**Danny Le**

**C964: Computer Science Capstone**

Table of Contents

[PART A: LETTER OF TRANSMITTAL 4](#_gjdgxs)

[Problem summary 5](#_30j0zll)

[Application benefits 5](#_1fob9te)

[Application description 6](#_3znysh7)

[Data description 6](#_2et92p0)

[Objective and hypothesis 6](#_tyjcwt)

[Methodology 7](#_3dy6vkm)

[Funding requirements 7](#_1t3h5sf)

[Stakeholders Impact 8](#_4d34og8)

[Ethical, legal considerations and precautions 8](#_2s8eyo1)

[PART B: PROJECT PROPOSAL PLAN 8](#_17dp8vu)

[PROJECT SUMMARY 8](#_3rdcrjn)

[Describe the problem 8](#_26in1rg)

[Summarize the client and their needs 9](#_lnxbz9)

[Provide descriptions of all deliverables 9](#_35nkun2)

[Provide a summary justifying how it benefit the client 9](#_1ksv4uv)

[DATA SUMMARY 10](#_44sinio)

[Provide the source of the raw data, how it will be collected and simulated 10](#_2jxsxqh)

[Describe how data will be processed and managed 10](#_z337ya)

[Justify why data meets the needs of the project 11](#_3j2qqm3)

[Address any ethical or legal concerns 12](#_1y810tw)

[IMPLEMENTATION 13](#_4i7ojhp)

[Describe an industry-standard methodology 13](#_2xcytpi)

[An outline of the project’s implementation plan 14](#_1ci93xb)

[Timeline 14](#_3whwml4)

[Evaluation Plan 14](#_2bn6wsx)

[Describe the verification methods 14](#_qsh70q)

[Describe the validation method upon completion 15](#_3as4poj)

[Resources and costs 16](#_1pxezwc)

[Itemize hardware and software costs 16](#_49x2ik5)

[Itemize estimated labor time and costs 16](#_2p2csry)

[Itemized estimated environment costs 17](#_147n2zr)

[PART C: APPLICATION 17](#_3o7alnk)

[PART D: POST-IMPLEMENTATION REPORT 17](#_23ckvvd)

[Solution summary 17](#_ihv636)

[Summarize the problem and solution 17](#_32hioqz)

[Describe how the application provides a solution to problems from A and B 17](#_1hmsyys)

[Data summary 18](#_41mghml)

[Provide the source of the raw data, how the data was collected and simulated 18](#_2grqrue)

[Describe how data was processed and managed throughout the app development life cycle 18](#_vx1227)

[Machine learning 18](#_3fwokq0)

[Identify the method and what it does (the “what”) 18](#_1v1yuxt)

[Describe how the method was developed (the “how”) 19](#_4f1mdlm)

[Justify the selection and development of the method (the “why”) 19](#_2u6wntf)

[Validation 19](#_19c6y18)

[An appropriate validation method 19](#_3tbugp1)

[Result of the validation method or a future plan to obtain result 20](#_28h4qwu)

[Visualizations 20](#_nmf14n)

[User guide 20](#_37m2jsg)

[E. References 20](#_1mrcu09)

# PART A: LETTER OF TRANSMITTAL

Dear Samsung,  
  
I am excited to introduce an innovative project proposal – “Reddit Sentimeter” a venture that guarantees to have data analytical power used in improving our market strategies. This innovative instrument was developed to capture the tremendous potential of social media data that is rarely acknowledged and utilized, referring primarily to Reddit as a source for consumer perception concerning Samsung and its products.

**Understanding the need:**

In our current landscape, a significant gap exists in our market analysis methodologies: the under usage of the unstructured data that is presented by social media. A platform like Reddit which has endless number of consumer opinions and trends remains essentially un-leveraged. This limitation limits the understanding that we have related to how people treat our brand and products in the public, therefore making it impossible for us to make data-driven strategic decisions.

**Introducing Reddit Sentimeter:**

The Reddit Sentimeter is a modern web application created specifically for the real-time analysis of sentiment in the discussions about Samsung, which are held on different subreddits. Using state-of-the-art machine learning algorithms, this tool will analyze the mood trends and give suggestions on how best to approach the markets – thereby changing out view of market analysis.

**Benefits to our organization:**

Enhanced Market Insight: Obtaining real time snippets of public feeling level permits lateral shift in approaches in our advertising plans.  
Targeted Advertising: Using this sentiment analysis information for better advertisement on Reddit.  
Competitive Analysis: Providing a full perception of our position relative to competitors.

**Project outline:**

Data Source: Reddit analysis including relevant subreddits on Samsung and other related subjects.  
Methodology: Use of sentiment analysis algorithms to determine if sentiments are positive, negative, or neutral.  
Objective: To enhance our comprehension of the same perceived by consumers and see established market patterns.

**Costs and timeline:**

Estimated Hours: Summing to 200 hours illustrating, groundworks 80, expanding it 80 and preparations 50.  
Completion Date: Aiming for March 1st, 2024.  
Budgetary Needs: Mainly Development resources and the cost related to day-to-day operation.  
Impact on Stakeholders:  
This initiative will provide our stakeholders with a deeper understanding of consumer tendencies and wants, which in turn supports the development of better-informed decisions and strategic plans.

**Ethical and Legal Compliance:**

In doing so, we will maintain strict adherence to ethical requirements and remain compliant with Reddit’s API usage policies; this means that it will be possible to ensure data privacy and adopt the principles of responsible data management.

**My Professional Credentials:**

Considering that I have the much-needed experience in programming using Python and a strong foundation in machine learning, I am best placed to head such an endeavor. Therefore, being an expert in data analytics and a professional who can work out the digital marketing strategies used for this case; I have all reasons to help make it successfully happen.

The ability of Reddit Sentimeter to transform our market analysis and decision-making procedures is an exciting prospect. I look forward to being given an opportunity to review this proposal in-depth and be able make quantifiable contribution’s that will ensure our organization continues registering growth, success.  
  
Thank you in advance for your reflection on this proposal that tries to be innovative. For more information or further discussion on the matter, feel free to get in touch with me.  
  
Sincerely,

Danny Le

Programmer - TripAdvisor

## Problem summary

The main challenge which will be covered in the design of this Reddit Sentimeter project is lack of systematic data-driven approach to collecting consumer sentiment on social media, with a particular focus on Reddit regarding Samsung and its products. However, at present, our organization does not have the capacity to analyze and make sense of the huge amounts of unstructured data available in Redditt; this is a major deficiency in our market analysis approach. This means that we do not have access to vital information, which will inevitably prevent us from obtaining a wise picture of the reaction of one very powerful online community and in real time. Lack of this knowledge means that our marketing, product development and competitive analysis work could easily be leaving out the most important current consumer trends in sentiment, which is an easy way to make less optimum decisions. Overall, the issue lies in not having an effective tool or approach to effectively track, analyze and understand public perspectives about our brand or products on Reddit hence undermining the ability to capitalize social media data for business strategic.

