UNIVERSITY OF NAIROBI

School of Computing and Informatics

SCI FOURTH YEAR PROJECT

Name: **William Otieno Ochomo**

Reg No.: **P15/1720/2016**

Statistical Forecasting of Vaccine Demand

Supervisor: **Prof Robert Oboko**

Written in: **March 2019**

Submitted as partial fulfillment for the Degree in Bachelor of Science: Computer Science of the University of Nairobi

Declaration

I hereby declare that this project work is a record of an original work done by me. I affirm that this project has not been presented for any other university award.

NAME: **William Otieno Ochomo**

REGISTRATION NUMBER: **P15/1720/2016**

SIGNATURE:

DATE:

This project has been submitted as partial fulfilment of the requirements of the Bachelor of Science in Computer Science of the University of Nairobi and has been done with the guidance of my supervisor.

NAME: **Prof. Robert Oboko**

SIGNATURE:

DATE:

Dedication

I also dedicate it to my peers in school who have walked with me on every step of the way in research and development of this project.

Abstract

Immunization has made significant contribution to public health in Africa and the world; eliminating, eradicating and controlling life threatening diseases. The immunization coverage around the globe has risen over the years. However, there are challenges limiting the greater potential of coverage that could be achieved. Some of these challenges include issues of funding and resources for immunization, vaccine stock outs and logistics.

The forecasting of vaccine demand is important in ensuring adequate immunization supplies and the foundation of Vaccine Security. The accuracy of the forecast is also key as underestimating the requirements may result in vaccine shortages and overestimating results in excess stock which increases the manufacturers' costs.

The key focus of this project is vaccine stock outs and inequity of distribution of vaccine resources. The project seeks to identify key parameters and algorithms that can be used to forecast the demand for the vaccines. This is so that with the data on the demand for vaccines and its proper presentation, the necessary stakeholders can have helpful insights that can aid in proper resource allocation necessary for higher positive immunization impact. (Mihigo Richard, 2017)

Acknowledgement

I am grateful to God for the good health and strength he grants me to learn for providing priceless people who without their effort, collaboration and dedication, this project would not have been a success.

I wish to personally thank the following people for their contributions to my inspiration and knowledge and other help in working through this project: my supervisor Prof. Robert Oboko for his unwavering support and guidance even where I thought it was not going to work.

My parents for standing with me through my education and affording me the chance to pursue this degree.

Mr. Moses Magero for his close tutelage in my software development journey.

All my lecturers at the School of Computing and Informatics who helped me in many ways and made my education journey at the University of Nairobi pleasant and unforgettable.

Table of Contents

List of Figures

INTRODUCTION

Background

For an effective national immunization programme, the supply chain has to ensure continuous and uninterrupted availability of the appropriate vaccines to the point of vaccination. Interruption of vaccine availability for any reason results in missed opportunities to vaccinate and the population runs the risk of being exposed to the deadly preventable vaccines.(Lydon Patrick, 2017)

Patrick Lydon of the World Health Organization (WHO) and his colleagues analyzed the GVAP vaccine stock out indicator in 194 countries WHO Member States. (Managing supplies of vaccines is a huge problem, 2017). The findings were that the incidences of national-level stock outs is most pronounced in Sub Saharan Africa where 38% of the 47 countries in the region are affected. On the other, South East Asia and Western Pacific regions seem to be the least impacted by vaccine stock outs. Additionally, stock outs at national level occur in countries of all income groups as seen in the Fig 1 below. Both lower and upper middle income countries reported the majority of stock outs – 31% on average for each group between 2011 and 2015.(Lydon Patrick, 2017)

National-level stockouts – Percent of countries by region, income and population.

	2010	2011	2012	2013	2014	2015	Avg. 2011–15
Summary indicators							
No. countries reporting stockouts	67	66	57	54	50	65	58
% of countries	35%	34%	29%	28%	26%	34%	30%
No. of stockout events	153	148	120	112	110	113	121
Average duration of stockouts (days)	42	33	32	29	45	44	37
No. UNICEF procuring countries reporting stockouts	44	39	28	52	36	22	35
% of UNICEF procuring countries	49%	43%	31%	58%	40%	24%	39%
% countries by stockout frequency							
One stockout event	43%	39%	40%	41%	50%	60%	46%
Two stockout events	19%	24%	25%	31%	18%	20%	24%
Three stockout events	13%	19%	24%	13%	12%	9%	15%
Four or more stockout events	25%	18%	11%	15%	20%	11%	15%
% stockout events by WHO Region[a]							
Americas (AMR = 35)	19%	18%	16%	17%	32%	17%	20%
Sub-Saharan Africa (AFR = 47)	39%	38%	38%	46%	32%	34%	38%
East Mediterranean (EMR = 21)	7%	11%	5%	6%	6%	9%	7%
Europe (EUR = 53)	18%	17%	18%	7%	14%	26%	16%
South-East Asia (SEA = 11)	4%	2%	5%	7%	4%	5%	5%
Western Pacific (WPR = 27)	13%	14%	18%	17%	12%	9%	14%
% stockout events by Income[b]							
Low income (n = 31)	21%	23%	23%	26%	26%	20%	24%
Lower-Middle income (n = 50)	36%	29%	35%	41%	24%	28%	31%
Upper-Middle income (n = 58)	33%	32%	32%	29%	36%	27%	31%
High income (n = 55)	10%	16%	10%	4%	14%	25%	14%
% stockout events by Vaccine							
BCG	33%	28%	34%	33%	25%	34%	32%
DTP containing vaccine	47%	45%	42%	35%	40%	51%	42%
Measles containing vaccine	7%	14%	9%	14%	14%	5%	11%
Polio	13%	13%	15%	18%	21%	10%	15%

* Numbers in brackets indicate the total number of WHO Member States for each region respectively.
** According to the World Bank classification of countries. The middle income group includes countries that are classified as lower-middle income countries and upper-middle income countries. Numbers in bracket indicate the number of WHO Member States in each income group.

Figure 1: National-level stock outs - Percent of countries by region, income and population

