Data Mining and Integration of Real-time Sensory Information in a Cloud Framework

By
Mansoor Rehman
NUST201362718MSEECS60013F

Supervisor
Dr. Safdar Abbas Khan
Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Information Technology (MS IT)

In
School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(July 2017)

Approval

It is certified that the contents and form of the thesis entitled "**Data Mining and Integration of Real-time Sensory Information in a Cloud Framework**" submitted by **Mansoor Rehman** have been found satisfactory for the requirement of the degree.

Advisor: **Dr. Safdar Abbas Khan**

Signature: ____________________

Date: ____________________

Committee Member 1: **Dr. Sohail Iqbal**

Signature: ____________________

Date: ____________________

Committee Member 2: **Dr. Omar Arif**

Signature: ____________________

Date: ____________________

Committee Member 3: **Dr. Anis ur Rahman**

Signature: ____________________

Date: ____________________

Abstract

In present times data is a ubiquitous resource, and its analysis to deduce highly selective and relatively more useful bits of information often holds great significance. However, in most advanced systems, the deduction of useful insights from a large dataset denotes only part of the required functionality; because the information derived from the input data needs to be integrated back into the system to increase the overall efficiency of operations in terms of feedback-based predictive capabilities or intelligent decision making.

Preferably, this integration of information in a dynamic environment needs to be carried out autonomously with minimal human intervention in order to gain maximum advantage of adopting this approach. For this purpose, agent-based systems are most suited as they typically offer features such as adaptability, pro-activeness, negotiation and learning with a certain degree of autonomy to achieve pre-defined goals.

Thus, the aim of the research was defined as the integration of information derived from real-time inputs in aiding intelligent decisions with a significant degree of autonomy. For this purpose, long term management of nutrition and health needs in modern day relief camps was chosen as a practical use case since it is an often over-looked subject compared to immediate disaster management where a lot of work has been done in developing advanced systems. Moreover, given the current scope of things where over 40 million people have been displaced world over as a result of natural disasters, conflicts or violence (with more than 1.5 million displaced persons registered in Pakistan alone), the research holds great significance.

In order to enhance the stability, cost-effectiveness and reliability, the use of cloud services – in particular infrastructure as a service (IaaS), is highly desirable. Hence, this research endeavor tends to focus on methods for analysis of real-time sensory data, and the development of an automated mechanism for integrating the select pieces of information within a cloud-based system using multi-agents.

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Mansoor Rehman**

Signature: ________________

Acknowledgment

Beginning with Allah almighty, I wish to thank my supervisor, Dr. Safdar Abbas Khan for his encouragement and having faith in me for completing this research endeavor. Many thanks to Dr. Peter Bloodsworth for providing me with initial direction and valuable suggestions for this thesis.

I also immensely appreciate the support I received from my committee members Dr. Sohail Iqbal, Dr. Anis ur Rahman and Dr. Omar Arif. Here, I would also gratefully acknowledge the support of my wife Tooba Samad who has always been source of great help and motivation throughout the completion of this dissertation.

Finally, many thanks to my parents who have always inspire me from day one. I would like to especially applaud my mother for her constant support and for always being there when I needed a listening ear. Thank you.

Dedication

With immense gratitude and love, I would like to dedicate this thesis to my parents; Mr and Mrs Malik Iqbal Nawaz who have been source of inspiration and encouragement to me throughout my life.

Table of Contents

List of Figures

Chapter 1

Introduction

This chapter introduces the core concepts forming the basis of this research exercise. It provides the background and motivation for this study, along with a brief description of working methodologies involved.

1.1 Introduction

In present times data is a ubiquitous resource, and its analysis to deduce highly selective and relatively more useful bits of information often holds great significance. However, in most advanced systems, the deduction of useful insights from a large dataset denotes only part of the required functionality; because the information derived from the input data needs to be integrated back into the system to increase the overall efficiency of operations in terms of feedback-based predictive capabilities or intelligent decision making.

Preferably, this integration of information in a dynamic environment needs to be carried out autonomously with minimal human intervention in order to gain maximum advantage of adopting this approach. For this purpose, agent-based systems are most suited as they typically offer features such as adaptability, pro-activeness, negotiation and learning with a certain degree of autonomy to achieve pre-defined goals.

Thus, the aim of the research was defined as the integration of information derived from real-time inputs in aiding intelligent decisions with a significant degree of autonomy. For this purpose, long term management of nutrition and health needs in modern day relief camps was chosen as a practical use case since it is an often over-looked subject compared to immediate disaster management where a lot of work has been done in developing advanced systems. Moreover, given the current scope of things where over 40 million people have been displaced world over as a result of natural disasters, conflicts or violence (with more than 1.5 million displaced persons registered in Pakistan alone), the research holds great significance.

In order to enhance the stability, cost-effectiveness and reliability, the use of cloud services – in particular infrastructure as a service (IaaS), is highly desirable. Hence, this research endeavor tends to focus on methods for analysis of real-time sensory data, and the development of an automated mechanism for integrating the select pieces of information within a cloud-based system using multi-agents.

1.2 Motivation

While extensive efforts have been made towards the development of highly efficient disaster management systems, sustainable software solutions for long-term relief camp management are potentially non-existent. Furthermore, most commercially available software either caters for leisurely camp management or remote site management.

Therefore, there is a crucial need for IT-assisted relief camp management systems which can:

- Track long term relief efforts.
- Have predictive/self-learning capabilities to assess basic needs of a (large) displaced population.
- Can aid relief camp management agencies with administrative decisions.
- Ensure considerably better utilization of limited resources over extended time periods.
- Function with minimal human intervention.

1.3 Problem Statement

Basic purpose of this research exercise is to make use of data analytics and cloud computing technology for handling sensory data streams in unstructured/probabilistic environments in order to extract meaningful insights; and integrating information within a cloud platform to facilitate decision-making using a multi-agent system. Thus, making evidence-based predictive decisions suited to the real world problem of smart management of relief camps over extended periods of time.

1.4 Challenges

The major challenges involved in this research activity include the integration of different paradigms i.e. data mining, cloud computing and multi-agent technology in order to support a set of key administrative decisions for long term relief camp management. The task also involved selection of these administrative aspects of relief camp management which would be focused upon during this exercise.

As already mentioned, commercially available relief camp management tools mostly cater for leisurely camping ventures or remote site management for specified time periods without emergency resource constraints; whereas the case explored via this research endeavor holds significantly greater proximity to real world problem where human lives are at risk.

Furthermore, intelligently catering for variables such as population increase, spread of diseases, lack of shelter, and exhaustion of natural sources of water in trying circumstances for a large group of displaced persons makes designing and development of such system a difficult task. Additionally, multi-agent systems operating in real environments are mostly dependent on continuous information in order to construct a global view of the environment, perform active learning for timely decisions and plan and execute actions accordingly. Therefore, effective integration of useful insights resulting from mining of data inputs are all major concerns that needed to be addressed.

1.5 Research hypothesis and questions

Following research hypothesis and questions were deduced based on the problem statement and challenges mentioned above.

1.5.1 Research hypothesis

Useful information obtained from real-time sensory data can be integrated within a computationally efficient cloud framework to aid evidence-based decisions suited to long term relief camp management.

1.5.2 Research questions

- How to effectively integrate useful insights determined from data mining practices?

- How effectively can the intelligent behavior of a multi-agent system be facilitated through the use of selective, more structured information?
- How to carry out effective resource allocation for food, water and medicine within a large group of displaced people with minimal human intervention over extended periods of time?
- How does the developed system fare against existing practices of relief camp management?

1.6 Methodology

A multi-agent based system would be developed to reside over a cloud framework to autonomously aid smart management of relief camps based on useful insights extracted from periodically collected data. The relief camps and supply warehouses would be geographically simulated via the GAMA platform with parameters for population and basic nutritional requirements estimated as per international humanitarian guidelines.

Summarily, the developed system would be based on the following fundamental principles:

- Long term management of relief camps with significant autonomy.
- Dealing with a continually evolving environment.
- Efficient rationing of limited resources.

Therefore, the multi-agent system will incorporate detection, monitoring, prediction, planning and management agents. These agents may provide required functionality alone or by creating other agents to perform a desired goal, depending on scale and complexity of a problem. Moreover, data mining algorithms would be applied to sensory data from relief camps and resulting information would be integrated back into the system **evidence-based** predictive decisions.

1.7 Objectives

Major objectives of this research exercise include application of data mining algorithms on sensory data collected periodically from simulated relief camps; and effective integration of resulting information within developed multi-agent system to aid intelligent administrative decisions. Therefore, the aim is to streamline long-term relief camp management efforts through the use of predictive methodologies and multi-agent technology for effective resource allocation for a large community of displaced persons. Moreover, the deigned system would be implemented over a cloud framework for increased robustness and avoidance of single point-of-failure problems.

1.8 Thesis outline

Rest of the thesis incorporates following chapters:

- Chapter 2: Literature Review. This chapter sheds details on various modern day applications of data mining, existing disaster and relief camp management systems, and the use of multi-agents and cloud computing to address real-world problems.
- Chapter 3:Proposed Architecture. This chapter explains different modules of the system and explains how they interact with each other to provide desired outcome.
- Chapter 4:Implementation. This chapter includes implementation details of the system.
- Chapter 5: Results and Analysis.
- Chapter 6: Conclusion and Future Work.

Chapter 2

Literature Review

The following chapter explores generic aspects of data mining, information extraction, and relief camp management; and establishes their relevance to the current research endeavor including comments on existing practices in long term relief camp management.

2.1 Data Mining and its Applications

Data mining is primarily a process of identifying useful patterns or insights, uncovering hidden relations between multiple attributes of a given dataset. Most datasets available today can be broadly be categorized as structured or unstructured. Interestingly, the growth of data in both categories over the years has followed distinctly different patterns. While the increase in the amount structured data has been mostly linear, the surge in unstructured data has followed an exponential progression [1]. By a safe estimate, nearly 90% of all data available today is unstructured, which essentially means that it lacks a certain degree of organization, cannot be augmented systematically and common search and analysis algorithms cannot be implemented on it directly without pre-processing. This monolithic volume of unstructured information is often synonymous with the term 'Big Data' [2] [3].

In such a scenario, techniques for data analysis, categorization and classification have inevitably evolved to meet future needs. According to Hsinchun Chen et al. [4], the terms business intelligence and big data analytics have advanced from DBMS and structured content also known as 'BI&A 1.0' frameworks to more dynamic 'BI&A 2.0' and 'BI&A 3.0' frameworks – the former (2.0) valid for web-based analytics, sentiment analysis, text mining; whereas the latter (3.0) valid for data content pertaining to mobile devices and the

Internet of things (IoT) ecosystem. The common denominator in all these frameworks, however, remains the objective of scrutinizing large data segments for pieces of information that can be used to streamline underlying processes, reduce operational delays, promote e-commerce applications etc. Aside from the business aspect, one of the goals is also to facilitate people in general, and afford them time to focus their efforts towards smarter goals as opposed to redundant tasks.

For example, a stock broker would prefer that a simple smartphone application would help him analyze the stock prices for share value holdings for the past six months to make quick sales and investments while remaining mobile. Similarly, home-based consumers for packaged and non-packaged household items would want to know the best available bargains to make necessary purchases. Alternatively, small and large scale business enterprises would like more accurate identification of their prospective target market in order to promote their products more effectively. The same applies to not-for-profit organizations offering humanitarian services; whereby they are interested in especially highlighting underprivileged segments of the population for providing necessary assistance.

Based on the abovementioned discussion, nearly the same amount of research is being carried out towards:

- Development of more advanced algorithms for data mining and analysis
- Finding new applications for data mining

In fact, similar to most scientific practices, the latter is truly driving the former. Another interesting aspect to consider here is pertaining to the effective integration of knowledgeable insights from large datasets into complex systems as a feedback mechanism for realizing areas of improvement. This can be in the form of informed/evidence based decision-making or significantly more precise predictive capabilities. Therefore integration of information resulting from data mining techniques is equally important to bring about improvements in overall performance, effectiveness and intelligence of complex systems [5] [6].

