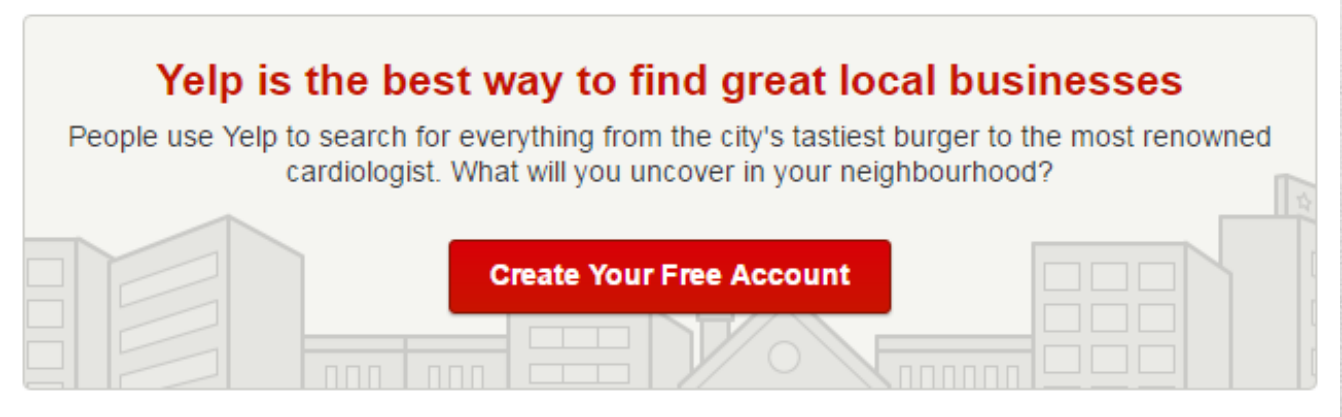
Yelp Restaurant Photo Classification

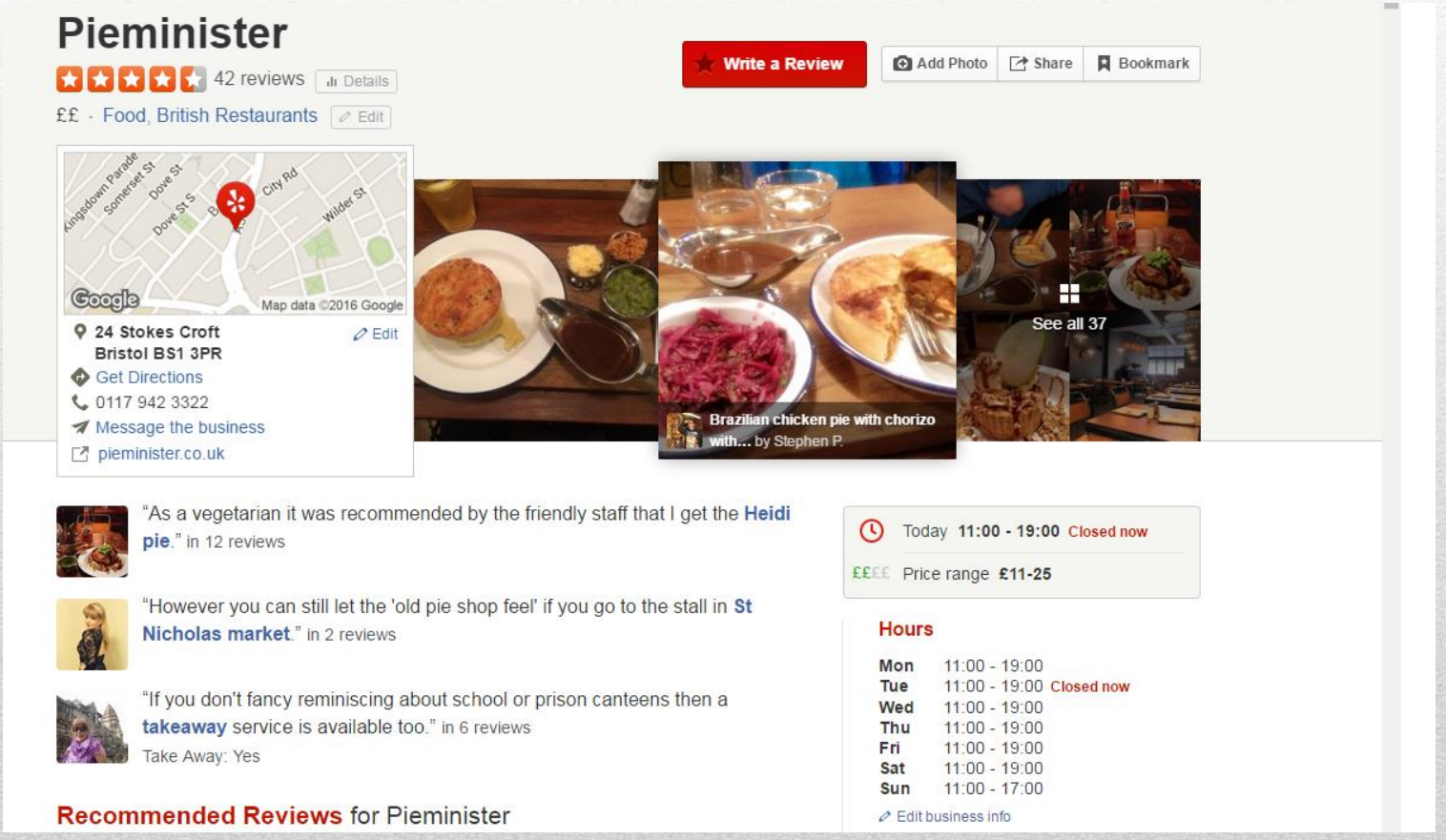
Anna Rumyantseva
Springboard Data Science Intensive Workshop
Spring 2016
Mentor: Alex Chao

