Data Science

Predicting Carbon Footprint For Food Products

Let’s find out if you buy in a sustainable way!

João Alves
Tomás Novo
Melvin Lundqvist

01 INTRODUCTION

02 MOTIVATION

03 DATASETS

04 LIMITATIONS

05 MODELS

06 FINAL PRODUCTS

07 IMPROVEMENTS

08 CONCLUSION

01

Introduction

The context

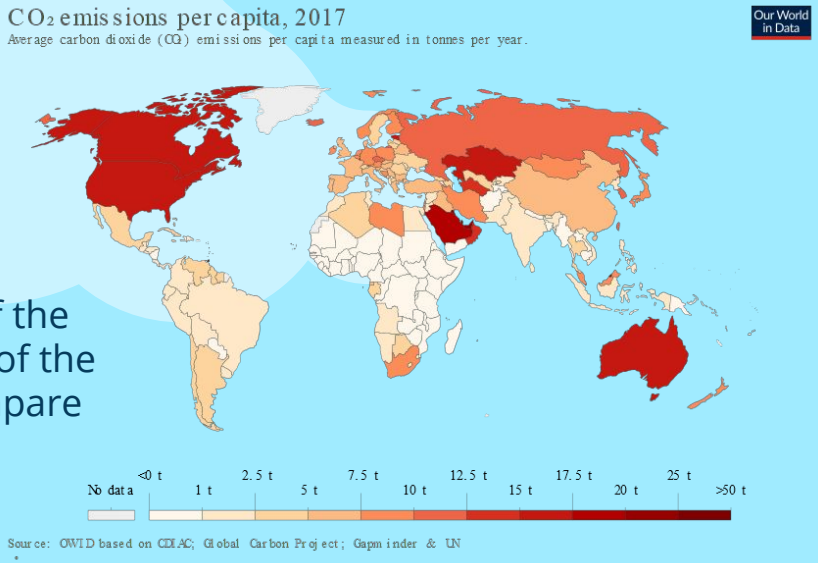
Introduction

What is a carbon footprint?

Total amount of greenhouse gas (GHG) emissions that a person/organization/product causes in its life cycle

Why calculate it?

To have a measure of the environmental impact of the products etc. and compare them



02

Motivation

How we ended up with this topic

Motivation

170.000.000

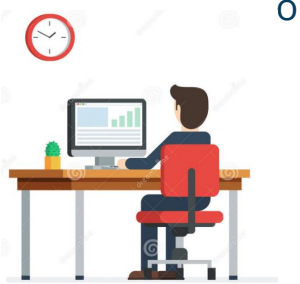
TONS OF CO2 EMITTED

Only of meat products in 2016

Motivation

Awareness

Easily accessible data



Create Interest

Is it possible to change our habits?

Change mindsets

Small actions can make big changes

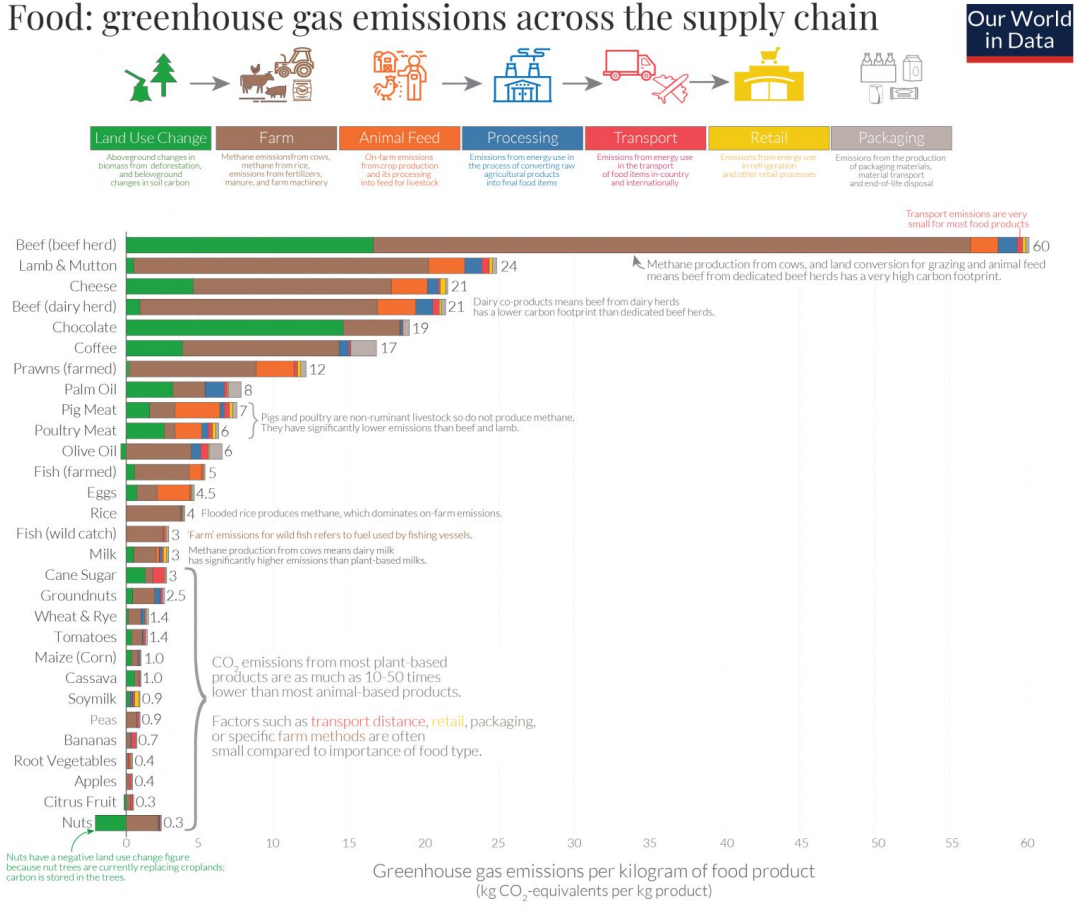
03

Datasets

The material

Datasets

Emission data



Note: Greenhouse gas emissions are given as global average values based on data across 38,700 commercially viable farms in 119 countries.
Data source: Poore and Nemecek (2018). Reducing food's environmental impacts through producers and consumers. *Science*. Images sourced from the Noun Project.
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Datasets

