41029 ENGINEERING RESEARCH PREPERATION
ASSESSMENT TASK 3 – Research Proposal

Assisted Posting for Small Business Owners on Instagram

Contents

Background 4
Research Question / Object of Design 6
Literature Review 6
Literature on the algorithm 6
Literature on small businesses 7
Existing Solutions & Alternatives 7
Methodology 8
Conceptual Design 8
Embodiment Design 8
Detailed Design 9
Project Management 11
High Level Scope 11
In Depth Analysis of Scope 11
Process and timeline 12
Milestones and Resources 13
Analysis 13
Design 14
Implementation 14
Testing and Project Finalisation 14
Uncertainties and Risks Control 14
Communication Management 16
Progress statement 18
Appendix 19
Appendix 1 – Model View Controller Architecture for assisted posting solution 19
Appendix 2 – Homepage & Parameter selection Mockup 19
Appendix 3 - Gantt Chart – Project timeline 21
Appendix 4 – Risk Management Register 22
Bibliography 23

Background

Social media and small businesses

Small businesses have undergone through a large shift as social media platforms have exponentially grown in recent years. This has forced marketers and business owners to shift their efforts from traditional means to online advertising. These platforms have performed extremely well as a channel to sell and expose products and services, as 83% of marketers use them for business-to-business advertising alone (Handley, et al. 2020). This exceeds search engine optimisation, another strategy for organic growth by almost 20% (**see Figure 1**).

Social media has become one of the largest platforms for businesses to communicate with their customers, the exposure has assisted greatly in growth and health of all types of businesses (Leibovitz, 2012). However, competition within social media advertising is skewed towards companies with a larger budget. This is due to the business model of these platforms being heavily reliant on paid advertising. Within Meta (formerly the Facebook company), paid advertisement revenue resulted in 98% of income during their Q2 quarterly report (Park, 2021). The revenue correlates to today's consumers that largely communicate with businesses through social media. A survey by New York university showed that 88% of consumers communicate with businesses through these platforms (Leibovitz, 2012).

Paid advertisement offers users various features and analytics such as targeted advertising, allowing marketers to directly expose their products to their intended consumers. These features make paid advertisement the most optimal way to reach a larger audience (Cooper, 2020). Users not paying for advertisement have access to some analytics however most of the small business do not pay for advertisement. Survey findings show 12% of businesses use paid advertisement compared to 52% exclusively (Anyan, 2016).

Organic vs Paid

Within this business to customer environment on social media, small business owners may face increasing competition in the face of monopolistic behaviour. Companies with large outside followings will likely find it easier to establish themselves by obtaining certified badges ("What is a verified badge on Instagram?", 2021). While Instagram states this does not endorse the brands with this privilege, consumers may look to verified badges as signs of approval. Larger businesses also have a larger budget to conduct paid advertisements within social media. While smaller businesses that need the capital to grow, are likely forced to use organic growth versus paid growth. This creates a paradox as small businesses need to grow to create capital however the environment requires capital to facilitate growth.

The algorithm favours paid users over free users by imposing this business model, as more likes and comments grant more exposure. Free posting only reaches 5.5% of the follower count a business has (Cooper, 2020). Embryonic businesses starting up cannot afford advertisement without capital which leaves them at the mercy of the algorithm. Fortunately, even within this unfavourable environment, some businesses have used organic growth strategies to establish a presence.

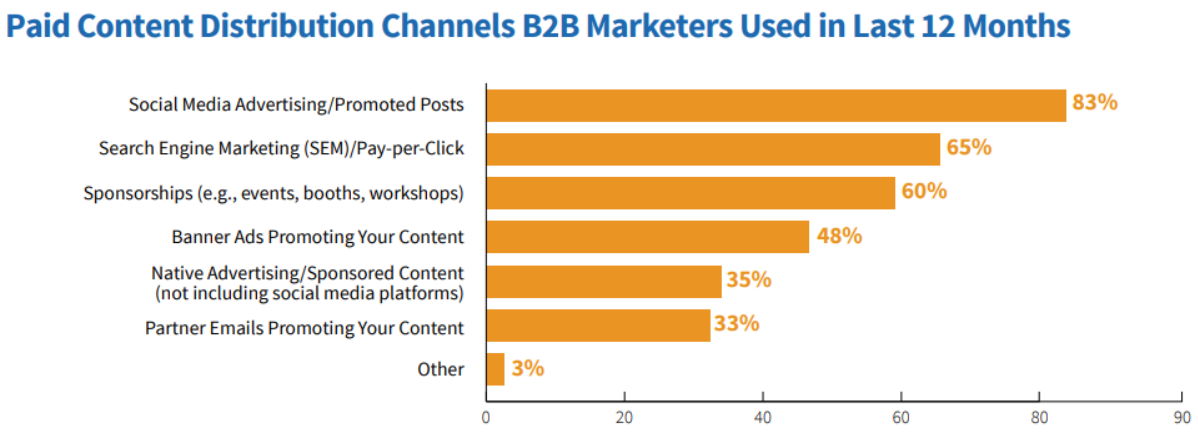


Figure 1 *- Social media advertising increased from 60% to 83% according to respondents.*

11th Annual Content marketing Survey: Content Marketing institute/MarketingProfs, July 2020

The economics

Small businesses account for 97% of all companies in Australia and represent 32% of its GDP ("Small Business Counts December 2020", 2020). Therefore, providing a healthy environment for these companies to grow without siphoning their resources is essential. Companies that sell products online has been growing, even more so throughout the covid-19 pandemic ("Online sales, May 2021 - Supplementary COVID-19 analysis", 2021). Social media startups such as artbyjazzd, a resin art business hosted on Instagram and forthesoul, a personalised frame business hosted on TikTok provide great examples of organically growth on social media due to their success without paid advertisement. Identifying successful concepts from these businesses will return useful data. This data can then be used to predict personalised suggestive measures to improve other social media accounts.

The proposal

This proposal will aim to assist small businesses in achieving optimal organic growth. By emulating already successful businesses within social media and analysing their data, a solution can be produced for new to market small businesses with necessary steps on how to improve. The solution will provide suggestive functionality that analyses data using machine learning and then processes those results for viewability. This in turn, will provide insight the user can incorporate within their own marketing strategy.

The solution will incorporate a stack of technologies appropriate to implement the solution intended for the end user. A virtually hosted application using a cloud database will provide global accessibility. The programs core will consist of an algorithm and library capable of retrieving, processing and returning data to produce results. This data can then be viewed with or without an extensive frontend via a terminal or UI. The proposal relies on the outcome of the effectiveness of its algorithm to produce insightful results, thereby being useful to small business owners.

Research Question / Object of Design

The objective of the project is to identify successful variables that lead to organic growth within Instagram. The following questions arise pertaining to this objective:

1. **What variables can be determined to improve engagement and retention?**
2. **What patterns correlate with successful free ad posting?**
3. **What is the most optimal method for gauging growth using organic posting methods?**

Answers to these questions lead to a system design that can input mass data and process them into results that are useful. The requirements that derive from these questions are as follows:

- Reliable data is needed for the input of the program.
- The algorithm must be pretrained with this data.
- Anomalies must be handled
- Processed results must be grouped into an insightful array of responses that the user can interpret.

Literature Review

Literature on the algorithm

According to the Elements of Statistical Learning by Stanford University, supervised learning uses a 'teacher' to modify outputs based on input, a process called 'learning by example' (Hastie et al., 2009). Supervised learning will be used as most of the data will vary and won't have any direct relation with each other. The algorithm will be given the key variables gauged for a successful Instagram account. These will include number of 'Likes' and 'Comments', 'Post Type', 'Captions' and Date variables. This provides a solution for research question **1**.

