

PROPOSTA DE ARQUITETURA DE PROCESSAMENTO DE DADOS PARA MANUTENÇÃO PROGNÓSTICA UTILIZANDO DATA ANALYTICS

Robson de Siqueira Oggioni

Master Business em Computação Aplicada pelo IITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
rsoggioni@gmail.com

Manoel Villas Bôas Júnior

Mestre em Computação Aplicada, IITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
mvbjunior@poli.ufrj.br

Edilberto Strauss

Ph.D. em Ciência e Tecnologia, IITLab/POLI da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil
strauss@poli.ufrj.br

RESUMO

O cenário tecnológico atual remete a indústria a integração denominada Cyber Physical System, considerando a interface entre o universo físico de produção e a conectividade em redes no ambiente virtual. Sensores de diversas finalidades, pulverizados por todo o ambiente fabril, permitem que controladores digitais e monitores computadorizados mapeiem e registrem todas as atividades fabris, gerando um enorme volume de dados, estocados em bancos de dados. Utilizando de Data Analytics, como ferramenta de apoio na Engenharia Manutenção, este artigo apresenta uma proposta de Pipe Line, em arquitetura de gestão de dados, para a realização de manutenção preditiva baseada em prognósticos, em uma estrutura pré-existente de uma Indústria 4.0. Com uso de técnicas de Machine Learning, baseado em um processo de Knowledge Discovery in Databases, será apresentada análise de banco histórico de dados real, demonstrando a viabilidade e simplicidade da manutenção baseada em prognósticos, proporcionando uma antecipação da falha em 48h e 96h com desvio RMSE menor que 0,1.

Palavras-chave: Cyber Physical System, Machine Learning, Industria 4.0.

DATA PROCESSING ARCHITECTURE PROPOSAL FOR PROGNOSTIC MAINTENANCE USING DATA ANALYTICS

ABSTRACT

The current technological scenario leads the industry to the so-called Cyber Physical System integration, considering the interface between the physical universe of production and network connectivity in the virtual environment. Sensors for different purposes, scattered throughout the manufacturing environment, allow digital controllers and computerized monitors to map and record all manufacturing activities, generating a huge volume of data, stored in databases. Using Data Analytics, as a support tool in Maintenance Engineering, this article presents a Pipeline proposal, in data management architecture, to carry out predictive maintenance based on prognoses, in a pre-existing structure of an Industry 4.0. With the use of Machine Learning techniques, based on a Knowledge Discovery in Databases process,

analysis of a real historical database will be presented, demonstrating the viability and simplicity of maintenance based on prognoses, providing a failure anticipation of 48h and 96h with RMSE deviation less than 0.1.

Keywords: Cyber Physical System, Machine Learning, Industry 4.0.

1 INTRODUÇÃO

O cenário atual de desenvolvimento tecnológico, junto ao recente fenômeno da globalização, vem forçando as indústrias a cada vez mais adquirirem tecnologia para se manterem competitivas. Na busca de aumento produtivo e atender as necessidades de qualidade de um mercado consumidor exigente, os sistemas de controle e monitoramento baseados em sensores optoeletrônicos e sistemas computacionais vem ganhando espaço. Conforme o quadro 1 de Macedo (2020), o desenvolvimento tecnológico é algo natural na indústria, sendo dividido em 4 fases, conhecidas como revoluções industriais.

Quadro 1 – Revoluções Industriais

Revoluções Industriais	Indústria 1.0	Indústria 2.0	Indústria 3.0	Indústria 4.0
Características	Mecanização Poder do Vapor Tear de Tecelagem	Produção em Massa Linhas de Montagem Energia Elétrica	Automação Computadores Eletrônicos	Cyber Physical IoT Redes Computação em Nuvem
Tipos de Manutenção	Corretiva	Preventiva	Produtiva Total	Preditiva
Inspecção	Inspecção Visual	Inspecção Instrumentada	Monitoramento por Sensores	Preditiva e Análises
OEE	< 50%	50% - 70%	75% - 90%	> 90%
Equipe de Manutenção	Treinado Artesãos	Inspetores	Engenharia de Confiabilidade	Dados Inteligentes

OEE – OVERALL EQUIPMENT EFFECTIVENESS = DISPONIBILIDADE x DESEMPENHO x QUALIDADE

Fonte: Macedo (2020)

Ao fim dos anos 70 e início dos anos 80, a crescente aplicação de dispositivos optoeletrônicos para controle substituiu os antigos painéis eletromecânicos, dando destaque para a inserção de sistemas digitais de controle, como os Controladores Lógicos Programáveis (CLPs) e redes industriais de dados, como as redes profibus, modbus e arqinet. Com isso, houve um incremento de princípios básicos para a utilização da automação de equipamentos e processos, induzindo a recursos integrados informatizados nas fábricas.

Oggioni (2023) enfatiza que, hoje o desenvolvimento das tecnologias de análise computacional de dados, somado a crescente migração dos processos industriais para sistemas cada vez mais automatizados, onde se geram um volume considerável de dados por sensores instalados de forma pulverizada em equipamentos e sistemas, com um registro em bancos de dados históricos gigantescos, tornou-se factível o mapeamento de toda vida funcional de equipamentos ou sistemas, com seu histórico de degradação e quebras.

