

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 14/05/2024.



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Faculdade de Engenharia e Ciências de Guaratinguetá

JOÃO OTÁVIO BELIZÁRIO TONHÃO

**Sistema Em Python Para Monitoramento De Processos De Laminação Utilizando
Machine Learning**

Guaratinguetá
2023

João Otávio Belizário Tonhão

**Sistema Em Python Para Monitoramento De Processos De Laminação Utilizando
Machine Learning**

Dissertação apresentada à Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, para obtenção do título de Mestre em Engenharia, área de Produção.

Orientadora: Profa. Dra. Paloma Maria Silva Rocha Rizol

Guaratinguetá
2023

T665s	Tonh�o, Jo�o Ot�vio Beliz�rio Sistema em Python para monitoramento de processos de lamina��o utilizando machine learning / Jo�o Ot�vio Beliz�rio Tonh�o - Guaratinguet�, 2023. 95 f : il. Bibliografia: f. 92-95 Disserta��o (Mestrado) – Universidade Estadual Paulista, Faculdade de Engenharia e Ci�ncias de Guaratinguet�, 2023. Orientadora: Prof�. Dr�. Paloma Maria Silva R. Rizol 1. Python (Linguagem de programa��o de computador). 2. Processo decis�rio. 3. Controle de processo. I. T�tulo. CDU 658.511.3(043)
-------	---

Luciana M ximo
Bibliotec ria/CRB-8 3595

IMPACTO POTENCIAL DESTA PESQUISA

Esse trabalho possui um impacto estratégico para a sustentabilidade do setor industrial de laminados no Brasil diante da concorrência externa, uma vez que a otimização de processos e aumentando a eficiência, além de garantir o volume de produção, impacta diretamente em redução de custos de produção, capacidade de produção, emissão de poluentes, consumo de energia, consumo de matéria prima, entre outros.

POTENTIAL IMPACT OF THIS RESEARCH

This work has a strategic impact for the sustainability of the rolling products industrial sector in Brazil in the face of external competition, since optimizing processes and increase material recovery, besides to guarantee the production capacity, has a direct impact on reducing production costs, emissions pollutants, energy consumption, raw material consumption, among others.



JOÃO OTÁVIO BELIZÁRIO TONHÃO

**ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO TÍTULO DE
"MESTRE EM ENGENHARIA"**

**PROGRAMA: ENGENHARIA
CURSO: MESTRADO**

APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO

 Documento assinado digitalmente
MANOEL CLEBER DE SAMPAIO ALVES
Data: 02/01/2024 09:00:18-0300
Verifique em <https://validar.jf.gov.br>

Prof. Dr. Manoel Cleber de Sampaio Alves
Coordenador

BANCA EXAMINADORA:

 Documento assinado digitalmente
PALOMA MARIA SILVA ROCHA RIZOL
Data: 04/12/2023 11:17:40-0300
Verifique em <https://validar.jf.gov.br>

Prof. Dr. PALOMA MARIA SILVA ROCHA RIZOL
Orientador - UNESP

 Documento assinado digitalmente
MARCELA APARECIDA GUERREIRO MACHADO DE
Data: 11/12/2023 07:53:40-0300
Verifique em <https://validar.jf.gov.br>

**Profa. Dra. MARCELA APARECIDA GUERREIRO MACHADO DE
FREITAS**
UNESP

 Documento assinado digitalmente
FABIO RODOLFO MIGUEL BATISTA
Data: 06/12/2023 13:38:00-0300
Verifique em <https://validar.jf.gov.br>

Prof. Dr. FABIO RODOLFO MIGUEL BATISTA
USP

NOVEMBRO de 2023

DADOS CURRICULARES

JOÃO OTÁVIO BELIZÁRIO TONHÃO

NASCIMENTO 03.02.1991 – Andradas / MG

FILIAÇÃO João Batista Tonhão
 Maria Aparecida Belizário Tonhão

2010/2014 Curso de Graduação
Engenharia de Controle e Automação – Universidade Federal de Ouro Preto - UFOP

Dedico este trabalho de modo especial, à minha família e esposa que sempre me apoiou e me incentivou para que essa conquista fosse alcançada.

AGRADECIMENTOS

Agradeço primeiramente a Deus por me dar condições para buscar a realização dos meus objetivos;

a minha esposa Paula pelo companheirismo e porto seguro;

aos meus Pais pelo suporte, exemplo e orientação;

e a professora Paloma Rizol pelo apoio e orientação ao longo deste trabalho.

