Hate Speech Filtering Web Extension

- Yulong Cui 190198653
- Arkaitz Zubiaga

Background

- ‘Racism’ is one of the many expressions of our evolved capacity to live and work in groups.
- Prejudice based on race, skin colour, gender and sexual orientation.
- Most common form of racism is textual based.
- Euro 2020 Final - three England players targeted on social media.
  - More than 2000 racist tweets/posts deleted
  - More than 200+ social media accounts investigated
- Poor mental and physical health related to racism.

Project Aims

Minimise user's exposure to online racism

Reduce reliance on big social media companies to filter out harmful content

User protecting themselves from hate speech, be proactive rather than reactive

Project Objectives

Text	Fetch textual content from websites, store locally as text files using BeautifulSoup
Algorithm	Create a Machine Learning algorithm that performs analysis on text files using Python NLTK and Python Profanity library
Extension	Create a user-friendly web extension using JavaScript, Django, and HTML
Integration	Integrate ML algorithm with the web extension

Computing/Engineering problem

- Filtering textual based content requires:
  - Natural Language Processing (NLP)
    - Variations of words
    - Combinations of phrases
    - Understanding in context
  - Machine Learning algorithm
    - Pattern recognizing
    - 'Training data' oriented

Literature Review Findings

Python & related libraries

Machine Learning

Existing systems

Python & related libraries

Python programming language

- Object-oriented
- Dynamic typing, binding
- Extensive selection of Machine Learning libraries

Natural Language Processing (NLTK libraries)

- Text processing libraries:
  - Classification
  - Tokenization
  - Stemming
  - Tagging
  - Parsing
  - Semantic reasoning

Profanity filter

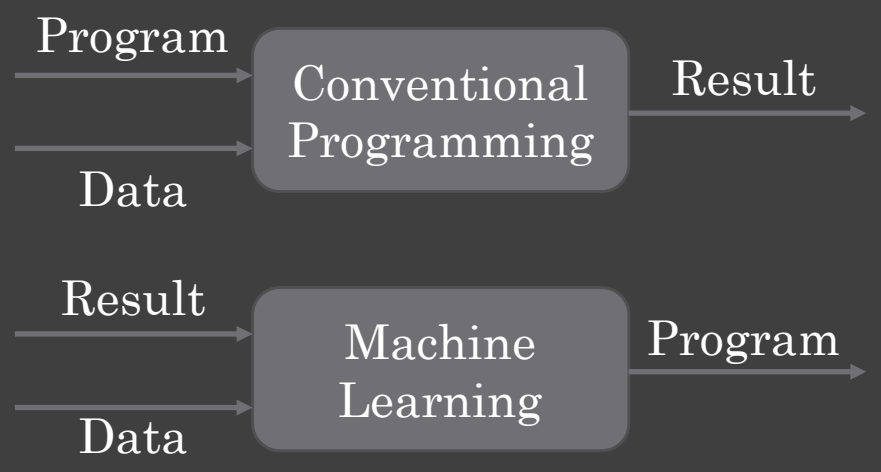
- Word censoring
- Deep analysis
- RESTful web service