Segundo Tropias et al (2017), “a Indústria 4.0 considera a interface entre o universo físico de produção e a conectividade em redes no ambiente virtual, permitindo que recursos, informações, objetos e pessoas estejam conectados”.

Neste cenário, continua Oggioni (2023), fica claro um novo ramo a ser desenvolvido nas diretrizes da manutenção industrial, por meio de uma estratégia baseada em um novo conceito que vem se difundindo nas últimas décadas, o Cyber-Physical System – sistemas industriais onde processos físicos são controlados e supervisionados por sistemas computacionais interligados por rede de dados – levando a um novo passo de prognósticos baseado em dados.

A Indústria 4.0 introduziu o Cyber-Physical System nas grandes indústrias, tornando possível a utilização das ferramentas e técnicas de análise de dados. Baldisarelli e Fabro (2019) ressaltam que visto a própria natureza dos sistemas de aquisição de dados de um sistema automatizado industrial, já que para o caso de processos industriais, uma abordagem bastante natural consiste na utilização de dados estruturados, uma vez que os sistemas de automação e informação fazem a geração e aquisição de dados de uma forma sistêmica a partir de fontes bem conhecidas (instrumentos, equipamentos, processos, etc.). Ressalta Fonseca (2014) que, a vantagem da estruturação dos dados está numa maior facilidade para habilitação de análises mais simples com ferramentas e técnicas menos sofisticadas, uma vez que o tratamento de dados não-estruturados implica normalmente em algoritmos e técnicas mais complexas.

A indústria torna-se cada vez mais cibernética, conforme Baheti e Gill (2011) e Lee et al (2015), onde a automação dentro da indústria torna-se característica intrínseca, estando as várias camadas de controle e registro interligadas por redes de dados na velocidade de Terabytes por segundo (TB/s). Neste cenário, a interação com os dados gerados se torna valioso recurso para aquisição de conhecimento e maximização da produtividade.

Fonseca (2014), traz um pequeno vislumbre do volume de dados, considerando uma planta industrial de porte médio, com um sistema de monitoramento e controle, que utiliza uma quantidade média em torno de 15.000 sensores para monitoramento e armazenamento, em uma taxa de atualização de 1 segundo, sendo 4 bytes de memória para o registro de cada variável

monitorada por sensor. Esta configuração ‘simplista’ nos traria um volume da ordem de 20 Mbytes por dia e 7,2 GBytes por ano.

2 CONTEXTUALIZAÇÃO DO CENÁRIO

‘A Engenharia de Manutenção tem como funções fundamentais a realização sistemática da coleta, o tratamento dos dados para subsidiar os estudos e as análises de manutenção, bem como a proposição de melhorias para o processo de gestão e execução da manutenção e melhoria do desempenho operacional dos equipamentos e instalações’ (BALDISSARELLI e FABRO, 2019)

Progressivamente as técnicas produtivas e os requisitos de qualidade, na geração de produtos e serviços, vem se desenvolvendo tecnologicamente em velocidade crescente. Naturalmente a disciplina de Manutenção necessita se desenvolver em suas técnicas e processos para garantir o máximo desempenho e qualidade em um mercado competitivo, sendo um fator determinante de lucratividade.

Marcorin e Lima (2013) discutem o erro do entendimento da Manutenção como custo indireto, onde se tenta minimizar ao máximo os seus custos. O investimento em Manutenção deve ser adequado ao retorno da lucratividade, onde se apresenta o ‘ponto ótimo’ na curva de gastos de manutenção e lucratividade, visto na figura 1.

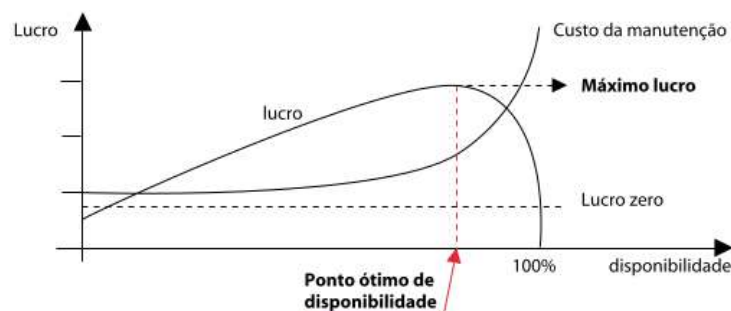


Figura 1 – Gráfico da Curva de Idealização de Custos de Manutenção e Lucratividade
Fonte: MARCORIN e LIMA, 2013

Observando como este processo de análise de dados ocorreria, Lee et al (2014) faz uma comparação entre as disciplinas de Medicina e Engenharia, sendo que a Medicina estuda o comportamento e relações biológicas do corpo humano, entendendo o seu perfeito

funcionamento, os fatores que o desregulam ou adoecem e os sintomas das doenças, e a Engenharia realiza o mesmo para máquinas e sistemas eletromecânicos. Desta comparação se herda os termos prognóstico e diagnóstico. Assim, o diagnóstico é a determinação da “disfunção da saúde” do equipamento ou sistema, através da leitura das indicações dos sensores no momento da falha, enquanto o prognóstico é a determinação de indícios ou fatores, na leitura dos sensores, que possam vir a se desenvolver um “disfunções da saúde” do equipamento ou sistema futuramente.

