

PRECISÃO NA DOSAGEM DE MATÉRIAS-PRIMAS EM FÁBRICAS DE NUTRIÇÃO ANIMAL UTILIZANDO TÉCNICAS DE MACHINE LEARNING

Filipi Bitencourt Piucco¹

Yuri Crotti²

Resumo: Empresas que atuam no desenvolvimento de soluções para processo *batch* voltado a nutrição animal tem-se dedicado a tornar o *MES* (Sistema de Execução da Manufatura) juntamente ao controle de automação em um sistema especializado, onde busca-se melhor precisão na dosagem dos produtos. Porém a parametrização ainda é realizada pela equipe de operação de forma manual. Este cenário indica vulnerabilidade dos processos produtivos relacionado a precisão na dosagem. Por este motivo propomos utilizar machine learning, seguindo uma abordagem supervisionada, onde submetemos os dados de um sistema de dosagem aos modelos de regressão linear, árvore de decisão e redes neurais artificiais. Sendo possível a partir do *setpoint* de dosagem prever os melhores parâmetros para garantir a precisão na dosagem. Dos modelos utilizados a árvore de decisão foi o modelo que teve o pior desempenho. Podemos indicar que tanto a regressão linear quanto a rede neural apresentaram bons resultados de predição. Atingindo valores de diferenças máximas e mínimas abaixo de 1% para rede neural, e entre 1,57% e 3,42% para a regressão linear, mas com diferença média abaixo de 1% para ambos os modelos. Utilizando machine learning este trabalho apresentou uma solução que facilitou a configuração e melhorou a precisão na dosagem de matérias-primas em fábricas de nutrição animal. Porém a solução não se limita a este processo, podendo ser utilizado a qualquer processo produtivo que utilize sistemas de dosagem por batelada. Além disso o conceito pode ser utilizado amplamente para outras soluções como indicação de necessidade de manutenção do sistema de dosagem, entre outros.

Palavras-chave: Dosagem. Precisão. Machine learning. Batelada. Nutrição animal.

1 INTRODUÇÃO

Em fábricas de nutrição animal utiliza-se o processo de fabricação *batch* que se referencia na norma ISA S88 e se faz o uso de receitas. Esta por sua vez é composta por fórmula, onde indica-se o percentual de inclusão das matérias-primas, informações de segurança e conformidade da linha de produção e o procedimento da produção. Estes são parâmetros utilizados para cumprir cada etapa do processo produtivo, como tempo de mistura e limpeza de linha, intertravamentos e tolerância

¹ Graduando em Engenharia da Computação. Ano 2022-2. E-mail: filipipiucco@hotmail.com

² Professor do Centro Universitário UniSATC E-mail: yuri.crotti@satc.edu.br

de dosagens. *Batch* ou batelada como é conhecido, é o resultado da formulação e segue um procedimento que tem como característica realizar etapas de produção com uma sequência definida, onde se busca padronizar o processo produtivo através da repetição.

Em um processo de fabricação cada etapa tem sua importância para a qualidade do produto acabado, porém, se destaca a dosagem adequada de matérias-primas (LEMOS et. al., 2016). Considerado este um momento crítico do processo, a conformidade com as especificações da fórmula deve ocorrer de forma precisa, o que na maioria das vezes, pode criar um gargalo no processo produtivo (NETO, 2006).

Nos últimos anos, empresas que atuam no desenvolvimento de soluções para processo *batch* voltado a nutrição animal tem-se dedicado a tornar o *MES* (Sistema de Execução da Manufatura) juntamente ao controle de automação em um sistema especializado, onde se busca otimizar a dosagem dos produtos com o objetivo de se obter maior precisão. Os softwares de controle oferecem ferramentas que tornam possível ajustar as dosagens e obter bons resultados, porém, esta parametrização ainda é realizada pela equipe de operação de forma manual. Estes sistemas utilizam da experiência do operador em relação ao produto, equipamentos e demais variáveis do processo para encontrar a melhor configuração. Sempre que uma condição do processo muda a equipe de operação precisa atuar, buscando um equilíbrio entre precisão e produtividade (FEHRENBACH, 2017).

O processo de construção de fórmula não é intuitivo, se consideram questões técnicas, seja em processos de nutrição animal, através de variáveis nutricionais, biológicas ou em demais processos industriais. Sendo assim a qualidade do produto acabado está diretamente ligada aos requisitos de fórmula.

Levando em conta a realidade dos controles de automação e a importância de manter preservados os requisitos de fórmula, é preciso considerar que o fator humano influenciará no resultado final. Um operador pode ser mais experiente que outro, e normalmente só atua quando o erro ocorre, não existindo uma predição do problema.

Através da análise da base de dados histórica, utilizando técnicas de Inteligência Artificial (IA) com foco em aprendizado de máquina supervisionado como regressão lineares (RL), árvore de decisão (AD) e rede neural artificiais (RNA), além

da análise exploratória torna-se possível traçar a correlação entre as variáveis do processo que impactam na eficiência e precisão das dosagens de matérias-primas. A partir da correlação pode-se extrair as principais características, treinar o modelo e conseguir um dado de saída capaz de otimizar o processo de dosagem de forma preditiva, ou seja, otimizar e alcançar os melhores índices de precisão na dosagem de matéria prima em fabricas de nutrição animal.

2 REVISÃO BIBLIOGRÁFICA

A fabricação de ração em ambiente industrial entre outros objetivos procura garantir a segurança alimentar e equilíbrio nutricional. As metodologias e procedimentos realizados neste processo produtivo, representam geralmente até 4% do custo de fabricação da ração. Porém esta pequena cota de custos influenciará diretamente os outros mais de 90% do custo da alimentação, que são representados pelo valor das matérias-primas (Duarte, S.d.).

A qualidade da ração é observada por dois requisitos, sendo: sanitário e nutricional. Abordaremos neste trabalho métodos para garantir o requisito nutricional que por sua vez é alcançado através da dosagem correta (com precisão) das matérias-primas (Duarte, S.d.). Para atingir este objetivo aplicaremos técnicas de IA.

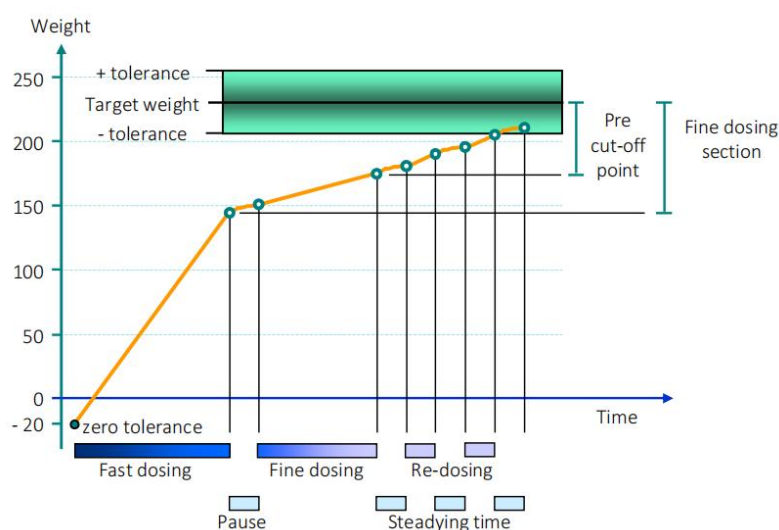
2.1 DOSAGEM DE PRECISÃO

A dosagem de precisão em processos industriais baseados em bateladas é tão importante na continuidade do negócio quanto à estratégia de mercado ou o planejamento financeiro de uma empresa. A competitividade entre as empresas para ganhar espaço ou manter-se no mercado está cada vez maior, isso pode causar impactos nos lucros. O sucesso da organização na maioria das vezes estará relacionado a eficiência do processo produtivo e a dosagem da matéria-prima com precisão garante o planejamento da produção, seus custos, além de manter o padrão de qualidade (ALVES et. al., 2016).

Para alcançar uma dosagem precisa o processo de dosagem vem sendo automatizado. Ele tem como objetivo que a dosagem de matéria-prima permaneça dentro das margens de tolerância e ainda assim mantenha-se um ritmo de produção

constante, minimizando as falhas e interferências humanas. Os sistemas de dosagens automatizados são compostos por atuadores mecânicos acionados por componentes elétricos e estes por sua vez são comandados por Controladores Lógicos Programáveis (CLP). Compõe ainda o sistema, balanças industriais que medem o peso do produto dosado (FEHRENBACH, 2017). Abaixo conforme Fig. 1, podemos observar um exemplo de algoritmo de dosagem e seus parâmetros.

