

Project Proposal 2023

23/11/23

Zak - 26577836

Max - 26584263

Jackson - 26580601

Hubert - 26505141

Alfie - 26584784

Dan - 26569011

Project Proposal

Project Scope:

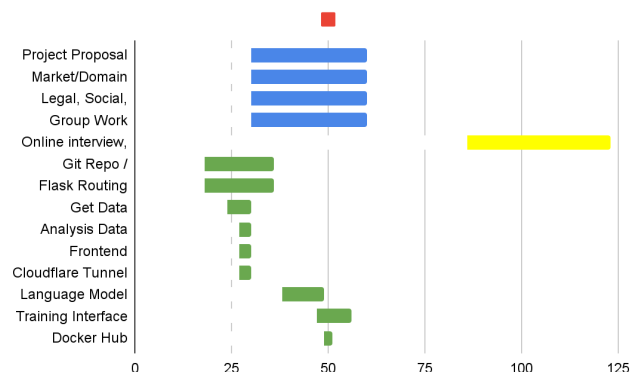
A web-accessible application, that allows users to query a tweet. The query will calculate a sentiment analysis for the data from our trained ML model and display it in an appropriate format. The sentiment analysis will be a score from 1-5 determining the emotion of the tweet.

Extensions for the project could be:

- Allowing user input on the sentiment to further train our ML model and comparing it to third-party ML models.
- Adding the ability to analyse Twitter's 'Timelines', trending Hashtags, individual users and comment threads.
- Performance optimization, testing for latency and improvements.
- Caching data/session handling with cookies
- Allowing for personal profile analysis

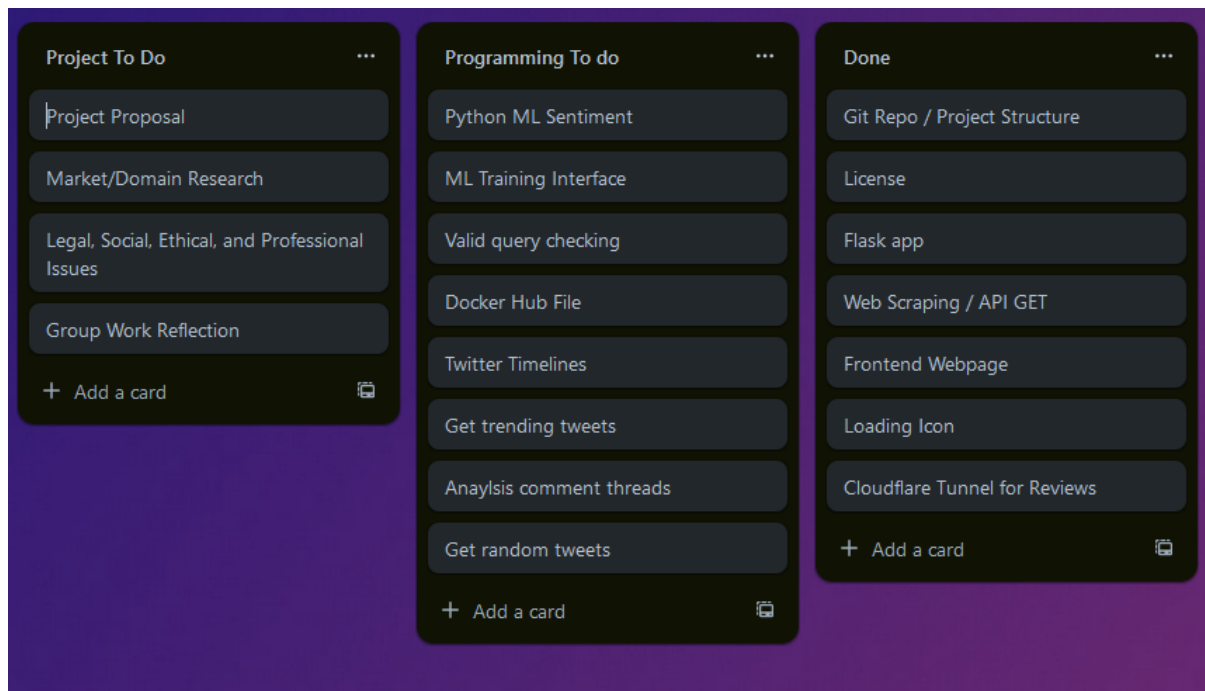
Project Plan:

The group decided that the agile/scrum model would most likely suit. This is because we had fast-growing ideas and frequent structural changes which were likely to cause problems in the waterfall model. Using iterative releases handled by GitHub makes the 'sprint review' very beneficial for maintaining targets. As a group, we found that the Gantt chart was limiting when trying to approximate time goals for complex tasks, where some might be much quicker or longer than expected. This is why we opted to only create a simplistic Gantt chart and opted to mostly use the website 'Trello' to track progress.



Monitoring and Evaluation:

By defining certain objectives in the 'scope' we can properly manage the progress of the project. We will measure progress by using feature milestones, e.g. a UI update or ML model implementation. Monitoring that the project remains on schedule is a very high priority for the team and will be done in the weekly reviews and by the Project Coordinator. Using the project management tool 'Trello' we can appropriately organise tasks and assign priority to ensure important tasks are completed.



Communication Plans:

The Team will have two weekly meetings to review progress. Every Tuesday at 12pm on campus and an online Microsoft Teams on Friday at 6pm with the Project Supervisor. For team discussion, a more frequent means was needed, so a group chat on the platform 'Discord' was created, allowing for time auditing and voice calls.

The stakeholders will primarily be contacted in the means of interviews and questionnaires to further our understanding of the project, mainly communicating with the researcher to then relay to the rest of the group to alter targets, such as wanting a callable REST API instead of a website interface.

Risk Management and Limitations:

The largest risk assessed is time management, allocating too much work in the given time frame. Even so after some group discussion we believe the targets we have set are appropriate and achievable. Limitations for the project are differences in knowledge, where some members are more familiar with the technology chosen so they have to learn new skills to be able to contribute to the codebase.

Role Definitions:

→ *Max* - **Team Leader / Lead Programmer**

The lead programmer's role will be responsible for the decision making and technical leadership, dictating the code reviews and ensuring that code standards are met. Providing efficiency of the codebase and guidance to other team members.

→ *Dan* - **Project Coordinator**

The Project coordinator's role will be responsible for ensuring that the other members are present in meetings, including recording absences.

→ *Hubert & Zak* - **Researcher**

The Researcher's primary role will be to ensure that the current project is meeting target stakeholder needs. The researchers will also be analysing other competitors to inform the team where our application may be inferior.

→ *Jackson & Alfie* - **Quality Assurance / UX Tester:**

The QA/UX Tester will insure that the software meets specified requirements, providing feedback to the overall quality of the software

Project Overview Decisions:

Frontend:

- Web Server: The Python web framework 'Flask' was picked because of its easier integration with the ML model and data processing even though its performance is slightly inferior to 'Node.js' and asynchronous application.
- Webpage: HTML 'div' elements are used to define the sections of the document, whilst using CSS grid to place our content where we need it. A mix of Javascript and Jinja are used to embed our Python content into the page and handle loading animations.

Backend:

- ML Model: The team has yet to pick a specific ML library as we still need to run performance testing on popular models following our agile model, but due to our current project structure it's most likely to be 'scikit-learn' in Python.
- Async Handling: Python 3's inbuilt library 'asyncio' was used for its 'await' function as we need to wait a small amount of time to retrieve and process the Twitter data.
- Data Retrieval: We found Twitter data to be quite challenging to retrieve after the recent API changes, so we first started with scraping the data, but this process was very slow as Twitter forces javascript enabled browsers, meaning an entire browser process had to be created, waited to load and then scraped. This required lots of compute resources and wait time. After reviewing their API for a second time, we found a method to retrieve tweet data from their 'embed api' which was meant for web pages to be able insert specific tweets, not requiring authentication or subscription. The data inside of the JSON contains the full HTML data including the text - this was a massive performance optimization.

Market Research

Provide evidence to demonstrate the need?

Twitter is a large opportunity for advertisement, as one of the largest social media platforms in the world it allows businesses to reach an audience as large as “550 million monthly active users” [1]. The use of sentiment analysis within Twitter would allow brands to manage customer satisfaction, using positive feedback and statistics would allow them to maintain happy customers, whereas negative would allow the growth and adaptation of the brand to suit the preferences of the customers, this is how they would manage their brand image.