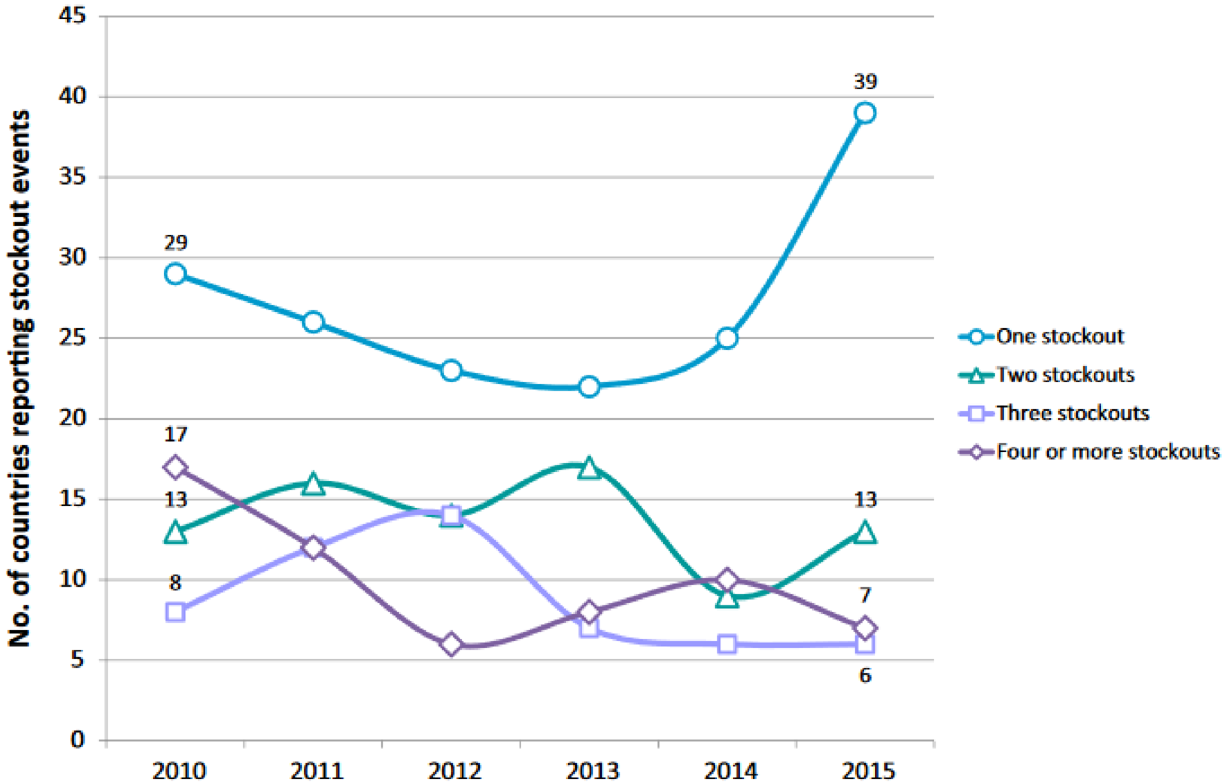


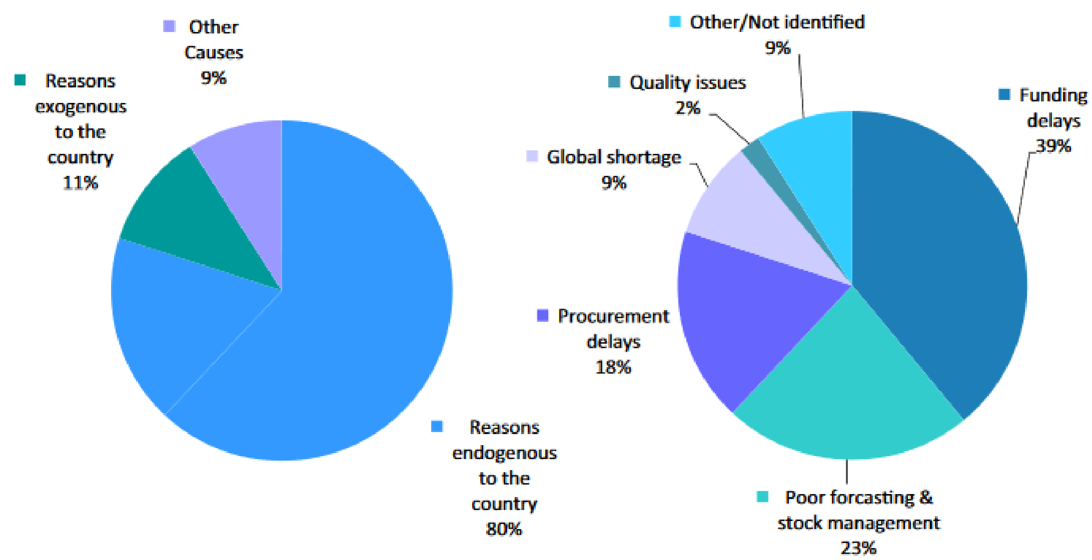
Figure 2: Frequency of stock outs

In countries reporting national-level stock outs, multiple events often occur within the year when one or more vaccines are affected. In some cases more than 4 stock out events are reported by countries for a given year (Fig. 2).

During the first half of GVAP, 43% of stock outs events concerned DTP-related vaccines. This proportion increased in 2015 when 51% of all reported stock out events included DTP-containing vaccines (often combined with HepB and Hib vaccine). The other vaccine most affected by stock outs is BCG representing another 31% of stock out events reported at national level. (Lydon Patrick, 2017)

An analysis of the UNICEF Supply Division information uncovered some of the causes of national level stock outs: 39% government funding delays, 18% delays in the procurement processes, 23% poor forecasting and stock management as seen in Figure 3 below.

With the great advancements in Machine Learning, it can be used to analyze historical supply chain data and public data to predict shifts in utilization and recommend corresponding stock levels and vaccine selection for the upcoming months. BCG and DPT being one of the most impacted by national level stock outs, this project intends to focus on child vaccines this is with the aim of maximizing childhood vaccination coverage and minimizing vaccine wastage.



Problem Statement

Child Vaccination is one of our SDG Health Goals. The immunization coverage has increased over the years. Technology has been used to improve credibility of data collected in remote areas for example use of USSD by ChanjoTrack in Nyandarua County. However, the positive efforts are drawn back by the stock outs due to improper planning by the government and also inequality in vaccine distributions because of lack of accurate information of where the vaccine demand is. Therefore combining the knowledge of Machine Learning, databases and statistics, the fairly credible data collected using systems like ChanjoTrack can be exploited to help magnify the areas with greater demand for vaccines in order to give them priority and also help governments and donors in informed planning when they are to disburse funds to purchase vaccines ensuring delivery of the right quantity of vaccines minimizing wastage.

Goals and Objectives

The aim of this project is to maximize childhood vaccination coverage by ensuring the various health facilities have adequate stock of child vaccines and minimizing vaccine wastage due to oversupplying or supplying to areas with less demand.

Research Objectives

1. To find out existing systems that also help in vaccine forecasting
2. To identify the various parameters that could be useful in predicting the demand for vaccines in a particular region
3. To identify the appropriate statistical and machine learning algorithms that can be used to forecast the demand for child vaccines
4. To find out the best way to present and visualize the information generated by machine learning algorithms

System Objectives

1. To clean the data gathered
2. To model the data gathered, identifying key features to enable a machine learning algorithm to make predictions of vaccine demand
3. Develop a prototype of a system for predicting vaccine demand
4. Test the prototype to establish accuracy
5. Visualize the output of the model
6. To develop a platform that various stakeholders; government, donors and manufactures can use to help in the decision making processes

Project Scope

This project will entail development of a web based platform which will entail the following:

1. Data Acquisition and Management Platform - it's main functionality is to enable users import bulk data, de-identify the data and restructure it in a way that can easily be represented in a relational database.
2. Data Processing Engine: this is where the raw data will be used to train and test algorithms that will learn to identify predictive patterns to forecast demand and recommend the optimal delivery of vaccines to the various regions.
3. Interactive User Interface: this will be a GIS interface which will provide a familiar and easy to understand and evaluate interface for use by the various stakeholders. The interface will also provide the users with the various graphs of vaccine data for the various locations and also enable exporting of the data and graphs

Project Justification

"For too many countries, it is nearly impossible to collect and harness data to forecast vaccine requirements and deliver vaccines when and where they're needed," said Jean Marie Okwo-Bele, MD, director of WHO Department of Immunization, Vaccines, and Biologicals, in the press release. "But there are innovations such as electronic data systems that are being piloted and scaled up in developing countries that could and should be more widely adopted." (Soucheray, 2017) (Singh, 2017)

It is evident from the previous sections that good decision making in any field depends on availability of credible information as a basis for the decision to be made. Vaccine stock outs have been seen to be a major challenge, affecting maximizing of childhood vaccination coverage. There is hope though, as said by the Director of WHO Department of Immunization, Vaccines, and Biologicals that with credible data collection systems being piloted such as ChanjoTrack, the data can be harnessed to help forecast vaccine requirements and deliver vaccines when and where they're needed helping to ensure minimal vaccine wastage and increase immunization coverage.

Additionally, a project by Mueller L.E, to evaluate the potential impact and value of applications (e.g. adjusting ordering levels, storage capacity, transportation capacity, distribution frequency) of data from demand forecasting systems implemented in a lower-income country's vaccine supply chain with different levels of population change to urban areas yielded the following conclusion that demand forecasting systems have the potential to greatly improve vaccine demand fulfilment, and decrease logistics cost/dose when implemented with storage and transportation increases. (Mueller Leslie E, 2016)

LITERATURE REVIEW

Vaccine Forecasting

Vaccine, Device and Cold Chain Forecasting is the first step in ensuring adequate immunization supplies and is the foundation of Vaccine Security. The accuracy of the forecast is important - underestimating the requirements results in vaccine shortages, overestimating results in excess stock - increasing the manufacturers' costs, which in turn increases the cost to UNICEF.