Figura 1: Algoritmo de dosagem de matéria-prima



Fonte: Baseado de Ammann (2020)

Atualmente no ambiente industrial, ainda que se tenha equipamentos que permitam realizar uma dosagem sem que o homem tenha contato com equipamentos ou matéria prima, os parâmetros ou dados de entrada do sistema que regem a dosagem, ou seja, suas configurações ainda dependem da interação humana (FEHRENBACH, 2017). Neste trabalho serão abordados, para efeito de revisão bibliográfica, técnicas de Machine Learning que serão utilizadas para tratar os dados de entrada em um sistema de dosagem e a partir destes encontrar a melhor configuração para atingir a precisão na dosagem de matérias primas.

2.2 SISTEMA DE DOSAGEM

O sistema de dosagem é composto por uma balança industrial que por sua vez é um conjunto de equipamentos responsáveis por realizar as pesagens das

matérias-primas de forma automática (RICARDO, 2017). No exemplo ilustrado através da Fig. 2, podemos observar um sistema de dosagem composto por dosadores do tipo rosca helicoidal com variação de velocidade, acoplados aos silos onde encontram-se as matérias-primas. O produto dosado por sua vez é pesado em um recipiente metálico, a balança propriamente dita, que está conectada as células de carga que são os instrumentos que fazem a medição do peso, tanto da estrutura, quanto da matéria-prima.

Figura 2: Sistema de dosagem de matéria-prima



Fonte: Baseado de Ammann (2020)

2.2 INTELIGÊNCIA ARTIFICIAL

Inteligência artificial (IA) é um campo prospero, contém muitas soluções práticas e campos de pesquisas ativos. Inicialmente a IA abordou soluções para problemas intelectualmente difíceis para os seres humanos, mas simples para computadores, como problemas envolvendo regras matemáticas. Porém, o principal desafio parece ser as atividades que os seres humanos realizam intuitivamente,

como se fossem automáticas. Problemas como reconhecimento de palavras faladas e rostos em imagens. Os sistemas baseados em IA precisam adquirir seu próprio conhecimento e realizam isso através da identificação de padrões em dados brutos. Essa capacidade é conhecida como aprendizado de máquina (GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A, 2016).

2.3 MACHINE LEARNING

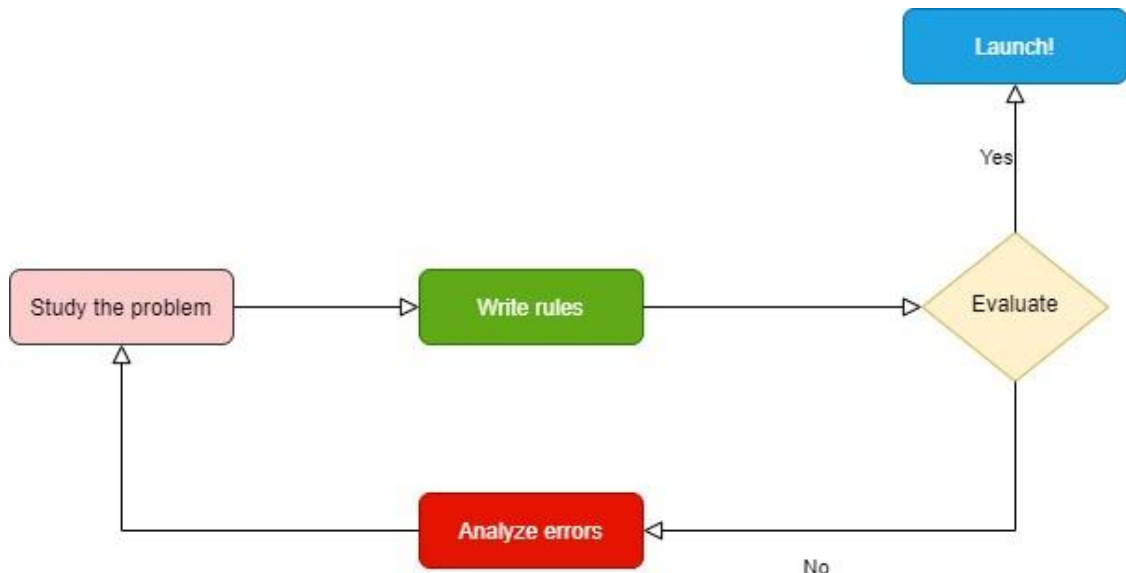
O conceito de Machine Learning (aprendizagem de máquina), causa certa confusão no entendimento do tema e isso pode criar pré-conceitos em relação a sua aplicabilidade.

“Quando a maioria das pessoas ouve “Aprendizado de máquina”, elas imaginam um robô” (GÉRON, 2017, p. 3).

Conforme Burkov (2019) Machine Learning pode ser considerado um processo para resolver problemas práticos. Atualmente na indústria é considerado um dos pilares deste novo momento em que as soluções são obtidas a partir da análise de dados. Isso porque a abordagem de machine learning permite a extração de informação utilizando dados de forma eficiente e eficaz. A eficiência alavanca-se muito devido ao baixo custo dos dispositivos voltados a computação de alto desempenho, quantidade e qualidade dos dados gerados, bem como a quantidade de modelos de aprendizagem utilizados para representar as aplicações. Esses modelos têm-se demonstrado cada vez mais sofisticados e hábeis, resolvendo problemas complexos, utilizando áreas da ciência como matemática e estatística (FREITAS, A. L.; SANTANA JUNIOR. O. V, 2019).

Em uma abordagem tradicional os termos de entrada são à base de dados e as regras necessárias para processar a mesma. Quando considera-se trabalhar com grandes volumes de dados, esta abordagem se torna inviável devido à complexidade de programar ou prever todas as regras (GÉRON, 2017). A seguir a Fig. 3 demonstra o fluxo de uma abordagem tradicional.

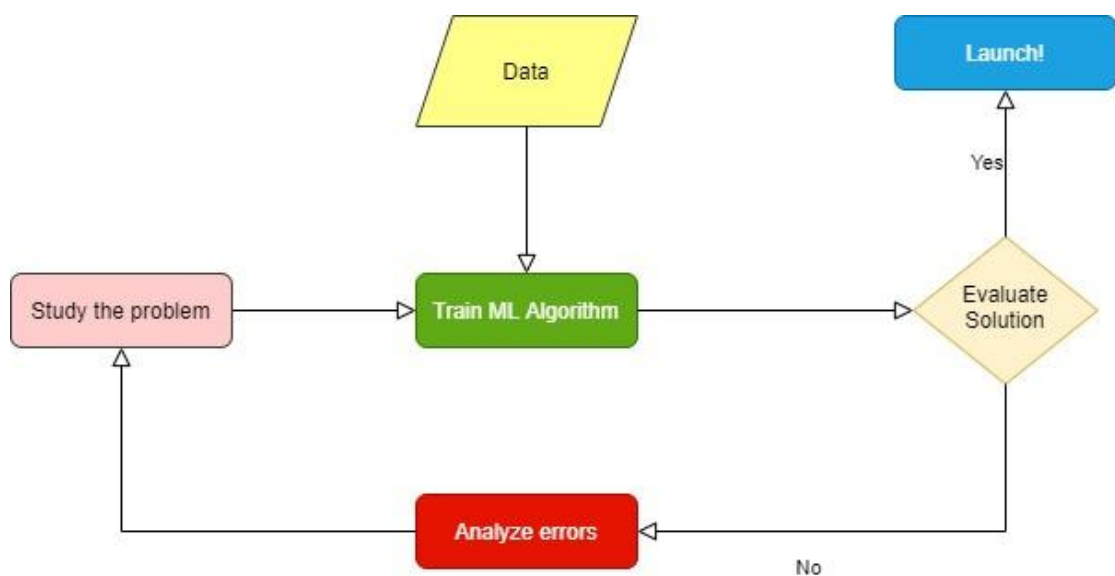
Figura 3: Abordagem tradicional de processamento de dados



Fonte: Baseado de Géron (2017)

Já a abordagem de machine learning utiliza-se dos dados de entrada para treinar e a partir deste ponto desenvolver os padrões e entregar os melhores dados de saída para o modelo (GÉRON, 2017). A seguir a Fig. 4 demonstra o fluxo de uma abordagem utilizando machine learning.

