



## Análise preditiva de um sistema de falha de máquinas

### Predictive analysis of a machine failure system

Lucas Henrique Silva Ribeiro<sup>1</sup>, Prof. Dr. Edson Pinheiro de Lima<sup>2</sup>

#### RESUMO

O grande volume de dados gerado a partir das últimas duas décadas precisou de um novo jeito de ser armazenado e processado, métodos tradicionais de banco de dados já não suportam a quantidade de dados gerados, não só pela quantidade, mas também pela frequência de novas informações, e com o suporte de pesquisa, ferramentas de *data science*, e em paralelo, o avanço da infraestrutura de telecomunicação, tornou possível acompanhar tudo em tempo real. Monitorar, descrever, analisar, prever e até tomar decisões sobre uma série de comportamentos, mudou o modo de como o mercado enxerga a gestão de manufatura, energia, mercado financeiro, comportamento humano e de máquinas. Este trabalho estuda, com base na literatura de análise de dados gestão de manufatura, uma análise descritiva/preditiva de um sistema de falha de máquina que, com base na descrição de comportamento, permite identificar sintomas de potenciais problemas na linha e agir preventivamente.

**PALAVRAS-CHAVE:** análise de dados; big data; gestão de manutenção.

#### ABSTRACT

The large volume of data generated over the last two decades has required a new way of storing and processing it. Traditional database methods can no longer cope with the amount of data generated, not just because of the amount but also because of the frequency of new information. With the support of research, data science tools, and in parallel, advances in telecommunications infrastructure, it has become possible to monitor everything in real-time. Monitoring, describing, analyzing, predicting, and even making decisions about various behaviors has changed how the market sees the management of manufacturing, energy, the financial market, and human and machine behavior. Based on the literature of data analysis in manufacturing management, this paper studies a descriptive/predictive analysis of a machine failure system, which, based on the description of behavior, makes it possible to identify symptoms of potential problems on the line and take preventive action.

**KEYWORDS:** data analysis; big data; maintenance management.

## MATERIAIS E MÉTODOS

O big data somado a análise de dados (BDA) desempenha um papel crucial na gestão de manutenção, ajudando as empresas a tomar decisões informadas e eficazes para garantir que seus equipamentos e ativos operem de forma eficiente e confiável, segundo McAfee e Brynjolfsson (2012), “[t]he more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results”. A seguir será apresentado um resumo dos conceitos citados.

Big data: se trata de um conjunto de dados extremamente grandes e complexos que estão além da capacidade das ferramentas e métodos tradicionais de processamento de

<sup>1</sup> Bolsista da Fundação Araucária. Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: [lucashenrique@alunos.utfpr.edu.br](mailto:lucashenrique@alunos.utfpr.edu.br).

<sup>2</sup> Docente no Curso de Engenharia Elétrica/DAELT/PIBIC. Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil. E-mail: [pinheiro@professores.utfpr.edu.br](mailto:pinheiro@professores.utfpr.edu.br) ID Lattes: 7304491503357135.



dados para gerenciar, analisar e extrair insights significativos. Algumas literaturas (Gartner, 2012; Kwon and Sim, 2013; McAfee and Brynjolfsson, 2012) caracterizam o big data por três principais atributos, frequentemente referidos como os três Vs:

**Volume:** O big-data envolve vastas quantidades de dados. Isso pode variar de terabytes a petabytes ou até mesmo exabytes de informações. O volume de dados em si é uma das características definidoras do big data.

**Velocidade:** Os dados no contexto de big data são gerados, coletados e processados em alta velocidade. Isso inclui fluxos de dados de várias fontes, como sensores, mídias sociais, aplicativos da web e mais.

**Variedade:** O big data é diversificado em termos de tipos e formatos de dados. Ele inclui dados estruturados (como dados em bancos de dados relacionais), dados semi-estruturados (como documentos XML ou JSON) e dados não estruturados (como texto, imagens, vídeos e postagens em mídias sociais). Lidar com essa variedade requer ferramentas e técnicas especializadas.

**Análise de dados:** é o processo de coletar, limpar, transformar e interpretar dados para descobrir informações úteis, padrões, tendências e insights que possam orientar a tomada de decisões informadas. Existem diversos tipos de análises de dados, cada um com seus objetivos e métodos específicos. Aqui estão alguns dos principais tipos de análises de dados:

**Análise descritiva:** É o tipo mais básico de análise de dados, que se concentra em resumir e descrever os dados. Isso inclui calcular médias, medianas, moda, desvio padrão e criar visualizações simples, como gráficos de barras e histogramas.

