

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Estimativa de desperdício de alimentos com o uso de
Aprendizado de Máquina: um estudo de caso numa
empresa brasileira de serviços de refeições coletivas**

Jorge Luís Cordenonsi

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Jorge Luís Cordenonsi

Estimativa de desperdício de alimentos com o uso de Aprendizado de Máquina: um estudo de caso numa empresa brasileira de serviços de refeições coletivas

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Prof. Dr. Heloísa de Arruda Camargo

Versão original

São Carlos

2024

Jorge Luís Cordenonsi

**Estimativa de desperdício de alimentos com o uso de
Aprendizado de Máquina: um estudo de caso numa
empresa brasileira de serviços de refeições coletivas**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Heloísa de Arruda Camargo

Original version

São Carlos

2024

RESUMO

Cordenonsi, J. L. **Estimativa de desperdício de alimentos com o uso de Aprendizado de Máquina: um estudo de caso numa empresa brasileira de serviços de refeições coletivas**. 2024. 50p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Desperdício de alimentos é um tema cada vez mais crítico na agenda dos administradores de empresas de serviços de alimentação, pois além de representar uma possibilidade de redução de custos, está relacionado à governança corporativa e ao programa de ESG (*Environmental, Social, Governance*) que praticamente é uma exigência de governança principalmente para as empresas de capital aberto.

Este trabalho possuiu como objetivo o desenvolvimento de um modelo de estimativa de desperdícios de alimentos com o uso de tecnologias de Inteligência Artificial e Big Data. Este desenvolvimento foi elaborado com base no uso de técnicas e algoritmos de Aprendizado de Máquina, bem como de técnicas estatísticas de correlação de dados. O foco foi de analisar o desperdício de alimentos na fase na cadeia de produção dos alimentos pelos restaurantes da empresa usada como estudo de caso.

Constatamos que a estimativa de desperdício futuros com o uso de aprendizado de máquina é possível implementar um sistema de gestão que possa ser a base de um melhor planejamento para a redução de desperdícios. Essa foi a principal contribuição a ser incorporada no modelo atual da empresa de administração dos desperdícios de alimentos.

Palavras-chave: Administração de desperdício de alimentos; Aprendizado de Máquina; Inteligência Artificial; Modelo de estimativa de desperdício de alimentos.

ABSTRACT

Cordenonsi, J. L. **Food waste estimation using Machine Learning: a case study in a brazilian food services company**. 2024. 50p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Food waste is an increasingly critical topic on the agenda of any food service company, as in addition to representing a possibility of cost reduction, its relation with corporate governance and the ESG program (*Environmental, Social, Governance*) which is practically a governance requirement mainly for public traded companies.

This research objective was to develop a food waste estimation model using Artificial Intelligence and Big Data technologies. This development was based on the use of Machine Learning techniques and algorithms, as well as statistical data techniques. The focus was to analyze food waste in the food production chain phase by the company's restaurants used as a case study.

We found that, by estimating future food waste based on machine learning algorithms, it is possible to deploy a management system that can be the basis for a better planning to reduce food waste. This was the main contribution of this work to be incorporated into the company's current food waste management model.

Keywords: Food waste management; Machine Learning; Artificial Intelligence; Food waste Estimation Model

LISTA DE FIGURAS

Figura 1 – Histograma – Largura de Pétalas com 10 e 5 cestas - adaptado de (ROMERO, 2023)	19
Figura 2 – Boxplot - Pressão Sistólica - adaptado de (PERES, 2024)	20
Figura 3 – Hierarquia dos tipos de aprendizado de máquina - adaptado de (FACELI <i>et al.</i> , 2021)	21
Figura 4 – Sistema de Aprendizado de Máquina - adaptado de (KUNG, 2014)	21
Figura 5 – Algoritmos de Aprendizado de Máquina - adaptado de (AMBIKA, 2020)	21
Figura 6 – Ilustração simplificada do procedimento realizado pelo SVM - adaptado de (FACELI <i>et al.</i> , 2021)	24
Figura 7 – Representação de uma RNA (Rede Neural Artificial) - adaptado de (SILVA, 2023)	25
Figura 8 – Cadeia de Geração de Desperdício de Alimentos (AUTOR)	27
Figura 9 – Diagrama de Blocos o desenvolvimento do Modelo de Estimativa de Desperdício de Alimentos - adaptado de (REZENDE, 2003)	31
Figura 10 – DashBoard de Controle de Desperdício de Alimentos - (AUTOR)	34
Figura 11 – Estrutura do conjunto de dados de desperdício de alimentos	37
Figura 12 – Descrição inicial do conjunto de dados de desperdícios	38
Figura 13 – Dados nulos identificados no conjunto de dados	38
Figura 14 – Descrição ajustada dos dados de desperdícios	39
Figura 15 – Boxplot da quantidade total de desperdício de alimentos	40
Figura 16 – Histograma da quantidade total de desperdícios de alimentos	40
Figura 17 – Total de desperdícios de alimentos por região	41
Figura 18 – Estatística do atributo chave: quantidade total de desperdícios	42
Figura 19 – Gráfico de dispersão dos dados de desperdícios de alimentos	42
Figura 20 – Resultados dos erros dos algoritmos de aprendizado de máquina	44
Figura 21 – Diferença percentual por região da quantidade de desperdícios planejada e realizada	44

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Contextualização, Motivação e Lacunas	13
1.2	Questões e Objetivos	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Conceitos	17
2.1.1	Exploração de dados	17
2.1.1.1	Média, Mediana e Percentil	17
2.1.1.2	Desvio Padrão, Variância, Skewness (Obliquidade) e Kurtosis (Curtose)	18
2.1.2	Aprendizado de Máquina	19
2.1.2.1	Árvore de Decisão	22
2.1.2.2	KNN - <i>K-Nearest Neighbors</i>	22
2.1.2.3	SVM - <i>Support Vector Machine</i>	23
2.1.2.4	Regressão Linear	23
2.1.2.5	Random Forest	24
2.1.2.6	Redes Neurais	25
2.1.3	Trabalhos relacionados com o tema desenvolvido	25
3	METODOLOGIA E DESENVOLVIMENTO	31
3.1	Metodologia definida para o desenvolvimento do modelo de estima- tiva de desperdício de alimentos	31
3.1.1	Coleta de dados	31
3.1.2	Exploração de dados	32
3.1.3	Pré-processamento dos dados	32
3.1.4	Processamento dos dados	32
3.1.5	Pós-Processamento dos dados	32
3.1.6	Utilização do conhecimento	33
3.2	Desenvolvimento	33
3.2.1	Descrição do estudo de caso	33
3.2.2	Aplicação da metodologia	34
4	ANÁLISE DOS DADOS E RESULTADOS	37
4.1	Conjuntos de dados	37
4.2	Análise exploratória do atributo chave	40
4.3	Análise dos resultados dos algoritmos de aprendizado de máquina	41
5	CONCLUSÕES	47

Referências	49
------------------------------	-----------

1 INTRODUÇÃO

1.1 Contextualização, Motivação e Lacunas

De acordo com GARRE, RUIZ e HONTORIA (2020), PRINCIPATO *et al.* (2023), BUSTAMANTE, AFONSO e RIOS (2018), DELIBERADOR *et al.* (2021), ENNIS (2023), YUSOFF, WOOLLEY e GODSELL (2023), 1/3 da produção de alimentos para seres humanos são desperdiçados no mundo o que gera inúmeros desafios na administração da cadeia de produção e abastecimento, cerca de 1300 milhões de toneladas de desperdício de alimentos por ano.

Desperdício de alimentos é um tema cada vez mais crítico na agenda dos administradores de empresas de serviços de alimentação, pois além de representar uma possibilidade de redução de custos, está relacionado à governança corporativa e ao programa de ESG (*Environmental, Social, Governance*) que praticamente é uma exigência de governança principalmente para as empresas de capital aberto.

BULHÕES *et al.* (2022) aponta num estudo feito que ter uma solução para o desperdício de alimentos é complexo, pois envolve uma análise de diversas dimensões. Além disso, destaca que as causas principais atribuídas ao desperdício de alimentos são: tamanhos dos pratos usados para servir, tamanhos das porções servidas que não seguem estimativas precisas, ausência de métodos de classificação dos alimentos e ausência de campanhas de conscientização.

A gestão do desperdício de alimentos exige muita disciplina para acompanhar de maneira detalhada tanto o aspecto de execução de uma operação, como também identificar as variáveis envolvidas e causas dos desperdícios que podem estar do lado da própria empresa fornecedora dos serviços de alimentação, quanto dos próprios usuários do serviço, ou de ambos.

Para entender estas variáveis, o problema e possíveis soluções este trabalho foi desenvolvido para analisar a operação de uma empresa com operação no Brasil que faz parte de um grupo internacional britânico, líder global do setor de serviços de alimentação e que está presente em cerca de 50 países. Esta empresa possui cerca de 30000 funcionários e possui cerca de 2000 unidades operacionais localizadas em vários Estados brasileiros.

Um dos aspectos críticos de um negócio de serviços de refeições coletivas é garantir a qualidade e segurança dos serviços prestados. Estas disciplinas são tradicionalmente endereçadas pelas empresas através de processos, tecnologia, pessoas e dados. Na dimensão dos processos, os processos de planejamento, compras, logística e produção são os mais importantes. No elemento de tecnologia os processos normalmente são automatizados com o uso de sistemas de informações desenvolvidos pela própria empresa ou o uso de soluções

de software prontas de mercado.

As pessoas são compostas predominantemente por profissionais em cada área de especialização, porém destaco a função do nutricionista, que é o responsável em garantir um planejamento de cardápios eficiente para atender às operações das unidades operacionais (restaurantes) clientes da empresa. O processo de planejamento é o mais importante, pois ele é o responsável em definir o cardápio que será executado diariamente nos restaurantes.

O restaurante deve seguir obrigatoriamente um contrato definido que estabelece de uma forma geral os tipos de ingredientes, marcas, quantidades e restrições gastronômicas específicas para a operação. De uma forma padronizada, o restaurante é administrado por um nutricionista que é o responsável pela operação do dia-a-dia. A produção do restaurante segue o planejamento de cardápio definido e o nutricionista deve garantir a entrega do serviço de acordo com o estabelecido em contrato.

Além da importância da garantia de um estoque de matéria-prima suficiente para garantir a produção necessária, o nutricionista deve comandar a produção de acordo com o plano de quantidade de pessoas que serão servidas. De uma forma geral os serviços são divididos em almoço, jantar, e ceia, porém isso pode variar entre os restaurantes. Importante destacar que cada restaurante não segue uma padronização de equipamentos, infraestrutura, capacidade logística, pois isso é de acordo com a operação do cliente.

A empresa fornecedora dos serviços de refeições coletivas instala uma infraestrutura básica de conectividade, hardware, software para que a unidade de operação local possa usar os sistemas de informação necessários para a operação diária. Um dos problemas críticos identificados nesta empresa é a dificuldade em gerenciar cerca de 700.000 receitas, 65.000 produtos e 6.000 fornecedores para atender 2.000 unidades de negócio (Restaurantes) em vários Estados. Cardápios e receitas customizadas com custo alto de matéria-prima e com uma logística de distribuição complexa.