It can be safely stated that increasing inclination towards data mining and greater need of finding undiscovered relations within data has led to the development of smarter systems [7]. Moreover, the development of such systems encourages the use of data mining practices in combination with other technologies which makes it an interesting research area. One such combination involves data mining and data analysis implemented through cloud computing platforms [8] [9], which proves to be an advantageous approach as it benefits users with cost effective, efficient and reliable solutions that help them to make knowledge driven decisions. Furthermore, cloud services help to address the problem of handling computationally intensive tasks through seamless scalability of resources. These obstacles are often faced while analyzing large pools of data to find potential hidden insights. Therefore, the combination of data mining and cloud computing proves to be highly valuable in fields such as medicine, genetics, e-commerce, agriculture, manufacturing industry etc. and holds equally valid for organizations around the globe ranging from small companies to huge enterprises [10].

Besides cloud computing, data mining techniques are also being implemented in conjunction with multi-agent systems as these techniques improve decision making process. A multi-agent system is a network of individual agents working together to achieve a global goal. Therefore, characteristics like communication, learning, decision making, negotiation and problem solving become essential features that should be present in every multi-agent system. In order to enhance learning feature of a multi-agent system, Hani K. Mahdi et al. [11] has used different data mining techniques to enable agents to get trained through its reaction and communication with environment. For this purpose, different classifiers were used and comparisons were drawn which determine that data mining techniques can help MAS in effective decision making with comparatively high degree of accuracy. In contrast, Longbing Cao et al. [12] have explored how agents can be used for data mining process which includes data extraction, selection, integration and pre-processing. Its large focus revolves around agent driven data mining as it ensures isolation of different data sources, dynamic selection of data sources and data gathering and security. Apart from this several challenges of agent driven data mining has been highlighted which include constant modification of agent behaviors when working in real time and development of systematic approach for complex rules in multi-agent to ensure dynamic incorporation.

2.2 Relief Camp Management

Controlling natural disasters such as earthquakes, floods, hurricanes are difficult for human beings. The uncertainty and complexity involved in minimizing the degree of damage can be done through rapid response and recovery through intelligent allocation of resources. For this purpose effective management of relief resources is considered a crucial element in saving lives or preventing any other damage. Recent, case studies reveal that failure to assign needed resources in timely manner has been a common problem which cost many lives. Therefore, there is a need for better support and management system to address the allocation problems [13] [14].

Mitigation of resource allocation problems also requires proper management of camps after disaster over extended time periods. As these camps play vital role in providing safe and secure environment to the victims, so proper designing and maintenance of camp should be taken under consideration. In addition to this, identification of suitable location of camps, development of check lists and monitoring forms, water accessibility, availability of fuel wood and presence of basic health services are few goals that should be met for effective camp management [15].

2.3 Fundamental Principles of Relief Camp Management

The basic principles of relief camp management are based on key phrase “right to life with dignity” as dictated by the international humanitarian charter. Hence, relief camp management revolves around the provision of basic human needs in terms of food, water, shelter and safety. These factors are primarily categorized under “administrative” needs of an affected population.

Besides administrative factors, social factors are equally important in terms of fulfilling the needs of a displaced population. Protection of cultural values of a community who are adjusting within an alien environment (especially in the case of refugees) is quite a challenging task because the local populace can often be un-accommodating. Similarly, difference in religious inclinations is an equally sensitive topic which can even result in inter-community rifts and requires careful management on part of relief camp coordination teams.

While social factors can only be managed through collective human efforts, the use of modern technology can aid with the smooth management of administrative needs of a displaced population especially when it concerns long term resource management. The key point here is to segregate and manage long term relief efforts as opposed to immediate relief and response associated with disaster management systems.

2.4 Existing Systems for Disaster Management

Efficiency of disaster management systems should be enhanced in order to improve coordination between rescue and relief efforts. Therefore, integration of latest technology in disaster management system for resource allocation can prove extremely beneficial. For this purpose, different management systems were surveyed to review current state of the art.

2.4.1 SAHANA

SAHANA [16] is an open source disaster management systems freely available. It is a web based solution particularly designed to help during relief stage of a disaster. Key features of the system include flexibility, as the system can be scaled up down depending on the requirements, information portability to areas which lack internet connectivity and ensures consistency, security and privacy of data. However, this system lacks timeliness characteristic in terms of response of overall system because data is manually recorded and possible delays during its translation to electronic format may cause someone's life.

2.4.2 Mobile disaster management system using android technology

Jovilyn and Carlos [17] developed an android based application named 'MyDisasterDroid' (MDD). This application uses geographic locations of victims through manual inputs in the application or SMS and then using travel salesman problem as its basis, it calculates optimum routes considering geographic locations as cities and rescuers as travelling salesman. Moreover, different genetic algorithms were also used to determine an optimum route and its parameters were varied to get the best combination for ideal results. Developed application is flexible and can help during response phase of the disaster when time is crucial. However, it fails to address relief resource allocation problems which is also an important element of response phase of a disaster, otherwise poor resource allocation would result in poor response which may cost human lives.

2.4.3 Decision support for disaster management

Response and recovery after any natural disaster are two elements that should be met as early as possible to prevent any other damage. Hence, smart and effective decisions to cope with uncertainty and complexity can produce significant results. For this purpose Erik Rolland et al. [18] proposed a decision support system that uses hybrid meta-heuristics for disaster response and recovery. They have explored scheduling problem during disaster. For this purpose they have proposed resource-constrained project scheduling models. In addition to this they have also developed an efficient algorithm to solve disaster response scheduling problem within limited time constraints. However, their models and algorithms only focuses on allocation of respondent teams at disaster site. Whereas decisions regarding placement of relief camps and supply food, water and medicine are equally crucial.

2.4.4 Emergency resource allocation for disaster response

Mohammed Muaafa et al. [19] has developed an optimization model and an algorithm to provide medical response during an emergency which identifies location of emergency units, dispatching strategies of emergency vehicles for evacuation and transportation purposes and finally number of victims that should be evacuated to each unit. Their major focus was to optimize response time and cost and it also fails to address management problems occurs during allocation of resources.

2.4.5 Optimal Model for Relief Resources Allocation using Dynamic Programming and Spatial Analysis Methods

According to [20] allocation of relief resources should be done prior to the disaster based on risk level. Therefore, a model was presented which combines spatial analysis method and dynamic programming to allocate relief resources to different store houses before occurrence of catastrophic event. It forecasts requirement of each store house based on service location and number of residents in particular area. Moreover, their dynamic program model also identifies residents at high risk areas and calculates the amount of relief resources required by them.

Early assumptions about catastrophic can result in wastage relief resources because food and medicine can be stored for particular time period and after that they can expire. Therefore, stuffing storehouses may seem to be an intelligent proactive approach but it is not a reliable solution. Therefore, there is an immense need of a system that can address allocation and management problems of relief resources simultaneously during a response and recovery phase of a disaster.

2.4.6 The LSS/SUMA Platform

The humanitarian supply management system SUMA [21] was conceived in the 1990s by Pan-American humanitarian agencies for close monitoring of relief goods. It involves registering items received for relief work including their categorization, storage and disbursement to keep both donors/donor agencies, and the requesting authorities apprised of the movement of relief goods.

It was aptly renamed as the "logistics support system" [22], because it was used to primarily manage logistics majorly based on human intervention. While SUMA/LSS offer a great means to adequately store, ration and disburse relief items, there remains a dire need for autonomy. In this context, the ability to timely predict the resource needs of an affected population, and timely disbursement can help save human lives. Furthermore, in times of disaster, system reliability is a major concern and the use of cloud computing technology can sufficiently aid in overcoming problems such as single point of failure and lack of system scalability – especially since the scale of natural or man-made disasters can vary drastically under different circumstances.

Chapter 3

Proposed Architecture

This chapter provides a detailed overview of proposed system architecture using data mining and information integration for managing limited resources for long term relief camp management. It highlights novel algorithms for using insights from simulated sensory data to aid evidence-based decisions suited to long term relief camp management. Furthermore, the use of a cloud computing platform to support the system is also discussed along with the systems various components and interactions between them.

3.1 Proposed Framework

As already stated vide previous chapters, the aim of this research is to present a viable case for innovatively managing selective administrative aspects of relief camp management mainly in terms of food, water and shelter for an affected community of people who have been forced to spend their livelihoods in relief camps owing to natural disasters, armed conflict or political insurgence. Hence, the proposed framework for our system relies on multi-agents for analyzing sensory data from simulated relief camps based on guidelines provided by international humanitarian agencies [23]. Furthermore, the same multi-agents monitor consumption and availability of resources, predicting future needs of a displaced population, and management of shortages faced by affected communities and planning timely disbursement of relief goods to ensure long term survival of these communities; while also preventing the incidence and spread of diseases resulting from malnutrition, lack of sufficient hydration and being exposed to natural elements in the absence of adequate sheltering mechanisms. This framework is diagrammatically represented in Fig 3.1.

3.2 Description of Architectural Modules

The various modules developed to compartmentalize tasks outlined for accomplishing our research objectives have been explained subsequently.

3.2.1 Environment Monitoring Module (EMM)

This module serves as the source of sensory data obtained from simulated relief camps. For purpose of this research, each relief camp is emulated by a relief camp agent (RCA) which relays requirements of food, water and shelter for a relief camp in real time based on periodic inputs received from field teams. Since relief camps can be geographically distant, these requirements are relayed to an environment monitoring agent (EMA) which consolidates them and relays them to the resource management module (RMM). Additionally, the EMM is also responsible for relaying information from the RMM back to field teams through their respective RCAs pertaining to the supply of necessary relief items.

3.2.2 Resource Management Module (RMM)

This module performs the following functions through respective agents as highlighted vide figure 3.1.

i Resource Monitoring

ii Relief Items Supply and Distribution

iii Conflict Management for Relief Camps competing for resources

The resource monitoring agent (RMA) is responsible for receiving sensory inputs from the EMA of the environment monitoring module (EMM) pertaining to the requirements of food, water and shelter for different relief camps. These requirements are registered in a centralized database and communicated to the warehouse facility through the agent tasked with provisioning relief supplies – the supply agent (SA). The supply agent maintains constant communication with the warehousing facility for provisioning relief goods strictly as per the requirement. In case the availability of certain items becomes scarce against growing demand, the RMA is kept fully informed by the supply agent.

Both the RMA and supply agent are overseen by the Conflict Management Agent (CMA) within the RMM. Essentially, when the stock of resources becomes depleted at the warehousing facility, and multiple relief camps are

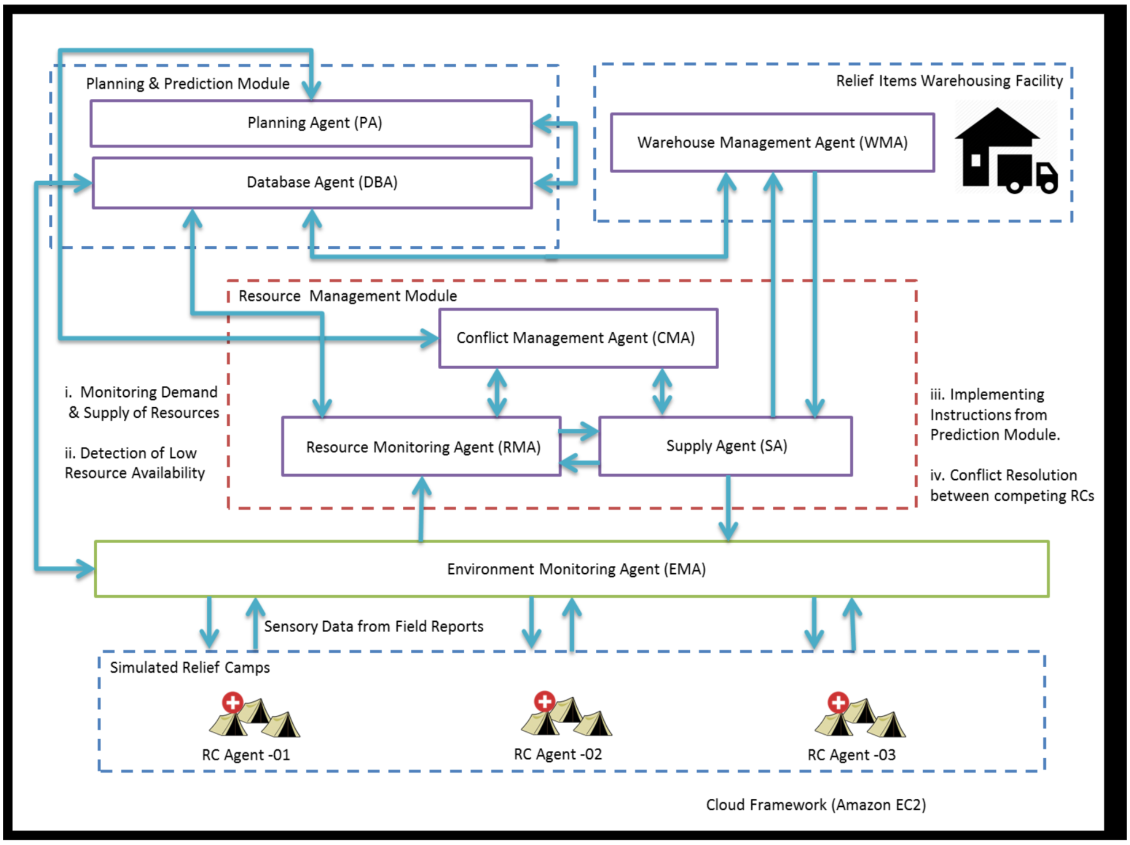

Figure 3.1: A high-level diagram of the proposed system architecture.

competing for the same resources, the CMA's responsibility is to ensure survival of maximum number of people living in relief camps through emergency rationing of services; meaning thereby that thresholds of food, water or shelter items are reduced for all relief camps from normally acceptable levels to emergency circumstances, such that the survival of people in different relief camps is not jeopardized by disproportionate distribution of resources. The algorithms used to accomplish this are detailed at length in the following sections.