A paper showing the empirical comparison between supervised learning algorithms reached a conclusion with Boosted Decision Trees and Random Forests ranking top two in performance (Caruana et al., 2006). Further, a post detailed an elaborated comparison between the Boosted Decision Tree and Random Forest algorithms exclusively. With support of the authors of the original paper, it is shown that the BDT algorithm is much more time consuming than the Random Forest ("What is better: gradient-boosted trees, or a random forest?", 2016). This is because the tuning of the results to reduce variability error requires a longer process. Random Forest is also shown to perform better with larger data sets and increased levels in nodes. This algorithm will be able to find relationship patterns with low variance, which addresses research question **2**.

A recent project predicting the likelihood of growth in user engagement between Instagram accounts found Random Forest to be the most suitable machine learning model for Instagram (Martulandi, 2020). This reinforces the selection of this algorithm over other methods as Martulandi uses similar variables to predict his changes in engagement.

In conclusion, as this project has multiple variables, large datasets of Instagram accounts, both qualitative and quantitative measures and potential for node error, the Random Forest algorithm is determined to be the most appropriate to calculate insightful results. This addresses research question **3**.

Literature on small businesses

In 2011, online retailing represented $8.4 billion AUD domestic revenue and $2.4 billion international revenue ("Economic Structure and Performance of the Australian Retail Industry", 2011). This number was reportedly lagging most developed countries at the time. According to the International Trade Administration, Australian e-commerce has increased dramatically, predicted to reach 32.3 billion by 2024 with 89% coming from domestic revenue ("Australia ECommerce", 2020). By analysing this data, Australian online retail has remained predominantly domestic for the past decade.

The growth in the sector has increased 11% from 2019 to 2020 and is expected to grow almost 26% from 2020 to 2024 ("Economic Structure and Performance of the Australian Retail Industry", 2011). As small businesses account for the majority ("Small Business Counts December 2020", 2020), aiding small businesses online can boost economic productivity. This will likely have a favourable effect on international sales due to online accessibility. Australian small business owners can potentially target foreign customers to bring in more revenue. The algorithm might benefit in making the results more insightful if it targets domestic successful businesses for initial success. Once they have received considerable domestic income, the algorithm can start sourcing data from around the world.

Existing Solutions & Alternatives

There is currently no program that exists to aid small business owners or free users with assisted social media posting. There is also no solution that exists that uses complex algorithmic models or machine learning to produce insightful data for assisted posting. There does exist, however, solutions that provide automated posting. This automated posting allows users

Later.com

Later.com (formerly Latergramme) is a Canada based software company that provides analytics and scheduling for social media users on Facebook, Instagram, Pinterest, and Twitter ("World's Favourite Instagram Marketing Platform | Later", 2021).

The program provides a UI for users to schedule posts as well as view analytics. The software main advantage is convenience, which mainly aids already established businesses. It does not contain predictive features or focus on assisting free posting. This is because it charges a premium subscription fee to use all their features. This is counterproductive to small businesses with low capital. It cannot suggest or guide their posts with emphasis on free to post strategies.

Hootsuite

Hootsuite is like Later.com, with the addition of controlling multiple social media platforms simultaneously ("Social Media Marketing & Management Dashboard - Hootsuite", 2021). This program gives more customer feedback analytics. Its main advantage is also convenience. It is an unviable solution for small business owners due to the fees. It also does not offer any predictive measures.

Methodology

Conceptual Design

The program will comprise of two core functionalities, the frontend and the backend. The frontend will serve to enter input values prompted from the user – this will define the algorithm parameters moving forward. Finally, the frontend will display results. The backend will contain all the logic, integrated library packages and the algorithm. These packages will host the machine learning model, communication with the Instagram API and be able to update the frontend. The architecture will follow a Model-View-Controller pattern (**see Appendix 1**).

Model: Algorithm, Data Sourcing, Database, API logic & more.

Controller: Page routers, Handles input and output between frontend and backend.

View: What the end user sees. Displays prompts and results.

Embodiment Design

The system can be further broken down into sub-functions. Each of these functions is responsible for a key component within the backend. The model class will contain three main functions to source and prepare the data to be consumed by the algorithm. The Industry and Small business Owner class will serve as the controller, prompting the model to process results based on the users' parameters. These classes will act as an intermediary between the frontend and the backend. They inherit from the Account parent class. Finally, once the Model returns with processed results, the View class will handle how the data is presented to the end user (**see Figure 2**).

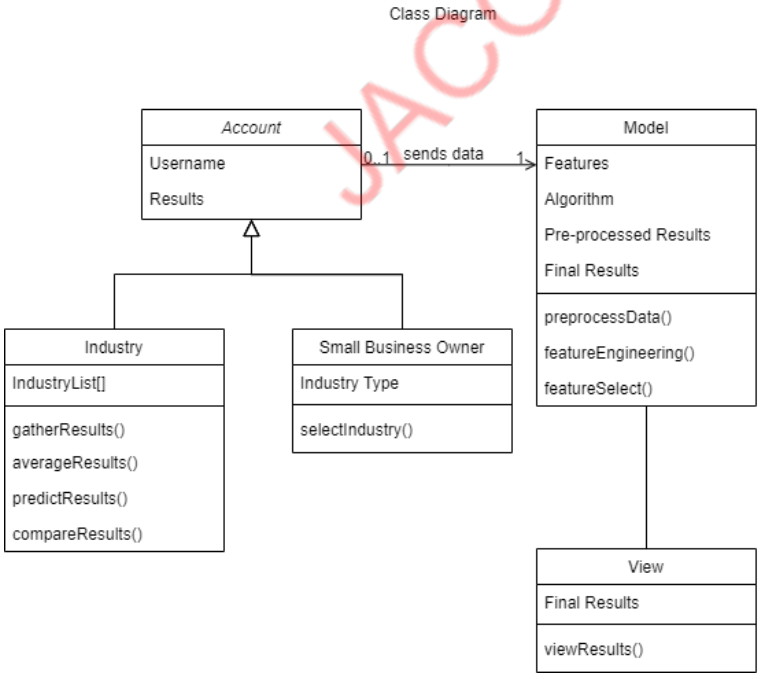


Figure 2 *– Class Diagram. Source - (Original, 2021)*

The process of the application starts with the user with initial parameters set such as setting industry type, entering username and basic details. This will be handled by a controller method that requests an account list to be processed. The account list will then prompt raw data to be retrieved from the

Instagram API. The data is then processed by the Model. The processed data is then retuned, formatted into interpretable results. Finally, it is viewable to the end user (**see Figure 3**)

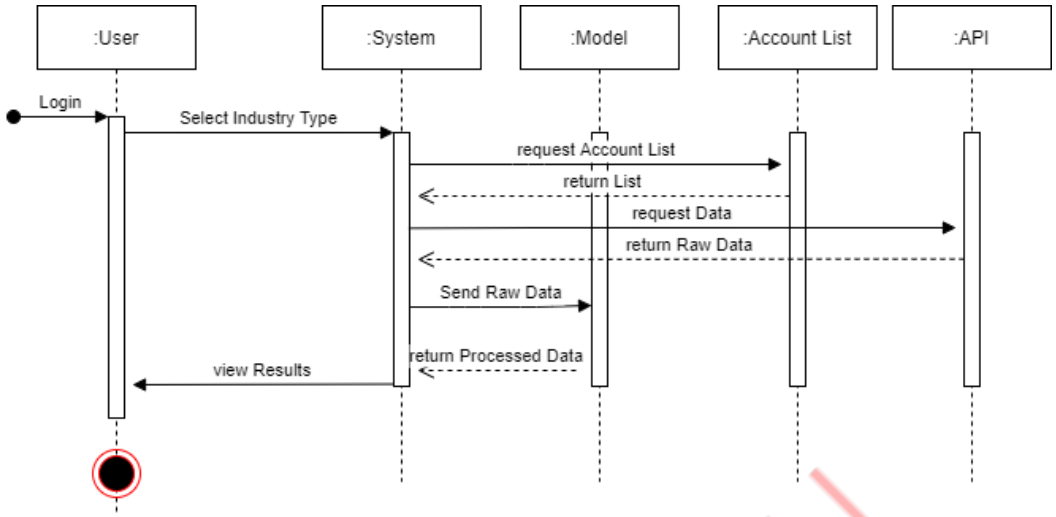


Figure 3 *– Sequence Diagram. Source - (Original, 2021)*

Detailed Design

Usability

The program will initialise by prompting the user for their basic credentials to analyse their Instagram account. This can include but is not limited to username and industry type (**see Appendix 2**). Data must be sourced from Instagram accounts with a certain number of likes, followers and comments. These variables will serve as the gauge for success within Instagram accounts. The range of each variable will be determined based on the user's discretion (**see Appendix 3**). The range will define what the algorithm sources as 'successful' accounts for pretraining.

