**ANÁLISE COMPORTAMENTAL EM MULTIPLAS VARIAVES DE PROCESSO DE MÁQUINAS INDUSTRIAIS**

***BEHAVIORAL ANALYSIS IN MULTIPLE PROCESS VARIABLES OF INDUSTRIAL MACHINES***

Denilson Rogozinski

rogozinskideni@hotmail.com

Diego Ribeiro Hobold

diego.hobold@hotmail.com

Centro Universitário Facens – Sorocaba, SP, Brasil

Submetido em: 21 set. de 2024. Aceito em: 28 set. de 2024.

# RESUMO

Com o advento da indústria 4.0 e os esforços para conectar os equipamentos industriais a rede de dados das empresas, uma infinidade de dados e informações podem e devem ser extraídas de cada equipamento. Assim esses dados são coletados, tratados e analisados caso a caso e com isso podem gerar informações valiosas para a tomada de decisão do processo fabril permitindo uma redução de custo e um aumento da produtividade significativa de forma a manter as empresas competitivas no mercado. Por meio da utilização de algumas técnicas de extração, tratamento e avaliação dos dados como por exemplo análise das médias, desvio-padrão, variância, mediana, foi possível avaliar diversas variáveis coletadas do equipamento industrial de maneira rápida de forma a evidenciar o comportamento dela ao longo do tempo e com isso permitir que os engenheiros e analistas do setor produtivo possam tomar decisões assertivas de forma a corrigir e manter o processo estável.

**Palavras-chave:** Estatística. indústria 4.0. Ponto de Mudança. Série Temporal.

# ABSTRACT

With the advent of Industry 4.0 and efforts to connect industrial equipment to companies' data networks, a multitude of data and information can and should be extracted from each piece of equipment. This data is then collected, processed and analyzed on a case-by-case basis and can generate valuable information for decision-making in the manufacturing process, allowing for a reduction in costs and a significant increase in productivity to keep companies competitive in the market. By using some data extraction, processing and evaluation techniques, such as analysis of means, standard deviations, variances and medians, it was possible to quickly evaluate several variables collected from industrial equipment to demonstrate their behavior over time and thus allow engineers and analysts in the production sector to make assertive decisions in order to correct and maintain the process stable.