Also, the CMA is directly connected with the Planning Agent (PA) from the planning and prediction module of our proposed architecture. Its responsibilities are also highlighted in the subsequent sections.

3.2.3 Warehouse Management Module (WMM)

Before moving towards the explanation of our planning and prediction module employing various data mining techniques to predict key administrative requirements of long term relief camps, an explanation of the warehousing modules function is warranted.

As already mentioned, the warehousing module emulated through a warehouse management agent (WMA), communicates through the supply agent to the RMM. It is primarily responsible for keeping sock of the relief items in terms of inventory management, as well as periodically placing orders for timely replenishing necessary goods. Furthermore, the WMA is also directly connected with the database agent (DBA) for reporting and reconciliation purposes pertaining to the amount of relief items supplied and the amount of relief items received by individual relief camps.

3.2.4 Planning and Prediction Module

The planning and prediction module (PPM) of the proposed architecture forms the core of our research exercise. It basically oversees all sub-functions in terms of WMM, RMM and EMM explained earlier and is closely integrated with them for:

- Analyzing sensory data received from the EMM
- Evaluating responses of the RMM against requests for relief items
- Ensuring active functioning of the conflict management agent (CMA) of the RMM.
- Predicting future needs of relief camps based on resource consumption patterns from RCAs

Sr no.	Agent Name	Functionality
1	Platform Agent (PtA)	Create and initialize other agents.
2	Environment Monitoring Agent (EMA)	Ensures timely correspondence from all relief camps.
3	Database Agent (DBA)	Retrieves data from database and responsible for database connection.
4	Warehouse Management Agent (WMA)	Manages inventory of food, water and shelter items in warehouse.
5	Relief Camp Agent (RCA)	Represents,an individual relief camp and corresponds continually with WMA regarding status of resources.
6	Conflict Management Agent (CMA)	Identify high priority RCs based on affected population and resolve conflicting requests for resources.
7	Supply Agent (SA)	Works for warehouse agent and is responsible for supplying activities.
8	Planning Agent (PA)	Prediction of future needs based on resource demand and supply.

Table 3.1: Description of Agents in Developed Framework

- Providing valuable input based on resource consumption patterns for emergency rationing
- Ensuring proportionate distribution of limited relief goods among the affected population.

Prior to proceeding with the discussion on the above-stated functionalities associated with the PPM, it is pertinent to mention that long term relief camp management is a hugely challenging task where needs of people (especially for survival under harsh circumstances) can mostly be estimated qualitatively owing to a large number of underlying variables involved.

For instance, the average daily water requirements of two relief camps with the same count of individuals can drastically vary based on variance in demographics. Similarly, the presence or absence of natural water sources in spatially distant relief camps operating under the same administrative jurisdiction can also result in resource management problems. In the same manner, the incidence of malnutrition or lack of certain type of micronutrients may affect one community more than the other, simply because displaced communities can often belong to considerably different climates. Therefore, their nutritional needs can vary accordingly.

Nevertheless, these challenges lend further credibility to this research endeavor as a novel architecture has been proposed for the developed system to analyze the problem of long relief camp management on a quantitative basis.

These modules consists of different agents specialized to achieve individual goals of each module and finally interaction among these modules provides overall functionality of system. Following table 3.1 provides the details of agents involved for relief camp management.

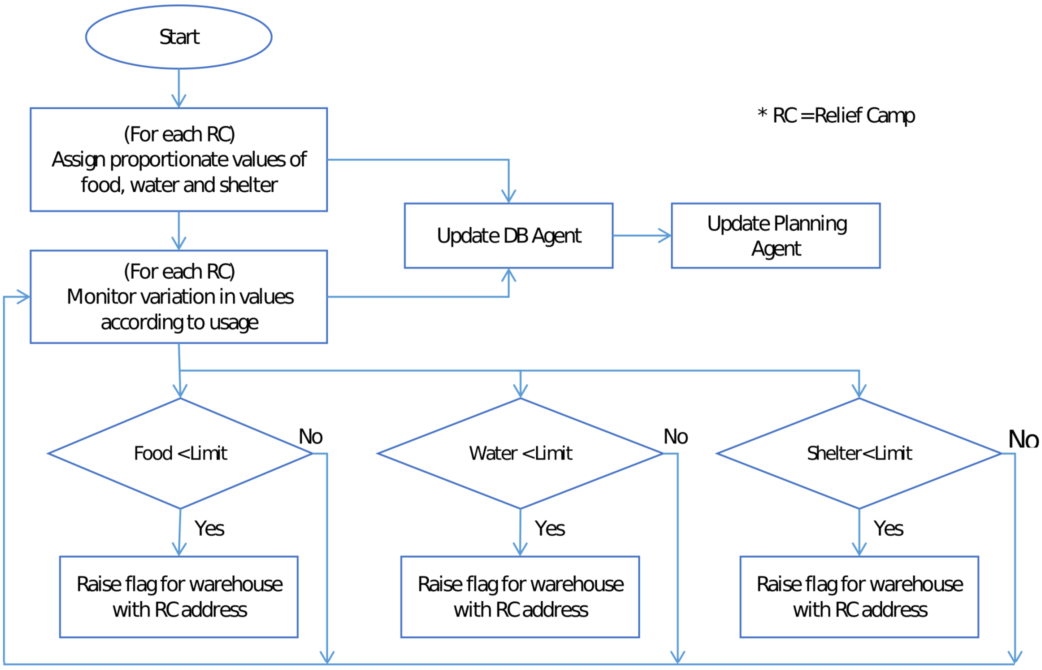

Figure 3.2: Initial Assignment and Monitoring Algorithm

3.3 Procedural Algorithms

Different algorithms were developed during this research exercise for better relief camp management.

3.3.1 Initial Assignment and Monitoring Algorithm

The algorithm begins with assignment of proportionate ranges of food, water and shelter items to each relief camp (RCs) based on their respective populations. Then resource consumption is closely monitored based on reports from the environment monitoring agents. In case the resource consumption falls below the threshold specified by the international humanitarian charter, a flag is raised at the warehouse by the respective RC for sending additional supplies. The database and planning agents remain apprised of these activities at all times. The aforementioned algorithm is summarized in Fig 3.2.

3.3.2 Warehouse Assignment Algorithm

The following figure 3.3 below explains the assignment of resources at warehouse end to the requesting RCs. Once the resource shortage flag ap-

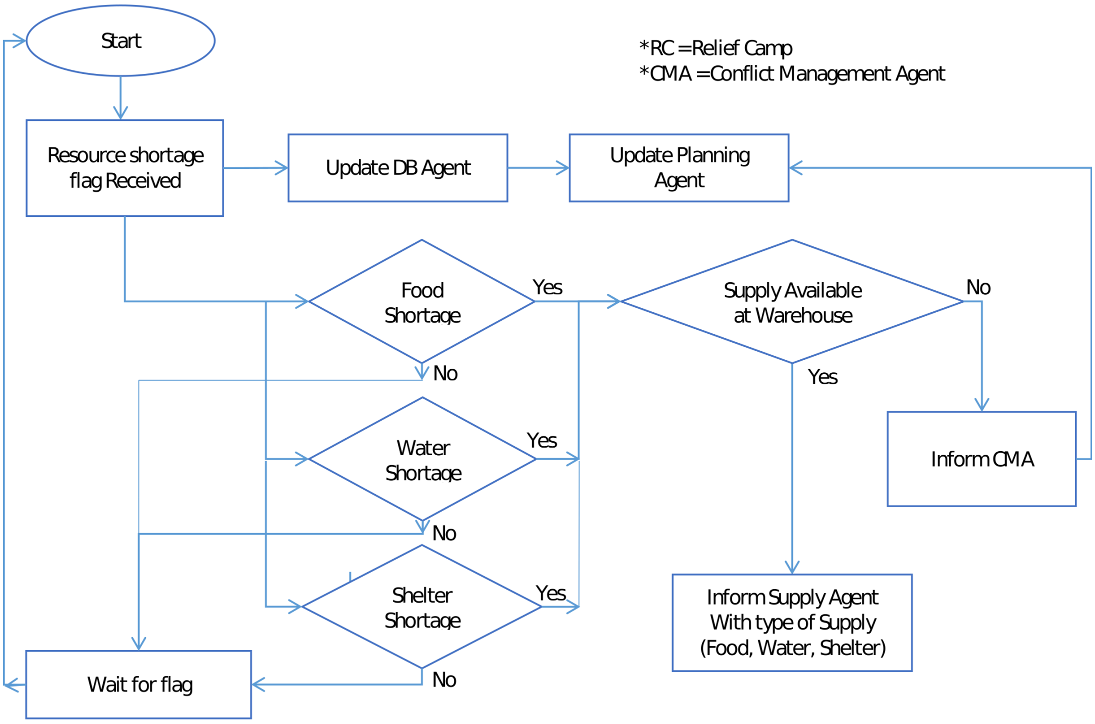

Figure 3.3: Warehouse Assignment Algorithm

pears, the WMA contacts supply agent (explained in section 3.2.3) otherwise conflict management agent plays its role (explained in section 3.2.4) in case there is resource shortage at the supplying warehouse.

3.3.3 Supply Algorithm

Fig 3.4 gives a summary of the supply algorithm. Basically, it sends the required supplies to the requesting relief camps (given the availability of items in stock). The counters for remaining supplies are duly monitored here as well for earliest identification of shortages while further supplies are ordered by the supplying warehouse.

3.3.4 Conflict Management Algorithm

Based on the earlier information, the conflict management algorithm has been designed such that it is invoked once an RC is requesting a particular supply of either food, water or shelter items and the warehouse is also short of stock. In this scenario we have segregated the management of existing resources through the following two cases.

