# **AAI-540 ML Design Document Template**

## **Team Info**

Project Team Group #:

Authors:

Business Name:

Publication Date:

## **Team Workflows**

GitHub Project Link:

Asana Board Link:

Team Tracker Link:

## **Project Scope**

**Project Background:**

1-2 paragraphs describing, at a high level, the problem you are trying to solve. This should provide the reader context for the technical solution they will review. This should be a quick elevator pitch for your project. Be sure to answer the following:

* What is the model’s objective?
* What type of Machine Learning problem will you be solving?

**Technical Background:**

1-2 paragraphs describing the technical details of the problem you are trying to solve. This should help the reader understand the project constraints. Be sure to answer the following:

* How will you evaluate your model?
* What is your data source?
  + How will you need to prepare your data?
  + How will you explore your data?
  + What do you hypothesize your main features will be?
* What type of model do you want to use?

**Goals vs Non-Goals:**

Write a bulleted list of (3-5 points each) of goals and non-goals. This should help the reader understand the context that would factor into solution selections and trade-offs. Goals will help the reader understand what a successful outcome looks like. Non-goals will help limit the scope of your project and prevent scope creep.

### **Solution Overview**

1-2 paragraphs summarizing your ML System. Include a system architecture diagram containing the components you use to store data, pre-process data, engineer features, build/train/debug your model, and deploy your model. Also add notes on what you will monitor and what you will test prior to releasing a new model.

**Data Sources:**

What is your data source?

What is your data volume?

Why did you select this data set?

Any risks (bias, sensitive features, etc)?

**Data Engineering:**

How will you store this data?

What data pre-processing do you need to do before you feed it into your ML system?

**Training Data:**

How will you split your data into training, test and validation?

Will you use any data labeling techniques?

**Feature Engineering:**

Which fields from your data will you use or exclude?

Which fields will be combined or bucketed?

What other data transformations will you apply to your data?

**Model Training & Evaluation:**

How will you train your model?

What algorithm will you use?

What parameters will you use?

How will you evaluate your model?

**Model Deployment:**

What instance size will you use?

Will your model function as a batch or real time model? Why?

**Model Monitoring:**

How will you monitor your model?

How will you monitor your infrastructure?

How will you monitor your data?

**Model CI/CD:**

What checkpoints will your CI/CD pipeline contain?

What tests will your CI/CD pipeline contain?

**Security Checklist, Privacy and Other Risks:**

Will this store or process Personal Health Information (PHI)?

Will this store or process Personal Identifiable Information (PII)?

Will user behavior be tracked and stored?

Will this store or process credit card information?

If you answered yes to any of the above questions, please justify.

What S3 buckets will this application read from or write to?

What data bias should be considered?

Will your model have potential for bias along sensitive features (race, ethnicity, gender, age, religion, disability, sexual orientation, or other personal attributes)

Are there any ethical concerts with the data or business problems that should be addressed?

**Future Enhancements:**

Provide at least 3 ways you would improve your ML system if you had more time or additional resources.