**Keywords:** statistics. Industry 4.0. changepoint. Timeseries.

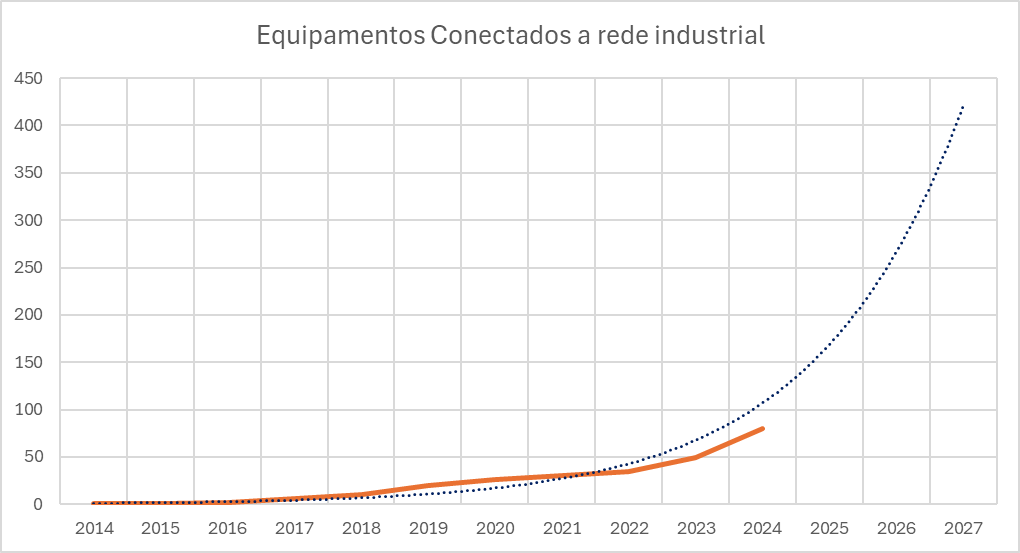
**1 INTRODUÇÃO**

Desde a revolução industrial a cerca de 300 anos atrás, a economia mundial vem se desenvolvendo de uma maneira cada vez mais veloz, buscando alavancar a produtividade das indústrias de forma a atender à crescente demanda de consumo de bens materiais. Recentemente com a chegada da 4ª revolução industrial também conhecida como Indústria 4.0, está sendo introduzido mais inteligência e integração entre máquinas e sistemas computadorizados buscando um ganho de produtividade e consequentemente redução de custos. Segundo CGEE, são exemplos de tecnologias envolvidas na indústria 4.0: Inteligência Artificial, Big Data, Internet das Coisas (IoT), Computação em nuvem, Sistema Ciberfísicos, Manufatura Aditiva, Robótica Avançada, Realidade virtual e aumentada, Sensores inteligentes etc.

Visando estar preparada e colher os frutos dessa nova revolução industrial, as empresas veem buscando cada vez mais conectar seus equipamentos industriais para obter dados de seus maquinários em tempo real e assim garantir uma produção eficiente de seus produtos. Segundo a ABDI, Agência Brasileira de Desenvolvimento Industrial, a adoção de conceitos da indústria 4.0 na matriz produtiva brasileira pode gerar uma economia de R$73 bilhões de reais ao ano. Essa economia se deve a redução dos custos de reparos dos equipamentos e do ganho de eficiência produtiva.

Diante desse cenário desafiador, uma indústria da região de Sorocaba-SP vem investindo em diversos projetos alinhados com os conceitos da indústria 4.0, uma das frentes de investimento está em conectar seus maquinários a sua rede de dados e que vem realizando desde 2016. Recentemente tem acelerado esses investimentos visando obter o máximo de proveito da indústria 4.0 para com seu parque fabril. Conforme gráfico abaixo nota-se que os equipamentos conectados praticamente quadruplicaram desde 2019 e que a tendencia é de possuir todos equipamento conectados no ano de 2027.

Gráfico 1: Equipamento conectados à rede industrial



Fonte: elaborado pelos autores

Com a ascensão dos equipamentos conectados à rede industrial, diversos dados antes desconhecidos podem agora ser estudados e compreendidos de uma maneira melhor. Cada equipamento conectado possui uma variedade de parâmetros e variáveis de processo que podem ser analisados, alguns equipamentos podem possuir mais de 1.000 parâmetros, porém a forma de análise utilizada até então se torna inviável, visto que a empresa não dispõe de uma infinidade de engenheiros e técnicos que possam acompanhar todos esses dados em tempo real. Cada equipamento conectado à rede, também está interligado a um sistema supervisório que já coleta e armazena os dados em um banco de dados relacional.

Com isto, este trabalho tem por objetivo desenvolver uma ferramenta para extrair os dados do banco de dados, analisar, aplicar técnicas estatísticas e de inteligência artificial e gerar relatórios que possam auxiliar os engenheiros e técnicos na tomada de decisão, analisando o comportamento das variáveis e segmentando-as para que os responsáveis possam avaliar e corrigir o processo produtivo de maneira rápida e eficiente.

**2 MÉTODO**

A metodologia do projeto se baseia em algumas técnicas já conhecidas para consumo e análise de dados, como ETL (*extract, transform and load)*, em português: Extrair, Transformar e Carregar, Análise descritiva dos dados, Análise de séries temporais e cálculos estatísticos.