Atualmente, o processo existente consiste na coleta de dados das unidades operacionais, através da pesagem diária dos desperdício de alimentos, sobra de alimentos, que ficam nas rampas de produção, nos pratos dos usuários e também sobras na linha de produção. Estes dados são armazenados numa base de dados de desperdício de alimentos e relatórios são gerados, através do uso de uma ferramenta de apresentação de dados, o *Microsoft Power BI*. Os administradores usam estes dados e geram gráficos para análise dos desperdícios, porém não existe um modelo de estimativa que possa auxiliar tanto a unidade operacional como os administradores a desenvolver campanhas ou ações para a redução do desperdício de alimentos.

Neste cenário, o problema analisado neste trabalho foi o de desenvolver um modelo de estimativa do desperdício de alimentos com base no processo e sistema atual usado pela empresa.

Este modelo de estimativa de desperdício de alimentos poderá auxiliar com que a empresa possa melhorar as suas margens operacionais, através de uma redução dos custos de matéria-prima que é um dos custos críticos, além do custo de mão-de-obra. Atualmente, o custo de matéria-prima gerado pelos desperdícios de alimentos representa na empresa cerca de 2 (dois) por cento do custo total de matéria-prima. Isso impacta diretamente o resultado operacional da empresa que possui uma margem operacional bruta de cerca de 6 (seis) por cento.

1.2 Questões e Objetivos

Este trabalho buscou responder as seguintes questões chave no uso de tecnologias de inteligência artificial e Big Data para a redução de desperdício de alimentos:

a) Como reduzir o desperdício de alimentos, através do uso e apoio das tecnologias de inteligência artificial e Big Data?

b) Quais são os dados fundamentais que devem ser coletados e analisados de maneira frequente para que seja alcançado uma melhoria no sistema de controle de desperdício de alimentos?

c) Como estimar o volume de desperdícios das unidades operacionais (restaurantes) para poder tomar decisões proativas no processo de planejamento e/ou ações de conscientização dos usuários?

d) Quais são as variáveis fundamentais que determinam o volume de desperdício de alimentos?

Este trabalho possuiu como objetivo o desenvolvimento de um modelo de estimativa de desperdícios de alimentos com o uso de tecnologias de Inteligência Artificial e Big Data. Este desenvolvimento foi elaborado com base no uso de técnicas e algoritmos de Aprendizado de Máquina, bem como de técnicas estatísticas de correlação de dados.

O modelo desenvolvido identificou as variáveis fundamentais para que se melhore a administração atual de desperdício de alimentos, assim como ter uma estimativa de desperdícios e com isso seja possível desenvolver campanhas de conscientização junto aos usuários dos restaurantes, bem como iniciativas de melhorias nos processos operacionais junto aos colaboradores com o objetivo de reduzir o desperdício de alimentos.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo descrevemos os conceitos, modelos e técnicas de aprendizado de máquina que foram utilizados para o desenvolvimento do modelo de controle de desperdício de alimentos. Além disso, fizemos uma análise dos trabalhos relacionados com o tema desenvolvido.

2.1 Conceitos

A Inteligência Artificial tem sido amplamente usada como forma de apoio aos processos de predição para atividades críticas de negócio. Estimar cenários futuros de apoio a um processo de decisão é um elemento cada vez importante no dia-a-dia dos administradores (FERREIRA *et al.*, 2022).

Segundo VIEIRA *et al.* (2021), não é uma tarefa simples pesquisar sobre o controle de desperdício de alimentos. Como resultado da Covid-19, a cadeia de suprimentos foi alterada e gerou impactos no fornecimento de alimentos no mundo, pois temos muitas diferenças culturais, geográficas e tecnológicas aplicadas pelas organizações.

Existe uma perspectiva que deve ser analisada na ótica do consumidor e da cadeia de suprimentos. Do lado do consumidor são necessárias campanhas de conscientização para a redução do desperdício dos alimentos e do lado da cadeia de suprimentos existe uma preocupação na análise dos impactos ambientais do desperdício dos alimentos.

Para este trabalho foi usado alguns conceitos de Estatística e, principalmente de uma área da Inteligência Artificial conhecida como Aprendizado de Máquina.

2.1.1 Exploração de dados

Para a exploração de dados usamos os conceitos de Média, Mediana e Percentil. Isso foi usado para o entendimento e análise inicial dos dados coletados. (FACELI *et al.*, 2021).

2.1.1.1 Média, Mediana e Percentil

O valor da média de um conjunto de dados com n elementos é calculado pela seguinte equação:

$$\text{Média} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

Para o cálculo do valor da mediana se usa a seguinte equação:

$$\text{Mediana} = \begin{cases} x_{\frac{n+1}{2}} & \text{se } n \text{ é ímpar} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{se } n \text{ é par} \end{cases} \quad (2.2)$$

A mediana corresponde ao valor do meio do conjunto de dados ordenado.

O percentil é calculado após o conjunto de dados ser ordenado. Ele é calculado pela seguinte equação:

$$\text{Percentil de um valor } k = \left(\frac{k}{n} \right) \times 100 \quad (2.3)$$

Onde k é o número de observações menores ou iguais ao valor dado de k e n é o total de observações.

2.1.1.2 Desvio Padrão, Variância, Skewness (Obliquidade) e Kurtosis (Curtose)

O desvio padrão é uma medida de espalhamento de um conjunto de dados que mede o quanto um valor difere, em média, da média desse conjunto de dados.

Ele é calculado pela seguinte equação:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.4)$$

Onde σ é o desvio padrão do conjunto de dados, n é a quantidade de observações, x_i valor de cada observação e $\bar{x}$ é a média das observações. Um desvio padrão baixo indica que os valores do conjunto tendem a estar próximos da média.

A variância é a medida mais utilizada para avaliar o espalhamento dos dados e é calculada através da seguinte equação:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.5)$$

Onde σ^2 é a variância, n é o número de observações, x_i é o valor de cada observação e $\bar{x}$ é a média das observações.

O Skewness (Obliquidade) mede a simetria do espalhamento dos dados em torno da média e é calculada pela seguinte equação:

$$\text{Skewness (Obliquidade)} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3 \quad (2.6)$$

Onde o *Skewness(Obliquidade)* é a simetria, n é o número de observações, x_i é o valor de cada observação, $\bar{x}$ é a média das observações e o σ é o desvio padrão.

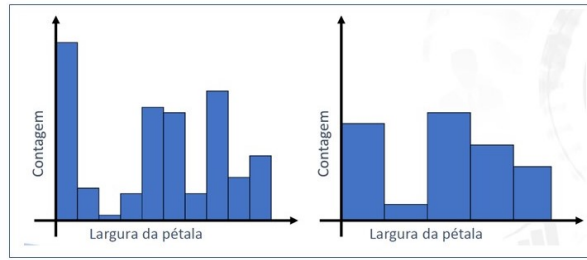


Figura 1 – Histograma – Largura de Pétalas com 10 e 5 cestas - adaptado de (ROMERO, 2023)

A Kurtosis (Curtose) é a medida de achatamento da função de distribuição. Se utiliza para verificar se os dados apontam para um pico ou são achatados em relação a uma distribuição normal. O cálculo é feito pela seguinte equação:

$$\text{Kurtosis (Curtose)} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - 3 \quad (2.7)$$

Onde n é o número de observações, x_i é o valor de cada observação, $\bar{x}$ é a média das observações e o σ é o desvio padrão.

Para a visualização dos dados deste trabalho usamos algumas técnicas de visualização, tais como o Histograma e Boxplot. A figura 1 ilustra um exemplo de Histograma, o primeiro gráfico mostra a quantidade de pétalas, distribuídas em 10 cestas e o segundo gráfico ilustra a quantidade de pétalas em 5 cestas. A figura 2 ilustra um exemplo de Boxplot.

O Histograma divide os valores de um conjunto de dados em grupos representados por barras que têm tamanho proporcional ao número de elementos de cada grupo. O Boxplot permite a visualização do primeiro, segundo (mediana), terceiro quartis e os limites inferiores e superiores dos dados.

2.1.2 Aprendizado de Máquina

De acordo com KUNG (2014), o aprendizado de máquina, na maioria dos métodos, consiste na extração de conhecimento genérico a partir de dados. Os algoritmos treinam os modelos, que depois serão utilizados para realizar alguma tarefa. Essas tarefas podem ser de diferentes tipos: preditivas ou descritivas e resolvem diferentes tipos de problemas. Este conhecimento adquirido é usado para auxiliar no processo de tomada de decisão.

Segundo FACELI *et al.* (2021), os algoritmos de aprendizado de máquina podem ser divididos para a execução de tarefas preditivas e descritivas. A tarefa preditiva possui duas características principais: o tipo do conjunto de dados, que deve ter um atributo alvo, e o tipo do modelo gerado, que é um modelo preditivo que vai prever um valor para o atributo alvo, com base no conhecimento extraído dos dados de treinamento. As etapas de

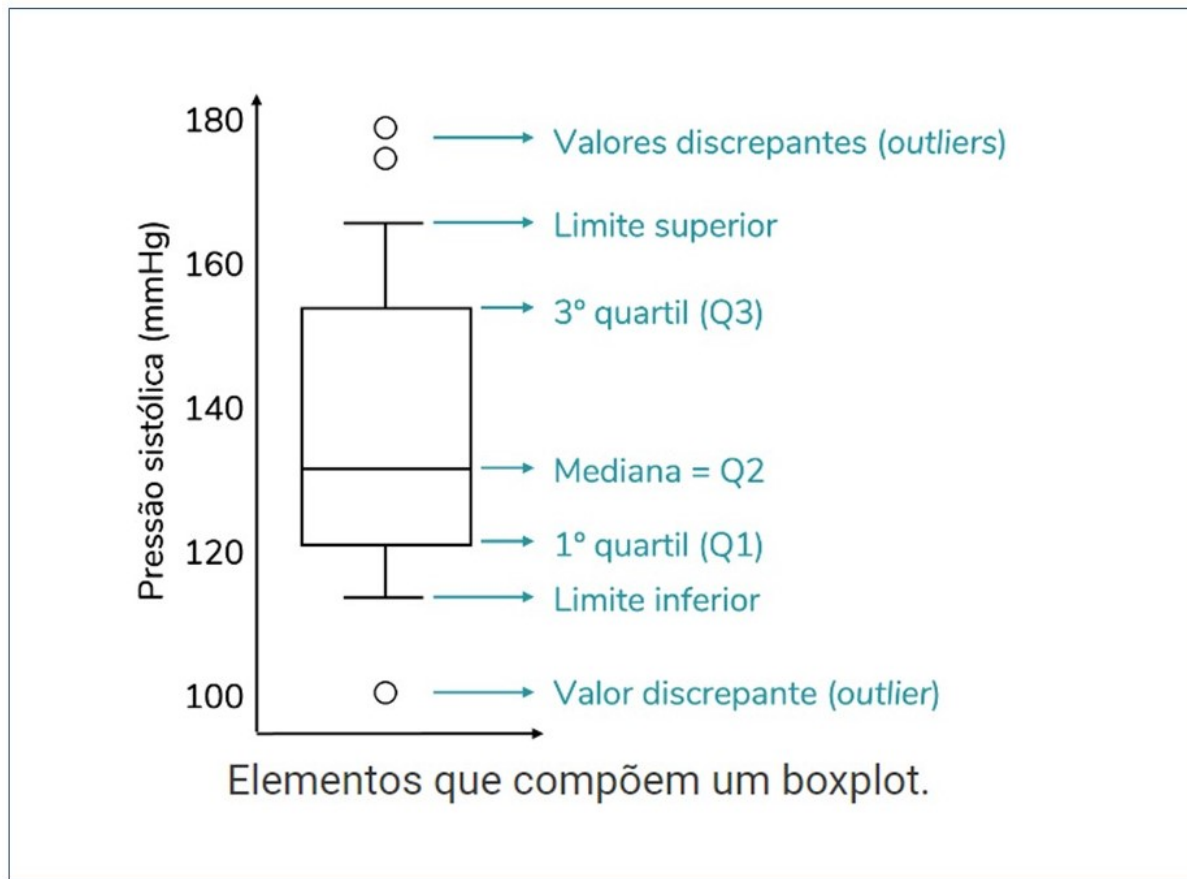


Figura 2 – Boxplot - Pressão Sistólica - adaptado de (PERES, 2024)

