

# Sistema de Predição de Valores Nutricionais em Receitas Culinárias através de Métodos de Aprendizagem de Máquina

Luciano D. S. Pacifico<sup>1</sup>, Emília G. de Oliveira<sup>1</sup>, Larissa F. S. Britto<sup>1</sup>,  
Teresa B. Ludermir<sup>2</sup>

<sup>1</sup>Departamento de Computação (DC) – Universidade Federal Rural de Pernambuco (UFRPE) – Recife – PE – Brazil

<sup>2</sup>Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE) – Recife - PE - Brazil

{luciano.pacifico, emilia.galdino, larissa.feliciana}@ufrpe.br

tbl@cin.ufpe.br

**Abstract.** *In this work, a system to predict nutritional values for food recipes is proposed. An automatic and data-driven approach to estimate nutritional values for new recipes is employed, developed through Machine Learning techniques. The proposed prediction system will be a central piece on the design of new Recipe Recommendation Systems, where complete personal dietary plans, contemplating both the user nutritional needs (according to daily nutrition standards) and expectations, are going to be fulfilled.*

**Resumo.** *Neste trabalho, uma proposta de sistema de predição de valores nutricionais para receitas culinárias será apresentada. Uma abordagem voltada aos dados (data-driven) para a estimação automática dos valores nutricionais de novas receitas será empregada, desenvolvidas através de técnicas de Aprendizagem de Máquina. O sistema de predição proposto será usado como base para a elaboração de Sistemas de Recomendação de Receitas Culinárias, que consigam estimar as necessidades nutricionais de seus usuários, de modo a gerar dietas completas que contemplem tanto os padrões nutricionais diários recomendados, quanto às expectativas desses usuários.*

## 1. Introdução

Nos dias atuais, o desenvolvimento tecnológico possibilitou uma enorme transformação no hábito de compartilhamento de receitas culinárias. Se antes receitas eram compartilhadas via oral (como tradições familiares) ou através de livros e revistas, atualmente é cada vez mais comum que as pessoas procurem por receitas através de sistemas computacionais, seja em *web sites* ou aplicativos especializados.

Porém, se por um lado o compartilhamento de receitas foi facilitado em decorrência do advento desses sistemas, por outro, a qualidade das receitas disponibilizadas, do ponto de vista nutricional, é bastante questionável. Isso ocorre pelo fato de que a grande maioria desses *web sites* e aplicativos é desenvolvida de forma colaborativa com sua comunidade de usuários, onde qualquer pessoa com acesso aos mesmos pode produzir e fazer *upload* de documentos, sem que haja qualquer tipo de cuidado com a qualidade nutricional do conteúdo compartilhado [Trattner and Elsweiler 2017]. Como as receitas

submetidas pelos usuários não são checadas por especialistas, os sistemas atuais geralmente acabam não oferecendo qualquer tipo de informação nutricional das mesmas, o que pode vir a acarretar em prejuízos à saúde de seus usuários, dado o fato de que existem diversos problemas de saúde crônicos relacionadas à má alimentação, como doenças cardíacas, obesidade, diabetes, hipertensão, e alguns tipos de câncer [Toledo et al. 2019].

Como uma tentativa de auxílio aos usuários desses repositórios, surgiram os Sistemas de Recomendação de Receitas Culinárias (SRRCs) [Gorbonos et al. 2018, Mokdara et al. 2018, Nirmal et al. 2018, Oliveira et al. 2019, Pacifico et al. 2019, Toledo et al. 2019]. Porém, poucos trabalhos na área de SRRCs focam em técnicas que, além de gerarem recomendações que satisfaçam às expectativas de seus usuários (com técnicas tradicionais de filtragem), ofereçam também receitas e dietas nutricionalmente equilibradas, como em [Toledo et al. 2019]. Mesmo para os poucos SRRCs que levam informações do usuário e nutricionais em consideração, há ainda a necessidade de que todas as receitas de suas bases de dados sejam categorizadas nutricionalmente *a priori* (antes da etapa de treinamento/filtragem) por especialistas, o que exige uma grande demanda de esforço humano, podendo gerar um gargalo no que diz respeito à disponibilização de novas receitas, além de tornar os custos de manutenção desses sistemas elevados.