Client

- **Yelp** is an American multinational corporation headquartered in San Francisco, California.
- **Yelp.com** and the **Yelp mobile app**, which publish crowd-sourced reviews about local businesses



Review example



Task overview

- Currently, restaurant labels are manually selected by Yelp users
- But Yelp's users upload an enormous amount of photos every day alongside their written reviews
- Selecting the labels is optional, leaving some restaurants un- or only partially-categorized.

<u>AIM</u>: build a model that automatically tags restaurants with multiple labels using user-submitted photos

Data set

- photos of the training set (235841 images; 6.64 GB)
- photos of the test set (474304 images; 6.71 GB)
- maps the photo id to business id
- maps the photo id to business id
- maps the business ids to their corresponding labels.

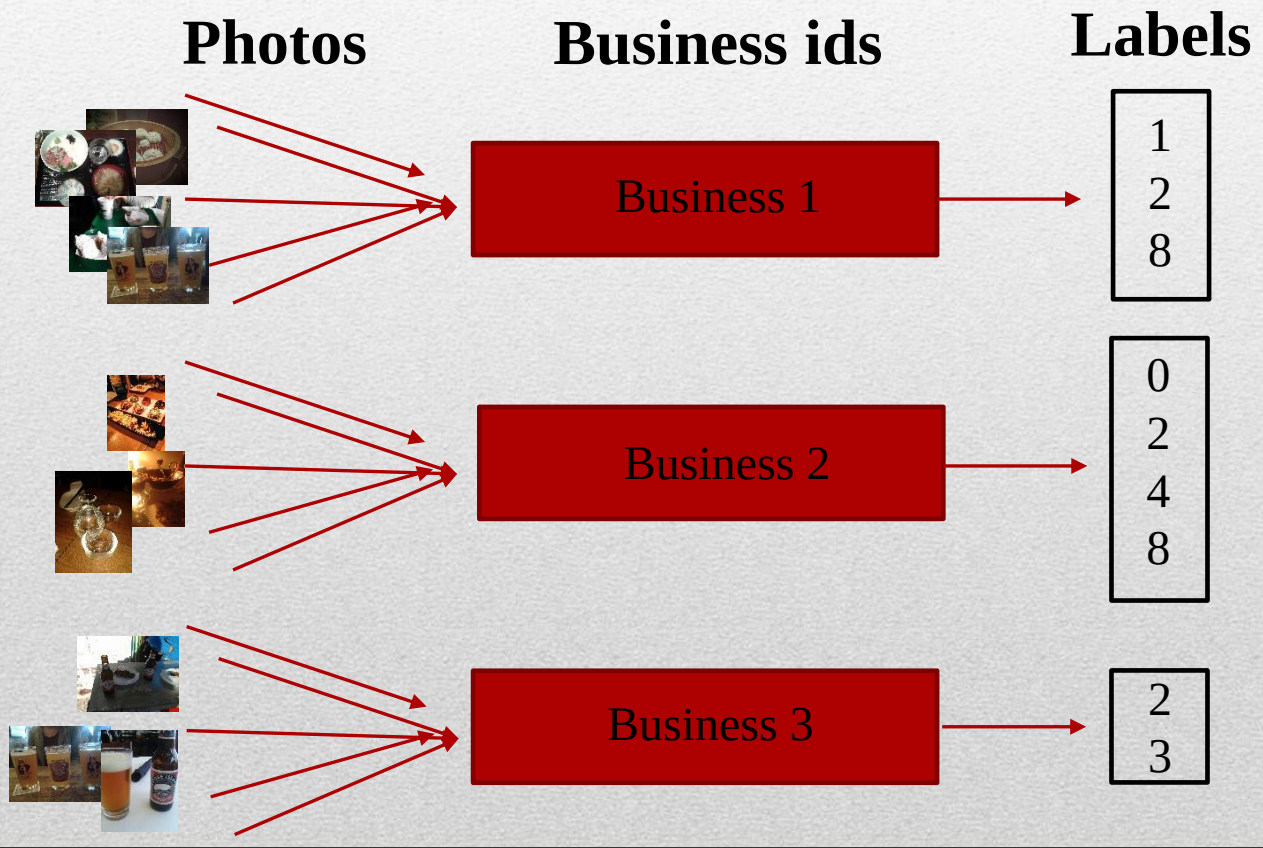
Restaurant labels

Label	Description
0	good_for_lunch
1	good_for_dinner
2	takes_reservations
3	outdoor_seating
4	restaurant_is_expensive
5	has_alcohol
6	has_table_service
7	ambience_is_classy
8	good_for_kids