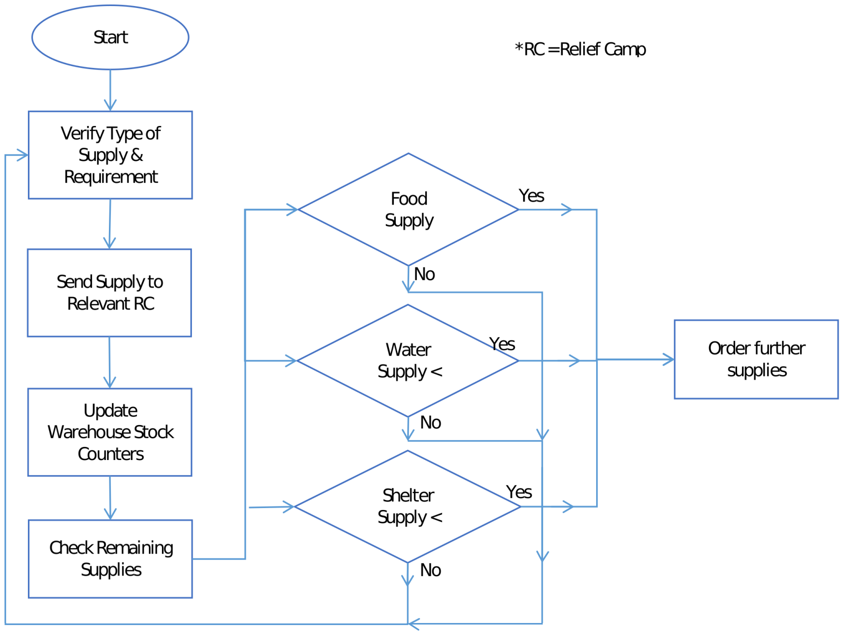

Figure 3.4: Supply Algorithm

Conflict Management Algorithm – Case 01

In the first case, we assume that only one RC is facing a shortage of relief goods which are also out of stock in the supplying warehouse. Given this arrangement, it is imperative to manage existing resources available at other RCs such that the people living in the requesting RC are saved from suffering while additional supplies are awaited by the warehouse.

Essentially, the resource availability at every other RC is calculated and they are subsequently ranked in order of resource availability. The planning agent then decides to route excess from the highest ranked (first) RC to the requesting RC. The counters are updated at each end and supply requirement is verified. In case the requirement is met, the conflict management algorithm is disengaged and the control is reverted back to the warehouse management algorithm by the planning agent. In case the shortage still persists at the requesting RC, the excess from the next highest ranked RC (second RC) is routed to the requesting RC until the resource shortage is met and the warehouse management algorithm is engaged. This process is summarized by Fig 3.5.

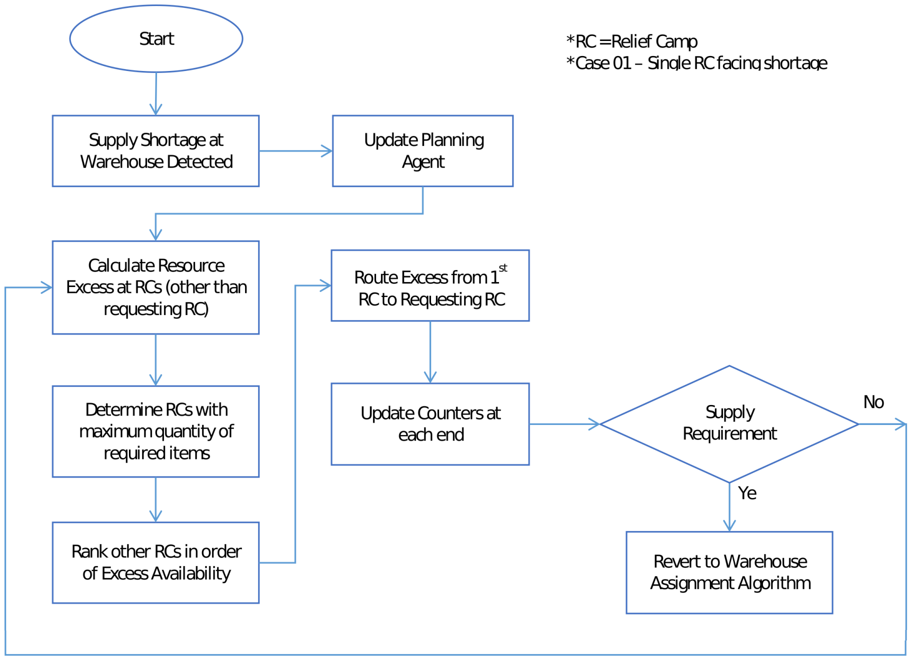

Figure 3.5: Conflict Management Agent Case-1

Conflict Management Algorithm – Case 02

Given the unpredictable variation in the resource needs of an affected population, it is very much possible that multiple relief camps suffer from resource shortages that are also out of stock at the supplying warehouse. Therefore, the second case for conflict management algorithm is designed to cater resource shortages at multiple relief camps.

Similar to the first case, RCs other than the requesting RCs are ranked in order of resource availability. However, routing of excess available at other RCs to requesting RCs is based on an additional metric which is also carefully determined by the Planning Agent. Essentially, the planning agent develops "supply ratios" from the total requirement of requesting RCs such that the excess can be routed in accordance with the requesting RC needs. This is done to avoid equal assignment of excess relief goods against a varying demand. Again, similar to the first case, this process is managed iteratively until the immediate resource shortage is met at requesting relief camps and the routine warehouse management algorithm is invoked by the planning agents. This algorithm is also summarized vide figure 3.6

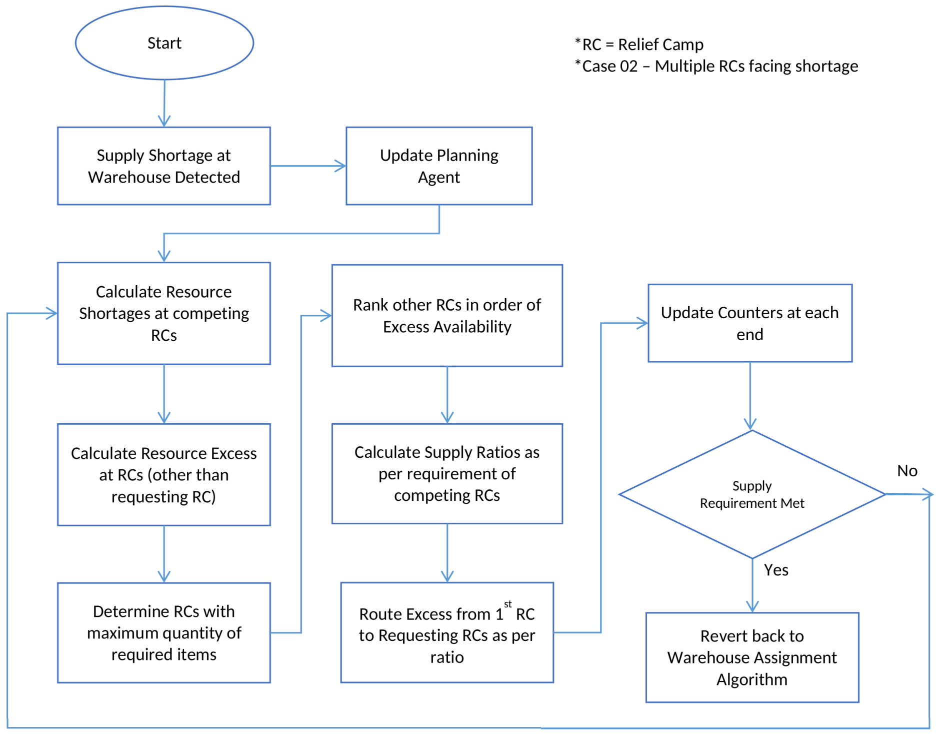

Figure 3.6: Conflict Management Agent Case-2

Chapter 4

Implementation

This chapter explores implementation details of proposed system architecture discussed in previous chapter. It provides overview of development methodology and platform used for implementation purpose. Next, it also discusses the simulation developed for data collection purposes. Towards the end, it provides details of cloud framework for which proposed system was developed.

4.1 Introduction

Implementation of a proposed system concerning real-time information exchange presents a challenging task. This is because developing systems that involve various dynamic factors requires detailed deliberation during the design phase – especially when the slightest of delays can endanger someone's life. Therefore, in order to achieve our implementation goal, we have used combination of different technologies such as data mining, multi-agents and cloud computing.

The use of multi-agent technology was an obvious choice as agents are inherently designed to operate with minimal human intervention. Moreover, they can be especially useful when it comes to being effective in decision making with relatively higher degree of accuracy. Additionally, multi-agent systems can aid with the effective segregation of underlying tasks such that a collective goal can be achieved smoothly.

For prediction purposes certain data mining algorithms had to be chosen such that they facilitate the aforementioned multi-agent system (MAS) in quick decision making. The use of the term 'quick' is imperative here since we are aiming for process automation of long term relief camp management where people's lives are at stake. At the moment, these tasks are largely accomplished through strenuous manual efforts which are inherently prone to human error and unnecessary delays [24]. Therefore the use of the 'eager' learning algorithms had to be given preference which also yielded high accuracy.

Next, the choice of a cloud framework was obvious because of its property of being flexible and to accommodate extensive workloads according to a pay as you go model; so this makes overall system reachable, cost effective and an efficient solution to address problems related to management of relief resources.

After identification of technologies to be used for implementation purposes, next major step to develop system was to select suitable programming language. For this purpose, Java programming language was chosen as it is platform independent, supports real-time coding and programming multi-agents in java are relatively easier than other programming languages.

Finally, for effective decision making and management of relief resources a simulation is developed to emulate data collection in relief camps and predict nutritional and shelter needs of victims based on observed patterns of resource consumption.

4.2 Multi-agent platform

Several, multi-agent platforms were explored and JADE was selected among them as it supports programming of multi-agent according to FIPA specification. Moreover, its flexible nature, support for peer to peer communication among agents, portability feature made it suitable platform for our implementation purposes.

4.3 Data mining Algorithms

In order to ease decision making process and identification of resource consumption patterns to predict nutritional and shelter needs of an affected population different data mining algorithms were used and their results were compared. Following are the details of algorithms used for implementation purposes.

4.3.1 Naïve Bayesian Classifiers

Bayesian classifiers belong to the class of 'eager' predictive learning algorithms. As the name implies, they are based on Bayes Rule to predict or determine the classification of an unknown dataset (test dataset) based on the relation between attributes and classification of a known dataset (training dataset) [25].

In particular, the Naïve Bayes Classifiers are defined as a family of deterministic classifiers based on applying Bayes' theorem where the attribute values are assumed to be conditionally independent of one another [26] [27]. Furthermore, the predictive algorithm is highly scalable and is hence well suited to our implementation scenario where the population living in relief camps can suddenly grow in case of an unforeseen disaster situation.

For purpose of our implementation, we will use the Gaussian model of the Bayesian classifier to determine whether a particular quantity of food, water and shelter items supplied to an affected community residing in multiple relief camps would or would not suffice the needs of the aforementioned. The initial estimates for the supplies would be based on international humanitarian. Hence we would be calculating posterior probabilities of the occurrence of shortages given a particular amount of supplies.
The following equations form the basis of our implementation:

$$\delta = Requirement - Supplies \tag{4.1}$$

Where 'δ' denotes the shortage faced by an affected community,therefore, the standard Bayesian equation can be used in our context as:

$$P(\delta|S) = \frac{P(S|\delta) \cdot P(\delta)}{P(S)} \tag{4.2}$$

Where$P(\delta|S)$ denotes the posterior probabilities we would be using to denote the occurrence of resource shortages while the remaining equation denotes the prior and class conditional probabilities. Since the denominator $P(S)$ is only used to normalize the equation, it may be treated as a constant.
Therefore we are left with the following generalized form of the equation.

$$P(\delta|S) = P(S|\delta) \cdot P(\delta) \tag{4.3}$$

Next, we elaborate the equation in terms of its components, meaning thereby that we need to state the underlying variables involved in the composition of the aforementioned equation.
As stated earlier in this section, the requirement of supplies for each relief camp is denoted by 'S'. If we consider 'S' in terms of each relief camp, we can use the expression – '$S(RC)$'. Since we are dealing with multiple relief camps in our simulation, supplies for each relief camp can be denoted as:

$$S(RC) = \sum_{i=1}^{n} S_i(RC_i) \tag{4.4}$$

Where 'i' denotes the relief camp number.
The above equation is further elaborated for the component $S(RC_i)$ to denote the relief resources monitored for each relief camp in our simulation exercise. Thus, $S(RC_i)$ represents the set of quantities for food, water and shelter items supplied to each relief camp according to pre-determined estimates based on humanitarian grounds.

$$S(RC_i) = \{F_i, W_i, S_i\} \tag{4.5}$$

Since the relief goods supplied based on estimates can vary from the actual demand, for instance if the population in general suffers from extreme malnutrition, they would require additional food. Similarly, if the climate is too hot or humid, water shortages can be frequent and very common. In the same manner, if the area where the relief camps are established experiences frequent rainfall, the dry items of shelters such as tents and blankets can become a scarce quantity.