RESUMO

Com o pleno avanço da Indústria 4.0 nos mais variados setores industriais, desponta como uma das principais ferramentas a utilização de técnicas de *Machine Learning*. À medida que os processos ficam mais rápidos e assertivos, aumenta a necessidade de maiores aquisições de dados para monitoramento, classificação e aperfeiçoamento do produto. Este alto volume de dados torna inviável a aplicação de análises manuais por seres humanos e destaca-se assim algoritmos com técnicas capazes de trabalhar com diferentes bases de dados simultaneamente. As técnicas de *Machine Learning*, como Árvore de Decisão, são altamente recomendáveis para sistemas de classificação de produto. É possível, por meio de critérios pré-estabelecidos ou por regressão, analisar conjuntos de variáveis pré-definidas de acordo com padrões desejáveis e auxiliar na tomada de decisão por meio das suas saídas. Neste contexto, a proposta deste trabalho é desenvolver um algoritmo em Python utilizando a técnica de árvore de decisão para o controle de qualidade no processo de laminação, comparando sua eficiência frente aos sistemas utilizados atualmente para este mercado. A utilização da linguagem Python permitiu a manipulação de grande quantidade de dados por meio de ferramentas estatísticas, geração de gráficos, simulação e supervisorio. Com isso, foi criado um sistema de monitoramento de processos de laminação baseado na técnica de *Machine Learning* Árvore de Decisão Classificatória, possibilitando o cálculo das regiões a serem descartadas no processo e o correto direcionamento do produto de acordo com as especificações dos clientes, optando por seguir o processo original, redirecionar para outro cliente ou sucatear o produto caso não haja a possibilidade de aproveitamento. O modelo apresentou uma acurácia de 83,33% e profundidade de 9 com ccp_alpha de 0,012. Resultando em um modelo otimizado com precisão de 83,58%, recall de 83,33% e f-score de 82,07%, onde concluiu-se ser possível construir um sistema eficiente para aplicação na indústria usando software de código aberto com baixo custo de implementação e manutenção, incluindo uma alta flexibilidade para fazer adaptações para cada demanda de processo.

PALAVRAS-CHAVE: Aprendizado de Máquina; Tomada de decisão; Python; Laminação; Controle de Processos.

ABSTRACT

With the full advancement of Industry 4.0 in the most varied industrial sectors, the use of Machine Learning techniques emerges as one of the main tools. As processes become faster and more assertive, the need for greater data acquisition for monitoring, classification and product improvement increases. This high volume of data makes the application of manual analyzes by humans unfeasible and thus highlights algorithms with techniques capable of working with different databases simultaneously. Machine Learning techniques, such as Decision Tree, are highly recommended for product classification systems. It is possible, through pre-established criteria or regression, to analyze sets of pre-defined variables according to desirable standards and assist in decision-making through their outputs. In this context, the purpose of this work is to develop an algorithm in Python using the decision tree technique for quality control in the lamination process, comparing its efficiency compared to the systems currently used for this market. The use of the Python language allowed the manipulation of large amounts of data through statistical tools, graph generation, simulation and supervision. With this, a lamination process monitoring system was created based on the Machine Learning Classification Decision Tree technique, enabling the calculation of the regions to be discarded in the process and the correct direction of the product according to customer specifications, opting for follow the original process, redirect to another customer or scrap the product if it cannot be used. The system presented an accuracy of 83.3%, where it was concluded that it was possible to build an efficient system for application in industry using open-source software with low implementation and maintenance costs, including high flexibility to make adaptations to each process demand.

KEYWORDS: Machine learning; Python; decision tree; rolling; quality control.

LISTA DE FIGURAS

Figura 1 - Classificação da pesquisa científica.....	20
Figura 2 - Número de publicações ao longo dos últimos anos com as palavras-chave <i>quality control</i> e <i>machine learning</i>	23
Figura 3 - Publicações por Área de Atuação	23
Figura 4 - Países Coautores	24
Figura 5 - Coocorrência de palavras-chave	25
Figura 6 - Processo de vazamento de placas	32
Figura 7 - Processo de lingotamento	33
Figura 8 - Diagrama de laminador "4-High"	33
Figura 9 - Fluxo de dados	35
Figura 10 - Fluxograma do software	36
Figura 11 - Fluxograma da pesquisa	37
Figura 12 - Demonstração IbaAnalyzer.....	40
Figura 13 - Demonstração IbadatCoordinator	40
Figura 14 - Demonstração de aplicação do PostgreSQL	41
Figura 15 - Interface Software Pycharm.....	42
Figura 16 - Exemplo de análise de descarte por espessura.....	43
Figura 17 - Exemplo de análise de descarte por temperatura.....	43
Figura 18 - início da gravação do arquivo	44
Figura 19 - Representação do evento StopTrigger	45
Figura 20 - tarefa de conversão - ibadatCoordinator	46
Figura 21 - Geração da fila de processamento	48
Figura 22 - Nomenclatura dos arquivos de produção.....	49
Figura 23 - Dataframe Principal	51
Figura 24 - Representação do descarte inicial por espessura	52
Figura 25 - Selecionando a região de descarte inicial	53
Figura 26 - Seleção dos valores de espessura fora dos limites	54
Figura 27 - Representação gráfica dos valores para descarte inicial de temperatura	56
Figura 28 - Filtrando a região de descarte inicial	57
Figura 29 - Seleção dos valores de temperatura fora dos limites	58
Figura 30 - Representação do descarte final por espessura	59
Figura 31 - Gráfico com a região de descarte final por espessura.....	60

