LAB 4 PART 2

MINUTES OF MEETING

Table of Contents

Team Details

Team Name: SSK

Name	USC ID
Shreyansh Baredia	3739756887
Soma Meghana Prathipati	4226812081
Kayvan Shah	1106650685

Meeting Notes for Day 1

Date: 2/13/2024

Agenda:

- Review of Assignment Requirements
- Assignment Task Delegation
- Setting Up Weekly Meeting Schedule

Meeting Summary:

- A detailed review of assignment requirements, focusing on web scraping, data preprocessing, clustering algorithms, and automation.
- Identified key challenges and potential solutions for each aspect.
- Assigned roles and responsibilities based on individual expertise and preferences to maximize efficiency and productivity.

Action Items:

- Begin development of web scraping script
- Research and evaluate various clustering algorithms such as K-means, hierarchical clustering, and DBSCAN, considering their suitability for our dataset and objectives.
- Investigate automation frameworks and tools for periodic data updates, focusing on scalability and reliability.

Meeting Notes for Day 2

Date: 2/14/2024

Agenda:

- Progress Update on Assigned Tasks
- Clustering Algorithm Selection
- Data Preprocessing Techniques Discussion

Meeting Summary:

- Comprehensive progress update on the development of web scraping script, including data collection strategies, handling of dynamic web content, and error handling mechanisms.
- In-depth exploration of various clustering algorithms, evaluating their performance metrics, computational complexity, and suitability for our dataset characteristics.
- Intensive discussion on data preprocessing techniques, considering text cleaning, tokenization, stemming, and TF-IDF or Doc2Vec vectorization methods to extract meaningful features from unstructured text data.

Action Items:

- Continue refining the web scraping script and implement advanced features such as multi-threading or asynchronous processing for improved efficiency.
- Conduct a comparative analysis of clustering algorithms based on clustering quality metrics.
- To optimize clustering performance, experiment with different data preprocessing pipelines, including text normalization techniques and feature engineering approaches.

Meeting Notes for Day 3

Date: 2/15/2024

Agenda:

- Chosen Clustering Algorithm Presentation
- Data Preprocessing Method Finalization
- Integration Plan Discussion

Meeting Summary:

- Implement the K-means clustering algorithm for its simplicity, scalability, and interpretability.
- Detailed presentation of the chosen data preprocessing method, encompassing text preprocessing steps such as stop word removal, punctuation removal, and lemmatization to enhance feature extraction accuracy.
- Extensive discussion on the integration plan, outlining the workflow for seamless integration of web scraping, data preprocessing, and clustering components, ensuring data consistency and integrity throughout the process.

Action Items:

- Implement web scraping script enhancements based on feedback and requirements gathered during the meeting.
- Fine-tune K means clustering parameters and optimizing the clustering performance.
- Finalize the data preprocessing pipeline and conduct comprehensive testing to validate the preprocessing accuracy and efficiency.

Meeting Notes for Day 4

Date: 2/16/2024

Agenda:

- Implementation Progress Review
- Troubleshooting and Issue Resolution
- Automation Integration Planning

Meeting Summary:

- A thorough review of individual implementation progress, highlighting successful milestones achieved and challenges encountered during the development process.
- Collaborative troubleshooting session to address technical issues and resolve implementation bottlenecks, leveraging collective expertise and problem-solving skills.
- Strategic planning for automation integration, including the design of robust error-handling mechanisms, logging strategies, and scalability considerations to ensure seamless and reliable data updates.

Action Items:

- Address any outstanding issues and finalize the web scraping script for production deployment.
- Optimize clustering algorithm performance and conduct extensive testing to validate clustering quality and scalability.
- Implement automation scripts for periodic data updates, incorporating error handling mechanisms and scheduling optimizations for efficient data processing.

Meeting Notes for Day 5

Date: 2/17/2024

Agenda:

- Final Implementation and Integration
- Submission Preparation
- Presentation and Video Planning

Meeting Summary:

- Completed the final implementation and integration of all components, ensuring seamless interaction and functionality across the system.
- Prepared submission materials, including codebase documentation, technical specifications, and demonstration videos, adhering to submission guidelines and requirements.
- Devised a comprehensive presentation and video plan outlining key project milestones, technical insights, and performance metrics to effectively communicate the project's objectives and achievements.

Action Items:

- Perform a final round of system testing and validation to ensure readiness for submission.
- Compile and organize submission materials into a cohesive package, including code repositories, documentation, and files.