The goal of vaccine forecasting is to estimate the quantity of goods and financial needs necessary to conduct immunization programmes. The value of the forecast depends on the accuracy - taking into consideration the type of vaccine, the presentation (vial size), the quantity and the timing of delivery of the vaccine. Poor forecasting may result in delays or shortfalls in delivery, additional costs, and reduces UNICEF credibility with manufacturers. (Supplies and Logistics - Vaccine Forecasting, 2017)

Throughout the year the forecast demand requirements are monitored and followed up by Supply Division with each country. Suppliers are informed of changes to the forecast demand requirements on a monthly basis. Effective forecasting is not an event - it is an ongoing process.

Advantages of obtaining accurate forecasting of vaccine needs:

1. It leads to efficient management of vaccines and immunization sessions
2. It eliminates shortages or overstocking of vaccines
3. It improves vaccine use and reduction of wastages
4. It helps to monitor the progress of immunization in relation to target coverage

There are three methods commonly used to estimate vaccine needs:

1. Target Population
2. Previous Consumption
3. Size of immunization sessions

Target Population Method

The target population is the main consumer of the particular vaccine. For example, BCG the target population is the number of children under one year and women of bearing age (15-49 years old).

To estimate vaccine needs on the basis of target population a number of parameters are necessary which are:

a. Target Population

Identify and quantify the target to be covered in a health facility

b. Immunization Schedule

Immunization schedule determine the age limits and the number of doses required to be fully immunized among each target group (children under one and women of childbearing age).

Vaccines	Number of Doses
BCG	1
Polio	4
Pentavalent	3
PCV	3
Yellow Fever	1
Measles	1
Tetanus Toxoid for women of child bearing age (15-49 years)	5

c. Immunization Coverage Target

The national policy is to reach every child. The Immunization coverage target for each antigen is dependent on the health facility and district micro plans and work plans respectively. These plans indicate the attainable percentage coverage at the end of current year.

d. Wastage rate and wastage factor

During immunization, the number of vaccine doses used is generally higher than the number of individuals immunized. The number of doses in excess represents "lost doses "or vaccine wastage. These may include:

- The remainder of doses discarded with vials after the immunization session
- Doses given outside the target
- Doses spoilt for one reason or the other e.g. VVM reached discard point, breakdown in the cold chain, frozen DTP+ HepB and TT or removed labels.
- Doses from vials broken during transport and handling
- Missing doses from vaccine stock ledgers etc

Number of unopened vaccines vials lost should be documented in the ledger books to facilitate calculations of wastage rate and factor.

Vaccine wastage can be explained into two ways:

1. Wastage rate

Vaccine wastage rate should be taken into account in the estimation of vaccine needs. Knowing the wastage rates helps to determine the wastage factor, which is one of the parameters used to estimate vaccine needs.

Vaccine wastage rates are not standard. Every district and health facility must calculate its monthly vaccine wastage rates of antigens and by the end of year know their vaccine wastages, which would be used for estimation of the vaccines.

Formula for Wastage rate (%):

$$Wastage\ Rate\ (\%) = \frac{\text{Doses Used} - \text{ Doses administered} * 100}{\text{Doses used}}$$

Doses Used *include vaccines administered and wasted doses*

Doses Administered *are doses which have been received by the targeted group.*

2. Wastage factor

Vaccines Wastage Factor is a multiplier used to order vaccines to cater for the targeted population and wastage.

The total number of vaccines supplied within given period is referred to as 100% supply.

Formula for calculating wastage factor

$$Wastage\ Factor = \frac{100\%\text{ supply}}{100\%\text{ supply} - \text{Wastage rate}}$$

Calculating the vaccine needs for a Health Facility

$$\text{Total Doses} = Target\ Population * No\ of\ Doses\ * Target\ coverage * Wastage\ Factor$$

Note: *Target coverage for the health facility level is 100% this is in line in reaching every child in the catchment area. Therefore the target coverage is* ***1***

VACCINES FORECAST SHEET FOR YEAR _________2005____

LEVEL (Health facility/District) NAME:______________________

RIOMA____________

	BCG	OPV	DPT+HEPB - HIB	MEASLES	TT
1. ANNUAL NEEDS (DOSES) BASED ON TARGET POPULATION (ALL CHILDREN UNDER 12 MONTHS OF AGE; ALL CHILD BEARING AGE WOMEN)					
[A] Target population	14,000	14,000	14,000	14,000	84,000
[B] Doses in immunization schedule	1	4	3	1	3
[C] Wastage factor	1.33				
[D] Total dose required this year = (A x B x C)	18,620				

Figure 3: Sample Vaccine Forecast Sheet

Estimating vaccine needs on the basis of previous consumption

The method of estimating vaccines needs based on previous vaccines consumption consists of calculating the quantity of vaccines consumed during the previous period. The resulting quantity is thereafter adjusted, for instance when there is increase in the population for the current period by 10%.

This method is based on reliable stocks management data. It is suitable therefore for use in health facilities sharing the catchment area and where the stock management is good but there is insufficient information on immunization objectives and targets for the implementation of the immunization session.

The data required for estimating vaccines needs on the basis of previous consumption are:

- Number of children immunized previously
- Wastage factor for the specific antigen
- Immunization schedule for the antigen.

After calculating the total estimated doses an additional 10% of the total doses is added to cater for unexpected increase in population.

Some Machine Learning Algorithm currently in use in the prediction of vaccine demand are:

1. Auto-Regressive Integrated Moving Average (ARIMA)

Box & Jenkins (1970) have proposed a methodology to forecast a variable using as a database only it's past and present. These models are very popular due to:

- The quality of the generated forecasts;
- The flexibility of the models;
- The rigor of the mathematical base of the model;
- The fact that this is an appropriate method for predicting variables with an irregular trend, too.

An autoregressive-moving average ARMA (*p,q*) model has an autoregressive type component or an average moving type component:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \ldots - b_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

Where *p* is the order of the autoregressive part, *q* is the moving average order and is a white noise type process (a sequence of independent and identically distributed random variables with zero mean).

To elaborate such models the following reasons are advised:

- The evolution of the economic phenomena is fueled by the existing resources, the already created capacities, the experience and the tradition. Variables in economy have an inertial character, a strong autoregressive component being presented (mainly in macroeconomic indicators);
- Moving average type component is the effect of unpredictable events on the variable, effects gradually assimilated in time. This component is justified by the intervention of sudden unexpected changes among the external factors correlated to the variable. For example, strikes effect in Greece during the last two years led to the deterioration of the tourism activity indicators of the country. The moving average part captures the gradual assimilation of shocks (accidental deviations) from outside the system.

The ARMA models are suitable for stationary series. These were generalized for non-stationary series that become stationary by differentiation; the resulting models are called autoregressive integrated-moving average ARIMA (*p,d,q*) where d is the order of differentiation required for stationary series.

Stages (the methodology) to develop an ARIMA (*p,d,q*) model are:

- Identifying the model → specifying the appropriate values for *p, d* or *q*;
- Estimation of model parameters → estimating the coefficients a_i, b_i;
- Testing the validity of the model. If the model is not valid then the model is once more specified (other plausible values for *p,d,q*) and the previous stages are repeated;
- Using the model to generate predictions (after it passed the validation tests).

2. Human-in-the-loop Machine Learning (HITL)

Evidence suggests that a variant of Pareto's famous 80:20 leads to some of the most accurate machine learning systems to date, with 80% computer AI-driven, 19% human input, and 1 % unknown randomness to balance things out.