Primeiramente será realizada a ingestão dos dados, para efeito de estudos os dados já foram extraídos do banco de dados relacional e foram salvos em diversos arquivos no formato CSV. Após o processo de carga dos dados, será realizada uma análise descritiva para que se possa entender como os dados são dispostos e quais suas características. A seguir será então implementado técnicas para tratar esses dados e depois realizar as análises estatísticas e aplicação de algoritmos para classificar os dados em normais ou suspeitos (dados que possuem pontos de anormalidades) gerando assim um relatório para que os engenheiros e técnicos possam atuar nos equipamentos.

2.1 COLETA DE DADOS

Como mencionado anteriormente, os dados já são coletados automaticamente por um sistema supervisório que a empresa possui. A cada 1 minuto o sistema supervisório acessa os controladora lógicos programáveis, CLP e salva esses dados no banco de dados relacional, ou seja, são armazenadas cerca de 1.440 informações por dia para cada variável e parâmetro da máquina.

2.2 DADOS DE ENTRADA

Como os dados já foram extraídos do banco de dados relacional e está armazenado no formato CSV, para ler esses dados foi utilizado a biblioteca Pandas para importação do arquivo. Na figura 1 podemos visualizar rapidamente como os dados estão armazenados. Já na figura 2, foi realizado um breve tratamento dos dados para junta-los com base na variável de processo coletada, isso foi feito através da concatenação dos campos nivel\_1, nivel\_2,nivel\_3,nivel\_4,tipo,unidade, e em seguida os valores foram substituídos por um nome genérico (de X0 até X240) e logo foi realizado o processo de *pivot table* onde os dados das linhas viraram novas colunas e assim pode-se ser aplicada a função *describe* da biblioteca pandas em que retorna as principais estatísticas dos dados.

Figura 1: Dados de entrada

Tabela

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Figura 2: Quantidade de parâmetros armazenados

Texto

Descrição gerada automaticamente

Fonte: elaborado pelos autores

2.3 ANÁLISE DESCRITIVA DOS DADOS

Como observado na Figura 2, as colunas iniciam em X0 e vão até X240, ou seja, a base de dados possui 241 variáveis e parâmetros armazenados. Sendo que são distribuídos nos grupos de variáveis de processos do tipo: temperatura, tempo, frequência, rotação, posição, batidas. Para efeito de estudos iremos focar nos dados armazenados do tipo Temperatura que no processo produtivo não deveria variar ao longo do tempo a menos que algum problema ou ajuste operacional tenha acontecido.

Nas figuras 3 e 4 apresentadas abaixo, podemos observar exatamente o ponto focal desse estudo, detectar uma variação dos dados ao longo do tempo. Note que a figura 3 apresenta um comportamento estável durante toda a janela de tempo, já a figura 4 apresenta uma oscilação dos valores de temperatura, permanece assim, e depois retorna aos valores iniciais.

Figura 3: Variável sem desvio no comportamento

Gráfico

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Figura 4: Variável com desvio no comportamento

Gráfico

Descrição gerada automaticamente

Fonte: elaborado pelos autores

2.4 MODELO DE ANÁLISE

De forma resumida, o modelo de análise e classificação dos dados se baseia em segregar os dados de entrada em diversos seguimentos menores, então compará-los um a um e verificar se há evidências estatísticas para afirmar que os dois seguimentos são iguais. Ao fim desse processo espera-se obter uma lista com todos os pontos de ruptura, ou seja, pontos em que o comportamento dos dados é diferente entre um seguimento e outro. A seguir é apresentado todas as etapas do processo.

