|  |
| --- |
| **Estudo do algoritmo de aprendizado de máquina *Support Vector Regression* (SVR)como ferramenta de apoio na gestão de refeitórios**  **Autores (as):**  **Igor Martins Ferreira**  **Lucas Silva Sousa**  **Renan Dias de Oliveira**  **Orientador (a):**  **Eduardo Savino Gomes** |

**Resumo**

Em português, texto justificado, fonte Times New Roman, corpo 12, espaçamento simples (entre caracteres, palavras e linhas), no mínimo 100 e no máximo 250 palavras. Deve conter objetivo geral do artigo, o marco teórico que sustentou o estudo (principais conceitos, fundamentos), a descrição básica da metodologia, principais resultados e conclusões em texto corrido. O texto deve ser apresentado em parágrafo único. Xxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

**Palavras-chave**: no mínimo três (3); no máximo cinco (5).

**Abstract**

Em inglês, deve ser registrado nos mesmos parâmetros do resumo em português. Xxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

**Keywords**: Em inglês; no mínimo três (3); no máximo cinco (5).

Conteúdo da primeira página do arquivo:

1. Template (obrigatório - Todas as páginas do artigo deverão conter o template)
2. Título
3. Resumo
4. Palavras-chave
5. Abstract
6. Keywords

Na sequência (mesma página), início do texto propriamente dito

Atenção:

O uso do template é obrigatório. Este modelo já possui a formatação solicitada. Utilize-o para escrever o seu artigo.

1. **Introdução**

Uma grande parcela da comida produzida no mundo é jogada fora sem ter a chance de ser ingerida. A Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), alertou que, anualmente são desperdiçadas ou se perdem ao longo das cadeias produtivas cerca de 1,3 bilhão de toneladas de comida. O órgão ainda afirma que este volume representa 30% da comida produzida por ano no planeta, comida essa que poderia ser destinada às 821 milhões de pessoas que ainda passam fome no mundo (FAO, 2018).

Segundo a FAO (2018) o desperdício e perda estão presentes em todos os locais que produzem, vendem ou transportam alimentos, e até mesmo na mesa do consumidor, sendo que este último grupo representa 28% do desperdício. Dentro deste grupo estão as Unidades de Alimentação e Nutrição (UAN’s), que são entendidos como locais de produção e distribuição e alimentação para coletividades, tais como refeitórios escolares, restaurantes, refeitórios industriais e praças de alimentação (TEIXEIRA, 2007).

Como é dito por Júnior e Teixeira (2010) e Abreu et al (2012), a quantidade de sobras está relacionada com o planejamento inadequado de refeições e o número de pessoas que frequentam o estabelecimento. Por isso, é importante a realização de um planejamento com antecedência, usando registros das quantidades produzidas anteriormente para adequar a produção futura.

Com esses números em mente, não se pode ignorar o fato de que se estes alimentos puderem ser melhor aproveitados, mais pessoas teriam acesso aos mesmos, podendo assim reduzir a quantidade de pessoas que passam fome no mundo, mas seria possível utilizar a tecnologia para auxiliar esse processo que hoje ocorre de forma empírica, ou seja, pela experiência da pessoa que irá estimar quanto de comida produzir.

Dentro do cenário de evolução tecnológica, procura-se associar a tecnologia às técnicas já existentes visando uma melhoria de desempenho dos métodos já existente. Sendo assim o presente trabalho pretende criar um modelo preditivo com uso de Inteligência Artificial para auxiliar o gestor da cozinha na escolha da quantidade de comida ou número de refeições que devem ser produzidas para um determinado dia da semana através de técnicas de aprendizado de máquina que utilizarão dados do passado para prever o futuro.

1. **Referencial Teórico**
   1. **Lixo alimentar**

Dentro da literatura, existem diferentes definições para o lixo alimentar. Por esta razão, esse capítulo pretende descrever os conceitos que serão utilizados como referência a esse assunto ao longo do trabalho, para se ter um alinhamento melhor.

Basicamente, o lixo alimentar tem dois tipos principais, o desperdício e a perda. Essa última é definida como a diminuição involuntária da disponibilidade do alimento, seja na produção, no armazenamento ou transporte dos mesmos (FAO, 2020).