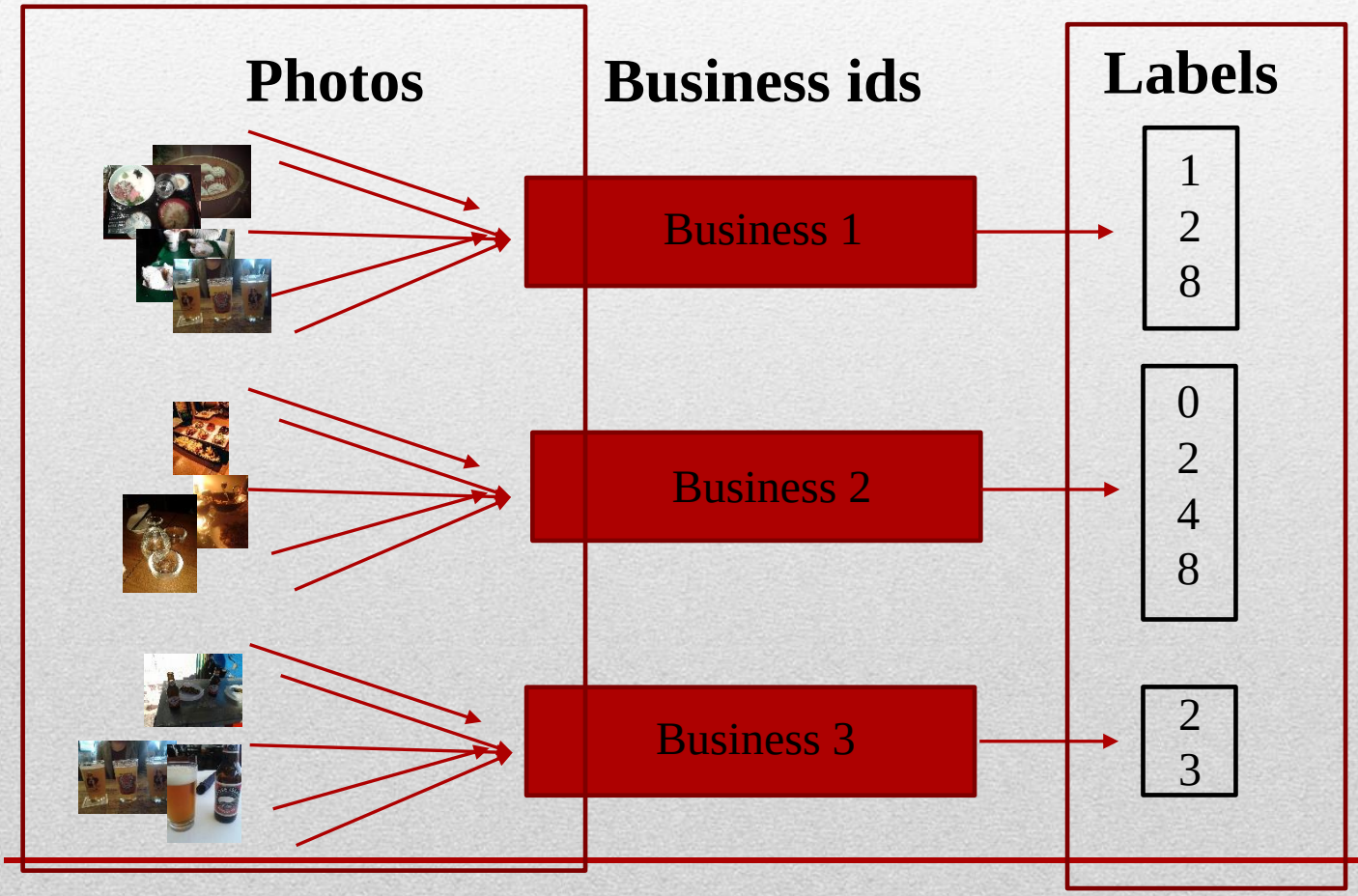
An example

- good for dinner,
- takes reservations,
- has alcohol,
- has table service,
- ambience is classy

What kind of problem is it in Machine Learning?



What kind of problem is it in Machine Learning?