LISTA DE FIGURAS

Figura 32 - Seleção dos valores de espessura fora dos limites.....	61
Figura 33 - Filtrando a região de descarte final por temperatura	63
Figura 34 - Seleção dos valores de temperatura final fora dos limites.....	64
Figura 35 - Trecho da tabela clientes.....	67
Figura 36 - Trecho da tabela resultado	69
Figura 37 - Dataframe de features do modelo originado da tabela "resultado"	74
Figura 38 - Estrutura da função train_test_split	75
Figura 39 - Visão gráfica da árvore de decisão inicial	79
Figura 40 - Visão gráfica da árvore após Pruning	84
Figura 41 - Aspectos de profundidade e nós da árvore de acordó com os valores de ccp_alpha	86
Figura 42 - Alfa vs acurácia	87
Figura 43 - Alfa vs acurácia (zoom)	87
Figura 44 - Ranking de importância das features	88
Figura 45 - Tela principal do supervisório	90

LISTA DE QUADROS

Quadro 1 - Código para listagem dos arquivos do diretório padrão.....	47
Quadro 2 - Código para identificação e classificação dos arquivos de produção	50
Quadro 3 - Armazena o nome dos arquivos e faz a leitura dos dados.....	51
Quadro 4 - Código para a seleção da região de descarte inicial	53
Quadro 5 - Código para identificação do comprimento do descarte inicial por espessura ..	55
Quadro 6 - Definição dos limites e criação de <i>dataframe</i> para cálculo do descarte inicial por temperatura	55
Quadro 7 – Código para a seleção região para descarte inicial por temperatura.....	56
Quadro 8 - Código para seleção dos valores de erro de temperatura fora dos limites especificados.....	57
Quadro 9 - Código para identificação do comprimento do descarte inicial por temperatura	58
Quadro 10 - Código para determinar a região de descarte final por espessura	59
Quadro 11 - Código para seleção dos valores fora dos limites especificados para a região de descarte final por espessura	61
Quadro 12 - Código para determinar o comprimento do descarte final por espessura	62
Quadro 13 - Código para determinar a região de descarte final por temperatura	62
Quadro 14 - Código para seleção dos valores fora dos limites especificados para a região de descarte final por temperatura	63
Quadro 15 - Código para determinar o comprimento do descarte final por temperatura.....	64
Quadro 16 - Código para cálculo do comprimento final eliminando as regiões de descarte	65
Quadro 17 - Script para a criação do banco de dados	66
Quadro 18 - Script para a criação da tabela clientes.....	66
Quadro 19 - Script para a criação da tabela resultado	68
Quadro 20 - Código para gerenciamento de arquivos processados.....	70
Quadro 21 - Código para a criação da conexão com o banco de dados	71
Quadro 22 - Código para a inserção de dados na tabela resultado	72
Quadro 23 - Código para criação dos conjuntos de dados de teste e treinamento	76
Quadro 24 - Código para a criação da árvore de decisão inicial	77
Quadro 25 - Código para a verificação de acurácia do modelo inicial.....	78
Quadro 26 - Código para a plotagem gráfica da árvore utilizando graphviz	79
Quadro 27 - Trecho do código para testes de profundidade com o critério Gini	80

Quadro 28 - Trecho do código para testes de profundidade com o critério Entropia.....	82
Quadro 29 – Códigos para a construção do modelo com os parâmetros otimizados e plotagem gráfica	84

LISTA DE TABELAS

Tabela 1 - Publicações em Scopus utilizando as palavras-chave <i>Machine Learning</i> , <i>Decision Making</i> , <i>Rolling</i> , <i>Decision Tree</i> e <i>Quality Control</i>	22
Tabela 2 - Pesquisa Aprofundada (Scopus).....	26
Tabela 3 - Resultados de acurácia para o critério Gini.....	81
Tabela 4 - Resultados de acurácia para o critério Entropia	82
Tabela 5 - Comparativo de acurácia para os critérios Gini e Entropia.....	83

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	QUESTÃO DA PESQUISA E OBJETIVOS.....	18
1.1.1	Questão da pesquisa	18
1.1.2	Objetivo Geral.....	18
1.1.3	Objetivos Específicos	18
1.2	CLASSIFICAÇÃO DA PESQUISA CIENTÍFICA	19
1.3	DELIMITAÇÃO DA PESQUISA	20
1.4	ESTRUTURA DA TESE.....	20
2	DESENVOLVIMENTO	22
2.1	REVISÃO DE LITERATURA.....	22
2.1.1	Justificativa e oportunidades.....	22
2.2	REFERENCIAL TEÓRICO	26
2.2.1	Redes neurais artificiais	27
2.2.2	Naive Bayes	27
2.2.3	Regressão linear.....	28
2.2.4	Máquinas De Vetores De Suporte (SVM).....	28
2.2.5	K-Vizinhos mais próximos.....	28
2.2.6	Árvores de decisão	29
2.2.7	Métricas de avaliação	30
2.2.8	O Processo de laminação.....	31
3	MATERIAIS E MÉTODOS	34
3.1	SISTEMÁTICA DO TRABALHO.....	34
3.2	FERRAMENTAS UTILIZADAS	39
3.2.1	Softwares iba	39
3.2.2	Arquivos com extensão “parquet”	40
3.2.3	Banco de dados PostgreSQL.....	41
3.2.4	Linguagem Python.....	41
3.2.5	Software Pycharm para programação em Python	42
3.3	TRATATIVA DE DADOS.....	42