Já o desperdício é resultado da decisão voluntária de se jogar fora a comida, depois de preparada e apta para consumo (Peixoto & Pinto, 2016). O desperdício de alimentos pode ser dividido entre dois subgrupos distintos para ajudar a identificar melhor o problema com base na sua causa, são eles:

Sobra Limpa: É a comida que preparada, porém não consumida, ou seja, são os alimentos que ficam na cuba de alimentos (SÃO PAULO, 1999).

Sobra Suja: é a sobra de alimento no prato do comensal, também conhecido como resto ingesta (SCOTTON, 2010).

O presente trabalho trata apenas da sobra limpa, pois ela está diretamente ligada a quantidade de comida que o gestor da UAN decidiu produzir para determinado dia, segundo Pereira (2018) quase todos os serviços de refeitórios e restaurantes se baseiam na intuição dos gestores para estimar a quantidade de comida que deve ser produzida, esse método de estimativa pode levar a superestimação do número de refeições que devem ser produzidas gerando desperdício na forma de sobra limpa.

* 1. **Inteligência Artificial**

A inteligência artificial é uma vertente da ciência da computação na qual se realiza o estudo das faculdades mentais através de modelos computaciona e de cálculos que possibilitam perceber, raciocinar e agir (CHARNIAK e MCDERMOTT, 1986; WINSTON, 1992).

A palavra inteligência é proveniente do latim *inter* (entre) e *legere* (escolher), ou seja, é aquilo que possibilita escolhas sendo elas racionais ou não, é a habilidade de corporificar determinada tarefa de forma eficiente, já a palavra artificial originou-se do latim *artificiale*, reflete algo não natural, feito pelo homem. (FERNANDES, 2005).

Dessa forma, é possível dizer que inteligência artificial é o esforço para fazer com que a ciência da computação desenvolva computadores e modelos para pensarem, de maneira semelhante ao funcionamento da mente humana no sentido completo e/ou literal (HAUGELAND, 1985), para exercer atividades como tomada de decisão, resolução de problemas e de aprendizado (HELLMAN, 1978).

Uma das vertentes existentes na inteligência artificial é o aprendizado de máquina, que lida com a implementação de programas de computador que podem aprender de maneira autônoma (<https://www.britannica.com/technology/machine-learning>) e executar a tarefa aprendida. Esses programas são subdivididos de acordo com o tipo de tarefa executada:

* Classificação: É um algoritmo que tenta aprender a relação entre um grupo de variáveis de base e a variável alvo, com a finalidade de quando receber novos dados de base, acerte qual é o alvo automaticamente, pois já está treinado para isso. (AGGARWAL, 2014)
* Regressão: É um metodo de investigar as relações funcionais entre variáveis de base e variáveis alvo, mas neste caso a ligação entre elas é matemática, feito por uma equação, e visa gerar um valor numérico na variável alvo, ao invés de uma classe. (CHATTERJEE, 2015)

Os programas também são divididos entre dois tipos. Supervisionados e não Supervisionados.

Algoritmos supervisionados utilizam resultados já conhecidos para aprender, enquanto não supervisionados aprendem de maneira autônoma de acordo com os dados informados. Cada algoritmo desempenha as tarefas acima de uma maneira diferente, entre eles podemos citar as *Support Vector Machines*, que podem ser usadas para tarefas de classificação e regressão.

* + 1. **Support Vector Machines**

É uma técnica de aprendizado de máquina utilizada para criação de um classificador, utilizado para separar e categorizar grupos de entradas de dados em uma vou mais categorias (PAULA, 2016), comumente muito útil nos casos em que os dados não são linearmente separáveis. Na figura 1 temos duas classes diferentes, os pontos vermelhos e azuis. Uma máquina de suporte vetorial adiciona um hiperplano que separa os pontos vermelhos dos azuis pela máxima distância dos pontos. O hiperplano em geometria é uma generalização do plano N–1 dimensões, na figura 1 os dados estão plotados nos eixos X e Y, ou seja, duas dimensões. Neste caso o hiperplano traçado pelo algoritmo é uma reta, porém ela não é capaz de separar os pontos.

