CAPSTONE PROJECT

FLORENCE:FIND BEST COUSINS TYPE AND AREA TO OPEN A RESTAURANT

Capstone project
FEDERICO PANNACCI

REPORT

In this project I'll try to find answers to a real case study that was presented by a client that is an entrepreneur that would like to open a restaurant in Florence. He's giving two options for his business: he would like to open a "Tuscan" restaurant in CAP area 50127 or a "Pizza" restaurant in CAP area 50145. He's asking which of this two restaurants would be much more appreciated and therefore which one will result as better business.

I started the project extracting data from Foursquare but I noticed that the number of restaurants returned was very small. For this reason I decided to build a python scraper in order to retrieve data of all restaurants in Florence area.

I used this data to retrieve latitude and longitude of all restaurant and then I proceed with the analysis.

Data source are essential Foursquare and TripAdvidor.

In order to reply to the customer question I proceeded dividing Florence in different areas represented by CAP codes and I searched for all nearby venues with FourSquare. Unfortunately I realised that few restaurant were returned by my query, therefore I decided to build a scraper to retrieve data from TripAdvisor as there isn't a free API for this.

Once I retrieved data I build an index that further I used as dependent variable in machine learning algorithm. The index was calculated multiplying number of review and rank number. The index was then normalised and finally I build different classes (from 1 to 10) for the index.
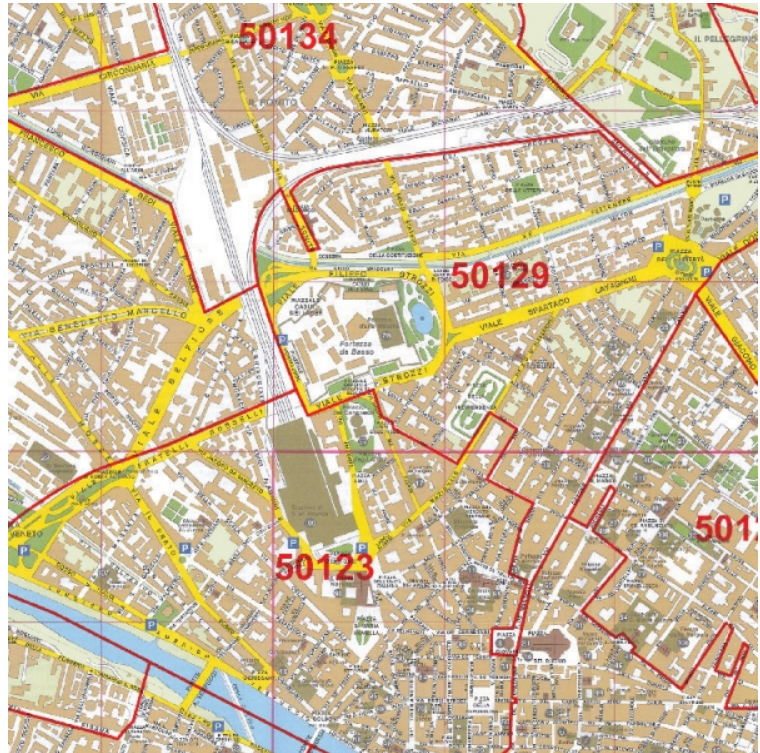
I constructed a database doing a "hot encoding" for all cousins type.

Then I created a Support Vector Machine Model in which and used as independent variables cousin types and CAP code, and quality class index as dependent variable.

I trained and tested the model and finally I evaluated class quality index for the two kind of restaurants that the client told: a "Tuscan" restaurant in CAP area 50127 or a "Pizza" restaurant in CAP area 50145.

Finally I choose the higher class value in order to suggest the kind of activity to open.