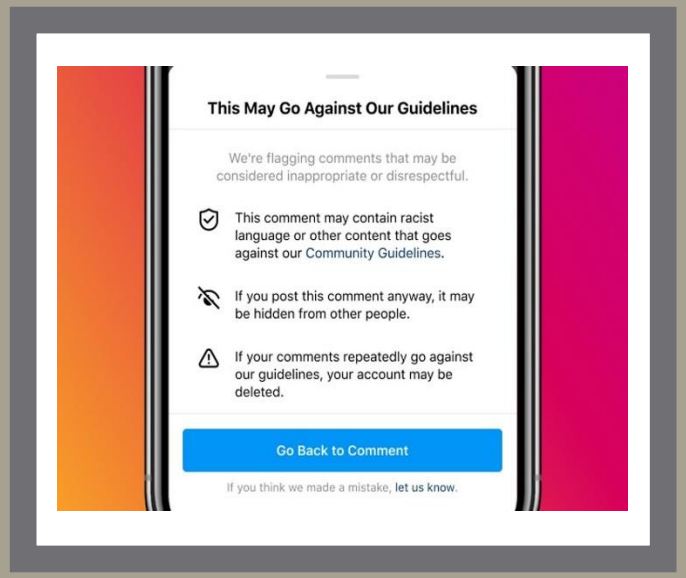
Beautiful Soup

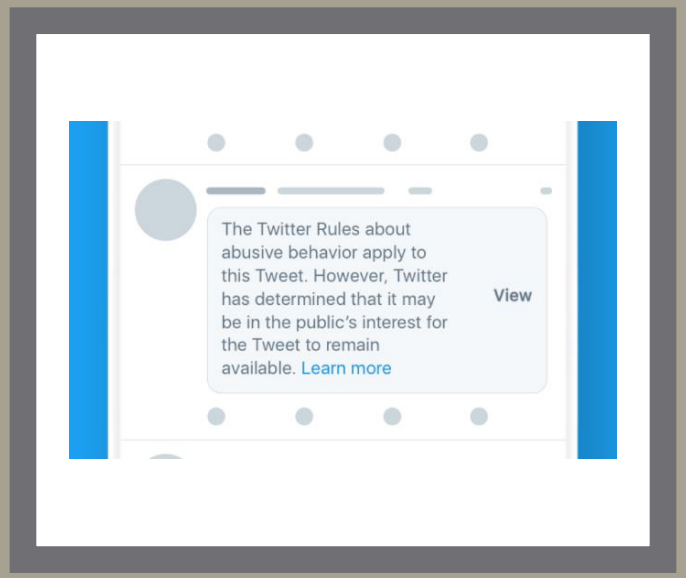
- Utilizes HTML/XML parser
- Web scraper

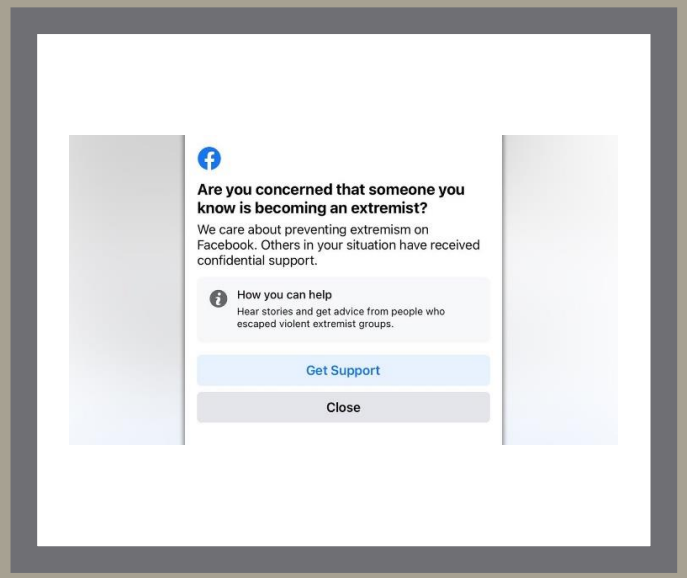
Machine Learning

- ML algorithms build a mathematical model/program using pre-written algorithm depending on 'training data'.
- Conventional programming outputs results depending on user written logic/programs.