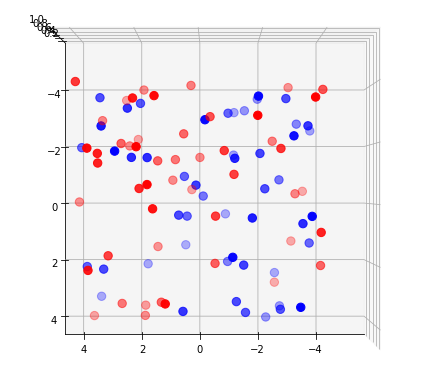


Figura - Ilustração do SVM, própria autoria.

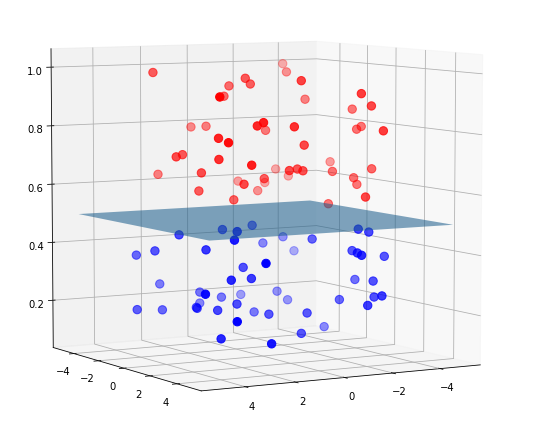
Para contornar esse problema, um modelo de *Support Vector Machine* organiza e esquematiza os dados de entrada em um espaço e cria um hiperplano para separar os dados como pode-se notar na figura 2.

Figura 2 - Ilustração do hiperplano, própria autoria

* + 1. **Support Vector Regression**

SVR (Support Vector Regression) é uma derivação do SVM (Support Vector Machine) que é um tipo de algoritmo utilizado para solução de problemas de regressão, isto é, prever valores de variáveis contínuas de uma função com múltiplas variáveis. Trata-se de um método de aprendizado de máquina supervisionado, ou seja, utiliza entradas de dados com resultados conhecidos para treinamento (AWAD e KHANNA, 2015). Uma maneira simples de entender o funcionamento dele é através de uma reta como exemplificado na figura 3. O épsilon é um dos hiperparametros para ajustar o modelo, ele é basicamente define um limite de tolerância, todos os dados que estiverem dentro da função da reta Ax + b + epsion, não são considerados como erro.

Uma imagem contendo mapa, pipa, texto, diferente

Descrição gerada automaticamente

Figura 3 - Ilustração de funcionamento do SVR, própria autoria.

O que levou o SVR à ser escolhido para ser aplicado neste caso é que de acordo com Awad e Khanna (2015), o SVR possui alta acurácia em suas previsões e alta capacidade de generalização, ou seja, descrever ou capturar os relacionamentos casuais não lineares entre as variáveis e também é muito utilizado para descobrir padrões não vistos anteriormente.

1. **Metodologia**

Se tratando de um problema de regressão, isto é, olhar para os dados do passado e tentar com base nisso prever o futuro, para que fosse possível atacar diretamente o problema do desperdício, seria necessário ter acesso à quantidade de comida produzida e a quantidade desperdiçada, no entanto, a estas informações não se teve acesso. Desta forma, o algoritmo irá gerar uma previsão de quantas pessoas irão à UAN com base nos dados históricos de refeições servidas, combinados com outros dados relevantes. A intenção é que essa estimativa auxilie a gestão do refeitório a planejar com antecedência o número de refeições que deverão ser produzidas naquele dia e assim o desperdício por sobras limpas diminua.

Para testar essa hipótese, o refeitório da Faculdade de Tecnologia Termomecânica foi escolhido como piloto para os testes. O refeitório serve as refeições no horário do almoço das 11:00 às 13:00 horas, e a janta das 18:15 até às 19:00 horas. A maior parte das refeições são servidas no período da manhã, onde estão presentes alunos e colaboradores em maior quantidade. O fluxo de alunos é mais ou menos constante na Instituição, já que uma das regras de aprovação nos cursos exige no mínimo 75% de presença de cada aluno nas aulas, fator esse que pode ter impacto na quantidade de pessoas que se alimentam na faculdade. Os alunos também são comunicados dos cardápios com antecedência, geralmente no começo da semana.

