

UNIVERSIDADE FEDERAL DO PARANÁ

WELLINTON DA SILVEIRA KIILLER

PREDIÇÃO DE VALORES EM UMA LINHA DE MONTAGEM DE INJETORES DIESEL:
APLICAÇÃO DE *Random Forest Regressor* PARA SELEÇÃO DE COMPONENTES

CURITIBA, PR
2023

WELLINTHON DA SILVEIRA KIILLER

PREDIÇÃO DE VALORES EM UMA LINHA DE MONTAGEM DE INJETORES DIESEL:
APLICAÇÃO DE *Random Forest Regressor* PARA SELEÇÃO DE COMPONENTES

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Especialista em Inteligência Artificial Aplicada no Programa de Pós-Graduação em Inteligência Artificial Aplicada, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná.

Orientador: Prof. Dr. Jaime Wojciechowski.

CURITIBA, PR
2023



MINISTÉRIO DA EDUCAÇÃO
SETOR DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CURSO DE PÓS-GRADUAÇÃO INTELIGÊNCIA ARTIFICIAL
APLICADA - 40001016348E1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação INTELIGÊNCIA ARTIFICIAL APLICADA da Universidade Federal do Paraná foram convocados para realizar a arguição da Monografia de Especialização de **WELLINTHON DA SILVEIRA KIILLER** intitulada: **Predição De Valores Em Uma Linha De Montagem De Injetores Diesel: Aplicação De Random Forest Regressor Para Seleção De Componentes**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de especialista está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 05 de Julho de 2023.

JAIME WOJCIECHOWSKI

Presidente da Banca Examinadora

RAZER ANTHOM NIZER ROJAS MONTAÑO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Predição de Valores em uma Linha de Montagem de Injetores Diesel: Aplicação de *Random Forest Regressor* para Seleção de Componentes

Wellinthon da Silveira Kiiller
Setor de Educação Profissional e Tecnológica
Universidade Federal do Paraná
Curitiba, Brasil
wellinthon.kiiller@hotmail.com

Prof. Dr. Jaime Wojciechowski
Setor de Educação Profissional e Tecnológica
Universidade Federal do Paraná
Curitiba, Brasil
jaimewo@ufpr.br

Resumo—A busca por soluções para problemas em ambientes produtivos tem crescido cada vez mais nas indústrias de fabricação [1]. Em linhas de montagem com alto volume de produção, por exemplo, a detecção antecipada de anomalias ou desvios acelera a tomada de decisão e pode reduzir falhas e evitar custos com o retrabalho de peças. A rastreabilidade torna possível identificar corretamente cada uma das peças e compreender em qual fase da montagem ocorreu uma falha. Neste contexto, este documento tem como objetivo analisar os dados gerados por estações de trabalho em uma linha de montagem de injetores diesel e desenvolver um modelo de *Machine Learning* (ML) capaz de prever os valores de espessura para a arruela de ajuste de pressão, um dos diversos componentes dos injetores. Como resultado, obteve-se um modelo com valor de $R^2=0.997$, $MAE=3.946$ microns, $RMSE=5.935$ microns e $MSE=35.223$ microns², que destacam a eficiência do modelo na explicação dos dados observados.

Palavras-chave—rastreabilidade, injetor diesel, linha de montagem, regressão

Resumo—The demand for solutions to problems in production environments has grown within the manufacturing industries [1]. In assembly lines with high production volume, early detection of anomalies or deviations speeds up decision making and can also reduce failures and avoid costs associated with reworking parts. Traceability makes it possible to identify each part correctly and understand at which assembly stage a failure occurs. In this context, the objective of this document is to analyze the data generated by the stations on a diesel injector assembly line and develop a Machine Learning (ML) model to predict thickness values of pressure adjustment washer, one of the various components of the injectors. As a result, a model with $R^2=0.997$, $MAE=3.946$ microns, $RMSE=5.935$ microns e $MSE=35.223$ microns² was obtained, highlighting the performance of the model in explaining the observed data.

Index Terms—traceability, diesel injector, assembly line, regression

I. DESENVOLVIMENTO

Com um mercado cada vez mais competitivo as indústrias têm procurado aumentar seus índices de produtividade. A eficiência nas linhas de montagem impacta consideravelmente estes números e a redução de falhas no processo é um aspecto importante durante as etapas de fabricação dos produtos [2].

As linhas de montagem são compostas por diversos processos sequenciais que variam em número e complexidade, de acordo com o ramo da indústria em que estão inseridas. Nas indústrias, estas linhas podem compreender técnicas de fixação, soldagem, conformação etc., que finalizam seu ciclo em uma estação para inspeção visual ou teste funcional.