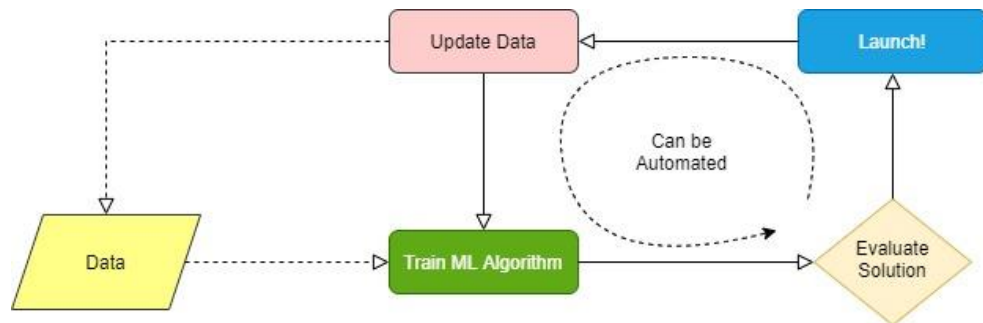
Figura 4: Abordagem Machine Learning



Fonte: Baseado de Géron (2017)

Já a Fig. 5 demonstra o fluxo de uma abordagem machine learning que retroalimenta a base de dados a partir dos resultados obtidos pelo modelo de modo a otimizar a aprendizagem através dos resultados obtidos pelo próprio modelo.

Figura 5: Abordagem Machine Learning com retroalimentação



Fonte: Baseado de Géron (2017)

Existem diferentes tipos de aprendizado de máquina, sendo aprendizado supervisionado, não supervisionado e por reforço (PEREZ, 2017). Abordaremos nesta revisão e nos procedimentos experimentais o aprendizado supervisionado.

2.3.1 Aprendizado supervisionado

No aprendizado supervisionado os dados são rotulados, ou seja, com identificação, as informações são conhecidas e sabe-se qual o *target*, variável de saída. Por este motivo o modelo recebe as variáveis de entradas e os resultados obtidos através das mesmas como forma de treinamento (GÉRON, 2017). O processo inicia-se com a coleta dos dados que são obtidos em pares (entrada e saída). A entrada pode ser qualquer informação, já as saídas são geralmente números reais ou rótulos (Burkov, 2019). Os principais métodos de aprendizagem são:

- Regressão Linear;
- Regressão Logística;
- KNN (vizinhos mais próximos);
- Máquinas de vetor de suporte (SVMs);
- Árvores de Decisão e Florestas Aleatórias;

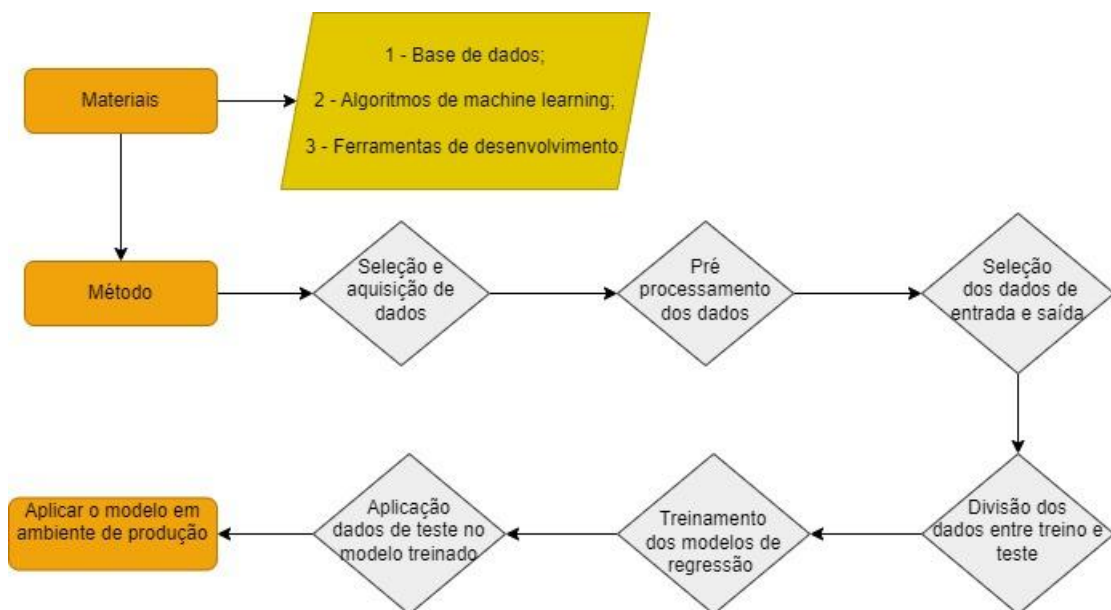
➤ Redes Neurais Artificiais.

Esses métodos podem resolver problemas de regressão e classificação, A regressão é utilizada geralmente para obter como resultado um valor numérico, por exemplo prever o valor médio de um carro. Já classificação entrega como variável de saída uma categoria, como por exemplo a classificação de defeitos em ambiente de fábrica. (Burkov, 2019).

3 PROCEDIMENTOS EXPERIMENTAIS

Este trabalho seguiu as etapas de desenvolvimento conforme fluxo demonstrado na Fig. 6.

Figura 6: Fluxograma do procedimento experimental



Fonte: Do autor (2022)

3.1 MATERIAIS

A seguir serão abordados os materiais utilizados no desenvolvimento deste trabalho.

3.1.1 Base de dados

As bases de dados utilizadas neste trabalho foram geradas a partir de um simulador de dosagem de propriedade da empresa AGPR5, localizada na cidade de Criciúma - SC. As bases contêm as mesmas informações quando comparadas com a base histórica de um sistema de dosagem real. O simulador é configurado com todos os parâmetros disponíveis no software que a AGPR5 comercializa e que configuram uma dosagem real de uma fábrica de nutrição animal. Além disso existe um setup que permite que as dosagens possam ser simuladas, que é a taxa de dosagem (kg/s). Utilizando a taxa de dosagem e variando as configurações, podemos chegar a valores dosados, muito próximos ao que ocorre em um sistema de dosagem real. A taxa de dosagem é gerada a partir de informações reais, que são extraídas da base de dados histórica de um sistema de dosagem.

3.1.2 Algoritmos de machine learning

Os algoritmos relacionados na Tab. 1 foram utilizados durante procedimento experimental e tiveram seus resultados comparados, de modo a identificar qual o mais eficiente.

Tabela 1: Algoritmos de machine learning.

Aprendizagem	Tipo de aprendizagem	Algoritmo
Supervisionada	Regressão	Linear
Supervisionada	Regressão	Árvore de Decisão
Supervisionada	Regressão	Rede Neural Artificial

Fonte: Do autor (2022)

As principais bibliotecas utilizadas para preparação dos dados, aplicação dos algoritmos e visualização dos dados estão relacionadas na Tab. 2.

Tabela 2: Bibliotecas.

Principais Bibliotecas	Função
Pandas	Aquisição e exploração da base de dados
Numpy	Preparação base de dados
Scikit - Learn	Preparação dos dados, treinamento, teste e métrica de algoritmo
Seaborn	Visualização dos dados através de imagens
Matplotlib	Visualização dos dados através de imagens

Fonte: Do autor (2022)

3.1.3 Ferramentas de desenvolvimento

As ferramentas relacionadas na Tab. 3 a seguir foram utilizados durante procedimento experimental.

Tabela 3: Ferramentas de desenvolvimento.

Ferramenta	Linguagem	Desenvolvedor
SQL Server	SQL	Microsoft
Visual Studio Code (VSCode)	Python	Microsoft
WPS Office	-	WPS
Google Colab	Python	Google
Notepad ++	-	Don Ho

Fonte: Do autor (2022)

3.2 MÉTODO

A seguir serão descritos os métodos abordados no desenvolvimento deste trabalho.

3.2.1 Seleção e aquisição de dados

Para seleção da base de dados, analisamos bases contendo histórico de dosagens de fabricas de nutrição animal, visando implementar a solução proposta neste trabalho. As principais informações que a base de dados forneceu foram as parametrizações da dosagem no sistema de automação e o histórico dos valores requisitados e dosados das matérias-primas. Estes dados estavam em um arquivo

de banco do Microsoft SQL SERVER e puderam ser visualizados utilizando SQL Server Management Studio.

Já com a base selecionada, porém antes de realizar qualquer preparação para executar modelos de aprendizagem, foram identificados dois pontos principais que impactariam no desenvolvimento do projeto, eles são:

- **Baixa variabilidade dos dados:** Os valores de referência para o *setpoint* de dosagem e para os valores dosados continham pouca variação. Isso se deve porque os processos de fabricação seguem uma padronização no tamanho das fórmulas (t) buscando sempre extrair a máxima eficiência dos equipamentos, ou seja, cada matéria-prima tem seu percentual de inclusão e tem pequenas alterações quando é necessário um balanceamento nutricional.