SUMÁRIO

3.3.1	Parâmetros para monitoramento e controle de qualidade	42
3.3.2	Coleta de dados	44
3.3.3	Conversão de arquivos .dat para .parquet.....	46
3.3.4	Gerenciamento dos arquivos .parquet.....	47
3.3.5	Identificação do produto	49
3.3.6	Cálculo das regiões de descarte	50
3.3.7	Cálculo do comprimento final após descartes.....	65
3.3.8	Armazenamento de dados.....	65
3.3.9	Gestão dos arquivos processados	69
3.3.10	Ingestão e consulta de dados.....	70
3.4	CONSTRUÇÃO DO MODELO.....	72
3.4.1	Filtragem de dados	73
3.4.2	Separação de dados (treinamento e teste)	74
3.4.3	Construindo a árvore de decisão classificatória	76
4	RESULTADOS E DISCUSSÕES	79
4.1	OTIMIZAÇÃO DO MODELO (PRUNING)	79
4.2	MÉTRICAS DE VALIDAÇÃO	88
4.3	SUPERVISÓRIO ITERATIVO PARA O USUÁRIO FINAL	89
5	CONCLUSÃO.....	91
	REFERÊNCIAS	92

1 INTRODUÇÃO

O desenvolvimento da indústria 4.0 tem se apresentado substancialmente presente nas indústrias de transformação. Ou seja, os processos dessa nova revolução estão associados as denominadas fábricas inteligentes, nas quais fazem uso da tecnologia para otimizar suas atividades e garantir resultados precisos em comparação com as empresas que não estão inseridas no modelo da indústria 4.0 (STOCK; SELIGER, 2016).

Que tem como um dos pilares a coleta, armazenamento e processamento de grande quantidade de dados simultaneamente no intuito de utilizá-los para a criação de algoritmos e modelos que possam prever resultados de um determinado processo, encontrar correlações entre variáveis críticas para o processo, criar modelos compatíveis com o funcionamento de equipamentos reais e simular ajustes sem qualquer intervenção real, analisar conjuntos de variáveis como largura, espessura, composição química e comprimento final após remoção das regiões de transiente e classifica-las de acordo com padrões pré-estabelecidos pelos clientes ou para a próxima etapa do processo.

O processo de laminação tem várias particularidades de acordo com o tipo de produto produzido e sua aplicação, sendo necessária a utilização de equipamentos complexos (laminadores) que exigem o emprego dos mais diversos tipos de instrumentos, Controladores Lógico Programáveis (CLPs) e sistemas de aquisição de dados para garantir a qualidade do material ao longo do comprimento. Porém, qualquer desvio no processo pode causar defeitos graves no material, necessitando também de um sistema eficaz para análise dos KPOs (*Key Process Outputs*) de modo a classificar os possíveis defeitos e fazer o direcionamento para a tratativa adequada (descarte da região do defeito, direcionamento para outro tipo de produto (reclassificação) ou até mesmo a sucata total da bobina.

Atualmente, é comum no mercado o envolvimento de especialistas para a análise dos arquivos de produção e tomadas de decisão para o material. Além da necessidade de profissionais com vasta experiência para garantir a análise e tratativa correta, este processo demanda muito tempo da equipe, sobrecarregando a rotina e impedindo-os de executar outras ações estratégicas para o negócio e muitas vezes gerando classificações fora do padrão devido a erros de interpretação ou informações divergentes entre os analistas. Esse fato leva a impactos negativos em todo o fluxo da cadeia, prejudicando a entrega aos clientes e aumentando os custos de produção devido a necessidade de produzir novamente o item que foi sucateado.

Conforme características do produto se torna mais complexo, a configuração até um limite apropriado, como um controle de espessura, para identificar variação desconhecida dentro da tolerância torna-se difícil tarefa. Esta é uma das razões pelas quais a indústria ainda usa inspetores humanos para realizar tarefas implícitas de inspeção para detectar defeitos não mapeados previamente como acabamentos ruins, amassados, orientações erradas, e muitas outras falhas mal definidas. Contudo, em relação aos inspetores humanos, muitos sistemas de inspeção automatizados possuem melhor desempenho em termos de precisão, consistência e velocidade de inspeção sem a necessidade de definir um buffer para inspeção, além de reduzir os custos trabalhistas (YACOB; SEMERE; NORDGREN, 2019).