Lee et al (2014) trata este contexto ao trabalhar os conceitos de diagnóstico, ao classificar a falha para correção, e prognóstico, ao determinar os indícios de degradação para evitar a falha (figura 2). Durante as três primeiras revoluções industriais a possibilidade de prognóstico era limitada, possibilidade que se ampliou com a utilização de Cyber-Physical Systems na quarta revolução industrial.

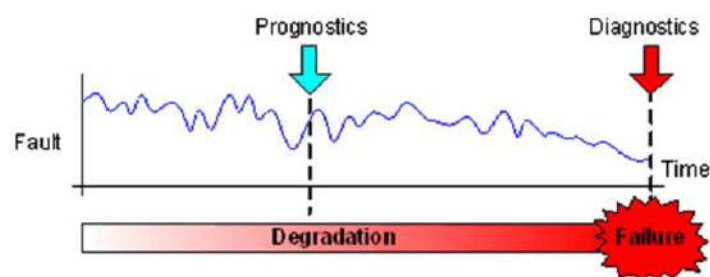


Figura 2 – Diferentes percepções entre Prognóstico e Diagnóstico

Fonte: LEE et al (2014)

Nesta contextualização, se traz o conceito de Prognóstico, onde a monitoração de deterioração de peças, previsão de futuras faltas e mensuração de expectativa de vida útil de equipamentos e sistemas se torna o caminho natural da Manutenção na Indústria 4.0. O Data Analytics torna-se ferramenta na adequação dos planejamentos de manutenção das indústrias já inclusas no ciclo da Indústria 4.0, sendo base de uma manutenção preditiva baseada em prognósticos, desempenhando a função de detectar os problemas e falhas antes que os mesmo se tornem impeditivos ao bom funcionamento do equipamento, sistema ou plantas industriais.

Com a evolução tecnológica, os projetos de equipamentos e sistemas industriais também crescem em complexidade, e conseqüentemente, as incertezas do processo produtivo aumentam. A adoção de sistemas de gestão de Manutenção baseados em prognóstico é uma necessidade atual. Nessa perspectiva, entende-se adequada a implementação da filosofia de manutenção baseada em Prognostics and Health Management (PHM), onde objetiva-se a antecipação de falhas

e problemas ocultos através de análise de dados, culminando em um processo de Knowledge Discovery in Databases (KDD) ou, traduzindo, conhecimento descoberto em bases de dados. A utilização de técnicas de Data Analytics e Machine Learning torna possível a criação de modelos de análise de tendências e classificação de desvios, antecipando problemas e quebras antes que se tornem iminentes. Os resultados destes modelos alimentarão sistemas de apoio a decisão voltados à engenharia de manutenção, garantindo ganhos de confiabilidade, disponibilidade e lucratividade.

3 MANUTENÇÃO PREVENTIVA BASEADO EM PROGNÓSTICOS

O conceito de prognóstico na Engenharia de Manutenção está muito além de ser apenas uma manutenção preditiva. Lee et al (2014) destaca que o conceito de prognóstico na Manutenção já é aplicado há mais de 20 anos, no entanto, “a maioria dessas aplicações aborda apenas curva de degradação ou previsão de vida útil restante (‘Remaining Useful Life’ – RUL), que é apenas uma faceta de prognósticos e gerenciamento de estado.

Lee et al (2014) descreve o PHM como uma disciplina da Engenharia de Manutenção, que visa fornecer aos usuários uma visão integrada do estado de saúde de uma máquina ou de um sistema geral. O desenvolvimento do conceito e a estrutura do PHM foram baseados em metodologias de manutenção e técnicas de diagnóstico largamente utilizadas nos dias atuais, como manutenção preventiva (Preventative Maintenance – PM), manutenção centrada na confiabilidade (Reliability Centered Maintenance – RCM) e manutenção baseada na condição (Condition Based Maintenance – CBM).

As filosofias de prognóstico, exemplificada pelo PHM, vem ganhando corpo e oportunidade ao se aumentar a complexidade e a incerteza do equipamento ou sistema, conforme a figura 3.

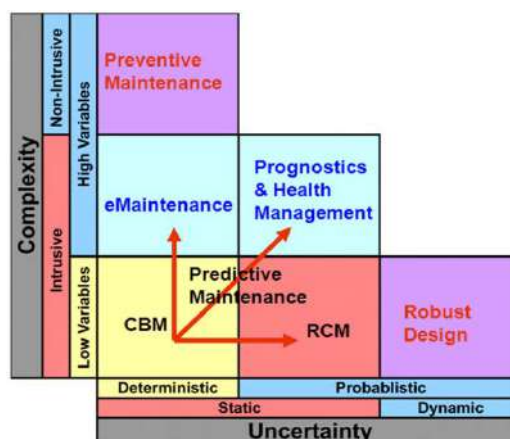


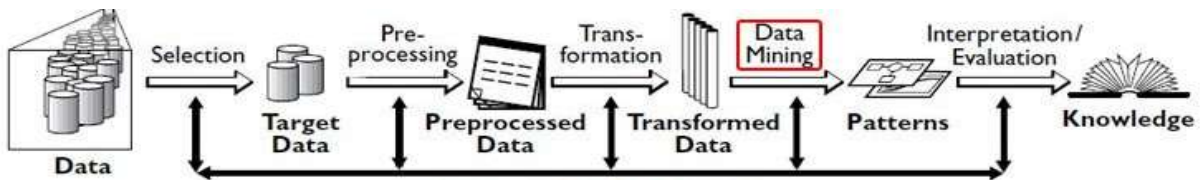
Figura 3 – Mapa de Complexidade e Incerteza por Metodologia de Manutenção

