**Software Engineering Standards**

**System Requirements and Analysis Specification (RAS)**

**Social Network Analysis with Deep Learning**

**Version 1.1**

Document Number: RAS-001

Project Team Number: A9

Project Team Members: Anish Malhorta (am7161)

Chun Zen Marc Tse (czt210)

Mehmed Mladenov (mm7277)

**REVIEW AND APPROVALS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Printed Name and**  **Title** | **Function**  **(Author,**  **Reviewer,**  **Approval)** | **Date** | **Signature** |
| Team A9 | Author | 2/25/2018 |  |
| Professor | Reviewer | 2/25/2018 |  |
| Team A9 | Author | 3/6/2018 |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**REVISION LEVEL**

|  |  |  |
| --- | --- | --- |
| **Date** | **Revision Number** | **Purpose** |
| Feb 25, 2018 | Version 1.0 | Initial Release |
| March 6, 2018 | Version 1.1 | Fixed Defects from V 1.0, Added Section 7 and 8 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

### TABLE OF CONTENTS

[1. INTRODUCTION 1](#_Toc29376)

[1.1 PURPOSE 1](#_Toc29377)

[2. SCOPE 1](#_Toc29378)

[2.1 IDENTIFICATION 1](#_Toc29379)

[2.2 BOUND 1](#_Toc29380)

[2.3 OBJECTIVES 1](#_Toc29381)

[3. OVERALL SYSTEM OVERVIEW 2](#_Toc29382)

[3.1 CONTEXT DIAGRAM 2](#_Toc29383)

[3.2 ADDITIONAL DESCRIPTIVE Items 3](#_Toc29384)

[4. DOCUMENT OVERVIEW 3](#_Toc29385)

[5. REFERENCE DOCUMENTS 4](#_Toc29386)

[6. BUSINESS REQUIREMENTS 4](#_Toc29387)

[6.1 TECHNOLOGY 4](#_Toc29388)

[6.2 ECONOMICS 4](#_Toc29389)

[6.3 REGULATORY AND LEGAL 4](#_Toc29390)

[6.4 MARKET CONSIDERATION 4](#_Toc29391)

[6.5 RISKS AND ALTERNATIVES 5](#_Toc29392)

[6.6 HUMAN RESOURCES AND TRAINING 5](#_Toc29393)

[7. SPECIFIC REQUIREMENTS (DESCRIPTIVE FUNCTIONAL REQUIREMENTS) 6](#_Toc29394)

[7.1 FUNCTIONAL DESCRIPTIVE DETAILED REQUIREMENTS 6](#_Toc29395)

[7.2 REQUIREMENT USE CASES 8](#_Toc29396)

[8. NON-FUNCTIONAL DESCRIPTIVE DETAILED REQUIREMENTSNON-FUNCTIONAL REQUIREMENTS DEFINITIONS 10](#_Toc29397)

[9. APPENDICES 11](#_Toc29413)

[SCHEDULE TRACKING 11](#_Toc29414)

[DEFECT TRACKING 13](#_Toc29415)

1. **INTRODUCITON**
   1. **Purpose**

The purpose of this document is mainly to define the domain, requirements and analysis for the project and the entirety of its development cycle. The document provides an overview of the objective of our team’s software and other facts pertaining to the use of the software, as well as the technical requirements for the software to be developed and operated successfully.

The intended audience for this document is mainly but not limited to, the client of the software product, the management level personnel involved with the software team, developers of the software, and the testers and end-users of the software.

1. **SCOPE**

There is a lack of simple, efficient and modularized support for determining the efficacy of marketing campaigns by operating on different data sets and performance metrics. The system will address these concerns by allowing the plugging-in of various data sets and the standards used to determine their success, as well as predict future performance. The system will provide an easy-to-use interface for this purpose, which will show results of big data analytics in visual formats, such as GIS maps, charts and spreadsheets. The system, and the algorithm that supports it, will be fine-tuned to support modularity for a variety of data sets that may be used with it as well as the performance indicators used to measure potential success. Such a system would have tremendous benefit for many organizations, simplifying the process of using Deep Learning greatly while also reducing the cost of implement such techniques.

* 1. **Identification**

Social Network Analysis with Deep Learning Requirements and Analysis Specification Version 1.1.

* 1. **Bounds**

The system boundary includes the software of the social network analysis itself, the boundary between end-users (researchers, data scientists, marketers in companies) and the software and the boundary between the software and external databases and data metrics from other companies (social media companies, digital marketing companies, in house data from client).

