

PATRYK LISIK

MATEUSZ BOROWIEC

Agile Methods

Assignment #2

1 Introduction

The tool would use natural language processing techniques to analyse text data from social media platforms and determine the sentiment (positive, negative, neutral) of the posts. The user could input a specific topic or hashtag to search for, and the tool would return a breakdown of the sentiment of the posts related to that topic or hashtag. Additionally, the tool could also display the posts themselves, along with their sentiment labels, in a visual format such as a bar chart or word cloud.

The sentiment analysis tool would use techniques such as natural language processing and machine learning to analyse text data from social media platforms like Twitter, Facebook, or Instagram. The user would have the ability to input a specific topic or hashtag that they are interested in, and the tool would then search for and gather all relevant posts from that topic or hashtag.

Once the relevant posts have been collected, the tool would use NLP (Natural Language Processing) techniques such as tokenization, POS tagging, and sentiment analysis to determine the sentiment of each post. Sentiment analysis is the process of determining whether a piece of text is positive, negative, or neutral in tone. There are various algorithms and models that can be used to perform sentiment analysis, such as using pre-trained word embeddings or training a machine learning model on a labelled dataset.

The tool would then present the sentiment analysis results in a visual format, such as a bar chart or word cloud. This would allow the user to quickly understand the overall sentiment of the posts related to the specific topic or hashtag they searched for, as well as the distribution of positive, negative, and neutral posts. Additionally, the tool could also display the posts themselves, along with their sentiment labels, in a list or table format, which would allow the user to read the posts and see the sentiment analysis results for each individual post.

Contents

1	Introduction	1
2	Project phases	1
3	Work breakdown structure and Gantt chart	2
4	Resources	4

2 Project phases

There are several phases that would be involved in building a sentiment analysis tool for social media posts, including:

Planning and research: This phase would involve defining the project scope, researching and identifying the specific social media platforms to be used, and gathering a dataset of labelled sentiment data for training the machine learning model.

Data collection: In this phase, the tool would gather the relevant social media posts from the specified platforms and hashtags. The data would be collected using the social

media platform's API, which allows the developer to fetch the data in a structured format.

Data preprocessing: The collected data would be preprocessed to remove any irrelevant information such as URLs, emojis, and mentions. The text data would be cleaned, tokenized and transformed into numerical values that the model can understand.

Model Training: Here, the labelled dataset would be used to train a machine learning model. The model would learn the patterns and correlation between the words and the sentiments. This can be done by using various algorithms such as Logistic Regression, Naive Bayes, Random Forest, etc.

Model evaluation: The trained model would be evaluated on a separate dataset to measure its performance. This phase would help to identify any errors or inaccuracies in the model, and make adjustments as needed.

Deployment: Once the model has been trained and evaluated, it would be deployed as a web application or API. This would allow users to input their desired topic or hashtag and receive sentiment analysis results in real-time.

Maintenance: The tool would require regular maintenance and updates to ensure that it continues to function correctly and effectively. This includes monitoring the performance, testing and fixing bugs, as well as updating the dataset and model to improve the performance.

3 Work breakdown structure and Gantt chart

1. Project initiation
 - (a) Define project scope and objectives
 - (b) Research social media platforms and dataset
 - (c) Gather project requirements
2. Data collection
 - (a) Develop code to access social media platform's API
 - (b) Collect social media posts using API
 - (c) Preprocess data (remove irrelevant information, tokenize etc.)
3. Model development
 - (a) Train machine learning model using labelled dataset
 - (b) Evaluate model performance
 - (c) Refine model as needed
4. Deployment
 - (a) Design web application
 - (b) Design web API
 - (c) Develop web app and API
 - (d) Test application

- (e) Deploy application to hosting platform
- 5. Stabilization
 - (a) Monitor application performance
 - (b) Fix user bugs
 - (c) Update dataset and model to improve performance
- 6. Project close
 - (a) Review project progress and deliverables
 - (b) Archive project documents and materials