Human input can come in two different forms.

- Helping label the original dataset that will be fed into a machine learning model
- Helping correct inaccurate predictions that arise as the system goes live.

Similar Systems and Research Work

1. A Forecasting Model for Deciding Annual Vaccine Demand

This is a research paper by Ruey Kei Chiu that presents a computer-based forecast model for building a decision support system for forecasting the annual vaccine demand of a specific vaccine. The model is formatted by employing a combination technique including the Neural Network and Auto-Regressive Integrated Moving Average. The result generated from the system were to be taken by the Taiwanese governmental immunization authority to make a better decision for budgeting and purchasing the annual requirement of specific vaccines.

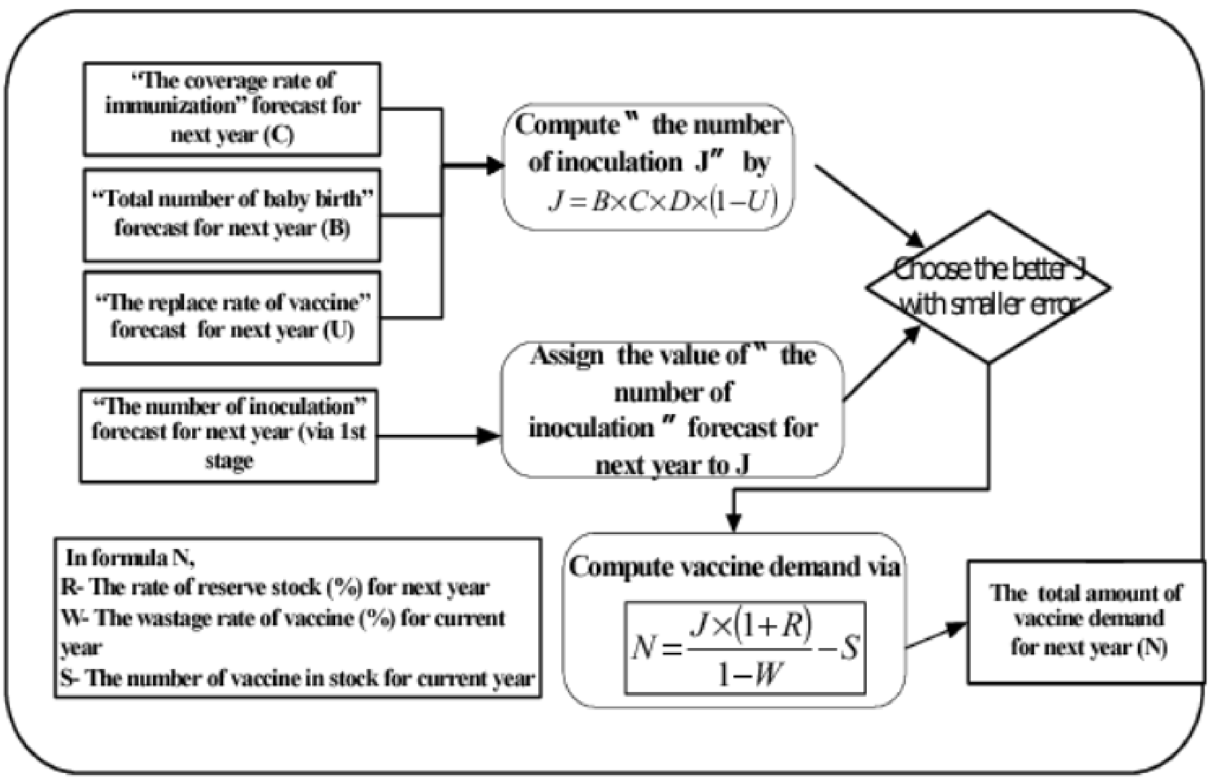


Figure 4: Model for computing vaccine demand

	Decisive Variables	Explanations and Data Sources
Vaccine demand relative variables	Total number of inoculation	If also means workload of vaccine inoculation. The value is obtained from NIIS database.
	Immunization coverage rate	The average coverage rate in last 3 years for MMR vaccine released by CDC immunization office is different from year to year..
	Vaccine wastage rate	The value is thought to be constant. Because there is no related records in NIIS database, so we adopt 25% as the general wastage rate of MMR by WHO's reports.
	The number of vaccine in stock	The value is the surplus stock in the end of every year.
	Vaccine reserved stock	The value is thought to be constant. The percentage of vaccine reserved stock is set to be 25% by CDC decision.
	The number of doses per fully immunized child	This value is thought to be constant. A new-born baby of Taiwan must be inoculated MMR total four doses which come to 2, 4, 6, 18 months.
	The price of single/multiple dose	Because the price of vaccine is different from every year, so the prices are set to purchase MMR from year to year.
Population growth	The number of baby birth	The value may be obtained form ministry of

Figure 5: Variables used to predict vaccine demand

2. How to deliver the right vaccine, in the right amount, at the right time.

Macro-eyes was awarded Global Grand Challenges funding from the Bill & Melinda Gates Foundation and USAID to design and pilot the predictive supply chain for vaccines in East Africa.

Macro-eyes is working with PATH, ministries of health, and district leadership. Macro-eyes machine learning technology is analyzing historical supply chain data, bio-surveillance data, and public data to predict shifts in utilization and recommend corresponding stock levels and vaccine selection for the upcoming months.

Macro-eyes will deploy human-in-the-loop machine learning to engage front line caregivers and gather information from critical health workers. Health worker insight on populations and demand, conveyed via text message, will programmatically augment the analysis of supply chain and immunization data.

Direct engagement increases the accuracy and precision of prediction, can resolve gaps in the data, and empowers champions at the point of care. The predictive supply chain for health will constitute one of the first deployments of machine learning – specifically human-in-the-loop machine learning – for global health.

Health supply chains are increasing in complexity and scale. Global coverage for basic childhood vaccines has reached a record 86%, but there has been a parallel increase in vaccine wastage, decreasing resource efficiency.

Vaccine stock-outs compound the problem by wasting opportunities for immunization that can resonate throughout a community. Macro-eyes technology creating the predictive supply chain for vaccines is breaking the link between higher rates of immunization and increased wastage.

Summary of Literature Review

The systems mentioned above are both not implemented in Kenya. The former is implemented in Taiwan while the Macro-Eyes system is yet to be piloted in Tanzania. This leaves Kenya having no such similar system, a void that can be filled with my proposed solution.

Macro-Eyes is yet to set up SMS systems to be used to collect the data which will be used by their model to predict vaccine supply. In Kenya, there is a USSD based system that has already been piloted in Nyandarua County called ChanjoTrack which readily provides some of the data that can be used in the prediction.

More research is yet to be carried out on the advantages of using HITL in prediction and the comparison with ARIMA and Back Propagation which is stated as the model intended to be used in the Taiwanese Research Paper. However ANN has been seen to provide more accurate results than the ARIMA.

METHODOLOGY AND APPROACH

Introduction

This is the study of the current system so as to come up with the right requirements for the proposed system and to understand how the proposed system should operate in its environment. Also it involves the study of the chances the proposed system has of success.

Methodology

This project will be carried out using Dynamic System Development Methodology (DSDM). The execution of the project will involve segmenting of data into learning and validation datasets to develop highly accurate prediction algorithms. Once the algorithms have been developed, they are applied to the full data set to do the prediction. The idea is that one should be able to perform these classifications without human intervention, and the methods should also be able to operate on very large data sets and be very fast.

System Development Methodology

The methodology that will be used in this project will be the Dynamic System Development Methodology (DSDM). DSDM is an agile project delivery framework, primarily used as a software development method. It is a framework which embodies much of the current knowledge about project management (Dsdmofagilemethodology.wikidot.com, 2015). DSDM is rooted in the software development community, but the convergence of software development, process engineering and hence business development projects has changed the DSDM framework to become a general framework for complex problem solving tasks. The DSDM framework can be implemented for agile and traditional development processes.

