**Capstone Project Overview**

In this capstone project, you will leverage what you’ve learned throughout the Nanodegree program to solve a problem of your choice by applying machine learning algorithms and techniques. You will first **define** the problem you want to solve and investigate potential solutions and performance metrics. Next, you will **analyze** the problem through visualizations and data exploration to have a better understanding of what algorithms and features are appropriate for solving it. You will then **implement** your algorithms and metrics of choice, documenting the preprocessing, refinement, and postprocessing steps along the way. Afterwards, you will collect **results** about the performance of the models used, visualize significant quantities, and validate/justify these values. Finally, you will construct **conclusions** about your results, and discuss whether your implementation adequately solves the problem.

**Capstone Project Highlights**

This project is designed to prepare you for delivering a polished, end-to-end solution report of a real-world problem in a field of interest. When developing new technology, or deriving adaptations of previous technology, properly documenting your process is critical for both validating and replicating your results.

Things you will learn by completing this project:

* How to research and investigate a real-world problem of interest.
* How to accurately apply specific machine learning algorithms and techniques.
* How to properly analyze and visualize your data and results for validity.
* How to document and write a report of your work.

**Capstone Description**

Think about a technical field or domain that you are passionate about, such as robotics, virtual reality, finance, natural language processing, or even artificial intelligence (the possibilities are endless!). Then, choose an existing problem within that domain that you are interested in which you could solve by applying machine learning algorithms and techniques. Be sure that you have collected all of the resources needed (such as data sets) to complete this project, and make the appropriate citations wherever necessary in your report. Below are a few suggested problem areas you could explore if you are unsure what your passion is:

* [Robot Motion Planning](https://docs.google.com/document/d/1ZFCH6jS3A5At7_v5IUM5OpAXJYiutFuSIjTzV_E-vdE/pub)
* [Healthcare](https://docs.google.com/document/d/1WzurKKa9AX2DnOH7KiB38mvozdOSemfkGpex8hdTy8c/pub)
* [Computer Vision](https://docs.google.com/document/d/1y-XfjkPFgUQxFIQ9bBncUSjs4HOf5E-45FrLYNBsZb4/pub)
* [Education](https://docs.google.com/document/d/1vjerjRQnWs1kLbZagDYT6rNqiwAG23Yj45oUY88IAxI/pub)
* [Investment and Trading](https://docs.google.com/document/d/1ycGeb1QYKATG6jvz74SAMqxrlek9Ed4RYrzWNhWS-0Q/pub)

In addition, you may find a technical domain (along with the problem and dataset) as *competitions* on platforms such as [Kaggle](http://kaggle.com), or [Devpost](http://devpost.com). This can be helpful for discovering a particular problem you may be interested in solving as an alternative to the suggested problem areas above. In many cases, some of the requirements for the capstone project are already defined for you when choosing from these platforms.

***Note:*** *For students who have enrolled before* ***October 17th****, we strongly encourage that you look at the* ***Capstone Proposal*** *project that is available as an elective before this project. If you have an idea for your capstone project but aren't ready to begin working on the implementation, or even if you want to get feedback on how you will approach a solution to your problem, you can use the* ***Capstone Proposal*** *project to have a peer-review from one of our* ***Capstone Project*** *reviewers!*

For whichever application area or problem you ultimately investigate, there are five major stages to this capstone project which you will move through and subsequently document. Each stage plays a significant role in the development life cycle of beginning with a problem definition and finishing with a polished, working solution. As you make your way through developing your project, be sure that you are also working on a rough draft of your project report, as it is the most important aspect to your submission!

To determine whether your project and the problem you want to solve fits Udacity's vision of a Machine Learning Capstone Project , please refer to the [capstone project rubric](https://review.udacity.com/#%21/rubrics/108/view) and make a note of each rubric criteria you will be evaluated on. A satisfactory project will have a report that encompasses each stage and component of the rubric.

## Software Requirements

**Your project must be written in Python 2.7.** Given the free-form nature of the machine learning capstone, the software and libraries you will need to successfully complete your work will vary depending on the chosen application area and problem definition. Because of this, it is imperative that all necessary software and libraries used in your capstone project are accessible to the reviewer and clearly documented. Information regarding the software and libraries your project makes use of should be included in the README along with your submission. Please note that proprietary software, software that requires private licenses, or software behind a paywall or login account should be avoided.

## Data Requirements

Every machine learning capstone project will most certainly require some form of dataset or input data structure (input text files, images, etc.). Similar to the software requirements above, the data you use must either be publicly accessible or provided by you during the submission process, and private or proprietary data should not be used without expressed permission. Please take into consideration the file size of your data — while there is no strict upper limit, input files that are excessively large may require reviewers longer than an acceptable amount of time to acquire all of your project files and/or execute the provided development code. This can take away from the reviewer's time that could be put towards evaluating your submission. If the data you are working with fits the criteria of being too large, consider whether you can work with a subset of the data instead, or provide a representative sample of the data which the reviewer may use to verify the solution explored in the project.