A rastreabilidade é um conceito importante dentro das indústrias, diretamente relacionado aos tópicos de Indústria 4.0. Dentro das instituições, este conceito refere-se à prática de identificar um objeto ou componente e acessar todas as informações sobre ele, em qualquer fase do seu ciclo de vida [3]. Em uma linha de montagem de peças automotivas a rastreabilidade permite o acompanhamento completo do processo de construção do produto, desde o início da fase de montagem até o teste de funcionamento.

Em uma linha de montagem de produtos para motores diesel, o injetor é um dos diversos itens que compõem um sistema para injeção de combustível em alta pressão, chamado de *common rail*. Alguns componentes da estrutura de um injetor diesel estão descritos na Figura 1. Quando o combustível é submetido ao interior do corpo do injetor com pressão suficientemente alta para vencer a força da mola de pressão (6), a agulha movimenta-se, permitindo a saída de um *spray* de combustível através dos orifícios presentes no bico injetor (3). O curso da agulha é determinado por uma equação específica e pode ser ajustado de acordo com diferentes espessuras da arruela de ajuste de pressão (4), que funciona como uma espécie de calço.

Neste sentido, esse trabalho está concentrado em propor um modelo de regressão para prever valores de espessura da arruela de ajuste de pressão aplicado em uma linha de montagem de injetores diesel e, deste modo, reduzir a variação dos valores de curso de agulha nesse produto.

A. Descrição dos dados

A base de dados utilizada para construção do modelo de predição foi obtida dos servidores internos de uma empresa do ramo automotivo localizada na cidade de Curitiba, Paraná,

1. Porca de fixação do bico
2. Disco intermediário
3. Bico injetor
4. Arruela de ajuste de pressão
5. Canal de pressão
6. Mola de pressão
7. Pino de pressão

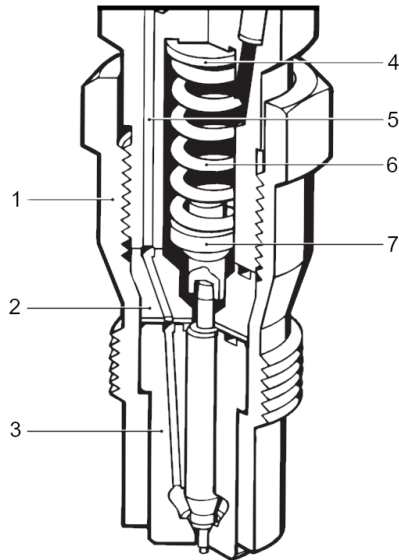


Figura 1. Elementos de um injetor para motores diesel. Adaptado de [4]

Brasil, onde os dados estão organizados de acordo com os conceitos de rastreabilidade do produto supracitados. Neste servidor todos os arquivos são armazenados no formato .DAT, uma extensão genérica que indica um arquivo de dados.

Para aplicação neste projeto foram coletados os dados de duas estações que fazem parte da linha de montagem dos injetores da empresa, sendo uma delas responsável pela seleção da arruela de ajuste de pressão para um determinado curso de agulha, baseado em um cálculo teórico de espessura. A outra estação é responsável por medir o curso da agulha e realizar uma verificação desse valor.

No total, o *dataset* contempla aproximadamente 923.000 peças produzidas no período de junho de 2022 a janeiro de 2023. Esses registros abrangem cerca de 158 modelos diferentes de produto e mais de 135 características específicas do processo.

Cada registro na base de dados contém um atributo identificador, juntamente com a data e hora de produção, para garantir a rastreabilidade das peças. Isso é de suma importância, especialmente porque serão tratados dados de duas estações diferentes e é crucial assegurar que os dados pertencem ao mesmo produto.

A regressão empregada neste trabalho considera a previsão de valores de espessura da arruela de ajuste de pressão como variável dependente e utiliza uma característica mecânica do injetor diesel e o valor de curso de agulha desejado em cada modelo de produto como variáveis independentes.

B. Métodos

Esta seção descreve os métodos aplicados para realizar a análise de dados dos injetores diesel. A análise está dividida cinco etapas: composição da base de dados; pré-processamento dos dados; preparação do treinamento; treinamento do modelo; e predição dos valores. A Figura 2 ilustra a sequência de passos realizada para obter o modelo de predição para diferentes espessuras da arruela de ajuste de pressão, enquanto a Figura 3 ilustra de forma mais detalhada os métodos empregados para pré-processamento e preparação dos dados. 843.940

