

Predicting Adherence to Canada's Food Guide Recommendations on Healthy Food Choices Using Machine Learning Algorithms

Mélina Côté, Didier Brassard, Julie Robitaille, Marie-Claude Vohl, Simone Lemieux, and Benoît Lamarche

Centre Nutrition, santé et société (NUTRISS), INAF, Université Laval

Objectives: Machine learning (ML) algorithms can potentially improve predictive performances compared to traditional statistical models. The aim of this study was to predict adherence to the 2019 Canada's Food Guide (CFG) recommendations on healthy food choices using ML and a large array of variables/features related to dietary habits.

Methods: In a sample of 1147 French-speaking adults (50% women) from the PREDISE study, Healthy Eating Food Index (HEFI-2019) scores were calculated using data from three unannounced web-based 24h recalls. Adherence to the 2019 CFG recommendations on healthy food choices (yes or no) was measured with the HEFI-2019 and arbitrarily defined as a score $\geq 46.7/80$ points. This value corresponds to the median HEFI-2019 score for adult women in Canada. A total of 2452 features encompassing individual, social and environmental characteristics related to dietary habits were retained as predictors in the analyses. Decision tree (DT) and Adaboost ML algorithms were developed, calibrated and then compared us-

ing accuracy score (proportion of correct predictions), precision score (positive predictive value) and recall score (sensitivity). All analytical steps were bootstrapped 100 times to generate 95%CI. The most important features retained by each ML algorithm were compared.

Results: The DT predicted adherence to the 2019 CFG recommendations on healthy food choices with an accuracy of 0.65 (95%CI: 0.59–0.71), a precision of 0.64 (95%CI: 0.44–0.84) and a recall of 0.31 (95%CI: 0.10–0.52). Adaboost had similar predictive performance metrics with an accuracy of 0.64 (95%CI: 0.59–0.69), a precision of 0.56 (95%CI: 0.45–0.67) and a recall of 0.49 (95%CI: 0.39–0.59). However, among the 15 most important features retained by each ML algorithm, only 6 features (40%) were shared by both.

Conclusions: The use of DT and Adaboost ML algorithms does not predict adherence to the 2019 CFG recommendations on healthy food choices measured by the HEFI-2019 score with high accuracy. The inconsistencies in the features retained by each ML algorithm also suggest that results are model-dependent. Further research is therefore necessary to successfully implement ML approaches that may help better predict adherence to dietary recommendations such as those found in the 2019 CFG.

Funding Sources: Instituts de la recherche en santé du Canada, Fonds de recherche du Québec - Santé.