
Dataimporta

Project Proposal

270957 - Machine Learning

Students

Sergio Postigo

Victor Diví

Professor

Bernat Coma



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**

April, 2022

1 PROJECT PROPOSAL

1.1 COMPOSITION OF THE TEAM

Victor Diví and Sergio Postigo

1.2 PROBLEM TO WORK ON

Obtaining categories for imported/exported goods based on the information registered by the customs office of the respective country (Peru).

1.3 REASON/MOTIVATION

Many countries base a big portion of their economies in foreign trade. Therefore, the customs agencies around the world collect data about every imported/exported good that passed across their ports, airports, borders, etc. In some countries, this data is open, and anyone can access to it to analyze it and make more informed decisions while importing or exporting goods. However, this data demands some challenges before its use. One of them is the labelling. For example, in Peru every time a good is imported/exported, a customs agent fills a form with the information of the product(s), where they include descriptions about it. Nonetheless, there isn't a proper labelling like for example, in the case that we wanted to aggregate amounts imported/exported by category.

1.4 REFERENCES OF PREVIOUS WORK

The labelling for goods is done manually mainly by consultancy agencies, who get this data to generate analytics reports for companies and institutions interested in foreign trade information of specific products. They usually use MS Excel spreadsheets to perform the labeling, which is not efficient and takes much time.

1.5 TITLE OF THE PROJECT

Dataimporta

1.6 INFORMATION ABOUT THE DATA

The data includes 41 columns, both numerical and categorical. The number of rows is 631394.