Fonte: LEE et al (2014)

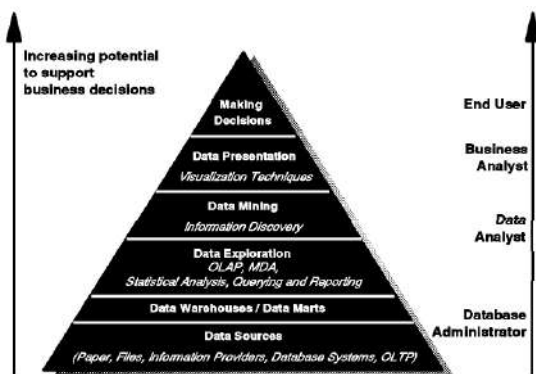
Os estudos em PHM, atualmente, vem se concentrando em prognósticos e gerenciamento de estado dos equipamentos e sistemas, com foco na detecção de falhas incipientes, avaliação de estado atual e previsão de vida útil restante. Devido a natureza da Indústria 4.0 a adoção do conceito PHM torna-se extremamente viável, trazendo a relação custos de manutenção e lucro para o ponto ótimo.

4 – O CONHECIMENTO DESCOBERTO EM DADOS

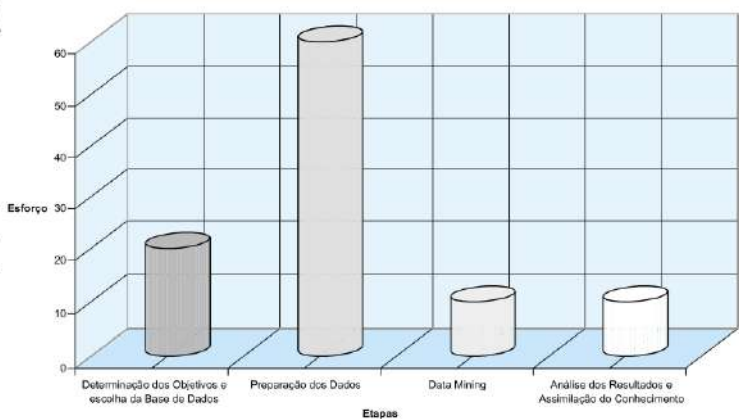
Conforme Boente et al (2008), o conhecimento descoberto em bases de dados ou, sua sigla em inglês, KDD é um campo de estudos originado da necessidade de se analisar grandes e múltiplas bases de dados. Dantas et al (2008) acrescenta que, o KDD pode ser visto como o processo de descoberta de padrões e tendências por análise de grandes conjuntos de dados, tendo por principal etapa o processo de mineração. Conforme figura 4(a). Dentro do processo de KDD, a mineração de dados é um processo de alto valor agregado, apesar de envolver pouco esforço na execução, conforme figuras 4(b) e 4(c).



(a)



(b)



(c)

Figura 5 - Processo do conhecimento descoberto em dados (a), gráficos de relação de valor agregado (b) de relação de esforço (c) nas fases de descoberta de conhecimento em bases de dados

Fonte: FAYYAD et al (1996), CADENA et al, 1999 e QUONIAM et al, 2001

Quando as bases de dados crescem em suas dimensões, tornando impraticável a exploração puramente pela capacidade humana, a utilização de recursos computacionais é uma realidade irreversível (FAYYD et al, 1996; BOENTE et al, 2008)

5 – DATA ANALYTICS

Conforme Runkler (2020), a análise de dados é definida como a aplicação de sistemas de computador para a análise de grandes conjuntos de dados para o suporte de decisões. A análise de dados é um campo muito interdisciplinar que adotou aspectos de muitas outras disciplinas científicas, como estatística, aprendizado, reconhecimento de padrões, teoria de sistemas, pesquisa operacional ou inteligência artificial. A literatura descreve diversas técnicas de análise, sendo as mais comuns apresentadas no quadro 2, destacando as filosofias para a aplicação de tarefas de Mineração de Dados.

Quadro 2 – Técnicas de Mineração de Dados

Técnica	Descrição	Tarefas
Árvore de Decisão	Baseada em estágios de decisão (nós) e na separação de classes e subconjuntos, organiza os dados de forma hierárquica.	- Classificação - Predição
Redes Neurais	Modelos inspirados na fisiologia do cérebro, nos quais o conhecimento é fruto do mapa de conexões neurais e de pesos dessas conexões.	- Classificação - Agrupamento - Predição
Raciocínio Baseado em Caos	Baseado em método do vizinho mais próximo, combina e compara atributos para estabelecer hierarquia de semelhança.	- Classificação - Agrupamento
Algoritmos Genéricos	Métodos gerais de busca e otimização, inspirados na Teoria da Evolução, em que cada nova geração, soluções melhores tem mais chance de ter “descendentes”	- Classificação - Agrupamento
Conjuntos Fuzzy	Oferece uma grande vantagem para classificar dados com um alto nível de abstração.	- Classificação - Agrupamento
Regras de Indução	Processo para obter uma hipótese a partir de dados já existentes.	- Classificação - Predição
Regras de Associação	Estabelece uma correlação estatística entre atributos de dados e conjunto de dados.	- Associação

Fonte: FURLAN (2018 apud GOLDSCHIMIDT, 2005)

Furlan (2018) diz que os métodos, também chamados de técnicas de Mineração de Dados, se estendem a qualquer teoria que possa fundamentar o processo de Mineração de Dados.