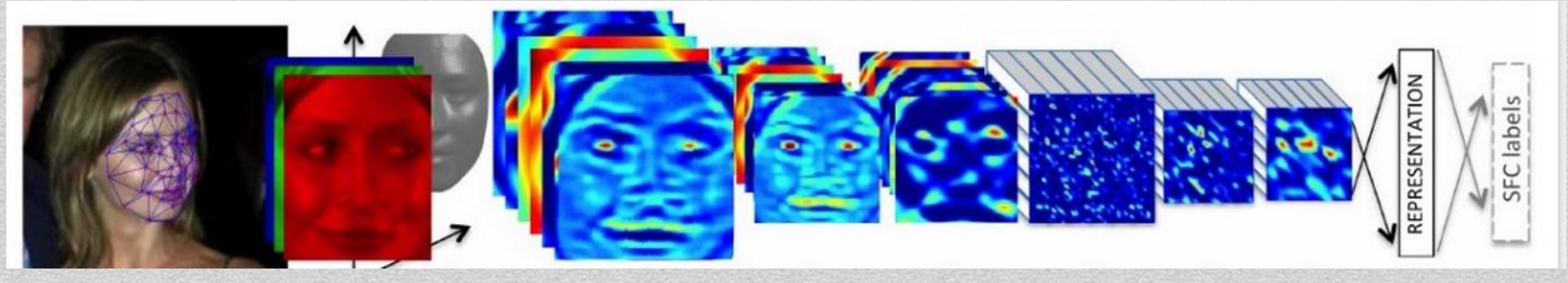
Approach

1. Implement transfer learning approach on Convolutional Neural Networks for image feature extraction
2. Create mean feature vector for each business id (dealing with **the multi-instance aspect** of the problem)
3. Train One vs Rest classifier on top of several supervised learning models (dealing with **the multi-class aspect** of the problem)
4. Try ensemble models for improvement of prediction

Convolutional Neural Networks (CNN)

- A very powerful deep learning technique in computer vision (fast and works well)
- CNN takes image as an input
- "Neurons" are organized in three dimensions
- **Pre-trained CNNs can be used as an image feature extractor**

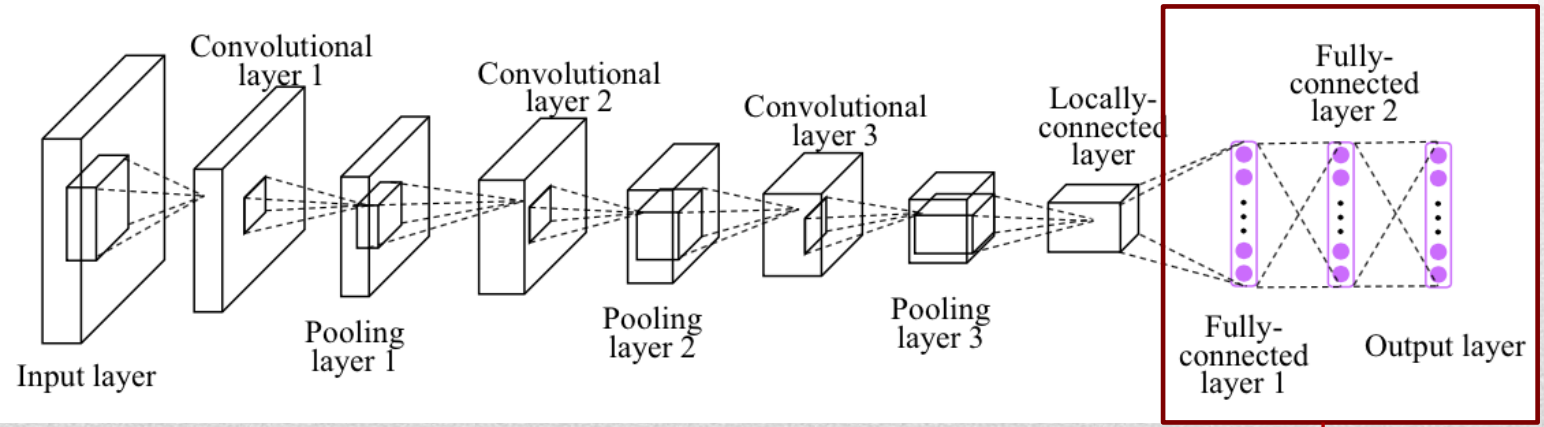
An image going through a CNN…



From http://karpathy.github.io/2015/10/25/selfie/

Convolutional Neural Networks (CNN)

CNN architecture example



from Deep Representation Learning with Target Coding
S. Yang, P. Luo, C. C. Loy, K. W. Shum, X. Tang
in Proceedings of AAAI Conference on Artificial Intelligence, 2015

Outcome from this layers can be used as an image feature vector (CNN code)!