Why use DSDM?

- Results of development are directly and promptly visible
- Since the users are actively involved in the development of the system, they are more likely to embrace it and take it on.
- Basic functionality is delivered quickly, with more functionality being delivered at regular intervals.
- Eliminates bureaucracy and breaks down the communication barrier between interested parties.
- Because of constant feedback from the users, the system being developed is more likely to meet the need it was commissioned for.
- Early indicators of whether project will work or not, rather than a nasty surprise halfway through the development
- System is delivered on time and on budget.
- Ability of the users to affect the project's direction.

The DSDM Process

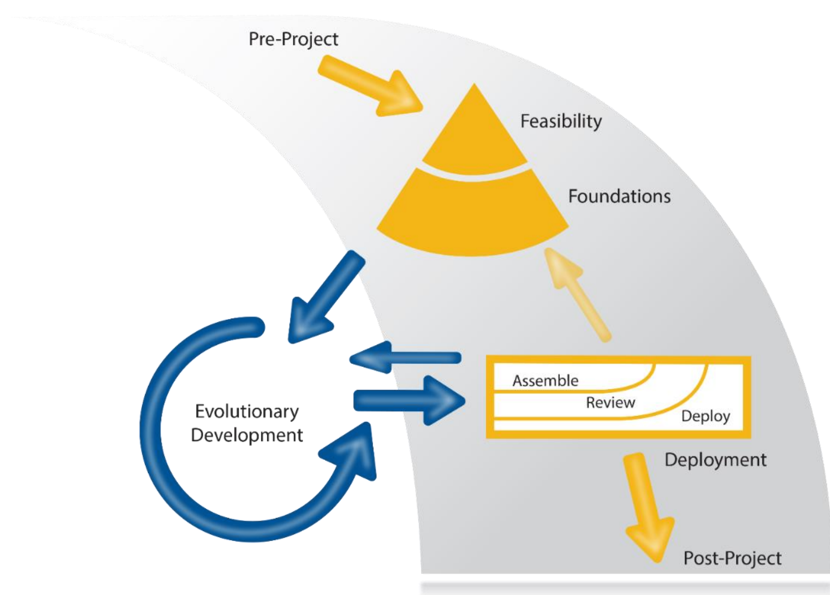


Figure 6: The DSDM Process

The DSDM process model comprises a framework which shows the DSDM phases and how they relate to one another. The phases are:

1. Pre-Project Phase

The Pre-Project phase ensures that only the right projects are started, and that they are set up correctly, based on a clearly defined objective.

2. Feasibility Phase

The Feasibility phase is intended primarily to establish whether the proposed project is likely to be feasible from a technical perspective and whether it appears cost-effective from a business perspective.

3. Foundations Phase

It is intended to establish a fundamental (but not detailed) understanding of the business rationale for the project, the potential solution that will be created by the project, and how development and delivery of the solution will be managed.

4. Evolutionary Development Phase

The purpose of the Evolutionary Development phase is to evolve the solution.

5. Deployment Phase

The objective of the Deployment phase is to bring a baseline of the Evolving Solution into operational use. The release that is deployed may be the final solution, or a subset of the final solution.

6. Post-Project Phase

After the final Deployment for a project, the Post-Project phase checks how well the expected business benefits have been met. Although it may be possible to highlight immediate benefits, most benefits will accrue over a pre-defined period of live use of the solution.

Project Activity Plan

Item	Start Date	End Date	Hours	Deliverables
Refining Requirements	15/1/2019	17/1/2019	10	Requirement specification document
Data acquisition and preparation	18/1/2019	21/1/2019	15	De-identified data in CSV format
Data Modeling	22/1/2019	29/1/2019	35	Database Schema
System and User Experience Design	28/1/2019	4/2/2019	40	System design documents
Data Importation	4/2/2019	11/2/2019	20	Scripts to import data
Data Mining Algorithms Implementation	28/2/2019	3/4/2019	150	Data Mining algorithms code
User Interface Design Implementation	3/4/2019	15/5/2019	225	User Interface Web Pages.
Data Import API	15/5/2019	21/5/2019	35	Data Import API
System Testing and Evaluation	23/5/2019	8/6/2019	75	Test and Evaluation Documents
Deployment	8/6/2019	9/6/2019	8	Setup server and Working URL address

SYSTEM ANALYSIS

Introduction

System analysis refers to the process of examining the proposed system to understand its constraints, procedures and functions. This involves gathering of information of how the system can be built to suit user needs and also defining the system requirements.

Feasibility Study

Operational Feasibility

It is the measure of how well the solution will work and also of how people feel about the proposed system.

Stakeholder meetings have also generated a lot of interest in the success of the project. Stakeholder are interested in seeing how their data is used and what further insights they can gain from the analysis of the data. This guarantees the incorporation of the portal into business activities. This was the case with ChanjoTrack.

Since the portal will be web-based, the reach and usability of the system is higher as no installation needs to be done, most of the people are already familiar with accessing of services through the web and the presence of capable web browsers.

Technical Feasibility

It is the measure of the practicality of the proposed system with regard to availability of technical resources and expertise. It is based on analysis of the hardware and physical requirements of the system and other characteristics such as network availability and strength, prevailing mobile technology etc. The technical assessment helped answer the question such as whether the technology needed for the system existed, how difficult it would be to build, and whether I had enough experience using that technology. This project was built on Python. This takes advantage of its richness in analytics and machine learning tools. Additionally, Django framework is very reputable for its convenience in building GIS Applications courtesy of the GeoDjango package. All these technologies are free and robust, making the system technologically feasible.

Schedule Feasibility

This is the measure of how reasonable the timeline for the proposed system is.

The project is attached to a very strict timeline of the next three months. The project is achievable within this timeline though against some challenges listed below.

1. Refinement of user stories. The data in the DSL warehouse is contributed by many stakeholder in the health sector. They have their own interpretation of how the data should be used. Some of these are not clear from the beginning, rather they are redefined after every iteration of development. Changes proposed to the system may cause time overheads, and this may slow down development.

2. Incomplete data. This is especially true concerning data that is relatively old. Some of the systems at that time did not capture data that is presently being captured now. This creates a scenario where data that is needed for processing is incomplete. Some of the missing data can be extracted or computed though not all. The system has to make tradeoffs between completeness and accuracy.

3. Cleaning of data.

Factoring in these challenges, the project will still be able to be done in the time allocated because of presence of presence of computing minimums.

Economic Feasibility

This is the measure of the cost-effectiveness of the proposed system in terms of building the system and the returns after the system is complete and functional.

Software tools used in this project were open source thus free. Benefits of the new system would therefore outweigh the cost of developing it. Upon completion, the system was hosted on Digital Ocean Cloud Servers, which are cost effective. A sizeable server space costs 10 USD per month

Information Gathering

This is the formal process of using various techniques to collect information about problems, requirements and preferences about the proposed system. There are various approaches used depending on the scope of the system and the sample space from which data is collected from. The information gathering method selected for this project include the following.

Stakeholder Meetings

Meetings were conducted with the stakeholders in ChanjoTrack and they were really open to the idea as Vaccine Demand Forecasting adds value to their system. They also shared on the important data required to do the forecasting and how they would make the data available for use by the system.

Research

A lot of research was carried out on the subject matter. I was able to identify the various parameters used to calculate vaccine demand by Ministry of Health for Health Facilities. I was also able to identify some similar systems, their approach to the problem and their system architecture.

Requirements Specification

Analysis of the information gathered is required to find out the system requirements. This is to clearly define what the proposed system must do. The functional and non-functional requirements were defined based on the results of the information gathering.

Functional Requirements

The functional projects provided for this project describe the tasks that the system must do. In the case of this system, the requirements revolved around the gathering, analysis and representation of data. These requirements are tailored to meet the needs of the users described primarily in the use-cases described further.