- **Ausência de histórico das parametrizações por dosagem:** Existia apenas o registro da última parametrização realizada, não contendo o histórico para cada dosagem.

Levando em consideração os pontos mencionados acima a base de dados selecionada para utilização neste trabalho foi gerada a partir do simulador de dosagens. O simulador é configurado com todos os parâmetros disponíveis no software que a AGPR5 comercializa e que configuram uma dosagem real de uma fábrica de nutrição animal. Além destas configurações, existe um *setup* que permite que as dosagens possam ser simuladas e desta forma popular uma base de dados, que é a taxa de dosagem (kg/s). Utilizando a taxa de dosagem e variando as configurações, podemos chegar a valores dosados, muito próximos ao que ocorre em um sistema de dosagem real. A taxa de dosagem é gerada a partir de informações reais, que são extraídas da base de dados histórica de um sistema de dosagem. Os dados gerados a partir do simulador continha a quantidade de 10 mil dosagens, o que representa uma boa base para treinamento de um modelo.

A Tab. 4 representa as informações verificadas tanto na base de dados de referência, extraída de um sistema de dosagem real, quando do simulador da AGPR5.

Tabela 4: Informações de dosagem

Parâmetro	Unidade
Setpoint	kg
Dosado	kg
P_a	kg
P_b	kg
P_c	%
P_d	%
P_e	%
P_f	kg
P_g	s
Tolerancia_mais	%
Tolerancia_menos	%
Diferenca_Percentual	%

Fonte: Do autor (2022)

3.2.2 Pré-processamento dos dados

O pré-processamento dos dados iniciou com a leitura da base de dados, armazenando as informações em um *dataset* o que permitiu realizar processos de limpeza, transformação ou enriquecimento quando necessário. Devido as regras existentes no simulador da AGPR5, não houve a necessidade de tratar campo vazio, dado não numérico ou valor atípico que foge da média (*outliers*). Porém foi necessário identificar o tipo de cada variável, onde analisado o *dataset* foram identificadas variáveis categóricas. Não sendo uma informação válida para treinamento dos modelos, foi realizado a limpeza do *dataset* removendo estas informações, o que trouxe maior qualidade aos dados utilizados para treinamento dos modelos.

3.2.3 Seleção dos dados de entrada e saída

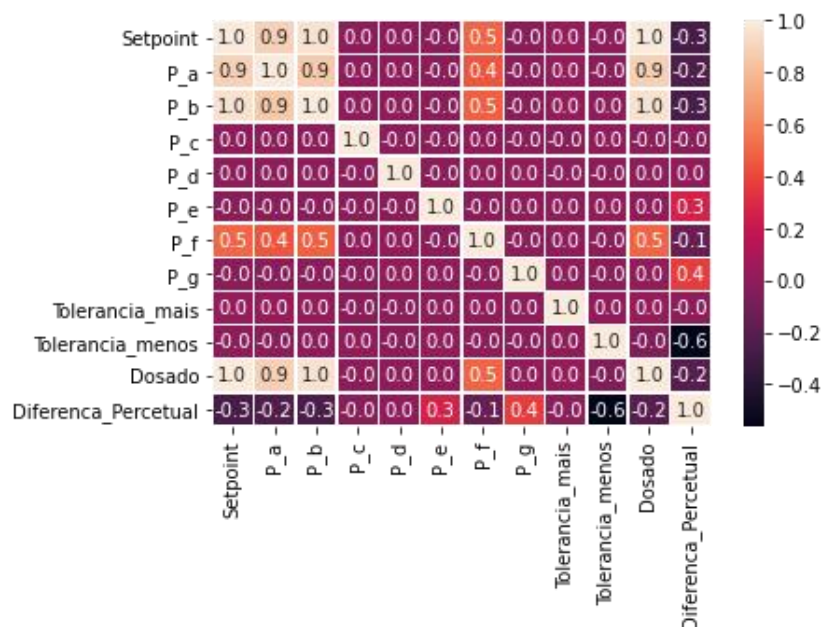
O objetivo com a base de dados é identificar a melhor configuração de dosagem para alcançar o valor de *setpoint* com a melhor precisão possível. Este objetivo é representado pelas informações abaixo:

- **Dosado:** Valor real dosado baseado no *setpoint*;
- **Diferenca Percetual:** diferenca percentual entre o *setpoint* e dosado.

Ainda que consideradas o objetivo, elas foram utilizadas como dados de entrada no modelo de treinamento e a partir delas encontramos os dados de saída, que são as configurações da dosagem.

Identificado os dados de entrada, foram verificados os índices de correlação com os demais dados da base. Para isso, utilizamos a correlação de Pearson e para poder realizar uma melhor análise visual do resultado geramos uma matriz de correlação utilizando o *heatmap* (mapa de calor) das correlações, conforme pode-se observar na Fig. 7.

Figura 7: Matriz de correlação dos dados de dosagem



Fonte: Do autor (2022)

Para avaliação da classificação da correlação, foram considerados as informações conforme a Fig. 8.

Figura 8: Classificação da correlação dos dados de dosagem

Valor de ρ (+ ou -)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Fonte: Baseado de SHIMAKURA (2006)

3.2.4 Divisão dos dados entre treino e teste

Com os dados de entrada (x) e saída (y) definidos, foram realizadas as separações de x e y para treino e teste dos modelos. A separação entre dados de treino e teste seguiu a proporção de 90% para treino e 10% para teste, gerando 4 novos conjuntos de dados. Quando necessário padronizar os dados, foi utilizado o método *standard scaler* (padronização).

3.2.5 Treinamento e teste dos modelos

Com os dados de treino e testes preparados, os mesmos foram submetidos aos modelos de regressão linear, árvore de decisão e rede neural artificial. Para comparar adequadamente o resultado, os modelos utilizaram os mesmos dados. O treinamento e teste dos modelos foi iniciado com a criação de suas instâncias e posteriormente com a aplicação dos dados separados para cada etapa. Os dados de saída preditos a partir do modelo treinado foram comparados com os dados de saída de teste já conhecidos.

Para cada modelo foram realizadas buscas de melhores hiperparâmetros utilizando uma busca em *grid* (*GRID SEARCH*).

3.2.6 Aplicação do modelo em ambiente de produção

Finalizado o treino e teste, aplicamos a mesma instância treinada a novos dados de entrada. A base utilizada continha 100 novos conjuntos de dados de entrada, representando 100 novos *setpoints* de dosagem para que o modelo pudesse prever os melhores parâmetros. Consideramos o simulador como sendo o ambiente de produção, visto que o mesmo contém as mesmas configurações e principalmente a mesma taxa de dosagem que deu origem a base de dados utilizada neste trabalho.

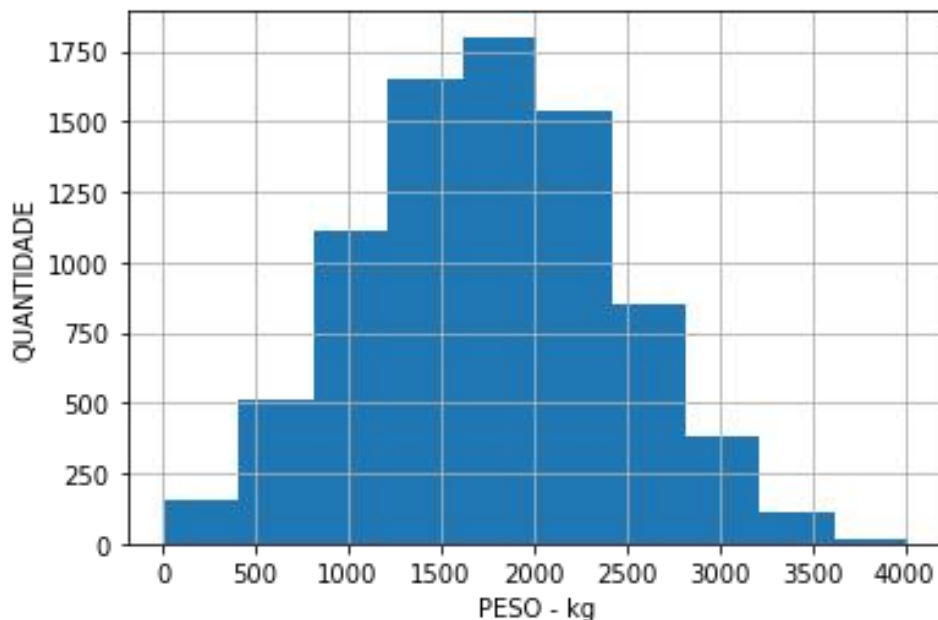
4 RESULTADOS E DISCUSSÃO

Finalizados os procedimentos experimentais, realizamos as análises de resultados das etapas descritas a seguir.