O estudo se baseará nos dados da janta, visto que a variação das refeições da janta é maior, pois há menos colaboradores se alimentando todos os dias.

* 1. **Coleta de dados**

Primeiramente foram coletados alguns dados junto a coordenação da faculdade e a gestão do Refeitório. Esses dados foram organizados em forma de tabela com cada coluna da tabela correspondendo a uma variável dependente ou independente que são utilizadas como treinamento para o algoritmo de previsão.

A escolha das variáveis independentes foi baseada em estudos anteriores e outros fatores relativos ao ambiente da faculdade também foram levados em consideração pelo seu potencial de influência na demanda por comida.

É possível observar uma introdução aos dados coletados na tabela 1, juntamente com uma breve descrição das variáveis.

|  |  |  |  |
| --- | --- | --- | --- |
| Variável | Tipo | Descrição | Fonte |
| Dia da semana | Valores 1 - 5 | Segunda, Terça, Quarta, Quinta e Sexta. | Calendário |
| Dia do bimestre letivo | 1 - n | Do primeiro dia do bimestre até a última avaliação do bimestre. | Calendário FTT |
| Precipitação | Escala em mm | Quantidade de chuva registrada para o determinado dia. | Instituto Nacional de Meteorologia |
| Refeições na janta | Valor numérico | Quantidade de pessoas ou, se preferir, refeições servidas. | UAN FTT |
| Guarnição | Variável categórica | Principal guarnição servida no dia. | UAN FTT |
| Semana letiva | 1 – n | Contagem de semanas até o fim do semestre. | Calendário FTT |
| Dia letivo | 1 – n | Contagem de dia até o fim do semestre | Calendário FTT |
| Temperatura mínima | Valor numérico | Temperatura mínima apresentada no dia vigente. | Instituto Nacional de Meteorologia |
| Precipitação binária | Variável categórica | 1 se choveu e 0 caso não tenha chovido. |  |
| Tipo preparo | Variável categórica | Modo de preparo, varia entre cozido, assado, frito e grelhado. | UAN FTT |
| Tipo refeição | Variável Categórica | Varia entre carne, frango, peixe e porco | UAN FTT |

Tabela 1 - Variáveis do Sistema

As variáveis **Dia letivo, Semana letiva** e foram utilizadas para que o modelo pudesse estabelecer uma relação de sazonalidade na frequência das pessoas no refeitório. Vale ressaltar também que nos dados de base que são usados testes e treinamento do código, estão sendo considerados apenas os dias letivos que foram apresentados como tal no calendário letivo fornecido pela própria faculdade.

Tipo\_dia - É provável que muitas famílias viagem durante as emendas de feriados prolongados ou que alguns colaboradores não trabalhem neste dia, por conta de alguma decisão da empresa, logo é esperado uma diminuição da frequência das pessoas em vésperas de feriado.

Precipitação – A escola fica localizada em uma região central do Grande ABC. Muitas vias de acesso que os alunos e colaboradores utilizam para descolar até a Faculdade podem ficar alagadas e obstruídas impedindo a chegada dos alunos até a faculdade.

Guarnição – As refeições servidas no refeitório seguem mais ou menos um padrão: É servido arroz, feijão, um ou dois tipos de guarnição, salada, suco ou sobremesa. Por conta da grande variedade de tipos de pratos que são servidos, foi decidido em agrupa-los de acordo com a principal guarnição servida. Os alunos e colaboradores são livres para decidir se vão almoçar no refeitório ou não. É provável que determinados tipos de pratos influenciem mais ou menos a frequência das pessoas em determinado dia.

Refeições – O número de refeições servidas representa o número de pessoas que frequentaram o refeitório. Esse dado é registrado todos os dias e trata-se da variável dependente, ou seja, o dado que vai nortear o treinamento do algoritmo.