**Análise Preditiva:** A análise preditiva usa técnicas de modelagem estatística e aprendizado de máquina para fazer previsões sobre eventos futuros com base em dados históricos. Isso é amplamente usado em previsão de vendas, análise de churn de clientes e previsões de demanda, entre outros.

**Análise Prescritiva:** A análise prescritiva vai além da previsão e sugere ações ou decisões a serem tomadas com base nos dados. Isso envolve a otimização de processos e a escolha das melhores opções em situações complexas.

Neste trabalho, apresentaremos um estudo que envolve análise descritiva e preditiva de uma base de dados obtida na plataforma Kaggle. Essa base de dados contém informações de 10.000 registros coletados de uma máquina industrial. Cada registro inclui os seguintes dados:

1. UID: Um número de identificação único para cada produto fabricado pela máquina.
2. Product ID: Uma classificação que indica a qualidade do produto, sendo representada pelas letras L, M ou H, que correspondem às categorias de qualidade: baixa, média e alta.
3. Air Temperature: A temperatura do ar em Kelvin (K).
4. Process Temperature: A temperatura do processo em Kelvin (K).



5. Rotational Speed: A velocidade de rotação da máquina em rotações por minuto (rpm).
6. Torque: O torque gerado pela máquina em Newton-metros (Nm).
7. Tool Wear: O tempo de desgaste da máquina, com variações de 2, 3 e 5 minutos, dependendo da categoria de qualidade L, M ou H, respectivamente.
8. Target: Uma indicação se a máquina falhou ou não.
9. Failure Type: A classificação do tipo de falha, quando ocorre.

Este estudo tem como objetivo realizar análises descritivas para entender a distribuição e comportamento desses dados. Em seguida, aplicaremos análises preditivas com técnicas estatísticas e de aprendizado de máquina para prever falhas na máquina com base nas informações disponíveis. O propósito final é aprimorar a manutenção preventiva da máquina, reduzir o tempo de inatividade devido a falhas e melhorar a qualidade dos produtos fabricados.

No âmbito da estatística e a probabilidade, ambas desempenham papéis cruciais e interconectados na análise de dados. Esses dois campos da matemática estatística fornecem as ferramentas essenciais para entender, interpretar e tirar conclusões a partir de conjuntos de dados complexos e frequentemente desordenados. Eles são as bases sólidas que sustentam o processo de análise de dados, proporcionando uma estrutura para a compreensão da incerteza, a identificação de padrões e relações ocultas.

A primeira etapa na análise dos dados envolveu a aplicação de técnicas estatísticas para compreender seu comportamento. Foram calculadas a média, o desvio padrão e a variância, permitindo-nos avaliar a tendência central e a dispersão dos dados. É importante destacar que os dados do Kaggle já estavam previamente depurados, não sendo necessário remover nenhum valor discrepante. Abaixo, apresentamos a tabela com os resultados dessas análises:

**Quadro 1: Análise Estatística dos dados**

	Média	Desvio Padrão	Variância
Air temperature(K)	300,00	2,00	4,00
Process Temperature (K)	310,00	1,49	2,23
Rotational Speed(rpm)	1538,78	179,28	32.1421,79
Torque (Nm)	39,99	9,97	99,38

Fonte: o autor, 2023

Em seguida, foi necessário verificar se os dados aderem a uma distribuição normal, que é também referida como uma distribuição gaussiana. A importância dos testes de normalidade reside na necessidade de verificar se essa suposição é válida. Quando os dados seguem uma distribuição normal, é possível fazer inferências mais precisas sobre a população a partir da amostra, além de aplicar técnicas estatísticas paramétricas com