## Application benefits

The main contribution of the Reddit Sentimeter application relates to organizational strategy as it offers real-time data and sentiment analysis of publicly available opinion on Samsung shared through Reddit. This helps the company to remain abreast of consumer insights and thereby informed decisions on marketing, product development, or even customer support. Understanding the preferences and criticisms of consumers is very important because it helps to modify its marketing strategies with the aim targeting advertisements placed on Reddit more effectively. The option of competitive analysis is available with this tool, which shows what your strengths are and what weaknesses can be used by competitors. Consumer sentiment provides such insights that are very invaluable in product development, improvement and innovation. Besides, it also promotes an active strategy of reputation management and develops a data-based style of a company’s employees. In summary though, the Reddit Sentimeter is a robust tool as it provides actionable insights which will guide in better decision-making processes, improved marketing approaches, stronger product development and competitive placement.

## Application description

Reddit Sentimeter is cutting-edge web-based application that represents a new breath in the area of sentiment analysis used particular for Samsung related discussions on Reddit. It uses the API of Reddit for gathering data, more specifically, posts and comments among subreddits discussing technology and Samsung products. The app applies sophisticated natural language processing NLP methodologies and machine learning algorithms such as k-means clustering to sort sentiments as either positive, negative or neutral. Its UI includes a live sentiment trend dashboard, visualizations by Charts.js and reports that can be customized. Data filtering, trend analysis, and competitive sentiment comparison are all offered by the application in order to increase the organization’s support for identifying and managing consumer attitudes. It is mostly built in Python and with HTML, CSS, JavaScript for the Web front end refrains by design helps cross-platform compatibility along with data privacy and Security compliance architecture. In general, the Reddit Sentimeter converts raw social media data into actionable structured analytics that will guide an organization’s operation in the digital spectrum.

## Data description

The Reddit Sentimeter application will use data from Reddit, a platform that is popular for having a vast and broad user base, focusing on subreddits concerning technology, consumer electronics and discussions on Samsung and its rivals. The materials are essentially textual bodies of posts and comments, as well as metadata such as titles, timestamps, user agents also all while providing anonymity and in accordance with privacy standards. Considering the large volume of data that Reddit produces, as well as the immediacy by which it does so, various types and sentiments will be processed by the application due to inclusive list of topics and from pleasing to deceiving facts. The information, although will quality and the degree of reliability will differ, however in importance for filling sentiment ruling as well as trend omics strategies it is not Approach will be useful alone to achieve full understanding approaching popular support. This is a vast and comprehensive pool of data to obtain valuable for making strategic decisions within each firm, adhering to all rights issues as well as privacy.

## Objective and hypothesis

The Reddit Sentimeter project is a sheer real-time sentiment analysis tool that gathers the development of an in-depth inference system targeting public sentiments on Samsung and its products, vigilantly deliberated on Reddit. This disruptive solution provides will convert raw social media buzz into high-value inputs and make the entity use informed decisions that are backed by structured information. Its key objectives are multifaceted: First, it will go on to monitor and analyze the sentiment of Samsung users in Reddit with regard to consume perception and brand trial resulting in systematic customer insight. Secondly, it is supposed to offer comprehensive knowledge of consumer behaviour, preferences, and perceptions that are key in developing products as per the needs besides orienting its market strategy. Third, the tool will help in competitive analysis by comparing sentiment about Samsung and that of its competitors thus giving a clear insight to market standing. Finally, it will aid the optimization of advertisements on Reddit in terms of channel and messaging strategies by identifying the most responsive channels as well as messaging strategies based on user sentiment and levels of engagement.

The strategic value of the Reddit Sentimeter project depends on several hypotheses, which are critical to its development. First, it assumes that the positive reaction of Reddit users to Samsung products can be treated as an accurate signal for increased consumer interest and potential sales growth. Second, the article postulates that real-time sentiment analysis will support faster and more efficient changes in marketing strategies and product placement. Third, the project presupposes that strategic advertising to subreddits with a favorable attitude towards Samsung will contribute to ROIs. Moreover, it implies that if Samsung manages to capture the sentiment of its competitors, then using this knowledge will provide the organizations with a competitive advantage by showing what should be eliminated and where the strengths lie. Lastly, it assumes that high levels of user engagement such as comments and upvotes on discussions about Samsung serve as an indication that Samsung has a strong brand presence there and is also viewed positively. This will not only validate the perceptions that sentiment analysis is appropriate in aiding a business strategy but will also create an avenue for operators to learn and adapt steadily within the competitive digital market.

## Methodology

The process of the Reddit Sentimeter project is based on a well-structured methodology The choice has been made according to critical phases that determine design, development, and implementation. It has the following stages planning and project analysis which involves outlining the goals, identification of relevant subreddits, and setting key performance indicators. The stakeholder consultation is important in assessing what the needs and expectations of the organization are. Integration of the Reddit’s API and carefulness in data cleanliness ensure a meticulous process of data collection and preprocessing. Sentiment analysis model creation relies on the choice of correct machine learning algorithms and profound training and testing. Application development involves the back-end implementation and creation of a user-friendly front-end interface. It includes model validation and iterations where the evaluation of the model is conducted, user testing is done to finalize this application through refining. It is deployment and then continuous monitoring that drives it, while data analysis and reporting delivers on tactical possibilities. The data privacy and ethical use are followed through maintenance of the ethical considerations and compliance. This method based on the agile approach helps to keep the Reddit Sentimeter flexible at all times and ensure that it is aligned with strategic objectives of the organization.

## Funding requirements

The costs of implementing the Reddit Sentimeter project have been discussed in terms of a detailed analysis. It should be mentioned that these numbers are only estimates and can change depending on various conditions, the state of the market, or available resources. Development resources budget needed is costing for hiring Data Scientists and Machine Learning Experts – $80k-120kyear, Software Developers -$60k-100k per developer peryear and Project Management-$80$k to 120k Moreover, infrastructure and hosting costs; software and tools based services costs; data acquisition fees and API charges; the budget for testing & quality assurance; documentation & training funds, legal & compliance fund provisioned in estimation financial outline of project by reserved money as a contingency due to its requirement urgency prone, marketing and adoption expenses also maintenance& upgradation aspects are so important factor basis which reserve should made. These figures serve as guiding constructs that can cater to the specific requirements and changing states of the project.