4.1 CARACTERÍSTICA DA BASE DE DADOS

As informações extraídas da base de dados foram importantes para entender o comportamento dos resultados em relação ao treino e teste dos modelos e principalmente quando colocado em produção. Analisando a base de dados utilizada durante desenvolvimento do trabalho foi possível identificar pontos importantes sobre suas características. A Fig. 9 demonstra a distribuição das faixas de *setpoint* de dosagem (kg), onde a principal distribuição está aproximadamente entre 1000 kg e 2500 kg, com maior densidade entre 1500 kg e 2000kg.

Figura 9: Distribuição *setpoint* de dosagem



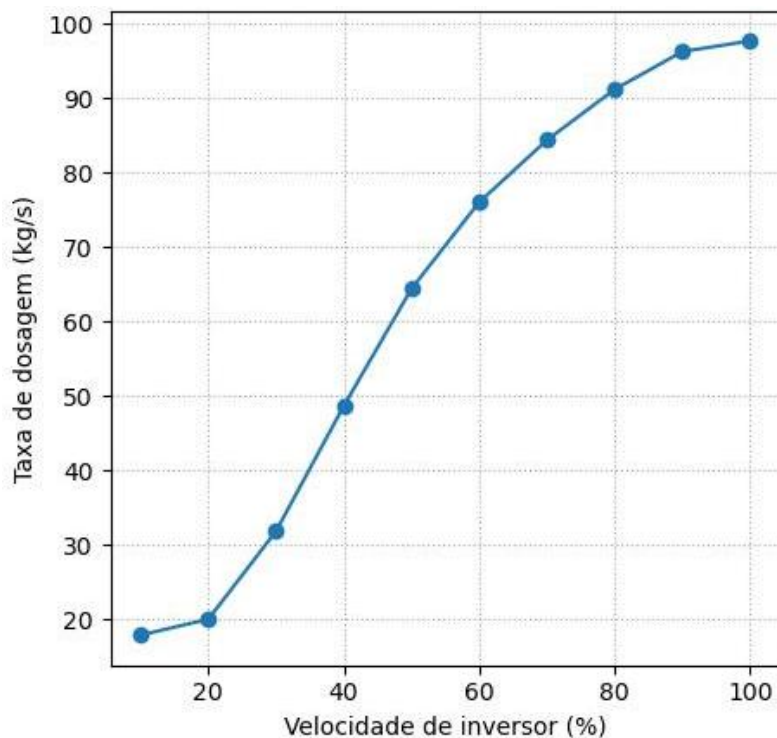
Fonte: Do autor (2022)

A análise e predição das configurações de um sistema de dosagem não deve ser analisada como um todo. Isto porque são aplicadas dosagens com matérias-

primas, componentes mecânicos, entre outras variáveis, com características diferentes. Desta forma a análise precisa sempre contemplar um silo ou tanque de cada vez. Para efeito de análise dos resultados deve-se considerar a faixa de *set point* de trabalho entre 1500 kg e 2000kg.

Já a Fig. 10 demonstra a taxa de dosagem (kg/s) do sistema, onde em 100% de velocidade pode dosar aproximadamente 98 kg/s. Esta taxa de dosagem é compatível com a faixa de trabalho de dosagem apontada na Fig. 9.

Figura 10: Conversão da velocidade de inversor em taxa de dosagem



Fonte: Do autor (2022)

Esta taxa de dosagem foi utilizada para configurar o simulador e foi extraída da base de dados do sistema de dosagem real ao qual alimentou inicialmente o simulador.

4.2 DADOS DE ENTRADA E SAÍDA SELECIONADOS

Foram analisadas as informações demonstradas pela matriz de correlação (Fig. 7). Em destaque os parâmetros P_a e P_b apresentaram correlação

positiva (diretamente proporcional) muito forte com o dado de entrada Dosado e uma correlação negativa (inversamente proporcional) fraca com Diferença Percentual. P_e e P_G apresentaram correlação bem fraca com Dosado e fraca e moderada respectivamente com Diferença Percentual. Já P_f apresenta correlação moderada com Dosado e bem fraca com Diferença Percentual. a Tab. 5 demonstra os parâmetros selecionados como dado de saída:

Tabela 5: Seleção de parâmetros por correlação

Parâmetro	V. de corr para Dosado	V. de corr para Diferença Percentual
P_a	0.9	-0.2
P_b	1.0	-0.3
P_e	0.0	0.3
P_f	0.5	-0.1
P_g	0.0	0.4

Fonte: Do autor (2022)

É possível notar que os parâmetros selecionados por correlação apresentam classificação diferente de bem fraca para pelo menos um dos dados de entrada.

A Tab. 6 demonstra os parâmetros selecionados por questões de processo, ainda que demonstrassem uma correlação bem fraca com os dados de entrada:

Tabela 6: Seleção de parâmetros por processo

Parâmetro	V. de corr para Dosado	V. de corr para Diferença Percentual
P_c	0.0	0.0
P_d	0.0	0.0

Fonte: Do autor (2022)

As variáveis P_c e P_d estão relacionados a configuração da velocidade dos dosadores. Mesmo não apresentando boa correlação foram selecionados para que o modelo também possa prever estes dados, mitigando as introduções manuais da operação nas configurações de dosagem.

Já o parâmetro Tolerância Menos que apresenta uma correlação moderada (-0,6), não foi selecionada por se tratar de um dado que é definido pelo cliente conforme regras do padrão de qualidade.

4.3 AVALIAÇÃO DO MODELO TREINADO

Analisado os resultados, o modelo de rede neural artificial apresenta os melhores valores para as métricas, seguido pela regressão linear e por último o modelo de árvore de decisão. As métricas aplicadas foram:

- Erro médio absoluto (MAE):

O erro médio absoluto é definido pela equação Eq. (1):

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x| \quad (1)$$

Onde:

n = número de erros (---);

$|x_i - x|$ = os erros absolutos (---).

- R-quadrado:

R-quadrado é definido pela equação Eq. (2):

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (2)$$

Onde:

x = Média de amostra valores conhecidos x (---);

y = Média de amostra valores conhecidos y (---).

A Tab. 7 demonstra os valores obtidos para o erro médio absoluto em todos os modelos aplicados. De forma complementar é indicado a precisão % gerada a partir do erro absoluto entre a média absoluta e o MAE.

Tabela 7: Erro médio absoluto

Parâmetro	média absoluta parâmetro	MAE RL	MAE AD	MAE RNA	Precisão RL (%)	Precisão AD(%)	Precisão RNA(%)
P_a	872,57	135,36	170,60	134,81	84,49	80,45	84,55
P_b	522,40	44,84	57,33	44,10	91,42	89,02	91,56
P_c	90,15	5,18	6,80	5,19	94,25	92,46	94,24
P_d	60,40	10,37	13,73	10,38	82,83	77,27	82,81
P_e	27,58	6,32	8,42	6,11	77,08	69,47	77,85
P_f	224,30	105,60	141,30	104,98	52,92	37,00	53,20
P_g	2,72	1,05	1,26	0,95	61,40	53,68	65,07

Fonte: Do autor (2022)

Conforme Tab. 7, regressão linear e a rede neural tem valores muito próximos. Já a árvore de decisão apresenta um erro maior quando comparado aos demais modelos. Quando verificado individualmente os valores absolutos do MAE a percepção é de que alguns parâmetros apresentaram um erro alto. Mas quando analisado em conjunto com a média absoluta do parâmetro é possível verificar que na maioria das vezes a Precisão é superior a 77%.

A Tab. 8 demonstra os valores obtidos para R-quadrado, onde foi identificado que o resultado acompanhou o índice de correlação dos parâmetros apontados na matriz de correlação (Figura 8):

Tabela 8: R-quadrado

Parâmetro	Regressão linear	Árvore de decisão	Rede neural artificial
P_a	0,78	0,60	0,79
P_b	0,92	0,85	0,92
P_c	-0,0011	-0,93	0,00056
P_d	-0,0027	-0,94	-0,0053
P_e	0,0263	0,93	0,07
P_f	0,26	-0,50	0,27
P_g	0,04	-0,56	0,20

Fonte: Do autor (2022)

Quando observado os modelos de regressão linear e rede neural, os parâmetros com correlação mais fortes apresentaram melhor resultado para R-quadrado e conseqüentemente pior resultado para correlações moderadas e fracas. Já a árvore de decisão apresentou certa incoerência, já que parâmetros

considerados com correlação muito fracas obtiveram melhor resultado do que os que apresentaram correlação muito alta.