The parameters will set, and the backend will run most of its functionality.

How the algorithm will work

Determining the right model is dependent on the data we can source from Instagram. Likes and comment count, while numerical, are not linear. Most data inputs from Instagram such as 'Captions' which includes a string (text) data type can only be filtered by Boolean conditions. This is because captions can have various meanings so the only way they can be measured, is either **a)** their length or **b)** whether they exist. Therefore, a binary classification algorithm is more appropriate over regression models as it addresses qualitative variables as well as qualitative variables.

This process applies to the project as we are giving it a set of conditions. If the input is defined by $\boldsymbol{y = f(x) + \varepsilon}$ and relationship is between $\boldsymbol{x}$ 'Followers' given $\boldsymbol{y}$ 'Likes'. The sourced data from successful Instagram accounts is input into multiple training sets $\boldsymbol{(x_i y_i) i = 1 \ldots N}$, with $\boldsymbol{N}$ being the number of accounts. The algorithm then produces an output $\boldsymbol{\hat{f}(x_i)}$ and modifies the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ for each node based on differences it perceives.

Decision Tree Algorithm

A decision tree is perfect to meet these criteria for a binary classification algorithm. The decision tree algorithm distributes data based on conditions into leaf's, with left leaf's representing true conditions while right leaf's represent false conditions (**see Figure 4**). However, while decision trees

can be simple to interpret, they are prone to data overfitting. In that, nodes can represent little data concentrations yet make conditional judgements for the entire system.

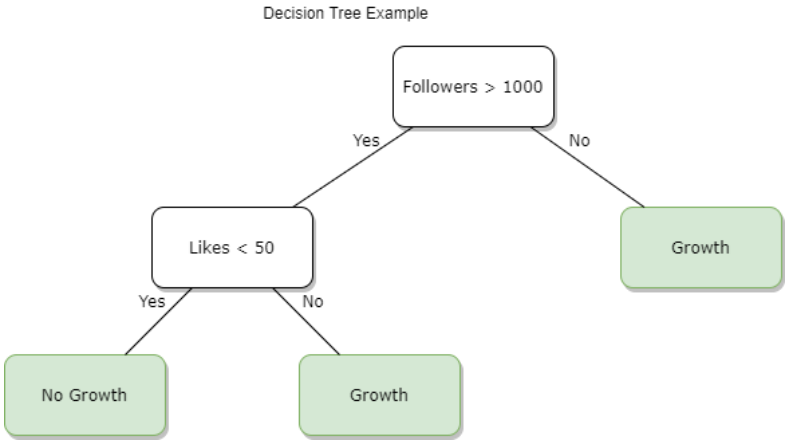


Figure 4 *– Decision Tree Example for assisted social media posting. Source - (Original, 2021)*

Variance Handling – Random Forest Algorithm

To introduce a method of reducing error, a more advanced algorithm must be used. A type of decision tree extension is the Random Forest algorithm. This method implements a solution by re-iterating over data sets and factoring in a measure of error such as statistical Entropy or the Gini Index. The Gini index will be used as it is simpler to implement. It calculates results within the leaf compared to their probability of happening (**see Figure 5**). This index is then used to structure the tree.

$$\text{Misclassification error:} \quad \frac{1}{N_m}\sum_{i\in R_m} I(y_i \neq k(m)) = 1 - \hat{p}_{mk(m)}.$$

$$\text{Gini index:} \quad \sum_{k\neq k'} \hat{p}_{mk}\hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk}(1-\hat{p}_{mk}).$$

$$\text{Cross-entropy or deviance:} \quad -\sum_{k=1}^{K} \hat{p}_{mk}\log\hat{p}_{mk}.$$

Figure 5 *– Decision Tree Example for assisted social media posting.*

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning Data

If a leaf has accumulated 5 True and 6 False, it is considered impure as no interpretable result can eb obtained. The Genie Impurity of this leaf is calculated by $\mathbf{1} - (\boldsymbol{P_T})^2 - (\boldsymbol{P_F})^2$ where $\boldsymbol{P_T}$ is the probability of True and $\boldsymbol{P_F}$ is the probability of False.

Eg: $\mathbf{1} - \left(\frac{5}{5+6}\right)^2 - \left(\frac{6}{5+6}\right)^2$

This is then repeated factoring in the leaf's counterpart (Eg No Growth vs Growth in Figure 4) to find the WGI (Weighted Gini Impurity).

The Random Forest algorithm will also compare various decision trees between variables and their WGI (Eg. Likes to Followers, Followers to Comments, Comments to Likes etc). Finally, the algorithm will rank features and relationships based on importance using the lowest WGI.

Technology and tools used:

- Python – Python will provide a fast and easy to use language. Python3 has many libraries available such as Pandas and SKLearn for data analytics and machine learning functionality.
- HTML - Html is the webpage standard markup language to display the program on a webpage.
- AWS – Amazon Web Services has a allows free with usage limitations. It will be more than enough to host the Algorithm Model and the Database.
- GitHub - The project will be uploaded and maintained on a version control software Github in order to keep a history of changes, make sure the project is saved on a cloud. This can prevent the project from being erased during the implementation process and if something goes wrong, it can be reverted to a previous version ("What is version control | Atlassian Git Tutorial", 2021). The website server can also be hosted via GitHub Pages seamlessly.

Project Management

High Level Scope

The project aims to identify methods associated with successful free posting. The aim is to develop a software that utilizes an algorithm to produce assistive results for small business owners. This includes but is not limited to defining the variables that need to be calculated for the user. Engineering a model that can retrieve mass amounts of targeted data from publicly available API's. Implementing an algorithm that can use this data to give the user insights on defined variables. Finally, these results should be returned programmatically or displayed on a frontend.

In Depth Analysis of Scope

Software Platform

The application will be a web-based application as it unlocks a range of modern and easy-to-use technologies such as frameworks, external API and has global accessibility. Frameworks will allow a quick and easy implementation of a frontend in order to focus on the much more complex algorithms implemented in the backend. External API is essential for obtaining data to compute for the algorithm from social media accounts. Web apps have the advantage over other platforms such as mobile apps because they can be used on virtually any device with web access (Atha, 2020) and don't have native limitations e.g.: IOS apps can only be used on iPhone.

Social Media Platform

The software will be targeted for Instagram. When considering how to benefit small business owners the most, the most used platform should be used. Instagram is considered the most effective tool for customer engagement and marketing a business (Miles, 2013). The platform was found to be popular for small business owners in Saudi Arabia, particularly with women (Alkhowaiter, 2016). This is due to Instagram being easy to use, inexpensive and accessible. In terms of sourcing successful users and their data, the API (Application Programming Interface) allows data to be collected more efficiently using GRAPHQL as opposed to REST ("Facebook for Developers", 2011). This means the program can collect specific data nodes rather than retrieving too much unnecessary data (Russell, 2019), which will make our program faster.