6 PROPOSTA DE APLICAÇÃO

Conforme exposto anteriormente, a adoção de técnicas de prognóstico (LEE et al, 2014) antecipando as falhas e os desvios nos equipamentos e nos sistemas é fundamental para a competitividade e rentabilidade (MARCORIN e LIMA, 2013), garantindo a máxima disponibilidade, qualidade desejada e a eficiência no processo industrial. Dessa maneira, a manutenção passa a ser considerada como ponto decisivo na disputa comercial, tornando-se investimento estratégico com retorno garantido e não mais como custo indireto produtivo.

Na Indústria 4.0, a adoção de técnicas de prognóstico se torna de aplicação natural, visto a necessidade de investimento relativamente baixo. As estruturas fabris baseadas na tecnologia de *Cyber-Physical System* (TROPIAS et al, 2017 e LEE et al, 2015) já estão preparadas para a adoção de sistemas de gerenciamento de Manutenção, tais como *Prognostics and Health Management* (LEE et al, 2014).

Através de métodos simples de *Data Analytics* e *Machine Learning*, como a adoção de uma estrutura em *pipe line*, conforme a figura 7, teríamos análises rotineiras, em uma visão de antecipação de falhas futuras, de acompanhamento da degradação de componentes e a previsão da vida útil de equipamentos e sistemas. Essas análises rotineiras são baseadas em métodos como a predição, que anteciparia o tempo para a falha, e a classificação, que caracterizaria onde a falha ocorreria (AMORIM 2006 apud CARVALHO, 2005, CAMILO et al, 2009 e FURLAN 2018 apud GOLDSCHIMIDT, 2005).

Os níveis de FONTE DE DADOS, TRATAMENTO e STORAGE apresentados na figura 5, já são pré-existent na concepção da Indústria 4.0, sendo o trabalho da Engenharia de Manutenção apenas a implementação das fases de ETL (*Extract Transform Load*) e *Machine Learning*. Com os históricos de dados acumulados de monitoramento e falhas anteriores, ou baseados em dados de ensaios e comissionamento (FONSECA, 2014), a utilização de aprendizado supervisionado é amplamente recomendável e o mais simples (CADENA et al, 1999). Ainda cabe ressaltar que a reiteração do aprendizado supervisionado, considerando os históricos pós implementação da aplicação de prognósticos, irá tornar cada vez mais precisos os prognósticos emitidos posteriormente.

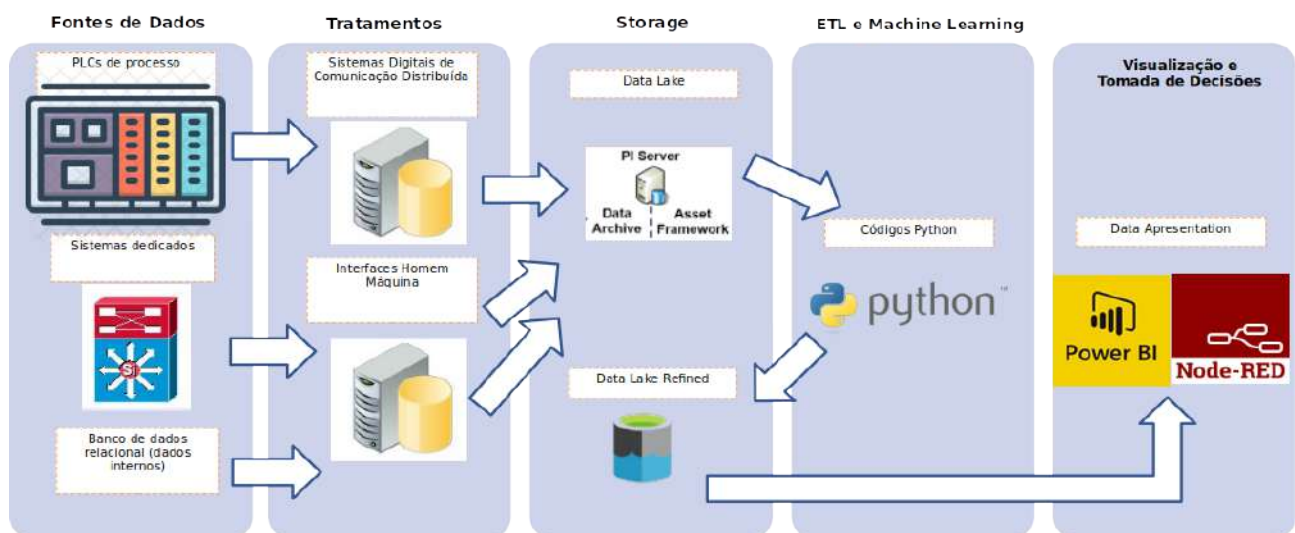


Figura 5 – Pipe Line proposto para implementação de Prognósticos de Manutenção

Fonte: Autor

7 ESTUDO DE CASO

Como forma de demonstração, foi selecionada uma base de dados pública, postada no site ‘Kaggle’, que se definia por ser uma base de dados não tratada de uma bomba de grande capacidade de água, contendo 52 sensores registrados em 153 dias, minuto a minuto, perfazendo 220.320 minutos sucessivos de monitoramento. Mesmo sendo uma base de dados pequena, a mesma foi escolhida por ser declarada de uma situação real e com um número satisfatório de elementos sensores atuantes, o que traria vantagens na demonstração dos conceitos apresentados. A desvantagem é a não classificação dos sensores e a descrição do sistema de forma detalhada, o que se apresentou como um desafio, pois não é possível se utilizar de conhecimento prévio técnico sobre a bomba de água e seus subsistemas.