Nutrition data

name	calories	saturated_...	cholesterol	sodium	choline
Cornstarch	381		0	9.00 mg	0.4 mg
Nuts, pecans	691	6.2g	0	0.00 mg	40.5 mg
Eggplant, raw	25		0	2.00 mg	6.9 mg
Teff, uncooked	367	0.4g	0	12.00 mg	13.1 mg
Sherbet, orange	144	1.2g	1mg	46.00 mg	7.7 mg
Cauliflower, raw	25	0.1g	0	30.00 mg	44.3 mg
Taro leaves, raw	42	0.2g	0	3.00 mg	12.8 mg
Lamb, raw, ground	282	10g	73mg	59.00 mg	69.3 mg
Cheese, camembert	300	15g	72mg	842.00 mg	15.4 mg
Vegetarian fillets	290	2.8g	0	490.00 mg	82.0 mg
PACE, Picante Sauce	25		0	781.00 mg	0
Goji berries, dried	349		0	298.00 mg	0

Datasets

- Cleaned the data
- Created a “library” to label products to a food category
  - E.g. food category “Wheat & Rye”:
    ((?=.*wheat|rye|durum|spelt|einkorn|emmer|khorasan|triticum)(?=.*bread|baguette|loaf|toast|brioche|bun|focaccia|ciabatta))|((?=.*dark|light|marbled)(?=.*rye))|(pumpernickel|crispbread)
- Used the emission values from the GHG emission dataset and added it to each product in the category

04

Limitations

What we couldn’t do

Limitations

Dataset searching

Limited to free datasets

Products->categories

Limitations in the process

Nutritional labels

Limitations in the real life usage of the prediction.

05

Models

What we did

Models

Baseline Algorithm Test

- To select a baseline algorithm for the regression problem, we started by running a collection of selected models with a standard set of data.
- But before starting going into the trial phase, there was a selection of relevant attributes by comparing the correlation values with the target variable (“best” 35 independent variables).
- By using K-fold cross-validation and implementing a pipeline to run all the selected algorithms we were able to identify the one that serves as the starting point.

Models

Linear Regression	Lasso	Elastic Net
Avg MAE 8.14	Avg MAE 9.08	Avg MAE 9.55

K-Nearest Neighbour	Decision Tree	Gradient Boosting
Avg MAE 5.40	Avg MAE 5.21	Avg MAE 5.15

Models

Baseline Algorithm Test

As we can see in the previous tables, Gradient Boosting got the best result, but we decided to go for more, it wasn't enough!

So, we changed the approach and used all the possible variables in context.

Models

Decision Tree	K-Nearest Neighbour	Bootstrap Aggregating (Bagging)
MAE 3.66	MAE 7.28	MAE 4.77

Random Forest	Adaptive Boosting (AdaBoost)	Gradient Boosting
MAE 4.49	MAE 7.26	MAE 4.83

Models

Best model

Despite getting a good result with the decision tree (**MAE: 3.66**), that wasn't the best trained model.

Trained under K-fold cross-validation and built with an 'Adam' optimizer and 'MAE' as loss function, the neural network was able to perform a prediction over the test set, with a value of **3.34** (MAE)

And by comparing with the MAE value (**12.93)** calculated using the mean value of the emissions for every food product, we conclude that our model has an impressive efficiency.

06

Final Products

What we can do with the outputs

Final Products

Mobile APP

A mobile App with a simple and intuitive interface that allows the consumer to verify the information of the carbon footprint of a specific product by scanning its barcode

Final Products

Software

A software that outputs the predicted carbon footprint in a screen or in a receipt could have impact in customers purchases.

07

Improvements

Where we can improve

Improvements

CATEGORIES

We could have used more and clearer specified food categories

LABELING

The labeling process could be more detailed and take more parameters than name into account

EMISSION GENERALIZATION

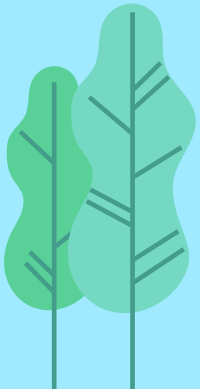
All products in a certain category do not cause the same amount of emission.

FEATURE EXPLORATION

We could have spent more time on understanding the nutritional facts.

FUTURE WORK

Expand the analysis and include the emission values for all the different stages



08

Conclusions

What we learnt

Conclusions

- The development of this project was beneficial for us due to the increased awareness of the need to buy sustainably

- We successfully developed models that accurately predict the greenhouse gas emissions for various food products using its nutritional composition, and the results obtained are better than what we were expecting.

- We realized the difficulty of manipulating data aiming to have an impact and to design a solid final product

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Thanks!

Do you have any questions?