ML Bootcamp Project Proposal
Kholoud Khateeb

Project objective

Develop a machine-learning model that analyzes a **nutrition dataset** to understand the features of the data and predict the countries producing food items with high nutrition scores based on the French scoring system "`nutrition-score-fr_100g`".

Machine Learning Algorithm

Two models will be evaluated: regression and neural networks

Possible data source:
https://www.kaggle.com/openfoodfacts/world-food-facts

Dataset structure

The dataset contains a single table, FoodFacts, in CSV form in FoodFacts.csv and in SQLite form in database.sqlite. [163 columns and 356,027 total values].
File name: en.openfoodfacts.org.products.tsv

After removing some of the irrelevant columns and rows the data size is now 49040 rows × 123 columns. Possibly this number will go lower with further cleaning for the data.

Appendix:
Original Dataset colonms:
The columns in Open Food Facts are as follows:

- code (text)
- url (text)
- creator (text)
- created_t (text)
- created_datetime (text)
- last_modified_t (text)
- last_modified_datetime (text)
- product_name (text)
- generic_name (text)
- quantity (text)
- packaging (text)
- packaging_tags (text)
- brands (text)
- brands_tags (text)
- categories (text)

- categories_tags (text)
- categories_en (text)
- origins (text)
- origins_tags (text)
- manufacturing_places (text)
- manufacturing_places_tags (text)
- labels (text)
- labels_tags (text)
- labels_en (text)
- emb_codes (text)
- emb_codes_tags (text)
- first_packaging_code_geo (text)
- cities (text)
- cities_tags (text)
- purchase_places (text)
- stores (text)
- countries (text)
- countries_tags (text)
- countries_en (text)
- ingredients_text (text)
- allergens (text)
- allergens_en (text)
- traces (text)
- traces_tags (text)
- traces_en (text)
- serving_size (text)
- no_nutriments (numeric)
- additives_n (numeric)
- additives (text)
- additives_tags (text)
- additives_en (text)
- ingredients_from_palm_oil_n (numeric)
- ingredients_from_palm_oil (numeric)
- ingredients_from_palm_oil_tags (text)
- ingredients_that_may_be_from_palm_oil_n (numeric)
- ingredients_that_may_be_from_palm_oil (numeric)
- ingredients_that_may_be_from_palm_oil_tags (text)
- nutrition_grade_uk (numeric)
- nutrition_grade_fr (text)
- pnns_groups_1 (text)
- pnns_groups_2 (text)
- states (text)
- states_tags (text)
- states_en (text)
- main_category (text)
- main_category_en (text)
- image_url (text)
- image_small_url (text)
- energy_100g (numeric)
- energy_from_fat_100g (numeric)
- fat_100g (numeric)

- saturated_fat_100g (numeric)
- butyric_acid_100g (numeric)
- caproic_acid_100g (numeric)
- caprylic_acid_100g (numeric)
- capric_acid_100g (numeric)
- lauric_acid_100g (numeric)
- myristic_acid_100g (numeric)
- palmitic_acid_100g (numeric)
- stearic_acid_100g (numeric)
- arachidic_acid_100g (numeric)
- behenic_acid_100g (numeric)
- lignoceric_acid_100g (numeric)
- cerotic_acid_100g (numeric)
- montanic_acid_100g (numeric)
- melissic_acid_100g (numeric)
- monounsaturated_fat_100g (numeric)
- polyunsaturated_fat_100g (numeric)
- omega_3_fat_100g (numeric)
- alpha_linolenic_acid_100g (numeric)
- eicosapentaenoic_acid_100g (numeric)
- docosahexaenoic_acid_100g (numeric)
- omega_6_fat_100g (numeric)
- linoleic_acid_100g (numeric)
- arachidonic_acid_100g (numeric)
- gamma_linolenic_acid_100g (numeric)
- dihomo_gamma_linolenic_acid_100g (numeric)
- omega_9_fat_100g (numeric)
- oleic_acid_100g (numeric)
- elaidic_acid_100g (numeric)
- gondoic_acid_100g (numeric)
- mead_acid_100g (numeric)
- erucic_acid_100g (numeric)
- nervonic_acid_100g (numeric)
- trans_fat_100g (numeric)
- cholesterol_100g (numeric)
- carbohydrates_100g (numeric)
- sugars_100g (numeric)
- sucrose_100g (numeric)
- glucose_100g (numeric)
- fructose_100g (numeric)
- lactose_100g (numeric)
- maltose_100g (numeric)
- maltodextrins_100g (numeric)
- starch_100g (numeric)
- polyols_100g (numeric)
- fiber_100g (numeric)
- proteins_100g (numeric)
- casein_100g (numeric)
- serum_proteins_100g (numeric)
- nucleotides_100g (numeric)
- salt_100g (numeric)

- sodium_100g (numeric)
- alcohol_100g (numeric)
- vitamin_a_100g (numeric)
- beta_carotene_100g (numeric)
- vitamin_d_100g (numeric)
- vitamin_e_100g (numeric)
- vitamin_k_100g (numeric)
- vitamin_c_100g (numeric)
- vitamin_b1_100g (numeric)
- vitamin_b2_100g (numeric)
- vitamin_pp_100g (numeric)
- vitamin_b6_100g (numeric)
- vitamin_b9_100g (numeric)
- vitamin_b12_100g (numeric)
- biotin_100g (numeric)
- pantothenic_acid_100g (numeric)
- silica_100g (numeric)
- bicarbonate_100g (numeric)
- potassium_100g (numeric)
- chloride_100g (numeric)
- calcium_100g (numeric)
- phosphorus_100g (numeric)
- iron_100g (numeric)
- magnesium_100g (numeric)
- zinc_100g (numeric)
- copper_100g (numeric)
- manganese_100g (numeric)
- fluoride_100g (numeric)
- selenium_100g (numeric)
- chromium_100g (numeric)
- molybdenum_100g (numeric)
- iodine_100g (numeric)
- caffeine_100g (numeric)
- taurine_100g (numeric)
- ph_100g (numeric)
- fruits_vegetables_nuts_100g (numeric)
- collagen_meat_protein_ratio_100g (numeric)
- cocoa_100g (numeric)
- chlorophyl_100g (numeric)
- carbon_footprint_100g (numeric)
- nutrition_score_fr_100g (numeric)
- nutrition_score_uk_100g (numeric)

carbon-footprint_100g : carbon footprint (as indicated on the packaging of some products)

nutrition-score-fr_100g : Nutri-Score - Nutrition score derived from the UK FSA score and adapted for the French market (formula defined by the team of Professor Hercberg)

nutrition-score-uk_100g : nutrition score defined by the UK Food Standards Administration (FSA)
https://static.openfoodfacts.org/data/data-fields.txt