Additionally, brands could target influencers who have high positive sentiment posts to improve their marketing performance. The results from Influencer advertising on social media platforms has proven to be very effective, for example, the game Raid Shadow Legends,[2] “amassing more than 10 million downloads worldwide as of December 2019...”. The study [3] also shows the impact of celebrity endorsements.

Will your project meet a niche gap in the market?

The market currently has some sentiment social media analysis tools but all of them are very costly, we hope by keeping our application free throughout its lifespan we can meet a user niche. The current top three tools stated by the blog ‘www.softwareadvice.com’ are ‘IntenCheck’, ‘SocialSearcher’ and ‘text2data’ which average 12.81 GBP monthly.

Individual Sentiment analysis tools exist for free, but this requires having to manually copy-paste the tweet text and replies, causing a large inconvenience for the user. Therefore if we provide a service that is directly suited for Twitter and free, we will be more valued than competitors.

<p>Strengths:</p> <ol style="list-style-type: none"> 1. Flexibility on what we can implement in our application, such as comment thread analysis and individual user analysis. 2. Much more affordable than other competitors. 3. Introduce new ideas and perspectives on how sentiment analysis does not have to be so expensive. 	<p>Weaknesses:</p> <ol style="list-style-type: none"> 1. Challenges in training ML models could take a long amount of time and resources. 2. Low amount of brand recognition and no money for marketing. 3. Smaller and less experienced team and less of an audience.
<p>Opportunities:</p> <ol style="list-style-type: none"> 1. As a relatively new market, this means that there is a high amount of opportunity to meet the high demand. 2. Working with companies and influencers based upon our target stakeholders, potential increase for awareness of our team. 3. Expand market / social media presence by collaborating with other people in the space. 	<p>Threats:</p> <ol style="list-style-type: none"> 1. The existence of previous text sentiment analysis tools could saturate the market and may make our application appear less unique. 2. Potential competition could develop similar features and match our prices, meaning high chance we could be beaten to it.

LSEP

Legal Issues:

GDPR is the most important regulation that should always be considered when discussing data processing, and our project is no exception. The data collected from our application to use in our ML model cannot be collected without appropriate consent from the user. The user needs to be correctly informed of what data will be collected from them. Legally we must also consider the terms of service of "X" (previously known as Twitter). If the TOS is violated the application could face a cease and desist and be shutdown. We will be able to avoid this as we have studied "X"s TOS, realising at the start of the project that we were not able to webscrape hence **"crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited."**[4] so the team switched to using an API. Further sentiment analysis other than positive and negative could also be breaching the *Defamation act 2013*. This is due to the program making potentially false analysis against a person or corporation.

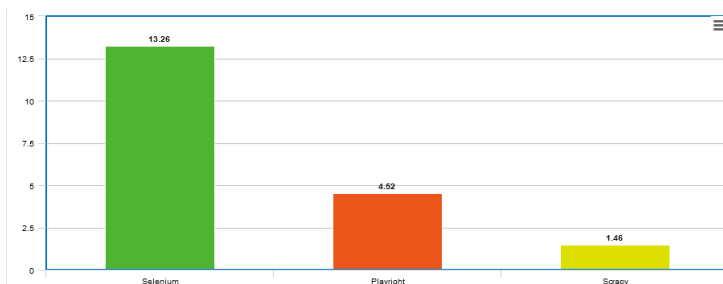
Ethical Issues:

By randomising tweets, we can obtain accurate data to train our ML model with little to no bias. We will still warn users due to the possible inaccuracies during training which could be from human error, too small strata or even social issues such as racism and sexism that could confuse the program's analysis.

What are the environmental and sustainability considerations of the development and deployment domains?

We decided to make our project a web application as this allows for a greater target market and greater scope. The sustainability of our project is much greater as it can be updated and maintained far easier than having to push an entire new version of a desktop program.