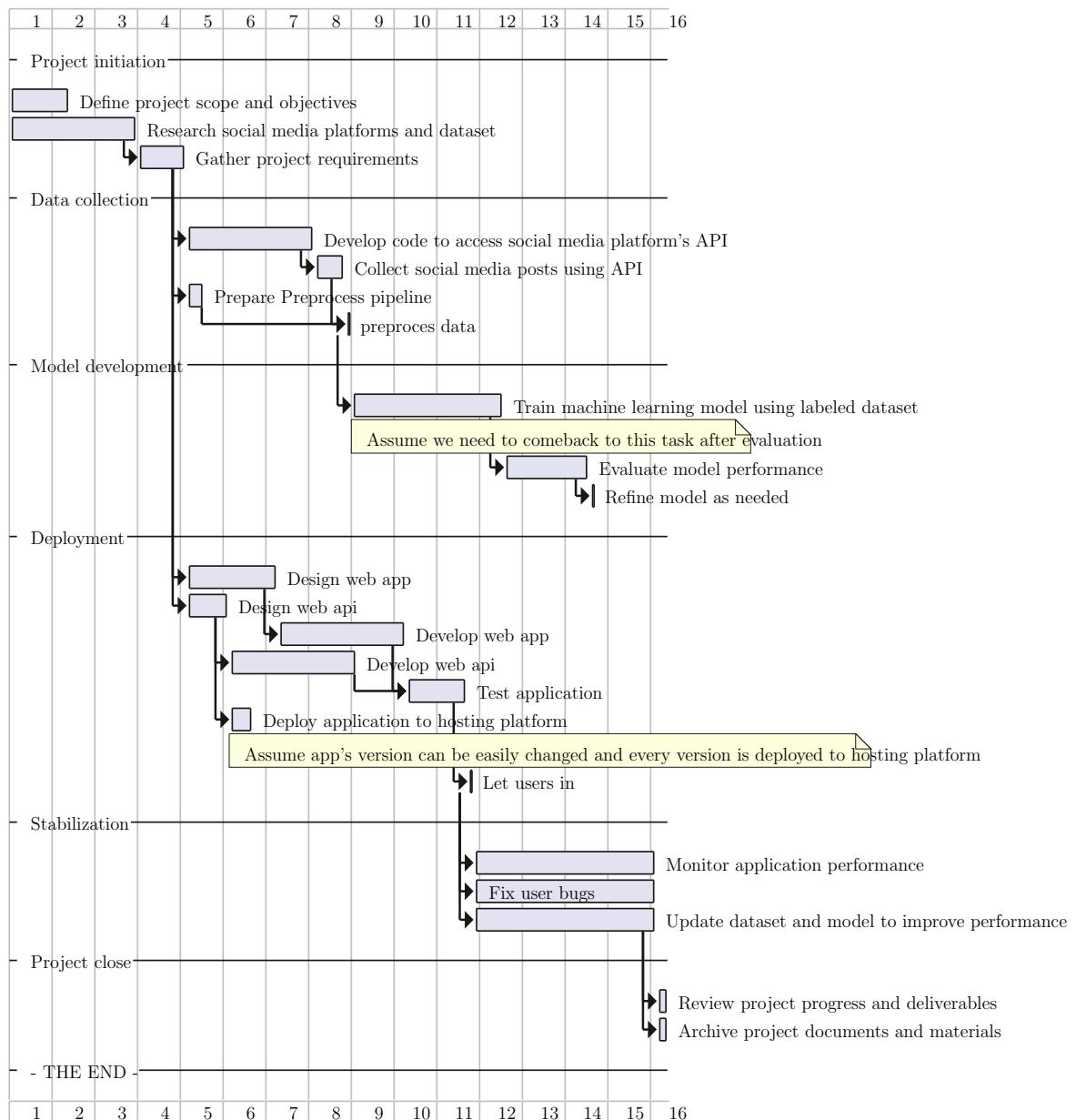


Figure 1: Gantt chart

4 Resources

Technical expertise: The project would require a team with knowledge and experience in natural language processing (NLP), machine learning, and web development. The team should have the ability to access social media platform's API, preprocess and extract features from the text data, train and evaluate machine learning models, and develop a web application or API.

Data: A labelled dataset of social media posts with sentiment labels would be required to train the machine learning model. This dataset could be collected from publicly available sources or by using the social media platform's API.

Hardware and software: The project would require access to a computer with sufficient processing power and memory to train machine learning models. The team would also need software for data preprocessing, model training and evaluation, and web development.

Cloud infrastructure: Would be required to host the application and make it available to users.

Time and budget: The project will require a significant amount of time and budget to complete. The team should have a clear understanding of the project schedule and budget, as well as any contingencies for delays or unexpected expenses.

Communication: Clear and regular communication between team members and stakeholders would be essential for the success of the project. This includes keeping everyone informed of progress and any issues that arise, as well as soliciting feedback and input as needed.

Continuous improvement: The project should be evaluated after completion and improve upon it. This would require monitoring the performance, testing and fixing bugs, as well as updating the dataset and model to improve the performance.