treinamento e uso dos modelos são distintas. Primeiro é feito o treinamento, depois o teste dos dados. Por fim, o modelo pode ser colocado para uso para fazer as previsões. Esses algoritmos realizam o aprendizado chamado de supervisionado.

Para as tarefas descritivas os algoritmos operam de maneira diferente e possuem também duas características principais: o conjunto de dados, que agora é não rotulado, porque não tem o atributo alvo, e o tipo de modelo, que não vai fazer previsões e sim estruturas descritivas e relacionamentos encontrados nos dados. Eles atuam para obter padrões descritivos e realizam o aprendizado dito não supervisionado.

A figura 3 ilustra a hierarquia dos tipos de aprendizado de máquina para a execução de tarefas preditivas e descritivas.

A figura 4 ilustra um sistema de aprendizado de máquina que consiste na geração do modelo a partir de um conjunto de dados com o posterior uso do modelo que pode ser preditivo ou descritivo sendo que os resultados são usados para dar suporte a tomada de decisões.

O foco deste trabalho foi o de explorar a execução de tarefas preditivas, mais especificamente de regressão onde o atributo alvo é contínuo, diferente dos casos para

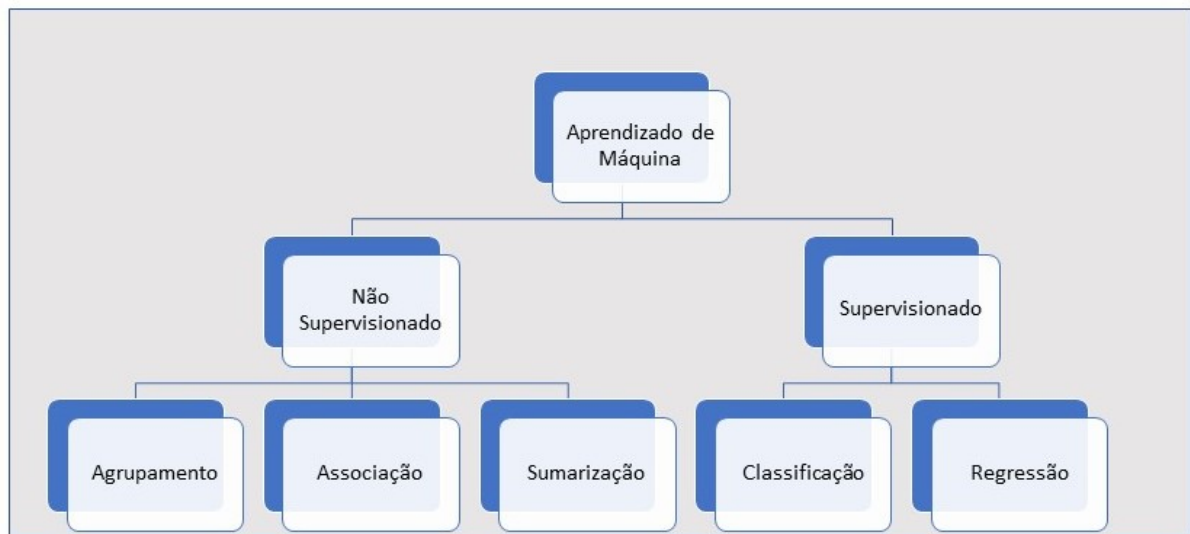


Figura 3 – Hierarquia dos tipos de aprendizado de máquina - adaptado de (FACELI *et al.*, 2021)

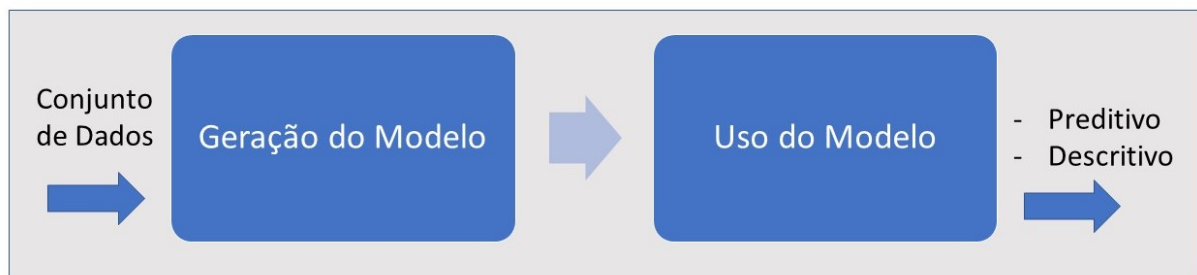


Figura 4 – Sistema de Aprendizado de Máquina - adaptado de (KUNG, 2014)

Algoritmo / Método	Tipo de Atividade
Arvore de Decisão	Classificação / Regressão
<i>K-Nearest Neighbor</i>	Classificação / Regressão
<i>Support Vector Machine</i>	Classificação / Regressão
Regressão Linear	Regressão
<i>Random Forest</i>	Classificação / Regressão
Redes Neurais	Classificação / Regressão

Figura 5 – Algoritmos de Aprendizado de Máquina - adaptado de (AMBIKA, 2020)

classificação que o atributo alvo é categórico.

AMBIKA (2020) descreve os tipos de algoritmos de aprendizado de máquina. A figura 5 ilustra os tipos de algoritmos e a atividade correspondente. (KUBAT; KUBAT, 2017), (MARCACINI, 2023), (FACELI *et al.*, 2021).

2.1.2.1 **Árvore de Decisão**

Árvore de decisão é um algoritmo de categoria simbólica e que busca dividir um problema para poder resolvê-lo. Podem ser usadas para problemas de classificação e regressão.

O critério de decisão de uma árvore de decisão é a medida que é utilizada para selecionar um atributo para um nó da árvore, que pode ser ganho de informação (baseado em entropia), índice gini. (FACELI *et al.*, 2021).

2.1.2.2 **KNN - K-Nearest Neighbors**

É um tipo de algoritmo de Aprendizado de Máquina que é baseado em distância, ou seja, que para o cálculo da predição consideram a distância entre os dados. Trabalham com a hipótese de que dados similares tendem a estar agrupados numa mesma região de espaço. Este algoritmo não generaliza o conhecimento, ele faz predições armazenando e comparando o dado com as amostras do conjunto de teste. (FACELI *et al.*, 2021).

Ele é considerado como o algoritmo mais simples a ser usado com uma boa acurácia para o cálculo da predição. As seguintes atividades são executadas:

- Cálculo da distância entre os objetos de teste e treinamento
- Seleção dos objetos vizinhos mais próximos dos dados de teste
- Alocação dos objetos de teste na classe mais frequente entre os objetos vizinhos mais próximos
- Para problemas de regressão, se calcula a média dos valores de saída dos objetos vizinhos mais próximos

Uma forma simples de calcular a distância de dois objetos $\mathbf{x}$ e $\mathbf{y}$ é através do cálculo da distância Euclidiana entre eles, através da seguinte equação:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.8)$$

2.1.2.3 SVM - *Support Vector Machine*

É um método de maximização de margens, que são os valores que determinam o grau de separação dos dados de diferentes classes, e que pode ser usado também para problemas de regressão que é caso deste trabalho para o desenho de um modelo de estimativa de desperdício de alimentos. Geralmente, apresenta um desempenho aceitável em problemas de cálculos preditivos, porém apresenta dificuldades para a definição dos valores dos parâmetros e dificuldade para a interpretação do modelo gerado por este método. (FACELI *et al.*, 2021).

A equação abaixo representa o algoritmo SVM para regressão:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2.9)$$

Com as seguintes restrições:

$$y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \quad (2.10)$$

$$\langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \quad (2.11)$$

$$\xi_i, \xi_i^* \geq 0 \quad (2.12)$$

x_i, y_i são exemplos de treinamento, w é o vetor de pesos, b é o viés (bias), $\epsilon + \xi_i$ e $\epsilon + \xi_i^*$ são variáveis de folga, ϵ é a margem de tolerância.

O objetivo é o de obter uma função linear que aproxime os pares (x_i, y_i) dos dados de treinamento com uma precisão de ϵ . A figura 6 ilustra de maneira simplificada o procedimento realizado pelo SVM.

2.1.2.4 Regressão Linear

A regressão linear pode ser uni variada ou multivariada e é um dos principais algoritmos de aprendizado de máquina. O objetivo é o de definir os melhores coeficientes para prever o valor conhecido da saída, através de uma análise da correlação de dados de entrada e saída.

Trata-se de uma abordagem simples para predição do valor da variável (y) com base num conjunto de variáveis observadas (x), ou seja, é um método estatístico que identifica a relação entre duas variáveis (FERREIRA *et al.*, 2022).

