2   Project Plan

2.1 PROJECT MANAGEMENT/TRACKING PROCEDURES

Which of agile, waterfall or waterfall+agile project management style are you adopting. Justify it with respect to the project goals.

Agile. Our project is highly dynamic and calls for changing requirements within each respective sprint. The waterfall method isn't incredibly useful to us as we must often go back to what we were doing and retest or reevaluate.

What will your group use to track progress throughout the course of this and the next semester. This could include Git, Github, Trello, Slack or any other tools helpful in project management.

Trello. This is a service that allows for a Kanban board and lets us see who is doing what on any given sprint.

2.2 TASK DECOMPOSITION

Each of our tasks will be completed on a rolling basis – time constraints aren't incredibly binding. Most of our tasks will be stretch goals.
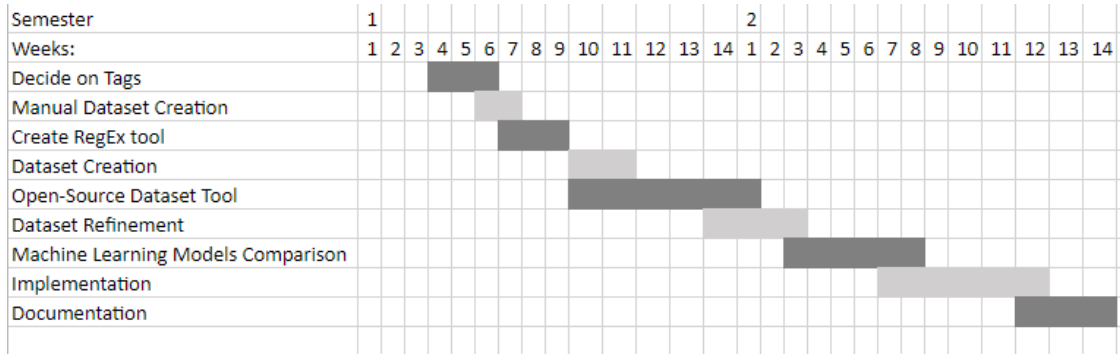
Timeline	Tasks
Semester 1	- Decide on tags for dataset - Create manually tagged dataset - Create RegEx (or other) tool to automatically tag source code - Collect large sample dataset - Evaluate how dataset will work; contextualize data and start work on neural network implementation - **Final deliverable**: open-source tool for source code dataset creation available to a wider public
Semester 2 (mostly stretch goals)	- Refine dataset/ how data is created - Experiment with several machine learning models and compare - Address limitations and create documentation for future work

2.3 PROJECT PROPOSED MILESTONES, METRICS, AND EVALUATION CRITERIA

What are some key milestones in your proposed project?

- The automatic dataset creation will tag 100% of tokens it comes across.

- The open-source tool will offer at least two ways of tagging Java source code.

- We will create ~100 tagged source code sets to allow for optimal training data.

2.4 PROJECT TIMELINE/SCHEDULE

Semester	1														2													
Weeks:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Decide on Tags				■	■	■																						
Manual Dataset Creation					■	■																						
Create RegEx tool							■	■																				
Dataset Creation									■	■																		
Open-Source Dataset Tool										■	■	■	■	■														
Dataset Refinement															■	■	■											
Machine Learning Models Comparison																		■	■	■	■							
Implementation																					■	■	■	■				
Documentation																										■	■	■

2.5 RISKS AND RISK MANAGEMENT/MITIGATION

Consider for each task what risks exist (certain performance target may not be met; certain tool may not work as expected) and assign an educated guess of probability for that risk. For any risk factor with a probability exceeding 0.5, develop a risk mitigation plan. Can you eliminate that task and add another task or set of tasks that might cost more? Can you buy something off-the-shelf from the market to achieve that functionality? Can you try an alternative tool, technology, algorithm, or board?

Task	Risks	Mitigation Efforts
Decide on Tags	N/A, largely a research task	N/A
Manual Dataset Creation	Incorrect Tagging, 0.2	N/A
Create RegEx Tool	RegEx not possible for a certain tag, 0.4

Tool is inaccurate, 0.6 | We have multiple ways for creating datasets (line by line vs word by word) which can help with RegEx limitations. With an inaccurate tool, this will delay the schedule. Time constraints aren't too dangerous as semester 2 is mostly stretch goals. |
Dataset Creation	Cannot find enough source code, 0.3	Create our own source code
Open-Source Dataset Tool	Tool not ready by Gantt chart deadline 0.2	Time limitations aren't crucial to deliverables
Dataset Refinement	Dataset doesn't fit into model, 0.7	Significant time setback. Readjust Gantt chart and reevaluate sprint deadlines.

ML Models Comparison	N/A	N/A
Implementation	N/A	N/A
Documentation	Documentation is lost, 0.1	Store digitally and rewrite.

2.6 PERSONNEL EFFORT REQUIREMENTS

Include a detailed estimate in the form of a table accompanied by a textual reference and explanation. This estimate shall be done on a task-by-task basis and should be the projected effort in total number of person-hours required to perform the task.

Member	Effort (weekly effort, task focus)
Robby Rice (Digital Content Coordinator)	Sem 1: ~3 hrs/ week, RegEx Tool Sem 2: ~2 hrs/week, Documentation
Maxwell Sutcliffe (Client Communication Coordinator)	Sem 1: ~3 hrs/week, Tag Deciding Sem 2: ~3 hrs/week, ML Models Comparison
Gavin Canfield (Quality Control)	Sem 1: ~4 hrs/week, Dataset Creation Sem 2: ~3 hrs/week, Open-Source Dataset Tool
Tanner Dunn (Agile Framework Organizer)	Sem 1: ~3 hrs/week, Dataset Creation Sem 2: ~4 hrs/week, Implementation
Amon McAllister (Individual Component Design)	Sem 1: ~3 hrs/week, RegEx Tool Sem 2: ~ 2 hrs/week, Implementation

All of these numbers are estimates and are not reflected in real hours worked per week.

2.7 OTHER RESOURCE REQUIREMENTS

Identify the other resources aside from financial (such as parts and materials) required to complete the project.

We may potentially need a GPU to run ML calculations on.