

Machine Learning Engineer Nanodegree

Capstone Project

Benjamin Rouillé d'Orfeuil

May 16, 2018

I. Definition

Project Overview

[Yelp](#) is a social networking site that publishes crowd-sourced reviews about local businesses. About two years ago, Yelp challenged Machine Learning practitioners to build a model that automatically tags restaurants with multiple labels using a dataset of user-submitted photographs. The goal of this project is to develop such a model.

The competition was hosted by [Kaggle](#), a platform where data scientist use their skills to produce the best models for predicting and describing the datasets uploaded by companies and users. The various datasets and inputs are accessible via the Yelp Restaurant Photo Classification competition webpage¹. Yelp provides a training dataset (234,842 photographs) and a test dataset (237,152 photographs) for this competition. Each photograph belongs to a business and the task is to predict the business attributes purely from the business photographs. There are 9 different attributes in this problem:

- good for lunch;
- good for dinner;
- takes reservations;
- outdoor seating;
- restaurant is expensive;
- has alcohol;
- has table service;
- ambience is classy;
- good for kids;

and a total of 2,000 and 10,000 businesses in the training and testing datasets, respectively. Note that the datasets are quite large. Both the training and test archive files are about 7 GB.

¹See <https://www.kaggle.com/c/yelp-restaurant-photo-classification>

Problem Statement

II. Analysis

III. Methodology

IV. Results

V. Conclusion