**Project Highlights**

This project is designed to get you acquainted with the many supervised learning algorithms available in sklearn, and to also provide for a method of evaluating just how each model works and performs on a certain type of data. It is important in machine learning to understand exactly when and where a certain algorithm should be used, and when one should be avoided.

Things you will learn by completing this project:

* How to identify when preprocessing is needed, and how to apply it.
* How to establish a benchmark for a solution to the problem.
* What each of several supervised learning algorithms accomplishes given a specific dataset.
* How to investigate whether a candidate solution model is adequate for the problem.

**Where are you in the course?**

* Machine Learning Brid's Eye View
* Linear Regression
* Perceptron Algorithm
* Decision Trees
* Naive Bayes
* Support Vector Machines
* Ensemble Methods
* Model Evaluation Metrics
* Training and Tuning
* **Project**

**Description**

CharityML is a fictitious charity organization located in the heart of Silicon Valley that was established to provide financial support for people eager to learn machine learning. After nearly 32,000 letters were sent to people in the community, CharityML determined that every donation they received came from someone that was making more than $50,000 annually. To expand their potential donor base, CharityML has decided to send letters to residents of California, but to only those most likely to donate to the charity. With nearly *15 million* working Californians, CharityML has brought you on board to help build an algorithm to best identify potential donors and reduce overhead cost of sending mail. Your goal will be evaluate and optimize several different supervised learners to determine which algorithm will provide the highest donation yield while also reducing the total number of letters being sent.

**Software and Libraries**

This project uses the following software and Python libraries:

* [**Python**](https://www.python.org/downloads/release/python-364/)
* [**NumPy**](http://www.numpy.org/)
* [**pandas**](http://pandas.pydata.org/)
* [**scikit-learn**](http://scikit-learn.org/0.17/install.html) (v0.17)
* [**Matplotlib**](http://matplotlib.org/)

You will also need to have software installed to run and execute a [**Jupyter Notebook**](http://ipython.org/notebook.html).

If you do not have Python installed yet, it is highly recommended that you install the [**Anaconda**](https://www.anaconda.com/products/individual)distribution of Python, which already has the above packages and more included.

**Starting the Project**

For this assignment, you can find the p1\_charityml folder containing the necessary project files on the [**Supervised Learning GitHub**](https://github.com/udacity/cd0025-supervised-learning), under the starter folder. This project contains three files that you will need to use:

* finding\_donors.ipynb: This is the main file where you will be performing your work on the project.
* census.csv: The project dataset. You'll load this data in the notebook.
* visuals.py: This Python script provides supplementary visualizations for the project. Do not modify.

In the Terminal or Command Prompt, navigate to the folder on your machine where you've put the project files, and then use the command jupyter notebook finding\_donors.ipynb to open up a browser window or tab to work with your notebook. Alternatively, you can use the command jupyter notebook or ipython notebook and navigate to the notebook file in the browser window that opens. Follow the instructions in the notebook and answer each question presented to successfully complete the project. A **README** file has also been provided with the project files which may contain additional necessary information or instruction for the project.

## Submitting the Project

### Evaluation

Your project will be reviewed by a Udacity reviewer against the [**Finding Donors for CharityML project rubric**](https://review.udacity.com/#!/rubrics/3336/view). Be sure to review this rubric thoroughly and self-evaluate your project before submission. All criteria found in the rubric must be meeting specifications for you to pass.

### Submission Files

Following files would be needed for evaluation:

* The finding\_donors.ipynb notebook file with all questions answered and all code cells executed and displaying output.
* An **HTML** export of the project notebook with the name **report.html**. This file must be present for your project to be evaluated.

When you are ready to submit your project, There are three ways in which your project can be submitted for evaluation.

1. If you run the notebook from your **local machine** collect the above files and compress them into a single archive for upload.
2. You could supply the above files on your **GitHub Repo** in a folder named finding\_donors for ease of access. This would build a good Github profile in parallel.
3. If you worked using the **workspace inside the classroom** you can submit your project directly for review using the submit button at the end of the project, just make sure you download the HTML report to your local machine and upload it back into the workspace BEFORE submitting your report.

As an added part of this project, we have partnered with Kaggle to get you started integrating into their amazing community which can lead to jobs, prize money, and most importantly - other data science friends!

Here, you can take your model one step farther by seeing how it performs on the test data available on the competition website. Notice, there are some additional difficulties in working with this new data, as there are a number of missing values.

Your current project gives you a great start, but you will have to perform a few additional cleaning steps to complete the competition! Once you have the data cleaned, see how far you can climb on the leaderboard!

[**ACCESS KAGGLE COMPETITION HERE**](https://www.kaggle.com/c/udacity-mlcharity-competition)

## Project Submission

CharityML is a fictitious charity organization located in the heart of Silicon Valley that was established to provide financial support for people eager to learn machine learning. After nearly 32,000 letters were sent to people in the community, CharityML determined that every donation they received came from someone that was making more than $50,000 annually. To expand their potential donor base, CharityML has decided to send letters to residents of California, but to only those most likely to donate to the charity. With nearly 15 million working Californians, CharityML has brought you on board to help build an algorithm to best identify potential donors and reduce overhead cost of sending mail. Your goal will be evaluate and optimize several different supervised learners to determine which algorithm will provide the highest donation yield while also reducing the total number of letters being sent.

### Evaluation

Your project will be reviewed by a Udacity reviewer against the [**Finding Donors for CharityML project rubric**](https://review.udacity.com/#!/rubrics/3336/view). Be sure to review this rubric thoroughly and self-evaluate your project before submission. All criteria found in the rubric must be meeting specifications for you to pass.

### Submission Files

When you are ready to submit your project, collect the following files and compress them into a single archive for upload. Alternatively, you may supply the following files on your GitHub Repo in a folder named finding\_donors for ease of access:

* The finding\_donors.ipynb notebook file with all questions answered and all code cells executed and displaying output.
* An **HTML** export of the project notebook with the name **report.html**. This file must be present for your project to be evaluated.