É importante ressaltar, ainda que em alguns pontos os modelos não apresentem métricas tão boas em relação a precisão quando comparado os dados de treino e teste, que não existe apenas uma combinação para atingir um valor de precisão na dosagem. O importante é que o modelo tenha aprendido como chegar no valor mais preciso em relação ao *setpoint*, indiferente da combinação. Por este motivo é importante dar maior foco aos resultados do teste de produção / simulação de dosagens.

4.4 AVALIAÇÃO DO MODELO EM AMBIENTE DE PRODUÇÃO

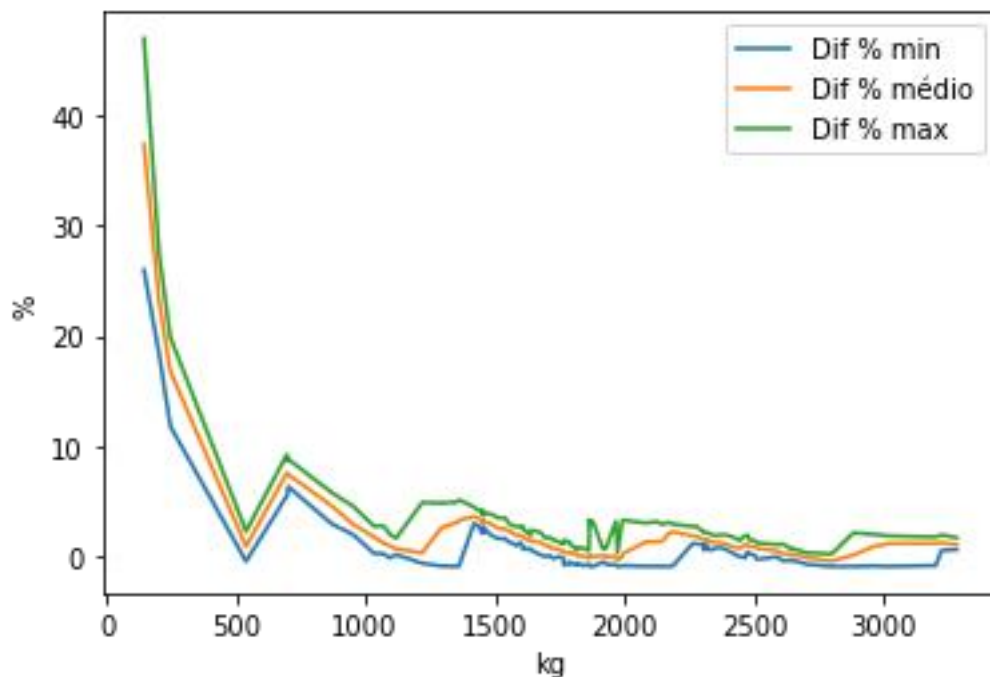
Conforme os resultados apresentados, pode-se realizar a avaliação final da comparação entre os modelos. Desta forma podemos indicar que tanto a regressão linear quanto a rede neural apresentaram bons resultados de predição quando aplicado a ambas, valores de *setpoint* entre a faixa de trabalho (1500 kg - 2000 kg) ou superior. Porém foi a rede neural que apresentou valores de diferenças máximas e mínimas sempre abaixo de 1%, ou seja, pouco mais de 99% de precisão nas dosagens, tendo o melhor desempenho entre os modelos utilizados. Enquanto a regressão apresentou diferenças máximas e mínimas, acima de 1%, entre 1,57% e 3,42%. Já a árvore de decisão foi o modelo que teve o pior desempenho, entregando como resultados valores destoantes para uma mesma faixa de dosagem. O que não permitiu extrair dados de precisão, o demonstra que não obteve um bom aprendizado do padrão contido na base de dados. Para cada conjunto de configuração utilizada foram realizadas 30 dosagens, gerando um total de 3 mil dosagens. Para melhor visualização os *setpoints* de dosagem foram agrupados por faixas de peso e foram selecionadas 15 amostras de 100 possíveis.

4.4.1 Regressão linear em ambiente de produção

Os resultados demonstram como se comportou o modelo em relação ao peso requerido na dosagem. Para *setpoints* até 500kg o erro médio obtido foi acima de 16%, tendo erro máximo de 47,08%. Conforme os pesos trabalhados aumentam,

os resultados melhoram. Entre 500kg e 1000 kg é possível perceber uma melhora significativa, com erro médio entre 2,8% e 7,5% aproximadamente. Dentro da faixa de trabalho proposta o erro médio esteve abaixo de 2,5%. Para *setpoints* acima de 2 mil kg o erro médio se manteve estável, muito próximo aos resultados da faixa de trabalho. Conforme Fig. 11 é possível visualizar graficamente o comportamento das dosagens realizadas a partir dos parâmetros gerados pelo modelo de regressão linear:

Figura 11: Resultado da diferença % - Regressão Linear



Fonte: Do autor (2022)

A leitura gráfica deixa o resultado mais claro, sendo possível visualizar como se comporta a precisão da dosagem através das faixas de *setpoint*. O eixo y representa o erro percentual em relação ao eixo x, que representa o *setpoint*.

A Tab. 9 demonstra os resultados de 15 dos 100 *setpoints* utilizados em ambiente de produção:

Tabela 9: Resultado da dosagem - Regressão Linear

Setpoint (kg)	Dosado média (kg)	Diferença % mínima	Diferença % média	Diferença % máxima
145.249	199.625	26.007	37.43	47.08
203.562	250.857	18.372	23.23	27.47
247.177	288.604	11.776	16.76	19.85
539.394	543.852	-0.531	0.82	2.21
696.730	749.068	5.43	7.51	9.19
870.861	910.174	2.87	4.51	5.72
953.149	980.165	1.90	2.83	4.47
1219.195	1222.133	-0.70	0.24	4.82
1419.270	1469.469	2.92	3.53	4.42
1539.518	1577.132	1.57	2.44	3.42
1638.439	1660.645	0.48	1.35	2.32
1748.296	1757.607	-0.08	0.53	1.20
1916.169	1916.509	-0.58	0.017	0.66
2417.073	2437.160	0.23	0.83	1.71
3201.714	3237.320	-0.90	1.11	1.67

Fonte: Do autor (2022)

Os valores demonstrados representam a distribuição dos setpoints utilizados. É possível perceber que conforme os valores de peso requisito aumentam os erros de dosagem diminuem. Cada dosagem aplicada em ambiente de produção teve seus dados de configuração preditos pelo modelo conforme demonstrado na Tab. 10:

Tabela 10: Configuração da dosagem - Regressão Linear

Setpoint (kg)	P_a (kg)	P_b (kg)	P_c (%)	P_d (%)	P_e (%)	P_f (kg)	P_g (s)
145.249	63.5	42.0	90.05	60.19	26.31	19.436	2.49
203.562	93.0	59.6	90.05	60.19	26.35	26.960	2.50
247.177	115.0	72.8	90.04	60.19	26.38	32.587	2.51
539.394	262.7	160.9	90.03	60.19	26.59	70.291	2.56
696.730	342.2	208.3	90.02	60.19	26.70	90.591	2.59
870.861	430.2	260.8	90.01	60.19	26.82	113.059	2.62
953.149	471.8	285.6	90.01	60.19	26.88	123.676	2.63
1219.195	606.2	365.8	89.99	60.20	27.07	158.003	2.68
1419.270	707.3	426.2	89.98	60.20	27.21	183.818	2.71
1539.518	768.1	462.4	89.98	60.20	27.29	199.333	2.73
1638.439	818.1	492.2	89.97	60.20	27.36	212.096	2.75
1748.296	873.6	525.4	89.97	60.20	27.44	226.271	2.77
1916.169	958.4	576.0	89.96	60.20	27.56	247.931	2.80
2417.073	1211.6	727.0	89.93	60.21	27.91	312.560	2.89
3201.714	1608.1	963.6	89.89	60.21	28.47	413.799	3.03