* 1. **Objectives**

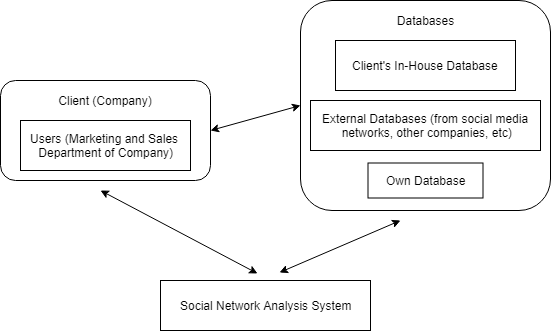
The highest priority objectives are the optimization of the algorithm necessary to perform analytics on data sets, as well as the actual representation of the analysis. As such, the visual component of the system will be crucial – making certain that the customer has the several interchangeable way to view the analytics they will use the system for. Making the system simple to use will also be a high priority deliverable, with a functional and self-descriptive dashboard being necessary. Another important, but lower priority objective will be modularization of the algorithm itself. Being able to use different data sets is an important long-term goal of the system, but not before the structure has been created for a general use case.

Because the system will require constant testing of the algorithm itself on various data sets, the project will run incrementally at each step of production, with texting and requirements analysis being done at every stage of development for maximum productivity and time-line benefit.

Initial Deliverables:

* Project Proposal: 2/08/18
* Software Requirements and Analysis Specification (RAS):
  + - Business & Project Definition: 2/20/18
    - Software Requirements: 3/06/18
    - Complete: 3/20/18
    - Software Analysis Specification: 4/17/18
* Software Project Management Plan (SPMP): 4/05/18
* Software Design Document: 5/01/18
* Presentation: 4/26/18 – 5/03/18

1. **SCOPE**
   1. **Context Diagram**

****

* 1. **Additional Descriptive Items**

Product Functions

* Collect large amounts of data on social network
* Analyze the sentiment of general public towards a specific topic
* Predict a trend towards a topic

User Characteristics

* Product will be used by marketing teams of firms, data scientists and researchers
* End-users can range from being not highly proficient with computers to extremely technical, but in general should know how to interact with a well-designed UI
* End-users is expected to know how to read and input data

Constraints

* If the software product is not allowed to input data from social network companies
* Computer processing power is not enough to compute the result of a large amount of data
* Public databases that are available for data mining may not have sufficient types of data
* Development language can be exploited to create a security loophole

Assumptions and Dependencies

* Availability and quality of the data types to be processed
* The optimization of analysis and prediction algorithm

Requirement Subsets

* User interface and general user-friendly control mechanism could be delayed until the algorithm of processing the data is done

1. **Document Overview**

The document is organized in such a way to first provide a top-level description and analysis of the need for this system as well as the general solutions being used for implementation in section 1 to 3. In section 6, the document further delves into the business applications of the system – the economic and technological drivers utilized by the system in order to be effective. The functional and non-functional requirements of the system are detailed in section 7 and 8 respectively and the analysis of these requirements is provided below in section 9. The manner of system testing in section 10 is described for the purpose of finding defects. In section 11 and 12, the document describes its own examination and validity, as well as the validity of each individual artifact stemming from its original functional requirement. In section 13 and 14, the document details the iterative approach taken in designing the system chronologically as it evolves, with new revisions of this document and their differences being described later. The appendix serves as a method of tracking schedule and defects encountered when designing the system.

1. **Reference Documents**

“Team A9 Project Proposal” Social Network Analysis with Deep Learning, A9, Version 2.0, February 16, 2018.

1. **Business Requirements**
   1. **Technology**

* Big data analytics helps support the business objective in easily scalable metrics for clients to improve upon their strategies. Being able to analyze large quantity of data can help gauge a better picture of consumers’ information and data related to them.
* Machine learning helps give a direction for clients in terms of how they can improve upon their strategies. Being able to predict a trend within social media can help clients create buzz-worthy marketing campaigns and generate useful leads for business.
  1. **Economics**
* Main financial income driver for the system is the motivation for clients to use the software to find the right targeted audience for their products/services, information about their consumers in order to optimize marketing strategies and fine tune companies’ products/services.
* Mode of financial income can be a subscription/service based model, where we act as a “consultant” and provide valuable insights and metrics to companies based on data sets available and optimize marketing strategies for a monthly or annual fee.
  1. **Regulatory and Legal**
* Privacy laws
* System manipulation and fraud
  1. **Market Considerations**
* Established names in the field, for example Google, who already does ad/marketing analytics and provides data to companies.
* Lots of competitions, especially because of the recent rise of popularity in big data and machine learning
* Lots of investors and VCs to provide funding for startups in the field, because of the popularity
  1. **Risks and Alternatives**

