Practicum Presentation - Due Date 13th - 15th August

15th - 21st July

- ➔ Started adding more sections to the presentation
- ➔ Started working on the script
- ➔ Wrote introduction slide
- ➔ Wrote research questions slide
- ➔ Wrote background research slides
- ➔ Wrote methodology and evaluation slides
- ➔ Wrote technical challenges and conclusion slides
- ➔ Finished first draft

8th - 14th July

- ➔ Starting working on practicum presentation
- ➔ Added figures and contributions as well as some sections

Project Work/Software - Due Date 23rd July

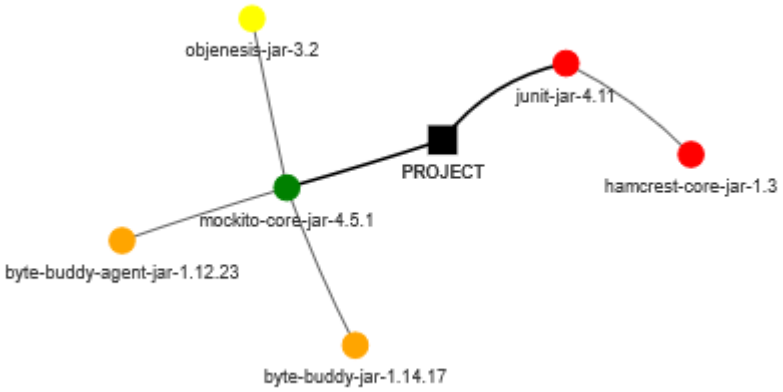
15th - 21st July

- ➔ Added to background research section
- ➔ Added more appendices
- ➔ Reread and corrected paper
- ➔ Reworded some sections
- ➔ Finished third draft
- ➔ Addressed all supervisor comments
- ➔ Added key terminology section
- ➔ Reworded some sections
- ➔ Rewrote conclusion and some parts of methodology
- ➔ Final reread
- ➔ Final draft

8th - 14th July

- ➔ Had peers look over the first draft and took their comments on board
- ➔ Added the appendices
- ➔ Added extra graphs and pipeline diagram
- ➔ Shuffled figures around to improve appearance
- ➔ Added more to Results and evaluation sections
- ➔ Finished Second Draft

- ➔ Minor bit of testing left to do and third draft
- ➔ Created maven project with two similar dependencies: Mockito and JUnit and checked results



- ➔ Updated Contributions Document

1st - 7th July

- ➔ Working on both the methodology section and the results/graphs section
  - ◆ Need to add more to the results/graphs section
- ➔ Working on the evaluation section - need to get average evaluation data for this
- ➔ Wrote a threats to validity section
- ➔ Reorganised the document according to the template sent
  - ◆ Need to fix website references though - done
- ➔ Working on rereading and adding some writing to some of the sections
- ➔ Finished First Draft
- ➔ Working on rereading and adding appendices
- ➔ Working on addressing recommendations from proofreaders
- ➔ Reworking graph locations