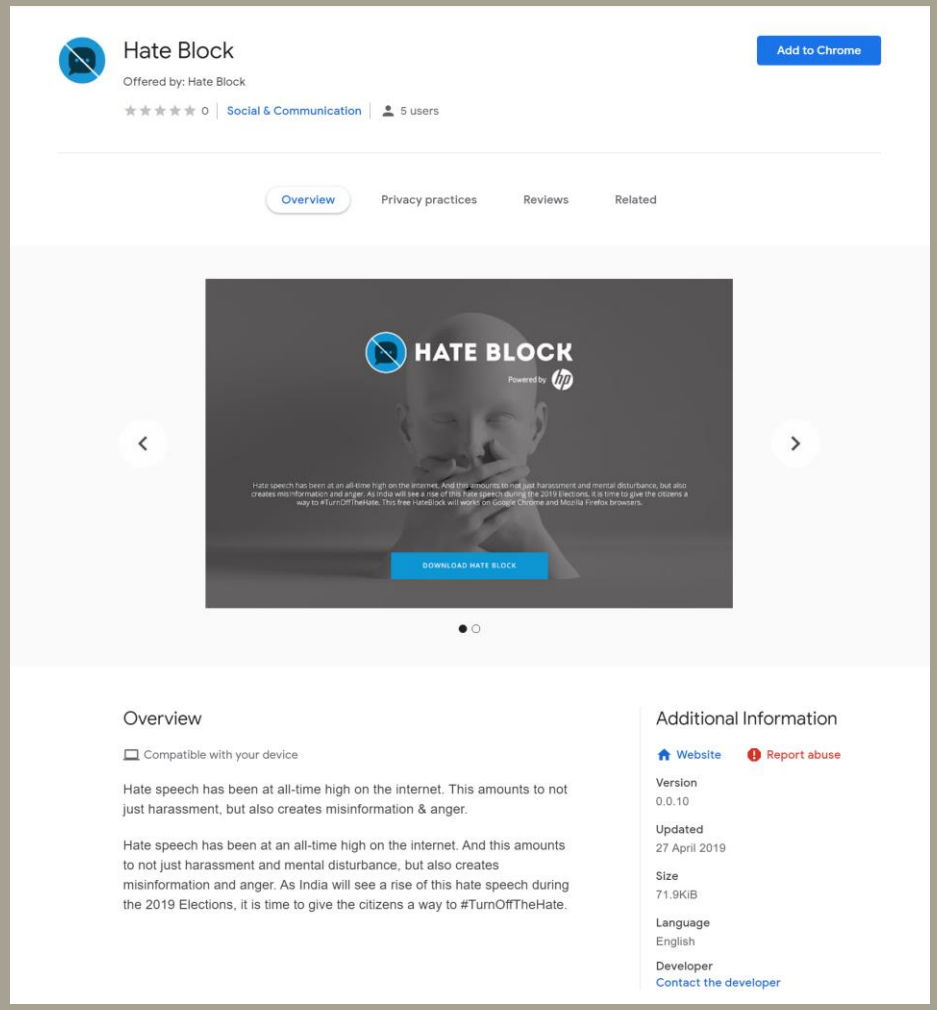


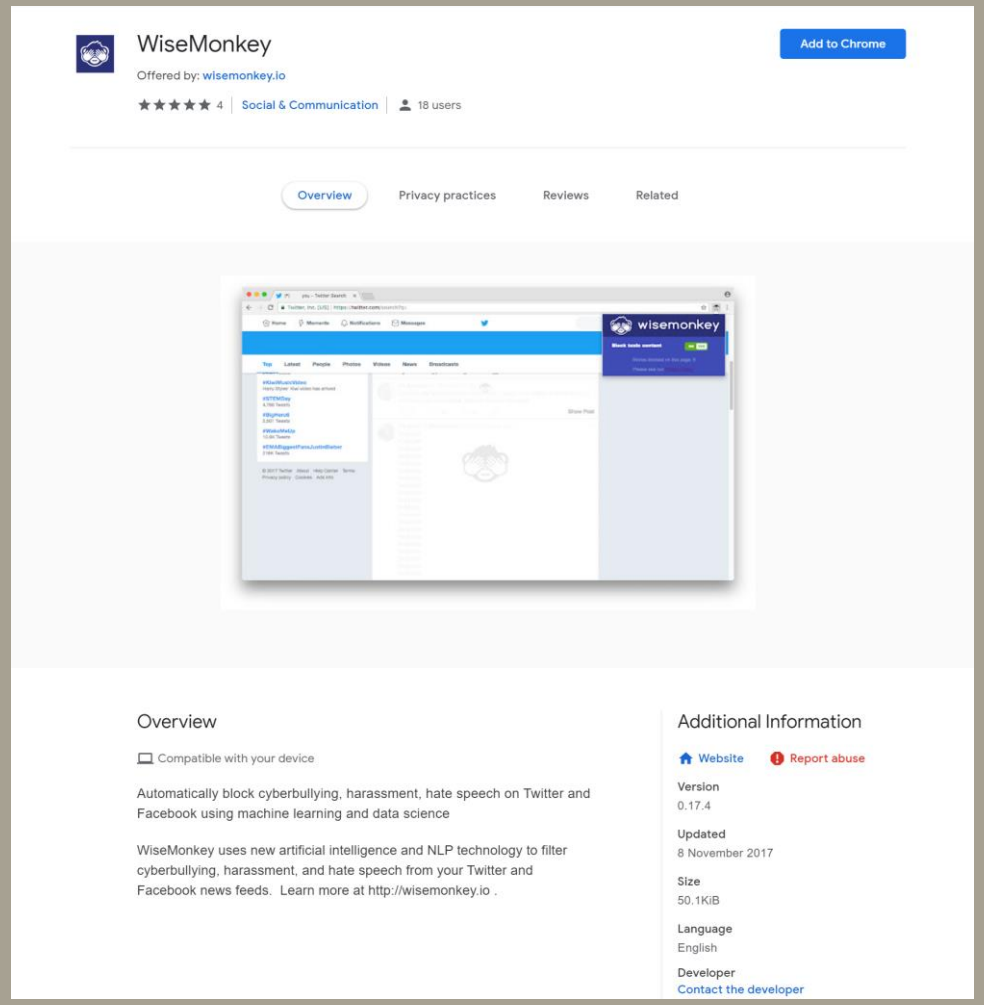




Existing systems

Meta (Facebook) social media platforms' content filtering system