The functional requirements include:

1. The system should be able to pull data from DHIS 2 and ChanjoTrack. This is as it forms the foundations of the analysis and communication.
2. The system should be able to preprocess the data collected making it portent for analysis.
3. A stakeholder should be able to view visualizations of the data before and after processing.
4. A stakeholder should be able to view forecasts of vaccine demands.

Non-Functional Requirements

These requirements provide guidelines on how the functional requirements will be met. They ensure not just that the system will work, it will meet the standards and the needs of the users

The non-functional requirements include:

1. Responsive. It cannot be assumed that the system will be accessed from a desktop or a machine with a large screen real estate. The system should cater for people accessing it from phones as will. The system should follow a mobile-first approach so as to create a system that works on all kinds of devices.

2. Load Time. Despite pulling and analyzing a lot of data, this is not meant to overwhelm the machine of the user. The system should provide information or respond to requests in a reasonable amount of time. This may involve using hardware that can provide the necessary computing power or tweaking the software to provide and analyze data faster.

3. Availability. The system should always be accessible at any time and from any point. This means putting in place infrastructure overheads to ensure availability such as server mirroring, load balancing and failover configuration.

4. Caching. This is meant to handle availability and load time at the same time. The system should be able to keep a local copy of the results of the data analysis locally so as to make it easier to get results without having to reprocess the data again. The results that are cashed should be prioritized to ensure that only the most requested for results are cached. This prevents the system from wasting storage space with data that is not frequently being accessed.

5. Versatile. The system should allow the user to intervene at any stage they feel so. The system should allow the use to define the kind of data that he wants to see. The system should not impose its perspective of the data on the user, he should feel free to generate his own data.

6. Progressive Web application. Progressive web app describe an application that is robust regardless of network failure. The application should not consume a large amount of data. Client side caching will help reduce the need to renew data when the pages are refreshed. Progressive web apps are also able to handle offline situations. Lack of network or good weak signals should not prevent the entire app from functioning, the app should be able to provide functionality albeit limited. Progressive web apps makes the application more practical to use and also crates the perception of high availability.

System Requirements

The system is broken into to applications, server-side and client-side applications. For smooth operations of both of these applications, the following hardware and software requirements need to be met.

Server-side application

It is required that the server application be very powerful so as to handle the processing of large amounts of data. The system is constantly acquiring large data sets from different systems, the size complexity of the data processing is always increasing. The system has to keep up with this growth in size while still performing efficiently

Hardware requirements

- Intel Core i-7 or higher
- 4 GB RAM or higher
- 1 Tb Hard disk drive

Software requirements

- Ubuntu operating system 18.04
- PostgreSQL database management system

- Django web framework 1.11
- GeoDjango framework and Leaflets Js. for GIS
- Pandas data modeling tools
- Numpy data analysis tool
- Python programming language
- Nginx web framework

Client-side application

The client applications are not loaded with a lot of requirements because it is desirable for the system to be accessible by any machine within the market.

Hardware requirements

- Intel core i3 or higher
- Any ARM or Mediatek mobile processor
- 1 Gb ram or higher

Software requirements

- Any operating systems
- Internet Explorer 10 or higher
- Es6 compatible browsers
- JavaScript enabled browsers