Frontend

The frontend of the application is what the user will see. First, the program must be useable to a non-technical user. This factor has to do with the UI (User Interface) of the program. As the frontend of the application is the least challenging part, it should take the least amount of time and shouldn't be overcomplicated. The scope of the frontend will be limited to basic functionality as it can always be expanded later. A framework will facilitate the maintainability and future expansion of the frontend (Pekarsky, 2021). Thus, framework technologies such as React, Vue or Angular will be considered. These technologies also have methods of seamless implementation to mobile such as React to React Native just for future product expansion into mobile.

Backend and Algorithm

The backend should be able to store data retrieved and processed by the algorithm though a database. When dealing with sourcing raw data, the results should be pre-processed so that it can be consumed by the model. This requires implementing Data cleansing, Feature engineering & Feature selection functionality. The most optimal machine learning algorithm should then be used, which will likely be the Random Forest algorithm (Martulandi, 2020) due to the qualitative nature of the results.

Process and timeline

The timeline and process of the application will follow the Software Development Life Cycle. This is structured method for developing software that consists of Planning, Analysis, Design, Implementation, Testing and Maintenance. Since Planning is completed before the project begins, it won't be considered within the timeline. Maintenance is also negligible as it does not relate to the immediate timeline of the project, only future sustainability. A complete timeline can be seen in **Appendix 3.**

Detailed descriptions of each process can be seen in Milestones below.

Milestones and Resources

Most resources will come from the documentation related to the technology used. Public forums such as Stack Overflow will be used for syntax assistance. Studying open-source code available online will also be used.

The major milestones outlined are the four highlighted components:

TASK	RESOURCE	CHANNEL
ANALYSIS		
Environment Research	Software versions, Technologies	Software Patches, Version lists, Updates
Define Variables	Planning	UML Diagrams software Project Research
Define Program Deliverables	Planning	Market Research, Project Research
Obtain API Permissions	Instagram Developer Form	https://developers.facebook.com/
DESIGN		
Select Tech STACK	Programming Languages, Database, Framework	https://docs.python.org/3/ https://developers.facebook.com/ https://docs.aws.amazon.com/
Design Architecture and Qualitative Features	Diagram	Draw.io (https://app.diagrams.net/) https://www.mural.co/ Microsoft Project
IMPLEMENTATION		
API Data Sourcing	Instagram Developer Documentation	https://developers.facebook.com/products/instagram/
Backend, Frontend	Documentation	Online
Data PreProcessing and Machine Learning Algorithm	Literature, Research, Code examples, Books	Online, Tutorials, UTS Library https://github.com/ https://stackoverflow.com/
TESTING	Testing methodology	Literature on Testing Methodology Books Research
Advice and Mentoring	Alan parr & CIC UTS	Meetings and Email Microsoft Teams Zoom

Analysis

Analysis of the project has four major milestones. Environmental research includes research of the latest tools and technology available to implement the software. Version control, latest updates and most suitable technologies should be considered. Defined variables and Program deliverables consist of the variables that should be returned by the program. These variables should be the most critical data points that can be predicted by the algorithm to influence a small business owners posting decisions. As Instagram will be the social media platform the application utilises, obtaining

developers permission API key will be required ("App Development - Documentation - Facebook for Developers", 2021). This report serves as most of the analysis milestone, it will be fully complete once Instagram API permissions have been granted.

Design

The design component of the project will focus on the foundation and skeleton of the application. The most appropriate technology stack will have to be chosen for suitability, efficiency, feasibility, and familiarity; Programming languages that are familiar and fast. Databases that are lightweight and easy to implement. These technologies should also synergise and be able to work within one integrated application. The architecture of the program will set the parameters, layout of technologies and necessary components within the application such as classes.

Qualitative features should also be known during week 4, these highlight what features, functionality and results should be in the program. This milestone is complete with this report.

Implementation

The implementation of the program will be the most time-consuming process and it is dependent on the design component. There are 6 major milestones: Data sourcing, Data pre-processing, Backend, Frontend, and the machine learning Algorithm. The functionality and deliverables will be coded during this time, taking about 4 weeks. The Integration milestone will make sure the entire project is brought together coherently and works sequentially. This milestone will be complete once the program has all functionality complete.

Testing and Project Finalisation

The testing process should consist of only a few days to put the application under various testing. Stress testing will ensure the program can run under copious conditions. Usability testing will ensure the application is simple and easy to use for non-technical users. Functionality testing, the most crucial part of the testing process will confirm all deliverables are functional and the algorithm is producing interpretable results. The project is then ready for handout. This milestone is complete once the project is tested thoroughly and is ready to be 'shipped'.

Uncertainties and Risks Control

Implementation Risks

There are functionality risks if the algorithm cannot provide insightful data, interpret data correctly or produce them in a human readable frame. The end goal is to give the small business owner ideas, strategy, and productive feedback for them to make their own creative decision on what to post. This should be based on results known that have proved successful for free posting businesses within their respective platforms. The main risk relies upon the accuracy of the results produced.

Project design and analysis will require the least amount of time and risk as it can easily be re-drawn and thought out. The worst that can happen is a few days delay, however most of the thought process should be completed by the analysis stage. The most risk and uncertainty lies within the implementation phase. According to the Systems Sciences Institute at IBM, the implementation phase is 6x more expensive than the design phase. This is because 35% of the defects are found within this phase. The deeper into the project, the more expensive and harder to fix it becomes (Burns, 2017).

To be able to source data isn't difficult, however, processing that into consumable data for the algorithm to process is. Machine learning is a complex algorithm to implement as algorithm correctness and debugging pose a problem (Enam, 2016). There is usually difficulty arising from the fact that the algorithm doesn't work well enough, or it doesn't work at all. The results returned must be interpretable and useful to the small business owner. Variables have to offer insights that aid the user into posting better content. There are many factors that contribute to the complexity (**see Figure 4)**

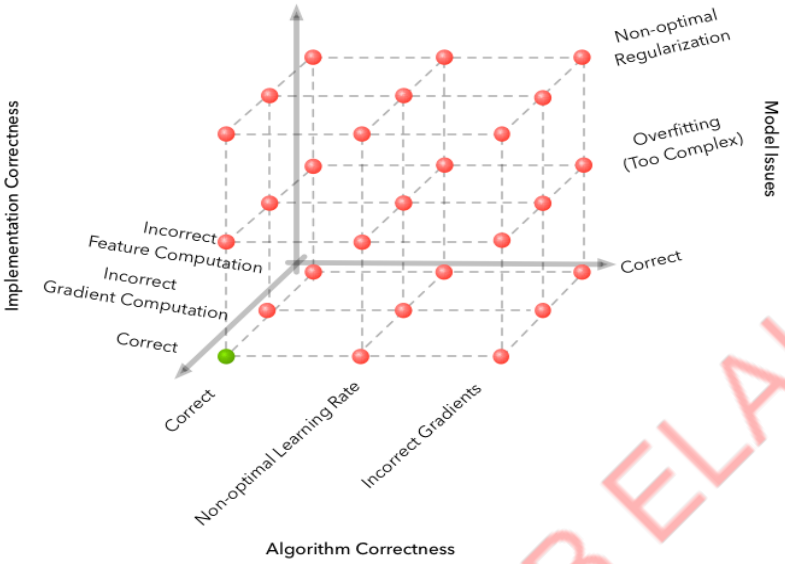


Figure 4 *– Machine Learning Algorithm Correctness (Source: Enam, 2016)*

If datasets are too large, the Random Forest algorithm chosen may require high loads of computational power.

