

MIDS 207: Introduction to Machine Learning | Section 5

Project 4 Specification

Project Team:

- A project team comprises 2 or 3 students.
- You can form a team yourself or ask your professor to assign you to a team.

Project Topic:

Choose a project topic from among these ...

- **Random Acts of Pizza**
<https://www.kaggle.com/c/random-acts-of-pizza>
Build a binary classification model to predict result of a pizza request post.
Dataset: 5671 observations
Example models provided.
- **Home Prices**
<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>
Build a regression model to predict sale price of a house.
Dataset: 2919 observations, 80 features
Example models provided.
- **Forest Cover**
<https://www.kaggle.com/c/forest-cover-type-prediction>
Build a multinomial classification model to predict a tree type.
Dataset: 581012 observations, 55 features
Example models provided.
- **Face Images**
<https://www.kaggle.com/c/facial-keypoints-detection>
Build a regression model to predict location of a keypoint on a face.
Dataset: 8832 observations, 15x2 keypoint features + 96x96 pixel features
Example models provided.
- **US Political Campaign Fundraising**
Propose a dataset to the professor.
- **US Political Election Results**
Propose a dataset to the professor.
- **Other Topic of Your Own Choosing**
Propose a topic and dataset to the professor.

Project Deliverables:

- One Jupyter notebook with text and python code
- One deck of presentation slides

Guidance:

A good notebook will include ...

- Discuss modeling approach.
- Prepare and explore dataset with some descriptive statistics and data visualizations.
- Try one or more feature selection methods.
- Try several data transformation methods.
- Try several model construction methods and several hyperparameter settings.
- Test models using train/dev/test holdout and/or cross-validation with several performance metrics.
- Report on any insights regarding the data, application, and/or approach with descriptive statistics, data visualizations, and commentary.

A good presentation will ...

- Use about 5 slides.
- Take 5-10 minutes.

Project Submission:

- Submit your deliverables at ISVC Gradebook W207.5.
- Each team member should submit his/her own copy of the deliverables.
- You are not required to submit an entry to a Kaggle competition.
- This project is graded by the professor.
- Notify the professor in advance in the unlikely event that your project will be unavoidably late.