Pre-trained CNN used in the project

Visual Geometry Group (VGG) CNN model developed by Karen Simonyan and Andrew Zisserman (http://www.robots.ox.ac.uk/~vgg/research/very_deep/)

- Trained on the ImageNet data set (1.2 million images with 1000 categories)
- Runner-up for the Large Scale Visual Recognition Challenge 2014
- VGG ConvNet features perform well in multiple transfer learning tasks

If you run Yelp images through VGG CNN model and get scores…

All of them are food related ☺

Extracting features from CNN

Photos corresponding to a business id

1

2

3

4

VGG CNN

CNN code 1

CNN code 2

CNN code 3

CNN code 4

Mean CNN code for a business id

Multi-class classification

- Logistic Regression
- Support Vector Classifier
- Decision Tree
- K Nearest Neighbour
- Gaussian Naïve Bayes

plus

One vs Rest Classifier

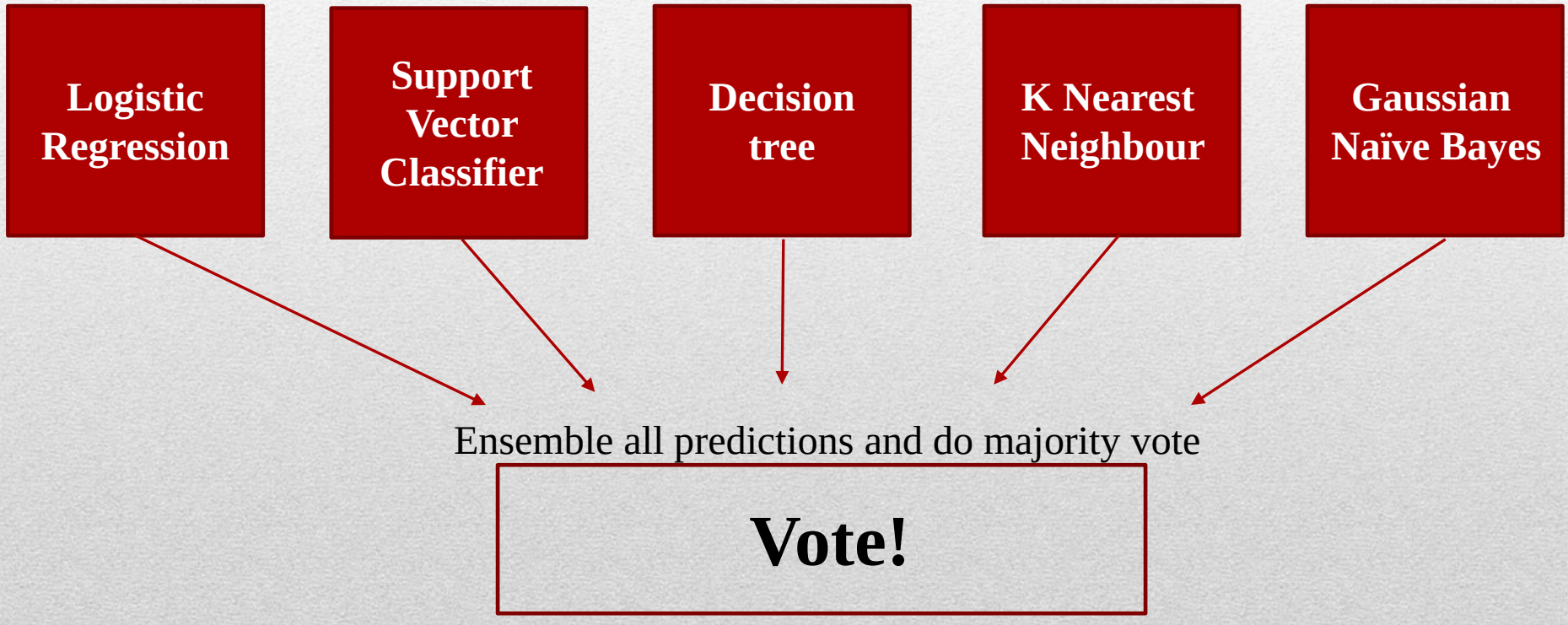
PCA decomposition on 100 components was applied (better performance and faster model training)