Machine learning becomes easier to maintain over time, as the ML model becomes larger and more accurate. Although this could result in large computing needs due to the complexity, resulting in more energy being used. Efficiency is very important for the project and the need to execute modular performance testing, to find the best frameworks and technology so we are best using the hardware. We have achieved this by testing different means of retrieving Twitter data and plotting it on a visual chart. This collected data shows the runtime for a single Twitter user query, and we have since improved on our performance further using asynchronous operations. Considering which platform to use when hosting the website can be a complex decision. There are a number of green or sustainable hosting companies out there that use sustainable resources to power their facilities which would be beneficial to reducing our programs' effect on the environment.



Identify one professional code of conduct and summarise how it applies to your development.

[3]The Association for Computing Machinery would be a good place to check for the norms of behaviour for sentiment analysis of tweets. The association holds a page which displays as well as maintains some of the key fundamental codes of ethics and professional conduct for all computing professionals. One good example of these codes is Avoiding Harm (1.2). This code highlights that Developers must be conscious of the potential societal impact of their work. The conduct applies to our topic as it could mean to ensure the technology is not used to manipulate public opinion, spread misinformation, or cause harm to individuals or communities. This could be done by manipulating users' results and spreading this data to the public causing fake information and calling them fact. This is why we will have to keep the user informed that the analysis isn't perfect data. This will depend on the machine learning progress and accuracy. If not informed this could cause harm to the users views.

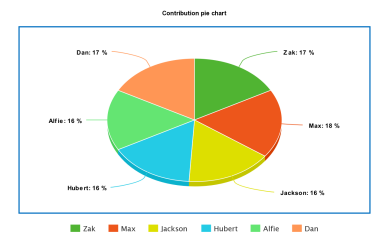
Group Work Reflection

When it comes to our group work, we think that we have done what we could together as a group and as individuals in this work. We managed to get together and discuss the issues that we had with the given task. The start of our project was the most challenging for our group as we had to reflect on which aspect of the project suited each person, as well as this, trying to find enough time within each of our time schedules was hard. However, after some time we managed to get all together and work on the project as a team, and even if some of the team members couldn't come and meet in person, we managed to connect missing members online, so that we were able to fix the problems that we encountered together.

After initially making contact we created groups on social media allowing us to share our opinions with additional research material alongside it; this made it easier for all of us to work at the given tasks. In the situation where one of us missed a meeting it was then simple to contact one another and move forward efficiently as a team. After all this time working, our group was successful at resolving all issues we faced and answering questions collaboratively.

We decided to split the contribution equally between all team members because all of us worked and shared opinions on everything, however, we were left with an additional 4% across 6 people so we decided to give the additional percents to the people whose contributions were more effectively used and also to those who put more time into it. Considering all the above we all think that our teamwork is on an advanced level.

The task turned out to be a success for our group. We all enjoyed working together on it and have seen good results so far. We did however encounter some issues on the way. For example, initially we had difficulties in communicating each of our strengths as individuals and also some technical issues such as links to joint work not functioning. On reflection, in the future we should introduce ourselves more assertively, outlining our strengths and weaknesses from the start. This will allow for more productive cooperation and make establishing roles between team members faster and easier.



Name	Zak	Max	Jackson	Hubert	Alfie	Dan
Percentage of contribution	17	18	16	16	16	17

CONTRIBUTION TABLE

Reference List

[1]

“550 million monthly active users.”

Livemint, (2023). livemint.com. *X has been losing daily active users since Musk takeover: CEO Linda Yaccarino*. 01 October. Available at:

<https://www.livemint.com/technology/tech-news/x-loses-daily-active-users-since-musk-takeover-ceo-linda-yaccarino-11696153731149.html>

[Accessed: 22 November 2023]

[2]

“amassing more than 10 million downloads worldwide as of December 2019...”

Jack R, (2021). gamerempire.net. *Plarium’s Aggressive Marketing of RAID: Shadow Legends Pays Off* [blog]. 24 April. Available at:

<https://gamerempire.net/plariums-aggressive-marketing-of-raid-shadow-legends-pays-off/>.

[accessed 18 November 2023].

[3]

<https://www.acm.org/code-of-ethics> - **The Association for Computing Machinery- ACM Code of Ethics and Professional Conduct**

[4]

<https://twitter.com/TOS> - **Twitter terms of service** - *Using the service, 4.(iii)*