Machine learning models have their disadvantages. They can be susceptible to error, consume large computational resources and they need a large amount of data to properly interpret and produce results ("Advantages and Disadvantages of Machine Learning Language", 2021). Coming up with the correct algorithm and a way to pretrain the model will be the focus to reducing any technical risks that may occur.

Mitigation

To mitigate this risk, similar projects that have proven to work should be researched. Research should be done into the different types of algorithms and their use cases. Ideally, instead of organically testing and figuring out which algorithm suites my problem the best, trying to adopt a similar projects algorithm can prove time saving and reduce the risk of it not working.

The program can be further simplified by reducing functionality such as reducing the HTML and web page to a terminal. This will save time with implementation.

Ethical Risks

With all the data being publicly available on these platforms, analysing the data of other entities could pose serious ethical concerns on plagiarism and data privacy. As social media is largely public, the only way to make posts inaccessible is to limit who can view your profile. This applies almost all large platforms such as Facebook & Instagram, where their API restricts pulling content

programmatically or viewing it through their UI without the user's permission ("Access Levels - Facebook for Developers", 2021). Thereby, every business shares their content to the public, assumingly most of them, will forcibly consent to their content being available for other businesses to copy or analyse.

There should be significant emphasis on only analysing successful strategies, characteristics, and variables of these publicly available companies to not plagiarise their content.

Social Risks

The solution is based on a directly proportionate ratio between effectiveness and user adoption. The risk is the more users adopt the same algorithm, the less effective it is. This is because the data being implemented dilates between the number of small business owners using it. If everyone is using the same predictive results, they may post similar content, thus getting lost in the algorithm once again. This can be offset by potentially analysing different results for each user.

Risk Importance

The primary risks are determined to be during the implementation phase, specifically getting the algorithm to produce satisfiable results. A full risk management register and risk priority table can be seen in **Appendix 4.**

Communication Management

Subject	Chanel	Purpose	Date	Time	General Notes
Meeting with Academic Supervisor (Alan Parr) 1	Teams Meeting	Discuss project milestones	21/02/22	13:00	Produce todo list and talk about start of project
Interview with Small Business Owner	Teams Meeting	Gauge User Feedback	25/02/22	Dependant on Availability	1 to many participants.
Progress Email with Academic Supervisor (Alan Parr)	Email	Report Project Progress	28/02/22	17:00	Obtain any feedback and advice
Email with CIC UTS 1	Email	Guidance on program Algorithm before Implementation Phase.	07/03/22	09:00	Discuss project and ask for tips and guidance

Meeting with Academic Supervisor (Alan Parr) 2	Teams Meeting	Discuss project milestones and progress		14/03/22	13:00	Talk about progress and design phase
Progress Email with Academic Supervisor (Alan Parr) 1	Email	Report Project Progress	21/03/22	17:00	Obtain any feedback and advice	
Meeting with Academic Supervisor (Alan Parr) 3	Teams Meeting	Discuss project milestones and progress	28/03/22	13:00	Talk about progress and implementation phase	
Progress Email with Academic Supervisor (Alan Parr) 2	Email	Obtain todo list for deliverables. Report Progress	04/04/22	17:00	Plan and advice before STUVAC	
Email with CIC UTS 2		Guidance on program Algorithm during and after implementation	18/04/22	09:00	Discuss project and ask for tips and guidance	
Meeting with Academic Supervisor (Alan Parr) 4	Teams Meeting	Discuss project milestones and progress	25/04/22	13:00	Talk about progress and testing phase.	
Progress Email with Academic Supervisor (Alan Parr) 3	Email	Discuss program finalisation.	02/05/22	17:00	Discuss the end of the project, what should be complete, how to finalise everything	
Interview with Small Business Owner	Teams Meeting	Gauge User Feedback on Final Product	09/05/22	Dependant on Availability	Adjust in frontend or note likes/dislikes	

Final Report submission	Teams Meeting	Get consent from supervisor	16/05/22	13:00	Get supervisor signature

Progress statement

Throughout the analysis phase, there have been minor changes to the overarching scope of the program. The objective design remains the same however, in order to adhere to time constraints, some functionality has been removed. These include the following:

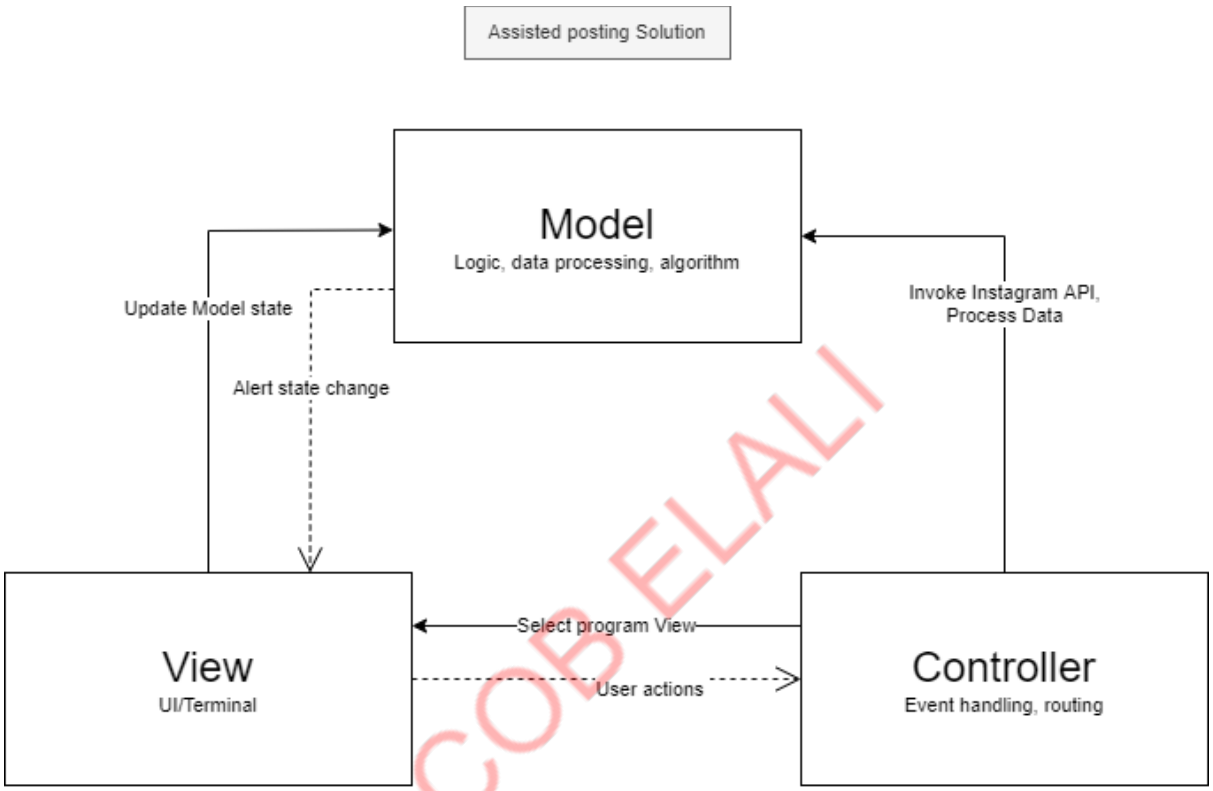
- **Results**: The initial design consisted of complex responses to the user which has been reduced to short responses. This will save development time and remove a major component of the program requiring artificial intelligence. These short responses will still be insightful and give the user more freedom in how they use the results for their business. The results will now be fact based and quantitative.
- **Frontend**: The need for an advanced user interface is also not necessarily given time constraints. As the main difficulty and usefulness of the program lies within the functionality, the user interface will be given a basic html structure. This rids of any complicated framework implementation such as react that is simply not needed. The structure of the frontend can be easily changed for further implementations of the program.
- **Tech Stack**: The technologies have mostly been determined with AWS to handle most cloud solutions. The cloud needs to have enough storage available for free to host the program, whether it be the database or hosting the server itself, configurations will be looked at closer to implementation phase.
- **Alternatives**: The program can be further simplified by reducing frontend to the terminal, removing the need to store data, changing data sourcing from the Instagram API to basic web scraping and locally hosting the program instead of the cloud. These options are available as a failsafe in order to ensure the project is completed by the end of the timeline. These changes will not affect the objective of design and desired core outcome. For the Random Forest algorithm, gauging differences in time can be complex as multiple decision trees need to be created for each Node. Thus, if this affects performance of implementation time significantly, it will be scrapped.

