Machine Learning Case Studies

## Project Overview

### Build Your Machine Learning Project

In this capstone project, you will leverage what you’ve learned throughout the program to build a machine learning engineer project of your choosing. Your project deliverables are:

1. A Github repository of your work.
2. A blog post written for a technical audience, or a deployed web application powered by data.

In this capstone project, you will leverage what you’ve learned throughout the Nanodegree program to solve a problem of your choice.

1. You will first **define** the problem you want to solve and investigate potential solutions.
2. 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.
3. You will then **implement** your algorithms and metrics of choice, documenting the preprocessing, refinement, and post-processing steps along the way.
4. Afterwards, you will collect **results** about your findings, visualize significant quantities, validate/justify your results, and make any concluding remarks about whether your implementation adequately solves the problem.
5. Finally, you will **construct** a blog post to document all of the steps from start to finish of your project, or deploy your results into a web application.

### Setting Yourself Apart

An important part of landing a job or advancing your career as a machine learning engineer is setting yourself apart through impressive projects. By now, you've completed several guided projects, and now's your chance to show off your skills and creativity. You'll receive a review and feedback from a Udacity mentor, and they will focus on how your project demonstrates your skills.

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 data science 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.

## Software Requirements

**Your project must be written in Python 3.x**. Given the free-form nature of the 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 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 has a segment that discusses [the sensitivity of data](https://classroom.udacity.com/courses/ud257/lessons/3998098714/concepts/39997087540923) (free course link) 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.

## Selecting a Project

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 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 Github (and your blog if that is the path you decide to pursue). 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/e/2PACX-1vT_MjTWMs1-668wgKALIK61AUHbqbKhp4UuUvYK99u0vc_R0K68Cqq5Br0Nf0aoGmg8XVAyfjd1MT8s/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/" \t "_blank). 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.

#### Udacity Specific Projects

* **Customer Segmentation Report for Arvato Financial Services**
* **Optimizing App Offers With Starbucks**
* **Use Convolutional Neural Networks to Identify Dog Breeds**  
  (If you decide to do complete this project, you can find additional content to assist in the extra-curricular portion of this program).

#### Check out Sample Projects

* Here are two projects that can give you an idea of what a final blog post might look like. Each of these meets the requirements for the capstone project: [project 1](https://github.com/udacity/machine-learning/blob/master/projects/capstone/report-example-1.pdf) and [project 2](https://github.com/udacity/machine-learning/blob/master/projects/capstone/report-example-3.pdf).

No matter what project you decide to complete, you will want to make sure to check the [project rubric here](https://review.udacity.com/#!/rubrics/2345/view).

## Create a Customer Segmentation Report for Arvato Financial Solutions

To introduce yourself to the scenario you'll be investigating in this capstone project option, take a look at the following video with Timo Reis from Arvato Financial Solutions.

<https://www.youtube.com/watch?v=qBR6A0IQXEE&feature=emb_logo>

### Steps to Complete This Project

The project has three major steps: the customer segmentation report, the supervised learning model, and the Kaggle Competition.

#### 1. Customer Segmentation Report

This section will be similar to the corresponding project in Term 1 of the program, but the datasets now include more features that you can potentially use. You'll begin the project by using unsupervised learning methods to analyze attributes of established customers and the general population in order to create customer segments.

#### 2. Supervised Learning Model

You'll have access to a third dataset with attributes from targets of a mail order campaign. You'll use the previous analysis to build a machine learning model that predicts whether or not each individual will respond to the campaign.

#### 3. Kaggle Competition

Once you've chosen a model, you'll use it to make predictions on the campaign data as part of a Kaggle Competition. You'll rank the individuals by how likely they are to convert to being a customer, and see how your modeling skills measure up against your fellow students.

## Aravato: Terms & Conditions

In addition to Udacity's Terms of Use and other policies, your downloading and use of the **AZ Direct GmbH** data solely for use in the **Unsupervised Learning** and **Bertelsmann Capstone** projects are governed by the following additional terms and conditions. The big takeaways:

1. You agree to **AZ Direct GmbH's** General Terms provided below and that you only have the right to download and use the **AZ Direct GmbH** data solely to complete the data mining task which is part of the **Unsupervised Learning** and **Bertelsmann Capstone** projects for the Udacity Data Science Nanodegree program.
2. You are prohibited from using the **AZ Direct GmbH** data in any other context.
3. You are also required and hereby represent and warrant that you will delete any and all data you downloaded within 2 weeks after your completion of the **Unsupervised Learning** and **Bertelsmann Capstone** projects and the program.
4. If you do not agree to these additional terms, you will not be allowed to access the data for this project.

The full terms are provided in the workspace below. You will then be asked in the next workspace to agree to these terms before gaining access to the project, which you may also choose to download if you would like to read in full the terms.

These same exact terms are provided in the next workspace, where you will be asked to accept the terms prior to gaining access to the data.

## Starbucks Project Overview

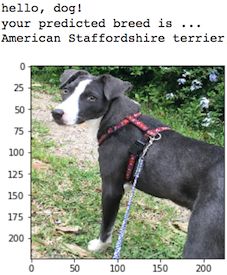
<https://www.youtube.com/watch?time_continue=33&v=bq-H7M5BU3U&feature=emb_logo>

## CNN Project: Dog Breed Classifier

### Project Overview

Welcome to the Convolutional Neural Networks (CNN) project! In this project, you will learn how to build a pipeline to process real-world, user-supplied images. Given an image of a dog, your algorithm will identify an estimate of the canine’s breed. If supplied an image of a human, the code will identify the resembling dog breed.

Along with exploring state-of-the-art CNN models for classification, you will make important design decisions about the user experience for your app. Our goal is that by completing this lab, you understand the challenges involved in piecing together a series of models designed to perform various tasks in a data processing pipeline. Each model has its strengths and weaknesses, and engineering a real-world application often involves solving many problems without a perfect answer. Your imperfect solution will nonetheless create a fun user experience!



Example dog breed output.

### Project Instructions

Clone the project from the GitHub [**repository**](https://github.com/udacity/deep-learning-v2-pytorch/tree/master/project-dog-classification). Follow the instructions in the README to complete the project.

### Evaluation

Your project will be reviewed by a Udacity reviewer against the CNN project [**rubric**](https://review.udacity.com/#!/rubrics/2259/view). Review this rubric thoroughly, and self-evaluate your project before submission. All criteria found in the rubric must meet specifications for you to pass.

### Project Submission

When you are ready to submit your project, collect the following files and compress them into a single archive for upload:

* The dog\_app.ipynb file with fully functional code, all code cells executed and displaying output, and all questions answered.
* An HTML or PDF export of the project notebook with the name report.html or report.pdf.
* Any additional images used for the project that were not supplied to you for the project. Please do not include the project data sets in the dogImages/ or lfw/ folders.

Alternatively, your submission could consist of the GitHub link to your repository.

### Ready to submit your project?

Click on the "Submit Project" button and follow the instructions to submit!