Veloxxity: Sentimental Analysis on Online Social Media Relating to the South China Sea

PREPARED FOR

Veloxxity LLC

Tom Coyle

CS4485

PREPARED BY

Ken Lin

Jack Moody

Collin Gibson

Riki Sasaki

UTD Senior Design Team

Abstract

This proposal serves to conceptualize crucial aspects of the development process of the UTD Senior Design project (sentiment analysis application). The goal of this project in accordance with the stakeholder (Sponsor: Veloxxity LLC) requirements is to develop a sustainable and interactive sentiment analysis that collects viral posts with regards to the international conflicts in the South China Sea. The application will be deployed as a Representational state transfer (REST) API with the objective of providing a solution to the distribution of misinformation on online social media platforms.

Table of Contents

EXECUTIVE SUMMARY

Under the guidance of the project sponsor (Veloxxity), our team is tasked to develop an application that identifies and appropriately documents the source of hostile messages on online social media (OSM) platforms. In order to satisfy the stakeholders requirements, the deployment of our application is done via a Sentiment Analysis API.

The API is accessible as a RESTful web service (Representative State Transfer) with the support of the IBM Cloud database. The advantages of deploying our sentiment analysis as an RESTful web service includes the ability of caching, uniform interface for data manipulation via standard HTTP methods (i.e POST, DELETE, PUT, etc) and the support of stateless protocol.

The output of our API includes the visualization and dynamic mapping of our measured dataset. The analysis of our dataset is achieved through the utilization of machine learning (ML) methodologies. Accurate execution of ML methodology requires the input of adequate training models. We can identify accurate and meaningful datasets from Kaggle. Kaggle is an open source platform that hosts non-proprietary datasets free to the public. Required datasets on major social media sites such as Twitter, FaceBook and Tiktok crucial for the development of our API can be collected from Kaggle's massive dataset repository.

The visualization of our outputs is established through plotly which is an interactive plotting library that supports different statistical graphs.

We are confident that society will benefit from the deployment of our application which includes the project stakeholders as well as the general public. With the support of our Sentiment Analysis API we are able to accurately evaluate hostil messages and provide dynamic outputs with interactive graphical visualizations coupled with important indicators including time, location or the underlying sentiment of the analyzed data.

1. Introduction

Veloxxity is a company that provides solutions to meet critical intelligence, information operations and cyber security requirements currently focused on the challenge of addressing hostile foreign influence. For this semester, our project is to develop an application that analyzes the sentiment of viral posts regarding the South China Sea, and with that analysis create a pie chart that displays the proportion of positive, neutral, and negative posts on the South China Sea.

2. Discussion

We are creating an application that measures the sentiment behind viral posts on Social Media that regard the South China Sea, and creating a pie chart that displays the proportion of positive, neutral, and negative posts on the topic of the South China Sea. To develop this application we are using the waterfall software model. We believe that the Waterfall is a simple and intuitive model with clear stages. It has clear stages, which makes it easy to arrange tasks due to the model's rigidity.

3. Resources

Software Resources

Plotly: Interactive plotting library that supports different graphs. The front end for ML and data science model, data visualization.

NLTK library: Provides Lexical analysis, n-gram, part-of-speech tagger, tree model and named-entity.

Kaggle: Data provider (Crowdsourced platform), non-proprietary datasets, support CSVs, JSON, SQLite, etc.

IBM Cloud: Deploy and visualize resources on this platform.

Hardware Resources

Data processor (Laptop/Desktop): Project development tool

Human Resources

UTD senior design team

4. Key Roles

Ken Lin: Point of Contact with Veloxxity, Meeting Planner and Team leader

Collin Gibson: Task tracker, Scrum Master

Jack Moody: Responsible for the appropriate documentation of the project development

Riki Sasaki: Responsible for the evaluation of requirements and the contributor to the appropriate estimation of costs

5. Communication Plans

*The table below details the Design Team's approach towards communication

What Information	Target Audience	When?	Method of Communication	Provider
Status Reports	Faculty Advisor and Sponsor	Weekly	Email	Team Leader (Ken Lin)
Project Proposal	Sponsor and Faculty Advisor	Week 7	Elearning Submission	Team Leader
Final Report	Sponsor and Faculty Advisor	Week 15	Elearning Submission	Team Leader
Poster	Faculty Advisor	Week 15	Email	Group
Accepted Change Requests (system design and architecture)	Sponsor	Week 8-14	Email/ Online Meeting (standup meetings)	Group
Final software presentation	Sponsor	Week 15	Online Meeting	Group

6. Risk Analysis

*The table (Risk Analysis Matrix) below highlights the potential risks that the project might encounter

Probability **Impact**	Low	Medium	High
High	Lack of Understanding Software		Scheduling Conflicts
Medium	Group Member contracts Covid-19	Failure to Satisfy System/ Architecture Requirements	
Low			Failure to respond in an expedited manner

7. Cost

The estimated cost tentatively based on the current state of software development. Currently, we have not encountered any monetary cost.