Overall, the project is on path to be implemented successfully. The project timeline and milestones have been carefully considered with constraints. The end of the project should produce a functional assisted posting solution for small business owners.

Appendix

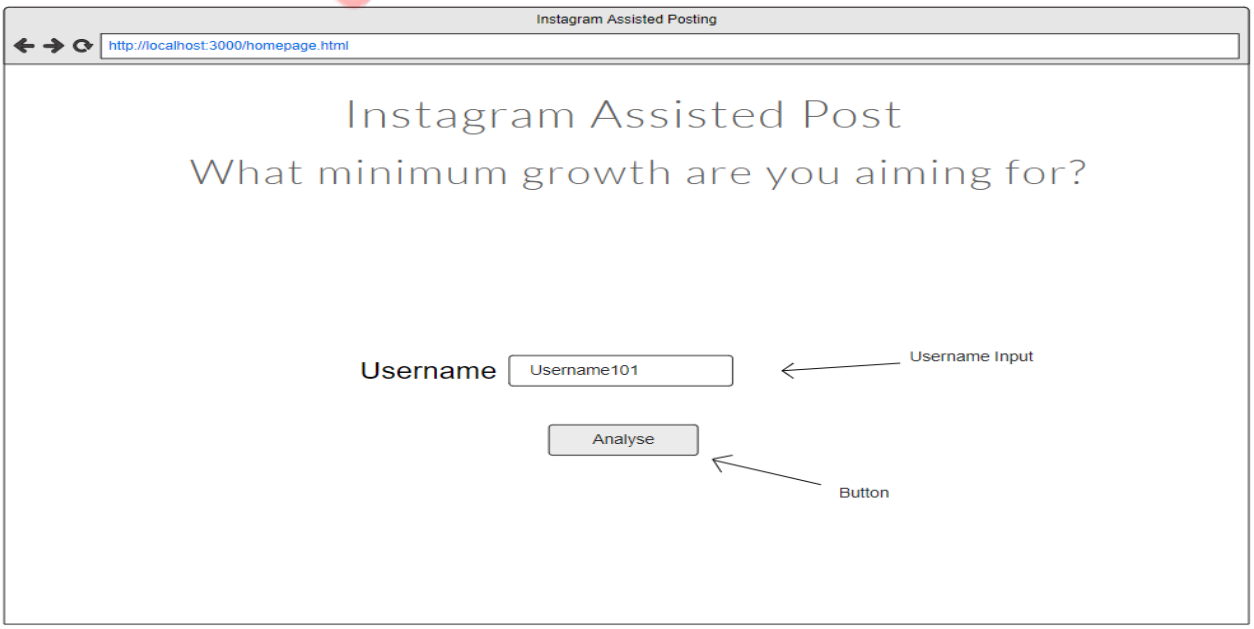
Appendix 1 – Model View Controller Architecture for assisted posting solution

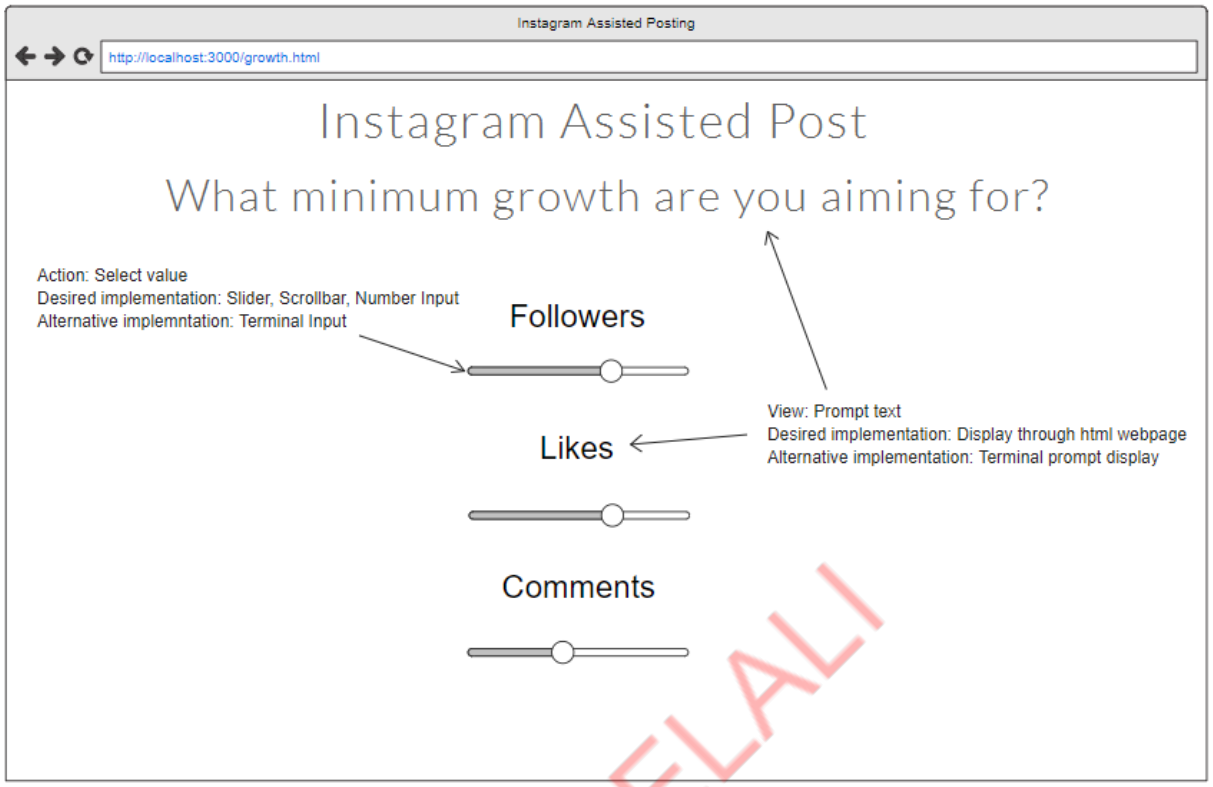
Source – (Original, 2021)