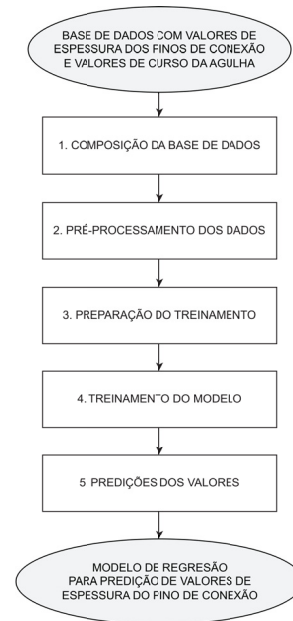


Figura 2. Fluxograma para obtenção do modelo de predição. Autoria Própria (2023)

1) Composição da base de dados: nesta etapa, os 923.000 registros que compõem o *dataset* coletado foram utilizados com o objetivo de criar um único conjunto com os atributos de ambas as estações. Esses registros foram submetidos ao pré-processamento, conforme descrito na seção a seguir.

2) Pré-processamento dos dados: nesta etapa os dados são preparados para que funcionem adequadamente quando forem aplicados ao modelo de predição. Com o objetivo de facilitar a visualização e compreensão neste trabalho, primeiramente os arquivos .DAT foram convertidos para o formato .CSV (*Comma Separated Values*).

No contexto deste projeto, para prever os valores de curso da agulha, primeiramente, foram removidas características que não estão relacionadas a esse valor, como o campo *Index*, uma vez que os índices de registro não são relevantes para a aplicação de uma regressão.

Em seguida, foram selecionados apenas os registros relacionados à peças aprovadas em ambas estações e removidos aqueles com duplicidade no atributo *Identifier*. Após essa etapa, os *datasets* das estações foram unidos em um único conjunto com todos os registros, utilizando o atributo *Identifier* para mesclá-los.

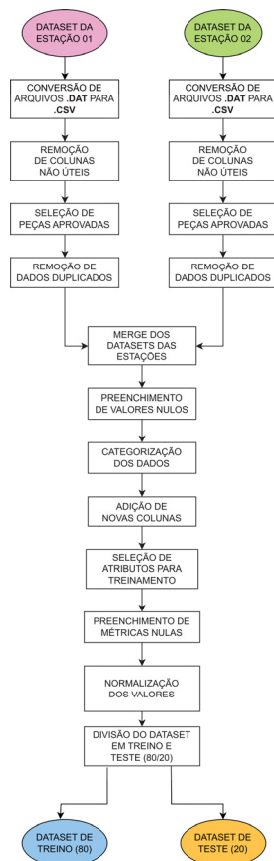


Figura 3. Etapas de pré-processamento e preparação dos dados. Autoria Própria (2023)

Na sequência realizou-se à busca por valores faltantes relacionados ao modelo de produto. Dentre as diversas técnicas disponíveis para o tratar esses valores, optou-se por preencher esses registros com 0 (zero). Os registros também foram categorizados da seguinte forma: uma categoria única para os modelos de produto que possuam menos de 400 amostras e categorias individuais para aqueles que com mais de 400 amostras.

Em relação à novas características, três novas colunas foram adicionadas à base da dados, contendo a média, a mediana e o desvio padrão dos valores de curso da agulha para as últimas 20 peças produzidas nas estações. Após todas estas etapas, os atributos descritos na Tabela I estavam disponíveis.

3) Preparação do treinamento: nesta etapa são escolhidos são escolhidos os atributos para treinamento do modelo e método para treinamento da base de dados pré-processada.

O primeiro passo é analisar se a tarefa a ser executada envolve classificação ou regressão. Quando os valores de saída são discretos tem-se uma tarefa de classificação, e quando os valores de saída são contínuos temos uma tarefa de regressão [5].

Uma vez que este trabalho tem o objetivo de desenvolver um modelo de *Machine Learning* para predição de valores de espessura da arruela de ajuste de pressão, tem-se uma tarefa

de regressão e, para isso, a Equação 1 serviu como base para a definição da Equação 2, que foi utilizada como alvo no treinamento do modelo. Entre os métodos disponíveis para modelos de regressão, escolheu-se o *Random Forest Regressor*, que é baseado em árvores de decisão [6] [7].

$$E = A - C \quad (1)$$

Onde: E = espessura da arruela de ajuste de pressão

A = distância entre válvula e pistão

C = curso da agulha

$$E_{ideal} = E_{atual} + (C_{medio} - C_{atual}) \quad (2)$$

Onde: E_{ideal} = espessura ideal da arruela de ajuste de pressão

E_{atual} = atual espessura da arruela de ajuste de pressão

C_{medio} = curso da agulha médio das últimas 20 peças

C_{atual} = último valor de curso da agulha medido

Para o treinamento do modelo foram selecionados apenas os atributos que influem na determinação da arruela de ajuste de pressão, conforme destacado na Tabela II:

Após a seleção desses atributos, os registros que possuam valores nulos para as métricas (média, mediana e desvio padrão) foram substituídos por 0 (zero), para garantir que estes não iriam interferir no treinamento do modelo.

