Product Requirements and Design - Version 1 (PRDv1)

Constructing Intelligence

Procore Technologies, Inc.

The Team:
Team Lead: Vivek Patel
Team Scribe: Nicole Moghaddas
Deanna Pham
Charlie Getzen
Lauren Dumapias

1. Introduction

Define problem, innovation, science, core technical advance (2-3 pages)
Define project specifics, team goals/objectives, background, and assumptions

The problem and importance of this project stems from unsafety in the workplace, specifically on construction sites. A significant amount of on the job accidents are construction related; understanding, identifying, and resolving the underlying causes will be extremely beneficial to increasing the safety of these projects. The goal in this project is to use predictive analytics to identify unsafe conditions before they actually cause accidents on the job. The goal is to create a predictive metric for the construction industry, that will be useful and critical for safety measure on projects.

Furthermore, extending beyond project safety, it would be ideal to also predict metrics on project performance and quality. A reach goal of this project is to predict additional project components, such as schedule, budget, and quality, answering questions as to whether the project will be on schedule or whether the project will be over or under budget. This would be a very helpful tool for companies to be able to identify certain problem aspects in their projects, and take action before any damage is done. A continuously updating prediction of these project characteristics would be very advantageous to the overall project performance.

The science behind this project involves big data and machine learning. Procore currently provides historical analytics in retrospect about each construction project, 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 project will delve into applying machine learning with data science to Procore's large database of project logs and project data, and explore what can be predicted from this information, what project attributes can be predicted from previous project data.

This is a critical advance in project planning and management tools, such as Procore's project tool for the construction industry. Instead of solely being able to use the product for documentation and storing project aspects, such as daily logs, project schedules, and project financials, this project will introduce novel and innovative perspectives for companies to analyze their current construction projects. In addition to being able to manage the cumulative progress and assess the current state of projects based on past data, this project will redefine how construction companies can manage and assess their projects. If they not only have information about the current state of their project, but also predictive qualities about project characteristics, they can use this information to make safer, smarter, and more efficient choices in regard to project management.

In this project, the goal is to create a predictive metric for the construction industry. The analysis will focus on Procore's quality and safety tools in order to foresee future hazards on project sites. The most important goal will be to provide a "Risk Score" in order to identify high risk projects and to help prevent accidents (perhaps sending alerts if the score goes above a certain level), allowing construction companies to take immediate action. The "Risk Score" would entail something such as a percentage that describes the safety level of the project. 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, or perhaps just whether or not the project will be over budget, under budget, on or off schedule. This will be accomplished by using the extensive data that Procore has from past projects to build machine learning models to predict these characteristics for new projects.

The project data we will be given will include written documentation, such as project logs, project conditions, project specifications, project teams, etc., and project images, such as blueprints, floor plans, etc. We will analyze this data, and extract what features will be useful in predicting the properties we want to forecast. Furthermore, we will need to determine what it is feasible to predict, what is possible to determine from the data. We will then build models, obtain parameters, and use these models for future data, to predict project attributes. Understanding what features will be useful and obtaining the correct models for the different kinds of data will be the challenge of this project.

3. Requirements (functional and non-functional)

Use cases:

1) Use case: Predicting Schedule Timeline
Actors: Product manager, Procore application
Precondition: Project has been going on for an x amount of time.
Flow of Events:
Basic Path:

1. Product manager checks the application to get up-to-date information about the project's schedule
2. System will give the product manager an accurate representation of how far along the project is and if the project itself is delayed or not.

Alternative Paths:

If the product manager doesn't check the application, they will not have a good idea on how far along the project is and that may cause them to make an uninformed decision.

Postcondition: Product manager is now able to make informed decisions based on the data that was presented

2) Use case: Employee on site
Actors: Employee
Precondition: Employee is unaware of project conditions
Flow of Events:
Basic Path:

1. Employee arrives on site, not up-to-date since the x amount of time he has been off site
2. System will send a daily alert of an overview of the project as of that moment
3. Employee will be able to prioritize certain things that need to be done and if it is safe to be on site

Alternative Paths:

If the employee was not informed about the project, they would not be able to make the best decisions for the project from that point forward or

Postcondition: The correct decisions were made in scheduling, budgeting, and other important aspects of the project.

3) Use case: Instant inspections update
Actors: Inspector
Precondition: Unsafe working conditions not updated in daily inspection log
Flow of Events:
Basic Path:

1. Inspector inspects project site and notices many hazardous details

2. Inspector checks application and notices it is not accurately portrayed
3. Inspector is able to quickly add data
4. Immediately, system is able to accurately update the application

Alternative Paths:

If new data isn't added regularly, the application predictions could become inaccurate.