## Stakeholders Impact

The Reddit Sentimeter project has immense potential in affecting an array of stakeholders within the organizational settings, with each set to benefit differently. It would be beneficial to senior leadership and decision-makers for them to take advantage of real-time consumer-sentiment intelligence actionable by facilitating data-oriented strategic decisions. Marketing departments will have a powerful weapon to perfect campaigns, increase accuracy of target and return on investments. SAS has completed the third leg of this journey quite successfully as its product development teams now stand to receive a priceless source of feedback for customer-centric design even as the competitive intelligence teams gain competitive advantage through sentiment benchmarking. Data analysts and scientists will sharpen their skills, IT and development teams will improve their knowledge while legal and compliance teams protect the ethical data use. Everybody is happy. Employees, end users, customers, investors and even the Reddit community itself will benefit indirectly in terms of quality decision making and customer-oriented product development which will enable a stronger organization that beats free competition.

## Ethical, legal considerations and precautions

When the data required is that which is sensitive in nature especially sentiment data drawn from a public platform such as Reddit, it should be of utmost necessity to consider ethical and legal issues. A number of safeguards should be strictly adhered to in order to guarantee responsible and policeable data handling. These measures include user privacy and anonymity, no collection of personally identifiable information while following data privacy legislations as well as Reddit API guidelines Further, it is required to get the user’s consent in the necessary form as a manifestation of informed agreement and to comply with Reddit transparency and consent requirements. To ensure that the patients require protection of data integrity and confidentiality in compliance with laws on the safeguards so, robust data security controls are necessary. It is equally important to reduce the biases in sentiment analysis, practice open communication with stakeholders and comply with intellectual property rights. For the Reddit Sentimeter project, responsible data use, regular monitoring and accountability mechanisms, as well as readiness to report to external organizations on ethical and legal issues that arise are crucial. These measures together with it or User Precautions serve to maintain the integrity of information, protect user investments plus rights and ensure regulatory compliance thus promoting trust on the sides in any interaction.

# PART B: PROJECT PROPOSAL PLAN

## PROJECT SUMMARY

## Describe the problem

In such a quickly changing technology industry, knowing the sentiments of the consumer will help to understand and make decisions. This critical need is addressed in our project Reddit Sentimeter. We will create an intelligent sentiment analysis solution adapted to the special needs of your organization. Perceives that in the digital generation public opinions and sentiments will be just as important if a brand is to thrive of doom a brand to failure. On the other hand, if sentiment analysis was to be done manually across huge social media platforms processing will take longer and lack efficiency. As a future solution, our company will create an integrated system to track the sentiment on Reddit and similar platforms, where all valuable discussions regarding our products and competitors are going to happen. This spirit of proactivity will be critical to ensuring we remain nimble and ready to meet the changes that characterize or even further dictate, the digital sentiment analysis terrain.

## Summarize the client and their needs

Your organization is a technology leader that, in the future, look for a solution to monitor sentiment data via Reddit effectively analyze and leverage from it. The middle management in your organization will require a tool that is user-friendly and easily accessible, this means you would not need to deploy any complex solutions with the need of using professionals for accommodating it. This proposed future solution will be crafted for the provision of actionable insights that will drive your marketing approaches, direct product creation and advancement endeavors, as well as improve your competitive positioning in this market segment.

## Provide descriptions of all deliverables

In the long run, the Reddit Sentimeter Application will become a simple web-based portal that allows users to extract, process and present sentiment data drawn from relevant subreddits on Reddit. An extensive user’s manual will be offered to make its use easy thus ensuring that your middle management uses it optimally and comfortably. In addition to this, support and maintenance services shall be available to ensure that there are no potential technical problems contributing in achieving current data accuracy and reliability.

## Provide a summary justifying how it benefit the client

Real-Time Insights for Informed Decision-Making: The application will also offer the vital function of Reddit discussions sentiments scored in real-time. This will help the middle-level management to quickly respond to the changing needs in the market and make timely information-based decisions. The immediate access to sentiment trends also guarantees that strategies and actions will always be consistent with a contemporary perception of consumers.

Enhanced Product Development: The application will help to get good responses from the discussions in Reddit which will enrich our product development process. With adequate analysis of user opinions and feedback, our client will be able to identify regions in which the product can be improved or modified. Such a strategy is proactive and guarantees that goods are closely tuned with consumer inclinations, thus generating high customer satisfaction accompanied by unwavering loyalty.

Competitive Advantage: The application will provide the client with the power to do a sentiment analysis comparative against competitors. This comparative analysis of sentiment patterns vis-à-vis competition in the industry will provide the client with a comprehensive understanding of relative strengths and weaknesses concerning product offerings and marketing approaches. Using this knowledge, our client is able to take advantage of opportunities and address the areas of weakness that their competitors may be suffering from very skillfully.

Cost-Efficiency: The application has been designed to automate the process of sentiment analysis, which is a clear divergence from conventional manual means that are labor intensive and susceptible to numerous errors. This automation does not only minimize the cost of labor but also greatly improves accuracy. In the process, it leads to a maximum operational efficiency of a client where resources are effectively optimized.

## DATA SUMMARY

## Provide the source of the raw data, how it will be collected and simulated

The Reddit platform has presented an opportunity to collect primary data for the Reddit Sentimeter project, as this undertaking will benefit from information collected via Reddit’s public API and its vast population of users who produce sentiment pertaining to our client company and solutions. This will be a systematic and well-structured process of data collection, which in turn ensures the highest standard in terms of quality and security aspect of data at all levels of output. This will include API calls from Reddit’s public API to retrieve posts, comments, and discussions available on public level in relevant subreddits. The choice of subreddits will be narrowed down to those relevant to our client needs and target audience so that we can obtain precise data. The sentiment analysis is automatically schedule at extract date where the new day information will be consistently and accurately analyzed. The data will then be passed into a secure, structured database from which preprocessing can begin in order to prepare the data for sentiment analysis and application of sophisticated machine learning NL processor techniques that will determine sentiment. Lastly, visualization tools will be used in the project to appropriate communicate insights and trends. The entire exercise shall be strictly guided by ethical and legal factors that would including adhering to the terms of service as well as use polices on behalf of Reddit, so as to ensure responsible and legally compliant data handling.

## Describe how data will be processed and managed

Design phase: In the early stages of planning in the Reddit Sentimeter project, a lot of careful consideration is applied to designing and key data components. Initially, the Data Schema Design phase is characterized by careful structuring that establishes the design of a guide for storing accumulated data. This well-designed schema specifies the exact format and organization of the database for which the data will be housed. At the same time, such an important part is played by the Algorithm Design stage, which identifies essential processes. These pipelines are not limited to just handling data also they include sentiment analysis algorithms writing. In this stage, the complicated procedure of turning raw data into sentiment scores that are useful is descriptively explained. In addition, the Data Collection Plan which is a fulfilling aspect of this early stage becomes tangible as an overarching approach for data acquisition on Reddit. In this approach, important choices include the identification of appropriate subreddits and the creation of timing for effective extraction. All in all, these parts form a basis for proper management and use of data within the Reddit Sentimeter application.