Contando com procedimentos sólidos e especificações detalhadas do material, surge a oportunidade de criação de um sistema capaz de analisar todo o material produzido de forma rápida, consistente e eficaz. Aliviando a carga dos profissionais envolvidos e garantindo análises padronizadas, eliminando possíveis divergências de interpretação entre diferentes profissionais.

Nos últimos anos, surgiram alguns novos métodos de aprendizagem de máquinas, como *Support Vector Machine* (SVM), *Classification* e *Árvore de Regressão* (CART), *Árvore de Regressão de Bagging* (BRT), *Redução Mínima Absoluta e Seleção Operador* (LASSO - *Least Absolute Shrinkage and Selection Operator*) e *Regressão Gaussiana do Processo* (GPR). Para a pesquisa no campo de laminação, percebeu-se que métodos de aprendizado de máquina podem ser utilizados para auxiliar os funcionários em tomadas de decisão referentes a classificação de produtos após o término do processamento, set up de equipamento de acordo com o material a ser produzido, manutenção preventiva, entre outros (XU, FENG e YAN, 2020).

Como ferramenta para a implementação de códigos utilizando técnicas de *machine learning*, tem-se a linguagem de programação Python, orientada a objetos, que ganhou popularidade devido ao fácil acesso a muitos pacotes de desenvolvimento, como Numpy, OpenCV, TensorFlow e PyQT. Python também inclui módulos bem conhecidos para aprendizado de máquina e computação científica (KIM, 2022). Além disso, com uma comunidade muito ativa, Python se tornou a linguagem de programação preferida para desenvolvimento de projetos nos últimos anos, ocupando o primeiro lugar no ranking interativo anual do IEEE Spectrum (CASS, 2021).

Objetivando a utilização de técnicas e ferramentas utilizadas atualmente na indústria 4.0, este trabalho consiste no desenvolvimento e implementação de uma ferramenta de análise para controle de qualidade de matérias laminados e bobinados, utilizando arquivos

de produtos contendo sinais do material e equipamento convertido em formato de texto (.parquet) para otimizar o processo de leitura, tratativa e processamento dos dados para a criação, treinamento e execução de um modelo de aprendizado de máquinas no software Python utilizando árvore de decisão.

Com este tipo de suporte automatizado para a tomada de decisão, a carga de trabalho e a carga cognitiva de os trabalhadores podem ser diminuídas e eles podem se concentrar na melhoria do processo quando ocorre um alarme de aumento do risco de falha durante a laminação, por exemplo. Os modelos de previsão estatística, tais como utilizando séries temporais, são especialmente adequados para a tomada de decisões; com sua capacidade de prever o resultado futuro, o tempo de reação após um alarme é maior (TAMMINEN et al, 2018).

1.1 QUESTÃO DA PESQUISA E OBJETIVOS

1.1.1 Questão da pesquisa

Quão eficiente seria o monitoramento do processo de laminação utilizando *machine learning*?

1.1.2 Objetivo Geral

Desenvolver um sistema supervisionado de monitoramento de um processo de laminação utilizando árvore de decisão classificatória para otimização da disposição do produto.

1.1.3 Objetivos Específicos

- Identificar dados dos recém-nascidos da cidade de São José dos Campos-SP, nos anos de 2016 a 2018
- Identificar relatórios diários dos poluentes do ar, PM₁₀ e NO₂, para o mesmo período
- Desenvolver um modelo baseado na lógica *fuzzy* e *neuro-fuzzy*, relacionando exposição das gestantes aos poluentes do ar com parto prematuro
- Realizar a comparação entre o modelo *fuzzy* e o sistema *neuro-fuzzy ANFIS* utilizados
- Estimar os parâmetros do processo;

- Desenvolver um algoritmo em Python utilizando técnica de *Machine Learning* (para monitorar o processo;
- Criar um sistema para geração automática e padronizada de Relatórios de Não Conformidade (RNC) para diferentes classes de material utilizando *Machine Learning* para tomada de decisão;