Hence there is every possibility that widely accepted standards of relief goods required by an affected population can vary drastically based on circumstances. For long term relief efforts, these estimates need constant monitoring and evaluation such that camp managers remain wary of shortages and make timely arrangements such that the affected community does not have to suffer.

Therefore, the use of effective prediction algorithms such as Bayesian classifiers to determine whether a relief camp will suffer shortages despite a seemingly adequate supply of resources is imperative for effective planning of relief efforts over long term.

4.3.2 Linear Regression

Once we are able to predict the probabilities of occurrence of resource shortages in various relief camps, data mining algorithms such as linear regression [28] [29] can help us formulate a minimization function such that the magnitude of resource shortage can be minimal. Essentially, the goal is to minimize the impact of shortages on communities forced to live in relief camps especially with regards to the provision of food, water and shelter items.

Therefore, based on the discussion in the previous section, we revert to our original equation (4.1) to denote resource shortages.

As stated earlier, for each relief camp the Supplied items (based on predetermined estimates can be elaborated as:

$$S(RC_i) = \{F_i, W_i, S_i\} \tag{4.6}$$

Where 'i' denotes the relief camp number. Similarly, the actual requirements can be represented as:

$$R(RC_i) = \{f_i, w_i, s_i\} \tag{4.7}$$

Now there are two ways to formulate our shortage minimization function using linear regression based on whether we want to collectively minimize the quantity of shortages 'δ' in which case we have the following equation;

$$\delta = \sum_{i=1}^{N} \{R(RC_i) - S(RC_i)\} \tag{4.8}$$

Or, alternatively we can look towards minimizing shortage at the level of individual relief camps as;

$$\delta_i = R(RC_i) - S(RC_i) \tag{4.9}$$

If we use the first case (collective shortage), the hypothesis can be modeled as;

$$\hat{h} = \hat{\alpha}_0 + \hat{\alpha}_1 \delta \tag{4.10}$$

Similarly the hypothesis can be modeled in the second case (individual relief camp) as;

$$\hat{h}_i = \hat{\alpha}_{i0} + \hat{\alpha}_{i1} \delta \tag{4.11}$$

In either case, the goal is to determine the coefficients through the sum of least square methods such that the cost function associated with each hypothesis respectively can be minimized. Thus, we can model the cost function 'J' for our simulation as;

$$J(\hat{\alpha}) = \frac{1}{2} \sum_{i=0}^{n} (\hat{h}_i - H_i)^2 \tag{4.12}$$

Once we are able to model the resource shortages faced periodically by people in relief camps as a cost minimization function; solving the equation while maintaining a certain degree of accuracy, we can achieve our goal of timely making arrangements to mitigate the adverse effects of these scarcities. Hence this information can be crucial in assisting relief camp managers for taking timely decisions.

4.4 System Simulation

To simulate data collection in relief camps and predict nutritional and shelter needs of victims based on observed patterns of resource consumption we chose GAMA [30] as simulation platform because it allows for easy handling of large number of heterogeneous agents which form the basis of our system design. Furthermore, GAMA offers support for GIS based simulated environments which offer close proximity to real world scenarios. There are also a number of sample maps already generated within the platform which can be molded to simulate different system requirements and run iterative procedures for testing purposes.

Additionally, GAMA also allows for definition of varying time scales such that simulation models can be run based on varying units of time. This is especially useful in our case where we are monitoring consumption patterns of food, water and shelter items. Thus, in our case, food, water and shelter items are consumed at varying rates. Therefore, the need to routinely monitor the consumption of these items has to be well thought out. For instance, it is commonly perceived that a human being can only survive three days without water whereas the same person can live for up to three weeks without food. Hence, for effective relief camp management, we have built a simulation which would monitor food and water supplies on a daily basis and evaluate requirements of shelter items on a weekly basis using GAMA.

4.4.1 Environment Simulation

In our simulation, the GAMA platform is used to denote a disaster affected population distributed in relief camps. For this purpose the following implementation tasks were highlighted:

- Creating a map in GAMA to emulate a real-world environment with high and low elevation areas and very few houses/buildings.
- Showing safety locations or relief camps on the map
- Showing one major warehouse/relief goods distribution point, where it is assumed that all relief goods are placed. Preferably the warehouse is at a central location such that the distance to each relief camp is the same.
- Depicting resource consumption patterns graphically for each relief camp for ease of monitoring.
- Building similar graphical representation for warehouse.

4.4.2 Design Assumptions

Next, the following algorithm design assumptions were made for simulation purposes:

- The centralized warehouse has a stock for 30 days for 5000 people.
- It takes 7 days to re-stock the warehouse.
- The warehouse sends out supplies for food on a weekly basis to all relief camps.
- Water supplies are sent every day owing to its necessity for human survival.
- Shelter items are monitored every day however supplies are disbursed on a bi-monthly basis.
- Minimum surface area for each relief camp is based on the assumption that $45m^2$ is the minimum required area with garden space for roads, services, shelter and family garden space, but not for livestock grazing.
- Avg. Population Growth Rate within the relief camps is 2.5-3.5
- Population Model for Camp Residents is as follows:
  - 1 Family – 4-6 persons
  - 1 Community – 16 families ∼ 80 Persons
  - 1 Village -4 Communities ∼ 320 Persons
  - 1 Block – 4 Villages ∼ 1280 Persons
  - 1 Sector – 4 Block ∼ 5000 Persons
  - 1 Camp – 4 Sectors ∼ 20,000 Persons
- Modeling Demographic Constraints such that vulnerable persons amongst the population should be given priority for housing near essential camp services such as clinics, distribution centers, community service offices

Apart from this, following key points were also considered for designing of relief camps:

- Camp management in refugee settings is different from camp management in IDP settings.
- Camp management in IDP settings is more difficult.

Figure 4.1: Spatial Representation of Relief Camps

- Key problem is long term assistance.
- Full spectrum of camp management activities.
- Three fundamental principles guiding all humanitarian action in support of IDPs:
  a) Right to life with dignity;
  b) Maintaining a clear distinction between combatants and non-combatants;
  c) Right of IDPs to protection against forceful return to places where their lives, safety, health and liberty would be placed at risk.

4.4.3 Developed Simulations

Based on the abovementioned discussion, Fig 4.1 shows the spatial distribution of relief camps and warehouse in a city simulated in GAMA. Next, the status of shelter items, food and water consumption can also be monitored through bar graphs as shown in fig 4.2 below.

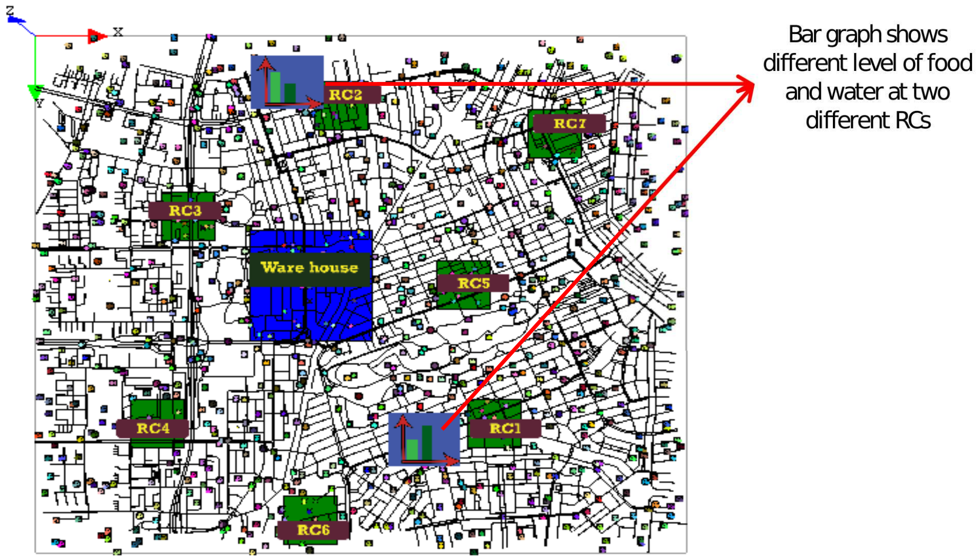

Figure 4.2: Graphical Representation of resource consumption patterns at relief camps

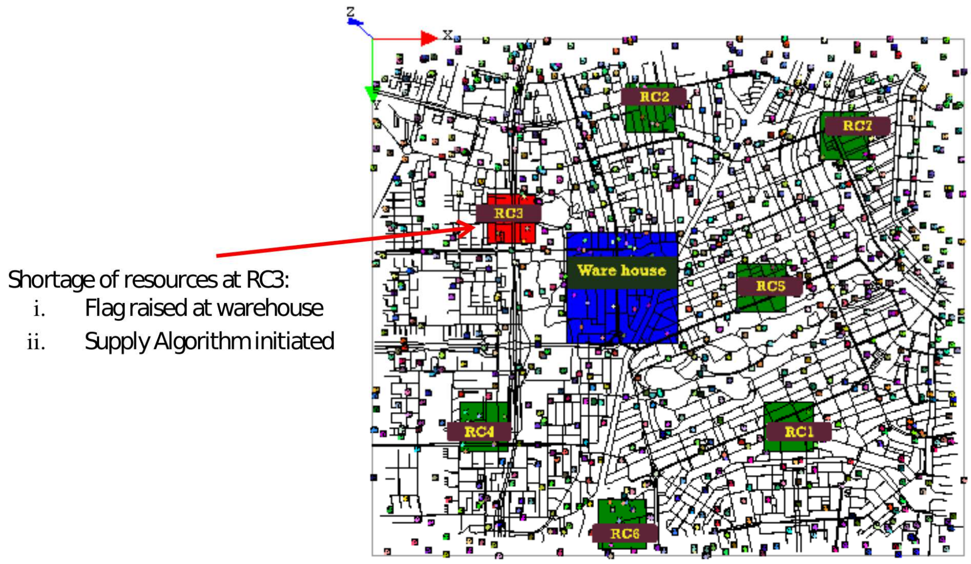

Figure 4.3: Resource Shortage Detected at One Relief Camp

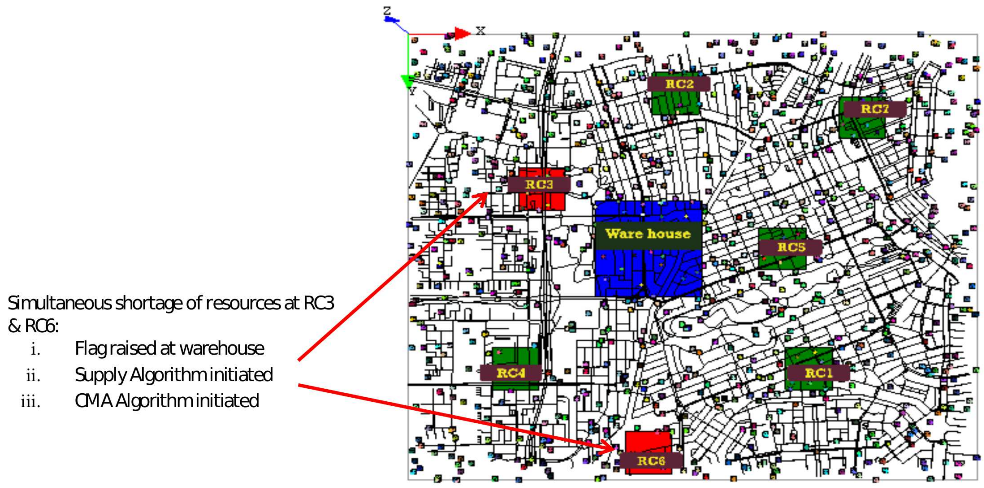

Figure 4.4: Simultaneous Resource Shortage Detected at two Relief Camps

4.5 Cloud Framework

Cloud services have been used to handle extensive workloads that occur in real time. Since, cloud services charge its users based on Pay-as you-go model which makes overall system a cost effective solution. For this purpose, we selected amazon EC2 cloud service provider because it is simple to use, scalability of system capacity changes according to the requirements; which means capacity of the system is scaled up or down automatically according to the workloads and finally its economical as it only charges their customer for the computing space which is actually being utilized [31] [32].