Dia\_Bimestre\_Letivo – COLOCAR EXPLICAÇÃO DA VARIÁVEL RENAN

Para aplicação deste algoritmo foi escolhido a linguagem de programação Python com a biblioteca Sklearn, que é uma biblioteca de aprendizado de máquina de código aberto que suporta aprendizado supervisionado e não supervisionado. Essa biblioteca também fornece várias ferramentas para ajuste de modelo, pré-processamento de dados, seleção e avaliação de modelo e outras utilidades (scikit-learn.org).

ADICIONAR A EXPLICAÇÃO PARA AS DEMAIS VARIÁVEIS DO ESTUDO

**Explicação do estudo de correlação – Breve explicação**

**Métodos Pearson e Spearman – Breve explicação**

**3.1 Treinamento**

Para treinar o algoritmo, a base foi dividida em duas categorias: Dados para treinamento e dados para teste. A base de teste é criada através da seleção aleatória de alguns registros da base de dados, ela possui 30% dos dados enquanto o treinamento do algoritmo é feito com os outros 70%. A figura 4 demonstra a incidência de registros de cada semana letiva na base utilizada para testes.

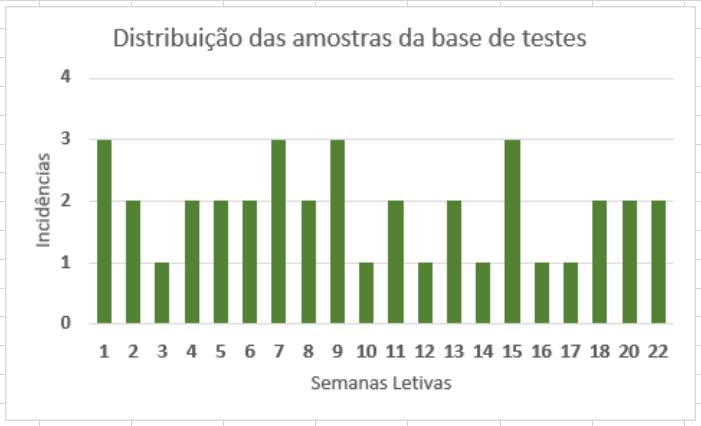


Figura 4 – Gráfico de distribuição da base de testes

* 1. **Otimização**

Existem alguns métodos utilizados para refinar ou melhorar a performance do algoritmo, entre elas podemos citar o GridSearch e a Permutation Feature Importance.

* ***GridSearch:*** O SVR possuí um conjunto de hiper parâmetros que devem ser otimizados para que o algoritmo obtenha melhor performance em sua precisão. Ela é implementada escolhendo um intervalo valores para os hiper parâmetros serem testado, eles são combinados de maneira aleatória entre si através de múltiplas iterações de treinamento com diferentes valores. A cada iteração o resultado de performance do algoritmo é calculada e o melhor conjunto de hiper parâmetros é selecionado.
* ***Permutation Feature Importance***: Trata-se de uma técnica utilizada para selecionar descobrir as variáveis mais importantes para a construção do modelo, permitindo a remoção das menos importantes que possam afetar negativamente o modelo (PEREIRA, 2018).

1. **Resultados obtidos**

**4.1 - Estudo de correlação**

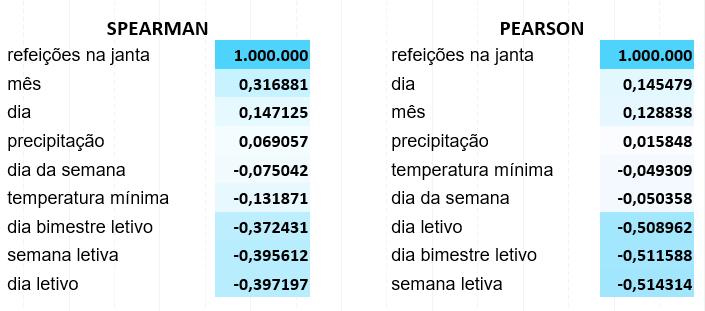
O estudo de correlação da variável “janta” com as demais não apresentou correlações fortes, como podemos ver na figura 4:

Tabela 2 - Resultado de estudo de correlação

Na figura 5 temos uma apresentação visual da correlação das variáveis, podemos perceber que os gráficos são muito dispersos, não apresentando nenhuma forma ou padrão definido.