A equação abaixo ilustra o modelo de regressão linear múltipla (AMBIKA, 2020).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

A variável dependente y será calculada por mais de uma variável independente x . A constante β_0 é o valor esperado de y quando todas as variáveis independentes forem

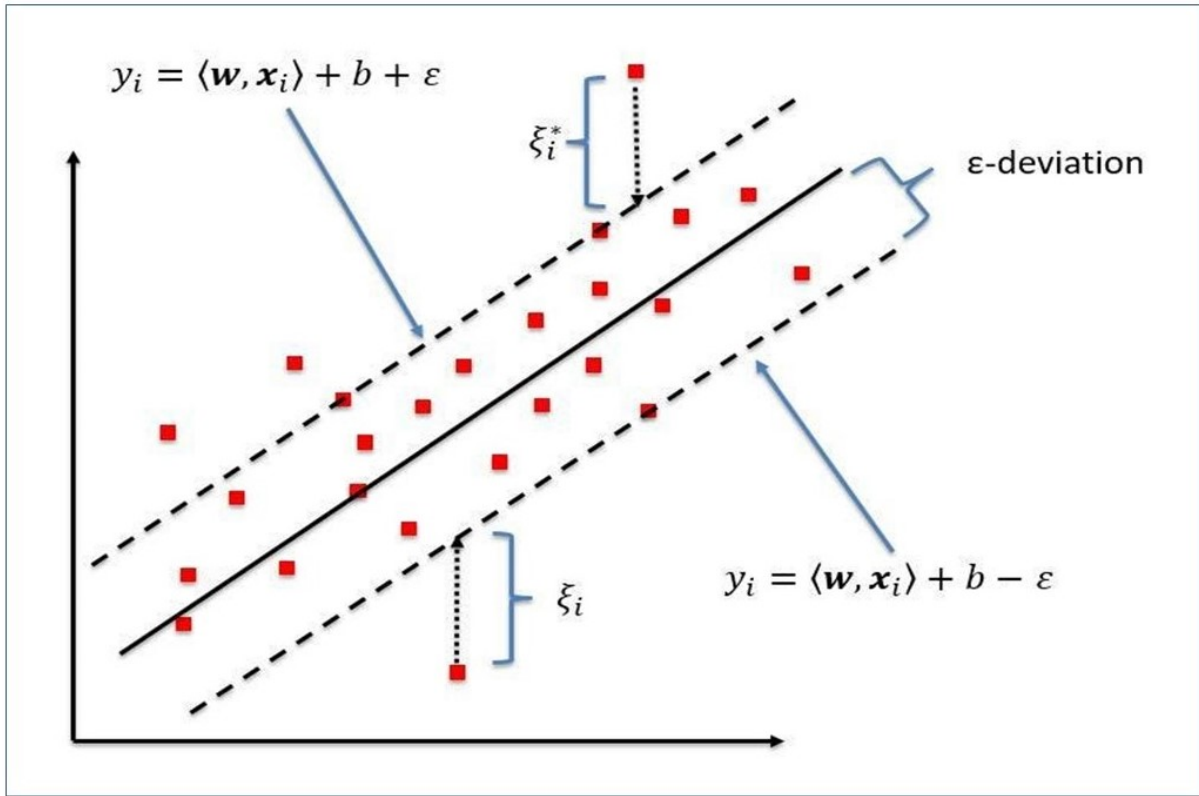


Figura 6 – Ilustração simplificada do procedimento realizado pelo SVM - adaptado de (FACELI *et al.*, 2021)

nulas. β_1 é a variação esperada em y dado um incremento unitário em x_1 . ϵ é o erro não explicado pelo modelo.

2.1.2.5 Random Forest

É um método simbólico do tipo comitê que usa um modelo de hierarquia de decisão que auxiliam resolução de problemas de regressão e classificação. Combina árvores de decisão onde cada árvore é induzida através do uso de um subconjunto aleatório de atributos preditivos. (ROMERO, 2023).

Para o caso de regressão a previsão final é geralmente a média das previsões de todas as árvores de decisão individuais na floresta e é calculada pela seguinte equação:

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B \hat{y}_i \quad (2.13)$$

Onde a previsão final $\hat{y}$ é a média das previsões $\hat{y}_i$ das B árvores individuais.

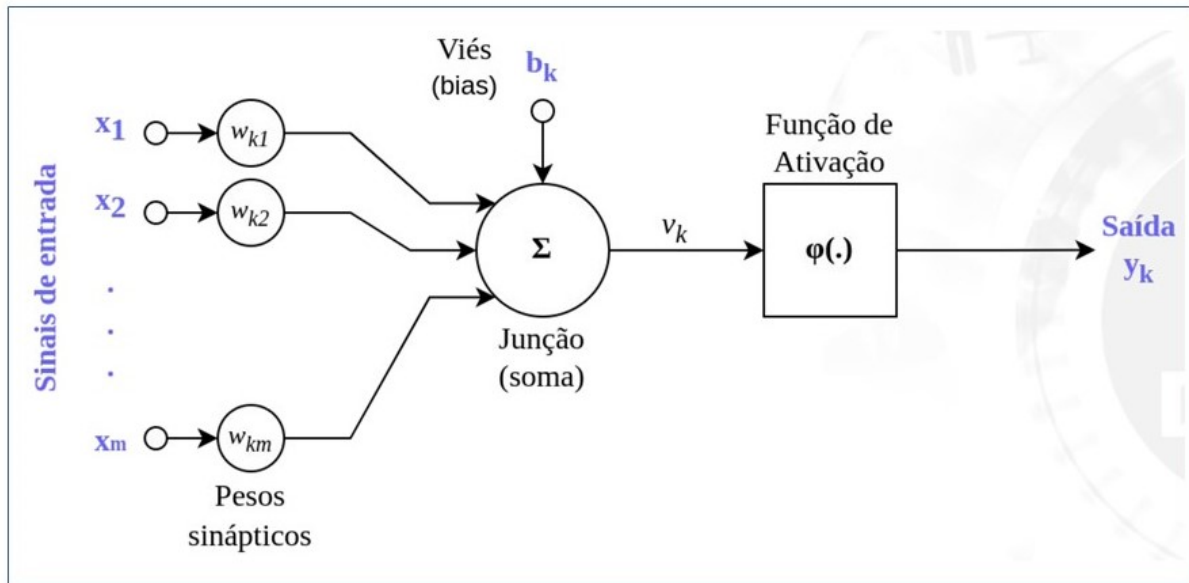


Figura 7 – Representação de uma RNA (Rede Neural Artificial) - adaptado de (SILVA, 2023)

2.1.2.6 Redes Neurais

É um método considerado conexionistas, pois consiste numa rede composta de camadas de neurônios, denominadas de RNAs (Redes Neurais Artificiais), que são sistemas de computação constituídos de unidades de processamento simples interconectados. As RNAs imitam a estrutura de um cérebro humano (FERREIRA *et al.*, 2022).

O conhecimento é adquirido por meio de um processo de aprendizado baseado na rede. Os pesos (parâmetros) das conexões são usados para armazenar este conhecimento. A figura 7 ilustra a representação de uma RNA. A saída de um neurônio é expressada pela variável y_k .

Esses algoritmos nos ajudam a resolver problemas de Regressão, ou seja, situações que temos como desafio determinar um valor específico com base numa análise de um conjunto de dados. Endereçam também a resolver os problemas de classificação, ou seja, resolver situações em que é necessário definir uma classe como resultado de uma análise de conjunto de dados.

Para este trabalho as técnicas exploradas foram as supervisionadas com foco em regressão, pois o objetivo foi a predição do volume de desperdícios de alimentos.

2.1.3 Trabalhos relacionados com o tema desenvolvido

PRINCIPATO *et al.* (2023) constatou que o uso de sistemas com base em TI (Tecnologia da Informação) auxilia na redução do desperdício de alimentos, assim como a importância de campanhas de treinamento dos colaboradores e programas de conscientiza-

ção dos comensais.

SOLE (2022) apontou que várias empresas investiram no desenvolvimento de soluções tecnológicas com o uso da Inteligência Artificial, com o foco na redução do desperdício de alimentos ao longo da cadeia de produção e abastecimento. Entre elas, a IBM, Gamaya, Tupl Agro, Seebo, Self Engine, Wasteless, Relex, Winnow, Phhod, Lumitisk, Orbisk e TotalCtrl. São soluções tecnológicas que buscam um monitoramento e controle do desperdício nas fases de produção, processamento, distribuição e consumo dos alimentos. Embora existam estas soluções disponíveis no mercado, a sua implementação não é fácil.

Atualmente, existe um processo amplo do uso de tecnologias de automação na agricultura, processamento, armazenamento e distribuição de alimentos. O desperdício ocorre em todas as etapas da cadeia de valor, portanto existe oportunidade de controle e redução em todas as etapas do ciclo de produção de alimentos. YUSOFF, WOOLLEY e GODSELL (2023) afirma que existe a necessidade de uma padronização na categorização do desperdício de alimentos. Isso facilitaria o uso da IA para buscar alcançar as otimizações necessárias.

ENNIS (2023) aponta que uma das problemáticas do desperdício de alimentos está relacionada com as políticas públicas que deveriam ser mais globais e efetivas quanto ao controle, do processo e de mecanismos que possam garantir um melhor aproveitamento dos desperdícios durante o processo de produção, distribuição e consumo. BULHÕES *et al.* (2022) conclui num estudo de desperdício de alimentos que uma solução efetiva requer uma abordagem de análise sob várias dimensões, não apenas com foco em medidas de redução dos desperdícios, mas também com foco em atividades e processos de produção e processamento dos alimentos.

Trabalhos com foco nas etapas da cadeia de geração de desperdícios

O desperdício de alimentos como a redução da quantidade de alimentos ocorre durante a cadeia de abastecimento alimentar. A maioria das pesquisas se concentram na análise de atividades executadas desde a produção dos alimentos até o consumo, e não na relação dos “stakeholders” envolvidos.

AHMADZADEH *et al.* (2023) apontou que o desperdício de alimentos é um problema global crescente com impactos negativos na economia dos países e que as tecnologias de IOT (*Internet of Things*) e Big Data estão sendo usadas para buscar soluções para a redução do desperdício de alimentos, através do uso de sensores (hardware) e Aplicativos (software) em toda a cadeia de abastecimento. Especificamente, descreve a possibilidade do uso dos diversos algoritmos de aprendizado de máquina como apoio para este desafio, porém não executou nenhum experimento com uma aplicação prática.

Para facilitar a identificação e análise dos trabalhos relacionados com o tema desenvolvido foi definido o seguinte critério de classificação que segue a estrutura de cadeia



Figura 8 – Cadeia de Geração de Desperdício de Alimentos (AUTOR)

de geração de desperdício de alimentos. A figura 8 ilustra esta cadeia.

A atividade de Planejamento de Cardápio consiste na elaboração do cardápio para cada unidade operacional (Restaurante), com base nos contratos estabelecidos e receitas existentes para os pratos. Segue a atividade de Operação, onde ocorre a produção dos pratos. Em seguida, a entrega é feita em balcões especializados onde os alimentos ficam armazenados. Nessa atividade, a entrega do alimento é feita na modalidade de auto serviço, onde o comensal seleciona livremente os alimentos desejados e na outra modalidade em que existe os profissionais que entregam os alimentos selecionados.

Trabalhos com foco na etapa de planejamento de cardápio

GARRE, RUIZ e HONTORIA (2020) realizou um estudo que identifica os desafios do processo de planejamento no processo produtivo de alimentos e o uso de algoritmos de aprendizado de máquina para reduzir o desperdício de alimentos com o foco específico no processo de planejamento como medida pró ativa. Como existem muitas incertezas neste processo o uso dos algoritmos se comprovou como uma maneira efetiva de melhorar o sistema de planejamento da produção de alimentos, pois o sistema analisado era baseado na experiência dos gerentes de produção. Nove modelos diferentes de regressão foram aplicados neste estudo que concluiu que o uso dos algoritmos de aprendizado de máquina auxilia na redução de incertezas do processo de planejamento da produção de alimentos.