The cost of work hours (effort) is dependent on the individual work packages detailed in the Work Breakdown Structure (WBS). Currently our team is averaging around 5-6 work hours each week. This information is recorded each week in our weekly report. Individual efforts are not recorded at this time.

The incurring cost of effort is expected to increase as the team transitions to the development phase of the project life cycle.

8. Time Table

*The Software Development life Cycle is partitioned into three phases detailed below

1. **Phase One**: acquire appropriate libraries and data sources.
    a. Python libraries applicable to sentiment analysis installed via pip.
    b. Data sources obtained via Kaggle (Data Science company that provides the required dataset needed for our application).
2. **Phase Two**: model development and training/validation of our sentiment analysis.
    a. Sentiment analysis can be modeled via the Plotly Python library (Python library that provides interactive plotting for statistical graphs).
    b. Training/validation of our AI (sentiment analysis) with appropriate training models.
3. **Phase Three**: completion of the application and provide the desired information/visualization.

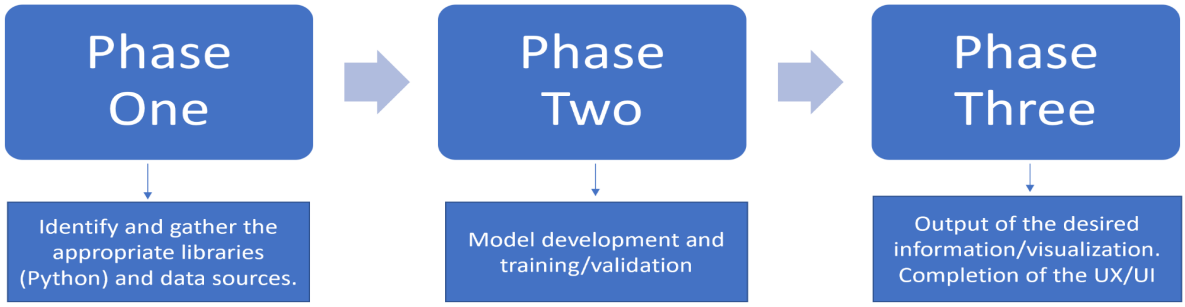

*Table below is a progress tracker for the Software Life Cycle

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15
Activity															
Kick - off Meeting	X														
Weekly Team/Sponsor Meeting		X													
Weekly Team/Sponsor Meeting			X												
Weekly Team/Sponsor Meeting				X											
Weekly Team/Sponsor Meeting					X										
Weekly Team/Sponsor Meeting						X									
Weekly Team/Sponsor Meeting							X								
Project Proposal							X								
Software Development															
Software Development															
Software Development															
Software Development															
Software Development															
Software Development															
Software Development															
Poster/Slide/Sponsor Approval Form															
Final Report															
Senior Design Day - Poster Session															
Senior Design Day - Oral Presentation															

9. Evaluation

Accurate evaluation on the current progress of the project development can be accessed by the number of completed software sprints. Efforts (Work hours) can be effectively measured by the number of work packages completed after each software sprint. Evaluation of the Software Development Cycle in its entirety can be evaluated by the number of completed software development phases. The final application may be deemed successful if it meets all the requirements provided by the stakeholder or if it finds a degree of popularity or demand in the marketplace. The assessment of the project success may not be fully realized until the conclusion of beta testing or the public deployment of the software.

10. Conclusion (Final Thoughts)

The development of this application can greatly reduce the spread of misinformation on online social media platforms as well as providing an automated solution towards hostile influences. With the support of our Sentiment Analysis API we can accurately determine the original source of misinformation. The outputs of our API provides both quantitative data and qualitative data (Graph visualization via Plotly), thus providing us with the appropriate resources necessary in establishing counter measure efforts.