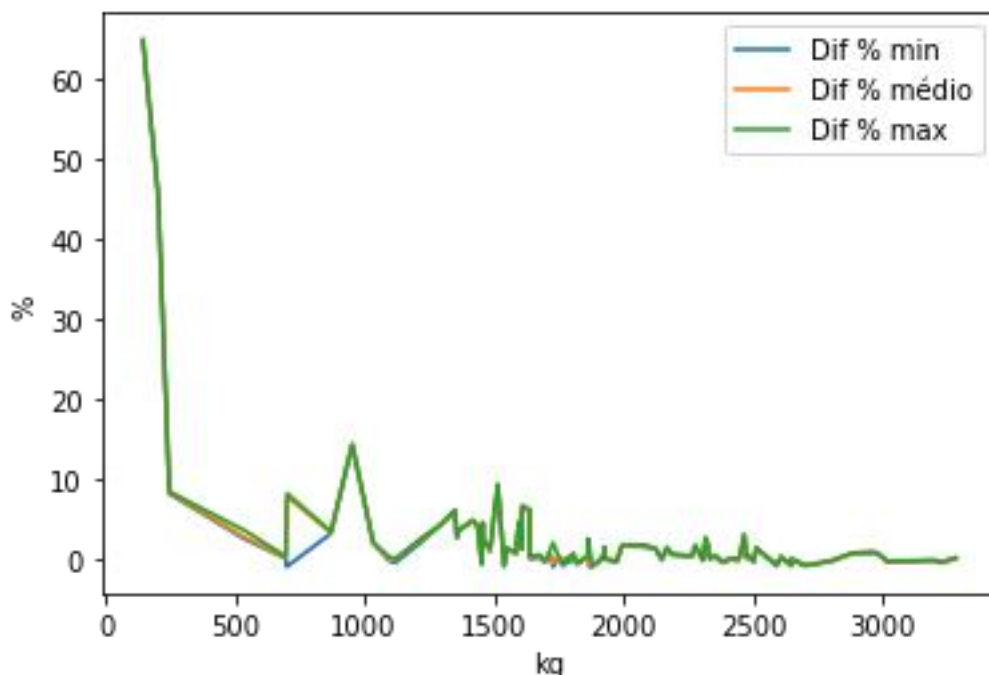
Fonte: Do autor (2022)

É importante perceber que os dados de saída representados pelos parâmetros P_a , P_b , P_c , P_d , P_e , P_f e P_g tem um comportamento linear, ou seja, segue uma linha de tendência com referência no *setpoint*. Assim como as informações demonstradas na Fig. 11 e Tab. 9 os dados demonstram estabilidade, ou seja não existe uma aleatoriedade nos resultados o que demonstra novamente que o modelo entendeu os padrões da base de dados.

4.4.2 Rede neural artificial em ambiente de produção

Para *setpoints* abaixo 500kg o erro médio obtido foi entre 8% e 64,7%, tendo erro máximo de 65,03% aproximadamente. Conforme os pesos trabalhados aumentam, os resultados melhoram. Entre 500kg e 1000 kg é possível perceber uma melhora significativa, com erro médio entre 2,6% e 3,2% aproximadamente. Dentro da faixa de trabalho proposta o erro médio esteve abaixo de 1%. Para *setpoints* acima de 2 mil kg o erro médio se manteve estável, muito próximo aos resultados da faixa de trabalho. Conforme Fig. 12 é possível visualizar graficamente o comportamento das dosagens realizadas a partir dos parâmetros gerados pelo modelo de regressão linear.

Figura 12: Resultado da diferença % - Rede Neural Artificial



A leitura gráfica deixa o resultado mais claro, sendo possível visualizar como se comporta a precisão da dosagem através das faixas de *setpoint*. O eixo y representa o erro percentual em relação ao eixo x, que representa o *setpoint*. A Tab. 11 demonstra os resultados de 15 dos 100 *setpoints* utilizados em ambiente de produção.

Tabela 11: Resultado da dosagem - Rede Neural Artificial

Setpoint (kg)	Dosado média (kg)	Diferença % mínima	Diferença % média	Diferença % máxima
145.249	239.510	64.73	64.89	65.03
203.562	296.408	45.48	45.61	45.77
247.177	267.685	8.15	8.29	8.39
539.394	553.769	2.45	2.66	3.50
696.730	697.815	0.04	0.15	0.27
870.861	899.366	3.16	3.27	3.40
953.149	1089.238	14.19	14.27	14.35
1219.195	1249.015	2.08	2.44	2.54
1419.270	1487.434	4.76	4.80	4.85
1539.518	1525.866	-0.95	-0.88	-0.81
1638.439	1642.588	-0.01	0.25	0.3
1748.296	1753.716	0.05	0.30	0.36
1916.169	1914.831	-0.14	-0.06	0.06
2417.073	2416.493	-0.09	-0.02	0.04
3201.714	3195.138	-0.24	-0.20	-0.14

Fonte: Do autor (2022)

Os valores demonstrados representam a distribuição dos *setpoints* utilizados. É possível perceber que conforme os valores de peso requisito aumentam os erros de dosagem diminuem. Cada dosagem aplicada em ambiente de produção teve seus dados de configuração preditos pelo modelo conforme demonstrado na Tab. 12:

Tabela 12: Configuração da dosagem - Rede Neural Artificial

Setpoint (kg)	P_a (kg)	P_b (kg)	P_c (%)	P_d (%)	P_e (%)	P_f (kg)	P_g (s)
145.249	13.5	4.6	89.7	56.0	25.0	2.5	2.2
203.562	10.2	8.7	89.0	62.9	30.05	-2.1	2.5
247.177	80.6	56.3	92.8	64.3	26.3	13.3	1.8
539.394	209.5	156.1	89.5	60.6	25.1	45.8	1.9
696.730	271.7	175.5	90.0	59.9	26.4	82.9	2.8
870.861	392.2	242.1	89.8	59.2	25.6	94.9	2.1
953.149	350.1	210.5	91.4	54.2	35.9	67.4	4.4
1219.195	585.2	353.6	89.4	60.4	24.7	151.9	2.2
1419.270	642.1	391.8	90.0	61.2	29.9	163.6	3.4
1539.518	697.3	428.7	90.3	61.3	28.1	204.8	3.1
1638.439	755.2	461.4	90.2	60.8	27.0	209.9	2.6
1748.296	843.7	503.8	90.0	59.8	25.3	192.6	2.4
1916.169	926.3	547.8	90.3	59.8	25.8	214.1	2.5
2417.073	1188.2	697.9	90.2	60.8	25.4	324.1	2.3
3201.714	1509.6	907.0	90.0	59.8	28.2	351.6	2.6

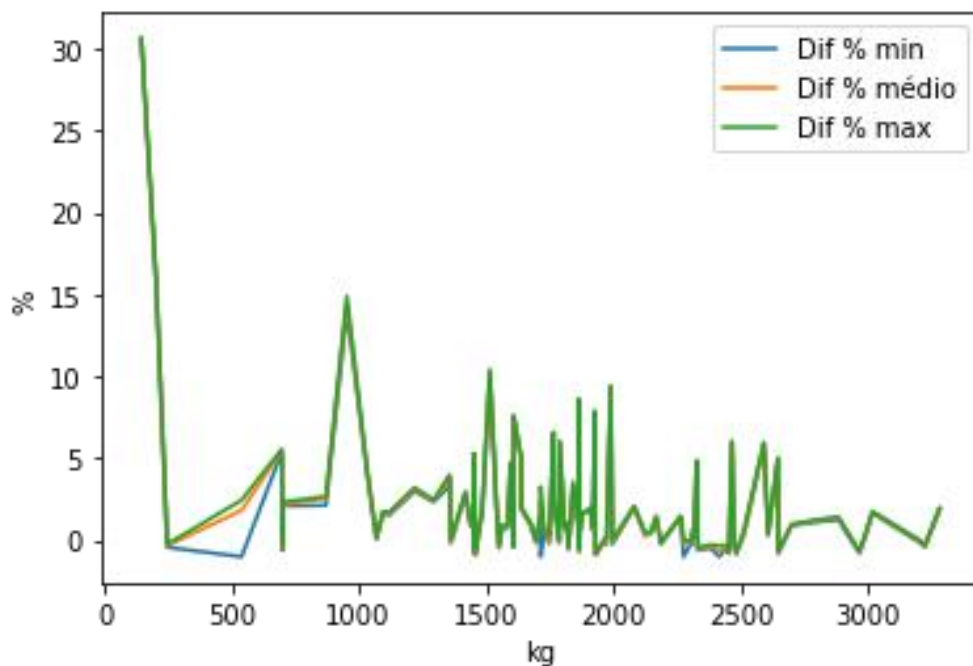
Fonte: Do autor (2022)

É importante perceber que os dados de saída representados pelos parâmetros P_a, P_b, P_c, P_d, P_e, P_f e P_g tem um comportamento linear, ou seja, segue uma linha de tendência com referência no *setpoint*. Assim como as informações demonstradas na Fig. 12 e Tab. 11 os dados demonstram estabilidade, ou seja, não existe uma aleatoriedade nos resultados o que demonstra novamente que o modelo entendeu os padrões da base de dados.