Use Case Diagram

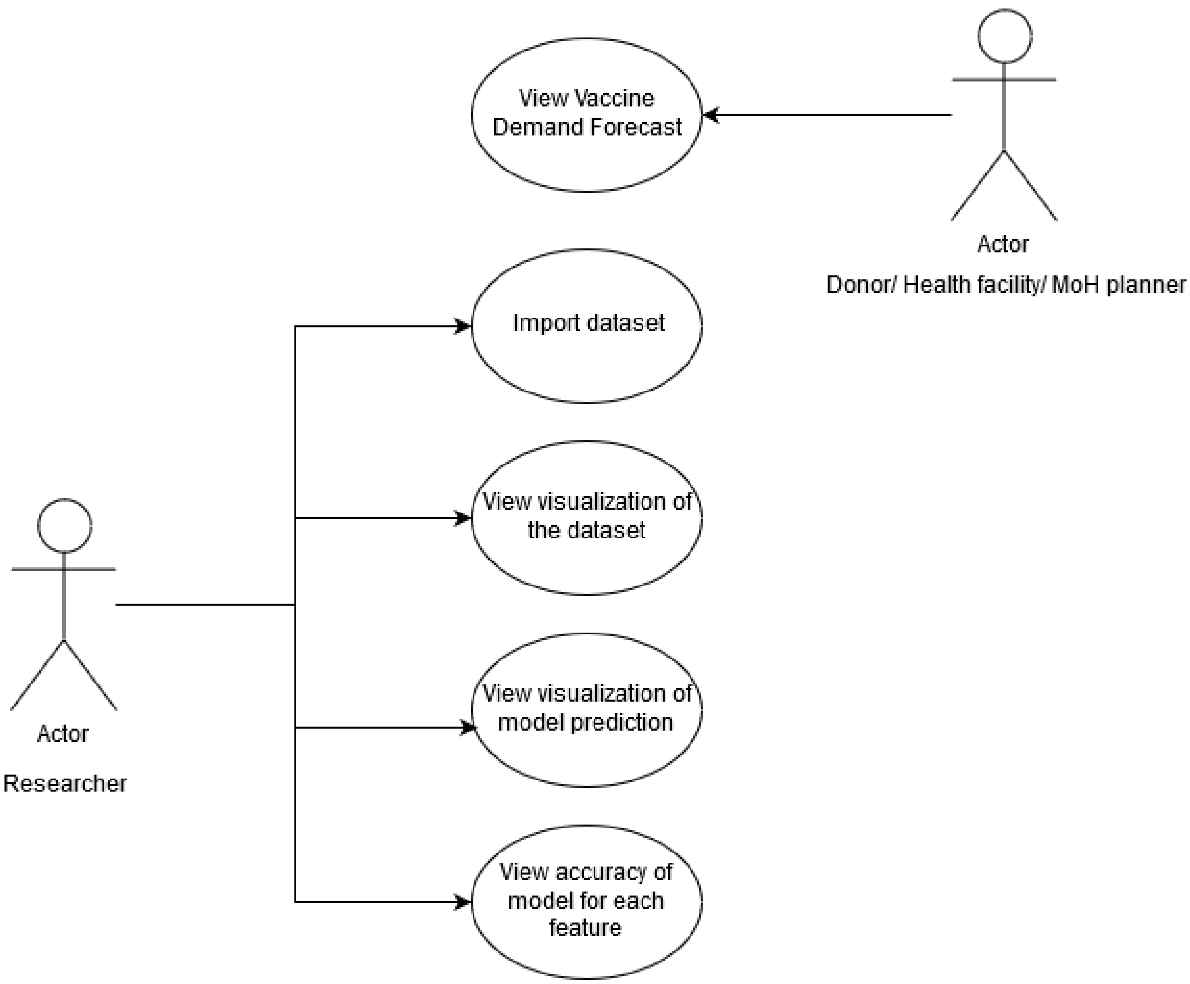


Figure 7: Use Case Diagram

Data Flow Diagram

Context Diagram

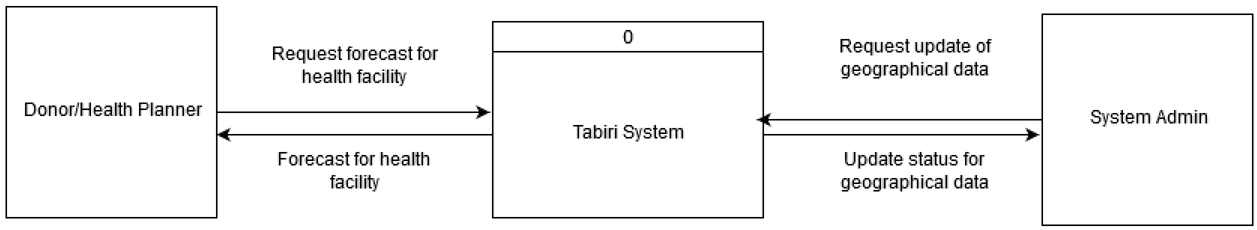


Figure 8: Context Diagram

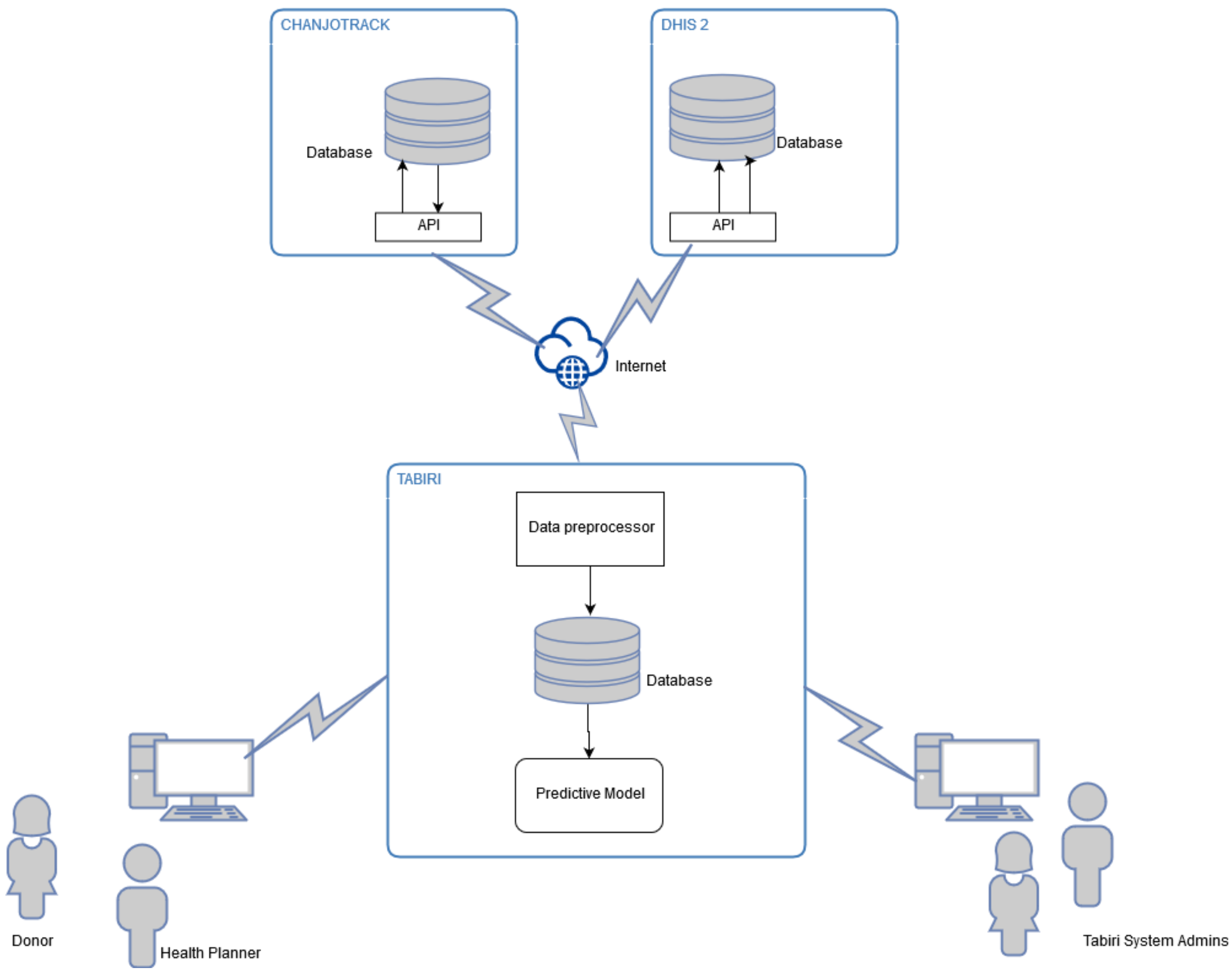
SYSTEM DESIGN

Introduction

System Design involves the conversion of the function models from System Analysis into models that represent the actual solution to the problem being solved.

From the functional requirements and non-functional requirements yielded from the analysis phase, designs for the system were made