11. Contact Information

Collin Gibson - Cag160330@utdallas.edu

Collin Gibson is a member of the UTD Senior Design Team. Collin is immensely sympathetic towards the field of Aviation. After graduating from UTD, he plans to fully immerse himself in this field of Aviation.

Jack Moody - Jtm160430@utdallas.edu

Jack Moody is a Senior Computer Science Student at The University of Texas at Dallas. He Graduates in December 2021.

Ken Lin - Kxl173530@utdallas.edu

Ken Lin is an aspiring software developer currently pursuing a Bachelor of Science degree in Computer Science at the University of Texas in Dallas. Ken is a member of the UTD Senior Design Team.

Riki Sasaki -Rxs190043@utdallas.edu

Riki Sasaki is a senior year student at The University of Texas at Dallas majoring in Computer Science. He is currently in UTD Computer Science Project class and is a passionate software developer.

12. Sources

*The references listed below is in the IEEE format

1. “Natural language toolkit¶,” *Natural Language Toolkit - NLTK 3.6.3 documentation*. [Online]. Available: https://www.nltk.org/. [Accessed: 08-Oct-2021].

2. “The front end for ML and data science models,” *Plotly*. [Online]. Available: https://plotly.com/. [Accessed: 08-Oct-2021].

3. “Your machine learning and Data Science Community,” *Kaggle*. [Online]. Available: https://www.kaggle.com/. [Accessed: 08-Oct-2021].

4. “IBM Cloud,” *IBM*. [Online]. Available: https://www.ibm.com/cloud. [Accessed: 08-Oct-2021].

13. Appendix

*The references listed below is in the IEEE format

1. “Natural language toolkit¶,” *Natural Language Toolkit - NLTK 3.6.3 documentation*. [Online]. Available: https://www.nltk.org/. [Accessed: 08-Oct-2021].

2. “The front end for ML and data science models,” *Plotly*. [Online]. Available: https://plotly.com/. [Accessed: 08-Oct-2021].

3. “Your machine learning and Data Science Community,” *Kaggle*. [Online]. Available: https://www.kaggle.com/. [Accessed: 08-Oct-2021].

4. “IBM Cloud,” *IBM*. [Online]. Available: https://www.ibm.com/cloud. [Accessed: 08-Oct-2021].

13. Appendix - Work Breakdown Structure (WBS)

*WBS provided below is in the indentation format

Sentimental Analysis on Online Social Media(OSM) Relating to the South China Sea

Prepared by: Ken Lin, Jack Moody, Riki Sasaki, Collin Gibson **Date: 9/30/21**

1.0 WBS Online Social Media(OSM) Relating to the South China Sea
- 1.1 Project Management
  - 1.1.1 Division of Responsibilities
  - 1.1.2 Weekly status Report
    - 1.1.2.1 Meeting Dates
  - 1.1.3 Risk Management
    - 1.1.3.1 Accounting for unforeseen illness/unresponsive members
    - 1.1.3.2 Accounting for possible internet outages
    - 1.1.3.3 Assuring validity of references
    - 1.1.3.4 Accounting for risk of lost data
- 1.2 Product Requirements
  - 1.2.1 Software Requirements
    - 1.2.1.1 Draft software requirements
    - 1.2.1.2 Final software requirements
    - 1.2.1.3 Software requirements approval
  - 1.2.2 Hardware Requirements
    - 1.2.2.1 Draft hardware requirements
    - 1.2.2.2 Final hardware requirements
    - 1.2.2.3 Hardware requirements approval
- 1.3 Detail Software Design
  - 1.3.1 Initial software design
  - 1.3.2 Final software design
  - 1.3.3 Software design approval
- 1.4 System Construction
  - 1.4.1 Configure software
  - 1.4.2 Install hardware
  - 1.4.3 Implementation
    - 1.4.3.1 Testing
  - 1.4.4 Deployment

12. Signature Page

*Please provide your electronic signature under your name. Thank you.

UTD Senior Designs Team Member **Ken Lin** 	**UTD Senior Designs Team Member** **Collin Gibson** 
UTD Senior Designs Team Member **Jack Moody** 	**UTD Senior Designs Team Member** **Riki Sasaki** 
Veloxxity LLC (Project Sponsor) **Company Mentor** **Tom Coyle** 	**Faculty Advisor** **Singh, Aishwarya**