4.4.3 Árvore de decisão em ambiente de produção

Para *setpoints* até 500kg o erro médio obtido varia entre 1,82% a 30,58% aproximadamente, tendo erro máximo também de 30,58%. Diferente dos demais modelos, conforme os pesos trabalhados aumentam, não há uma estabilidade no erro, seguiram ocorrendo resultados aleatórios, hora bons, hora ruins. Entre 500kg e 1000kg, o erro médio ocorre entre 1,82% e 14,5%. Conforme Fig. 13 é possível visualizar graficamente que as dosagens realizadas a partir dos parâmetros gerados pelo modelo de árvore de decisão não mantiveram um resultado linear, onde em valores de *setpoint* próximos os resultados foram muito diferentes. Desta forma as análises de precisão ficaram comprometidas.

Figura 13: Resultado da diferença % - Árvore de Decisão



Fonte: Do autor (2022)

A leitura gráfica deixa o resultado mais claro, sendo possível visualizar como se comporta a precisão da dosagem através das faixas de *setpoint*. O eixo y representa o erro percentual em relação ao eixo x, que representa o *setpoint*. A Tab. 13 demonstra os resultados de 15 dos 100 *setpoints* utilizados em ambiente de produção.

Tabela 13: Resultado da dosagem - Árvore de Decisão

Setpoint (kg)	Dosado média (kg)	Diferença % mínima	Diferença % média	Diferença % máxima
145.249	189.679	30.58	30.58	30.58
203.562	235.597	15.66	15.73	15.81
247.177	246.368	-0.42	-0.32	-0.24
539.394	549.246	-0.99	1.82	2.45
696.730	734.757	5.36	5.45	5.54
870.861	893.144	2.13	2.55	2.67
953.149	1090.9471	13.96	14.45	14.88
1219.195	1256.636	2.98	3.07	3.21
1419.270	1460.249	2.80	2.88	2.97
1539.518	1577.970	2.45	2.49	2.55
1638.439	1670.659	1.889	1.96	2.00
1748.296	1746.805	-0.19	-0.08	0.19
1916.169	1950.836	1.73	1.80	2.03
2417.073	2401.716	-0.99	-0.63	-0.31
3201.714	3200.776	-0.10	-0.02	0.03

Fonte: Do autor (2022)

Os valores demonstrados representam a distribuição dos *setpoints* utilizados. É possível perceber que conforme os valores de peso requisito aumentam os erros de dosagem não se comportam com a mesma estabilidade dos demais modelos. Cada dosagem aplicada em ambiente de produção teve seus dados de configuração preditos pelo modelo conforme demonstrado na Tab. 14:

Tabela 14: Configuração da dosagem - Árvore de Decisão

Setpoint (kg)	P_a (kg)	P_b (kg)	P_c (%)	P_d (%)	P_e (%)	P_f (kg)	P_g (s)
145.249	55.5	33.7	80.0	41.0	40.0	22.8	1.4
203.562	81.1	53.7	82.0	67.0	26.0	24.1	2.1
247.177	105.7	61.3	91.0	56.0	24.0	28.8	2.8
539.394	215.8	149.7	83.0	48.0	18.0	49.1	1.0
696.730	253.6	199.5	93.0	41.0	28.0	110.2	1.8
870.861	350.5	307.9	86.0	59.0	19.0	63.7	3.7
953.149	502.8	227.6	97.0	57.0	37.0	69.3	4.6
1219.195	437.1	330.4	96.0	73.0	25.0	35.6	2.4
1419.270	541.2	421.9	98.0	68.0	35.0	170.5	2.4
1539.518	955.5	521.7	91.0	40.0	36.0	357.8	3.0
1638.439	812.0	563.0	83.0	78.0	37.0	71.2	1.9
1748.296	994.9	518.0	82.0	67.0	31.0	193.3	2.6
1916.169	740.1	600.6	99.0	42.0	25.0	481.7	3.3
2417.073	1170.9	804.7	96.0	58.0	19.0	145.8	0.9
3201.714	1823.9	1000.9	88.0	77.0	19.0	491.5	2.1

Fonte: Do autor (2022)

É importante perceber que os dados de saída representados pelos parâmetros P_a, P_b, P_c, P_d, P_e, P_f e P_g tem um comportamento não linear. Assim com as informações demonstradas na Fig. 13 e Tab 13 os dados demonstram certa aleatoriedade nos resultados o que demonstra que o modelo não entendeu os padrões da base de dados.

4 CONCLUSÃO

A partir dos resultados obtidos pode-se afirmar que o objetivo proposto neste trabalho foi atingido, visto que foram alcançados ótimos índices de precisão nas dosagens testadas a partir de parâmetros sugeridos pelos modelos de aprendizado de máquina. Logo utilizando esta abordagem a equipe de produção pode testar de forma simples, qual a melhor configuração de dosagem para um valor

específico de *setpoint*, antes de colocá-lo em produção, evitando assim desperdícios e demora para encontrar o melhor ajuste, ou seja, redução de custos na operação.

Durante o desenvolvimento do trabalho, percebemos que existem grandes oportunidades de aplicação de machine learning em soluções industriais, mas os desafios e dificuldades parecem ser os mesmos. Isso por que o que chamamos atualmente de “indústria 3.0”, onde o foco foi automatizar equipamentos e processos, passou a gerar muitos dados, porém não se teve a preocupação de estruturar os dados gerados para que eles pudessem ser interpretados da forma que se propõem hoje (indústria 4.0).

Ainda que a proposta deste trabalho teve como foco principal sistemas de dosagens para fabricas de nutrição animal, ele pode ser aplicado a qualquer processo que utilize sistemas de dosagem por batelada. A proposta ainda pode ir além. Conhecendo as características do sistema de dosagem em condições normais, sistemas de predição para manutenção podem também ser implementados, já que ao detectar valores fora do esperado pode-se determinar a necessidade de manutenção.

Porém, como mencionado os dados precisarão passar por uma reorganização no seu armazenamento. Tudo isso para que tenhamos melhor desempenho na coleta, redução dos esforços no pré-processamento, dos recursos computacionais. Por fim, analisar se os dados salvos atualmente ainda fazem sentido e avaliar se novas informações devem ser armazenadas, isto pensando na limpeza ou enriquecimento da base de dados.

REFERÊNCIAS

Alves, Reinaldo Galo Febrônio; et. al. Dosagem programada de matéria-prima para os diversos processos da indústria. **Revista Eletrônica eF@tec**. São Paulo, v. 6, n. 1, 2016.

BURKOV, Andriy. The Hundred-Page Machine Learning Book. **Andriy Burkov**, 2019.

AMMANN, Alex. Dosing & Weighing. **Shule für futtermitteltechnik**, 2020.(Material fornecido no treinamento)

DUARTE, Karina Ferreira. **Indústria de ração**: Quais os maiores desafios e oportunidades? S.d. Disponível em: <https://nutricaoesaudeanimal.com.br/industria-de-racao>. Acesso em: 29, setembro 2022.

FEHRENBACH, Rodrigo Bonemann. **Desenvolvimento de sistema automatizado para controle de dosagem de ração animal**. 2017. Disponível em: <http://hdl.handle.net/10737/1665>. Acesso em: 02, novembro 2021.

FREITAS, A. L.; SANTANA JUNIOR. O. V. Machine learning: desafios para um brasil competitivo. **Computação Brasil**, Porto Alegre, v. 39, p. 8, jan. 2019. Disponível em: https://www.sbc.org.br/images/flippingbook/computacaobrasil/computa_39/pdf/Comp_Brasil_39_180.pdf. Acesso em: 23, novembro 2022.

GÉRON, Aurélien. **Hands-On Machine Learning With Scikit-Learn & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. O'Reilly Media, 2017.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. [S.l.]:MIT press, 2016.

LEMONS, Julio Cesar Lopes. **Otimização da dosagem de matérias-primas na produção do vidro plano incolor**. 2016. Disponível em: <https://www.aedb.br/seget/arquivos/artigos16/22324235.pdf>. Acesso em: 04, setembro 2021.

OLIVO, Rubison. **O mundo do frango**: Cadeia produtiva da carne de frango. Livraria Varela, Criciúma, 2006.

RICARDO, José Luís Anselmo. **Aplicações dos Sistemas de Pesagem na Indústria das Rações**. 2017. Disponível em: <http://hdl.handle.net/10737/1665>. Acesso em: 02, novembro 2021.

SHIMAKURA, Silvia Emiko. **Interpretação do coeficiente de correlação**. 2006. Disponível em: <http://leg.ufpr.br/~silvia/CE003/node74.html>. Acesso em: 06, novembro 2022.