**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 Project Description

Think about a problem domain that you are passionate about, such as healthcare, engineering, finance, robotics, marketing, bio-informatics, or something that excites you. Then, choose an existing problem within that domain 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.

For whichever application area or problem you ultimately investigate, there are five major phases to this capstone project which you will document. Each phase 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, keep track of these phases with the following questions provided, to ensure that each have been thoroughly documented and satisfy the requirements of the [**capstone project rubric**](https://docs.google.com/document/d/1Wxa-kLCKFkUEJBTWLFTvVbuwq2j5mj7TCYnxmFXiEQg/pub?embedded=true).

You can find a [**project report template here**](https://docs.google.com/document/d/1B-vEOscvfqctGEMHTFDS9Nw7aqcE2iuwPRfp0jK8nf4/pub?embedded=true), which can serve as a guideline for writing your report. **Your submitted report should follow this general structure to streamline the evaluation process.**

## Capstone Project Guidelines pt. 1

Use the following questions in each section to keep a mental checklist of if you are meeting the standards and requirements for each rubric item your project is graded against. You can find these questions in their respective sections of the project report template [**here**](https://docs.google.com/document/d/1B-vEOscvfqctGEMHTFDS9Nw7aqcE2iuwPRfp0jK8nf4/pub?embedded=true).

### Definition

**Project Overview**

* Has an overview of the project been provided, such as the problem domain, project origin, and related datasets or input data?
* Has enough background information been given so that an uninformed reader would understand the problem domain and following problem statement?

**Problem Statement**

* Is the problem statement clearly defined? Will the reader understand what you are expecting to solve?
* Have you thoroughly discussed how you will attempt to solve the problem?
* Is an anticipated solution clearly defined? Will the reader understand what results you are looking for?

**Metrics**

* Are the metrics you’ve chosen to measure the performance of your models clearly discussed and defined?
* Have you provided reasonable justification for the metrics chosen based on the problem and solution?

### Analysis

**Data Exploration**

* If a dataset is present for this problem, have you thoroughly discussed certain features about the dataset? Has a data sample been provided to the reader?
* If a dataset is present for this problem, are statistics about the dataset calculated and reported? Have any relevant results from this calculation been discussed?
* If a dataset is **not** present for this problem, has discussion been made about the input space or input data for your problem?
* Are there any abnormalities or characteristics about the input space or dataset that need to be addressed? (categorical variables, missing values, outliers, etc.)

**Exploratory Visualization**

* Have you visualized a relevant characteristic or feature about the dataset or input data?
* Is the visualization thoroughly analyzed and discussed?
* If a plot is provided, are the axes, title, and datum clearly defined?

**Algorithms and Techniques**

* Are the algorithms you will use, including any default variables/parameters in the project clearly defined?
* Are the techniques to be used thoroughly discussed and justified?
* Is it made clear how the input data or datasets will be handled by the algorithms and techniques chosen?

**Benchmark**

* Has some result or value been provided that acts as a benchmark for measuring performance?
* Is it clear how this result or value was obtained (whether by data or by hypothesis)?

### Methodology

**Data Preprocessing**

* If the algorithms chosen require preprocessing steps like feature selection or feature transformations, have they been properly documented?
* Based on the **Data Exploration** section, if there were abnormalities or characteristics that needed to be addressed, have they been properly corrected?
* If no preprocessing is needed, has it been made clear why?

**Implementation**

* Is it made clear how the algorithms and techniques were implemented with the given datasets or input data?
* Were there any complications with the original metrics or techniques that required changing prior to acquiring a solution?
* Was there any part of the coding process (e.g., writing complicated functions) that should be documented?

**Refinement**

* Has an initial solution been found and clearly reported?
* Is the process of improvement clearly documented, such as what techniques were used?
* Are intermediate and final solutions clearly reported as the process is improved?

## Capstone Project Guidelines pt. 2

Use the following questions in each section to keep a mental checklist of if you are meeting the standards and requirements for each rubric item your project is graded against. You can find these questions in their respective sections of the project report template [**here**](https://docs.google.com/document/d/1B-vEOscvfqctGEMHTFDS9Nw7aqcE2iuwPRfp0jK8nf4/pub?embedded=true).