1.2 CLASSIFICAÇÃO DA PESQUISA CIENTÍFICA

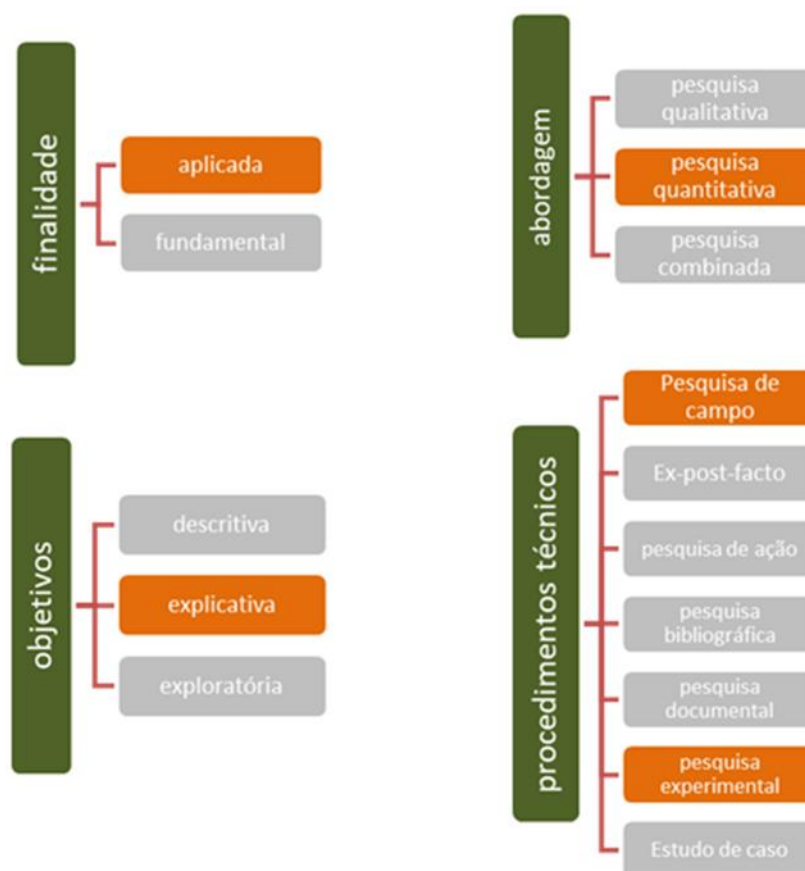
Conforme apresentado na Figura 6, trata-se de uma pesquisa aplicada, com objetivo explicativo, com abordagem quantitativa, de caráter explicativa, utilizando o procedimento experimental e pesquisa de campo.

Quanto à natureza é classificada como pesquisa aplicada, pois pretende gerar resultados para aplicação prática dirigida à solução de problemas específicos, neste caso relacionado ao controle de qualidade de materiais laminados. (LAKATOS; MARCONI, 2010).

Quanto à forma de abordagem, caracteriza-se como quantitativa devido às variáveis de decisão envolvidas no método proposto, com classificação através de registros de análises. (LAKATOS; MARCONI, 2010).

Quanto aos objetivos caracteriza-se como pesquisa explicativa pois visa identificar os fatores que determinam fenômenos e explica o porquê das coisas. Envolve levantamento de dados de processo, especificação de clientes, categoria de material e variáveis críticas de processo que determinam os fenômenos de defeito já descritos e detalhados. Quanto aos procedimentos técnicos utiliza-se a pesquisa experimental com base no levantamento das variáveis que seriam capazes de influenciar na qualidade do produto, e de campo por conta com a experiência de profissionais qualificados para confrontar os dados obtidos pelo modelo com as técnicas de análise de qualidade utilizadas atualmente (GIL, 2010).

Figura 1 - Classificação da pesquisa científica



Fonte: Adaptado de Kothari (2013).

1.3 DELIMITAÇÃO DA PESQUISA

A pesquisa está delimitada ao processo de laminação. Para o estudo e desenvolvimento foi utilizado um conjunto de 470 arquivos de produção de bobinas de material laminado no ano de 2022.

Os arquivos contendo os dados coletados do processo foram convertidos para uma extensão específica (.parquet) onde os sinais foram ser importados, tratados e utilizados como base para a criação de um modelo de árvore de decisão utilizando a ferramenta Python.

1.4 ESTRUTURA DA TESE

Este trabalho está organizado em 5 capítulos, o primeiro contém a introdução sobre o tema, assim como objetivo geral, específicos e oportunidades encontradas na pesquisa bibliométrica.

O segundo capítulo, Fundamentação Teórica, aborda os conceitos de *machine learning*, assim como as técnicas utilizadas, métricas de validação do modelo e o processo utilizado como base para este estudo.

O terceiro capítulo, Ferramentas Utilizadas, aborda detalhes sobre as ferramentas utilizadas para o desenvolvimento do projeto, assim como o passo a passo executados ao longo do desenvolvimento do código e funcionalidades de cada função.

No quarto capítulo, são apresentados os resultados obtidos durante a criação e após otimização do modelo, com detalhamento de métricas de validação do mesmo e comparações com os métodos de classificação convencionais já utilizados neste processo.

Por fim, o último capítulo expõe as conclusões alcançadas e contribuições deste estudo para a sociedade.

5 CONCLUSÃO

Neste estudo, foi possível avaliar dados de produção de uma empresa de laminação do período de 2022. Onde abordou a utilização da técnica de aprendizado de máquina conhecida como árvore de decisão classificatória para desenvolvimento de um modelo de caracterização e classificação de produtos de acordo com as variáveis de entrada liga, espessura, largura, descarte por espessura e temperatura na região final e inicial do produto, tendo como variável de saída categorias incluindo a ação tomada (seguir processo, reclassificar ou sucatear) seguidas do cliente sugerido de acordo com especificações pré definidas.

