MÉTODOS PREDITIVOS COMO FERRAMENTA PARA A PREVENÇÃO DE RESÍDUOS DE ALIMENTOS EM RESTAURANTE UNIVERSITÁRIO: ABORDAGEM BASEADA EM APRENDIZADO DE MÁQUINA

Palavras-Chave: [prevenção de resíduos], [resíduos de alimentos], [aprendizado de máquina]

Autores/as:

AMANDA LIZ MARSON DUTENHEFNER [FT-UNICAMP]
Prof. Dr. GUILHERME PALERMO COELHO (coorientador) [FT-UNICAMP]
Prof.ª Dr.ª CARMENLUCIA S. G PENTEADO (orientadora) [FT-UNICAMP]

INTRODUÇÃO:

Os resíduos de alimentos gerados no final da cadeia produtiva, especialmente no consumo, representam até 50% de todo o alimento perdido no mundo (GAIANI et al., 2018). Olhando para os restaurantes universitários, há relatos de que são a principal fonte de resíduos sólidos dentro de todo o campus (ABDELAAL, 2019). Os resíduos gerados nos restaurantes são classificados de duas formas: as sobras de bandejas de usuários (ELISSON et al., 2019), representando um terço de todo o alimento descartado, e as “sobras limpas”, alimentos preparados que não chegam a ser servidos, representando dois terços do total (ABDELAAL, 2019, DUTENHEFNER, 2021). Para a primeira fração, um método para reduzir a quantidade seria aumentar a conscientização entre os comensais, ou servir os alimentos em pratos menores, para evitar as sobras na bandeja e incentivar o consumo de tudo o que foi servido. Já para a segunda fração, de maior impacto, sua redução torna-se mais complexa pois envolve o planejamento de quantas refeições deverão ser servidas no dia e quanto alimento deve ser preparado para tal. Esta previsão contém muitas incertezas, pois a adesão do público é muito variável entre os dias da semana, do mês, período do semestre, eventos no dia, o cardápio sendo oferecido, entre outras possibilidades, o que é muito difícil de ser feito de forma manual.

Assim, para compreender o impacto real de cada possível fator na quantidade a ser prevista, e para conseguir estimar com mais precisão a quantidade de cada alimento a ser preparada para atender todo o público, um modelo de aprendizado de máquinas pode ser implementado. O aprendizado de máquinas vem ganhando espaço em muitos setores da sociedade e apresentando bons resultados para os fins desejados, seja no agrupamento e classificação de dados, na detecção de inconsistências nos registros, ou na previsão de cenários futuros (DE CASTRO; FERRARI, 2016), como é o nosso caso, mais especificamente relacionado à previsão de público em restaurantes e para a geração de resíduos sólidos.

Neste contexto, a presente pesquisa visa realizar uma revisão sistemática da literatura existente no âmbito da utilização do aprendizado de máquinas para otimização de preparo de refeições em restaurantes universitários. Almeja-se conhecer todos os trabalhos já realizados anteriormente neste contexto, entendendo quais algoritmos foram escolhidos e por quais motivos, quais parâmetros e atributos devem ser considerados, e qual acurácia de previsão pode ser esperada.

METODOLOGIA:

A pesquisa consistiu na realização de uma revisão sistemática de literatura, nas bases *Scopus* e *Web of Science*, com os termos de busca ("*machine learning*" AND ("*forecast**" OR "*predict**") AND ("*food waste*" OR "*restaurant*" OR "*catering*" OR "*dining*")). Tais bases foram escolhidas por abrangerem muitas outras bases como *SpringerLink*, *IEEE* e *ScienceDirect*, tornando a busca individual em cada uma destas redundante. Não foi definida restrição de período para a busca de artigos pois, dada a alta especificidade e novidade do assunto, era esperada dificuldade na obtenção de um grande volume de trabalhos de escopo correlato, então, optou-se por realizar a busca com todo o histórico disponível no tema – até dezembro de 2021, quando esta etapa da pesquisa foi concluída. Desta forma, 4064 artigos foram extraídos com as condições de pesquisa citadas, triados sequencialmente com base em suas palavras-chave, títulos e resumos, afunilando a pesquisa, conforme mostra a Figura 1.