Árvores de decisão são considerados os melhores algoritmos de aprendizagem de máquina supervisionada [8]. O método das árvores de decisão geralmente é o primeiro método a ser submetido a testes em muitos setores da indústria e do comércio, pois destaca-se por sua facilidade de compreensão, escalabilidade para grandes volumes de dados e versatilidade no tratamento de valores discretos e contínuos [5].

Diferente dos modelos lineares, os modelos gerados a partir de árvores de decisão conseguem se adaptar muito bem à cenários não lineares e, deste modo, são amplamente usados em problemas de classificação ou regressão [8].

Uma árvore de decisão representa um método de aprendizado capaz de retornar um valor de saída único a partir de alguns atributos de entrada. A árvore alcança a saída (decisão) realizando diversos testes sequenciais, que vão da raiz até o ramo final da árvore [5].

Um ponto negativo das árvores de decisão refere-se à sua instabilidade, uma vez que adicionar um novo exemplo ou conjunto de dados na raiz altera a árvore inteira [5]. Com o objetivo de corrigir esse e outros problemas, surgiu o método chamado de *Random Forest* [6].

Para normalização dos valores que compõem o *dataset*, foi aplicado o método *MinMaxScaler*, que consiste em atribuir novos valores para os registros de acordo com o intervalo desejado [9], conforme demonstra a Equação 3. Geralmente o valor de intervalo mais comum e padrão para o *MinMaxScaler* é entre 0 (zero) e 1 (um). No entanto, dessa forma, os valores médios teriam números próximos a 0.5. Neste trabalho, optou-

Tabela I
ATRIBUTOS DA BASE DE DADOS

Atributo	Descrição
<i>DateTime_Station01</i>	data e hora de produção na estação 01
<i>TypeNo_Station01</i>	modelo de produto na estação 01
<i>StationNo_Station01</i>	número da estação responsável pela seleção da arruela de ajuste de pressão
<i>Identifier</i>	identificador exclusivo da peça produzida
<i>Result_Station01</i>	resultado da operação na estação 01
<i>ResultDistanceValvePiston – Value</i>	distância entre válvula e pistão
<i>ResultDistanceValvePiston – Result</i>	resultado da inspeção de distância entre válvula e pistão
<i>ResultDistanceValvePiston – LoLim</i>	limite inferior da distância entre válvula e pistão
<i>ResultDistanceValvePiston – UpLim</i>	limite superior da distância entre válvula e pistão
<i>ResultDGVDimensionCalcVal – Value</i>	curso da agulha calculado
<i>DateTime_Station02</i>	data e hora de produção na estação 02
<i>StationNo_Station02</i>	número da estação responsável pela medição do curso da agulha
<i>Result_Station02</i>	resultado da operação na estação 02
<i>ResultHeid2 – Value</i>	curso da agulha medido
<i>ResultHeid2 – Result</i>	resultado da medição de curso da agulha
<i>ResultHeid2 – LoLim</i>	limite inferior do curso da agulha
<i>ResultHeid2 – UpLim</i>	limite superior do curso da agulha
<i>result_heid_2_med</i>	média do curso da agulha dos últimos 20 registros
<i>diff_values_Station02</i>	diferença entre o valor de curso da agulha calculado x medido
<i>Target</i>	valor alvo para regressão
<i>mean_diff</i>	média da diferença entre os últimos 20 valores de curso da agulha
<i>std_diff</i>	desvio padrão da diferença entre os últimos 20 valores de curso da agulha
<i>median_diff</i>	mediana da diferença entre os últimos 20 valores de curso da agulha

Tabela II
ATRIBUTOS UTILIZADOS NO TREINAMENTO DO MODELO

Atributo	Descrição
<i>TypeNo_Station01</i>	modelo de produto na estação 01
<i>ResultDistanceValvePiston – Value</i>	distância entre válvula e pistão
<i>mean_diff</i>	média da diferença entre os últimos 20 valores de curso da agulha
<i>std_diff</i>	desvio padrão da diferença entre os últimos 20 valores de curso da agulha
<i>median_diff</i>	mediana da diferença entre os últimos 20 valores de curso da agulha

se por normalizar os valores entre -1 e 1, para que os valores médios fossem próximos a zero.

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Onde: x'_i = valor normalizado

x_i = valor original

x_{min} = valor mínimo

x_{max} = valor máximo

Então, após as etapas de pré-processamento dos dados e da escolha do algoritmo a ser utilizado para regressão é necessário separar o *dataset* em um conjunto de treino e outro de teste. Geralmente, os valores de proporção mais comuns para divisão dos conjuntos são 60/40, 70/30 ou 80/20 [10]. Para este trabalho

o *dataset* foi dividido seguindo a proporção 80/20, ou seja, 80% dos dados foram usados no treinamento do modelo e os 20% restantes foram submetidos aos testes.

