Project Proposal Template

Project Name:
Team Name:
Team Members:

1. Define the problem

   1.1. Define the business problem
   1.2. What is the business goal or outcome?
   1.3. What is success look like
   1.4. What is the actual output you want to see from your model?
   1.5. What model do we choose?
2. Choosing Data
   2.1. Understand your data: How much, where, have access
   2.2. Have data you need? Is Data representative ?
   2.3. Evaluate the quality of your data
   2.4. Identifying features & labels you have
   2.5. Do you need a lot of labeled data?
3. Planning a Data Lake
   3.1. Ingestion needs (push / pull via streaming or batch)
   3.2. Security around data access
   3.3. Data retention and archival policies
   3.4. Encryption requirements
   3.5. Organization of data for optimal data retrieval
   3.6. Scheduling and job management
   3.7. Logging and auditing
   3.8. Technology choices comprising the overall data lake architecture (object store, HDFS, Hadoop components, NoSQL DBs, relational DBs, etc.)
   3.9. Overall design.

4. Identify Success
   4.1. Model performance  Metrics used during testing evaluation to express accuracy.
   4.2. Business Goal Metrics used after a model deployed measure model performance in the real world.