Postcondition: Everyone has access to a more updated application

4) Use case: General contractor supplies equipment
Actors: General contractor
Precondition: Project has been purchasing equipment without noticing the budget
Flow of Events:
Basic Path:

1. Employees under the general contractor have been making purchases as told without consulting the application
2. General contractor analyzes data given from application and realize the project has gone over budget

Alternative Paths:

General contractor never consults the application and the project becomes more expensive than the quota given at the beginning of the project

Postcondition: General contractor and their employees are now more aware of this when they purchase necessary materials

5) Use case: Updating data with daily alerts
Actors: Employees
Precondition: Data is not as accurate as it could be
Flow of Events:
Basic Path:

1. System alerts employees to update daily logs
2. Employee fills out daily logs and other features that will make predicting more precise daily
3. Application has the most up to date data to run on

Alternative Paths:

System does not alert employees and the project analysis is not accurate.

Postcondition: All employees have access to the most recent analyzation and will be able to make better decisions based on that

User stories:

1) As a developer, I want to be able to visualize the data features to see what features correlate together so that I can accurately base emphasis on needed areas of data.
2) As a developer, I need good real or simulated data that maps correctly so that I can accurately predict possible outcomes based on features and the map.
3) As a developer, I can predict whether the project is hazardous to the employees or not.
4) As a developer, I can predict on a scale from 1-10 whether the project is safe for a specific employee.
5) As a developer, I can go through millions of data sets and learn with each set to predict whether the project is on schedule, on budget, and safe so that construction companies can be more efficient.