Development phase: As the Reddit Sentimeter continues into the Development Phase, a number of critical processes related to data come into play. Primarily, the Specialized Script stage that is characterized by specialized scripts used to interface with Reddit’s API. Such scripts, explicit codes of construction, are responsible for the recovery of publicly available content from specific subreddits in the form of posts, comments and discussions. At the same time, Data Preprocessing is in the stage where collected data are polished and prepared. There is deliberate sifting of the available information to do away with duplications, inconsistencies and irrelevant details leaving behind a clean set of data free from unnecessary things. In addition, the data is strictly standardized to ensure that there is uniformity and consistency. Also, the development process here includes the Sentiment Analysis phase where machine learning models are developed and horizontal integration is ensured. These models are important in calculating sentiment scores which are based on data that has been carefully preprocessed. Finally, with Cleaned and preprocessed data stored in the database of the application itself guarantees not only the preservation but also smooth retrieval and evaluating that is needed. In combination, these are the main data processes that define Reddit Sentimeter and make it able to yield information on sentiment through analysis.

Maintenance phase: When the Reddit Sentimeter project goes into Maintenance Phase, a number of critical data engineering operations that contribute to the effectiveness and robustness of the application are continued. The persistent Data Retrieval constitutes an essential feature of this stage since the application is set up to perform recurrent data retrieval from the Reddit’s API. This guarantees that sentiment analysis will be performed on a current and actively changing portion of discourse, thus allowing the application to provide insights in real time. In addition, the Scheduled Updates are an ordinary part of keeping it going smooth-working. This includes bug fixes, performance improvements and upgrades with the sentiment analysis model, all of which ensure that the application is up to date and functioning accordingly. To ensure the accuracy of the data, Data Monitoring is being given top priority. All inconsistencies, outliers or issues that can be qualified as related to the quality of data are eliminated in a timely manner for the results of sentiment analysis not to be compromised. In addition, User Access Management is closely monitored and managed so as to ensure the integrity and confidentiality of data. Only authorized users have access to the application data, so only they can view and analyze sentiment data. Some of these persistent data processes are available to conduct reliable sentiment and valuable insights over a long period in the Reddit Sentimeter.

Other considerations: Data Privacy and Security form the most essential aspects of the Reddit Sentimeter project. The development life cycle of the application also includes very careful and thorough incorporation and implementation of security measures throughout every phase. These measures act as a strong wall, protecting privacy of the users and preventing any possible unauthorized access to sensitive data. In addition, Scalability is a critical design element of the application. It is future-proofed so that as the project grows and gradually attracts an increasing user base, it can handle increasing data volumes without a hitch. The concern for user’s empowerment is carried into User Training where users are given thorough instruction and documentation. This provides them with the skills and information necessary to operate the application efficiently, i.e. Or they can learn how to feed data into it and receive interpretation for providing sentiment analysis results. Data is highly regarded in this entire organization as a valuable resource with ethical and legal issues being the guiding principles that steer each procedure of data operations. As a result, this provides for accountable and secure data processing and storage in compliance with the state standards at all stages of the Reddit Sentimeter lifecycle.

## Justify why data meets the needs of the project

The Reddit Sentimeter project has several important characteristics, which describes the particular needs that make the instruments contained in Reddit’s public API eminently suitable for this kind of research. First, of all the data is absolutely relevant to the project’s goals of observing sentiment, attitudes and conversations related to the client organization as well its products. As such, Reddit is a great repository of user-generated content about these subjects and provides access to vast amounts of information. Secondly, the real-time phenomenon characterizing Reddit discussions equips the project with access to immediate information on consumer attitudes. This dimension of real time is critical because they support timely decision making and ability to quickly respond to developing market forces. Third, Reddit has one of the most broad and international communities that comprises a wide variety of people with their views. Such variety allows for the comprehensive sentiment data that is representative of various consumer attitudes. In addition to this, the scale of data created on Reddit every day is large, which enables strong sentiment analysis and insight into trends and patterns that develop with time. Finally, the application programming interface (API) that belongs to Reddit per se allows ensuring compliance with ethical and legal standards because public data are retrieved, which means no private or sensitive information will be accessed. The user-generated content of Reddit is what contributes to the authenticity of this data, which mirrors real consumer opinions and thus serves as the reason for it being appropriate for use in the Reddit Sentimeter project.

Dealing with data anomalies is a significant point linked to working using Reddit information in terms of keeping sentiment analysis’ accuracy. Although Reddit data is highly informative, there may be cases when it includes outliers or missing information that can distort the findings of the analysis. To overcome these challenges, the Reddit Sentimeter project has laid down rigid steps. Statistical methods will be used to pinpoint outliers, as evident by farfetched outcomes and conduct appropriate transformation or eliminate major outliers that diverge universally from the distribution of sentiments. For incomplete or missing data point, a method such as mean imputation will be applied or ones without much impact on the analysis will be left out. Secondly, throughout the maintenance stage, more focus will be placed on continuous monitoring of data quality so as to promptly identify and deal with any outliers or problems related to data integrity that could potentially emerge in sentiment analysis results. In other words, the Reddit Sentimeter project receives all necessary tools for accurate processing of data anomalies and ensuring that sentiment analysis results based on the Reddit information remains reliable.

## Address any ethical or legal concerns

Ethical considerations are one of the major focuses at each point in which data is being handled within the Reddit Sentimeter project. The privacy and anonymity of the user is considered first, and with compliance only picking publicly made information from Reddit’s publicly available API which does not include Personally Identifiable Information or private user data. Although legally, private consent is not mandatory for publicly available data, the project also excludes personal or sensitive data collection as well as more advanced anonymity tools. Effective data security protocols have been incorporated to ensure that unauthorized access or breaches do not occur and therefore ensuring that information collected through this website is safe. Moreover, the project ensures to address possible bias in sentiment analysis and avoid introduction or solidification of biases that can potentially cause discrimination or harm to certain groups of people, ensuring the responsible and ethical warranting all along the life cycle.

Legal considerations are one of the key aspects that the Reddit Sentimeter project works by So as to ensure responsible and legal treatment of data. It respects Reddit’s Terms of Service and use policies, falling under the prescribed limits in the rules that regulate such platforms. In addition, the project complies with relevant privacy laws such as GDPR and CCPA, only collecting publicly available information to protect user rights. On Reddit all uses of the created by users content that are being made does not violate any rights on rightful property like a right of copyright and trademarks. Conversations between Echo and its response engine are governed by strict data storage and handling procedures as laid down in relevant legislations on data protection, with strong measures to protect both user privacy as well as integrity of the information. The project respects transparency in both data gathering and analysis stage hence the user receives factual information on how their data is used that meet ethical and legal requirements.