|  |  |
| --- | --- |
| Business Risk: | Limited access to data from companies |
| Probability: | Medium |
| How discovered: | Other companies limit use of data |
| Responsible Party: | Management, Legal |
| Status: |  |
| Mitigation Plan: | Deal making, benefits to companies |

|  |  |
| --- | --- |
| Operational Risk: | Project Delays |
| Probability: | Medium |
| How discovered: | Schedule Tracking |
| Responsible Party: | Developers, PM |
| Status: |  |
| Mitigation Plan: | Have meetings to follow up with dev |

|  |  |
| --- | --- |
| Technology Risk: | Not enough quality data |
| Probability: | Medium |
| How discovered: | After getting results from data analysis |
| Responsible Party: | Developers |
| Status: |  |
| Mitigation Plan: | Extract more specific data types |

|  |  |
| --- | --- |
| Economic Risk: | Market too niche |
| Probability: | Low |
| How discovered: | Not many people willing to switch to system |
| Responsible Party: | Designers |
| Status: |  |
| Mitigation Plan: | Adopt more traditional elements |

* 1. **Human Resources and Training**
* Developers should have knowledge in data science, machine learning and statistics
* UI developers should have knowledge in front-end development and UI/UX design

1. **Specific Requirements (Descriptive Functional Requirements)**
   1. **Functional Descriptive Detailed Requirements**
2. Login
3. Dashboard
4. Social Network Tools
5. Graphical Results
6. Data Analysis Algorithm
7. Database
   1. **Requirement Use Cases**
8. **Login**

The login will provide access to the dashboard, which will serve as the main primary point of access for the client in order to use the Social Network Analysis System, and will be the first point of entry after logging in. The login screen will show the user the various projects they have created and allow the user to create new projects for analysis. Upon selecting the project or creating a new project, the user will be brought to the dashboard.

1. **Dashboard**

The dashboard will provide a clear mechanism for the user to upload custom data sets in a new project to satisfy the purpose of modularity (section 2.2). The dashboard will also provide the ability to begin a new project by providing integration with various social media platforms through the use of APIs. This tool will lead to the ‘Social Networks’ context (section 3.1). The dashboard will allow the user to continue analysis on a previous project and will clearly present the user configurable metrics for an analysis they have ran on a set of data. These metrics will be empirical and statistical data deemed useful by the algorithm (section 2) based on key analysis parameters set by the user. The dashboard will allow the user to access visual representations of data in the ‘Graphical Results’ definition (section 3.1).

1. **Social Networks Tool**

The ‘Social Networks’ context will provide the user with a search bar in order to create tags that will provide search parameters in order to build a data set. The tags will be searched for throughout widely-used Social Media platforms such as Twitter, Facebook and Instagram through API integration. The tool will also provide the user with demographics they can use to organize data, such as age, location, density. Each query resulting from specification of a tag or demographic can be ‘added’ to iteratively build a customized data set for analysis. These tags and demographics will provide the basis for analysis with the algorithm in the back end.

1. **Graphical Results**

The ‘Graphical Results’ context will make use of the algorithm heavily and provide the primary method of viewing data analysis. The tool will allow for the user to sort and organize the data based on various insights, tags and demographics the algorithm has determined to be pertinent toward the data set. The tool will make use of GIS maps, charts (using D3), and the other visual representations of data specified (section 2). The tool will allow exporting of raw data used to build these visuals, and will also allow the user to export the graphics themselves for outside use. The tool will provide a mechanism with which to compare analyses with other projects they may have created. Correlation and drawing of insights with respect to multiple projects is left to the user—the tool will simply provide a way to view graphical representations of multiple data sets they have saved in their account for side-by-side viewing.

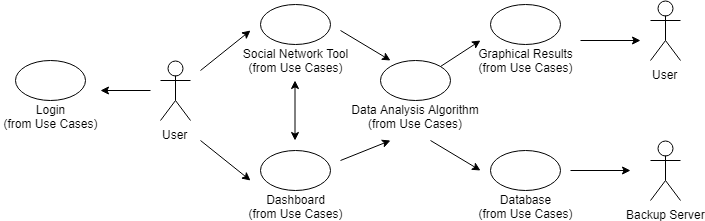
1. **Data Analysis Algorithm**

The algorithm as specified will make use of tags and demographics set by the user in analysis but will not refer only to those guiding principles. The algorithm will be built adaptively in order to use deep-learning techniques for Big Data analysis on vastly different sets of data. As such, the algorithm must provide non-trivial analysis based on trends that may not be prevalent to the user. The algorithm must provide various perspectives with which the user can view the data in the ‘Graphical Results’ and ‘Dashboard’ contexts (sections 2, 7.2.4, 7.2.5)

1. **Database**

The database will store user information, projects, analysis parameters and the results of running the algorithm. The database will not store graphical results built in the ‘Graphical Results’ context in order to save space and time, as an optimization trade-off. The charts and visual representations of analytics will be built in real time with the idea that users will be editing the points of focus in this analysis for each project multiple times.