8 APLICAÇÃO DE DATA ANALYTICS COM MACHINE LEARNING

No caso estudado, foi verificado que mesmo com a base de dados com valores não tratados, como por exemplo, não normalizado as casas decimais ou verificado valores outliers – que sem o conhecimento prévio da função de cada sensor torna-se extremamente difícil definir se é um erro ou uma medição real – é possível a utilização de métodos preditivos de machine learning. Ambos os métodos escolhidos – Árvore de Decisão e Regras de Indução – são baseados em aprendizado supervisionado, levando a divisão da base de dados em duas bases de igual tamanho (50% a 50%), sendo uma base de dados de treino e outra de teste, com o critério de 4

falhas estarem na base de treino, 3 falhas estarem na base de teste e cada base tendo uma minutagem de supervisão de 110.160 minutos, ou 76,5 dias, conforme a figura 6.

```
[ ] # Divisão dos grupos de treino e teste (meio a meio)
train_df = df.loc[df.index < "2018-06-09 10:40:00"]
test_df = df.loc[df.index >= "2018-06-09 10:40:00"]
X_train = train_df.drop(['machine_status', 'Operation'], axis = 1)
y_train = train_df.Operation
X_test = test_df.drop(['machine_status', 'Operation'], axis = 1)
y_test = test_df.Operation
```

Figura 6 – Divisão das bases de dados treino e teste

Fonte: Autor

Como método de Regras de Indução foi utilizado a função ‘LinearRegression’, ou Regressão Linear, da biblioteca ‘scikit-learn’, figura 7. Para avaliação do resultado, foi utilizada a função RMSE – ‘Root Mean Squared Error’, ou Raiz Média do Erro Quadrático, também da biblioteca ‘scikit-learn’. A função RMSE – ‘Root Mean Squared Error’ é uma métrica de avaliação amplamente utilizada e reconhecida para medir o desempenho de modelos de regressão. Ela é calculada tomando-se a raiz quadrada da média dos quadrados dos erros, onde o erro bruto é a diferença entre o valor previsto pelo modelo e o valor real.

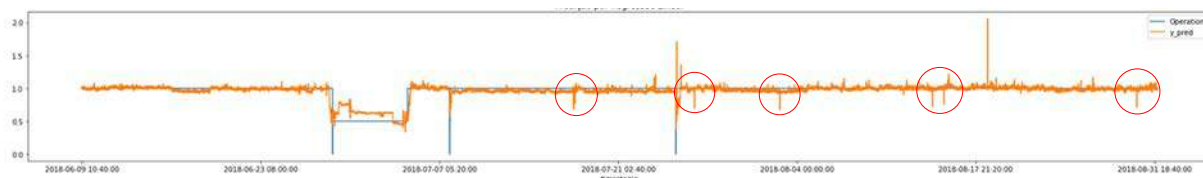


Figura 7 – Gráfico da indicação de falha por regressão linear da base de dados teste

Fonte: Autor

Com um resultado de RMSE de 0,059, é possível observar a boa resposta do método à previsão do estado do equipamento. Notável ainda, a existência de 9 pontos de ‘ruídos’, conforme indicados pelos círculos vermelhos na figura 6. Estes ‘ruídos’ apresentam grande possibilidade de sinalizarem problemas ocultos, que podem acarretar futuras falhas no equipamento.

Como método de Árvore de Decisão, foi utilizada a função ‘RandomForestRegressor’, ou Regressor por Floresta Aleatória, da biblioteca ‘scikit-learn’, figura 8. Para avaliação do resultado, foi utilizada a função RMSE, conforme anteriormente.

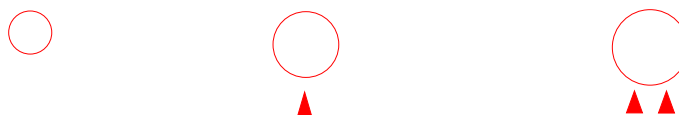


Figura 8 – Gráfico da indicação de falha por floresta aleatória da base de dados teste

Fonte: Autor

Com um resultado de RMSE de 0,037 e com uma resposta melhor que o método de Regras de Indução, também se demonstra uma boa resposta do método na previsão do estado do equipamento. Observa-se a presença de pontos de ‘ruído’ neste método, indicado pelos círculos vermelhos, contudo em menor quantidade, apenas 4 pontos de ‘ruídos’. Ressaltando a coincidência de 3 destes pontos de ‘ruídos’ com os pontos do método anterior, indicado pelas setas vermelhas.

