Nova Analytics Group is a students' organization founded in 2016 to integrate students, promote activities, and share knowledge about Data Science at Universidade Nova de Lisboa. Currently, the organization gathers more than 20 active members from several countries, including bachelors, masters and professors. By engaging students through the organization, we can promote personal and professional development, establish partnerships with local companies and develop projects with social impact.

The group is currently representing the Information Management School from UNL (Nova IMS) in a project with the Queen's University in Canada focused on food security and sustainability, based on guidelines proposed by the World Health Organization (WHO). In this context, Nova Analytics Group is currently developing a study for improving nutrition and modelling predictions for diagnosing diabetes in Brazil.

During an initial assessment, we could not find a significant prevalence of the disease associated with ethnicity, income, education, gender, among other factors; by using statistical methods, the most relevant variables identified were age and BMI (body mass index).

At the moment, the best-performing model has a sensitivity level (true positive rate) of 65% for sick individuals, and specificity level (true negative rate) at 77% for healthy individuals. Accuracy level wasn't considered for benchmarking the developed models given the unbalanced nature of the dataset (in our sample, less than 7 in every 100 records are identified diabetic individuals), high accuracy levels usually created a bias for predicting healthy individuals and worse results on predicting individuals with diabetes.

The lack of relevant variables and time series data certainly plays a huge role on the prediction power of a model. The project assessment will suggest using new approaches for continuously obtaining relevant medical data for improving prevention and diagnosis. However, sensitive patients' data should require specific policies for safeguarding data privacy, security, and liability for bad use of sensitive data.

In order to carry out this project, we would like to briefly discuss the following questions, focusing specifically on the case study at hand:

1) Overview about the disease
2) Effects of Diabetes on Brazil
3) Policies in place to improve prevention, treatment and diagnosis

Thanks!

Marcel Motta