4.6 Conclusion

In this chapter major implementation aspects of developed system have been discussed. First multi-agent platform used for agent's development has been discussed. Next, details of data mining algorithms used for decision making and prediction purposes have been provided. Moreover, details of simulation system which includes distribution of relief camps, monitoring of food, water and shelter items at various relief camps. Finally, cloud framework selected for implementation purposes has been discussed.

Chapter 5

Results and Analysis

The following chapter provides a detailed commentary on the evaluation methods used to test the efficacy of the developed system and the subsequent results obtained. It essentially builds upon the discussion in the previous section highlighting the implementation methodology, and provides the necessary validation for this research exercise on the whole.

5.1 Dataset and test Environment

The dataset used in our exercise was simulated based on the assumptions in the preceding section where a detailed test environment was constructed using the GAMA platform. This comprised of setting up a total of seven relief camps, each housing a combined average of approximately five thousand persons each. Next, the resource monitoring mechanism was setup such that the collective consumption of food, water and shelter items was depicted graphically at each relief camp level.

The monitoring process itself was based on a simple supply and demand algorithm. Each relief camp was initially assigned rations for food, water and shelter based on international humanitarian guidelines. The food items were denoted through 10kg bags comprising of wheat flour, pulses, rice, sugar, salt, cooking oil, spices and dry nuts so that they meet the necessary nutritional requirements in terms of daily caloric and mineral intake of a family of 4-6 persons per week.

Water requirement was defined in terms of gallons/day for each relief camp where estimates were based on the four broad categories of usage – domestic, primary care (medical) centers and community mosques. Consideration for the population demographic has also been factored in for purposes of accuracy because simple logic dictates that water consumption patterns for children between the ages of 4-9 will differ from those of adults falling within the age brackets of 10-18 or 19-65 years. Wastage is also a common occurrence in water distribution therefore it was also factored in as 10% of the overall requirement. Another key factor to consider here is that water requirements – even in disaster hit areas are always managed through a combination of relief supplies and natural resources such as wells, streams, lakes and rivers. This is because water is a key element for survival and relief camps sites are often chosen by camp managers in close proximity to natural water sources in order to meet at least a minimum of 50% of the overall requirements. Hence, we have also used the assumption that natural water sources meet 50% of each camp's requirement, while the remainder is supplied by relief supplies.

Shelter items were defined in terms of requirement for tents and blankets where one large tent is assumed to house 4-6 person (including children) and number of blankets required is equal to the number of person in each relief camp with a 10% contingency. Although setting up relief camps is a gradual process and supply of rations is initiated based on patterns of population growth, for purpose of our simulation we have considered that all relief camps are already populated and an initial supply of items is sent accordingly. Meanwhile the annual rate of increase of population will be at an average of 3.5% collectively across all seven camps. One can argue the effects of this assumption on the results obtained from the system evaluation, however given that the system is designed in such a way as to smoothly accommodate a real-time increase or decrease in population vis-à-vis resource consumption, the underlying assumption is acceptable for having minimal impact on the results. The count of population within relief camps and the corresponding demographic composition is depicted in Figure 5.1 and 5.2 below; whereas the initial requirement of food water and shelter items is depicted in Figure 5.3.

Along with the relief camps, the warehousing facility is set up to represent the various algorithms which govern the distribution of relief items through intelligent supply algorithms; as well as centrally monitoring resource consumption patterns for different relief camps. In this particular scenario, the relief camps represent the nerve center of our system from where the planning, prediction, supply and demand algorithms are implemented through the help of respective agents specified in Chapter 3.

As stated earlier, the developed system is designed for real-time monitoring of relief camps; however for testing purposes different supply schedules for food, water and medicine were defined. This approach was adapted to also factor in the availability of large transport vehicles for carrying out supplies. Food supplies were sent on a weekly basis, water supplies on a daily

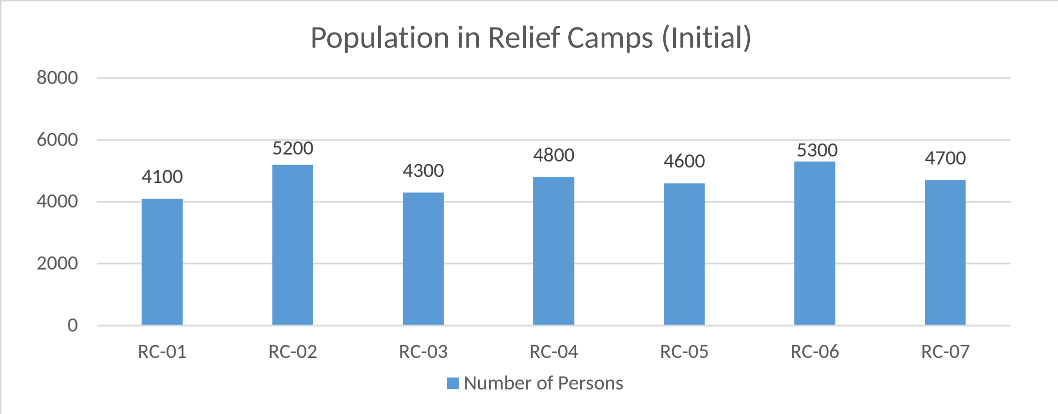

Figure 5.1: Population in relief camps

basis, and shelter items in the form of tents or blankets were supplied on a monthly basis based on the demand estimated by our system. The goal of the prediction exercise using a Naïve Bayesian Classifier and linear regression methods was set each time to minimize resource shortages; however the testing phase revealed another important aspect – ensuring the need to manage excess disbursement of resources. This is because the supply of relief items is limited and excessive supplies can result in shortages where there is an actual need. Another factor to consider here is that relief items for food in particular comprise of perishables, which have pre-defined shelf lives that can decrease significantly once it is taken outside the control environment generally maintained within warehousing facilities. Therefore, the designed algorithms were thoroughly tested to minimize both resource shortages and resource excesses in our research exercise.

5.2 Evaluation Criteria and Results

After qualifying necessary parameters for our simulated dataset and the test environment, the evaluation methodology was decided to test the system performance for effectively predicting resource consumption of food, water and shelter items. Since the distribution schedule varied from one resource to another (as pointed out in the preceding section), separate projections for resource consumption were developed for each resource.

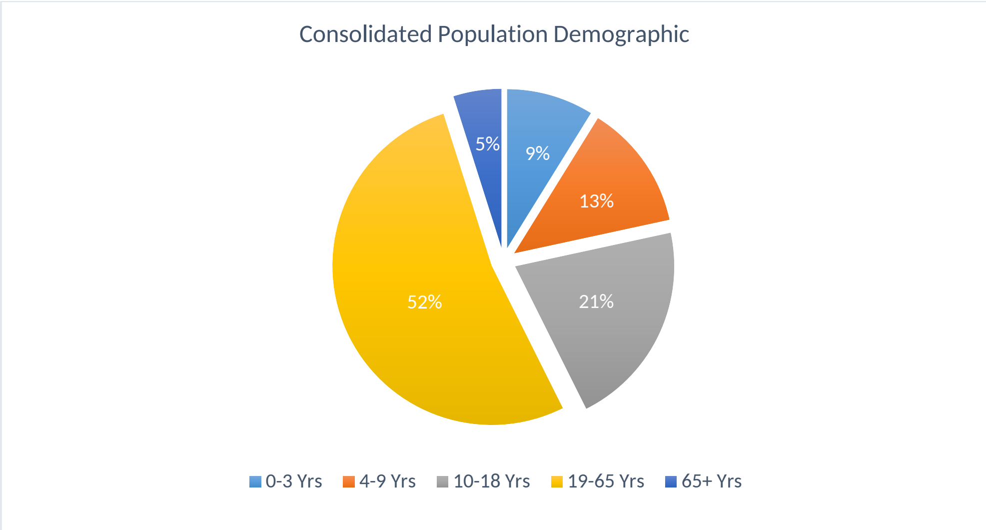

Figure 5.2: Demographic distribution of population living in Relief Camp

5.2.1 Management of Water Resources

Water is a key resource necessary for human survival. As stated vide earlier sections, a human being might survive without food for up to 21 days however without water, the same person might not be able to live past 3 days – even lesser if the climate is hostile. Therefore, we begin the evaluation exercise by simulating actual water requirements projected over a six-monthly period considering daily supply for all seven relief camps. This is represented in figure 5.4 where for purpose of clarity selected values were plotted over fortnightly intervals. In order to carry out a valid system evaluation, we have mapped the collective actual requirement along with our system predicted values, and a base line denoting the absolute minimal required as 95% of the minimum actual required supplies over the six month period. This is depicted in the following figure 5.5.

As evident from the graph above, the predicted values begin to conform to actual population needs as time progresses. Another encouraging factor which can be derived from the figure above is that over the six month projection, the predictive values never drop below the emergency threshold defined at 95% of the minimum actual requirement. Moreover, contrary to the seemingly large difference in terms of thousands of gallons/per day between actual and predicted values, it must be taken into consideration that this difference is averaged over all requirements of all relief camps and in fact the maximum error percentage observed is less than 8%.

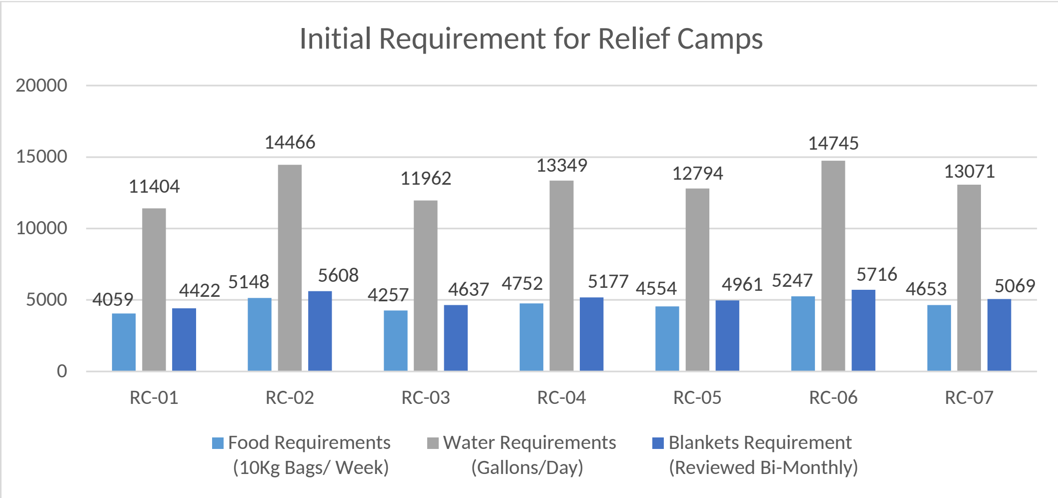

Figure 5.3: Initial resource requirements of population living in Relief Camps

5.2.2 Managing Food Consumption

Similar to the approach adopted for water resource management, the nutritional requirements for each of the seven relief camps in terms of 10kg bags supplied per week is evaluated. However, since the distribution schedule for food supplies is considered on a weekly basis, in order to have equivalent points of comparison in context of an increasing overall trend, the projection is determined for a period of twelve months with selected values graphically depicted in figure 5.6. Once again we proceed with the validation for our system for accurately predicting nutritional needs through the graphical representation of collective food requirements and predicted values. One key difference however, is that the baseline depicting the absolute minimum requirements threshold is set at 80% of the actual minimum requirement observed over the projection span since food can be rationed further in case of extreme shortages.

Based on the figure 5.7, the key points which can be derived are that even when the predictive value is least accurate in case of the second month – a maximum observed error of over 17%, the obtained value is still above the minimum defined threshold (80% of the minimum actual requirement). Furthermore, similar to the case observed for water supplies, the predicted values gradually converge towards actual values as time progresses with an average error percentage of less than 10%.