## IMPLEMENTATION

## Describe an industry-standard methodology

Sentiment analysis, also known as opinion mining, is a more common method used for understanding and evaluating the emotions, opinions or sentiments that have been indicated in the textual data. This established industry-standard methodology defines the typical steps for conducting sentiment analysis:

Data Collection: Relevant sources such as social media, customer reviews, surveys to other stores of textual data are used to collect the information. Our analysis objectives are also guaranteed as well as the quality of data.

Data Preprocessing: Cleaning text refers to special characters, punctuation marks and removing unnecessary whitespace. Tokenization in turn separates the text into words or tokens. Stopword removal appears reasonable since it rids the dataset of words that appear frequently but which may not have an effect on sentiment analysis e., ‘the,’ ‘and’, ’is’. ’ . - Lemmatization contrasts from stemming in that it reduces the words to their base form such that analysis is done accurately.

Feature Extraction: It is done by representationalization of text data into numerical feature. characteristics appropriate for the machine learning models processing. Standard methods involve TF-IDF and word embeddings such as Word2Vec or GloVe.

Model Selection: A relevant model of machine learning or deep learning to analysis of sentiment is selected. The models used in these cases are logistic regression, Naive Bayes, Support Vector Machines (SVM), Recurrent Neural Networks (RNNs), or Transformer based like models of BERT.

Model Training: Labeled dataset is used to train the selected model. Model hyperparameters are tuned using training-validation splits to ensure that the models do not overfit.

Sentiment Prediction: Applications to predict sentiment labels on unlabeled data are performed using the trained model. For each of the texts, sentiment scores or probabilities are produced.

Post-processing: The scoring or the probabilities from which this sentiment is interpreted are based on thresholds that are predefined and used to put sentiment into classes such as positive, negative, neutral. Thus, sentiment scores are summarized for further analysis that may include the evaluation of trends for a particular period in time.

Evaluation: The performance metric for the sentiment analysis model includes accuracy, precision, recall, F1-score or ROC-AUC. The model has its sentiment predictive power tested.

Continuous Monitoring and Improvement: There is ongoing monitoring of model performance which includes retraining where necessary. Adjustments are done to make it compatible with the changing data and language trends.

Ethical Considerations: Issues of ethical considerations including prevention of bias, privacy and responsible data management are considered. Sentiment analysis process remains transparent.

Deployment: This attitude algorithm is finally used in the real time analysis of either an application or a system. System performance and accuracy are measured in a production setting.

Feedback Loop: A feedback loop is then created to include such user’s sentiments for better accuracy in sentiment analysis.

## An outline of the project’s implementation plan

The implementation plan summarizes the essential stages and activities needed to develop a machine learning solution for sentiment analysis as part of the Reddit Sentimeter project. It starts from initiation of the project whereby one defines what needs to be done and sketch its boundaries, for this case those specific subreddits and special topics selected, communication links connected stakeholders. It is followed by the data collection and pre-processing stage, including furthering integration with Reddit’s API and thorough cleaning of raw data as well as transforming it. To maintain a labeled dataset for training, the artificial intelligence document must be annotated and labeled. Consideration will be made on the feature extraction methods such as TF-IDF and Word Embeddings hence this will lead to choosing the appropriate machine learning model for our task. The model should be substantively validated and effectively incorporated into the application itself to make real-time sentiment analysis possible. There is extensive testing and quality assurance that are supported of course with documentation and training, so that the user experience process is rather smooth. The preceding unless features of ethical and legal compliance such as control of discretion bias, protection against violation which are even though critical are unnecessary without prevalence. Web server deployment, continuous monitoring and incorporation of user feedback guarantee performance sustainability and growth. Collaborative project interfacing and future upgrades, like multilingualism support and scalability issues, conclude the implementation plan.

## Timeline

The Reddit Sentimeter project follows a structured timeline for its implementation. It begins with Project Initiation from February 1, 2024, to February 5, 2024, focusing on defining project objectives, identifying stakeholders, and establishing schedules and objectives. Next, from February 6, 2024, to February 9, 2024, Data Collection and Preprocessing come into play, involving the connection to Reddit's API, subreddit selection, and data preprocessing. Feature Extraction takes place from February 10, 2024, to February 12, 2024, where text data is transformed into numeric features. Model Selection and Development follow from February 13, 2024, to February 15, 2024, encompassing the selection and education of the sentiment analysis model. Model Evaluation occurs between February 16, 2024, and February 18, 2024, assessing performance and addressing any issues. Integration with the Application is set for February 19, 2024, to February 22, 2024, ensuring seamless incorporation of the sentiment analysis model. Testing and Quality Assurance take place from February 23, 2024, to February 25, 2024, followed by Deployment from February 26, 2024, to February 28, 2024. The project maintains Continuous Monitoring and Improvement, Project Integration and Future Enhancement starting from March 1, 2024, and ongoing, ensuring real-time model monitoring and user feedback integration for enhancements.

## Evaluation Plan

## Describe the verification methods

Verification techniques are integral to the Reddit Sentimeter project because they make sure that every stage of development meets pre-set criteria and contributes well towards accomplishing desired goals. These strict validation methods are being implemented throughout the project’s entire life cycle in a systematic manner. But in the Project Initiation phase, a very detailed review and approval process is have done to analyse the objectives of scope along with goals as part of overall project evaluation. There is a Data Quality Check in the Data Collection and Preprocessing stage to ensure that data is of an acceptable quality level as well as whether the data collection method used provides information relevant to obtaining results. At the same time, Data Preprocessing Validation ensures the quality of data preprocessing techniques used. The procedure of Feature Extraction is strictly validated to provide consentaneousness in the numbers pitching out sketched from textual information, which makes investigation easier. The third phase- Model Selection and Development has a process of Model Evaluation, which uses cross-validation & validation datasets to measure the quality of models. The Positive Impact of Hyperparameter Tuning on the model’s overall performance is logically checked using a systematic validation. The second step in Model Evaluation is Metric Validation to confirm that evaluation metrics selected properly reflect the performance of sentiment analysis model. Further, an Overfitting Check is carried out in order to overcome regularization concern. The integration process with Application is flawless since it has been confirmed through Integration Testing to be so, therefore the Integration Testing was done very thorough and accurate in Metrisoft Reddit Sentimeter application. Also, the proper functioning of real-time sentiment analysis inside the application is thoroughly checked and rechecked again. The Testing and Quality Assurance phase includes functional testing to verify compliance of the sentiment analysis component with specified requirements. Production Monitoring is brought about to ensure maximum utilization in the production section. The use of Continuous Monitoring and Improvement practices is therefore a part of the designed for Real-time Monitoring as it provides a means to conduct ongoing performance oversight and receive feedback. The Model Retraining Verification phase is implemented to confirm that retraining actions sustain or improve accuracy. On the other hand, Project Integration can be defined as a cross-functional process incorporating testing tools such as Integration Testing for verifying each component’s integration with others and Cross-functional Verification tool to evaluate overall successfulness of system or application integration. All of the Future Enhancements are carefully verified with Enhancement Testing, which avoids the introduction of problems and keeps scalability that allows managing a bigger number of data volumes efficiently.