Etapa 1: Segregação dos dados de entrada

Na etapa de segregação dos dados foi aplicada a técnica da análise da variação da média móvel e do desvio padrão móvel a cada janela de tempo. Dessa forma é possível observar quando há alguma grande oscilação no comportamento dos dados, tanto do ponto de vista da média, como do desvio-padrão. Os passos abaixo foram aplicados para chegar aos resultados:

1. Define-se uma janela de tempo para analisar os dados, nesse caso foi definido como 15 pontos de análise.
2. Calcula-se o desvio-padrão e a média móvel com base nessa janela.
3. Utilizando a função DIFF() e ABS() da biblioteca pandas, é calculada a diferença absoluta da média e do desvio-padrão móvel da linha anterior com a linha atual. Esses dados são armazenados em duas novas colunas: DIFF\_AVG e DIFF\_STD
4. Tomando por base essas novas colunas, foram então calculadas as médias e desvio-padrão de cada uma delas, dessa forma foram obtidos os comportamentos das variações das médias moveis e desvio-padrão ao longo do tempo. Esses dados foram salvos nas variáveis FATOR\_AVG e FATOR\_STD
5. Agora cada ponto da média e do desvio-padrão móvel é dividido por esses dois fatores e então armazenados nas colunas: CHANGE\_STD e CHANGE\_AVG
6. Onde o valor da coluna CHANGE\_STD ou CHANGE\_AVG for maior que um (1), entende-se que é um ponto de ruptura inicial.

A figura 5 mostra exatamente esse processo dos cálculos dos pontos de rupturas e na figura 6 podemos observar os pontos de ruptura aplicados aos dados originais e a criação dos seguimentos de dados identificados na coluna CLUSTER.

Figura 5: Detecção de rupturas na média móvel e no desvio padrão móvel

Interface gráfica do usuário, Aplicativo, Linha do tempo

Descrição gerada automaticamente Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Figura 6: Linha do tempo com os pontos de ruptura iniciais

Gráfico

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Etapa 2: Teste estatístico

Nesta etapa cada grupo de dados que fora segregado na etapa anterior e já classificado com um código na coluna ‘CLUSTER’, é então analisado um a um utilizando os testes estatísticos não paramétricos KRUSKAL-WALLIS e FLIGNER-KILLEEN, esses foram os testes escolhidos pois não se pode afirmar que todos os segmentos possuem uma distribuição normalizada dos dados, dessa forma é possível verificar se cada segmento difere dos demais em relação a mediana ou variância dos dados. Na figura 7 podemos observar o resultado dos testes não paramétricos aplicados a cada segmento.

Figura 7: Testes não paramétricos

Tabela

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Etapa 3: Plotar pontos de ruptura nos dados originais

Agora que os dados foram testados um a um e os pontos de ruptura iniciais foram revisados, restaram apenas alguns pontos de ruptura e que plotados nos dados originais pode-se observar claramente a segregação dos dados que possuem medianas ou variância diferentes. Na figura 8 podemos observar através do boxplot o comportamento desses segmentos o que serve de orientação para a tomada de decisão do usuário final.