PRINCIPATO *et al.* (2023) constatou que é fundamental medir o desperdício de alimentos e implementar medidas de redução, através do uso de sistemas (software e hardware) para fazer o monitoramento e identificar melhorias nos processos usados na operação da produção dos alimentos. Entender bem o volume do desperdício e as razões são importantes para a definição das ações necessárias de melhorias.

Os algoritmos/modelos de aprendizado de máquina usados foram o *Random Forest*,

Bagging e *Boosting*. Foi usado o volume diário, em KG (Quilogramas) de desperdício de alimentos como variável de resposta e o dia da semana, tipo de serviço, quantidade de serviços realizados e o tipo do serviço como variáveis exploratórias.

Uma das dificuldades encontradas para a medição do desperdício durante a preparação/operação está relacionado ao desafio de prever exatamente a quantidade a ser produzida. Foi identificado 3 momentos chave da ocorrência do desperdício: na preparação, na entrega e no consumo. Esta pesquisa apontou que o uso dos sistemas e soluções digitais gerou reduções no desperdício e como resultado benefícios financeiros para a empresa e para o meio ambiente.

RODRIGUES *et al.* (2024) usou o algoritmo de *Random Forest*, Redes Neurais, *Light Gradient-Boosting Machine* (*LightGBM*) e conseguiu medir um potencial redução de desperdício de alimentos de 14 % a 52 % diretamente com um melhor processo de planejamento da operação dos serviços de alimentação. Além disso, foi possível suprir cerca de 3 % a 16 % da demanda não atendida com a implementação de melhorias no processo de planejamento, ou seja, a melhoria na previsão da produção minimiza o desperdício e maximiza a oferta de serviços.

Concluiu que o uso de aprendizado de máquina do processo de planejamento resulta numa melhoria no nível de serviço de alimentação e na redução do desperdício. Aponta que o estudo do uso de melhores previsões de produção de alimentos para conseguir uma redução do desperdício dos alimentos está no estágio inicial de desenvolvimento.

Trabalhos com foco nas etapas de operação, entrega e consumo dos alimentos

Segundo PARISOTO, HAUTRIVE e CEMBRANEL (2013), a adoção de medidas de educação com os colaboradores e de conscientização dos comensais colaboraram efetivamente para a redução do desperdício de alimentos. Do lado dos colaboradores foi necessário a elaboração de fichas técnicas de preparação para padronizar o processo de operação. Do lado dos comensais, foram elaboradas campanhas de conscientização. Isso motivou a redução de resto-ingesta de 4,77 % para 3,99 %.

Trabalhos com foco em toda a cadeia de geração de desperdício de alimentos

BUSTAMANTE, AFONSO e RIOS (2018) executou um estudo de análise exploratória de desperdício de alimentos na Espanha, em restaurantes escolares, e pode observar mudanças no volume de desperdício antes e depois de atividades de conscientização dos comensais. Aponta que as causas do desperdício de alimentos estão mais centradas no comportamento dos colaboradores que trabalham na operação do restaurante e dos comensais quanto ao consumo. Alerta que é importante sempre ter claro a definição do prazo, destino, alcance e limite do desperdício. Além disso, resume que o desperdício é gerado

em função de erros no processo de estimativa da demanda, no processo de preparação, armazenamento e manipulação dos alimentos e, finalmente, no consumo pelo comensal.

Nestas etapas temos como resultado o desperdício por super produção, por produtos vencidos, por desperdício na cozinha, por produtos produzidos e não consumidos e por alimentos de sobra nos pratos dos comensais. A análise estatística dos dados foi feita com uma ferramenta denominada *Statgraphics XVII*. Foi realizado uma análise estatística de desperdício nos pratos, com base em teste de hipóteses em que se comprovou a relação do volume do desperdício com a localização geográfica do restaurante e com o perfil educacional do comensal.

DELIBERADOR *et al.* (2021) realizou uma pesquisa em restaurantes universitários no Brasil, utilizou a correlação de Spearman para verificar a correlação entre as variáveis de desperdício de alimentos e a razão dos quais os comensais geram estes desperdícios. Concluiu um desperdício médio de 68 gramas por comensal. Foi observado também um desperdício maior dos comensais que colocavam os alimentos na bandeja versus os que se serviam de pratos prontos. Foi considerado um intervalo de confiança de 95 %, erro amostral de 5 % e a participação de 342 comensais.

Esta pesquisa concluiu que existe uma correlação entre as variáveis de sabor, cheiro, aparência, textura e temperatura como causas potenciais do desperdício de alimentos. Isso foi possível apurar, pois foi utilizado uma técnica de questionário respondido pelos comensais para a coleta de dados.

Até o momento não foi identificado nenhum trabalho de pesquisa com o mesmo escopo e objetivo, ou seja, que tenha como escopo a análise de dados específica de desperdício na atividade de consumo dos alimentos em restaurantes de uma empresa prestadora de serviços. Isso fortalece a importância desta pesquisa para apoiar no desafio contínuo da redução do desperdício de alimentos.

3 METODOLOGIA E DESENVOLVIMENTO

3.1 Metodologia definida para o desenvolvimento do modelo de estimativa de desperdício de alimentos

A metodologia para este trabalho contém componentes que são os blocos que estão ilustrados na figura 5. Cada bloco possui entrada e saída de dados e seguem um fluxo bem definido.

O objetivo desta metodologia é o de orientar a execução das seguintes etapas e atividades:

- Coleta de Dados
- Exploração de Dados (medidas de localização, momentos e visualização)
- Pré-processamento de Dados (limpeza dos dados, transformação dos dados)
- Processamento de Dados (algoritmos de aprendizado de máquina, planejamento e execução da experimentação)
- Pós-Processamento de Dados (verificação da meta e critérios)
- Utilização do Conhecimento (aplicação das conclusões obtidas para a elaboração do Modelo de Estimativa de Desperdício de Alimentos)

3.1.1 Coleta de dados

Esta etapa consiste no levantamento de dados dos desperdícios de alimentos dos restaurantes da GRSA. Cada restaurante (unidade operacional) segue uma rotina diária

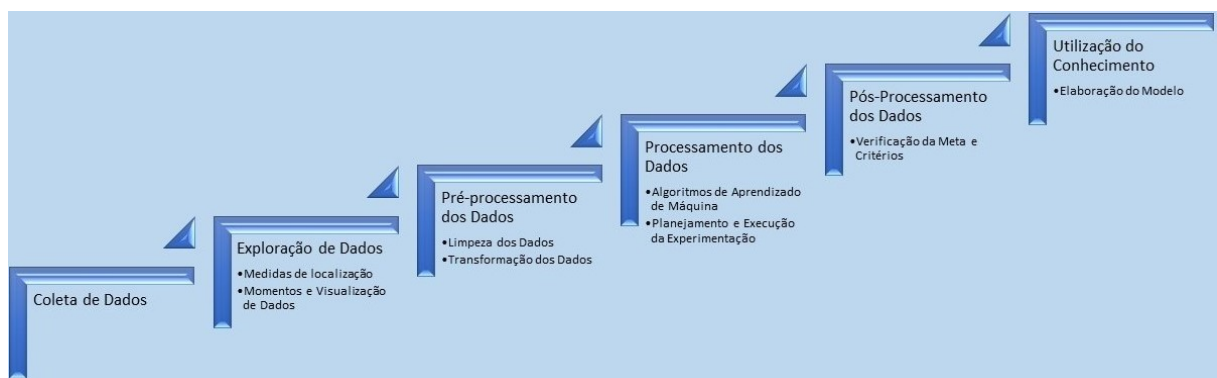


Figura 9 – Diagrama de Blocos o desenvolvimento do Modelo de Estimativa de Desperdício de Alimentos - adaptado de (REZENDE, 2003)

de coleta e pesagem dos desperdícios de alimentos da produção, distribuição, entrega e consumo dos alimentos. Os desperdícios são classificados pelos colaboradores, pesados e registrados num sistema de informação pela responsável do restaurante. Os dados estão registrados num *dataset* que é a base de dados para este trabalho.

3.1.2 Exploração de dados

Na etapa exploração dos dados de desperdício dos alimentos contém a atividade de gerar hipóteses, entender os dados por meio de técnicas e sumarizar as informações. Consiste no uso de medidas de localização: Média, Mediana e Percentil. Além disso, consiste na identificação dos momentos com o cálculo da Variância, Skewness, Kurtosis e Correlação.

Para a visualização inclui a atividade de apresentação dos dados que é feita com o Histograma e Box Plot. Inclui também uma atividade para a definição de uma estratégia de amostragem e uma sistemática de treinamento e teste dos dados, com base nos algoritmos selecionados.

3.1.3 Pré-processamento dos dados

Esta etapa é responsável pela limpeza dos dados, transformação dos dados, redução de dimensionalidade e balanceamento dos dados. A atividade de limpeza dados tem como objetivo identificar e analisar os dados ausentes, redundantes e inconsistentes. Além disso, inclui a atividade de transformação dos dados, através de codificação, normalização, padronização e redução de dimensionalidade, com o respectivo balanceamento, para que os dados possam ser processados de maneira adequada pelos algoritmos de aprendizado de máquina.

3.1.4 Processamento dos dados

Esta etapa tem como objetivo elaborar o planejamento, executar a experimentação, através do uso de algoritmos de aprendizado de máquina, como foco no objetivo principal deste trabalho que é de caráter preditivo (estimativa) de desperdício de alimentos. Está previsto o uso dos algoritmos de Árvore de Decisão, KNN, SVM, Regressão Linear, *Random Forest* e Redes Neurais e inclui também a atividade de ajustes dos parâmetros de cada algoritmo.

3.1.5 Pós-Processamento dos dados

Esta etapa contempla a atividade de verificação da meta definida de redução de desperdício de alimentos. Inclui uma atividade para avaliar a qualidade do modelo, através do cálculo de medidas de erro.

3.1.6 Utilização do conhecimento

Como objetivo central deste trabalho, esta etapa consiste na utilização do conhecimento adquirido, ou seja, nas estimativas de desperdícios de alimentos calculadas para que a empresa possa ajustar o modelo atual e implementar ações para melhorar a eficiência e eficácia do processo atual de controle dos desperdícios de alimentos.

Estas ações são classificadas em medidas quantitativas e qualitativas de gestão. Isso permite uma discussão com o time da gestão na empresa e identificação de ajustes a serem implementadas para que seja possível reduzir o problema de desperdício de alimentos.

3.2 Desenvolvimento

3.2.1 Descrição do estudo de caso

No caso específico desta pesquisa, o foco foi de analisar o desperdício de alimentos na fase na cadeia de produção dos alimentos pelos restaurantes da empresa usada como Estudo de Caso: GRSA Compass, denominada GRSA.