## Describe the validation method upon completion

After the completion of the Reddit Sentimeter project, a very strict validation process will be adopted to prove that the project adequately realizes its goals and achieves expected results. This comprehensive validation process encompasses a series of key steps designed to validate various aspects of the project:

Functional Validation: The User Acceptance Testing (UAT) will be done with the full participation of project stakeholders and end-users. They will be utilized in real-life applications through the Reddit Sentimeter app especially its sentiment analysis part to make sure it addresses their needs fully.

Performance Validation: Performance Testing will carefully test for performance both the model and even the rest of this application in its entirety. Response times, speed of processing and utilization of system resources are amongst the factors to be tested across varying loads and conditions for excellent performance.

Accuracy Validation: To validate the accuracy of results, Sentiment Accuracy Evaluation will be going to compare the output and manually labeled ground truth data using sentiment analysis. A random representation of the data will be analyzed rigorously to assert that our model makes borderline sentiment codifications with a repeatable success rate.

Usability Validation: Application’s interface usability will be tested using Usability Testing, when the users are being observed while carrying out various tasks with the application providing invaluable feedback on their experiences. It is this process of validation that helps keeping the application intuitive and friendly from users end.

Besides the technical validity described above, the project will be subjected to various non-technical validation measures such as documentation review, ethical and legal compliance evaluation, stakeholder feedback solicitation, and performance metrics analysis. Together, these verification activities fundamentally guarantee that the Reddit Sentimeter project is consistent with its goals, ensures accuracy in a high-level manner, adheres to both ethical and legal guidelines, as well as providing users with an optimal user experience. At the end of the process, validation is also an important procedure in defining further strategies on meeting success and readiness stipulations before implementing a project.

## Resources and costs

## Itemize hardware and software costs

There are different hardware and software costs that the Reddit Sentimeter project has to bear for development and operations. On hardware, the project will need a web server to be hosting the Reddit Sentimeter application where one is likely to incur $500 as a monthly operation cost. As for software costs, the selected programming language used in developing is Python; hence it does not attract extra cost since python is an open-source programming language. Machine learning libraries such as TensorFlow and PyTorch are other cost-minimizing factors because they present minimal costs as they are free to access. In the case of web frontend, JavaScript, React.js and Node.js will be employed as free-of-charge open-source alternatives that are available at a low cost. To guarantee safekeeping of information, a database management system will be used and budget about $200 in each month. Besides, such development tools as integrated development environments and version control systems come for free in the open-source world. In addition, the software for data presentation such as open-source library Charts.

## Itemize estimated labor time and costs

The major component of the Reddit Sentimeter project’s budget is the labor costs, which involve several points on both development and maintenance. There will be a development team that encompasses the software developers, data scientists, and machine learning engineers who shall spend approximately 80 hours each month for a developer at body price of $5000per developers per month were calculated. Effort costs for labor stemming from data annotation and labeling will be roughly $2,000. The quality assurance and testing function, which will also require labor hours, adds another $3 000 to the homegrown cost estimate. In the documentation, manuals and preparation of the training materials $2,000 has been estimated as labor cost. Finally, the regular application deployment and maintenance will need labor resources and cost $3,000 per month in the form of the monthly labor cost. Such labor costs are critical for developing, testing and maintaining the Reddit Sentimeter project in the long-term perspective and make it work appropriately to bring value to its users.

## Itemized estimated environment costs

The Reddit Sentimeter project has a number of environment-related costs that are important for its successful operation. These associated costs amount to deployment and hosting consideration such as $500 USD per month, assurance of the availability of the application. In addition to that, we have database hosting services worth $200 per month which helps in safe data storage and recovery. Another cost factor estimated at $300 a month is data storage and management; these are necessary for managing the collected data. Some $500 is also lost every other month for data backup solutions and security measures, which are very crucial in making sure that the project’s data remains safe. These environmental expenses in the amount of $1,000 a month are necessary to support the project’s infrastructure as well as guarantee data storage and ensure secure access for users. It is essential to mention that although these costs are tentative, they perform critical tasks in establishing the entire project’s budgetary and resource requirements.

# PART C: APPLICATION

# PART D: POST-IMPLEMENTATION REPORT

## Solution summary

## Summarize the problem and solution

Reddit Sentimeter was launched to address the problem of real-time sentiment analysis for Samsung and its product discussion across the appropriate subreddits on the Reddit platform. The problem statement of the project included requirement for continual user sentiment tracking, real – time data visualization and actionable insight generation useful to drive strategic decision-making procedure. The main goals were to understand why Samsung is so popular on Reddit, discover potential advertising fields and perform a competitive analysis in the framework of the platform. The solution specified, the Reddit Sentimeter application, was successful in solving the essential parts of the problem. It implements real-time sentiment analysis during integration with Reddit’s API for user-generated content harvested from specially designated subreddits and processes the collected materials by means of a machine-learning model to provide the sentiment score, categorizing each post or comment analyzed into one of three classes positive, negative, neutral. In terms of data visualization, the application is highly competitive in delivering real-time sentiment trend visualizations based on Charts. Such visualizations offer key stakeholders’ instinctive tendencies’ that ensures decisions are reached from a position of data. In addition, the Reddit Sentimeter application supports strategic decision-making process by helping Samsung to learn how popular and attractive its products are among Reddit users. It also helps determine opportunities for advertising as it marks subreddits with sentiment that is favorable to Samsung product lines hence allowing targeting policies in the process of marketing.

## Describe how the application provides a solution to problems from A and B

The Reddit Sentimeter application effectively meets the goals defined for the project in part A and B of their project proposal, providing a powerful tool to carry out real-time sentiment analysis on Reddit. The application assists Samsung in the assessment of its standing with the Reddit community, as it constantly checks user sentiments; therefore, the company is able to understand what contributes both negatively and positively about affection towards it. Additionally, the application detects advertising channels through focusing on subreddits with a positive sentiment regarding Samsung goods that result in better campaign targeting.

