**1. Project Overview**

The project plan must include the following:

**· A Project Title.**

Sentiment Analysis of Emotions in Twitter and Amazon Reviews

**· A short summary of the project topic and background.**

Sentiment analysis is the case study of computational study about opinions, sentiments and emotions expressed in the text. In this project, emotions are analyzed in textual data coming from two particular sources, Twitter and Amazon reviews. Irrespective of whether they are truncated or explicit, Twitter can be considered real-time expressions of sentiments on various topics and Amazon reviews are elaborated opinions from customers about products or services. The project will seek to understand and compare emotional tone of social media post and e commerce feedback writing these datasets.

**· A Research Question.**

Can machine learning enhance predictive accuracy beyond traditional statistical models?

**· The Project Objectives.**

* To create a robust model capable of accurately detecting and classifying emotions in both Twitter posts and Amazon product reviews.
* To analyze and contrast the emotional tones between social media content and e-commerce reviews to identify unique patterns or similarities.
* To assess the effectiveness of the sentiment analysis model across both datasets, ensuring its applicability and reliability in diverse textual contexts.

**· Reference List**

Arif, M., Hasan, M., Al Shiam, S.A., Ahmed, M.P., Tusher, M.I., Hossan, M.Z., Uddin, A., Devi, S., Rahman, M.H. and Biswas, M.Z.A., 2024. Predicting Customer Sentiment in Social Media Interactions: Analyzing Amazon Help Twitter Conversations Using Machine Learning. *International Journal of Advanced Science Computing and Engineering*, 6(2), pp.52–56.

Muthu Ruben, V., VijayaKumar, R. and Sateesh Kumar, T.K., 2024. Unveiling the Emotions: A Sentiment Analysis of Amazon Customer Feedback. In: R. El Khoury, ed. *Anticipating Future Business Trends: Navigating Artificial Intelligence Innovations*, Studies in Systems, Decision and Control. [online] Cham: Springer Nature Switzerland. pp.403–410. <https://doi.org/10.1007/978-3-031-63402-4_33>.

Wassan, S., Chen, X., Shen, T., Waqar, M. and Jhanjhi, N.Z., 2021. Amazon product sentiment analysis using machine learning techniques. *Revista Argentina de Clínica Psicológica*, 30(1), p.695.

**2. Project Plan: Task List and/or Project Timeline**

| **Task Number** | **Task Description** | **Start Date** | **End Date** | **Notes** |
| --- | --- | --- | --- | --- |
| 1 | Conduct Literature Review on Sentiment Analysis Techniques | Feb 10, 2025 | Feb 24, 2025 | Review existing methodologies and tools used in sentiment analysis. |
| 2 | Data Collection and Preprocessing | Feb 17, 2025 | Mar 3, 2025 | Acquire datasets from specified sources and perform necessary cleaning and formatting. |
| 3 | Model Development and Training | Mar 3, 2025 | Mar 31, 2025 | Develop and train the sentiment analysis model using the preprocessed data. |
| 4 | Model Evaluation and Validation | Mar 24, 2025 | Apr 7, 2025 | Test the model's performance and make necessary adjustments. |
| 5 | Comparative Analysis of Emotional Expressions | Apr 7, 2025 | Apr 21, 2025 | Analyze and compare the emotional tones in Twitter and Amazon review datasets. |
| 6 | Preparation for Final Assessment | Apr 21, 2025 | May 5, 2025 | Compile findings, prepare reports, and develop presentation materials. |
| 7 | Final Assessment Submission | May 5, 2025 | May 5, 2025 | Submit all project deliverables for evaluation. |

**3. Data Management Plan**

**Overview of the Dataset**

* This dataset is made up from classified tweets that is very substantial, labeled for positive or negative sentiment. The data was collected to assist in the research of sentiment analysis and is available at Hugging Face.
* This dataset contains comments and user ratings for Amazon product. It was curated to be used for sentiment analysis of e-commerce feedback and available at Hugging Face.

**Data Collection**

The Hugging Face URLs that are provided will be the datasets’ sources directly. The formats will be downloaded and will be subsequently preprocessed and analysed.

**Metadata**

* **Twitter Dataset**: More than a million tweets are labeled for sentiment. The dataset is in CSV format, with each entry comprising the tweet text and its corresponding sentiment label.
* **Amazon Reviews Dataset**: Comprises approximately 4,915 reviews. The dataset is in CSV format, with fields including reviewer name, review text, rating, and additional metadata.

**Document Control**

A GitHub repository will be established to manage the project's codebase and documentation. Regular commits will be made on a weekly basis to track progress and maintain version control. The repository will be organized with clear file naming conventions and versioning to ensure clarity and traceability.

ReadMe File:

Upon project completion, a comprehensive ReadMe file will be included in the GitHub repository. This document will provide an overview of the project, instructions on how to set up and run the code, explanations of the directory structure, and any other pertinent information to assist future users in understanding and utilizing the codebase.

Security and Storage:

Data and code will be securely stored in the GitHub repository and backed up on a secondary platform, such as OneDrive, on a weekly basis. Access to the repository will be managed to ensure that only authorized personnel can make changes, thereby maintaining data integrity and security.

Ethical requirements: You must address each of the following issues and state how your specific dataset meets these requirements, give evidence when possible (e.g. screenshots or references):

1. Does the data come under GDPR requirements?

The datasets utilized are publicly available and anonymized, ensuring compliance with GDPR regulations.

2. Does the project conform to UH ethical policies?

The project adheres to the ethical guidelines set forth by the University of Hertfordshire, including the use of publicly available data and the protection of individual privacy.

3. Do you have permission to use the data for your proposed research project?

The datasets are open-access and intended for research purposes, granting implicit permission for use in this project.

4. Are you assured that the data was collected ethical (i.e. by the original people who gathered/collected/ collated/made the data)?

The data was collected and made available by reputable sources, ensuring that ethical standards were maintained during the initial data gathering process.