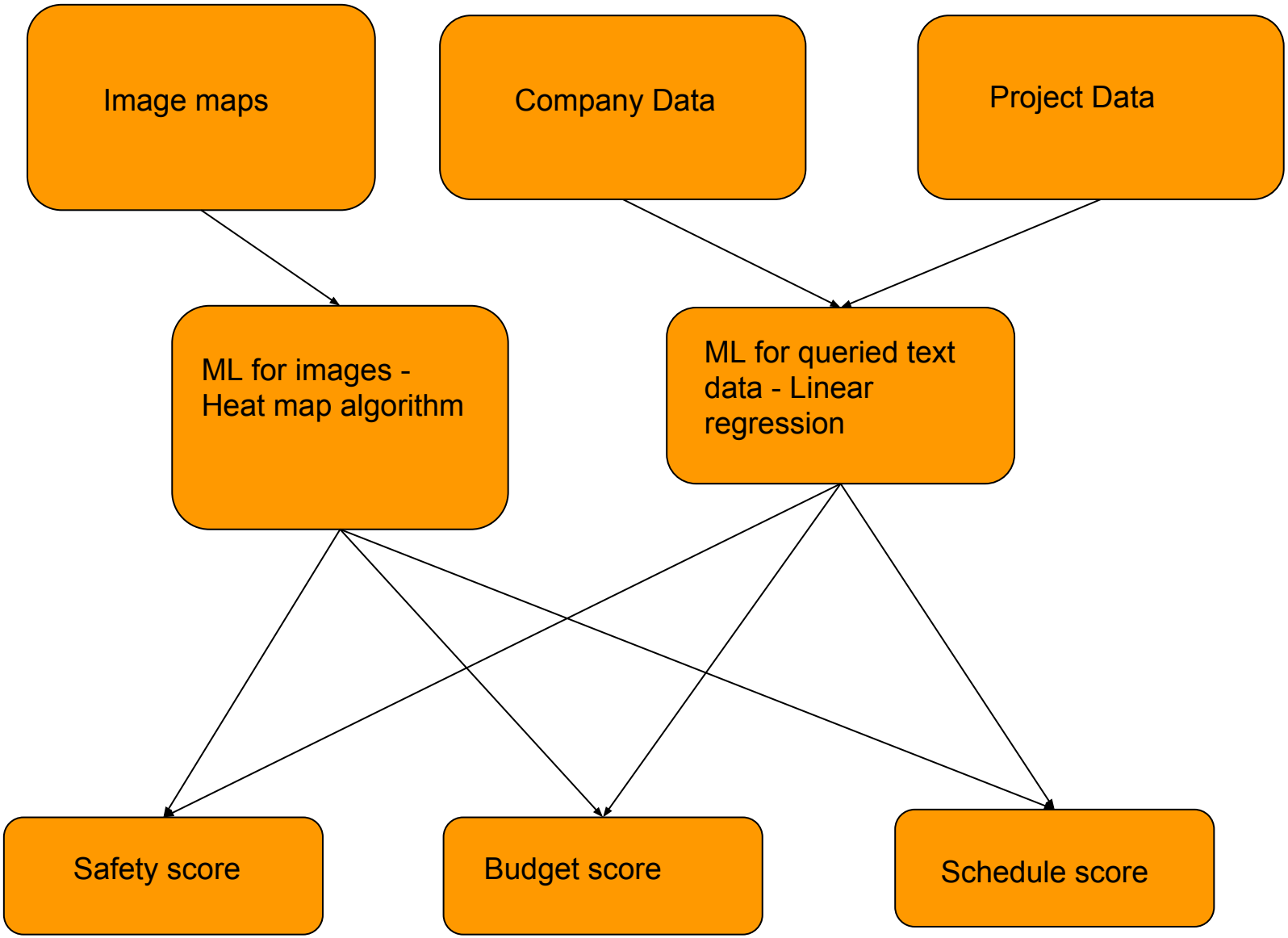
4. System Models

(PRDv2) Contexts, sequences, behavioral/UML, state

5. Appendicies

- Jupyter Notebook: A web application used to create and share “notebooks” containing code that can be compiled and ran by the application.
- Python: Our choice of programming language.
- Trello:A web application used to track our project’s progress with tasks under each ‘stage’ of our progress. Each member owns each card with certain tasks we each need to complete.
- Google Docs: Web application we use for our documentation.
- Machine Learning: A type of artificial intelligence we are using to train our program to predict many outcomes of a project.
- Python Data Science Libraries: Pandas, SciKit Learn, Numpy. These libraries will allow us to access their functions aiding us in our data analysis.
- Github: Our repository hosting service and our issue tracker.

2. System Architecture Overview



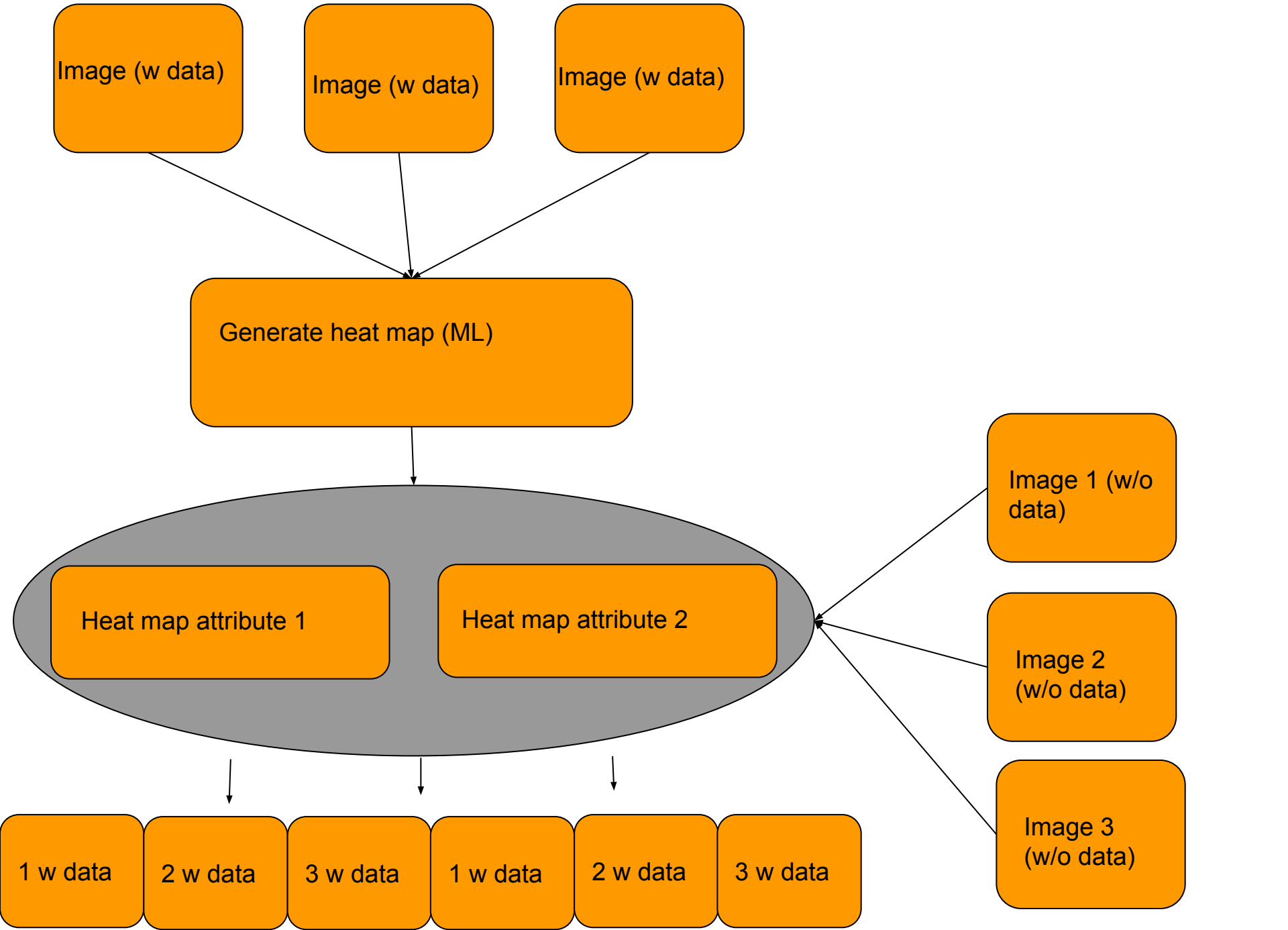


Image (w data)
Image (w data)
Image (w data)
Generate heat map (ML)
Heat map attribute 1
Heat map attribute 2
Image 1 (w/o data)
Image 2 (w/o data)
Image 3 (w/o data)
1 w data
2 w data
3 w data
1 w data
2 w data
3 w data