

In light of the global environmental crisis, people are encouraged to take steps to reduce their consumption of certain resources, such as electricity or gasoline. It is hoped that this, combined with an increase in the price of such resources, will actually bring about a change in consumers' habits, such as using public transportation instead of a private car or reducing the use of electricity in their homes.

The goal of this project is to develop a machine learning algorithm that can read the bills from most electrical companies in Spain. This will eventually help assess how electricity consumption has changed in a determined period of time in an accurate and fast way, especially when compared to transcribing the information manually.

Thus, we need someone to develop such an algorithm, which must read the electricity bills from several companies in Spain, identify the items contained in them and classify them properly. For instance, read and identify the kilowatts spent during the last billing period (and the previous ones, if available), and classify them as such as an outcome. For this, it should perform Optical Character Recognition (OCR).

We expect the algorithm to deal with the clients' private information contained in the bills, such as their name, address, and bank information. The algorithm should identify such information and propose a mechanism to discard it, in order to avoid ethic-related conflicts when subsequently analyzing the data. In that sense, the algorithm should be able to erase it from the image, without saving confidential information read through OCR or any other method.

An important consideration is that companies do not display the information equally among them, so the algorithm must be able to correctly classify the information by taking into account such differences. Bills' samples from different companies are found online and will be provided so they can be processed.

This thesis will be developed as part of the project "New opportunities to enhance or extend (mobile) web survey data and get better insights", conducted by Dr. Melanie Revilla from the Research and Expertise Centre for Survey Methodology (RECSM) of Pompeu Fabra University.