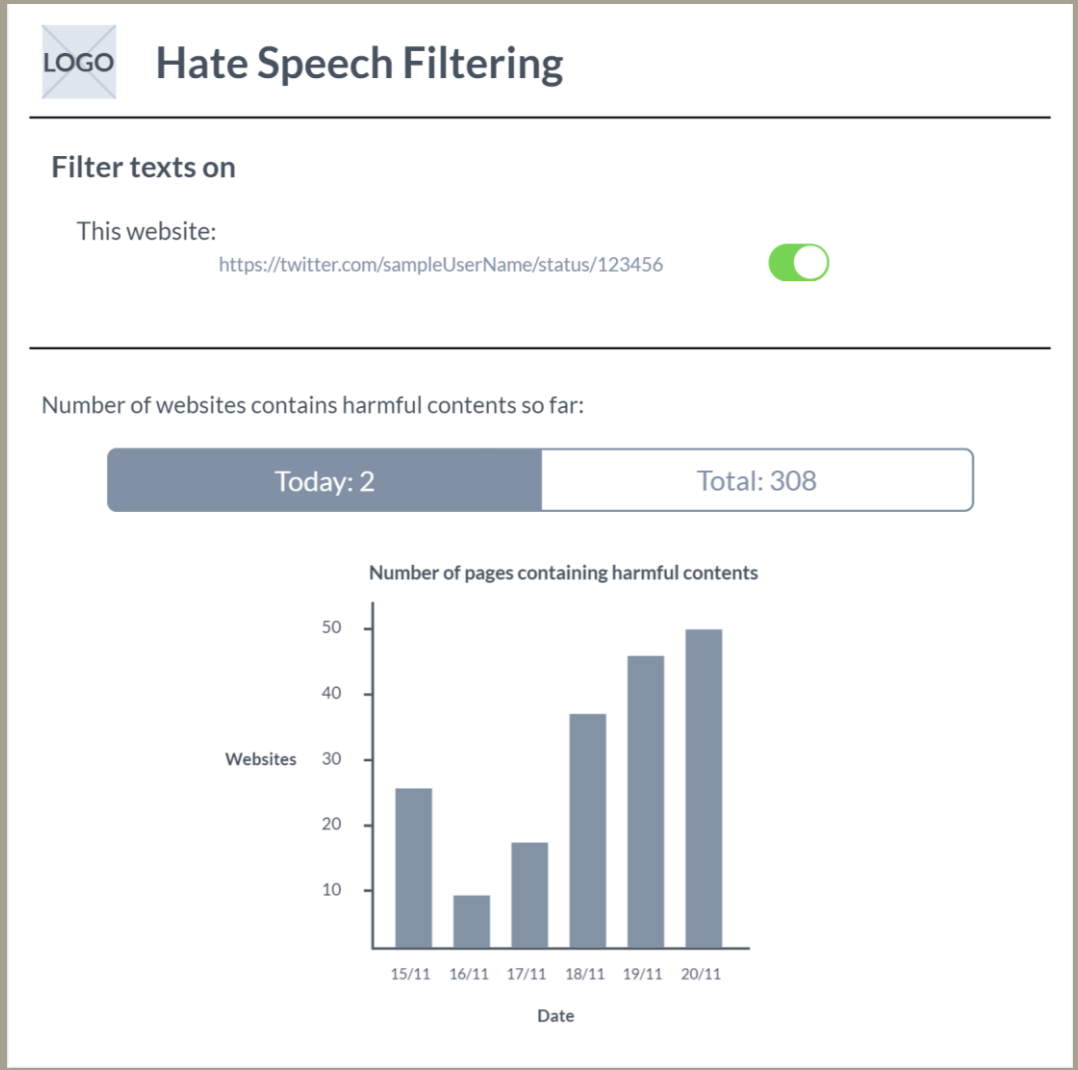


Existing systems

Similar Chrome web browser extensions

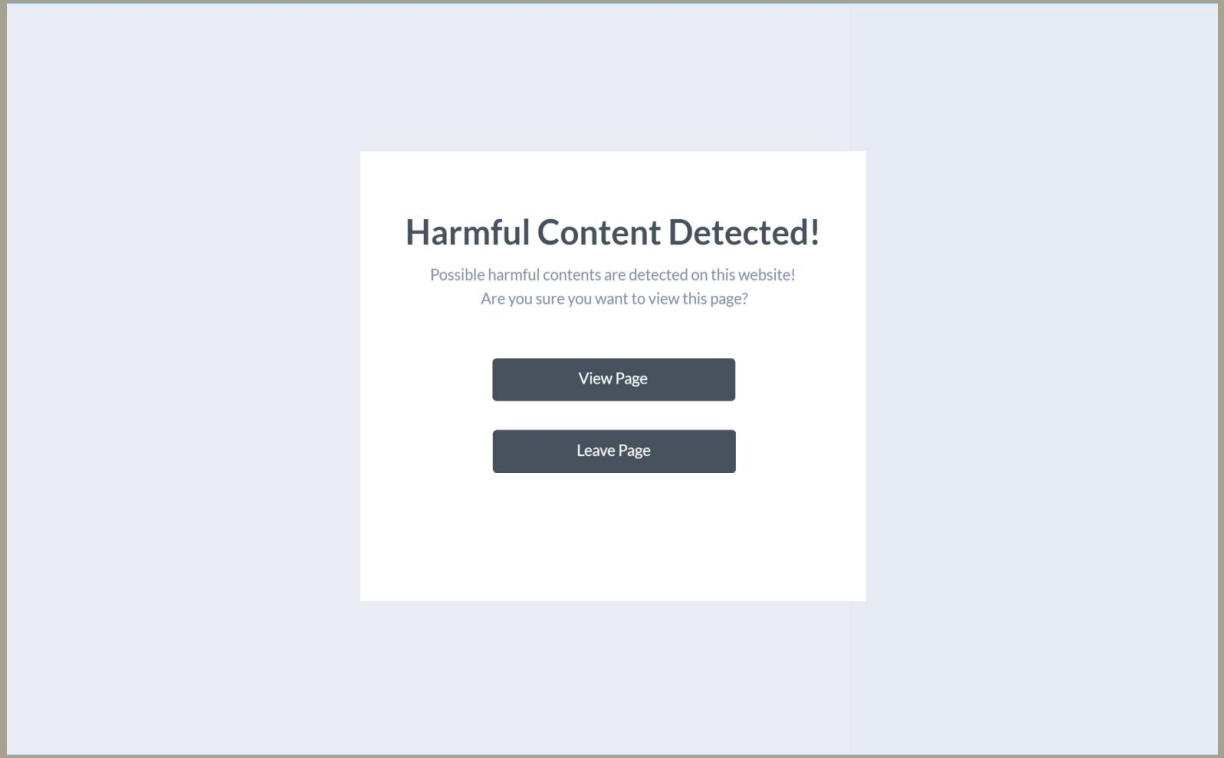
GUI Design

-Extension Interface