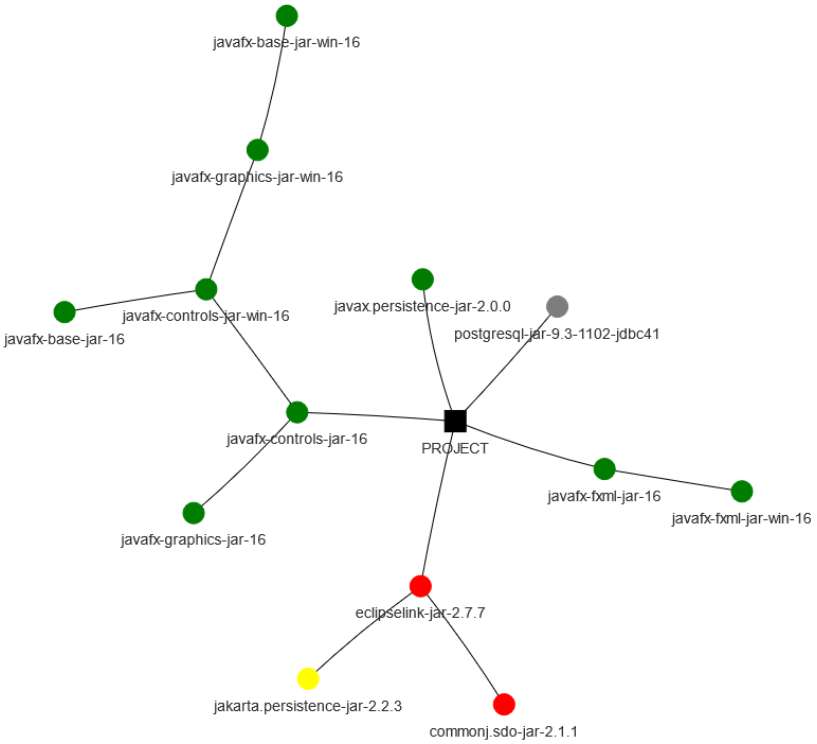
24th - 30th June

- ➔ Finished finding prediction evaluations including fixing inf issue
- ➔ Cleaned up code and updated task lists
- ➔ Changed variable names to more descriptive names
- ➔ Fixed issue with return commits risk score
  - ◆ Changed the calculation
- ➔ Emailed supervisor to ask for a meeting
- ➔ Fixed issue with vulnerability score always returning -5
- ➔ Fixed another typo issue
- ➔ Started working on the paper
  - ◆ Abstract
  - ◆ Introduction
  - ◆ Some of the calculations
- ➔ Added case studies section and redid the numbering

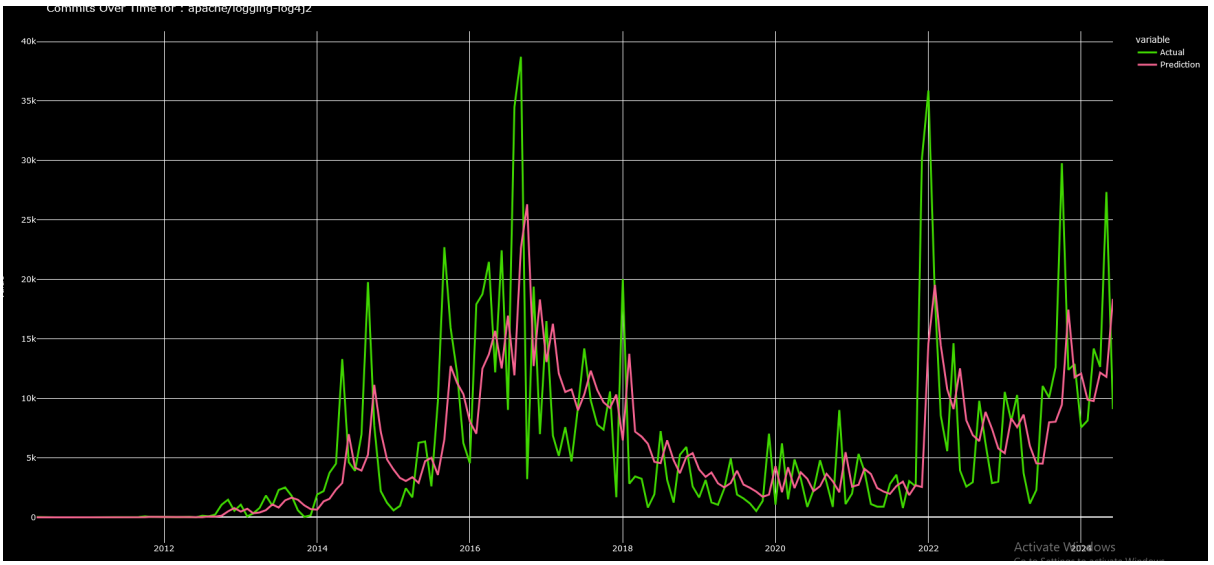
- ➔ Working on the related work section and the results and evaluation
- ➔ Redoing the abstract and the introduction
- ➔ Got the references working using Zotero
- ➔ Redid the appearance - two column, numbering
- ➔ Fully written first draft of - abstract, introduction, background research, case studies, started on the methodology section