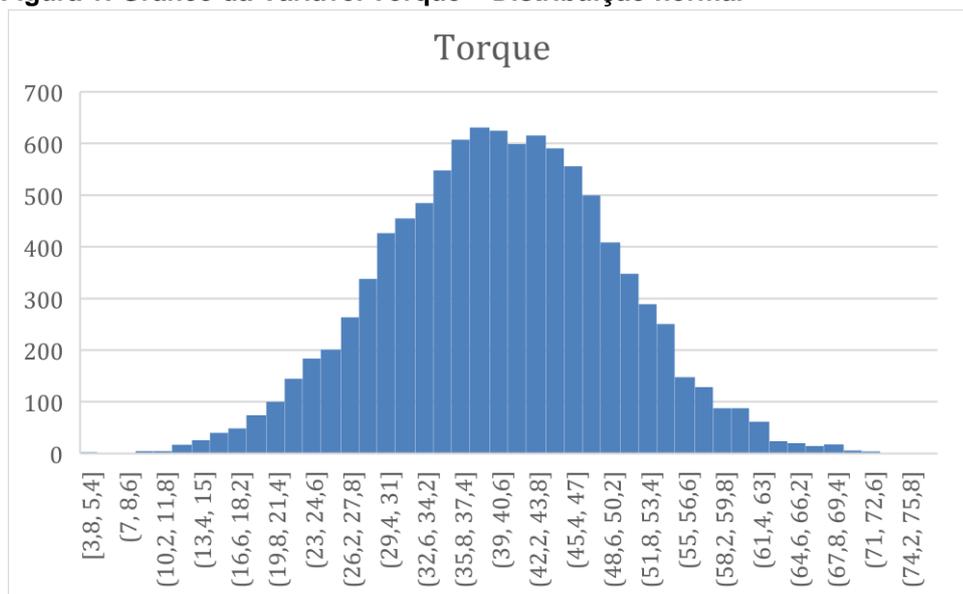
segurança. No entanto, se os dados não seguem uma distribuição normal, é necessário considerar métodos estatísticos não paramétricos ou transformações de dados.

O teste de Shapiro-Wilk é amplamente utilizado para verificar a normalidade dos dados. Ele calcula uma estatística de teste e compara-a a uma distribuição de probabilidade conhecida. Se o valor-p associado ao teste for maior que um nível de significância escolhido (geralmente 0,05), a hipótese nula é aceita, indicando que os dados provavelmente seguem uma distribuição normal.

Já o teste ANOVA funciona comparando a variabilidade entre os grupos (variabilidade entre as médias) com a variabilidade dentro dos grupos (variabilidade dentro de cada grupo). Se a variabilidade entre as médias for significativamente maior do que a variabilidade dentro dos grupos, isso sugere que pelo menos um dos grupos difere dos outros de maneira estatisticamente significativa.

Para realizar ambos os testes de Shapiro-Wilk, usou-se o auxílio do software PAST, segue abaixo um exemplo de gráfico de normalidade da variável torque.

Figura 1: Gráfico da Variável Torque – Distribuição normal



Fonte: o autor, 2023

## RESULTADOS E DISCUSSÕES

Com posse dos dados e de seus respectivos gráficos, e testes foi possível cruzar as informações das variáveis e convalidar os cinco modos de falha da máquina, sendo eles:

1. Desgaste da Ferramenta (Tool Wear Failure): Este motivo de falha ocorre quando a ferramenta usada na máquina está desgastada. A falha pode ocorrer aleatoriamente a partir de 200 minutos de uso da ferramenta.



2. Dissipação de Calor (Heat Dissipation Failure): A falha por dissipação de calor ocorre quando a diferença entre a temperatura do ar e a temperatura do processo é menor do que 8,6 Kelvin e a velocidade da máquina está abaixo de 1380 rotações por minuto (rpm).
3. Potência (Power Failure): Esta falha ocorre se o produto do torque e da velocidade de rotação (em radianos por segundo) estiver abaixo de 3500 Watts ou acima de 9000 Watts, o que indica que a máquina não está fornecendo a potência adequada para o processo.
4. Esforço Excessivo (Overstrain Failure): O motivo de falha por esforço excessivo ocorre quando o produto do desgaste da ferramenta e do torque excede 11.000 minuto-newtons metros (minNm) para a variante do produto L (ou 12.000 para M e 13.000 para H). Isso indica que a máquina está sendo submetida a um esforço excessivo.
5. Falha Aleatória: Cada processo tem uma chance de 0,1% de falhar aleatoriamente, independentemente de outros fatores.

Com base nessas informações, é proposto o desenvolvimento de um sistema supervisor avançado para monitorar a máquina e fornecer alertas em tempo real ao operador. Esse sistema terá a capacidade de monitorar diversos parâmetros críticos, incluindo tempo de operação, temperatura, contagem de produtos produzidos com informações sobre sua qualidade e velocidade de rotação (RPM).