Posterior a divisão do *dataset*, foi realizada a normalização dos valores entre -1 e 1 utilizando o *MinMax Scaler* e, em seguida, foi aplicada uma validação cruzada (*cross-validation*).

Conforme ilustra a Figura 4, na validação cruzada *k-fold* um conjunto de dados é dividido em subgrupos de tamanho *k*, cujos itens de cada grupo são selecionados aleatoriamente. Posterior a isso, o algoritmo separa um destes grupos ($1/k$) para testar o desempenho do modelo de aprendizagem e realiza o treinamento com os demais grupos, repetindo o processo até que tenham sido percorridos todos os subgrupos ($k-1/k$) [11] [12]. Os valores de *k* utilizados com maior frequência são 5 e 10 e a qualidade do modelo é estimada por meio da média dos resultados obtidos em cada um dos processos [11]. Neste trabalho foi utilizada a técnica *k-fold* com valor de $k=5$.

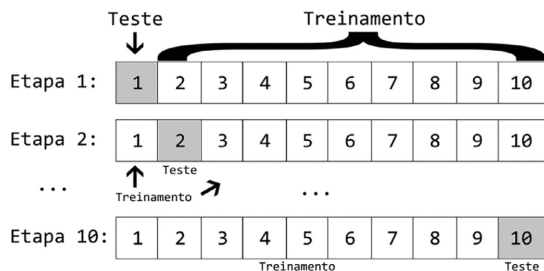


Figura 4. Exemplo de validação cruzada para valor de $k=10$. Adaptado de [11]

Na maioria dos casos uma única hipótese do modelo não é suficiente para se obter uma acurácia boa e, então, é necessário criar outras hipóteses e comparar os modelos. Nesse sentido, os modelos podem conter valores distintos para seus parâmetros, por sua vez chamados de "hiperparâmetros" [5]. Os hiperparâmetros podem ser ajustados de forma manual ou automática, conforme apresentado a seguir:

1) *Ajuste Manual de Hiperparâmetros*: O ajuste manual de hiperparâmetros refere-se à seleção de alguns valores de acordo com a experiência prévia de quem está treinando o modelo que, após algumas rodadas de treino e teste pode fazer novas alterações nos valores dos parâmetros até que se obtenha um desempenho satisfatório [5].

2) *Ajuste Automático de Hiperparâmetros*: Quando existe uma quantidade maior de hiperparâmetros ou uma larga faixa de valores para ajuste deles é recomendado utilizar uma busca automática, em grade ou aleatória. Estas buscas têm o objetivo de testar todas as combinações possíveis de valores e checar qual delas produz o melhor resultado [5]. Ambas as buscas demandam de maior custo computacional.

Para este projeto escolheu-se empregar o ajuste automático de hiperparâmetros por meio do *Randomized Search CV*, no qual destaca-se o uso dos parâmetros $n_jobs=-1$ para uso de todos os processadores do computador e $scoring=R^2$ como métrica de avaliação de performance do modelo.

4) *Treinamento do modelo*: conforme mencionado, o modelo utilizado para predição dos valores foi o *Random Forest Regressor*. Foram utilizados os hiperparâmetros destacados abaixo, e os valores padrões, bem como a faixa de valores para busca com o *Randomized Search CV* são apresentados na Tabela III.

- $n_estimators$: define o número de árvores na floresta ($default=100$);
- max_depth : define a profundidade máxima da árvore ($default=None$);
- $min_samples_split$: define o número mínimo de amostras para criar se criar um novo nó ($default=2$); e
- $min_samples_leaf$: define o número mínimo de amostras para ser um nó folha ($default=1$)

Em uma análise preliminar, com o objetivo de reduzir o tempo de treinamento do modelo foi utilizada uma faixa de valores com números baixos para uma primeira busca de hiperparâmetros com o *RandomizedSearchCV*. Os resultados

Tabela III
HIPERPARÂMETROS UTILIZADOS NA BUSCA COM O *RandomizedSearchCV*

Hiperparâmetro	Valor Padrão	Faixa de Valores
$n_estimators$:	100	entre 5 e 100
max_depth	<i>None</i>	entre 5 e 100
$min_samples_split$	2	entre 2 e 100
$min_samples_leaf$	1	entre 2 e 100

obtidos foram satisfatórios e, por esse motivo, a faixa de valores não foi ampliada.

5) *Predições dos valores*: nesta etapa foram feitas as predições dos valores para espessura da arruela de ajuste de pressão dos injetores, baseando-se no modelo treinado a partir da base de dados mencionada na Seção B.1.