O modelo apresentou resultados satisfatórios após processo de *prunning* utilizando avaliação de acuraria e otimização do fato *ccp_alpha*. Saindo de uma acurácia inicial de 61% e profundidade de 11 para uma acurácia de 83,33% e profundidade de 9 com *ccp_alpha* de 0,012. Resultando em um modelo otimizado com precisão de 83,58%, *recall* de 83,33% e *f-score* de 82,07%. Já avaliando os pesos de cada *feature*, os resultados obtidos após a otimização do modelo de árvore de decisão mostraram a largura do material como *feature* principal, seguidos de comprimento, espessura e largura.

O desenvolvimento de um supervisor para interface e supervisão operacional possibilita a otimização constante do modelo através de entradas manuais caso necessário, além da visualização clara através de cores caracterizando cada ação tomada e atualização em tempo real após cada etapa do processo.

Como trabalhos futuros, destaca-se a possibilidade de integrar o modelo ao sistema de controle da produção de modo a otimizar a tomada de decisão observando o fluxo interno de material e estoques, tanto da fábrica quanto dos clientes em si.

Neste trabalho, observou-se que é possível construir um sistema eficiente para aplicação real na indústria usando software de código aberto e baixo custo de implementação e manutenção, incluindo uma alta flexibilidade para fazer adaptações para cada demanda de processo e com uma máquina humana amigável interface tornando-se uma operação simples de todos os níveis de usuários.

REFERÊNCIAS

BRAMER, M. **Pre-pruning classification trees to reduce overfitting in noisy domains**. Portsmouth: Faculty of Technology, University of Portsmouth, UK, 2002.

CASS, S. **Top programming languages 2021**. Disponível em: <https://spectrum.ieee.org/top-programming-languages-2021>. Acesso em: 21 set. 2023.

COUTINHO, T. **O que é python matplotlib?** conheça a biblioteca de gráficos! Disponível em: <https://www.voitto.com.br/blog/artigo/o-que-e-python-matplotlib>. Acesso em: 03 fev. 2023.

DESHPANDE, B. **2 main differences between classification and regression trees**. Disponível em: <https://www.simafore.com/blog/bid/62482/2-main-differences-between-classification-and-regression-trees>. Acesso em: 11 out. 2022.

GIL, A. C. **Como elaborar projetos de pesquisa**. 3. ed. São Paulo: Editora Atlas, 1996. 159p.

GRAYSON, J. **Python and Tkinter programming**. Greenwich: Manning Publications, 2000.

IBA AG. **Measurement systems for industry and energy**. Disponível em: <https://www.iba-ag.com/en/>. Acesso em: 10 mar. 2021.

INTERNATION MACHINES BUSINESS CORPORATION. **Supervised Learning**. Disponível em: <https://www.ibm.com/br-pt/topics/supervised-learning>. Acesso em: 18 dez. 2023.

JETBRAINS Pycharm: **o IDE Python para desenvolvedores profissionais**. Disponível em: <https://www.jetbrains.com/pt-br/pycharm>. Acesso em: 03 fev. 2023.

KALATHAS, I.; PAPOUTSIDAKIS, M. Predictive maintenance using machine learning and data mining: a pioneer method implemented to greek railways. **Designs**, v.5, n. 5, 2021. Disponível em: <https://doi.org/10.3390/designs5010005>. Acesso em: 13 abr. 2023.

KIM, M. Guaranteeing that multilevel prioritized DNN models on an embedded GPU have inference performance proportional to respective priorities. **IEEE Embedded Systems Letters**, New Jersey, v. 14, n. 2, p. 83-86, 2022. Disponível em: <https://ieeexplore.ieee.org/document/9624962>. Acesso em: 13 abr. 2023.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010.

LAURETTO, M. S. **Árvores de decisão**. EACH-USP, 2010. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf. Acesso em: 15 mar. 2023.

LI, J. *et al.* Feature selection: a data perspective. **ACM Computing Surveys**, v.50, n.6, dez. 2016. Disponível em: <https://dl.acm.org/doi/10.1145/3136625>. Acesso em: 07 mar. 2024.

LOH, W. Classification and regression trees. **Wiley interdisciplinary reviews: data mining and knowledge discovery**, New Jersey, v. 1, p. 14–23, 2011.

MANUFACTURING GUIDE. **Ingot casting**. Disponível em: <https://www.manufacturingguide.com/en/ingot-casting>. Acesso em: 24 dez. 2023.

MILMANN K.J.; AVAIZIS M. Python for scientists and engineers. **Computing in science & engineering**, California, v.13, p. 9-12, 2011. Disponível em: <https://escholarship.org/uc/item/93s2v2s7>. Acesso em: 10 set. 2023.

MULINARI, B. **Pandas Python**: vantagens e como começar. Harve 2023. Disponível em: <https://pypi.org/project/graphviz/>. Acesso em: 12 mar. 2023.

SQL MANAGER. **Postgress database management application example**. Disponível em: <https://www.sqlmanager.net/products/postgresql/manager>. Acesso em: 05 fev. 2023.