Esta empresa atua no Brasil no segmento de serviços de refeições coletivas. Atualmente, a sua operação possui dois segmentos: Soluções de Alimentação e Serviços de Suporte e atua nos seguintes setores: Empresas, Saúde, Educação, Eventos, *Off Shore*, Serviços e Varejo. A GRSA possui atualmente cerca de 30.000 colaboradores e cerca de 2000 filiais (Restaurantes - Unidades Operacionais) distribuídos no Brasil. A GRSA é parte do Grupo Compass, presente em cerca de 50 países.

Uma das iniciativas estratégicas da GRSA é o de obter “Desperdício Zero” nas Unidades Operacionais. Isso faz parte do programa de ESG (*Environmental, Social and Governance*) do Grupo Compass.

O processo atual de “Controle de Desperdício de Alimentos” possui como objetivo reduzir o custo de matéria-prima e diminuir o desperdício em estágios críticos da operação. Além disso, o foco estendido é de envolver o cliente e os colaboradores para estarem cientes dos desperdícios e com isso gerar mudanças necessárias sempre na busca da meta de “Desperdício Zero”.

Este processo é seguido diariamente nos Restaurantes, denominados de “Unidades de Operação”. É de responsabilidade da Gerente da Unidade de Operação, que normalmente por necessidade regulatória local é uma Nutricionista. O processo é composto das seguintes atividades:

a) Coletar os resíduos de pré-preparo (aparas, talos e folhas de hortaliças), super produção (resto da distribuição e sobra limpa da cozinha) e resto de ingestão (sobra das bandejas dos consumidores);

b) Pesar os resíduos na balança em todos os serviços diários (Almoço, Jantar e

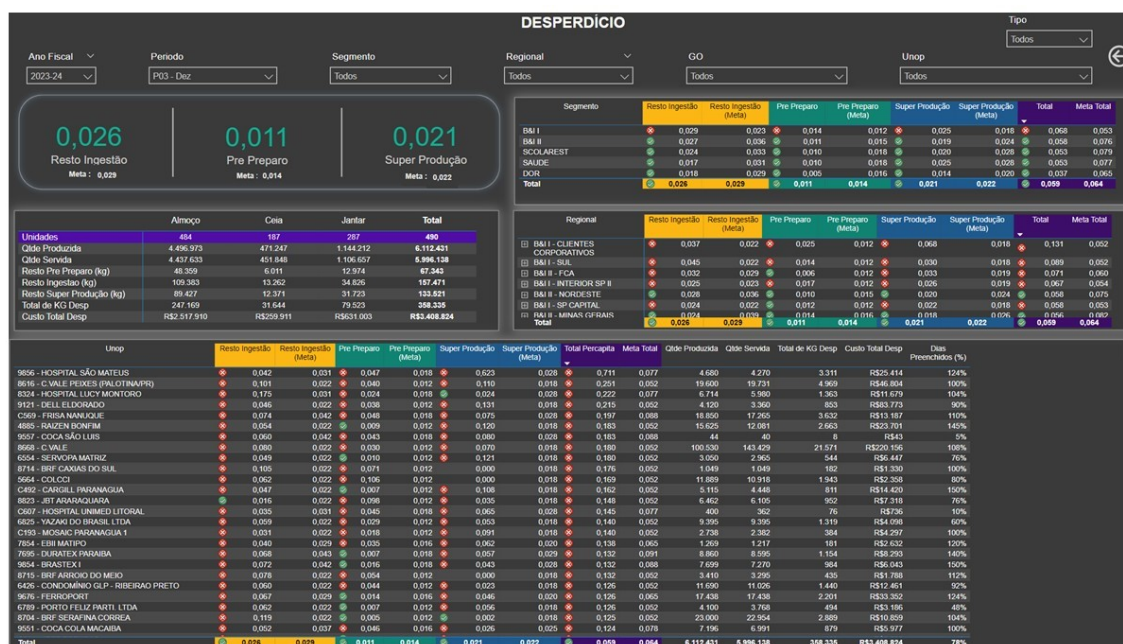


Figura 10 – DashBoard de Controle de Desperdício de Alimentos - (AUTOR)

Ceia);

d) Registrar o total em quilos de desperdícios no sistema “SGO – Desperdício Zero”;

e) Descartar o saco de lixo no recipiente de lixo orgânico assim que o desperdício for pesado;

f) Registrar no sistema SGO, o número de refeições planejadas, número de refeições efetivas, uso de proteínas com ossos, frutas com cascas e suas quantidades produzidas.

O sistema SGO deduz do Per Capita o peso dos ossos e cascas para apurar de maneira mais precisa o volume do desperdício.

A figura 10 ilustra o *DashBoard* de Controle de Desperdício de Alimentos que é usado pela área de qualidade e segurança alimentar da empresa. Este controle é usado para o monitoramento e administração dos níveis de desperdício para cada unidade operacional e setor. Este controle permite com que a gerente da unidade possa tomar medidas corretivas para estar de acordo com a meta estabelecida.

3.2.2 Aplicação da metodologia

Para o desenvolvimento deste trabalho foi usado uma metodologia com as seguintes etapas: coleta de dados, exploração de dados, pré-processamento de dados, processamento de dados, pós-processamento de dados e utilização do conhecimento.

A coleta de dados consiste na consolidação dos dados de desperdícios de alimentos

dos restaurantes e que formam o *data set* base para o trabalho. Na etapa de exploração dos dados são usadas as técnicas da ciência de dados para entender o comportamento dos atributos analisados, principalmente do volume, tipo de desperdício e localização geográfica do restaurante.

Na etapa de pré-processamento e processamento são usados os algoritmos de aprendizado de máquina com o foco no cálculo da estimativa de desperdício de alimentos. Este atributo é chave e serve como parâmetro para comparação com o valor atual planejado de desperdício para cada restaurante.

Para a etapa de pós-processamento é executado a avaliação dos modelos gerados pelos algoritmos, através do uso de medição de erros e com isso é definido a melhor opção para ser usada como estimativa de desperdício dos alimentos.

Finalmente, na etapa de utilização do conhecimento a empresa usa as estimativa de desperdício de alimentos obtidas para ajustar a sua estratégia e modelo atual de gestão de controle dos desperdícios de alimentos.

4 ANÁLISE DOS DADOS E RESULTADOS

4.1 Conjuntos de dados

Um conjunto de dados foi construído e estruturado para ser a base deste trabalho de estimativa de desperdício de alimentos. Isso foi possível, pois a GRSA, empresa do estudo de caso, iniciou no ano de 2021 um programa de conscientização para redução de desperdício de alimentos. Uma parte fundamental desse programa foi o desenvolvimento e implantação de um processo e sistema de informação.

Os dados foram extraídos desse sistema e consolidados num conjunto de dados que possui uma estrutura que é apresentada na figura 11. Esse conjunto de dados possui 268.778 registros e 15 colunas.

Esse sistema foi inserido num portal já existente nas unidades operacionais. Os dados de desperdícios são coletados diariamente ou nos dias de funcionamento da unidade operacional. Essa atividade é realizada pela equipe operacional da unidade que é responsável tanto pela coleta, pesagem e registro no sistema.

Uma análise inicial do conjunto dados apontou uma necessidade de pré-processamento para normalização e limpeza de dados nulos e incorretos. A figura 12 mostra estatísticas

Atributo	Tipo	Descrição
cdcr	object	código do restaurante - alfanumérico
nmcr	object	nome da unidade operacional
ano	int64	ano referência do sistema
mes	int64	mês referência do sistema
refeicao	object	tipo do serviço (Almoço, Jantar ou Ceia)
dt	object	data da coleta do desperdício no formato dia/mês/ano
refeicoesproduzidas	float64	quantidade de refeições produzidas (unidades)
refeicoesservidas	float64	quantidade de refeições servidas (unidades)
qttotaldesperd	float64	quantidade total de desperdícios (peso em KG)
custototaldesperd	float64	custo total em Reais dos desperdícios (valor calculado pelo sistema)
UF	object	unidade federativa
Divisao	object	divisão de negócios da empresa
Diretoria	object	diretoria da empresa
Contrato	object	tipo de contrato de prestação de serviços
GrupoEconomico	object	identifica o grupo econômico em que o restaurante faz parte

Figura 11 – Estrutura do conjunto de dados de desperdício de alimentos

	ano	mes	refeicoesproduzidas	refeicoesservidas	qttotaldesperd	custototaldesperd
count	268778.000000	268778.000000	267017.000000	266620.000000	268778.000000	2.687750e+05
mean	2023.475593	6.161818	307.344763	302.659898	15.709918	1.415881e+03
std	0.507959	3.494411	984.638898	1786.636939	972.734718	3.598363e+04
min	2021.000000	1.000000	0.000000	0.000000	-495739.975599	-2.907059e+03
25%	2023.000000	3.000000	68.000000	65.000000	3.000000	1.363519e+01
50%	2023.000000	5.000000	150.000000	150.000000	7.100000	4.660236e+01
75%	2024.000000	9.000000	335.000000	329.000000	17.000000	1.271240e+02
max	2024.000000	12.000000	270289.000000	861116.000000	72530.000000	6.921348e+06

Figura 12 – Descrição inicial do conjunto de dados de desperdícios

```
data.isnull().sum()

cdcr          0
nmcr          0
ano           0
mes           0
refeicao       1353
dt            252
refeicoesproduzidas  1761
refeicoesservidas  2158
qttotaldesperd    0
custototaldesperd    3
UF             0
Divisao       1117
Diretoria     1117
Contrato      1117
GrupoEconomico 1117
dtype: int64
```

Figura 13 – Dados nulos identificados no conjunto de dados

para algumas das variáveis do conjunto de dados e permite visualizar um exemplo tanto da existência de dados nulos, como de custos de desperdício negativo.

A figura 13 demonstra os dados nulos identificados no conjunto de dados. As variáveis mais críticas são: quantidade total de desperdício (**qttotaldesperd**), refeições produzidas (**refeicoesproduzidas**) e refeições servidas (**refeicoesservidas**).

Foi identificado no conjunto de dados 268.778 registros e após a limpeza de dados nulos, negativos e inconsistentes o conjunto passou a ter 255.920 registros com 12.858 registros excluídos, o que representa 4,8% do total de registros.

A figura 14 apresenta os dados estatísticos básicos do conjunto de dados. O atributo ano possui um intervalo de dados de 2021 à 2024. 2021 foi o ano que a GRSA implementou

	ano	mes	refeicoesproduzidas	refeicoesservidas	qttotaldesperd	custototaldesperd
count	256452.000000	256452.000000	256452.000000	256452.000000	256452.000000	256452.000000
mean	2023.473531	6.162576	299.829966	295.140848	16.387160	143.063444
std	0.507572	3.515758	554.206241	551.910666	29.722862	414.278850
min	2021.000000	1.000000	1.010000	1.010000	0.100000	0.000000
25%	2023.000000	3.000000	70.000000	65.000000	3.200000	15.140345
50%	2023.000000	5.000000	150.000000	150.000000	7.381616	47.509139
75%	2024.000000	9.000000	333.000000	327.000000	17.346170	125.437429
max	2024.000000	12.000000	15550.000000	18998.000000	500.000000	9991.852030

Figura 14 – Descrição ajustada dos dados de desperdícios

o processo de controle de desperdício de alimentos.