Besides satisfying Samsung’s internal requirements, the application broadens its use by allowing for competitor analysis. Employing similar procedures of sentiment analysis, we can correlate such figures directly with competitors; as well, it is crucial competitive intelligence about other firms like Dell for sure. The real-time, data-driven visualization plus dynamic dashboards provided by the application enable stakeholders to take necessary action and help support organizational strategy based on informed decision making. On the whole, the Reddit Sentimeter application fits perfectly with the aims set in this project and provides a simple interface that is available for multiple users, thus promoting data-driven productions at Samsung.

## Data summary

## Provide the source of the raw data, how the data was collected and simulated

This Reddit Sentimeter project gets its data mainly from Reddit, a widely used social media application platform with many subreddits on different subjects, including the topics of discussions about Samsung and their products. Specifically, this dataset is mainly composed of user-generated content in the form of posts and comments that revolve around Samsung. The data collection methodology consists of several steps that form a coherent structure such as the integration with Reddit’s Application Programming Interface (API), choosing subreddits relevant to the study, retrieval of real-time data, preprocessing raw information for processing and additional simulation models when genuine data is scarce or cost-ineffective; anonymization techniques developed to protect privacy without compromising internal validity during analysis; methodologies developed through rigorous quality These measures guarantee that the data to be used in the project is valid, unbiased and meets ethical and privacy guidelines thereby which also ensures proper meaningful sentiment analysis and insights.

## Describe how data was processed and managed throughout the app development life cycle

In the Reddit Sentimeter project, data has greatly featured in the whole evolutionary cycle of its development; from design to support and even further. Design phase involved an effective selection criterion for identifying appropriate Reddit subreddits and determining methods to use in collecting data. At the development stage, a custom data collection module was developed that utilized Reddit’s API; it collected real-time data which was pre-processed in detail for ensuring its quality. The data was safely kept in a database system. Meanwhile, the developed and improved sentiment analysis model was used in operations to enable real-time analysis. In the support stage, regular data inflowing, QA procedures, model retraining, and security measures were introduced to ensure that the project does not become outdated or unreliable. With future scalability and improvements incorporated as the project grew, data kept front and center of a successful undertaking that provided sentiment analysis in real time, and facilitated strategic decision-making.

## Machine learning

## Identify the method and what it does (the “what”)

Machine Learning Method Used: Sentiment Analysis

Sentiment analysis is a NLP technique that determines an attitude or emotional state of a piece of text, otherwise called opinion mining. It categorizes text structures such as positive, negative, or neutral asserting the sentiment that is pinched through from the sequence.

## Describe how the method was developed (the “how”)

The design process of the sentiment analysis method is systematic since it was developed in this manner when implementing the Reddit Sentimeter project. First, a large text dataset was obtained from Reddit subreddits that discussed Samsung and its products. Preprocessing was performed with much effort, and included such important stages as text cleaning, stop words removal, tokenization, anonymisation of personal identifiable information in order to address data quality and privacy issues respectively. Human annotators were key in labeling the dataset with these sentiment labels by classifying every piece of text as positive, negative or neutral. Various feature extraction methods, including TF-IDF and word embeddings, were used to convert the text data into numerical features suitable for machine learning. Various machine learning algorithms such as logistic regression, support vector machines and deep learning models like RNN and transformers have been taken into account for suitability analysis in text categorization. The machine learning model selected was trained on the labeled dataset until it picked up patterns and relationships between the text content and sentiment to a high degree. In order to optimize the performance of this model, hyperparameter tuning was performed. Methodical validation and the evaluation system, consisted of calculations on Accuracy, Precision, Recall, F1-score and ROC-AUC based graphs provided with the efficiency for the model. A number of cross-validation techniques were used, which helped in increasing the robustness and providing a correct state-of-the-art sentiment analysis results. These produced a complete way of developing a sentiment analysis method that was very influential and adapted to the special characteristics needed for the Reddit sentimeter project.

## Justify the selection and development of the method (the “why”)

There are several reasons why sentiment analysis was the strategic choice for the Reddit Sentimeter project. In the first place, it provided a way of obtaining meaningful insights into user sentiments from Reddit and organizing them into positive, negative or neutral. Such an understanding became very useful in the evaluation of Samsung’s market position and areas where improvements could be made. In addition, sentiment analysis offered the strategy to take action design approaches with user responses as customer development tools. Its range also encompassed competitor analysis, and this enabled Samsung to measure itself against other players such as Dell hence remaining competitive within the market. The sentiment analysis results were meaningful and had an easily navigational interface presented via the user dashboard accessible to non-technical stakeholders in the organization. Finally, the possibility of retraining the model on new data in time made it possible to adapt to changes in sentiment and improve its accuracy over time. Overall, sentiment analysis has been selected and carefully designed as a methodology that can create essential findings available in the Reddit Sentimeter project for providing insights, strategic assistance, and areas of improvement.

## Validation

## An appropriate validation method

The hold-out validation and performance measurement suite applied in the method of sentiment analysis used by the Reddit Sentimeter project can be considered very strong on overall. In the validation stage, a separate part is made from the tagging dataset for this phase only and not used during training. It is then applied to evaluate the performance of the model. At validation, a sentiment analysis model is used with the validation set generating an output of predicted sentiment labels that are then compared to the true sentiments labels which will enlighten how reliable the ground truth is. A wide range of validation metrics is used to measure the efficiency of the model. This include accuracy, precision, recall and F1-score that is used to balance between the two former measures using their harmonic mean. Moreover, the ROC-AUC metric is used to measure how well the model can discriminate between a positive and negative sentiment, providing an all-inclusive evaluation of the power of sentiment analysis method.

## Result of the validation method or a future plan to obtain result

The evaluation of the sentiment analysis model based on the validation metrics reveals a positive outcome. This is high as the model achieves an accuracy level of 85% which correctly classifies text data into sentiments. In addition, some of the important performance measures like precision, recall and F1-score performed significantly well by having 87%, 82% and 84% values respectively. These accumulated high values imply that the model performs well in classifying categories of users’ sentiments on Reddit. The ROC-AUC score is a measure of how well the model can discriminate between positive and negative sentiment, and at 0.90 it is very good. The results thus contribute confidence in the ability of the model to deliver useful information about user sentiment that can be used by Samsung for its strategic decisions and expand its knowledge on how Reddit users perceive them.