Figura 8: Série temporal com os segmentos bem definidos

Gráfico

Descrição gerada automaticamente

Fonte: elaborado pelos autores

Figura 9: Boxplot de cada segmento

**Gráfico, Gráfico de caixa estreita

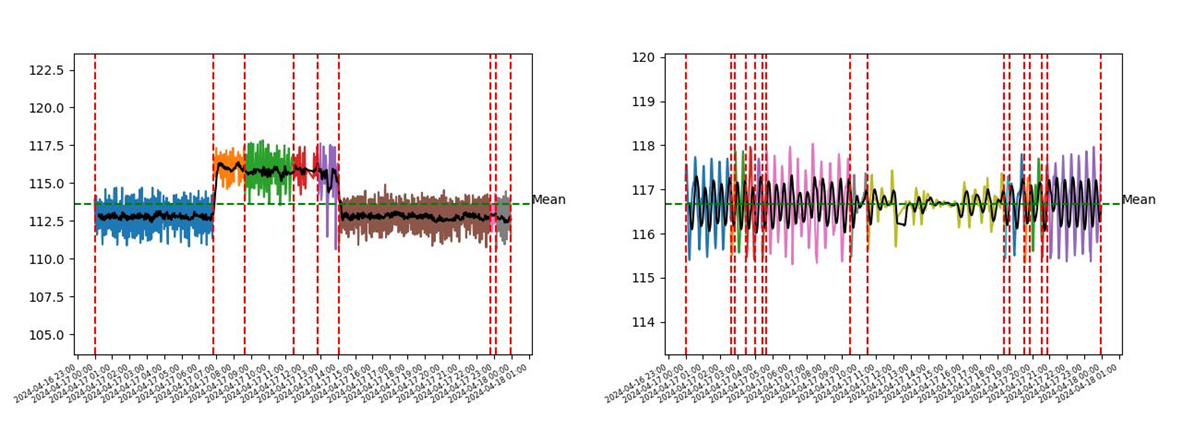
Descrição gerada automaticamente**

Fonte: elaborado pelos autores

2.5 APLICANDO A TÉCNICA PARA AS DEMAIS VARIAVEIS

Ao aplicar as mesmas técnicas para as demais variáveis do tipo temperatura, podemos observar que a técnica foi eficiente na detecção dos pontos de ruptura da série temporal. Ou seja, a técnica foi capaz de detectar e segmentar cada pedaço da série temporal em pedaços menores separando-os com base em duas características principais, no caso a mediana e a variância dos dados. A figura 10 mostra alguns resultados obtidos em outros conjuntos de dados do mesmo tipo de variável, no caso temperatura.

Figura 10: Detecção em outras variáveis.

Fonte: elaborado pelos autores

**4 RESULTADOS E DISCUSSÃO**

As técnicas aqui aplicadas se mostraram capazes de analisar os dados coletados dos equipamentos fabris, segmentar as séries temporais e evidenciar para os engenheiros e analistas o comportamento de cada variável. Como sequência desse estudo, se faz necessário aplicar a mesma técnica aos outros tipos de variáveis, e verificar se possuí a mesma eficácia que observado para a variável de temperatura. Outro ponto importante é que por se tratar de mais de 200 variáveis de processo, se faz necessário retornar uma métrica que indica quais variáveis possuem comportamentos preocupantes, como por exemplo aumento do desvio padrão, ou tendência de a variável sair do controle do processo.

**5 CONSIDERAÇÕES FINAIS**

Como discutido anteriormente, a indústria está em plena expansão e conectando seus diversos equipamentos a sua rede de comunicação e com isso uma infinidade de dados pode ser coletada, estudados e com isso ser transformados em melhorias que trazem redução de custos e aumento de produtividade.

Esse trabalho teve como proposta a coleta, análise e segmentação dos dados com base em suas características para que os engenheiros e analistas da área produtiva possam avaliar e tomar medidas corretivas e até preventivas para manter o processo produtivo estável e operante.

**REFERÊNCIAS**

ABDI. Agência Brasileira de Desenvolvimento Industrial. **Indústria 4.0 pode economizar R$ 73 bilhões ao ano para o Brasil.** https://www.abdi.com.br/industria-4-0-pode-economizar-r-73-bilhoes-ao-ano-para-o-brasil/. Acesso em: 07 set. 2024

CGEE. Centro de Gestão e Estudos Estratégicos. **Arcabouço normativo para implementação da Industria 4.0 no Brasil**. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivo-camara-industria/iniciativas/ci_nt_arcabouco-normativo.pdf>. Acesso em: 07 set. 2024

*CHANGEPOINT: An* ***R*** *Package for Changepoint Analysis. Rebecca Killick and Idris A. Eckley.* Disponível em: https://www.jstatsoft.org/article/view/v058i03. Acesso em: 10/08/2024

KRUSKAL-WALLIS: definição, fórmula e exemplo. Disponível em: https://statorials.org/pt/teste-kruskal-wallis/. Acesso em: 15/09/2024

PANDAS, Disponível em: https://pandas.pydata.org/docs/reference/index.html. Acesso em: 15/07/2024