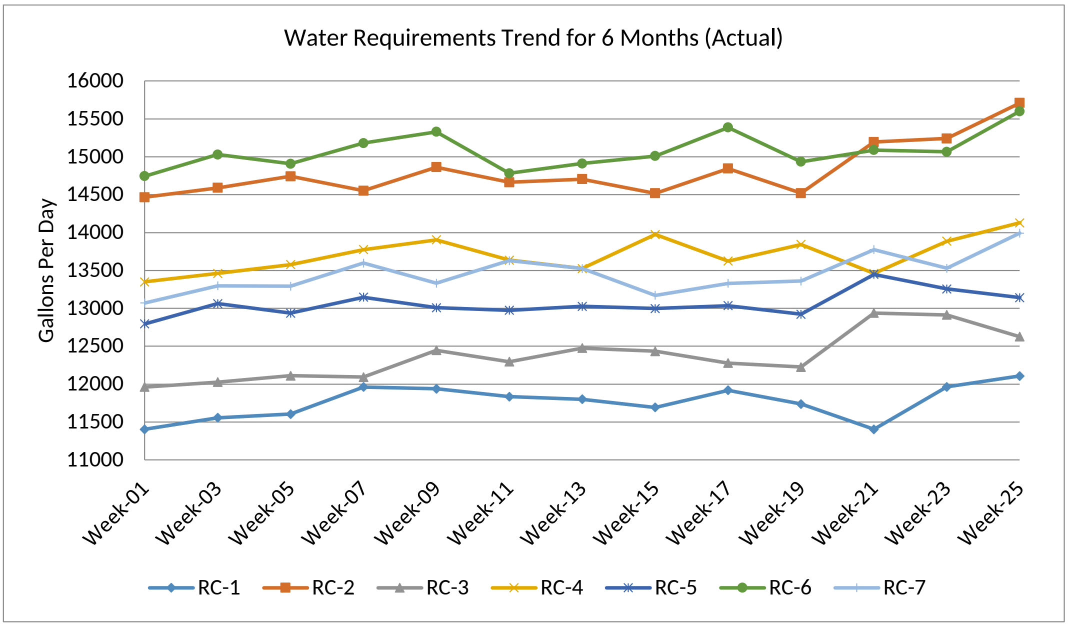

Figure 5.4: Water requirement trend for relief camps(six months projection)

5.2.3 Managing Shelter Items

The last case examined here is the demand for shelter items for communities living in relief camps. As opposed to food and water, the requirement for shelter is most significantly impacted by the population demographic. The demographic distribution which we have considered for our simulated relief camps as per commonly observed standards has been depicted earlier in figure 5.2.

Based on the abovementioned, and the fact that we are broadly considering shelter items in the form of tents and blankets, and that their distribution schedule is considered on a monthly basis; we have the following graphical depiction of actual requirements of tents and blankets for our relief camps projected over a two-year period in order to have equivalent points of comparison in context of an increasing overall trend.

Next, we proceed with the validation for our system for accurately predicting shelter needs through separate graphical representation of collective requirements for tenting and blankets, along with their respective predicted values. The baseline depicting the absolute minimum requirements threshold is once again set at 80% of the actual minimum requirement observed over the projection span of two years.

Based on the above-mentioned figures 5.10 and 5.11, it is clearly evident that obtained predictive values never fall below the minimum threshold, even

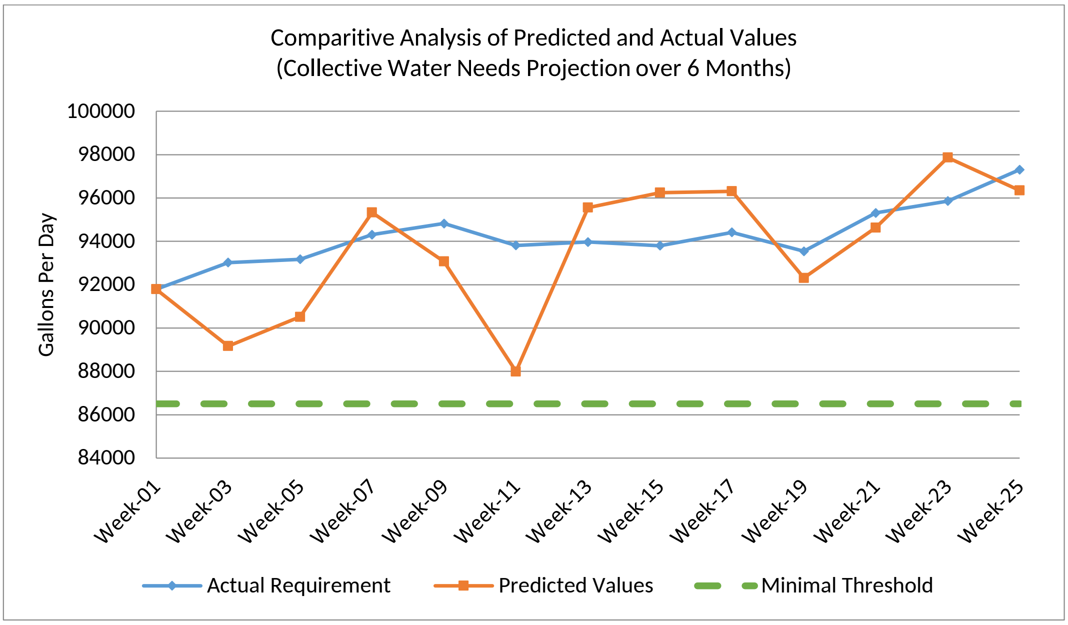

Figure 5.5: Comparison between actual and predicted water needs(six months projection)

during the initial validation phases when the error percentage is on the higher side – over 11% for the collective requirement of tents and almost 14% for the requirement of blankets. However, the error factor in both cases appears to converge with time as the system prediction accuracy improves to within an average error percentage of less than 10% in both scenarios.

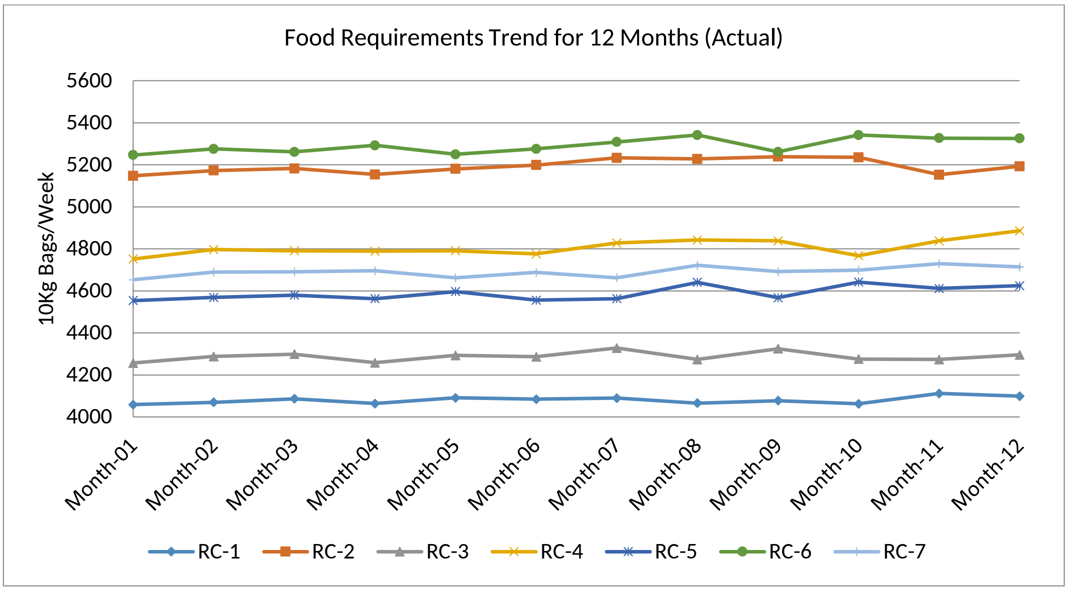

Figure 5.6: Food requirement trend for relief camps(12 months projection)

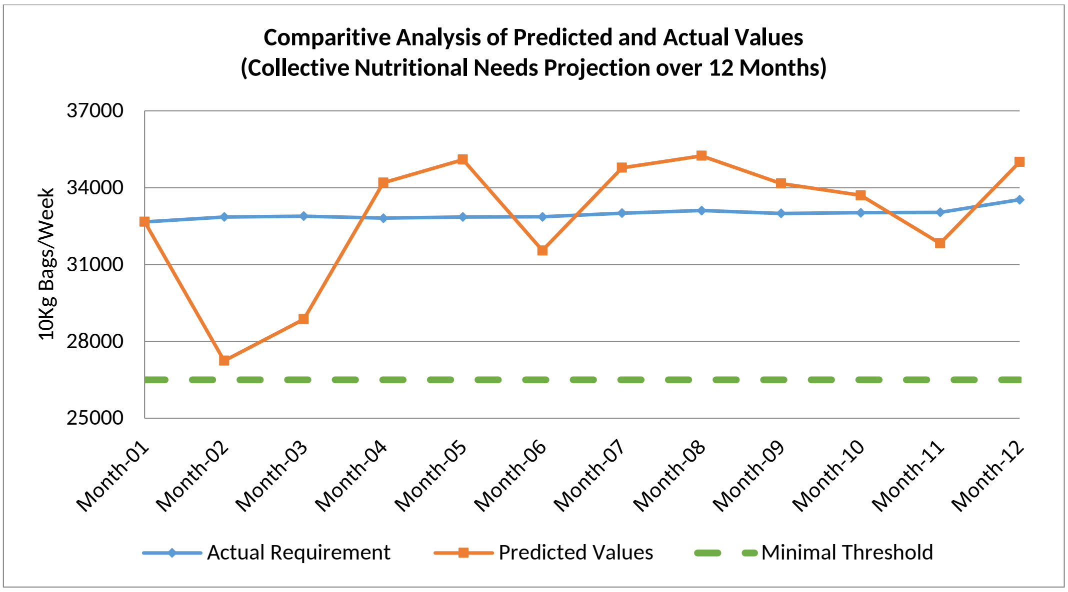

Figure 5.7: Comparison between actual and predicted nutritional needs(12 months projection)

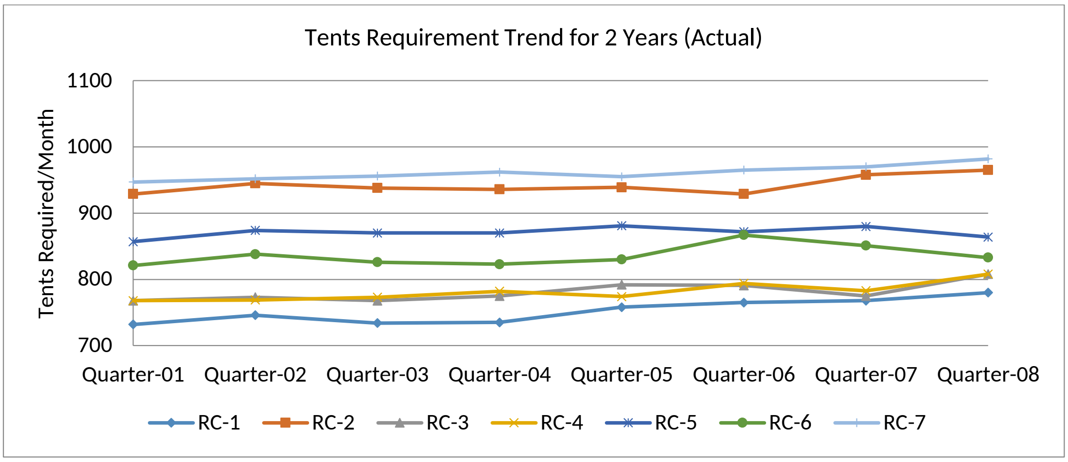

Figure 5.8: Tents requirement trend for relief camps (2 years projection)

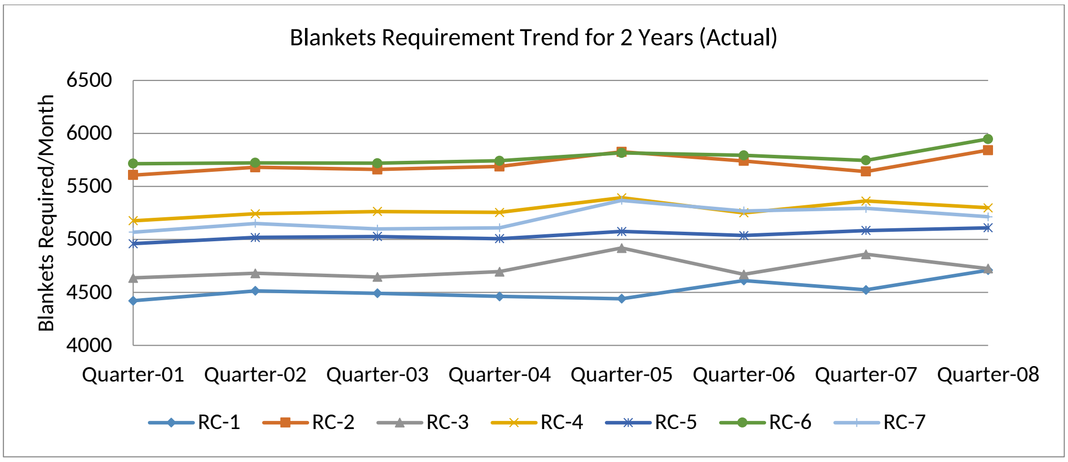

Figure 5.9: Blankets requirement trend for relief camps (2 years projection)