Neste trabalho, um sistema de predição de valores nutricionais de receitas é proposto, no intuito de realizar uma categorização nutricional automática de novas receitas. O sistema automático estimará o valor nutricional de novas receitas no momento de suas submissões pelos usuários, sem que haja a necessidade da consulta a especialistas. A abordagem adotada é voltada aos dados (*data-driven*), sendo necessária uma base de dados com **receitas categorizadas nutricionalmente inicialmente**, pela qual os modelos de Aprendizagem de Máquina serão **treinados**, e em seguida, novas receitas poderão ter seus valores nutricionais estimados. O presente trabalho dá continuidade e estende o projeto de Sistemas de Recomendação e Geração de Receitas iniciado em [Britto et al. 2019, Oliveira et al. 2019] pelo acréscimo de um módulo de estimação dos valores nutricionais. O trabalho está dividido da seguinte forma: a metodologia de desenvolvimento do sistema a ser proposto será apresentada na próxima seção (Seção 2), e em seguida serão apresentadas as conclusões e próximas etapas da projeto (Seção 3).

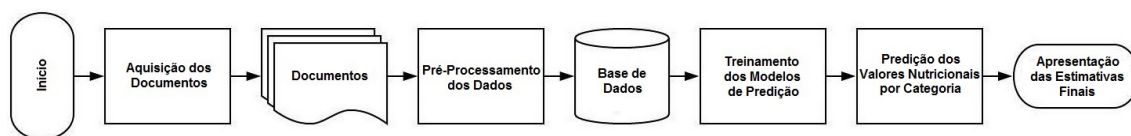
## 2. Metodologia

A metodologia adotada é apresentada na Fig. 1. O sistema proposto é dividido em duas etapas principais: a aquisição da base de dados inicial e o treinamento dos modelos de Aprendizagem de Máquina. A base de dados utilizada neste trabalho foi proposta em [Majumder et al. 2019], extraída da página Food.com<sup>1</sup>. Esta base de dados é composta por receitas no idioma Inglês, sendo as faixas nutricionais apresentadas de acordo com os padrões diários recomendados norte-americanos.

Embora a base de dados já apresente um pré-processamento dos documentos (as receitas na base de dados original já contém lista de ingredientes e medidas nas categorias nutricionais separadas), para o presente trabalho, uma etapa de extração e padronização das medidas dos ingredientes fez-se necessária. Essa padronização visa a elaboração de uma **matriz de frequência de ocorrência dos ingredientes** que será fornecida para o

---

<sup>1</sup><https://www.kaggle.com/shuyangli194/food-com-recipes-and-user-interactions>



**Figura 1. O Sistema de Predição de Valores Nutricionais das Receitas Proposto.**

treinamento dos algoritmos de aprendizagem de máquina. Como as receitas apresentam várias medidas diferentes para um mesmo ingrediente (por exemplo, "xícara", "colher de sopa", "grama", "copo americano", são medidas comumente atribuídas ao ingrediente "açúcar"), há a necessidade de eliminação de ambiguidades antes da etapa de treinamento dos modelos, de modo que uma medida única (*denominador comum*) foi adotada para cada ingrediente da base. Ao todo, um total de 102 medidas diferentes foram identificadas na base de receitas. Algumas medidas contidas nessa base são de difícil padronização (como por exemplo, "pitada" ou "a gosto"), tendo as receitas que contém ingredientes com esse tipo de medidas sido desconsideradas nesta etapa do projeto, porém, com a continuidade do processo de padronização de medidas, tais receitas poderão vir a ser consideradas em etapas futuras da pesquisa. Um total de 6226 ingredientes e 24079 receitas foram padronizados, o que é uma quantidade suficientemente grande para o início dos trabalhos.