17th - 23rd June

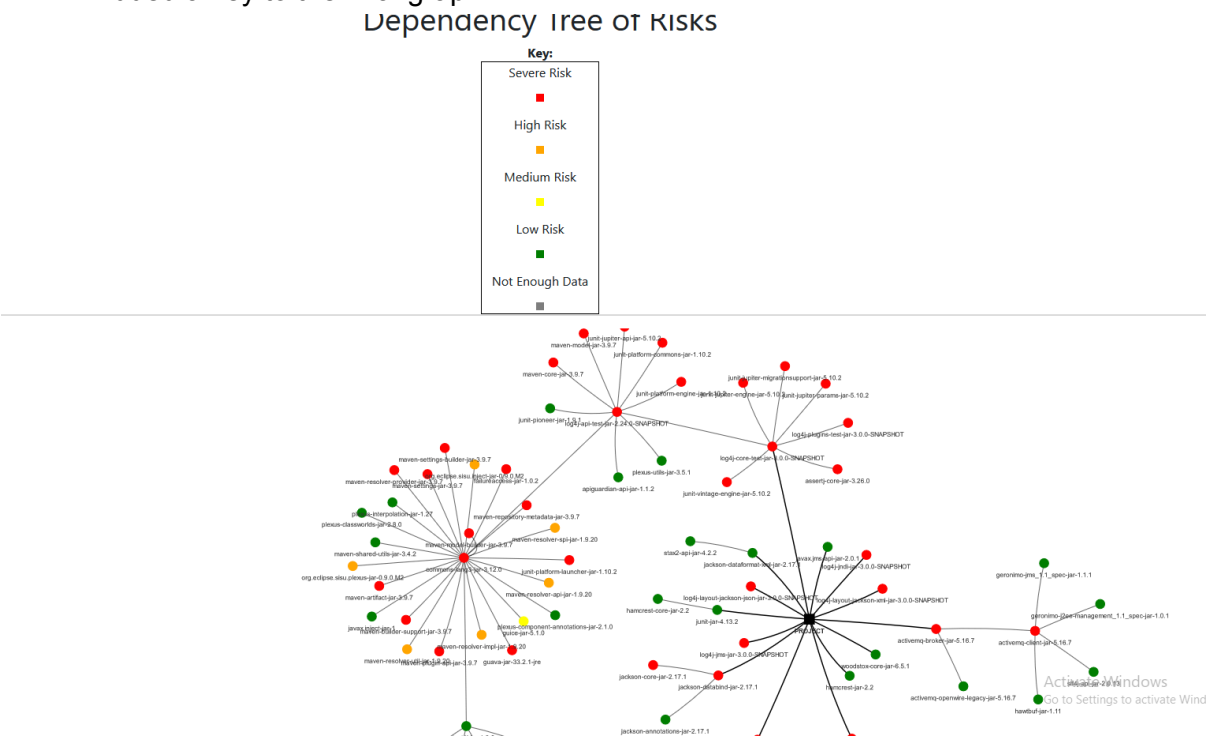
- ➔ Returning final value from predictions
- ➔ First look at graph



- ➔ This graph is just based on commits
- ➔ Have now gotten graph to display actual values vs commits - looking into changing the colour of second line
- ➔ The graphs now look really good like this:



- ➔ Updated the excel sheet of tasks
- ➔ Added to the current list of issues
- ➔ Added a key to the final graph



- ➔ Tested the project on another laptop to ensure portability
  - ◆ Reorganised data into folders
  - ◆ Created requirements to install for python
  - ◆ Wrote tutorial for partner to run code
- ➔ Finished keyword functionality
- ➔ Combined both prediction methods - graph now uses both
- ➔ Working on (finally) evaluating predictions - this may take a while
  - ◆ Done - using MAPE, MAE and RMSE

10th - 16th June

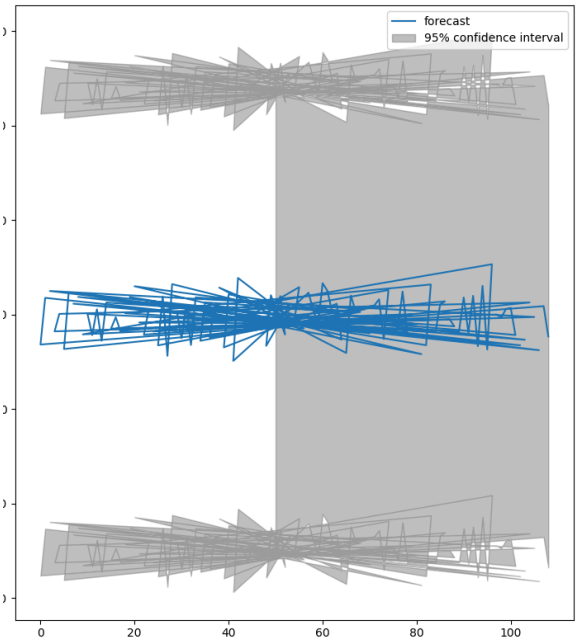
- ➔ Had meeting with supervisor
- ➔ Fixed auto parameters but some of the graphs look funky
- ➔ Working on adding github urls for log4j dependency tree - Done
- ➔ Started working on creating different configuration options - using a JSON
    - ◆ Finished working on integration
- ➔ Cleanup code/update description
- ➔ Next steps:
    - ◆ Fix graphs (Done)
    - ◆ Evaluate accuracy - using scores from summary/model
    - ◆ Redo commits section of code (Done)
    - ◆ Add actual score values - prediction values (Done)