Posteriormente à seleção dos artigos pertinentes, um fichamento dos trabalhos foi realizado visando a extração das principais informações de interesse para a construção de um modelo próprio, adaptado ao cenário dos Restaurantes Universitários da Unicamp de Limeira, como: o objeto de estudo (cenário, escopo, local), os tipos de algoritmos utilizados, os principais atributos considerados, o período dos dados para treino e teste do algoritmo, a eficiência das predições e as limitações de cada estudo. Estes dados foram tabelados, permitindo análise e conclusão sobre os caminhos possíveis a seguir no presente estudo.

RESULTADOS E DISCUSSÃO:

A pesquisa inicial resultou em 3959 artigos encontrados na base *Scopus* e 105 na *Web of* Science. Após as etapas de triagem e remoção de duplicatas, 40 artigos foram identificados como correlatos ao tema específico da presente pesquisa, e selecionados para leitura integral. Destes 40, a leitura completa indicou 11 artigos como relevantes para comporem a revisão sistemática, conforme Figura 1.

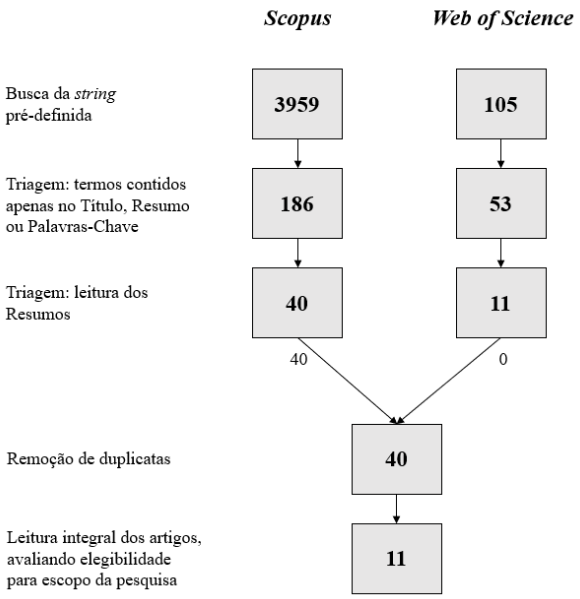

Figura 1: critérios de busca e seleção dos artigos

Conforme explicado anteriormente, o tema tem notável atualidade, de forma que todos os 11 artigos considerados pertinentes datam apenas dos últimos 4 anos. Além disso, apenas 1 deles é brasileiro, realizado na Universidade Federal de Uberlândia, publicado em 2021. Os demais, apesar de apresentarem metodologias e resultados interessantes, podem conter limitações para a aplicação no cenário aqui proposto, por se tratarem de culturas diferentes (Asia, Europa ou América do Norte) ou diferentes tipos de serviço de alimentação (restaurantes particulares, alimentação infantil ou produção industrial). Apenas 2 dos artigos selecionados estudaram o ambiente universitário, sendo esta então uma importante limitação para a interpretação dos resultados; este ambiente possui características particulares em comparação a outros restaurantes, sendo o fluxo de pessoas no campus muito dependente do calendário da graduação/pós-graduação, eventos no campus e do cardápio do dia.

Por fim, os métodos testados nos trabalhos para situações de predição como esta são:

- **Random Forest** (SANTOS et at., 2021; MALEFORS et al., 2020; TANIZAKI et al., 2020; GARRE et al., 2020; TANIZAKI et al., 2019; MA et al., 2018);
- **Redes Neurais Artificiais** (SANTOS et at., 2021; LUTOSLAWSKI et al., 2021; FAEZIRAD et al., 2021; MALEFORS et al., 2020;TSOUMAKAS, 2020);
- **ARIMA** (SU, 2020; TSOUMAKAS, 2020);
- **K-Nearest-Neighbors – KNN** (SANTOS et at., 2021; MA et al., 2018);
- **Linear Regression** (TANIZAKI et al., 2020; GARRE et al., 2020);
- **Boosting** (GARRE et al., 2020; MA et al., 2018);
- **STL decomposition** (DILKHUSH et al., 2020).