Multi-class classification: results on cross validation

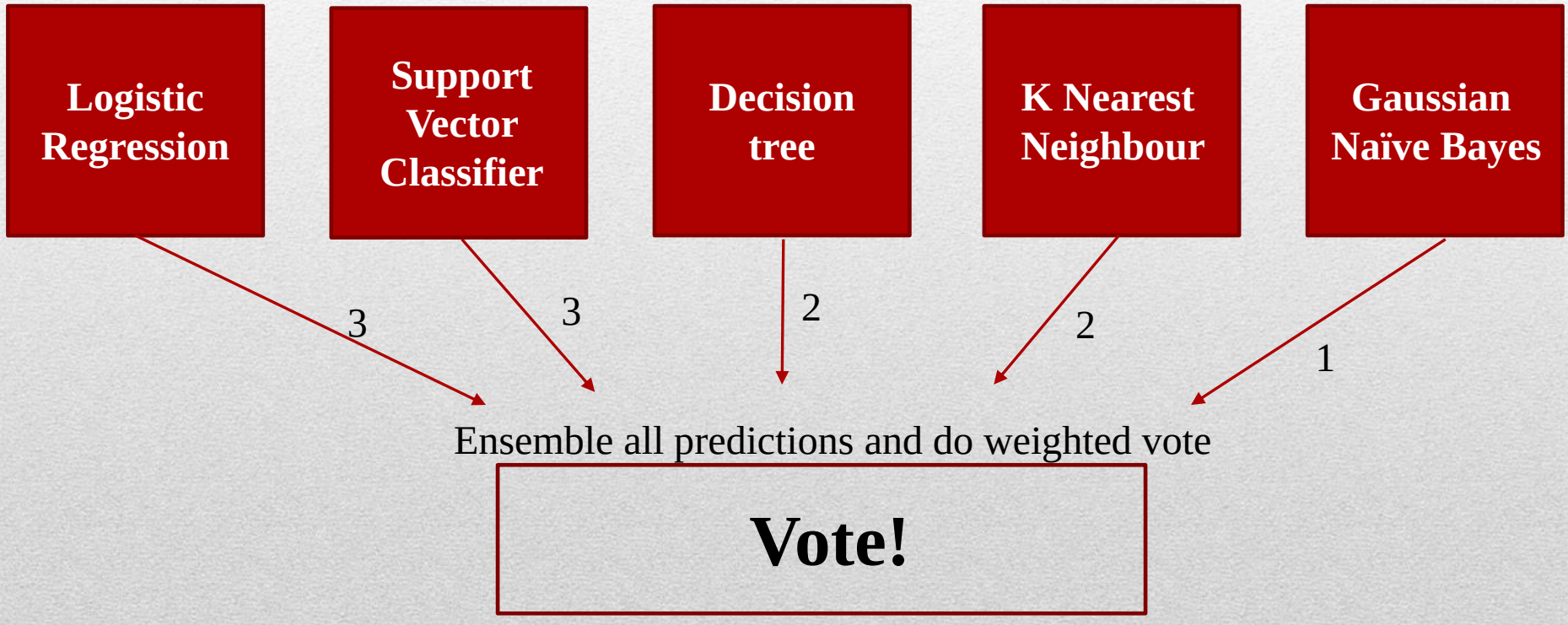
Classifier	Mean f-score
Logistic Regression	0.82818
Support Vector Classifier	**0.83254**
Decision Tree	0.77096
K Nearest Neighbour	0.80608
Gaussian Naïve Bayes	0.694705