C. Tecnologias

Para se atingir o objetivo citado neste documento, foi desenvolvido um *script* Python em um computador pessoal com: processador Intel Core i7-10850H de 6 núcleos a 2,70GHz; placa de vídeo NVIDIA Quadro RTX 3000 com 6GB de memória RAM dedicados; e sistema operacional Windows 10. Foram utilizadas as seguintes aplicações e bibliotecas:

- Python, versão 3.8.5 [13]
- Visual Studio Code, versão 1.78.2 [14]
- Pacote *pandas*, versão 1.5.3 [15]
- Pacote *numpy*, versão 1.23.5 [16]
- Pacote *scikit-learn*, versão 1.2.2 [17]
- Pacote *scipy*, versão 1.10.0 [18]

II. RESULTADOS E DISCUSSÕES

Esta seção apresenta numericamente e graficamente os resultados obtidos com os experimentos realizados.

As métricas utilizadas para avaliação de performance foram: *Mean Absolute Error* (MAE) - Erro Médio Absoluto, *Mean Squared Error* (MSE) - Erro Médio Quadrático, *Root Mean Squared Error* (RMSE) - Raiz do Erro Médio Quadrático e R^2 , também conhecido como Coeficiente de Determinação.

O coeficiente de determinação R^2 representa como um modelo de regressão se comporta na predição dos dados com relação aos dados observados, variando entre 0 (zero) e 1 (um). Quanto mais próximo de um melhor é a capacidade de predição do modelo em relação aos dados observados. Matematicamente o Coeficiente de Determinação é expresso de acordo com a Equação 4.

O Erro Médio Quadrático (MSE) calcula a média dos quadrados da diferença entre os valores reais e preditos [8]. Quanto maior é o valor de MSE, pior é a performance do modelo. Entretanto, como os valores obtidos são elevados ao quadrado, existe uma distorção da unidade de medida e isso pode provocar erros de interpretação. Com o objetivo de manter a unidade de medida original e facilitar a interpretação dos valores surgiu a Raiz do Erro Médio Quadrático (RMSE), que consiste na aplicação da raiz quadrática no cálculo do MSE.

Matematicamente, o MSE e o RMSE são escritos conforme as Equações 5 e 6.

Por outro lado, o Erro Médio Absoluto (MAE) fornece uma média da diferença entre o valor real e predito. A Equação 7 descreve o MAE.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

Onde: n = número de observações

y_i = valor real

\hat{y}_i = valor predito

Os melhores valores para hiperparâmetros encontrados por meio da busca utilizando o *RandomizedSearchCV* e a faixa de valores citada na Tabela III estão destacados na Tabela IV.

Tabela IV

MELHORES HIPERPARÂMETROS ENCONTRADOS COM *RandomizedSearchCV*

Hiperparâmetro	Melhor Modelo
<i>n_estimators</i> :	50
<i>max_depth</i>	28
<i>min_samples_split</i>	12
<i>min_samples_leaf</i>	2

Após o treinamento do modelo de regressão, os registros contidos no *dataset* de teste foram submetidos ao modelo para avaliação de desempenho, de acordo com as métricas supracitadas. Os resultados obtidos para estas métricas estão descritos na Tabela V e, para este trabalho, ressalta-se que o coeficiente de determinação R^2 foi a métrica escolhida para seleção do melhor conjunto de hiperparâmetros com o uso do *RandomizedSearchCV*.

Tabela V

RESULTADOS DA AVALIAÇÃO DO MODELO

Métrica	Resultado
<i>Mean Absolute Error (MAE)</i>	3.946
<i>Mean Squared Error (MSE)</i>	35.223
<i>Root Mean Squared Error (RMSE)</i>	5.935
R^2 (<i>Coefficiente de Determinação</i>)	0.997

Neste trabalho, fazendo-se o uso do modelo de regressão com os parâmetros apresentados na Tabela III foi possível obter

um valor de $R^2=0.997$, ou seja, uma excelente performance de predição do modelo. Isto indica que o modelo de regressão é bastante preciso na explicação da variação nos dados observados.

Para as métricas MAE e RMSE, foram obtidos os valores de 3.946 microns e 5.935 microns, respectivamente. Já para o MSE foi obtido o valor de 35.223 microns² que, conforme explicado, distorce a unidade de medida. Esses resultados revelam que a predição da espessura da arruela de ajuste de pressão obteve um erro médio significativamente baixo, destacando a eficiência do modelo desenvolvido no trabalho em questão.