E os atributos citados como importantes a serem considerados no modelo são: **Dia, Mês, Dia da semana, Datas especiais (feriados, férias, greves), Cardápio / tipo de alimento, Quantidade de produção, Quantidade de pessoas**, Localidade, Preço da refeição, Taxa de reserva, Clima.

Assim, constata-se que os algoritmos mais comumente usados atualmente são **Redes Neurais Artificiais** e **Random Forest**. Pela análise dos trabalhos, é possível entender o contexto em que cada atributo foi utilizado, e compará-los à aplicação específica nos restaurantes dos campi de Limeira. Alguns dos atributos considerados nos trabalhos não são pertinentes ao Restaurante Universitário, como preço (invariável), clima, taxa de reserva (inexistente) e localidade (população praticamente invariável por campus); em paralelo, como visto na etapa anterior deste trabalho (DUTENHEFNER, 2021), a aderência ao RU pode também variar de acordo com a aceitação do cardápio do dia, eventos estudantis e semanas de provas, além de poderem ocorrer variações de quantidade internas ao tipo de alimento (por exemplo, não se pode considerar que cada pessoa sempre consumirá a mesma massa de salada, visto que um dia a salada é composta por alface – leve – e em outro dia, por tomate – mais denso), logo, alguns atributos possivelmente pertinentes não foram abordados pelos artigos.

Para ambos os métodos, uma etapa de pré-tratamento importante deve ser considerada antes da implementação dos algoritmos, criando variáveis próprias que ajudem a identificar datas especiais, taxa de aceitação de cardápios e outros, que não estão contidos nos dados fornecidos diretamente pelo Restaurante. Lembra-se, contudo, que a quantidade de assuntos relacionados ao tema específico da pesquisa ainda é incipiente, logo, a adaptação deve ser abordada com cautela, e os resultados encontrados no cenário da Unicamp de Limeira pode diferir dos resultados encontrados na revisão de literatura. Além disso, os estudos aqui citados não consideram a quantidade de cada tipo de alimento (arroz, feijão, salada, proteínas...) a ser preparado no dia, mas apenas a quantidade total de refeições, de público, a ser prevista. Esta diferenciação resultaria numa maior complexidade do algoritmo, porém poderia trazer melhor precisão e controle do preparo. Para isso, poderá ser prevista então uma combinação de algoritmos, que possa prever em duas etapas a produção diária: primeiro, o público esperado, com base nas descobertas aqui citadas; segundo, a quantidade de cada item, com base no cardápio diário.

CONCLUSÕES:

Pela falta de conteúdo sobre o tema, ainda é difícil concluir com certeza qual o algoritmo mais indicado para o objetivo aqui proposto. Os dois estudos que foram realizados em restaurantes universitários (SANTOS et al., 2021; FAEZIRAD et al., 2021) relatam que as redes neurais artificiais foram bem-sucedidas na previsão. Santos et al. (2021) comparou algoritmos, indicando também Random Forest como um bom modelo, enquanto Faezirad et al. (2021) implementou apenas uma rede neural no projeto. As redes neurais artificiais são muito mais complexas que os demais algoritmos de

serem implementadas, logo, é necessário estudar se, para o cenário da Unicamp de Limeira, realmente seria interessante aplicá-la ou se um modelo mais simples já atenderia.

Dada a falta de unanimidade entre os autores sobre os modelos preditivos, o ideal seria testar um conjunto de algoritmos em paralelo (inclusive os menos testados pelos artigos citados, como KNN ou regressões lineares, dada sua relativa simplicidade de implementação) e comparar a facilidade de uso e a eficiência de predição, como foi realizado pela maioria dos autores citados. Todos os métodos possuem pontos positivos e negativos, mas como os estudos diferem muito entre si quanto ao escopo do projeto, ainda é difícil concluir sobre algum método sem que todos sejam testados no cenário correto.