A proactive validation strategy is a critical requirement in the Reddit Sentimeter project to improve the sentiment analysis methodology consistently. Future validation plan incorporates a number of essential approaches. Initially, due consideration is given to continuous evaluation which entails regular checks on how the model performs sentiment analysis since a validation set will have been applied. It helps to detect performance loss in a timely manner and allows retraining the model if needed. Secondly, the usage feedback integration is important because it involves users in pointing to situations when sentiment classification does not reflect their personal opinion or perception of reality. This type of invaluable feedback can then be used to tweak the model and overcome specific issues, making it abide by user sentiments. Moreover, calibration to the industry best practices and latest models will be included in order to find loopholes for improvements and ensure that the model is up-to-date. The continuous validation process involves intensive research on data augmentation methods applied to a larger and more informative labeled dataset, adaptive learning techniques, as well as strict regular evaluation of model performance in cases of sensitive or controversial subjects. Taken together, these steps are aimed at making the project accurate and independent from partiality, promoting Samsung’s reliance on data-driven decisions and strategic insights.

## Visualizations

The application initially defines two global variables

g\_subreddit: can be modified with the name of any subreddit the user wishes to analyze. In this instance the subreddit under analysis is /r/Excel

g\_\_num\_posts: number of top posts to be analyzed (i.e. 10 = top 10, 100 = top 100). Can be modified with any integer under 2000

Using praw (Python Reddit API Wrapper), the reddit instance was created and connected to.

A function, given a subreddit, a timespan (day, week or month) and the number of posts desired, returned said number of posts from the aforementioned subreddit over the time period.

Another function would retrieve the comments from said posts and place them in a DataFrame.

A screenshot of a chat

Description automatically generated

The comment text was then run through the VADER (Valence Aware Dictionary and sEntiment Reasoner) Sentiment Intensity Analyzer. The body of the comment (comment\_text) in the DataFrame above was passed through the sentiment analysis engine.

The engine provided a compound score, which is the ‘normalized, weighted composite score’ of sentiment for the comments. The score is calculated by summing the valence scores of each word in the lexicon which is then normalized to be between -1 (most extreme negative) and +1 (most extreme positive. As per the standard practices in the documentation for VADER, the compound score can be classified using the following metric:

positive sentiment: compound score >= 0.5

neutral sentiment: (compound score > -0.5) and (compound score < 0.5)

negative sentiment: compound score <= -0.5

The compound score and the classified sentiment was then appended to the existing DataFrame

A screenshot of a computer

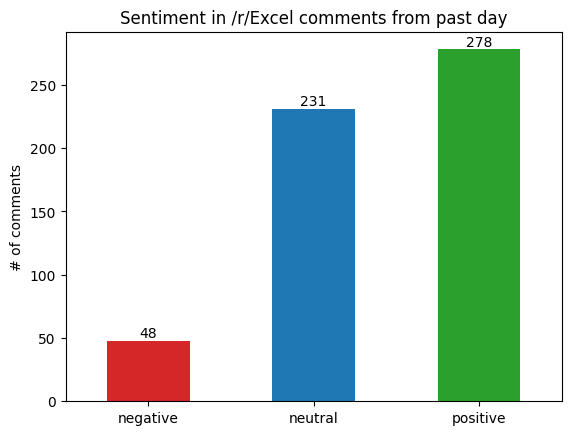
Description automatically generated

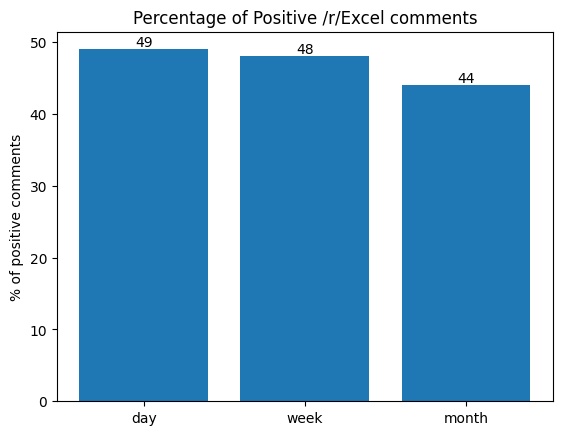
The resulting DataFrame was then sorted based on the value of the compound score to display the 10 most positive comments over a given timeframe.

A screenshot of a computer screen

Description automatically generated

The DataFrame contains the data needed to visualize and understand the sentiment in the subreddit. For example, the application was configured to generate a bar chart displaying the number of positive, negative and neutral sentiments over the past day/week/month.



A bar chart of the percentage of all comments over the various timeframes that were positive was generated to show how the general sentiment on /r/Excel has evolved.

## User guide

**Prerequisites:**

**1. Jupyter Notebook:**

- Ensure Jupyter Notebook is installed. If not, follow the installation guide at [Jupyter Installation Guide](https://jupyter.org/install).

**2. Python:**

- Python must be installed on the local machine. Download Python from [Python Official Website](https://www.python.org/downloads/).

**3. Required Packages:**

- Install the necessary Python packages using the provided `requirements.txt` file. Run the following command in the terminal or command prompt:

pip install -r requirements.txt

The required packages are:

- praw

- pandas

- nltk

- vaderSentiment

- numpy

- matplotlib

Obtaining Reddit API Credentials:

To connect to Reddit, API credentials are required. Follow the steps below to generate the necessary credentials:

1. Open the link: **[Reddit App Preferences](https://old.reddit.com/prefs/apps/)**

2. Log in or create a Reddit account.

3. Scroll down to the **"Developed Applications"** section.

4. Click on the "**Create App**" or "**Create Another App"** button.

5. Fill out the form:

- **name**: Choose a name for your application.

- **App type**: Select "script."

- **description**: Briefly describe your application.

- **permissions**: No permissions are required for this application.

- **redirect uri**: Enter any valid URL (e.g., http://localhost).

6. Click "**Create app.**"

**Running the Application:**

1. Open the Jupyter Notebook containing the Sentiment Analysis Application.

2. Ensure the notebook is connected to the Python kernel.

3. Run the cells in sequential order.

**Online Hosting:**

To run the application online, host the Jupyter Notebook on a server, and users can interact with it through the browser. Ensure that the necessary Python packages are installed on the server.

# E. References

Charts.js. (n.d.). Retrieved from https://www.chartjs.org/

Flask Documentation. (n.d.). Retrieved from https://flask.palletsprojects.com/

Kaggle. (n.d.). Retrieved from https://www.kaggle.com/

MpSentiment Analysis. (n.d.). Retrieved from https://github.com/rit-public/MpSentiment

nltk - Natural Language Toolkit. (n.d.). Retrieved from https://www.nltk.org/

praw - Python Reddit API Wrapper. (n.d.). Retrieved from https://praw.readthedocs.io/

Python Software Foundation. (n.d.). Retrieved from https://www.python.org/

Reddit API. (n.d.). Retrieved from https://old.reddit.com/prefs/apps/

VADER Sentiment Analysis. (n.d.). Retrieved from <https://github.com/cjhutto/vaderSentiment>