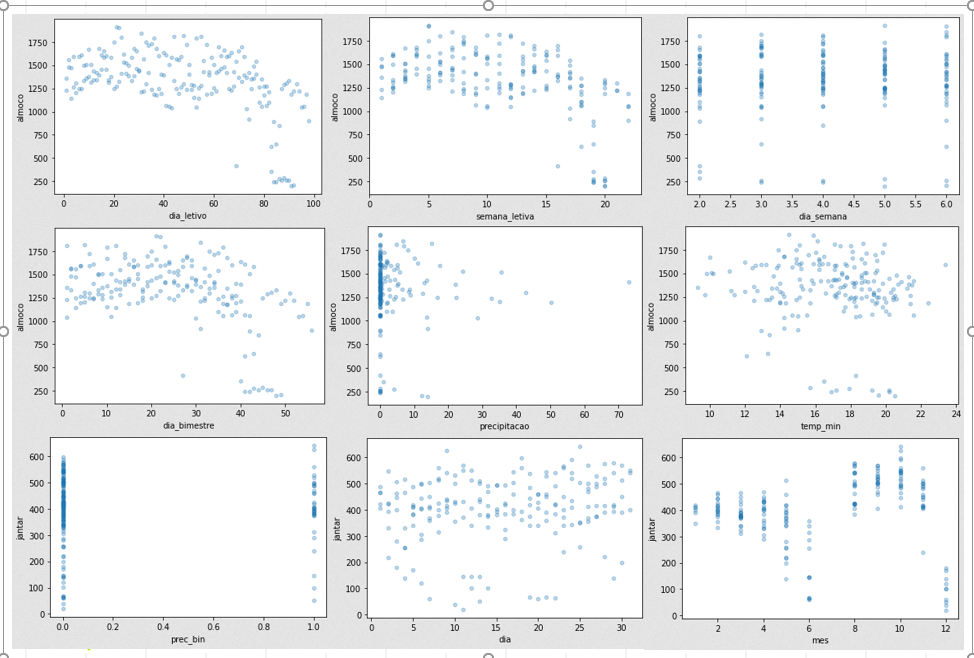


Figura 5 – Gráfico do estudo de correlação

É possível notar que alguns pontos ainda causam bastante estranheza devido ao fato de estarem muito fora da curva e sem motivo aparente. Estes pontos tão fora da curva e do esperado podem causar extrema confusão no algoritmo e prejudicar seu desempenho e precisão.

Abaixo temos um gráfico apresentando o desempenho atual do programa com a base de dados que está sendo usada. O número no canto superior esquerdo representa o score do algoritmo, ou seja, a sua precisão quando compara a previsão de refeições que deveriam ser servidas (que o algoritmo calculou) com as refeições que foram efetivamente servidas e que se encontram nos dados históricos da base.

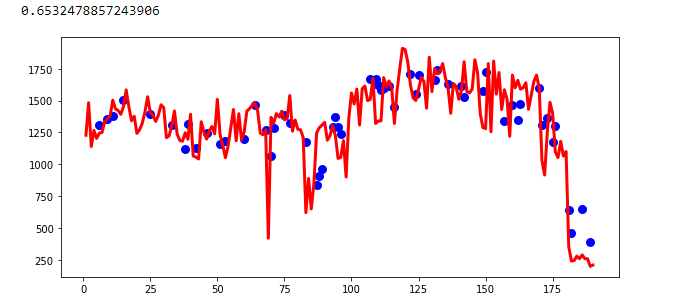


Figura 4 - Linha Vermelha o que foi produzido. Pontos azuis previstos pelo SVR

A linha em vermelho representa a quantidade de refeições servidas através dos dias no eixo X e os pontos em azul são a estimativa de refeições que deveriam ser preparadas que o software gerou para esses dias, após o treinamento com os dados. Na execução que gerou este gráfico, 70% dos dados foram usados para treinamento, e os outros 30% para testes, o que justifica a quantidade menor de pontos em azul que em vermelho.