O atributo **refeicoesproduzidas** possui um intervalo de 1 a 15.550 refeições produzidas. Isso representa o universo de tipos e tamanhos diferentes de unidades operacionais (restaurantes) da empresa. Nesse período de 2021 a 2024 foi constatado uma média de 288 refeições produzidas por dia.

No mesmo período, foi observado que o atributo **refeicoesservidas** com um intervalo de 1 a 18.998 refeições servidas. Valor bem próximo à quantidade de refeições produzidas e que está aderente ao processo atual de planejamento e estratégia de servir o que for produzido.

Quanto ao atributo **qttotaldesperd**, quantidade total de desperdício, que além de ser o principal atributo analisado nesse trabalho, possui um intervalo de 0.10 a 500 kg, o que representa uma variabilidade significativa e um foco prioritário para redução. A média de 16.38 kg por dia, serviço e unidade operacional (restaurante) justifica a realização desse trabalho que contribui para a existência de um processo de estimativa de desperdícios de alimentos, e também da busca da meta de zero Kg para desperdício de alimentos.

Para o atributo **custototaldesperd**, custo total de desperdício, representa o custo em Reais do desperdício que é calculado pelo sistema de controle de desperdícios. A variabilidade desse atributo é de 0 até R\$ 9.991,85 com um custo médio de R\$ 135,00 por dia e serviço (almoço, jantar ou ceia).

A figura 15 apresenta o boxplot do atributo **qttotaldesperd**, quantidade total de desperdícios, que indica a presença de *outliers* o que justifica a necessidade de ações gerenciais nos pontos de coleta de dados de desperdícios para buscar o objetivo da sua redução.

A figura 16 apresenta um histograma trimodal do atributo **qttotaldesperd**, quantidade total de desperdícios, que indica a existência de *outliers* e também uma concentração

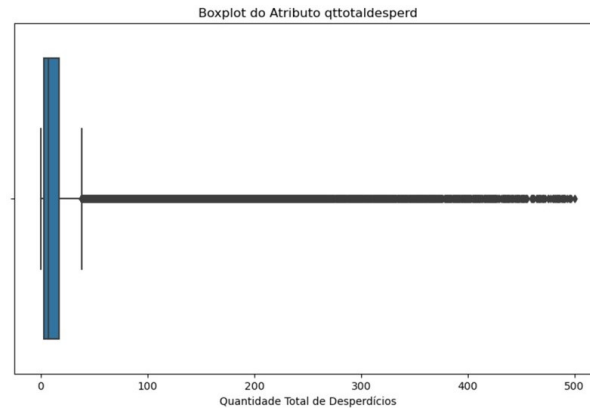


Figura 15 – Boxplot da quantidade total de desperdício de alimentos

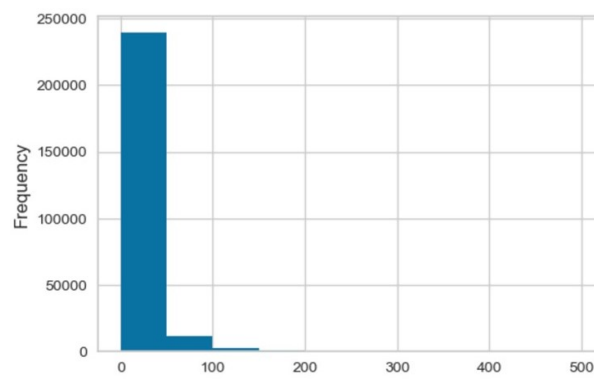


Figura 16 – Histograma da quantidade total de desperdícios de alimentos

do volume de desperdícios até 50 kg na maioria das observações realizadas.

Foi observado uma concentração dos desperdícios na região Sudeste, que além de ser a região com a maior quantidade de unidades operacionais, também gerou o maior volume de desperdícios e, conseqüentemente, o maior custo, conforme ilustra a figura 17.

4.2 Análise exploratória do atributo chave

O atributo chave do conjunto de dados analisado é a quantidade total de desperdícios, **qttotaldesperd**. A figura 18 representa o cálculo estatístico das métricas: média, mediana, percentis, desvio padrão, variância, obliquidade e curtose.

A média de 16.38 kg e a mediana de 7.38 kg demonstra uma diferença importante o que suporta uma simetria à direita. A mediana indica uma melhor representação do ponto médio sem a influência dos *outliers*.

O 25% percentil de 3.2 kg indica que 25% da quantidade total de desperdícios está abaixo de 3.2 Kg, 50% percentil de 7.38 kg indica que 50% da quantidade total de desperdícios está abaixo de 7.38 kg e 75% percentil de 17.34 kg indica que 75% da quantidade total de desperdícios está abaixo de 17.34 Kg.

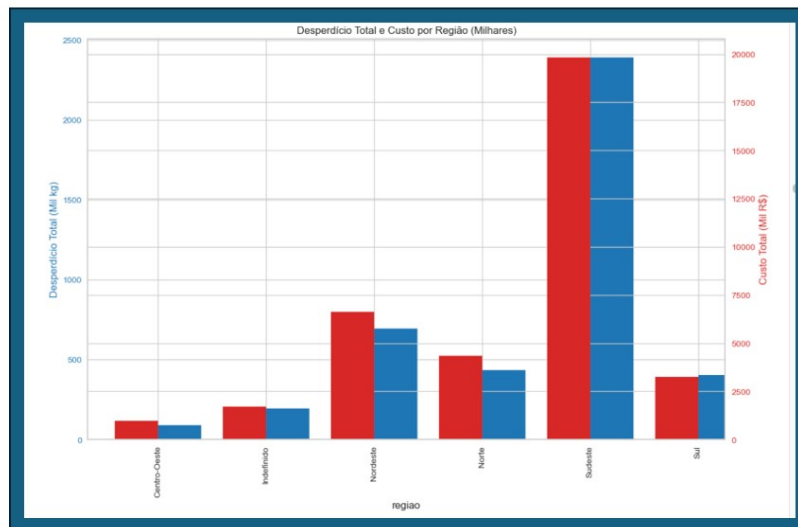


Figura 17 – Total de desperdícios de alimentos por região

O desvio padrão de 29.72 kg e a variância de 883.44 kg indicam valores altos o que demonstra uma distribuição de dados altamente assimétrica, com uma grande variabilidade e com presença de *outliers*.

A obliquidade de 6.41 indica um valor alto com uma distribuição assimétrica à direita e com uma cauda longa do lado direito da distribuição dos dados com a presença de valores extremos (*outliers*) maiores que a maioria dos dados. A curtose de 62.64 indica um valor muito alto e com a presença de muitos *outliers*.

Os dados do atributo **qttotaldesperd**, quantidade total de desperdícios, são altamente variáveis e assimétricos, com muitos valores baixos e com alguns valores extremamente altos que geram impactos na média. A mediana e os percentis indicam o que é o típico para a maioria dos dados, enquanto que a média, desvio padrão e variância indicam a influência dos *outliers*.

A figura 19 ilustra um gráfico de dispersão dos dados de desperdício de alimentos. O eixo x apresenta o atributo **refeicoesservidas**, quantidade de refeições servidas, e o eixo y o atributo **qttotaldesperd**, quantidade total de desperdício, o que indica que existe uma forte correlação entre esses atributos conforme o indicador pearson de 0.736.

4.3 Análise dos resultados dos algoritmos de aprendizado de máquina

Para a execução dos algoritmos de aprendizado de máquina foi necessário a normalização dos dados dos atributos numéricos para a melhoria da performance. Os algoritmos

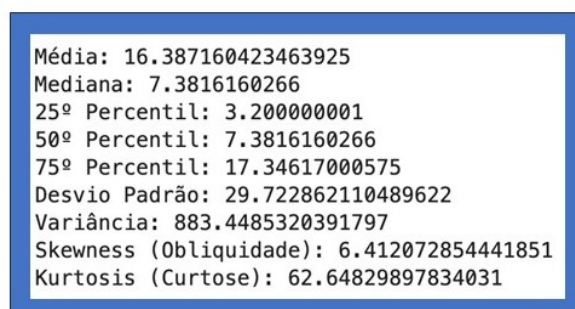


Figura 18 – Estatística do atributo chave: quantidade total de desperdícios

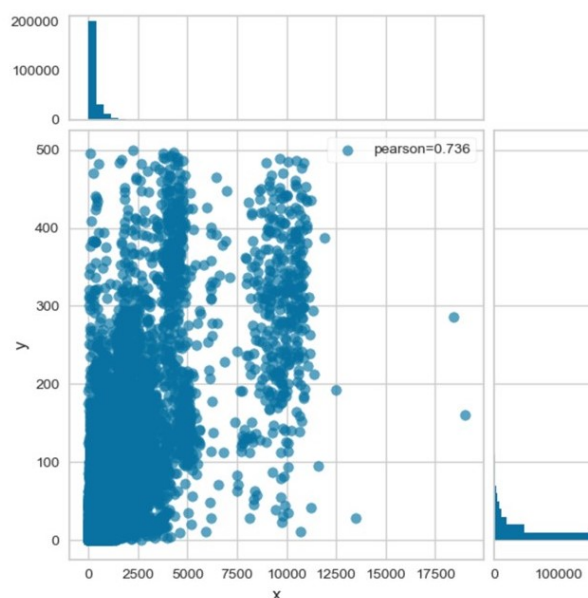


Figura 19 – Gráfico de dispersão dos dados de desperdícios de alimentos

usados foram:

1. Árvore de Decisão
2. KNN - K-Nearest Neighbors
3. SVM - Support Vector Machine
4. Regressão Linear
5. Random Forest
6. Redes Neurais

Para a execução dos algoritmos de regressão utilizamos os seguintes hiperparâmetros:

1. Arvore de Decisão: DecisionTreeRegressor, test_size=0.2, random_state=42

2. KNN - K-Nearest Neighbors: `KNeighborsRwregressor(n_neighbors=5), test_size=0.2, random_state=42`
3. SVM - Support Vector Machine: `kernel='rbf', C=100, gamma=0.1, epsilon=.1, test_size=0.2, random_state=42`
4. Regressão Linear: `LinearRegression(), test_size=0.2, random_state=42`
5. Random Forest: `RandomForestRegressor, test_size=0.2, random_state=42, n_estimators=100`
6. Redes Neurais: `activation='relu', optimizer='adam', loss='mean_squared_error', epochs=100, validation_split=0.2, verbose=1, test_size=0.2, random_state=42`

Inicialmente os algoritmos foram executados utilizando apenas os atributos refeições produzidas e servidas, com o objetivo de fazer uma análise preliminar. Posteriormente, adicionamos outros atributos como região e tipo de refeição. Para a definição da região fizemos uma conversão do atributo **UF**, unidade federativa, do tipo objeto para numérico (binário). Isso facilitou a conversão dos dados para as seguintes regiões: Norte, Nordeste, Centro-Oeste, Sul e Sudeste.

