Vision Statement

Project Title: Constructing Intelligence
Team Lead: Vivek Patel vivek.patel@procore.com
Team Scribe: Nicole Moghaddas nicolemoghaddas@gmail.com
Charlie Getzen cgetzen@umail.ucsb.edu
Lauren Dumapias laurendumapias@gmail.com
Deanna Pham deannahpham@gmail.com

What the project is about?
What problem the project is solving?

The project will create a predictive metric for the construction industry.   Procore currently provides historical analytics, and would like to determine the feasibility of training machine learning models on its data set so that they can also provide some predictive metrics   This project will demonstrate the feasibility of doing so, as well as identifying additional opportunities for use of machine learning.

- The analysis will focus on Procore's quality and safety tools in order to foresee future hazards on project sites.
    - Include some sort of "Risk Score" in order to identify high risk projects to help prevent accidents
- The reach goal would be to predict additional predictive metrics, such as actual likely budget and schedules of projects, and the overall quality of projects, updating continuously as the project advances.

Why the problem is important?

One in five on the job accidents are construction related, resulting in 1000 deaths yearly out of 5000. Procore feels that identifying unsafe conditions and notifying individuals can greatly reduce these accidents. Using predictive analytics to possibly output risk scores for construction projects will allow companies to reassess their practices. Collecting and analyzing all this data might illuminate commonalities between high risk projects, so that predictions can be made for new projects, and prevent accidents before they happen.
Identifying when projects are behind schedule and over budget will also greatly assist companies in increasing their quality and services.

How the problem is solved today?

This problem can be solved with machine learning models.. Depending on the depth and scope of the data involved and the analysis done, these models can range from linear regression to deep learning.

The outcome of the project

The outcome of the project will provide a predictive metric that matters to construction agencies, using predictive analytics and machine learning models that are created from lots of construction data and reports collected by Procore's team (data on site accidents, safety incidents, project qualities, project schedules, project budgets, etc.)

- Concretely, some sort of Risk Score or project score to be generated using these models on a per project basis, allowing the construction agencies to assess their project safety and adjust accordingly.
- Furthermore, predictive metrics on additional project attributes, such as financial budgets, project progress, and overall job quality that include likelihoods of these figures, taking into account day to day changes, and modifying accordingly.
- Basically, an easy to use tool for construction agencies to assess whether their project is on track in a certain area or not (safety, budget, schedule, quality).

Initial project milestones (specification, design, prototyping):

- Need to get set up, get into accounts
- Attain sample data from sample companies
    - Determine what format the data should be in to train the models
- Determine the type of model to be used
- From model, generate risk score or project score (updated daily?)
- Determine data set and how to validate and test models - what data to train models, what data to test models?
- Identify additional predictions and questions that can be answered via the data/models

Plan on how to articulate and design the solution:

Implementation platforms and technologies to use to develop the solution:

Platforms:

- Python for data science
- Jupyter notebook (similar to mathematica) http://jupyter.org
- Sci-kit learn

Technologies:

- Linear regression
- Determine the ML techniques to be used

Overview of the process model employed to achieve the milestones:

The process model will be an agile process with daily scrums and four two week sprints.