The Conceptual Model



System Flowchart

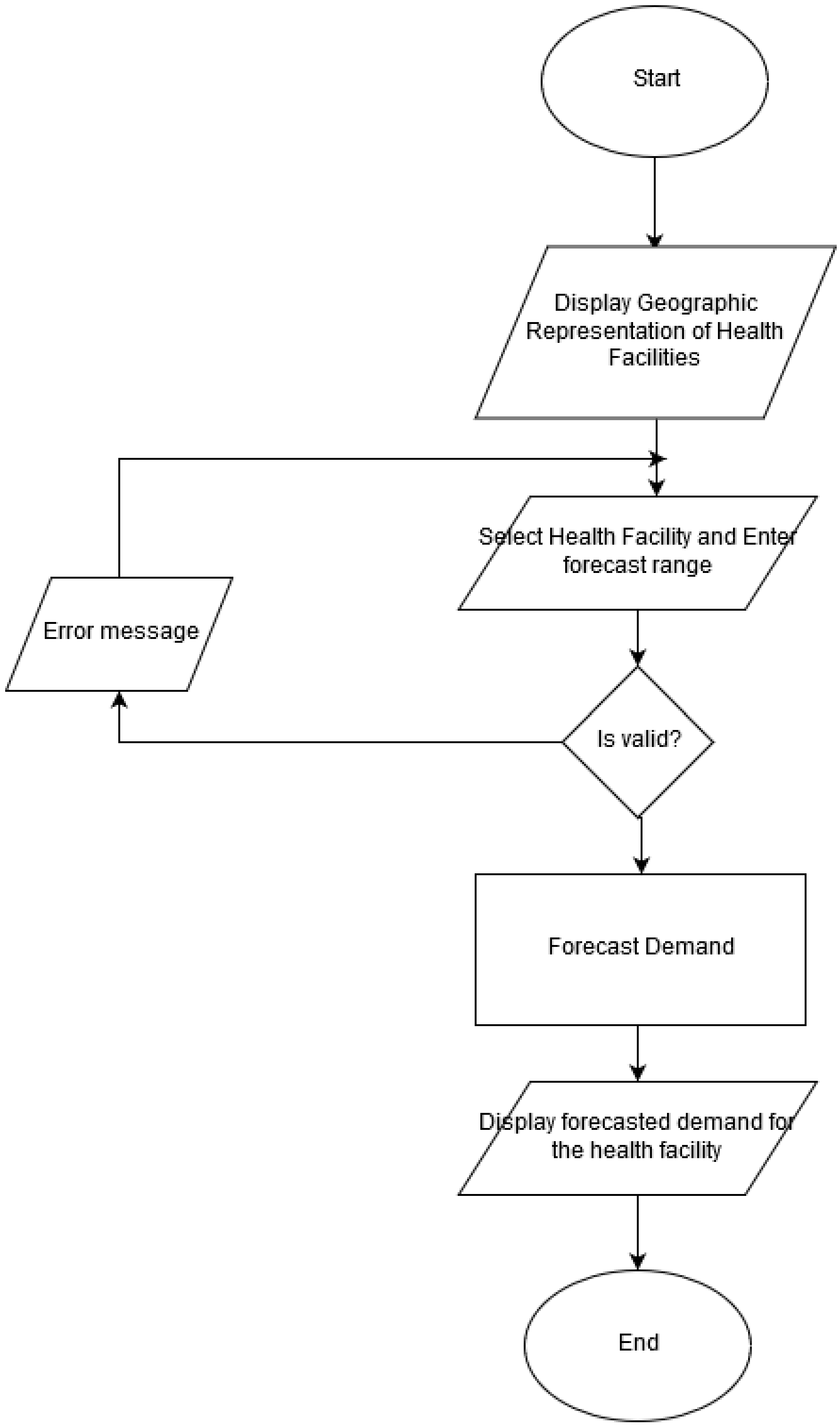


Figure 9: Donor or Planner Flowchart

User Interface Design

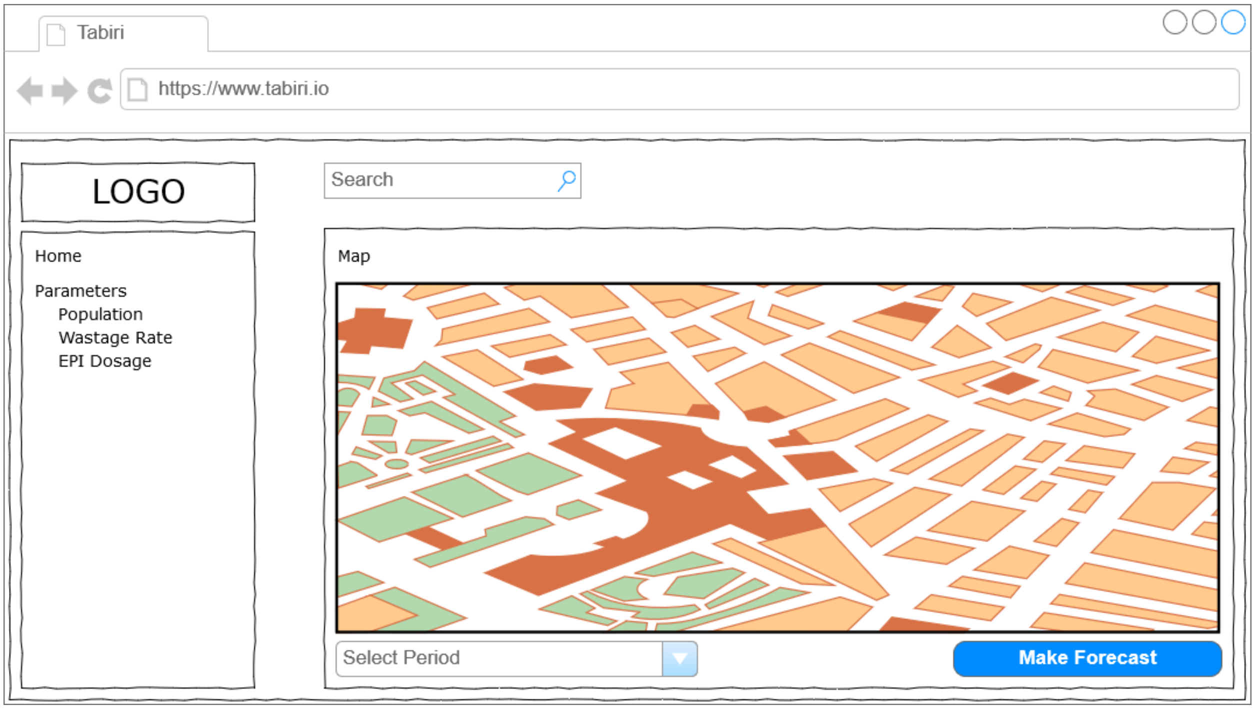


Figure 10: Home Page

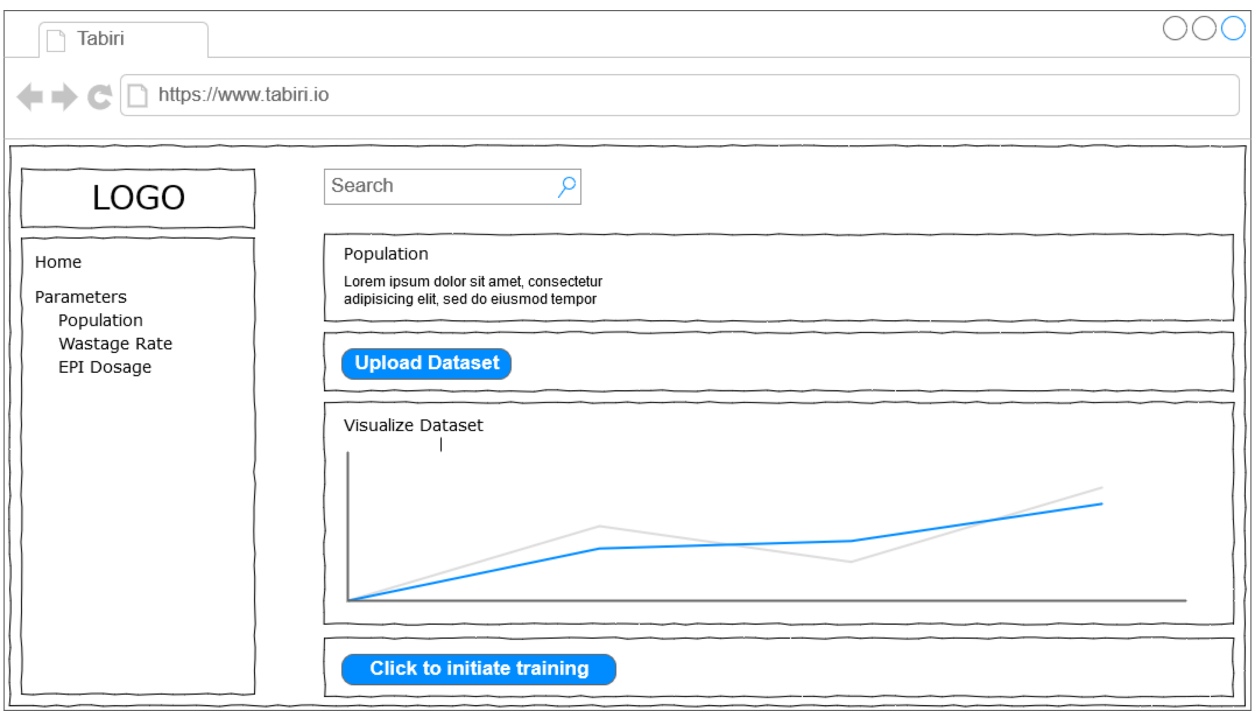


Figure 11: Population Page

References

Lydon Patrick, S. B. (2017). Vaccine stockouts around the world: Are essential vaccines always available when needed? *Vaccine, 35*, 2121-2126.

Managing supplies of vaccines is a huge problem. (2017, April 1st). Retrieved from The Economist: https://www.economist.com/science-and-technology/2017/04/01/managing-supplies-of-vaccines-is-a-huge-problem

Mihigo Richard, O. J. (2017, 6 21). Challenges of immunization in the African Region. *The Pan African medical journal, 27*(Suppl 3), 12.

Mueller Leslie E, H. L. (2016). The impact of implementing a demand forecasting system into a low-income country's supply chain. *Vaccine*, 3663-3669.

Singh, P. (2017, March 29). *Global immunization impact constrained by outdated vaccine delivery systems, researchers say* . Retrieved from EurekAlert! The Global Source of Science News: https://eurekalert.org/pub_releases/2017-03/b-gii032317.php

Soucheray, S. (2017, March 30). *Center for Infectious Disease Research and Policy (CIDRAP)*. Retrieved from Freezing, shortages among global vaccine challenges: http://www.cidrap.umn.edu/news-perspective/2017/03/freezing-shortages-among-global-vaccine-challenges

Supplies and Logistics - Vaccine Forecasting. (2017, July 27). Retrieved from UNICEF: For Every Child: https://www.unicef.org/supply/index_55506.html