* 1. **Use Case Diagrams**



**Use Case Descriptions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Login** | | | |
| **Description** | Primary access point for user to use the system | | |
| **Pre-Condition** | Permission to gain access | | |
| **Flows** | **Basic or Normal Flows** | 1 | User passing a lock mechanism to gain access |
| 2 |  |
| 3 |  |
|  | **Alternative Flows** | 1 |  |
| 2 |  |
| 3 |  |
| **Post Conditions** | Allows users to either create new projects or gain access to saved projects, or create new data sets | | |
| **Special Requirements** | None | | |
| **Extension Points** | Leads to Dashboard or Social Network Tool | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Dashboard** | | | |
| **Description** | Allows users to upload custom data sets/Begin new project/Continue from saved projects | | |
| **Pre-Condition** | Have already passed through login mechanism | | |
| **Flows** | **Basic or Normal Flows** | 1 | Transition from login |
| 2 | Create a blank project |
| 3 | Open past project |
|  | **Alternative Flows** | 1 | Import custom data sets from Social Network Tool |
| 2 | Go to Social Network Tool to create data sets |
| 3 |  |
| **Post Conditions** | Construct clear configurable metrics to users for analysis | | |
| **Special Requirements** | None | | |
| **Extension Points** | Leads to Social Network Tool or Data Analysis Algorithm | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Social Networks Tool** | | | |
| **Description** | Allows user to create tags that provide search parameters for social networks to build a data set | | |
| **Pre-Condition** | Have already passed through login mechanism/Created a blank project and attempt to build data sets | | |
| **Flows** | **Basic or Normal Flows** | 1 | Transition from login |
| 2 | Transition from dashboard to create data set for use |
| 3 | Modify data sets for projects |
|  | **Alternative Flows** | 1 | Failed to build data, retry |
| 2 |  |
| 3 |  |
| **Post Conditions** | Data set is built for projects to use/Failed to build data sets, requires more specific search parameters | | |
| **Special Requirements** | Search parameters must be feasible for social network API | | |
| **Extension Points** | Transition to Data Analysis Algorithm or Dashboard | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Analysis Algorithm** | | | |
| **Description** | Make use of tags and data sets to generate trends and useful metrics for users | | |
| **Pre-Condition** | Project has been created to go through analysis | | |
| **Flows** | **Basic or Normal Flows** | 1 | Transition from dashboard/social network tool to do analysis |
| 2 |  |
| 3 |  |
|  | **Alternative Flows** | 1 | Data analysis has failed |
| 2 | Back to social network tool/dashboard to change metrics |
| 3 |  |
| **Post Conditions** | Notification to let users know whether analysis has failed or not/Provide instructions | | |
| **Special Requirements** | Data sets are specific enough to generate useful metrics | | |
| **Extension Points** | Transition to Graphical Results and Database | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Graphical Results** | | | |
| **Description** | Provide visual data based on various insights, tags and demographics | | |
| **Pre-Condition** | Data analysis algorithm has been run | | |
| **Flows** | **Basic or Normal Flows** | 1 | Data analysis produces visual data based on metrics input |
| 2 | Transition back to dashboard to edit project for other analyzes |
| 3 | Allows users to download results |
|  | **Alternative Flows** | 1 |  |
| 2 |  |
| 3 |  |
| **Post Conditions** | Graphical results need to be visually interpretable/Lets users know when analysis fails | | |
| **Special Requirements** | Data analysis has been executed correctly | | |
| **Extension Points** | Back to users | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | | | |
| **Description** | Store data sets, projects, analysis parameters and results | | |
| **Pre-Condition** | Data analysis algorithm has been run | | |
| **Flows** | **Basic or Normal Flows** | 1 | Data analysis produces results to be saved |
| 2 | Users save data to database |
| 3 |  |
|  | **Alternative Flows** | 1 | Automatically save projects when user is in dashboard |
| 2 |  |
| 3 |  |
| **Post Conditions** | Provides indication that data has been saved | | |
| **Special Requirements** | Have access to backup server | | |
| **Extension Points** | Back to users | | |

1. **Non-Functional Requirements Definitions**
2. **System Capabilities**

The platform will be cloud based and saved on our servers, which would make it available for users with any hardware characteristics to have access.