Foi feito também a conversão do atributo **refeicao**, tipo da refeição servida que pode ser almoço, jantar ou ceia do tipo objeto para numérico (binário). Em ambos os casos, aplicamos o *OneHotEncoder* para essa conversão.

Com isso definimos para a variável dependente o atributo **qttotaldesperd**, quantidade total de desperdícios, e os atributos: **ano,mes,refeicoesproduzidas, refeicoesservidas, Regiao_Centro-Oeste, Regiao_Nordeste, Regiao_Norte, Regiao_Sudeste, Regiao_Sul, refeicao_Almoço, refeicao_Ceia, refeicao_Jantar**.

Além disso, o conjunto de dados foi segmentado numa proporção de 20% como base de dados de testes e 80% como base de dados de treino. Esses parâmetros foram usados para a execução dos algoritmos.

A figura 20 demonstra o resultado dos erros dos algoritmos com o cálculo do *Mean Squared Error* (MSE) e o *R-squared*. A execução de cada algoritmo indicou que o *Random Forest* apresentou um melhor desempenho com um (MSE) de 0.0010645290420613468 e um *R-squared* de 0.7076573973409781. O pior desempenho foi o algoritmo SVM - *Support Vector Machine*, com um (MSE) de 0.0062993473621917664 e um *R-squared* de -0.7299364603058511.

O algoritmo de Redes Neurais apresentou um bom desempenho com um MSE de 0.001240216931123677 e um *R-squared* de 0.659409719058105, bem próximo dos resultados do algoritmo KNN que apresentou um MSE de 0.0011846916756091351 e um *R-squared* de 0.674658150119221.

Algoritmo	Mean Squared Error (MSE)	R-squared R ²
Regressão Linear	0.0015596706348650347	0.571680851652127
KNN – K-Nearest Neighbors	0.0011846916756091351	0.674658150119221
SVM – Support Vector Machine	0.0062993473621917664	-0.7299364603058511
Redes Neurais	0.001240216931123677	0.659409719058105
Árvore de Decisão	0.0018081156942720202	0.5034524873567094
Random Forest	0.0010645290420613468	0.7076573973409781

Figura 20 – Resultados dos erros dos algoritmos de aprendizado de máquina

Região	Quantidade de Desperdício (Previsto x Realizado) (%)
Sul	15.56 %
Sudeste	23.12 %
Norte	1.61 %
Nordeste	34.92 %
Centro-Oeste	16.73 %

Figura 21 – Diferença percentual por região da quantidade de desperdícios planejada e realizada

Como o algoritmo *Random Forest* apresentou um melhor desempenho fizemos uma análise dos resultados obtidos de valores preditos de desperdícios de alimentos para este algoritmo. Foi possível observar uma diferença de comportamento entre as regiões. A figura 21 ilustra a diferença percentual entre a quantidade real e a quantidade prevista pelo algoritmo. A região Norte apresentou a menor diferença percentual e a região Nordeste a maior.

Essas diferenças indicam uma necessidade de uma revisão de processos, controles, e práticas para uma melhor administração dos desperdícios de alimentos, principalmente na estimativa do planejamento de desperdícios e no próprio processo de planejamento da produção de alimentos.

Numa análise geral dos resultados obtidos na execução dos algoritmos foi possível

observar que os algoritmos *Random Forest* e o KNN apresentaram os melhores desempenhos, o algoritmo Redes Neurais apresentou um bom desempenho, porém com potencial de melhoras, o SVM apresentou o pior desempenho e os algoritmos de Regressão Linear e Árvore de Decisão apresentaram desempenhos relativamente baixos o que indica uma complexidade da relação entre os atributos que possivelmente eles não foram capazes de capturar.

5 CONCLUSÕES

Para o desenvolvimento do modelo de estimativa de desperdício de alimentos foi usado os principais algoritmos de aprendizado de máquina, entre eles, a Regressão Linear, KNN, SVM, Redes Neurais, Árvore de Decisão e *Random Forest*.

Foi constatado de que o desempenho da execução dos algoritmos é diferente o que foi um indicativo para selecionarmos o melhor algoritmo. Um dos desafios encontrados para trabalhar o conjunto de dados selecionado foi o fato da existência de ruídos. Os mesmos foram trabalhados o que gerou a necessidade de um processo de limpeza de dados para transformar os dados inconsistentes. Além disso, gerou demandas de melhorias sistêmicas para a inclusão de regras a serem incorporadas no processo e sistema existente de controle de desperdícios.

O conjunto de dados de desperdício de alimentos analisado gerou indicativos de que foi viável a definição de uma estimativa futura de desperdícios de alimentos. Isso ajudou a operação em buscar e ter uma meta de planejamento de geração de desperdício de alimentos. Como resultado foi possível a adoção desse dado tanto na fase de planejamento de cardápios, com a respectiva definição dos custos operacionais, assim como os ajustes necessários para considerar o volume de desperdícios estimados.

Além disso, foi identificado a existência de unidades operacionais (restaurantes) com casos extremos de quantidade total de desperdícios o que apontou a necessidade do desenvolvimento de políticas e ações específicas para a redução do volume de desperdícios. Ações como treinamento das equipes operacionais e programas de conscientização com os usuários do restaurante.

Os dados analisados mostraram que o controle de desperdício é um desafio, pois além das etapas envolvidas no processo produtivo de alimentação, envolve atividades que podem ser padronizadas e também podem seguir melhores práticas, porém ter um objetivo de desperdício zero, desejado, ainda não é possível.

O que constatamos é que é possível reduzir o desperdício com a implementação de um sistema de gestão e que ter uma estimativa, com base em algoritmos de aprendizado de máquina, que possa ser a base de um planejamento, além de orientar as equipes, possibilita um monitoramento e a implementação de um ciclo de melhoria contínua. Essa foi a principal contribuição a ser incorporado no modelo atual da empresa de administração dos desperdícios de alimentos.

Além disso, o trabalho indicou que uma revisão e melhoria do processo de planejamento de cardápios é necessária, pois é nessa fase do processo que são definidos os detalhes das receitas, quantidades de pratos a serem servidos, produzidos e consequentemente

é possível um planejamento mais assertivo com o uso da tecnologia de aprendizado de máquina.

Para a continuidade de pesquisas na área de desperdício de alimentos, sugerimos as seguintes pesquisas futuras:

1. Análise de dados de desperdícios de alimentos na etapa de preparação e operação dos alimentos
2. Análise do comportamento dos usuários no uso dos serviços de alimentação com foco num consumo consciente
3. Desenvolvimento de melhores práticas para melhoria dos processos operacionais para a produção e entrega de serviços de alimentação
4. Desenho de um modelo de estimativa de desperdício de alimentos com uso de um estudo de caso múltiplo (uso de algumas empresas do segmento de serviços de refeições coletivas)
5. Desenho e implementação de um sistema de controle de desperdício de alimentos com o uso de algoritmos de aprendizado de máquina

REFERÊNCIAS

- AHMADZADEH, S. *et al.* A comprehensive review on food waste reduction based on iot and big data technologies. **Sustainability**, MDPI, v. 15, n. 4, p. 3482, 2023.
- AMBIKA, P. Machine learning and deep learning algorithms on the industrial internet of things (iiot). **Advances in computers**, Elsevier, v. 117, n. 1, p. 321–338, 2020.
- BULHÕES, M. dos S. *et al.* Evaluation of factors responsible for waste and environmental impacts arising from the production of meals in food services. **Food Science and Technology**, v. 42, 2022.
- BUSTAMANTE, M.; AFONSO, A.; RIOS, I. De los. Análisis exploratorio del desperdicio de alimentos en plato en comedores escolares en españa. **LA GRANJA. Revista de Ciencias de la Vida**, Universidad Politécnica Salesiana, v. 28, n. 2, p. 20–42, 2018.
- DELIBERADOR, L. R. *et al.* Desperdício de alimentos: Evidências de um refeitório universitário no brasil. **Revista de Administração de Empresas**, SciELO Brasil, v. 61, p. e2020–0271, 2021.
- ENNIS, R. C. La problematización de las pérdidas y desperdicios de alimentos desde los organismos internacionales entre 2011 y 2021. **De Prácticas y Discursos. Cuadernos de Ciencias Sociales**, v. 12, 2023.
- FACELI, K. *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.: s.n.], 2021.
- FERREIRA, J. *et al.* Inteligência artificial/análise preditiva para modelagem computacional: Uma revisão de literatura. **Salão do Conhecimento**, v. 8, n. 8, 2022.
- GARRE, A.; RUIZ, M. C.; HONTORIA, E. Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty. **Operations Research Perspectives**, Elsevier, v. 7, p. 100147, 2020.
- KUBAT, M.; KUBAT, J. **An introduction to machine learning**. [S.l.: s.n.]: Springer, 2017. v. 2.
- KUNG, S. Y. **Kernel methods and machine learning**. [S.l.: s.n.]: Cambridge University Press, 2014.
- MARCACINI, R. M. **Método KNN - Notas de Aula - Curso 2 - CD, AM e DM - MBA IA e Big Data - USPSC**. [S.l.: s.n.], 2023.
- PARISOTO, D. F.; HAUTRIVE, T. P.; CEMBRANEL, F. M. Redução do desperdício de alimentos em um restaurante popular. **Revista Brasileira de Tecnologia Agroindustrial**, v. 7, n. 2, 2013.
- PERES, F. F. **Estatística Aplicada a Vida Real**. 2024. Acesso em: 30 maio 2024. Disponível em: <<https://fernandafperes.com.br/blog/interpretacao-boxplot/>>.

PRINCIPATO, L. *et al.* Introducing digital tools for sustainable food supply management: Tackling food loss and waste in industrial canteens. **Journal of Industrial Ecology**, Wiley Online Library, 2023.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. [*S.l.: s.n.*]: Editora Manole Ltda, 2003.

RODRIGUES, M. *et al.* Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. **Journal of Cleaner Production**, Elsevier, v. 435, p. 140265, 2024.

ROMERO, R. A. F. **Notas de Aula - Curso 2 - CD, AM e DM - MBA IA e Big Data - USPSC**. [*S.l.: s.n.*], 2023.

SILVA, D. F. **Redes Neurais e Deep Learning- MBA IA e Big Data - USPSC**. [*S.l.: s.n.*], 2023.

SOLÉ, O. Valero i. Artificial intelligence and blockchain for food loss and waste prevention and reduction. 2022.

VIEIRA, L. M. *et al.* Abordagens metodológicas para enfrentar o desperdício de alimentos: avançando a agenda. **Revista de Administração de Empresas**, SciELO Brasil, v. 61, 2021.

YUSOFF, N. I. M.; WOOLLEY, E.; GODSELL, J. Food security via food waste prevention: Categorization of household food waste for artificial intelligence-driven interventions. **Engineering Proceedings**, MDPI, v. 40, n. 1, p. 4, 2023.