### Results

**Model Evaluation and Validation**

* Is the final model reasonable and aligning with solution expectations? Are the final parameters of the model appropriate?
* Has the final model been tested with various inputs to evaluate whether the model generalizes well to unseen data?
* Is the model robust enough for the problem? Do small perturbations (changes) in training data or the input space greatly affect the results?
* Can results found from the model be trusted?

**Justification**

* Are the final results found stronger than the benchmark result reported earlier?
* Have you thoroughly analyzed and discussed the final solution?
* Is the final solution significant enough to have solved the problem?

### Conclusion

**Free-Form Visualization**

* Have you visualized a relevant or important quality about the problem, dataset, input data, or results?
* Is the visualization thoroughly analyzed and discussed?
* If a plot is provided, are the axes, title, and datum clearly defined?

**Reflection**

* Have you thoroughly summarized the entire process you used for this project?
* Were there any interesting aspects of the project?
* Were there any difficult aspects of the project?
* Does the final model and solution fit your expectations for the problem, and should it be used in a general setting to solve these types of problems?

**Improvement**

* Are there further improvements that could be made on the algorithms or techniques you used in this project?
* Were there algorithms or techniques you researched that you did not know how to implement, but would consider using if you knew how?
* If you used your final solution as the new benchmark, do you think an even better solution exists?

### Quality

**Presentation**

* Does the project report you’ve written follow a well-organized structure similar to that of the project template?
* Is each section (particularly Analysis and Methodology) written in a clear, concise and specific fashion? Are there any ambiguous terms or phrases that need clarification?
* Would the intended audience of your project be able to understand your analysis, methods, and results?
* Have you properly proof-read your project report to assure there are minimal grammatical and spelling mistakes?
* Are all the resources used for this project correctly cited and referenced?

**Functionality**

* Is the code that implements your solution easily readable and properly commented?
* Does the code execute without error and produce results similar to those reported?

**Software and Libraries**

**Your project must be written in Python 2.7.** Given the free-form nature of this capstone project, 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 all necessary software and libraries used in your capstone project are accessible to the reviewer. An included README file which documents the software used, Python libraries (including version), and extraneous installation information is sufficient.

**Deliverables**

The following files should be included in your submission, and can be packaged as a single zip file for convenience:

* A 9 - 15 page report in PDF addressing the five major project development phases.
* All development code necessary for your solution in a clean, commented format.
* Any required supporting files for your code (.csv datasets, input files, etc.).
  + *If these files are too large, please instead reference appropriate download links in the README.*
* A README documenting the software and libraries used in your project, including any necessary instructions required to execute your code.

## Capstone Project Suggestions

If you have not been able to identify an application area or problem you feel passionate about, that’s okay! We have compiled a small list of project suggestions from different application areas which can be used to inspire your own ideas. Each project suggestion has its own series of instructions and guidelines, and a few include a supplementary course for prerequisite material. Before beginning one of the projects below, be sure to thoroughly read the included project description and report format in the respective link.

In addition, depending on your problem domain, you may find challenges and datasets on platforms such as **[Kaggle](http://kaggle.com/" \t "_blank)**, [**Devpost**](http://devpost.com/), etc. helpful for discovering a particular problem you may be interested in solving. In many cases, some of the requirements for this project are already defined for you when choosing from these platforms.

**You must provide the necessary citations if your problem comes from one of these sources.**

### Application Areas

* [**Deep Learning**](https://docs.google.com/document/d/1L11EjK0uObqjaBhHNcVPxeyIripGHSUaoEWGypuuVtk/pub)
* [**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)

### Evaluation

Your project will be reviewed by a Udacity reviewer against [**this rubric**](https://review.udacity.com/#!/rubrics/108/view). Be sure to review it thoroughly before you submit. All criteria must meet specifications in order to pass.

### Submission

When you're ready to submit your project go back to your **[Udacity Home](https://www.udacity.com/me" \t "_blank)**, click on Project 5, and we'll walk you through the rest of the submission process.

For students who have completed the Di-Tech Challenge and who wish to be considered for a position in Didi Chuxing’s Beijing or Silicon Valley Research Lab, please name your zip file***“Didi-Chuxing.zip”***

If you are having any problems submitting your project or wish to check on the status of your submission, please email us at **machine-support@udacity.com** or visit us in the [**discussion forums**](http://discussions.udacity.com/).

### What's Next?

You will get an email as soon as your reviewer has feedback for you. In the meantime, review your next project and feel free to get started on it or the courses supporting it!