; Zontul, M., Yelmen, I. Drug sales prediction with ACF and PACF Supported ARIMA Method, 2020 5th International Conference on Computer Science and Engineering (UBMK), 2020, pp. 243-247, doi: 10.1109/UBMK50275.2020.9219448.
- Lee, R. S. (2019); Chaotic interval type-2 fuzzy neuro-oscillatory network (cit2-fnon) for worldwide 129 financial products prediction. *International Journal of Fuzzy Systems*, 21(7), 2223-2244.
- Google Colab. Disponível em: <<https://colab.research.google.com/>>. Acesso em: 14 jun. 2021.
- Hattori, L.T., Benítez, C.M.V., Gutoski, M., Aquino, N.M.R., Lopes, H.S. A Novel Approach to Protein Folding Prediction based on Long Short-Term Memory Networks: A Preliminary Investigation and Analysis. In: *Proc. of IEEE World Congress on Computational Intelligence*, Rio de Janeiro, 2018. Doi: 10.1109/IJCNN.2018.8489514
- Humphreys, D. (2015). *The remaking of the mining industry* (pp. 61-87). Basingstoke: Palgrave Macmillan.
- Ismail, Z., and Efendi, R. (2011). Enrollment forecasting based on modified weight fuzzy time series. *Journal of Artificial Intelligence*, 4(1), 110-118.
- Qi, C. C. (2020). Big data management in the mining industry. *International Journal of Minerals, Metallurgy and Materials*, 27(2), 131-139.
- Qiu, Y., Zhou, J., Khandelwal, M. et al. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Engineering with Computers* (2021). <https://doi.org/10.1007/s00366-021-01393-9>
- Jahanshahi, E.; Sivalingam, S.; Schofield, J. B. Industrial test setup for autotuning of PID controllers in large-scale processes: Applied to Tennessee Eastman process. *IFAC-PapersOnLine*, Volume 48, Issue 8, p.469-476, 2015. <https://doi.org/10.1016/j.ifacol.2015.09.012>
- Luo, C., Tan, C., Wang, X. and Zheng, Y. (2019). An evolving recurrent internal type-2 intuitionistic fuzzy neural network for online learning and time series prediction applied soft computing, 78, 150-163.
- Sahin, E.K. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Appl. Sci.* 2, 1308, 2020. <https://doi.org/10.1007/s42452-020-3060-1>
- Satyanarayanan, M.; Bahl P.; Caceres R.; Davies N.; The case for VM-based cloudlets in mobile computing, *IEEE Pervasive Computing*, Oct 2009.
- Schwab, K. *The Fourth Industrial Revolution*. World Economic Forum, 2017.
- Silva, P. Scalable models for probabilistic forecasting with fuzzy time series. 2019. Tese (Doutorado em Engenharia Elétrica). Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: https://www.ppgee.ufmg.br/tese_defesas_detalhes.php?aluno=1416. Acesso em: 20 agosto 2022.
- Silva, P., Lucas, O., Sadaei H., and Guimarães F. . *pyFTS: Fuzzy Time Series for Python*, 2018. URL <https://doi.org/10.5281/zenodo.597359>.
- Stülp, V. J.; Fochezatto, A. A evolução das disparidades regionais no Rio Grande do Sul: uma aplicação de matrizes de Markov. *Nova Economia*, [S. l.], v. 14, n. 1, 2009. Disponível em: <https://revistas.face.ufmg.br/index.php/novaeconomia/article/view/424>. Acesso em: 8 maio. 2022.
- Tak, N. (2020). *Type-1 possibilistic fuzzy forecasting functions*. *Journal of Computational and Applied Mathematics*, 370, 112653.
- Vegapuls Radar. Datasheet: Vegapuls 69, medição contínua de enchimento de sólidos. Disponível em: <http://partnner.com.br/datasheets/ins/VEGA%20VegaPuls69_Datasheet.pdf>. Acesso em: 23 maio 2021.
- Yu, H. K. (2005). Weighted fuzzy time series models for taie forecasting. *Physica A: Statistical Mechanics and its Applications*, 349(3-4), 609-624.