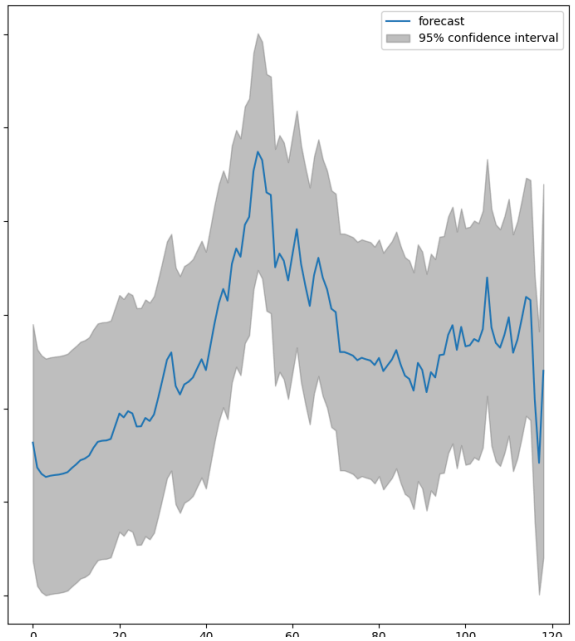
3rd - 9th June

- ➔ Ensured the ARIMA data is stationary
- ➔ Attempting to automatically find parameters:
    - ◆ p is the number of autoregressive terms
    - ◆ d is the number of nonseasonal differences - sorted
    - ◆ q is the number of lagged forecast errors in the prediction equation
- ➔ Wrote progress report

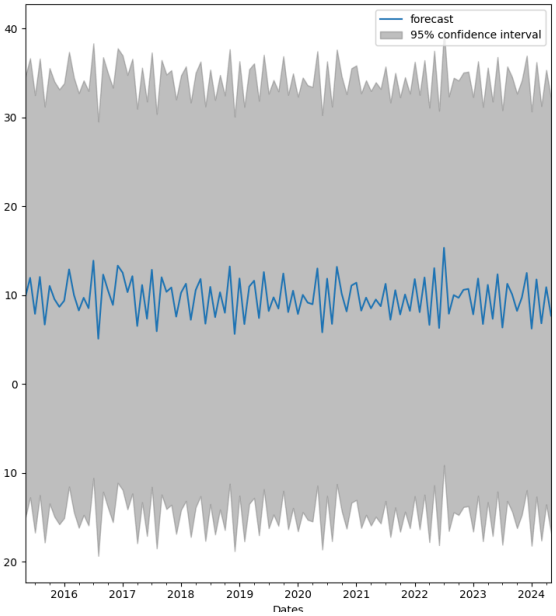
27th - 2nd June

- ➔ Working on ARIMA prediction for issue resolve time
- ➔ Meeting thesis partner to discuss progress
- ➔ Ensuring all of the graphs work correctly for project prediction
- ➔ Working on ARIMA prediction for now - working but some of them are wonky
- ➔ Evaluating the ARIMA predictions that we have so far
- ➔ I also have been working on using the dates for the ARIMA predictions - this will mean filling any empty sections in with a 0 or a -1 - decided on a 0
    - ◆ Maybe I will do it up to the current date actually
- ➔ Graph is also broken at the moment
    - ◆ Fixed the graph
- ➔ Now working on this tutorial to ensure values get fixed in the prediction





- ➔ Not sure why some of the graphs look like this and others are fairly accurate
- ➔ Managed to fix the predictions as well as add extra dates

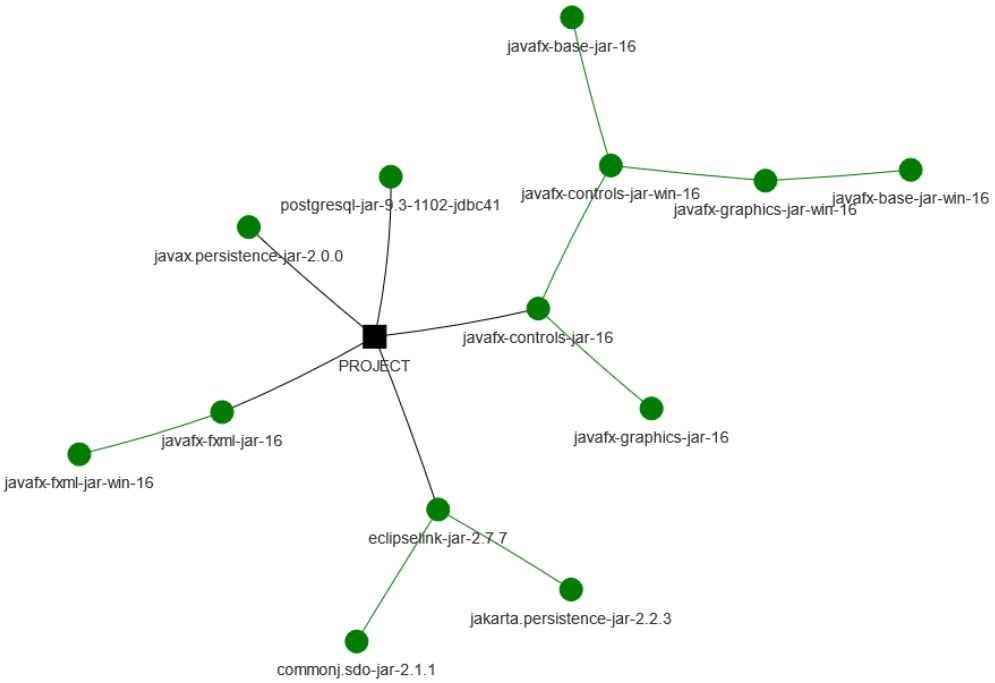


- ➔ This is the first prediction shown above fixed
- ➔ Have uploaded the current paper structure
- ➔ Now working on adding some to the methods section
- ➔ Created and assigned some more issues on Gitlab

20th - 26th May

- ➔ Creating a method to store the dependencies
- ➔ Working on creating a manual list of github links
- ➔ Investigated - maven, cmake and npm methods to gather dependencies
    - ◆ If maven dependency methods gets sorted I will create a method for these

- ➔ Made some progress on creating a graph of the dependencies - still needs to be improved



- ➔ Trying to find a better graphing method
  - ◆ For now sticking with the same method but using different colours for different risk levels
- ➔ Working on finding github links
  - ◆ Found github links for current dependency list
- ➔ Created function to populate dependency list
- ➔ Now working on getting a few different maven projects to gather more dependency trees
  - ◆ This is done - finished gathering github urls
- ➔ Next step working on graphing data for github issues
- ➔ Using a colour scheme to show risk in a package

13th - 19th May

- ➔ Starting working on predicting issue resolve time in GitHub Projects - using API
  - ◆ Not sure what prediction method to use here
  - ◆ Might also be worth adding open issues
- ➔ Organised GitLab and updated descriptions and READMEs
- ➔ Gathered data for all closed issues in apache spark repo
- ➔ Working on getting the time to close issues
  - ◆ For now will use ARIMA
- ➔ Met supervisor to discuss plan of action
- ➔ Started working on a method to extract dependencies from maven output
  - ◆ Running into an issue here where it is difficult to extract github links
    - Maybe worth creating a manual list
- ➔ Read up on log4j vulnerability

6th - 12th May

- ➔ Started gathering data for both the project prediction and vulnerability prediction section
- ➔ Graphed both of the data over time for apache spark
- ➔ Started the vulnerability prediction section using ARIMA (as this was highly recommended in papers)
- ➔ Found p, d and q using various testing on the data
- ➔ Started organising better sections in the code
- ➔ Wrote a draft plan of action for the project and emailed supervisor to sort a meeting

14th - 15th April

- ➔ Set up some functions for separating each section of the code
- ➔ Started working on some of the project prediction section - using github api
- ➔ Organised folders for data, code and docs
- ➔ Started gathering commit counts per day and creating dataframes
- ➔ Printed graphs of commit counts over time

11th March

- ➔ Setting up the overleaf for the project - including a basic structure
- ➔ Looking into linking zotero
- ➔ Started researching methods of extracting software metrics

Literature Review - Due Date 20th February

19th - 20th February

- ➔ Organised meeting with supervisor to discuss literature review draft
- ➔ Received feedback on draft
- ➔ Restructured draft as advised and fixed citations
- ➔ Added title and names as well as a table of contents
- ➔ Made sure to make any corrections as advised by supervisor
- ➔ Uploaded Literature Review to gitlab

8th - 16th February

- ➔ Had meeting with supervisor to discuss literature review draft
- ➔ Restructured the themes into two themes
- ➔ Continued researching papers and read another 17 papers during this period while writing the review
- ➔ Wrote vulnerability prediction studies section
- ➔ Wrote time series forecasting and vulnerability prediction studies section
- ➔ Wrote text mining and software metrics section
- ➔ Rewrote research question section

- ➔ Finished off section on time to fix and project activity predictions
- ➔ Added in statistics in introduction and research question section
- ➔ Wrote analytic conclusions
- ➔ Put all the papers in the taxonomy of papers in both themes
- ➔ Made last few edits on literature review draft
- ➔ Sent literature review draft to professor

29th January - 4th February

- ➔ Started organising papers read into different themes
- ➔ Came up with draft of planned headings and descriptions for literature review as requested by supervisor
- ➔ Wrote section on time lags in engineering process
- ➔ Sent headings draft to supervisor and organised meeting
- ➔ Wrote section on project activity
- ➔ Read another paper on lag in a network

25th January

- ➔ Sent an email to supervisor asking to discuss structure for the literature review

3rd - 11th January

- ➔ Started putting researched papers into themes
- ➔ Began working on extracting some CVE data from NVD API
- ➔ Reworded some of the theme headings
- ➔ Read another 8 papers and summarised their contents in the excel spreadsheet
- ➔ Managed to extract some necessary dates from CVEs

20th - 30th December

- ➔ Created and started organising excel sheet for research
- ➔ Organised sections in lit review as in drive
- ➔ Wrote introduction and survey question/motivation section
- ➔ Started coming up with some theme and subtheme ideas
- ➔ Read 7 papers and described the different points in the excel sheet

Practicum Proposal Presentation - Due Date 24th November

28th - 29th November

- ➔ Wrote structure of a script for proposal presentation
- ➔ Came up with bullet points for research question and plan
- ➔ Helped edit script sections for partner

- ➔ Had the proposal approved after the presentation.

Practicum Proposal

21st November

- ➔ Incorporated any last bits of feedback on proposal from supervisor

5th - 14th November

- ➔ Rewrote the section on the data we plan to use and the experiments
- ➔ Generalised the plan instead of focusing on time series to give a more general machine learning plan
- ➔ Updated the papers
- ➔ Sent the proposal draft to supervisor for review
- ➔ Incorporated feedback into report including rewriting several sections
- ➔ Resent proposal draft to supervisor

31st October

- ➔ Had meeting with supervisor to discuss proposal

22nd - 26th October

- ➔ Came up with some research questions. Wrote out a description of the topic.
- ➔ Described the planned exploration of these questions.
- ➔ Finished description of the topic and the research questions as well as the plan.
- ➔ Rewrote section on how the proposal relates to existing work.
- ➔ Sent draft proposal to supervisor.

11th - 13th October

- ➔ Started coming up with some research questions.
- ➔ Started on the proposal form and came up with an official topic for the proposal.

6th - 7th October

- ➔ Read some notes and did some research on the practicum including reading some papers and came up with some structure ideas for the proposal/project.

3rd October

- ➔ Read a paper and took notes in terms of structure and ideas for the project.

26th - 30th September

- ➔ Began researching three different project ideas.
- ➔ Read all papers that were in the project idea descriptions.
- ➔ Read extra papers to get a better idea of what the research in the area is like.
- ➔ Came up with some structure ideas for each of the projects.
- ➔ Decided which of the projects to email professors about.
- ➔ Emailed the professor and set up a repository.