O objetivo primordial desse sistema supervisor é fornecer ao operador uma visão abrangente e imediata do estado da máquina, permitindo uma tomada de decisão ágil e eficaz para evitar falhas e interrupções não planejadas na produção. Aqui estão alguns aspectos-chave desse sistema:

- a) em Tempo Real: O sistema supervisionará constantemente os parâmetros essenciais da máquina em tempo real. Isso incluirá a contagem precisa dos produtos fabricados, identificando cada um deles quanto à sua qualidade e registrando o tempo de operação contínuo.
- b) Alertas Imediatos: Sempre que qualquer um dos parâmetros monitorados se aproximar ou ultrapassar os limites críticos estabelecidos, o sistema emitirá alertas instantâneos ao operador. Esses alertas serão visuais e sonoros, garantindo que o operador possa responder prontamente.
- c) Registros e Históricos: O sistema também manterá registros detalhados de todas as atividades e eventos relevantes, incluindo falhas anteriores, alertas acionados e ações tomadas para resolver problemas. Isso permitirá uma análise posterior para identificar tendências e padrões.
- d) Integração com Qualidade do Produto: Além de monitorar a produção, o sistema supervisionará a qualidade dos produtos, registrando informações detalhadas sobre



cada item fabricado. Isso ajudará a identificar se há uma correlação entre falhas na máquina e produtos de qualidade inferior.

- e) Interface de Usuário Intuitiva: O sistema contará com uma interface de usuário intuitiva e de fácil compreensão, exibindo gráficos, alertas e informações essenciais em tempo real. Isso permitirá que o operador tenha uma visão clara do status da máquina.
- f) Capacidade de Análise de Dados: O sistema poderá analisar os dados coletados ao longo do tempo para identificar possíveis melhorias no processo de produção e no desempenho da máquina.
- g) Integração com Manutenção: O sistema também poderá ser integrado com o departamento de manutenção, permitindo que as equipes de manutenção sejam acionadas automaticamente em caso de falha grave ou necessidade de manutenção preventiva.

Em resumo, o sistema supervisorio proposto será uma ferramenta fundamental para otimizar a eficiência operacional, minimizar o tempo de inatividade e garantir a qualidade do produto. Ao proporcionar um monitoramento abrangente e a capacidade de resposta imediata a problemas potenciais, ele contribuirá significativamente para a operação bem-sucedida da máquina.

## CONCLUSÃO

A aplicação da análise de dados em conjunto com o Big Data desempenha um papel fundamental na gestão eficaz da manutenção industrial, respaldando a afirmação de McAfee e Brynjolfsson (2012) de que empresas que se autodenominam orientadas por dados geralmente alcançam melhores resultados financeiros e operacionais. Nesse contexto, o conceito de Big Data, caracterizado pelos três Vs - volume, velocidade e variedade, representa a vasta quantidade de informações, sua rápida geração e a diversidade de formatos. Paralelamente, a análise de dados abrange desde a análise descritiva até a análise preditiva e prescritiva, permitindo extrair insights valiosos a partir de dados complexos.

Um estudo de caso prático foi apresentado, envolvendo uma base de dados de uma máquina industrial. Inicialmente, realizaram-se análises estatísticas descritivas para entender a distribuição dos dados, seguidas de testes de normalidade para validar a suposição de que os dados seguem uma distribuição normal. Essas etapas preparatórias foram cruciais para identificar e validar os principais modos de falha da máquina.

Propôs-se, então, o desenvolvimento de um sistema supervisorio avançado que monitora em tempo real os parâmetros críticos da máquina, emite alertas imediatos ao operador, mantém registros detalhados e oferece análises de dados para melhorar a eficiência operacional e a confiabilidade da máquina. Em um mercado industrial cada vez mais competitivo, a capacidade de tomar decisões informadas com base em dados se revela essencial para otimizar processos, reduzir custos e aprimorar a qualidade dos



produtos, consolidando-se como uma tendência essencial para empresas que almejam se destacar globalmente.

### **Agradecimentos**

Agradeço ao meu orientador Prof. Edson Pinheiro de Lima e a Fundação Araucária pela dedicação, paciência, incentivo e oportunidade.

### **Conflito de interesse**

“Não há conflito de interesse”.

### **REFERÊNCIAS**

GARTNER, 2012. Big Data. Available from: (<http://www.gartner.com/it-glossary/bigdata/>) (retrieved 09.07.13)

KWON, O., Sim, J.M., 2013. Effects of data set features on the performances of classification algorithms. *Expert Syst. Appl.* 40 (5), 1847–1857. <http://dx.doi.org/10.1016/j.eswa.2012.09.017>.

MCAFEE, A., Brynjolfsson, E., 2012. Big data: the management revolution. *Harv. Bus. Rev.* October, 61–683

WALLER, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77–84.