Graficamente, os resultados obtidos a partir dos experimentos realizados são apresentados nas Figuras 5, 6 e 7. As Figuras ilustram os valores de espessura da arruela de ajuste de pressão calculados na linha de montagem para diferentes modelos de produto, em comparação com os valores preditos pelo modelo. Pode-se observar uma redução na variação dos valores, aproximando-os da média e alcançando o objetivo proposto por este trabalho.

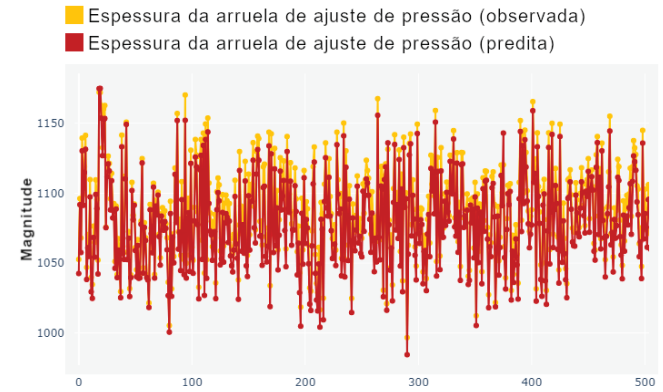


Figura 5. Espessuras da arruela de ajuste de pressão E observada x predita. Autoria Própria (2023)

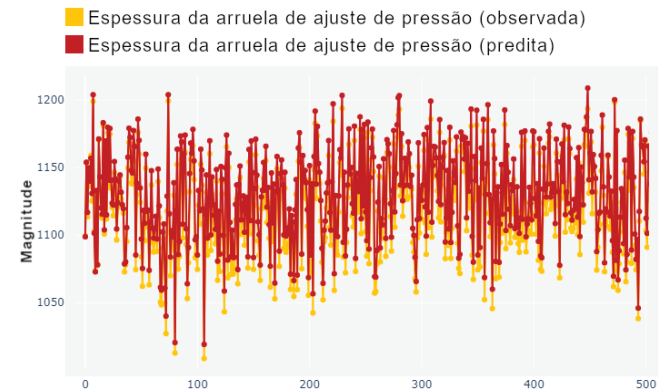


Figura 6. Espessuras da arruela de ajuste de pressão E observada x predita. Autoria Própria (2023)

Em conclusão, este trabalho abordou a importância da busca por soluções em ambientes produtivos, especialmente nas indústrias de fabricação. Na indústria automotiva na qual o trabalho foi aplicado, notou-se que a detecção antecipada dos

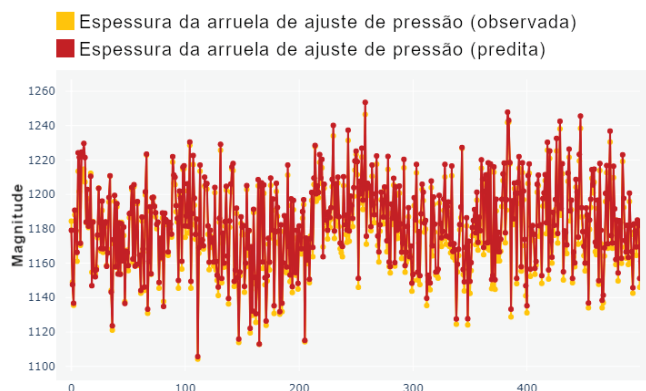


Figura 7. Espessuras da arruela de ajuste de pressão E observada x predita. Autoria Própria (2023)

desvios para os valores de espessura da arruela de ajuste de pressão nas linhas de montagem da empresa é crucial para acelerar a tomada de decisão e reduzir falhas e retrabalhos. Com o modelo de *Machine Learning* obtido a partir da análise dos dados realizada, foi possível prever valores de espessura para a arruela de ajuste de pressão mais adequados a cada um dos injetores.

A aplicação deste modelo permitiu destacar valiosos pontos para o aprimoramento contínuo do processo de montagem dos injetores diesel. Ao identificar corretamente cada peça e prever valores para que os funcionários agissem de maneira rápida foi possível otimizar a eficiência e qualidade do processo de fabricação, resultando em benefícios significativos para a indústria.

Quando uma peça é reprovada no teste funcional, é necessário desmontá-la para reaproveitar ou descartar os componentes. Essa tarefa consome tempo e resulta no desperdício de componentes que poderiam não ter sido utilizados em peças que apresentariam resultados negativos. Portanto, este projeto contribui significativamente para a redução de peças reprovadas e, conseqüentemente, para a redução de retrabalhos na linha de montagem.

III. TRABALHOS FUTUROS

A partir das discussões e resultados alcançados com este trabalho, fica evidente a importância e o potencial de aplicações como esta em linhas de montagem presentes na indústria.