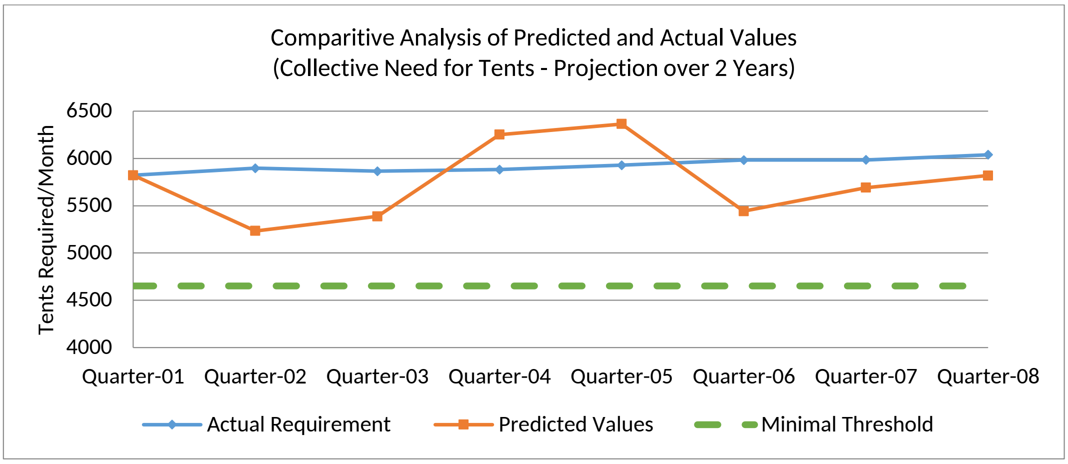

Figure 5.10: Comparison between actual and predicted tenting needs (2 years projection)

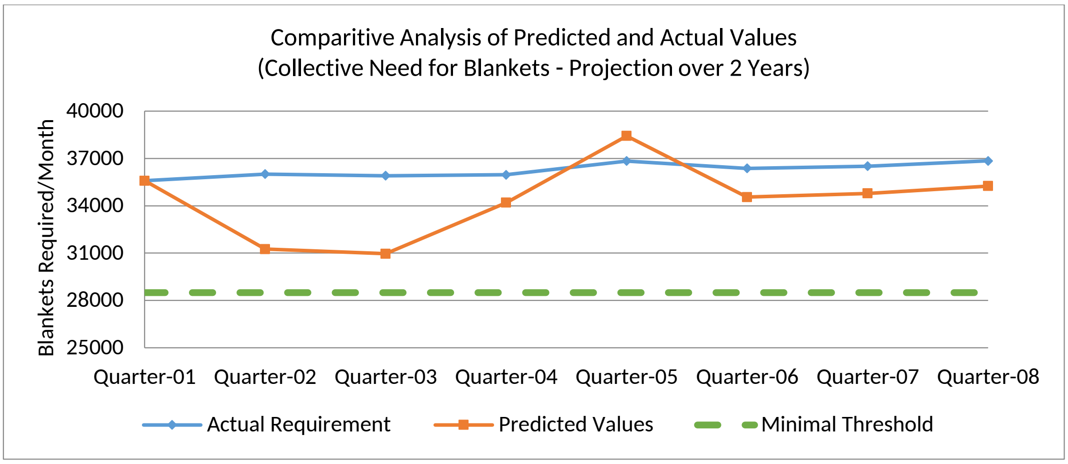

Figure 5.11: Comparison between actual and predicted need for blankets (2 years projection)

5.3 Summary

The preceding sections of this chapter provide a comprehensive overview of the results obtained after a detailed objective evaluation of our system. Essentially, a systematic approach was adopted to prove how selective data mining algorithms can be used for aiding effective decisions for relief camp management based on simulated real-time sensory inputs. Hence it was concluded that administrative aspects of relief camp management pertaining to the provisioning of food, water and items of shelter items can be intelligently managed through our developed system implemented over an instance of the Amazon EC2 cloud platform for increased accessibility, robustness, and scalability.

Chapter 6

Conclusion and Future Work

The following chapter goes on to summarize the work carried out over the course of this detailed research exercise. The draft presented in the preceding chapters is reviewed in the following sections and future directions of work are stated.

6.1 Conclusion

This research thesis deals with an often overlooked domain of long term relief camp management as a viable and novel use case for data mining and integration of real-time sensory information within a cloud framework. The term 'domain' is used here since relief camp management comprises of numerous social and administrative aspects. However, for this research endeavor, selective administrative aspects of relief camp management were considered in terms of timely meeting food, water and shelter needs of an affected community.

A simulation was constructed based on a comprehensive system design using the GAMA platform to depict various relief camps housing affected communities and a centralized warehousing facility from where relief goods disbursement would be managed. Next, software agents were assigned for real-time environment monitoring in terms of remotely monitoring consumption of food, water and shelter items in each relief camp. Separate agents were also used to oversee warehousing and supplies. All functions within the complex environments were in turn administered by a planning a prediction module which could timely identify potential resource shortages and minimize their impact on an affected population.

For prediction purposes, the Gaussian model of the Naïve Bayes Classifier was used to determine whether a particular relief camp would face scarcity of food, water and shelter items. Once the possibility of shortages was identified, linear regression methods were used to quantify future shortages by denoting them as a cost function which could be minimized using sum of least squares. The resulting information was fed back to the developed system for knowledge-driven decisions pertaining to effective distribution of limited relief resources. This was possible through the development of systematic algorithms which could be used to manage supplies as a routine function, as well as handle conflicts between relief camps competing for limited resources. Therefore, these novel algorithms help integrate useful information within the simulated environment; thus helping relief camp managers in making effective and timely decisions which can impact the lives of huge population living under stressful circumstances.

In order to address scalability concerns of the developed system, the use of a cloud platform such as Amazon EC2 was ideal because such platforms not only provide the necessary infrastructural support at a reasonable cost, but also help mitigate concerns regarding single-point-of-failure. This increased robustness is especially useful in our context because the system is designed to be used in prolonged emergency circumstances often encountered in long term relief camp management.

Finally, the system performance was evaluated through iterative testing procedures which showed a distinct decrease in the magnitude of resource shortages as well as the probability of shortage occurrence when smarter assignment of limited resources was carried out in our developed simulation.

6.2 FutureWork

The future directions of research are multidimensional in our case. Firstly, the system has been developed using Bayesian Classifiers and Linear Regression methods. Hence there is room to implement other supervised learning algorithms for predictive purposes. Furthermore, we have utilized the eager class of learning algorithms in our research exercises citing the need for quick decision making. A similarly detailed study can be carried out to prove whether slow or lazy learning algorithms perform equally well or impact the system performance in terms of more accurate predictions of resource shortages.

From the perspective of the use case explored in this research endeavor pertaining to long term relief camp management, we have only monitored limited administrative aspects from a broad vista – especially the factor concerning nutrition needs. While we have only considered general estimates based on international humanitarian standards, nutritional requirements can vary drastically for different populations and it is not necessary that commonly available food groups meet all deficiencies. In fact, a separate system can be developed having similar complexity just for monitoring nutrition needs.

Moreover, another major problem faced by communities forced to live in relief camps is managing the supply and demand of medicine. Like varying nutritional requirements, medicinal requirements can vary drastically for people affected by drought as opposed to people affected by floods.

Therefore, based on the abovementioned factors, there are numerous research directions which can be possibly pursued following the completion of this research exercise.

References

[1] R. Primmer, "Structured vs. unstructured data — robert primmer," 2017. [Online]. Available: http://www.robertprimmer.com/blog/structured-vs-unstructured.html

[2] P. Chandarana and M. Vijayalakshmi, "Big data analytics frameworks," in *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 international conference on*. IEEE, 2014, pp. 430–434.

[3] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014.

[4] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact." *MIS quarterly*, vol. 36, no. 4, 2012.

[5] M. Rychly *et al.*, "Scheduling decisions in stream processing on heterogeneous clusters," in *Complex, Intelligent and Software Intensive Systems (CISIS), 2014 Eighth International Conference on*. IEEE, 2014, pp. 614–619.

[6] D. Petrou, G. A. Gibson, and G. R. Ganger, "Scheduling speculative tasks in a compute farm," in *Proceedings of the 2005 ACM/IEEE conference on Supercomputing*. IEEE Computer Society, 2005, p. 37.

[7] C. Apte, B. Liu, E. P. Pednault, and P. Smyth, "Business applications of data mining," *Communications of the ACM*, vol. 45, no. 8, pp. 49–53, 2002.

[8] R.-S. Petre *et al.*, "Data mining in cloud computing," *Database Systems Journal*, vol. 3, no. 3, pp. 67–71, 2012.

[9] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[10] R. Vrbić, “Data mining and cloud computing,” *JITA-JOURNAL OF INFORMATION TECHNOLOGY AND APPLICATIONS*, vol. 4, no. 2, 2012.

[11] H. K. Mahdi, H. K. Mohamed, and S. S. Attia, “Data mining for decision making in multi-agent systems,” in *Multi-Agent Systems-Modeling, Interactions, Simulations and Case Studies*. InTech, 2011.

[12] L. Cao, V. Gorodetsky, and P. A. Mitkas, “Agent mining: The synergy of agents and data mining,” *IEEE Intelligent Systems*, vol. 24, no. 3, 2009.

[13] S. Thompson, “Community-based camp management,” *Forced Migration Review*, vol. 30, no. April, pp. 26–28, 2008.

[14] W. Lee, S. J. Stolfo, P. K. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop, and J. Zhang, “Real time data mining-based intrusion detection,” in *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX’01. Proceedings*, vol. 1. IEEE, 2001, pp. 89–100.

[15] N. R. Council, “Camp management toolkit,” 2004.

[16] P. Currion, C. d. Silva, and B. Van de Walle, “Open source software for disaster management,” *Communications of the ACM*, vol. 50, no. 3, pp. 61–65, 2007.

[17] J. T. B. Fajardo and C. M. Oppus, “A mobile disaster management system using the android technology,” *WSEAS Transactions on Communications*, vol. 9, no. 6, pp. 343–353, 2010.

[18] E. Rolland, R. A. Patterson, K. Ward, and B. Dodin, “Decision support for disaster management,” *Operations Management Research*, vol. 3, no. 1-2, pp. 68–79, 2010.

[19] M. Muaafa, A. L. Concho, and J. Ramirez-Marquez, “Emergency resource allocation for disaster response: an evolutionary approach,” 2014.

[20] X. Su, X. Zhang, W. Su, S. Yang, J. Huang, and Y. Chen, “The study on optimal model for relief resources allocation using dynamic programming and spatial analysis methods.”

[21] 2017. [Online]. Available: http://www.disaster-info.net/SUMA

[22] 2017. [Online]. Available: http://www.lssweb.net

[23] 2017. [Online]. Available: http://www.spherehandbook.org

[24] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive datasets*. Cambridge university press, 2014.

[25] C. K. Williams and D. Barber, "Bayesian classification with gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1342–1351, 1998.

[26] K. M. Leung, "Naive bayesian classifier," *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, 2007.

[27] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in neural information processing systems*, 2002, pp. 841–848.

[28] R. B. Darlington and A. F. Hayes, *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford Publications, 2016.

[29] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 936.

[30] 2017. [Online]. Available: https://github.com/gama-platform/gama/wiki

[31] 2017. [Online]. Available: https://aws.amazon.com/ec2/

[32] D. Talia, "Cloud computing and software agents: Towards cloud intelligent services." in *WOA*, vol. 11, 2011, pp. 2–6.