1. **Security**

The platform will provide the users with the ability to control the access to their own data. This includes controlling who may view and alter data uploaded to the platform. The database would be encrypted with Advanced Encryption Standard (AES), which is one of the standard encryption algorithms by the U.S. Government. Also, extra layer of protection will be added by providing users with Secure Hash Algorithm-2 (SHA-2). Two-factor authentication (2fa), will be added as optional configuration for users with sensitive data uploaded to our database. Supported 2fA third party apps will be limited to *Google Auth, Authy and Duo*, since these are the most trusted authentication apps.

1. **Performance/Availability**

The database would not save any parameters for developing visual results. These parameters will be input by the user and/or calculated real time in order to save space on the platform servers, which would allow increased space usage for detailed searches, with the cost of time. However, the platform uses several of the fastest sorting algorithms with worst case time complexity of O(n log n), which compensates the time used by the parameters.

1. **Recoverability/Backup**

The user would be provided with storage on our cloud servers to save their data. The platform will also automatically make daily backups of this data on an offline server and provide users with a private key or a secrete seed, so they can retrieve their data securely on other account in case of hack. If the data is retrieved on another account, it would be deleted from the old one. If any of these keys are lost, the data becomes irretrievable. The backups will be synchronized daily and be available for downloads on users’ devices.

1. **Analysis**
2. **System Test Play Requirements**
3. **Qualification Provisions**
4. **Requirements Traceability**
5. **Evolution of the RAS**
6. **Rationale**
7. **Notes**
8. **Appendices**

**Schedule Tracking**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Business & Project Definition | Anish | 4 Hours | 6.5 Hours | 2.5 Hours |
|  | Mark | 4 Hours | 6.5 Hours | 2.5 Hours |
|  | Mehmed | 4 Hours | 6.5 Hours | 2.5 Hours |
|  | Summary (People Hours) | 12 Hours | 19.5 Hours | 7.5 Hours |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Software Requirements | Anish | 3 Hours | 5 Hours | 2 Hours |
|  | Mark | 3 Hours | 5 Hours | 2 Hours |
|  | Mehmed | 3 Hours | 5 Hours | 2 Hours |
|  | Summary (People Hours) | 9 Hours | 15 Hours | 6 Hours |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Complete | Anish | 7 Hours |  |  |
|  | Mark | 7 Hours |  |  |
|  | Mehmed | 7 Hours |  |  |
|  | Summary (People Hours) | 21 Hours |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| Software Analysis Specification | Anish | 10 Hours |  |  |
|  | Mark | 10 Hours |  |  |
|  | Mehmed | 10 Hours |  |  |
|  | Summary (People Hours) | 30 Hours |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| SPMP | Anish | 8 Hours |  |  |
|  | Mark | 8 Hours |  |  |
|  | Mehmed | 8 Hours |  |  |
|  | Summary (People Hours) | 24 Hours |  |  |

**Cumulative**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
|  | Anish | 32 Hours |  |  |
|  | Mark | 32 Hours |  |  |
|  | Mehmed | 32 Hours |  |  |
|  | Summary (People Hours) | 96 Hours |  |  |

**Defect Tracking**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Business & Project Definition | Anish | 7 | 12 | 5 |
|  | Mark | 8 | 12 | 4 |
|  | Mehmed | 6 | 12 | 6 |
|  | Summary (Avg.) | 7 | 12 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Software Requirements | Anish | 6 |  |  |
|  | Mark | 12 |  |  |
|  | Mehmed | 5 |  |  |
|  | Summary (Avg.) | 8 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| RAS Complete | Anish | 16 |  |  |
|  | Mark | 16 |  |  |
|  | Mehmed | 14 |  |  |
|  | Summary (Avg.) | 15 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| Software Analysis Specification | Anish | 8 |  |  |
|  | Mark | 12 |  |  |
|  | Mehmed | 11 |  |  |
|  | Summary (Avg.) | 10 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
| SPMP | Anish | 11 |  |  |
|  | Mark | 5 |  |  |
|  | Mehmed | 9 |  |  |
|  | Summary (Avg.) | 8 |  |  |

**Cumulative**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Artifact or Deliverable | Who (Individual or Team) | Estimated | Actual | Difference |
|  | Anish | 48 |  |  |
|  | Mark | 53 |  |  |
|  | Mehmed | 45 |  |  |
|  | Summary (Avg.) | 49 |  |  |