Com a revisão, contudo, foi possível entender a importância dos atributos relacionados às datas para que se tenha uma previsão eficiente. Para qualquer algoritmo escolhido, deverão ser incluídos atributos relacionados então à temporalidade, sazonalidade, e às características intrínsecas ao campus, assim como atributos relacionados à aceitabilidade dos alimentos, se possível. Nota-se também a importância de se obter uma base de dados robusta, idealmente com pelo menos um ano completo de registros, para que se tenham dados suficientes relativos aos diferentes períodos possíveis, e disponibilidade de dados para treino do modelo, e em seguida para testes.

Para a continuidade e futura conclusão desta pesquisa, dados mais completos e padronizados devem ser coletados, por um período também mais longo, permitindo que, uma vez os modelos construídos, eles possam ser testados e implementados eficientemente.

BIBLIOGRAFIA

ABDELAAL, A. H.; MCKAY, G.; MACKEY, H. R. **Food waste from a university campus in the Middle East: Drivers, composition, and resource recovery potential**. Waste Management, v. 98, p.14-20, 2019. DOI: 10.1016/j.wasman.2019.08.007

DE CASTRO, L. N.; FERRARI, D. G. **Introdução à Mineração de Dados - Conceitos Básicos, Algoritmos e Aplicações.** São Paulo: Saraiva. 351 p. 2016.

DILKHUSH, D.; JOSEPH, J. T.; RENISHA, B. **Time series analysis for supply chain planning in restaurants**. ICCCS. 2020.

DUTENHEFNER, A. L. M. **Prevenção de Resíduos de Alimentos: O Campus Universitário como um laboratório integrador de pesquisa e ação**. Projeto de Iniciação Científica, Universidade Estadual de Campinas, 2020-2021.

ELLISON, B.; SAVCHENKO, O.; NIKOLAUS, C. J.; BRITTANY, R.L.; DUFF, B. R. L. **Every plate counts: Evaluation of a food waste reduction campaign in a university dining hall.** Resources Conservation & Recycling, v. 144, p. 276-284, 2019. DOI: 10.1016/j.resconrec.2019.01.046

FAEZIRAD, M; et al. **Preventing food waste in subsidy-based university dining systems: An artificial neural network-aided model under uncertainty**. Waste, Management & Research, v. 39, p. 1027-1038. 2021.

GAIANI, S.; CALDEIRA, S.; ADORNO, V.; SEGRÉ, A.; VITTUARI, M. **Food wasters: profiling consumers' attitudes to waste food in Italy**. Waste Management, v. 72, p. 17-24, 2017.

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, v. 7. 2020.

LUTOSLAWSKI, K; et al. F**ood Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network**. IEEE Access, v. 9. 2021.

MA, X. et al. **PREDICTING FUTURE VISITORS OF RESTAURANTS USING BIG DATA**. International Conference on Machine Learning and Cybernetics, 2018.

MALEFORS, C. et al. **Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in Swedish public catering during the Covid-19 pandemic**. Socio-Economic Planning Sciences. 2021.

SANTOS, Y. D. C.; et al. **Um método baseado em aprendizado de máquina para previsão da produção de refeições em restaurantes universitários**. SBSI, 2021.

SU, C. **Passenger Flow Forecast of Catering Business based on Autoregressive Integrated Moving Average and Smoothing Index Prediction Model**. International Signal Processing, Communications and Engineering Management Conference (ISPCEM), 2020.

TSOUMAKAS, G. **A survey of machine leaning techniques for food sales prediction**. Artif Intell Rev, v. 52, p. 441-447. 2019.

TANIZAKI, T. et al. **Demand forecasting in restaurants using machine learning and statistical analysis**. Procedia CIRP, v. 79, p. 679-683. 2019.

TANIZAKI, T. et al. **Forecasting customers visiting using machine learning and characteristics analysis with low forecasting accuracy days**. International Federation for Information Processing, 2020.