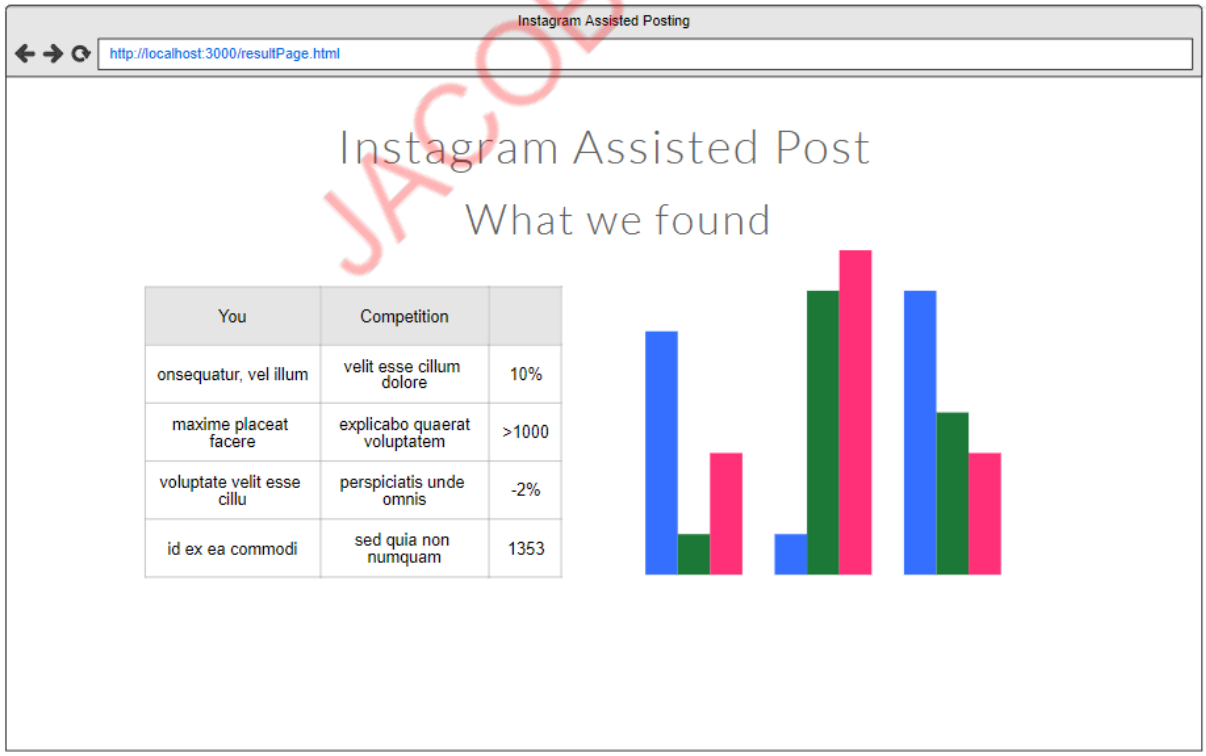


Appendix 2 – Homepage & Parameter selection Mockup

Source – (Original, 2021)



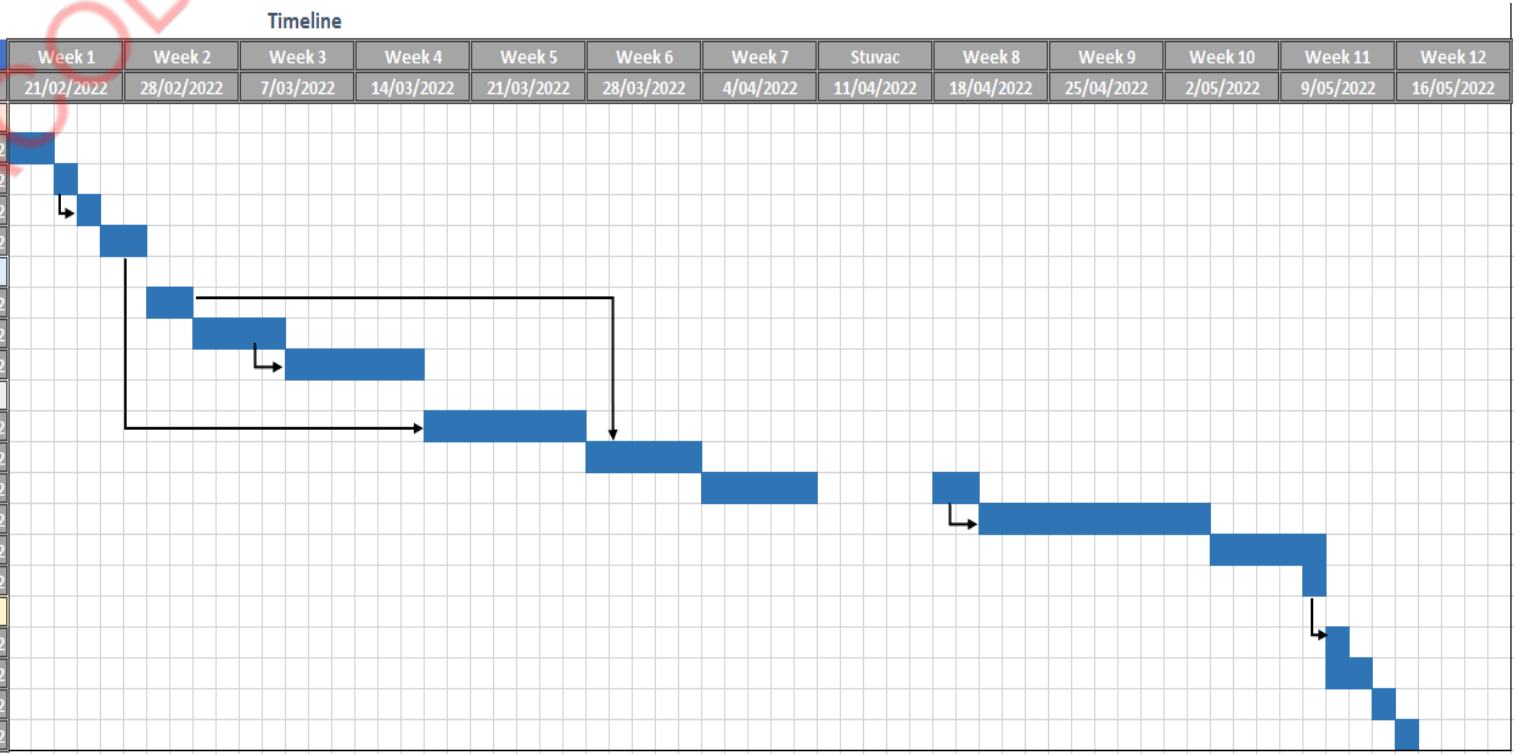




You	Competition	
onsequatur, vel illum	velit esse cillum dolore	10%
maxime placeat facere	explicabo quaerat voluptatem	>1000
voluptate velit esse cillu	perspiciatis unde omnis	-2%
id ex ea commodi	sed quia non numquam	1353

Appendix 3 - Gantt Chart – Project timeline

No.	Description	START	END
1	**ANALYSIS**		
1.a	Environment Research	21/02/2022	22/02/2022
1.b	Define Variables	23/02/2022	23/02/2022
1.c	Define Program Deliverables	24/02/2022	24/02/2022
1.d	Obtain API Permissions	25/02/2022	28/02/2022
2	**DESIGN**		
2.a	Select Tech STACK	1/03/2022	2/03/2022
2.b	Design Architecture	3/03/2022	6/03/2022
2.c	Design Qualitative Features	9/03/2022	16/03/2022
3	**IMPLEMENTATION**		
3.a	API Data Sourcing	17/03/2022	25/03/2022
3.b	Backend	28/03/2022	1/04/2022
3.c	Data PreProcessing Functionality	4/04/2022	19/04/2022
3.d	Algorithm	20/04/2022	3/05/2022
3.e	Frontend	4/05/2022	10/05/2022
3.f	Integration	10/05/2022	10/05/2022
4	**TESTING**		
4.a	Functionality testing	11/05/2022	11/05/2022
4.b	Usability testing	11/05/2022	12/05/2022
4.c	Stress testing	13/05/2022	13/05/2022
5	**Project Finalisation**	16/05/2022	16/05/2022



Appendix 4 – Risk Management Register

Risk likelihood: 1 = lowest probability, 5 = highest probability

Risk consequences: 1 = lowest risk, 5 = highest risk

Level of risk: Risk likelihood * Risk consequence.