PSYCOPG ORGANIZATION. **Basic module usage**. Disponível em: <https://www.psycopg.org/docs/usage.html>. Acesso em: 12 mar. 2023.

PYDATA ORGANIZATION. **Pandas**: ready parquet. Disponível em: https://pandas.pydata.org/docs/reference/api/pandas.read_parquet.html. Acesso em: 04 fev. 2023.

PYTHON ORGANIZATION. **Os miscellaneous operating system interfaces**. Disponível em: <https://docs.python.org/3/library/os.html>. Acesso em: 08 fev. 2023.

RUEDIGER P. *et al.* Combining visual analytics and machine learning for reverse engineering in assembly quality control. **Journal of imaging science and technology**, v. 64, n. 6, p. 060405-1-060405-13, 2020. Disponível em: https://web.cs.ucdavis.edu/~hamann/RuedigerClausHamannHagenLeitteJIST_VDA2021PaperAsPublishedB12312020.pdf. Acesso em: 06 set. 2023.

PENG, R. Nonlinear vibration characteristics and time-delayed displacement control of rolling mill under dynamic rolling force. **Journal of Vibroengineering**, v. 23, n. 7. p. 1535–1548, ago. 2021.

SCIKIT-LEARN ORGANIZATION. **sklearn.model_selection.train_test_split**. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html. Acesso em: 24 mar. 2023.

SCIKIT-LEARN ORGANIZATION. **sklearn.tree.plot_tree**. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.tree.plot_tree.html: Acesso em: 13 mar. 2023.

SCIKIT-LEARN ORGANIZATION. **Decision trees**. Disponível em: <https://scikit-learn.org/stable/modules/tree.html>. Acesso em: 24 ago. 2023.

SCIKIT-LEARN ORGANIZATION. **Post pruning decision trees with cost complexity pruning**. Disponível em: https://scikit-learn.org/stable/auto_examples/tree/plot_cost_complexity_pruning.html#sphx-glr-auto-examples-tree-plot-cost-complexity-pruning-py. Acesso em: 12 dez. 2023.

SHAMRAT, F. M. J. *et al.* A comprehensive study on pre-pruning and post-pruning methods of decision tree classification algorithm. *In: INTERNATIONAL CONFERENCE ON TRENDS IN ELETRONICS AND INFORMATICS*, 5., 2021, Tirunelveli. **Proceedings [...]**. Tirunelveli, 2021. p. 1339-1345.

SINGER, G.; COHEN, I. **An objective-based entropy approach for interpretable decision tree models in support of human resource management: the case of absenteeism at work**. Ramat Gun: Faculty of Engineering, Bar-Ilan University, 2020.

SINGER, G.; COHEN, I. An objective-based entropy approach for interpretable decision tree models in support of human resource management: the case of absenteeism at work. **Entropy**, Calgary, v.22, n. 8, p.821, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7517405/pdf/entropy-22-00821.pdf>. Acesso em: 18 set. 2023.

SONGINI, M L. Put in plain language: the high portable, object-oriented python language moves into enterprise application development. **Computerworld**, 12 set. 2005. Disponível em: <http://www.computerworld.com/softwaretopics/software/story/0,10801,104484,00.html>. Acesso em: 30 maio 2022.

STOJILJKOVIĆ, M. **Split your dataset with scikit-learn's train_test_split()**. Disponível em: <https://realpython.com/train-test-split-python-data/>. Acesso em: 13 mar. 2023.

STOCK, T.; SELIGER, G. Opportunities of sustainable manufacturing in industry 4.0. **Procedia Cirp**, v. 40, p. 536-541, 2016.

TAMMINEN, S. *et al.* From measurements to knowledge: online quality monitoring and smart manufacturing. *In: PERNER, P. (ed.) Advances in data mining: applications and theoretical aspects*. Oulu: Springer, 2018. v. 10933, p. 17-28.

TEC-SCIENCE. **From steel to semi-finished products**. Disponível em: <https://www.tec-science.com/material-science/steel-making/steel-semi-finished-products-continuous-ingot-casting>. Acesso em: 13 dez. 2023.

WANG, Y.; WITTEN. Induction of model trees for predicting continuous classes. *In: THE EUROPEAN CONFERENCE ON MACHINE LEARNING*, 1997, Prague. **Proceedings [...]**. Prague: University of Economics, Faculty of Informatics and Statistics, 1997.

WHITE, T. **Hadoop: the definite guide**. 4th ed. O' Reilly Media, 2015.

XU, L.; FENG, L.; YAN, W. A comparative assessment of six machine learning models for prediction of bending force in hot strip rolling process. **Metals**, v.10, n.5, 2020.

YACOB, F.; SEMERE, D.; NORDGREN, E. Anomaly detection in Skin Model Shapes using machine learning classifiers. **The International Journal of Advanced Manufacturing Technology**, v.105, p. 3677-3689, 2019.