9. ANTECIPAÇÃO DE FALHAS

Conforme Browlee (2020), é possível montar um modelo de base de dados com ‘janela deslocada’, em uma série temporal, de modo a se usar um método de treinamento supervisionado, dando a este método as indicações e os valores de forma antecipada na janela de tempo. Ou seja, ao replicar a base de dados e deslocar os valores das colunas referentes ao tempo de referência e ao estado de operação para cima ou para baixo, adiantando ou atrasando os valores das medições dos sensores no tempo e causando o adiantamento ou o retardo da análise de estado do equipamento, torna-se possível o treinamento da previsão da falha pelos métodos. Para este estudo, foi utilizado o algoritmo proposto por Browlee (2020), conforme figura 9:

```

0s ✓ ▶ for i in df.columns:
        if i == 'timestamp':
            continue
        else:
            for t in [2880, 5760]:
                df[f'{t}-{i}'] = df[i].shift(t)

```

Figura 9 – Algoritmo para ‘deslocamento de janelas’ replicantes

Fonte: Adaptação do Autor baseado em Browlee (2020)

O algoritmo proposto por Browlee (2020) é simples e funciona por meio da comparação das colunas da base de dados, com a condição de serem diferentes de 'timestamp', que é a base de tempo de referência. Satisfazendo a condição, esta coluna é replicada e deslocada em 180, 1440, 2880 e 5760 posições, sendo cada posição referente a 1 minuto do tempo de referência, através do comando 'pn.shift()' da biblioteca Pandas.

9.1 – Antecipação 2880 minutos (48 horas)

Utilizando a base de dados montada com o deslocamento de 2880 minutos. Os resultados foram semelhantes para a Regressão Linear da base de dados original no que se trata dos 'ruídos', conforme a figura 10. O RMSE apresentou resultado de 0,060.

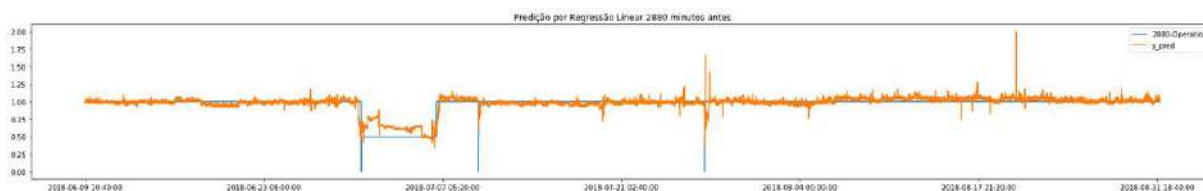


Figura 10 – Predição por regressão linear 2880 minutos antes da falha

Fonte: Autor

Para o método de Floresta Aleatória foi verificado que os 'ruídos' basicamente se mantiveram, conforme figura 11, também apresentando pequena melhora do RMSE de 0,035.



Figura 11 – Predição por floresta aleatória 2880 minutos antes da falha

Fonte: Autor

9.2 – Antecipação 5760 minutos (96 horas)

Utilizando a base de dados montada com o deslocamento de 5760 minutos, o equivalente a 96 horas ou 4 dias de antecipação, foi verificado que o resultado se assemelhou aos apresentados pela aplicação do método nas bases de dados original e deslocadas em 2880

minutos, conforme a figura 12, inclusive nos ‘ruídos’ apresentados. Contudo o resultado de RMSE foi o mais alto, sendo 0,070, mesmo assim atendendo a adequação do proposto.



Figura 12 – Predição por regressão linear 5760 minutos antes da falha

Fonte: Autor

Para o método de Floresta Aleatória foi verificado que os ‘ruídos’ basicamente se mantiveram, conforme figura 13, também apresentando pequena piora do RMSE de 0,035 se igualando ao resultado da aplicação do método para a base deslocada em 2880 minutos.

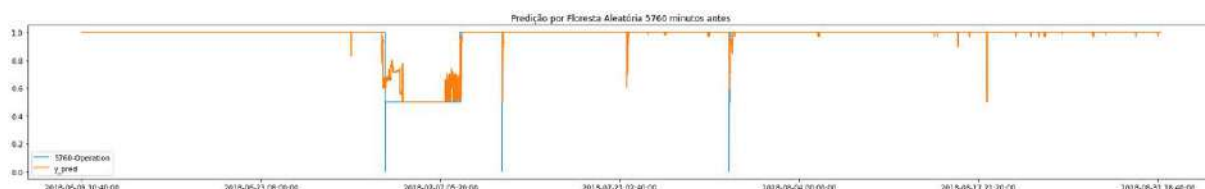


Figura 13 – Predição por floresta aleatória 5760 minutos antes da falha

Fonte: Autor

10 CONCLUSÃO

Finalizando a análise, podemos ver que a antecipação de falhas é eficaz e apresenta grande robustez, uma vez que os ensaios realizados utilizaram um base de dados real – sem o conhecimento prévio do projeto, da estrutura de processo e das grandezas mensuradas – onde, através de machine learning, foi realizado a mineração de dados obtendo-se resultados satisfatórios, comprovando as vantagens da arquitetura de gestão de dados, conforme quadro 3.