GUI Design

-Warning splash screen



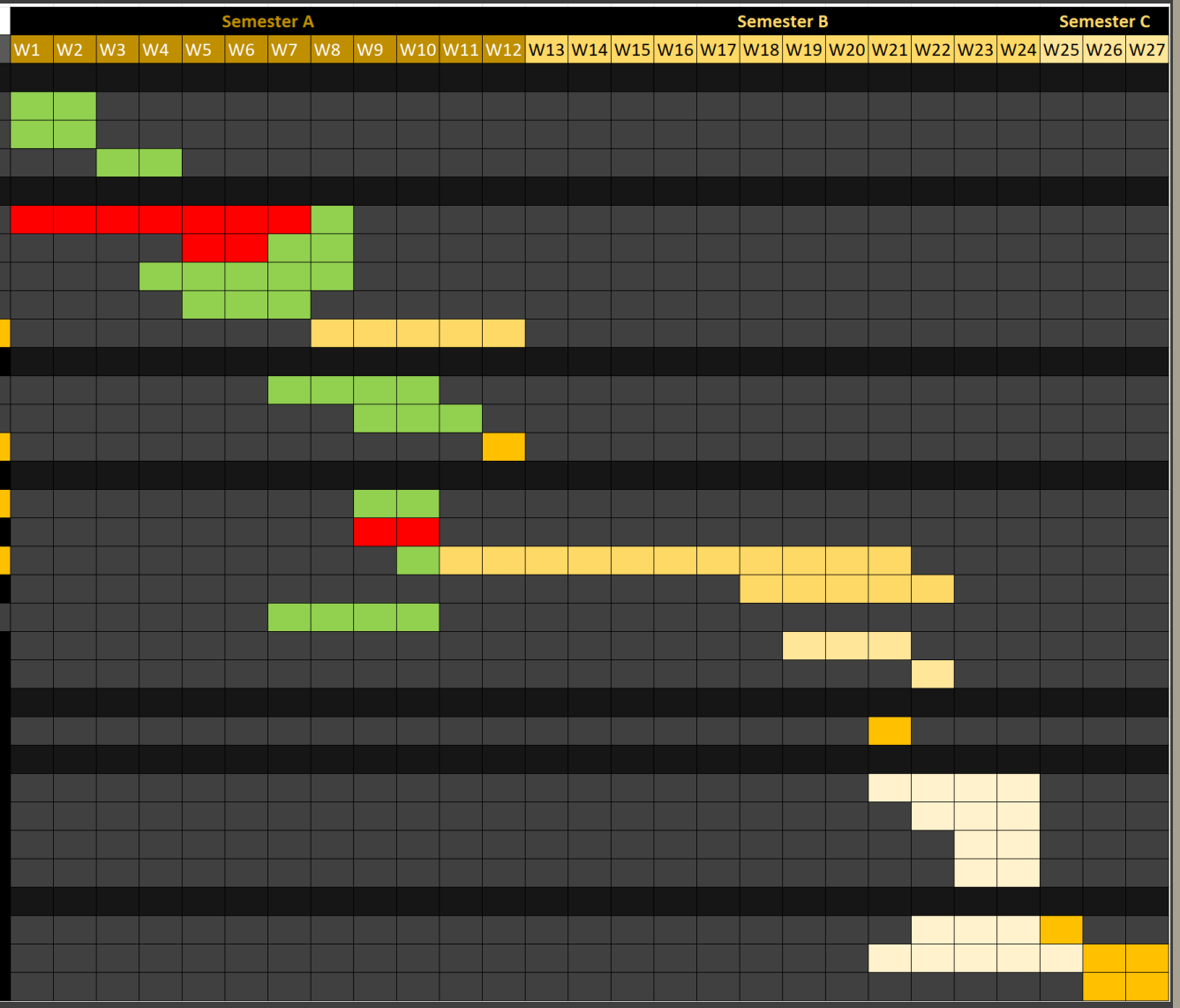
Queen Mary
University of London

Evaluation

- Text scraper
  - Social media websites
    - Tweets/posts
    - Replies/comments
  - Published documents
- Machine Learning algorithm
  - Training data
  - Existing results
- Web extension
  - Developer testing
  - Sample user testing