Para validar a performance do modelo na linha de montagem são necessários testes no modo *shadow mode*. Neste modo o modelo realiza predições para os dados, mas não as usa para tomada de decisões e, assim, os resultados do modelo não interferem no fluxo de trabalho da linha de montagem [19].

Para trabalhos futuros sugere-se:

- Preparar o ambiente fabril para aplicação real;
- Correlacionar as predições do modelo com os resultados obtidos na linha de testes funcionais, de forma a agir nas estações anteriores;
- Balancear o *dataset* para verificar a influência no treinamento do modelo;

- Realizar o treinamento do modelo com novas faixas de valores para busca com o *RandomizedSearchCV*; e
- Realizar o treinamento do modelo utilizando outras técnicas de aprendizado como, por exemplo, Redes Neurais Artificiais e *Support Vector Machines* (SVM).

REFERÊNCIAS

- [1] M. Meiners, A. Mayr, and J. Franke. "Process curve analysis with machine learning on the example of screw fastening and press-in processes," *Procedia CIRP*, vol. 97, 2021, pp.166-171. DOI:10.1016/j.procir.2020.05.220.
- [2] D. Ribeiro, L. M. Matos, P. Cortez, G. Moreira and A. Pilastrri. "A comparison of anomaly detection methods for industrial screw tightening," 2021 21st International Conference on Computational Science and Its Applications (ICCSA), vol. 12950, 2021, pp.485-500. DOI: 10.1007/978-3-030-86960-1_34
- [3] R. Schuitemaker and X. Xu. "Product traceability in manufacturing: A technical review," 2020 53rd CIRP Conference on Manufacturing Systems, vol. 93, 2020, pp.700-705. DOI: 10.1016/j.procir.2020.04.078.
- [4] R. Bosch. "Manual de tecnologia automotiva," São Paulo: Blucher, 2005. 2005. ISBN 978-85-212-0378-0.
- [5] S. J. Russell and P. Norvig. "Inteligência Artificial: uma abordagem moderna," Grupo Editorial Nacional S.A.: Rio de Janeiro, 2022 - 4 ed. Publicado pelo selo LTC. 2022. ISBN 9788595159495.
- [6] L. Breiman. "Random Forests," 2001, pp.45, 5-32. DOI:10.1023/A:1010933404324.
- [7] J. Grus. "Data Science do Zero: Noções Fundamentais com Python," Rio de Janeiro: Alta Books, 2021. 2021. ISBN 978-8-550-81646-3.
- [8] A. Netto. "Python para Data Science e Machine Learning: Descomplicado," Rio de Janeiro: Alta Books, 2021. 2021. ISBN 978-65-5520-317-2.
- [9] L. B. V. Amorim, G. D. C. Cavalcanti and R. M. O. Cruz. "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, 2023. DOI:10.1016/j.asoc.2022.109924.
- [10] S. Raschka and V. Mirjalili. "Python Machine Learning," Birmingham: Packt Publishing Ltd., 2019. 2019. ISBN 978-1-78995-575-0.
- [11] D. C. B. Mariano, L. T. Marques and M. S. Silva. "Data Mining," Porto Alegre: SAGAH, 2020. 2020. ISBN 978-65-5690-029-2.
- [12] T. Hastie, R. Tibshirani and J. Friedman. "The Elements of Statistical Learning: Data Mining, Inference and Prediction," New York: Springer, 2009. 2009. ISBN 978-03-8784-857-0.
- [13] Python. "Python 3.8.5", 2023. Disponível em: <<https://www.python.org/downloads/release/python-385/>>. Acesso em 11 jun. 2023.
- [14] Visual Studio Code. "Documentation for Visual Studio Code", 2023. Disponível em: <https://code.visualstudio.com/updates/v1_78/>. Acesso em: 11 jun. 2023.
- [15] Pandas. "Pandas documentation - pandas 1.5.3 documentation", 2023. Disponível em: <<https://pandas.pydata.org/pandas-docs/version/1.5.3/index.html/>>. Acesso em: 11 jun. 2023.
- [16] NumPy. "NumPy documentation - NumPy v1.25 Manual", 2023. Disponível em:
- [17] Scikit-learn. "Scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation", 2023. Disponível em: <<https://scikit-learn.org/1.2/>>. Acesso em: 11 jun. 2023.
- [18] SciPy. "SciPy documentation — SciPy v1.10.0 Manual", 2023. Disponível em: <<https://docs.scipy.org/doc/scipy-1.10.0/index.html/>>. Acesso em: 11 jun. 2023.
- [19] B. Mikulski. "Shadow deployment vs. canary release of machine learning models," 2022. Disponível em: <<https://www.qwak.com/post/shadow-deployment-vs-canary-release-of-machine-learning-models/>>. Acesso em: 11 jul. 2023.