Risk	The risk – what can happen and how.	Risk likeli-hood	Risk conce-quences	Risk Level	HA Priority	Mitigation Strategy
	Internal					
Implementation	Algorithm can't produce results	4	5	20	1	Simplify algorithm -> Use basic decision tree.
Implementation	Severe delays during integration and coding	3	5	15	2	Reduce functionality -> Host everything locally, reduce frontend to terminal, remove database
Ethical	Project might violate serious ethical concerns	2	3	6	4	Limit sourced data to remove any ethical concerns.
Social	Too many users cause reduce effectiveness	1	5	5	5	Revise algorithm
Testing	Bugs arising hindering final product	3	3	9	3	Use various testing methodology.

Bibliography

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition (2nd ed. 2009.). Springer New York. https://doi.org/10.1007/978-0-387-84858-7
- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Proceedings of the 23rd International Conference on Machine Learning, 161–168. https://doi.org/10.1145/1143844.1143865
- What is better: gradient-boosted trees, or a random forest?. (2016). [Blog]. Retrieved 27 October 2021, from http://fastml.com/what-is-better-gradient-boosted-trees-or-random-forest/.
- Pekarsky, M. (2020). *Does your web app need a front-end framework?*. Stack Overflow Blog. Retrieved 15 September 2021, from https://stackoverflow.blog/2020/02/03/is-it-time-for-a-front-end-framework/.
- Atha, H. (2020). *Web Application Development – The Definitive Guide for 2020*. Moveo Apps. Retrieved 15 September 2021, from https://www.moveoapps.com/blog/web-application-development-guide/.
- Miles, J. (2013). *Instagram power : build your brand and reach more customers with the power of pictures : Miles, Jason, 1970- : Free Download, Borrow, and Streaming : Internet Archive*. Retrieved 15 September 2021, from http://videoplus.vo.llnwd.net/o23/digitalsuccess/SUCCESS%20Book%20Summaries/2014%20March%20SBS/InstagramPower_Review.pdf.
- Russell, D. (2019). *GraphQL vs REST: What You Need to Know*. Rubrik. Retrieved 15 September 2021, from https://www.rubrik.com/blog/technology/19/11/graphql-vs-rest-apis.
- Alkhowaiter W. (2016) The Power of Instagram in Building Small Businesses. In: Dwivedi Y. et al. (eds) Social Media: The Good, the Bad, and the Ugly. I3E 2016. Lecture Notes in Computer Science, vol 9844. Springer, Cham. https://doi-org.ezproxy.lib.uts.edu.au/10.1007/978-3-319-45234-0_6
- *Facebook for Developers*. Developers.facebook.com. (2011). Retrieved 15 September 2021, from https://developers.facebook.com/blog/post/616/.
- Martulandi, A. (2020). *Predicting Instagram Influencers Engagement with Machine Learning in Python*. Medium. Retrieved 15 September 2021, from https://medium.datadriveninvestor.com/predicting-instagram-influencers-engagement-with-machine-learning-in-python-726c68bda29b.
- *App Development - Documentation - Facebook for Developers*. Developers.facebook.com. (2021). Retrieved 16 September 2021, from https://developers.facebook.com/docs/development/.
- *What is version control | Atlassian Git Tutorial*. Atlassian. (2021). Retrieved 16 September 2021, from https://www.atlassian.com/git/tutorials/what-is-version-control.
- Enam, S. (2016). *Zayd's Blog– Why is machine learning 'hard'?*. Ai.stanford.edu. Retrieved 16 September 2021, from https://ai.stanford.edu/~zayd/why-is-machine-learning-hard.html.
- Burns, E. (2017). *The cost of fixing bugs throughout the SDLC - Tech Monitor*. Tech Monitor. Retrieved 16 September 2021, from https://techmonitor.ai/techonology/software/cost-fixing-bugs-sdlc.
- Sydney resin artist Jasmine from https://www.instagram.com/artbyjazzd/
- Tiktok entrepreneur Maddison from https://www.tiktok.com/@forthesoul.au
- Park, M. (2021). *Facebook Reports Second Quarter 2021 Results*. https://investor.fb.com/. Retrieved 24 August 2021, from https://investor.fb.com/investor-news/press-release-details/2021/facebook-reports-second-quarter-2021-results/default.aspx.
- HANDLEY, A., MURTON BEETS, L., STAHL, S., ROSE, R., MOUTSOS, K., McPHILLIPS, C. and KALINOWSKI, J., 2020. *B2B CONTENT MARKETING*. [online] Available at: <https://contentmarketinginstitute.com/wp-content/uploads/2020/09/b2b-2021-research-final.pdf> [Accessed 23 August 2021].

- About.instagram.com. 2016. *See Posts You Care About First in Your Feed | Instagram Blog*. [online] Available at: <https://about.instagram.com/blog/announcements/see-posts-you-care-about-first-in-your-feed> [Accessed 24 August 2021].
- Newsroom.tiktok.com. 2020. *How TikTok recommends videos #ForYou*. [online] Available at: <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you> [Accessed 24 August 2021].
- Hainsworth, A. (2018). #pictureperfect. Real Estate Institute Of New South Wales, 69(2), 34-35. Retrieved 25 August 2021, from.
- Leibovitz, L. (2012). Social Media and Holiday Shopping: An Analysis of Consumer Attitudes. Offerpop Tumblr. Retrieved 23 August 2021, from https://offerpop-blog-blog.tumblr.com/post/35564895137/social-media-and-holiday-shopping-an-analysis-of.
- Cooper, P. (2020). Organic vs. Paid Social Media: How to Integrate Both into Your Strategy. Social Media Marketing & Management Dashboard. Retrieved 25 August 2021, from https://blog.hootsuite.com/organic-vs-paid-social-media/.
- Anyan, S. (2016). *Social Media Marketing Survey 2016: Paid Versus Organic Social Media*. Clutch.co. Retrieved 30 August 2021, from https://clutch.co/social-media/resources/social-media-marketing-survey-2016.
- *Online sales, May 2021 - Supplementary COVID-19 analysis*. Australian Bureau of Statistics. (2021). Retrieved 27 August 2021, from https://www.abs.gov.au/articles/online-sales-may-2021-supplementary-covid-19-analysis.
- *Small Business Counts December 2020*. Asbfeo.gov.au. (2020). Retrieved 29 August 2021, from https://www.asbfeo.gov.au/sites/default/files/ASBFEO%20Small%20Business%20Counts%20Dec%202020%20v2.pdf.
- *Access Levels - Facebook For Developers*. Developers.facebook.com. (2021). Retrieved 27 August 2021, from https://developers.facebook.com/docs/graph-api/overview/access-levels.
- Advantages and Disadvantages of Machine Learning Language. Data Flair. (2021). Retrieved 31 August 2021, from https://data-flair.training/blogs/advantages-and-disadvantages-of-machine-learning/.
- What is better: gradient-boosted trees, or a random forest?. (2016). [Blog]. Retrieved 27 October 2021, from http://fastml.com/what-is-better-gradient-boosted-trees-or-random-forest/.
- *What is a verified badge on Instagram?*. Help.instagram.com. (2021). Retrieved 19 October 2021, from https://help.instagram.com/733907830039577.
- *Economic Structure and Performance of the Australian Retail Industry*. Pc.gov.au. (2011). Retrieved 27 October 2021, from https://www.pc.gov.au/inquiries/completed/retail-industry/report.
- *Australia ECommerce*. International Trade Administration | Trade.gov. (2020). Retrieved 28 October 2021, from https://www.trade.gov/market-intelligence/australia-ecommerce#:~:text=Australia's%20eCommerce%20market%20is%20growing,billion%20in%202020%20and%20USD32.