Após a geração da matriz de frequência de ocorrência dos ingredientes (uma matriz de dimensões 24079 x 6226, inicialmente), os algoritmos de aprendizagem de máquina serão treinados da seguinte forma: para cada categoria nutricional (**gorduras, fibras, carboidratos, proteínas, sais, colesterol e calorias**), um algoritmo supervisionado de regressão (como a Regressão Linear, Máquina de Vetores de Suporte, ou Redes Neurais Artificiais)[Haykin 1994] será treinado individualmente, para a estimação de valores em cada uma dessas categorias. A estimativa final gerada pelo sistema para receitas novas será dada pela apresentação dos valores individuais estimados em cada categoria nutricional.

Atualmente, o projeto se encontra em fase de padronização de novas medidas (o que expandirá a base de dados inicial, tanto em número de receitas, quanto em número de ingredientes) e de treinamento dos algoritmos de regressão. Uma abordagem de **Validação Cruzada** será usada para avaliar a qualidade de cada regressor em cada uma das medidas, de modo que o sistema final será composto pela seleção dos melhores algoritmos para cada categoria nutricional. Ainda neste sentido, uma abordagem derivada, que leva em consideração apenas os ingredientes como forma de predição dos valores nutricionais em cada categoria em faixas pré-estabelecidas (*baixo, normal e alto*, de acordo com os padrões estabelecidos pelo FDA<sup>2</sup>), está em andamento, seguindo uma abordagem de **problema de classificação de padrões**.

### 3. Conclusões

Neste trabalho, uma proposta de sistema de predição automática de valores nutricionais em receitas culinárias é proposto, baseado em técnicas de aprendizagem de máquina. O passo-a-passo do desenvolvimento do novo sistema é apresentado, assim como o estado atual da pesquisa. A continuidade do trabalho será dada pela avaliação de diferentes algoritmos de regressão na geração das estimativas dos valores nutricionais, sendo o

<sup>2</sup> <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>

melhor modelo obtido escolhido para compor o sistema de predição final. Além disso, informações presentes nos modos de preparo das receitas serão extraídas e processadas para complementar o treinamento dos algoritmos adotados. Após a conclusão dos testes iniciais, o sistema de predição de valores nutricionais elaborado será usado como parte integrante dos sistemas de recomendação e geração de receitas culinárias desenvolvidos no decorrer do projeto maior dos autores.

## Agradecimentos

Os autores gostariam de agradecer à FACEPE, ao CNPq e a CAPES pelo suporte financeiro.

## Referências

- Britto, L. F. S., Oliveira, E. G., Pacifico, L. D. S., and Ludermir, T. B. (2019). A text analysis approach for cooking recipe classification based on brazilian portuguese documents. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)*, 1:1–12.
- Gorbonos, E., Liu, Y., and Hoàng, C. T. (2018). Nutrec: Nutrition oriented online recipe recommender. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 25–32. IEEE.
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- Majumder, B. P., Li, S., Ni, J., and McAuley, J. (2019). Generating personalized recipes from historical user preferences. *arXiv preprint arXiv:1909.00105*.
- Mokdara, T., Pusawiro, P., and Harnsomburana, J. (2018). Personalized food recommendation using deep neural network. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*, pages 1–4. IEEE.
- Nirmal, I., Caldera, A., and Bandara, R. D. (2018). Optimization framework for flavour and nutrition balanced recipe: A data driven approach. In *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pages 1–9. IEEE.
- Oliveira, E. G., Britto, L. F. S., Pacifico, L. D. S., and Ludermir, T. B. (2019). Recipe recommendation and generation based on ingredient substitution. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)*, 1:1–12.
- Pacifico, L. D. S., Oliveira, E. G., Britto, L. F. S., and Ludermir, T. B. (2019). Sistemas de recomendação e geração de receitas através da categorização ontológica dos ingredientes. In *Symposium in Information and Human Language Technology (STIL 2019)*, volume 1, pages 81–85. SBC.
- Toledo, R. Y., Alzahrani, A. A., and Martínez, L. (2019). A food recommender system considering nutritional information and user preferences. *IEEE Access*, 7:96695–96711.
- Trattner, C. and Elswiler, D. (2017). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498. International World Wide Web Conferences Steering Committee.