1. **Considerações Finais**
2. **Referencial Bibliográfico**

PAULA, Ruben. MÁQUINAS DE SUPORTE VETORIAL COMO INSTRUMENTO DE PRIORIZAÇÃO DE INVESTIMENTOS APLICADO AO PROGRAMA

DE INVESTIMENTO E LOGÍSTICA – AEROPORTOS.

Orientador: DR. DONALD MATTHEW PIANTO. 2016. 268 p.

Dissertação (Mestrado em Economia do Setor Público) - Universidade de Brasiília, [S. l.],

2016.

----------------------------------------------

AWAD, Mariette; KHANNA, Rahul. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. [S. l.]: Apress, 2015. 268 p. ISBN 9781430259909. E-book.

----------------------------------------------

AWAD, Mariette; KHANNA, Rahul. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. [S. l.]: Apress, 2015. p 4. ISBN 9781430259909. E-book.

HAUGELAND, John. Artificial Intelligence: The Very Idea. Massachusetts: The MIT Press, 1985

FEIGENBAUM, A. E. Expert Systems, principles and pratice in the Encyclopedia of computer Science and Engineering, 1992

FERNANDES, Anita Maria da Rocha - Inteligência Artificial noções gerais – 2°, Florianópolis: VisualBooks, 2005

CHARNIAK, E. e McDERMOTT, D. (1985). Intradaylion to Artificial Intelligence. Addison-Wesley

WINSTON, P. H. (19921, Artificial Intelligence (Third -edition) Add.son-Wes ley.

JUNIOR, E. A.; TEIXEIRA, R. P. A. Manual de procedimentos para utilização de sobras alimentares. Modelo nutrição: módulo programação. Rio de Janeiro: SESC, 2010.

ABREU, E.S.; SIMONY, R.F.; DIAS, D.H.S.; RIBEIRO, F.R.O. Avaliação do desperdício alimentar na produção e distribuição de refeições de um hospital de São Paulo. Simbio-logias, v.5, n.7, p.42-50. 2012.

PEIXOTO, M.; PINTO, H. S. Desperdício de Alimentos: questões socioambientais, econômicas e regulatórias. Brasília: Núcleo de Estudos e Pesquisas/CONLEG/ Senado, fevereiro/2016 (Boletim Legislativo nº 41, de 2016). Disponível em: www.senado.leg.br/estudos. Acesso em 11 de Maio de 2020.

PEREIRA, D. X. R. Going zero waste in canteens: Exploring food demand using data analytics. Faculdade de Engenharia da Universidade do Porto, 18 de julho de 2018.

SCOTTON, V. et al. Desperdício de Alimentos em Unidades de Alimentação e Nutrição: a contribuição do resto-ingestão e da sobra. Revista Higiene Alimentar, v. 24, n. 186/187, p. 19-24, 2010.

Conselho Federal de Nutricionistas. RESOLUÇÃO CFN Nº 600, DE 25 DE FEVEREIRO DE 2018. Dispõe sobre a definição das áreas de atuação do nutricionista e suas atribuições, indica parâmetros numéricos mínimos de referência, por área de atuação, para a efetividade dos serviços prestados à sociedade e dá outras providências. Diário Oficial da União [Internet]. 2018 Maio. Disponível em: https://www.cfn.org.br/wp-content/uploads/resolucoes/Res\_600\_2018.htm

TEIXEIRA, S. M. F. G. et al. Administração aplicada às unidades de alimentação e nutrição. São Paulo: Atheneu, 2007.

COLOG, Boas Práticas na Manipulação de Alimentos, Gestão das Sobras – Como reduzir às Sobras, 2018;

AGGARWAL, Charu C. (Ed.). Data classification: algorithms and applications. CRC press, 2014.

CHATTERJEE, Samprit; HADI, Ali S. **Regression analysis by example**. John Wiley & Sons, 2015.

Para a parte textual do artigo, recomenda-se a seguinte estrutura:

1. Introdução OK
2. Referencial Teórico
3. Metodologia
4. Análise dos resultados
5. Considerações finais
6. Referências