## Ethics

Udacity's A/B Testing course, as part of the Data Analyst Nanodegree, has a segment that discusses [the sensitivity of data](https://classroom.udacity.com/nanodegrees/nd002/parts/00213454013/modules/411033896375460/lessons/3998098714/concepts/39997087540923) and the expectation of privacy from those whose information has been collected. While most data you find available to the public will not have any ethical complications, it is extremely important that you are considering where the data you are using came from, and whether that data contains any sensitive information. For example, if you worked for a bank and wanted to use customers' bank statements as part of your project, this would most likely be an unethical choice of data and should be avoided.

## Report Guidelines

Your project submission will be evaluated primarily on the report that is submitted. It is expected that the project report contains enough detail, documentation, analysis, and discussion to adequately reflect the work you completed for your project. Because of this, it is extremely important that the report is written in a professional, standardized way, so those who review your project submission are able to clearly identify each component of your project in the report. Without a properly written report, your project cannot be sufficiently evaluated. A [project report template](https://github.com/udacity/machine-learning/blob/master/projects/capstone/capstone_report_template.md) is provided for you to understand how a project report should be structured. We strongly encourage students to have a report that is approximately **nine to fifteen pages in length**.

The Machine Learning Capstone Project report should be treated no different than a written research paper for academics. Your goal is to ultimately present the research you've discovered into the respective problem domain you've chosen, and then discuss each stage of the project as they are completed. The narrative found in the [project report template](https://github.com/udacity/machine-learning/blob/master/projects/capstone/capstone_report_template.md) provides for a "report checklist" that will aid you in staying on track for both your project and the documentation in your report. Each stage can be found as a section that will guide you through each component of the project development life cycle. Please make use of this resource!

## Example Machine Learning Capstone Reports

Included in the project files for the Capstone are three example reports that were written by students just like yourselves. Because the written report for your project will be how you are evaluated, it is absolutely critical that you are producing a clear, detailed, well-written report that adequately reflects the work that you've completed for your Capstone. Following along with the **Capstone Guidelines** will be very helpful as you begin writing your report.

[Our first example report](https://github.com/udacity/machine-learning/blob/master/projects/capstone/report-example-1.pdf) comes from graduate Martin Bede, whose project design in the field of computer vision, named "Second Sight", was to create an Android application that would extract text from the device's camera and read it aloud. Martin's project cites the growing concern of vision loss as motivation for developing software that can aid those unable to see or read certain print.

[Our second example report](https://github.com/udacity/machine-learning/blob/master/projects/capstone/report-example-2.pdf) comes from an anonymous graduate whose project design in the field of image recognition was to implement a Convolutional Neural Network (CNN) to train on the [Cifar-10](https://www.cs.toronto.edu/%7Ekriz/cifar.html) dataset and successfully identify different objects in new images. This student describes with thorough detail how a CNN can be used quite effectively as a descriptor-learning image recognition algorithm.

[Our third example report](https://github.com/udacity/machine-learning/blob/master/projects/capstone/report-example-3.pdf) comes from graduate Naoki Shibuya, who took advantage of the pre-curated robot motion planning ["Plot and Navigate a Virtual Maze" project](https://docs.google.com/document/d/1ZFCH6jS3A5At7_v5IUM5OpAXJYiutFuSIjTzV_E-vdE/pub). Pay special attention to the emphasis Naoki places on discussing the methodology and results: Projects relying on technical implementations require valuable observations and visualizations of how the solution performs under various circumstances and constraints.

Each example report given has many desirable qualities we expect from students when completing the Machine Learning Capstone project. Once you begin writing your project report for which ever problem domain you choose, be sure to reference these examples whenever necessary!

## Submitting the Project

### Evaluation

Your project will be reviewed by a Udacity reviewer against the [**Machine Learning Capstone project rubric**](https://review.udacity.com/#%21/rubrics/108/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

At minimum, your submission will be required to have the following files listed below. If your submission method of choice is uploading an archive (\*.zip), please take into consideration the total file size. You will need to include

* Your **capstone proposal** document as **proposal.pdf** if you have completed the pre-requisite Capstone Proposal project. Please also include your review link in the student submission notes.
* A project report (**in PDF format only**) addressing the five major project development stages. The recommended page length for a project report is approximately nine to fifteen pages. **Please do not export an iPython Notebook as PDF for your project report**.
* All development Python code used for your project that is required to reproduce your implemented solution and result. Your code should be in a neat and well-documented format. Using iPython Notebooks is strongly encouraged for development.
* A README documentation file which briefly describes the software and libraries used in your project, including any necessary references to supporting material. If your project requires setup/startup, ensure that your README includes the necessary instructions.
* Any additional supporting material such as datasets, images, or input files that are necessary for your project's development and implementation. If these files are too large and you are uploading your submission, **instead provide appropriate means of acquiring the necessary files in your included** README.

Once you have collected these files and reviewed the project rubric, proceed to the project submission page.