Project Planning

Phase	Task	Due Date	Status
Research Phase	Project Definition		
	Background Research & Context	2-Jul-21	Completed
	Aims & Objectives	6-Sep-21	Completed
	Problem Definition Submission	18-Oct-21	Completed
	Research Methodology		
	Literature Review	25-Oct-21	Completed
	Primary Research	28-Oct-21	Completed
	Testing Criteria Research	3-Nov-21	Completed
	Technology implementation research	5-Nov-21	Completed
	Machine learning implementation research	17-Nov-21	Started
	Interim Presentation		
	Interim Report	29-Nov-21	Completed
	Progress Slides	6-Dec-21	Completed
	Progress Presentation	13-Dec-21	Started
Machine Learning Phase	**Design & Development**		
	Environment setup	31-Jan-22	Started
	Data gathering implementation	7-Feb-22	Not Started
	Implementation of algorithm	22-Mar-22	Started
	train algorithm	27-Mar-22	Not Started
Application Phase	design user interface	25-Mar-22	Completed
	implement design	6-Apr-22	Not Started
	Integrate algorithm with UI	6-Apr-22	Not Started
Finalisation Phase	**Report**		
	Draft Report	21-Mar-22	Not Started
	Testing		
	Write tests for python backend	21-Apr-22	Not Started
	Write tests for javascript frontend	22-Apr-22	Not Started
	Test & document for backend	30-Apr-22	Not Started
	Test & document for frontend	30-Apr-22	Not Started
	Conclusion		
	Project Video Submission	6-May-22	Not Started
	Final Report Submission	4-May-22	Not Started
	Viva	09 to 20 May 22	Not Started



Risk Register

Risk Description	Likelihood rating	Impact rating	Impact	Preventative actions
Poor time management	Medium	High	Affect entire project timeframe, resulting in unfinished final product or a product with less standards	Utilise Gantt chart to track progress regularly, ensure sufficient progress is made on time or ahead of the time
Personal illness and unforeseen events	Low	Medium	Cause delays in final product	Ensure healthy personal schedule and diet, protect mental health
Lack of technical expertise	Medium	High	Cause delays in final product and lower standards in final product	Ensure regular contact with advisor, seek help from experienced peers if necessary
Loss of code/progress	Low	High	Loss of large portions of work if forgot to save progress/computer problems	Use cloud services such as GitHub and OneDrive for code backups
Project development becomes too large to finish before deadline	Low	Medium	Project overextends to unnecessary territories causing loss focus on main functionalities	Stick with original plan for project, minor changes can be done by myself, consult experienced peers and advisor for any major changes to project

References

1. Brooks, R., 2021. The Origins of Racism. [online] The Conversation. Available at: <https://theconversation.com/the-origins-of-racism-8321> [Accessed 26 November 2021].
2. Perrigo, B., 2021. *Twitter Offers Transparency on Racist Abuse, But Few Solutions*. [online] Time. Available at: <https://time.com/6089289/twitter-racist-abuse-anonymity/> [Accessed 26 November 2021].
3. Lewsley, J., (2020). The effects of racism on health and mental health [online]. *Medical and health information*. Available at: https://www.medicalnewstoday.com/articles/effects-of-racism#adults [Accessed 26 November 2021].
4. What is python? Executive summary [online]. *Python.org*. [Viewed 27 November 2021]. Available from: https://www.python.org/doc/essays/blurb/
5. Python AI: why is python so good for machine learning? [online]. *Custom Software Development | Netguru*. [Viewed 27 November 2021]. Available from: https://www.netguru.com/blog/python-machine-learning
6. NLTK : natural language toolkit [online]. (no date). *NLTK : Natural Language Toolkit*. [Viewed 27 November 2021]. Available from: https://www.nltk.org/
7. Profanity-filter [online]. (no date). *PyPI*. [Viewed 27 November 2021]. Available from: https://pypi.org/project/profanity-filter/#deep-analysis
8. Beautifulsoup4 [online]. *PyPI*. [Viewed 27 November 2021]. Available from: https://pypi.org/project/beautifulsoup4/
9. How different are conventional programming and machine learning? - [online]. *Tech Blogger*. [Viewed 28 November 2021]. Available from: https://contenteratechspace.com/how-different-are-conventional-programming-and-machine-learning/