

## Use of Machine Learning for the study of Obesity in Chile

**Problem:** 74.2% of Chilean population is overweight or obese<sup>1</sup>, risk factor for non-communicable diseases that causes premature mortality and Years of Potential Life Lost<sup>2</sup>. Obesity is due to an imbalance between energy consumption and expenditure, which is affected by individual physiological factors and interaction with the environment<sup>3</sup>. Interventions carried out to control obesity that aimed to modify food's consumption have not produced the expected effect<sup>4</sup>. Due to the problem's complexity new approaches might give greater precision.

**Solution:** Machine Learning (ML) methods emerge as useful tools in the study of obesity and diet, since they have the potential to consider a large number of predictor variables in order to generate a model flexible enough to represent a given phenomenon, with predictive properties generated from data<sup>5,6</sup>. The use ML methods will be used to determine whether food consumption (FC) and sociodemographic (SOC) variables are associated with overweight or obesity in the Chilean population.

**Methods:** A national diet assessment for Chile was available from the 2008 National Survey of Food Consumption (ENCA)<sup>7</sup>. ML algorithms were implemented to predict the nutritional status of the respondents via 5 classification and 5 regression algorithms, with and without sociodemographic variables as part of the predictor variables.

**Variables and Metrics:** ENCA includes anthropometric information, food consumption and sociodemographic variables. The algorithms were compared using ANOVA test according to performance metrics for classification (accuracy) and regression (root-mean-square error). Exploratory analysis included correlation, variance and principal component analysis.

**Hypothesis:** Nutritional status of the Chilean population can be predicted from the consumption of FC and/or SOC of the Chilean population.

**Objective:** Implement ML algorithms for classification and regression to predict the nutritional status (normal/overweight/obese or BMI) of the Chilean adult population, based on data on FC, SOC characteristics or the combination of both.

**Results:** FC variables are characterized by low correlation and low explained variance. For classification the best accuracy was  $72.2\% \pm 2.6$  obtained for FC+SOC using Support Vector Machines. Regression showed the best RMSE:  $5.2 \pm 0.4$  with FC+SOC with Linear Regression. Statistical tests for comparison between algorithms were not significant.

**Outlook:** Self-reported diet is noise data<sup>8</sup>, and even if it was not possible to predict to predict accurately nutritional status, these algorithms can be applied to a variety of public health surveys.

## REFERENCES

1. Ministerio de Salud Chile, Departamento de Epidemiología. *ENCUESTA NACIONAL DE SALUD 2016-2017 Segunda Entrega de Resultados.*; 2018. [http://www.minsal.cl/wp-content/uploads/2017/11/ENS-2016-17\\_PRIMEROS-RESULTADOS.pdf](http://www.minsal.cl/wp-content/uploads/2017/11/ENS-2016-17_PRIMEROS-RESULTADOS.pdf). Accessed July 10, 2018.
2. Bedregal P, Margozzini P, González C. *Informe Final: Estudio de Carga de Enfermedad y Carga Atribuible.*; 2008. [http://www.cienciasdelasalud-udla.cl/portales/tp76246caadc23/uploadimg/File/Informe-final-carga-Enf-2007.pdf](http://www.cienciasdelasalud.udla.cl/portales/tp76246caadc23/uploadimg/File/Informe-final-carga-Enf-2007.pdf).
3. Butland B, Jebb S, Kopelman P, et al. *Foresight Tackling Obesity: Future Choices – Project Report.*; 2007. Government Office for Science, UK.
4. Lee BJ, Kim KH, Ku B, Jang J-S, Kim JY. Prediction of body mass index status from voice signals based on machine learning for automated medical applications. *Artif Intell Med.* 2013;58(1):51-61. doi:10.1016/J.ARTMED.2013.02.001
5. James G, Witten D, Tibshirani R, Hastie T. An Introduction to Statistical Learning with Applications in R. *Book.* 2013:431. doi:10.1007/978-1-4614-7138-7
6. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA.* 2018;319(13):1317. doi:10.1001/jama.2017.18391
7. Departamento de Nutrición, Escuela de Nutrición, Escuela de Salud Pública, Centro de Microdatos. *Encuesta Nacional de Consumo Alimentos.*; 2014. [http://web.minsal.cl/sites/default/files/ENCA-INFORME\\_FINAL.pdf](http://web.minsal.cl/sites/default/files/ENCA-INFORME_FINAL.pdf).
8. Becker W, Welten D. Under-reporting in dietary surveys--implications for development of food-based dietary guidelines. *Public Health Nutr.* 2001;4(2B):683-687. doi:10.1079/PHN2001154