

Project Proposal: LiDAR data to refine open-source geospatial datasets

Matteo Biglioli - 938199

October 29, 2021

In the last decade, thanks to the cloud and to the progress made both in memory/storage and in computational power, even small businesses have finally the tools to work with complex geospatial information. One of the main topic of this field is building-related information that can be leveraged for multiple reasons, most of them related to the energy sector, like balancing energy grids.

There are already different actors providing this kind of data: [Google](#) and [OpenStreetMap](#) are the most known, but lately different open source projects are growing popular.

The problem of these datasets is that they are complete and detailed in the most populated and important cities, but they lack a lot of information when we look for smaller towns.

The goal of this project is to leverage [LiDAR data provided by the Ministry of Environment](#)¹ to provide an open-source geospatial dataset with information regarding the geometry of buildings and their primary use (residential, industrial, ...).

In order to estimate the primary use of a building we will develop a machine learning algorithm² that will take as input different features of the building itself³.

For the scope of this project we will consider both as train and test sets different neighborhoods of Milan due to the fact that is one of the most detailed datasets provided free of charge.

For what it concerns the dimensions of the data, the dataset provided by the Ministry contains xyz coordinates of points with a density greater than 1.5 points per square meter covering the whole nation. It is difficult to provide a certain estimate of the cardinality of the dataset that will be used in this project due to the fact that:

- We will consider only the points covering buildings.
- We still need to define which algorithm to use and, based on the computational resources required, we will be able to process neighborhoods of different cardinalities.

¹In the second page we provide a brief visual representation of the dataset.

²We can work with different ideas, from decision trees to CNN that analyze the roof conformation as an image.

³Mainly area, height and roof conformation but we can add others like distance from city center, distance from the nearest building, ...

Here we show a few images to provide a graphic explanation of the dataset; the points are colored based on their altitude:

