**Project Timeline**:

11th October – 8th November (4 weeks):

- **Data collection** – this involves both actively looking for online lists of data, and inactively collecting data from VirusTotal using their API to periodically receive a live feed of submitted URLs (https://developers.virustotal.com/reference#url-feed).
- **Reading papers** – making detailed notes on papers which have investigated similar concepts and understanding their approach. These can be tested when I have the required data to see if I support their findings or identifying any improvements over their work.
- **Data understanding** – depending on the format of the data I will be using I can begin to pre-process/visualise the data to determine if I can identify any patters/clusters using basic statistical technique.

8th November – 10th December (4 weeks):

- **Feature engineering** – pre-process the collected data and select appropriate features based on understanding of data to be used in algorithms. Can implement GET request functionality at this stage to retrieve metadata but may not be includes in initial algorithms.
- **ML algorithms** – explore many different approaches detailed in other papers/learned in MLT module. Decide on a performance metric and at what point the solution can be classified as acceptable/successful (can define in terms of precision and recall, or as an accuracy metric).
- **Progress report** – begin write-up of progress up to this point.

10th December (hand-in):

- **Progress report** – finished by now.

10th December – 24th January (6 weeks)

- **Continue actions** – keep working on feature engineering and ml algorithms.
- **Revise for exams** – so I pass my modules.

24th January – 21st February (4 weeks)

- **ML algorithms** – refine algorithm selection, choosing the best from the set of possible algorithms and fine-tuning their hyperparameter values where appropriate to optimise performance (could explore other features).
- **Evaluate performance** – visualise results and analyse statistical elements of performance using determined performance metrics. Compare performance directly to those discovered from papers and comment on how similar/different results are.

21st February – 28th April (9 weeks)

- **Extension** – begin developing a chrome extension or some other user interface to be integrated with the model, prototyping different designs and seeing if model can be made more efficient.
- **Final report** – begin write-up of final report up to this point.

28th April (hand-in):

- **Final report** – finished by now.