Quadro 3 – Tabela comparativa dos resultados nas bases de dados estudadas

Base de Dados	Regressão Linear RMSE	Floresta Aleatória RMSE
ORIGINAL	0,059	0,036
2880 Minutos Antecipada	0,060	0,035
5760 Minutos Antecipada	0,070	0,035

Fonte: Autor

Em todas as simulações os resultados foram adequados à predição da falha, com resultados abaixo de 0,1 de desvio, oferecendo oportunidade de melhor preparo de recursos e manobras operacionais para minimizar ou eliminar os impactos advindos desta falha.

Também, a presença dos 'ruídos' coincidentes, entre os resultados dos métodos aplicados, demonstra possíveis problemas ocultos que deverão ser tratados, sendo necessário nesta parte uma melhor descrição da estrutura de projeto do sistema da bomba de água e as funções dos sensores aplicados a mesma, possibilitando inclusive a etapa de classificação das faltas.

REFERÊNCIAS

AMORIM, Thiago; **Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de dados**; Trabalho de conclusão de curso de Ciências da Computação da Universidade Federal de Pernambuco; 2006.

BALDISSARELLI, Luciano.; FABRO, Elton; **Manutenção Preditiva na indústria 4.0**; Scientia Cum Industria, V. 7, N. 2, pp. 12-22, 2019.

BAHETI, Radhakisan; GILL, Helen; **Cyber-physical Systems**; The Impact of Control Technology, 2011.

<http://ieeecss.org/sites/ieeecss/files/2019-07/loCT-Part3-2CyberphysicalSystems.pdf>

BOENTE, Alfredo Nazareno P.; GOLDSCHMIDT, Ronaldo R.; ESTRELA, Vania Vieira; **Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados**; V SEGeT Simpósio de Excelência em Gestão e Tecnologia; 2008.

BROELEE, Jason; **Random Forest for Time Series Forecasting**; site: Machine Learning Mastery; 2020.

<https://machinelearningmastery.com/random-forest-for-time-series-forecasting/> (Acessado em 24.02.2023)

BROELEE, Jason; **How to Convert a Time Series to a Supervised Learning Problem in Python**; site: Machine Learning Mastery; 2020.

<https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/> (Acessado em 24.02.2023)

CABENA, Peter; Hyun CHOI, Hee; KIM, Il Soo; OTSUKA, Shuichi; REINSCHMIDT, Joerg; SAARENVIRTA, Gary; **Intelligent Miner for Data Applications Guide**; IBM International Technical Support Organization; 1999.

CAMILO, Cássio Oliveira; SILVA, João Carlos da; **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**; Technical Report - RT-INF_001-09 - Relatório Técnico - Agosto 2009; Instituto de Informática da Universidade Federal de Goiás, 2009.

DANTAS, Eric Rommel G. Dantas; PATRÍCIO JÚNIOR; José Carlos Almeida; LIMA, Daniel Silva de; AZEVEDO, Ryan Ribeiro de; **O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões**; V SEGeT Simpósio de Excelência em Gestão e Tecnologia; 2008.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P.; **The KDD Process for Extracting Useful Knowledge from Volumes of Data**; Communications of the ACM; 1996.

FONSECA, Marcos de Oliveira; **Aplicação de Técnicas de Data Analytics para Melhoria do Desempenho de Processos**; Contribuição técnica ao 18º Seminário de Automação e TI Industrial, 23 a 26 de setembro de 2014, São Paulo, SP, Brasil.

FURLAN, Matheus Batista; **Algoritmos e Técnicas para Mineração de Dados**; Trabalho de conclusão de curso de Ciências da Computação do Instituto Municipal de Ensino Superior de Assis; 2018.

LEE, Jay; BAGHERI, Behrad; KAO, Hung-An; **A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems**; Research Gate: Manufacturing Letter, 2015.

LEE, Jay; WU, Fangji; ZHAO, Wenyu; GHAFARI, Mahsa; **Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications**; Research Gate: Mechanical System and Signal Processing 42, 2014.

MACEDO, Letícia Costa; **Manutenção Preditiva no Contexto da Indústria 4.0: Um Modelo Preditivo em Uma Fábrica do Ramo Metalúrgico**; Monografia de Graduação em Bacharelado em Engenharia Metalúrgica, IFES; Vitória Espírito Santo; 2020.

MARCORIN, Wilson Roberto; LIMA, Carlos Roberto Camello; **Análise dos Custos de Manutenção e de Não-Manutenção de Equipamentos Produtivos**; Revista de Ciência & Tecnologia • v. 11, Nº 22 – pp. 35-42, 2013.

BASE DE DADOS: <https://www.kaggle.com/datasets/nphantawee/pump-sensor-data?select=sensor.csv> (Acessado em 11.01.2023).

QUONIAM, Luc; TARAPANOFF, Kira; ARAUJO JÚNIOR, Rogério Henrique de; ALVARES, Lillian; **Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil**; - Ci. Inf., Brasília, v. 30, n. 2; 2001.

RUNKLER, Thomas A.; **Data Analytics: Models and Algorithms for Intelligent Data Analysis**; Editora Springer Vieweg, 3º edition; 2020.

TROPIAS, Célio Eduardo Zacharias; SILVA, Pedro Paulo; DIAS, Ana Valéria Carneiro; **Indústria 4.0: Uma Caracterização do Sistema de Produção**; ALTEC 2017 – XVII Congresso Latino-Americano de Gestão de Tecnologia, Cidade do México, 2017.