K-fold cross validation using 5 folds

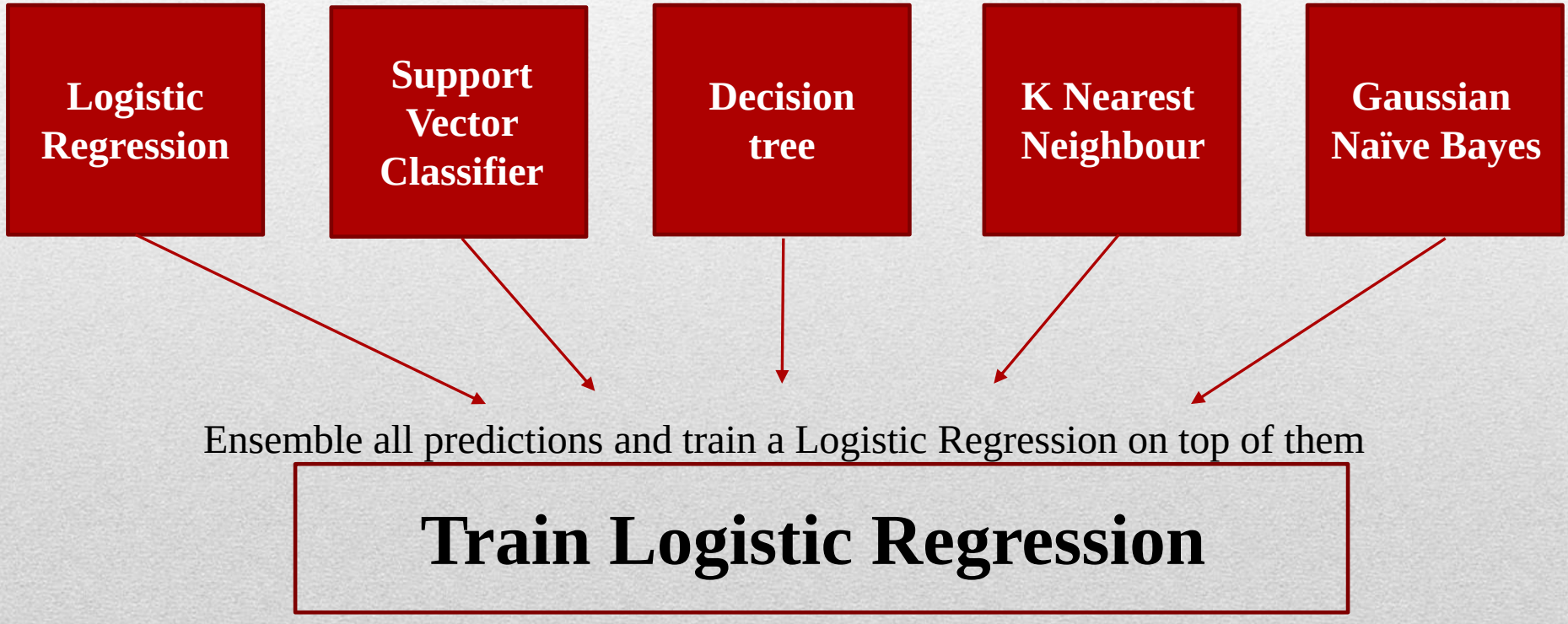
Ensemble learning: Approach 1



Ensemble learning: Approach 2



Ensemble learning: Approach 3



Results on the Kaggle leaderboard

Model	f-score	Place on the leaderboard
Logistic Regression Classifier	0.80184	95
Support Vector Classifier (rbf kernel)	**0.80957**	**76**
Ensemble Learning based on majority vote	0.74898	146
Ensemble Learning based on weighted vote	0.62099	280
Ensemble learning stacked with Logistic Regression	0.80538	86

Support Vector Classifier performed better then the ensemble models

Trying the prediction model on my photos

Beers in a bar in New Orleans: **good for lunch**, **has alcohol**

Trying the prediction model on my photos

Having coffee in Cafe du Monde in New Orleans: **good for lunch, good for kids**

Trying the prediction model on my photos

Fancy dinner with a friend:
good for dinner,
takes reservations,
restaurant is expensive,
has alcohol,
has table service

Trying the prediction model on my photos

Beer in a street café in Spain:
has alcohol, has table services

Final remarks

- Classification of images using the outcome just from one CNN model results in relatively good f-score (~0.81)
- Simple model (SVC classifier) outperformed the ensemble models.
- 76th place on the Kaggle leaderboard
- Further improvements in the model performance can be achieved by advanced feature engineering
- The model gives sensible labels for my photos